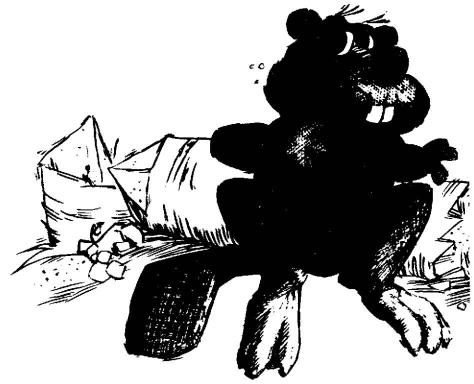
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



The Decidability of The DTOL Sequence Equivalence Problem and Related Decision Problems

Karel Culik II
Juhani Karhumäki

CS-85-05

May, 1985

The Decidability of the DTOL Sequence Equivalence Problem and Related Decision Problems*

Karel Culik II

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

Juhani Karhumäki†

Department of Mathematics
University of Turku
Turku, Finland

ABSTRACT

We show that it is decidable whether or not two given morphisms agree word by word on a given DTOL language. Hence, a nontrivial generalization of the famous DOL sequence equivalence problem, namely DTOL sequence equivalence problem is decidable. We also show that our main decidability result holds for some larger families of (morphically defined) languages.

Key words: formal languages, decision problems, DTOL systems, morphic equivalence.

CR Categories: 5.23, 5.27

---

* This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant A-7403.

† This work has been done during the second author's visit at the University of Waterloo.

1. Introduction

One of the most interesting problems in the theory of formal languages in the 1970's was the DOL sequence equivalence problem, i.e., the problem of finding an algorithm to decide, whether or not for a given word w and for two morphsims h and g the equation $h^n(w) = g^n(w)$ holds for all $n \geq 0$. Not only the problem itself and its ultimate solution, but also techniques developed to attack this problem as well as new problems encountered turned out to be of a crucial importance, cf. [3], [12].

We recall that it was the DOL sequence equivalence problem which created the notion of an equality language of two morphisms and subsequently lead to many interesting representation results of language families, see e.g. [4], [9], and [21]. The problem of morphic equivalence on languages, a very natural equivalence problem of (deterministic) translations, was defined in [7] and has been studied since that in many papers, see e.g. [2] and [13]. Finally, the origin of the famous and important Ehrenfeucht Conjecture, see [12], seems to have a connection to the DOL sequence equivalence problem.

The DOL sequence equivalence problem has been solved in [5]. The algorithm given in [5] and also the one given later in [8] are based on the "bounded balance" property of two equivalent DOL sequences. This property does not hold for HDOL sequences, hence these proofs cannot be generalized to HDOL sequence equivalence problem, i.e. to the problem of deciding whether or not for a word w and for three morphisms h, g, and f the equation $f(h^n(w)) = f(g^n(w))$ holds for all $n \geq 0$. Neither can they be modified for the DTOL sequence equivalence problem (defined in Section 2). Only in the special case when morphisms are defined in a binary alphabet an (actually optimal) algorithm for the DOL sequence equivalence problem is known such that it can be generalized for HDOL and DTOL cases as well, see [11]. A surprising connection between the DOL sequence equivalence problem and the Ehrenfeucht Conjecture was found in [6], where it was shown that if the Ehrenfeucht Conjecture holds true even noneffectively for all DOL languages, then the DOL sequence equivalence problem is decidable. Moreover, this approach can be generalized to solve HDOL sequence equivalence problem, too. Quite recently, it was shown in [19] according to these lines that HDOL sequence equivalence problem is indeed decidable.

The goal of this paper is to show that this approach actually gives a solution to the DTOL sequence equivalence problem as well. First, contrary to the DOL case, the DTOL sequence equivalence problem is equivalent, as was observed already in [7], to the problem of deciding whether two morphisms agree word by word on a given DTOL language. Secondly, the arguments of [6] can be generalized to show that this latter problem is decidable providing the Ehrenfeucht Conjecture holds for DTOL language. In [6] it has also been shown that the Ehrenfeucht Conjecture can be translated into the compactness problem for systems of equations over a free monoid. Recently [1] used this interpretation to prove that the Ehrenfeucht Conjecture does hold.

Our approach is not restricted to DTOL languages only. Our Theorem 4 gives a considerably larger family of languages for which the morphic equivalence problem is decidable. Finally, in Section 4 we discuss about the possibilities (and the difficulties) of generalizing our results for finite substitutions instead of morphisms.

2. Preliminaries

We assume that the reader is familiar with the basic notions of family languages, see e.g. [10], or in the case of L systems [20]. Consequently, the following lines are mainly to fix the terminology as well as to state the problems.

The basic object of our study is the morphism from a free monoid into itself (or if clearer into another free monoid). We say that two morphisms h and g of Σ^* are *equivalent* or *agree* on a language L, in symbols $h \stackrel{L}{=} g$, if the equality $h(x) = g(x)$ holds for every word x in L. We also say that a word x is *morphically forced* by a language L if for any two morphisms h and g the equality $h(x) = g(x)$ holds whenever $h \stackrel{L}{=} g$.

Let L be an arbitrary language over Σ. We say that a finite subset F of L is a *test set* (for morphic equivalence) of L if whenever two morphisms agree on F they agree on L as well, or in other words, L is morphically forced by F. The *Ehrenfeucht Conjecture* states that each language possesses a test set.

In this paper we are mainly dealing with languages generated in a "morphic way", the simplest being so-called DOL languages. A *DOL system* G is a triple (Σ, h, w), where Σ is a finite alphabet, h is a morphism of Σ^* and w is a nonempty word of Σ^*. A DOL system (Σ, h, w) defines the language $L(G) = \{h^n(w) \mid n \geq 0\}$ and the sequence $E(G) = w, h(w), h^2(w), \cdots$. Languages and sequences thus defined are called *DOL languages* and *DOL sequences*. An *HDOL sequence* (resp. *HDOL language*) is obtained from a DOL sequence (resp. DOL language) by applying another morphism to that sequence (resp. language). Finally, a *DTOL system* is a $(k+2)$-tuple $(\Sigma, h_1, \ldots, h_k, w)$ where, for each $i = 1, \ldots, k$, (Σ, h_i, w) is a DOL system. A DTOL system defines in a natural way a complete n-ary tree (called a *DTOL tree* or *sequence*) shown in Figure 1.

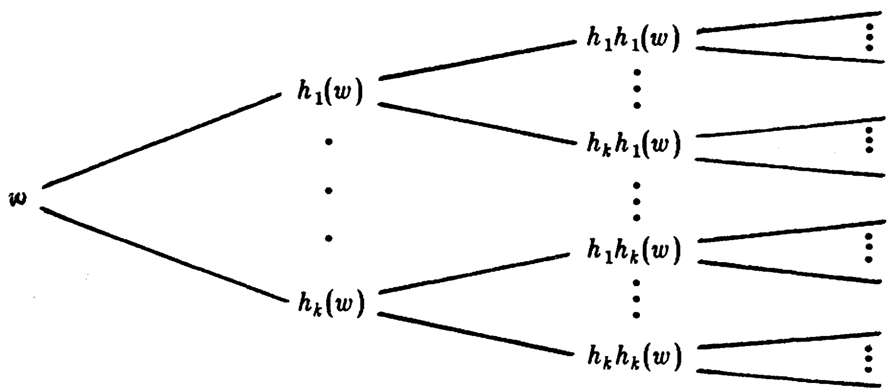


Figure 1

The set of all nodes of this tree forms a *DTOL language*. *HDTOL sequences* and *languages* are defined as morphic images of DTOL sequences and languages.

Obviously a DTOL language is a generalization of a DOL language. We still enlarge the language family as follows. For a language L and a morphism h we call the language $\{h^n(x) \mid x \in L, n \geq 0\}$ *h^*-closure of L*. If instead of one morphism h we use k morphisms $h_1, \ldots, h_k$ we obtain a $(h_1, \ldots, h_k)^*$-*closure* of L: $\{h_{i_1} \cdots h_{i_n}(x) \mid x \in L$ and, for $n \geq 0$, $i_1, \ldots, i_n \in \{1, \ldots, k\}\}$. Clearly, each DOL language is h^*-closure of a singleton set. Now let L be a family of languages. *DOL-closure* of L, in symbols $H^*(L)$, is the family of all h^*-closures of languages in L. Similarly, the *DTOL-closure* of L, in symbols $H^*_{fin}(L)$, is the family of all languages obtained as $(h_1, \ldots, h_k)^*$-closures of L in L, for some $k \geq 1$. Clearly, for any language family L, $L \subsetneq H^*(L) \subsetneq H^*_{fin}(L)$. The inclusions are proper for example for L equal to the family of finite sets.

Now we are ready to define our basic problems:

Problem 1. The DOL (resp. HDOL, DTOL, HDTOL) *sequence equivalence problem* is the problem of deciding whether or not two given DOL (resp. HDOL, DTOL, HDTOL) sequences coincide.

Problem 2. The *morphic equivalence problem* for a family $\mathcal{L}$ of languages is the problem of deciding, given a language L in $\mathcal{L}$ and two morphisms h and g, whether or not h and g are equivalent on L, i.e., whether or not $h \overset{L}{=} g$ holds.

Let $H = (\Sigma, h, w)$ and $G = (\Sigma, g, w)$ be DOL systems. Clearly, they define the same sequence, i.e., are equivalent, if and only if the morphism h and g are equivalent on a DOL language generated by H (or by G). Consequently, our Problems 1 and 2 are closely related in the case of DOL systems. They are also closely related to the equations of free monoids. Indeed, two morphisms h and g agree on a word, say aab, if and only if the quadruple $(h(a), h(b), g(a), g(b))$ is a solution of the equation $xxy = uuv$. This leads us to consider the systems of equations over a finitely generated free monoid.

Let Σ be a finite alphabet and N another finite set disjoint from Σ. The *equation* over Σ^* with unknowns N is a pair $(u,v) \in (\Sigma \cup N)^* \times (\Sigma \cup N)^*$, usually written as $u = v$. A *system of equations* is any collection of equations. A *solution* of a system S of equations is a morphism $h : (\Sigma \cup N)^* \to \Sigma^*$ such that $h(a) = a$ for all a in Σ and $h(u) = h(v)$ for (u,v) in S. Since $h(a) = (a)$ for a in Σ any solution can be identified with an n-tuple from $(\Sigma^*)^n$, where n denotes the cardinality of N. Finally, we say that two systems of equations are *equivalent* if they have exactly the same solutions.

3. Results

In this section we show that our Problem 2 is decidable for quite a large family of languages, and that this implies that all the variations of Problem 1 are decidable as well.

We start with a result in [6]:

Theorem 1. *The equivalence problem for finite systems of equations with a finite number of unknowns is decidable.*

The proof of Theorem 1 is based on a deep decidability result of Makanin, see [15], stating that it is decidable whether a given equation possesses a solution, and on an observation that the nonequality $u \neq v$ can be stated in the form $u_1 = v_1 \vee \cdots \vee u_n = v_n$ for some finite n.

As an immediate consequence of Theorem 1 we obtain the following very useful auxiliary result.

Theorem 2. *For two finite languages L_1 and L_2 with $L_1 \subseteq L_2$, it is decidable whether or not L_1 is a test set for L_2.*

Another important result for us is the following very interesting and deep recent result, see [1]:

Theorem 3. *The Ehrenfeucht Conjecture holds true for all languages.*

Using the above theorems we are able to prove our main result:

Theorem 4. *Let L be a family of languages satisfying the following two* *conditions: (i) $\mathcal{L}$ is effectively closed under union and morphisms, (ii) for each language in $\mathcal{L}$ there effectively exists a test set. Then for each language in $H^*_{fin}(\mathcal{L})$ there also effectively exists a test set. Consequently, the morphic equivalence problem for $H^*_{fin}(\mathcal{L})$ is decidable.*

Proof: Clearly, the last sentence of the theorem follows from the previous one. Let us fix $H = \{h_1, \ldots, h_k\}$ to be a set of morphisms and L a language in $\mathcal{L}$. We define

$$L_0 = L$$

$$L_{i+1} = h_1(L_i) \cup \cdots \cup h_k(L_i) \cup L_i\,, \quad \text{for } i \geq 0,$$

and further

$$L' = \bigcup_{i \geq 0} L_i.$$

We have to show that L' possesses a test set and that such can be found effectively.

Now, by (i), each L_i is in $\mathcal{L}$ and hence, by (ii), for each $i \geq 0$ a test set F_i for L_i can be effectively found. We set

$$\hat{F}_i = \bigcup_{j \leq i} F_j$$

and claim that, for some i_0, all words of $\hat{F}_{i_0+1}$ are morphically forced by $\hat{F}_{i_0}$. Indeed, if this were not the case, then the infinite language

$$\bigcup_{i \geq 0} \hat{F}_i$$

would not possess a test set, a contradiction with Theorem 3. Since each $\hat{F}_i$ can be found effectively, it follows from Theorem 2 that the above mentioned $\hat{F}_{i_0}$ can be found effectively, too.

We claim that $\hat{F}_{i_0}$ is a test set for L'.

To prove this claim it is enough to show that all words of L' are morphically forced by $\hat{F}_{i_0}$. Assume that this is not the case, and let j_0 be the smallest integer such that there exists in L_{j_0} a word, say x, such that it is not morphically forced by $\hat{F}_{i_0}$. This means that there exist morphisms g and f such that

$$g(x) \neq f(x) \text{ with } x \text{ in } L_{j_0}, \text{ and } g(y) = f(y) \text{ for all } y \text{ in } \hat{F}_{i_0}. \qquad (1)$$

Since $\hat{F}_{i_0}$ is a test set for L_{i_0}, we must have $j_0 > i_0$. Further from the minimality of j_0 it follows that (1) can actually be rewritten as

$$g(x) \neq f(x) \text{ with } x \text{ in} L_{j_0}, \text{ and } g(y) = f(y) \text{ for all } y \text{ in } L_{j_0-1}. \qquad (2)$$

Now, let $m = j_0 - i_0 - 1$ and $h_{i_1}, \ldots, h_{i_m}$ be a sequence of morphisms in H such that

$$x = h_{i_1} \cdots h_{i_m}(x') \text{ for some } x' \text{ in } L_{i_0+1} \qquad (3)$$

Let further

$$g' = gh_{i_1} \cdots h_{i_m} \text{ and } f' = fh_{i_1} \cdots h_{i_m}.$$

Then we conclude from (2) and (3) that

$$g'(z) = f'(z) \text{ for all } z \text{ in } L_{i_0}$$

and

$$g'(x') = gh_{i_1} \cdots h_{i_m}(x') = g(x) \neq f(x) = fh_{i_1} \cdots h_{i_m}(x') = g'(x')$$

with x' in L_{i_0+1}. These relations mean that all words of L_{i_0+1} are not morphically forced by L_{i_0}, which, in turn, implies that neither are all words of $\hat{F}_{i_0+1}$ morphically forced by $\hat{F}_{i_0}$. This last implication follows since $\hat{F}_i$ is a test set for L_i for all i.

So we derived a contradiction with the choice of i_0, and hence the theorem holds. □

Theorem 4 has several interesting corollaries. First of all the assumptions (i) and (ii) are certainly satisfied by the family ***Fin*** of finite languages. Hence, from the fact that $H^*_{fin}(\mathbf{Fin})$ equals the family of all DTOL languages we obtain

Corollary 1. *The morphic equivalence problem for the family of DTOL languages is decidable.*

Corollary 1 implies, as was already noticed in [7], the following important result.

Corollary 2. *The DTOL sequence equivalence problem is decidable.*

Actually, the Corollary 2 holds for morphic images of DTOL sequences as well. Hence, we have obtained as a special case the decidability of HDOL sequence equivalence problem which was for a long time open, until it was solved quite recently in [19]. We also want to emphasize that the only special case when Corollary 2 was known to hold was the simple case of binary alphabet, see [11]. It is also interesting to note that the DTOL language equivalence problem is undecidable, see [18].

Corollary 1 can be stated in a stronger form. Taking L equal to the family of DTOL languages and applying Theorem 4 iteratively we obtain

Corollary 3. *Each HDTOL language possesses effectively a test set.*

As another example we set $L = CF$, the family of context-free languages. Clearly, this family satisfies the condition of (i) and the rather difficult result that it also satisfies (ii) has been shown in [AKC]. Hence, we have the following

Corollary 4. *Each language in $H^*_{fin}(CF)$ has effectively a test set.*

Observe that this result is a proper strengthening of Corollary 3. Indeed, $H^*_{fin}(CF)$ is a proper superfamily of the families of context-free and HDTOL languages, cf. [20].

It was shown in [17] that the family of supports of F-rational formal power series, with F a field, satisfies the condition (ii) of Theorem 4, for definitions see [22]. Since it also satisfies the condition (i) we can use this family as our starting family in Theorem 4 and derive a new larger family of languages for which we know that each language of this family possesses effectively a test set. Recall that the family of supports of F-rational formal power series is incomparable with the family of context-free languages and contains, e.g., the language $\{a^n b^m \mid n = m^2\}$.

4. Concluding Remarks and Open Problems

We have proved that the problem of deciding whether or not two given morphisms h and g are equivalent word by word on a given language L is decidable for quite a large class of languages. An interesting question is in which extent, if of all, these results can be generalized to finite substitutions, which are natural generalizations of morphisms.

Of course, the essential difference between morphisms and finite substitutions is that the former are deterministic while the latter are nondeterministic. This nondeterminism seems to make the problem very difficult to attack. Indeed, we were even not able to show that it is decidable whether or not two finite substitutions τ and δ are equivalent on a given regular language L, i.e., whether or not the equality $\tau(x) = \delta(x)$ holds for all x in L. However, we believe that this is the case:

Conjecture 1. It is decidable whether or not two finite substitutions are equivalent word by word one given regular language.

To emphasize that the above conjecture is probably much more difficult than it seems we state

Theorem 5. *Given two finite substitutions τ and δ and two regular languages R and L, it is undecidable whether or not $\tau(x) \cap R = \delta(x) \cap R$ holds for all x in L.*

Theorem 5 is a straightforward consequence of the result stating that it is undecidable whether many-valued mappings defined by inverses of two finite substitutions are equivalent word by word on a given regular language, see [16] as well as Theorem 4.1 in [14].

The reason why we believe that Conjecture 1 holds is that, it is difficult — even more difficult than in the case of morphisms — to make two finite substitutions to agree on a given word. In other words, if they do agree they force a "lot of periodicity". Based on this intuition we think that not only Conjecture 1 but also much stronger conjecture holds true:

Conjecture 2. For each language L over a finite alphabet there exists a finite subset F such that whenever two finite substitutions agree word by word on F, they agree on L as well.

Of course, the above conjecture is the extension of the Ehrenfeucht

Conjecture to finite substitutions. A weaker form of the conjecture is obtained if only such substitutions τ are considered that satisfy the property that the cardinality of $\tau(a)$ for all letters a is uniformly bounded by a fixed constant.

Although we have little support for the validity of Conjecture 2, we do have a connection between Conjectures 1 and 2.

Theorem 6: *If the Conjecture 2 holds (even noneffectively and in the weaker form) then it is decidable whether or not two finite substitutions agree word by word on a given HDTOL language.*

Idea of Proof: The proof of Theorem 6 is a suitable adaptation of that of Theorem 4 (or, in fact, Corollary 3). □

Since each regular language is an HDTOL language, cf. [20], we have the following.

Corollary 5. *Conjecture 2 implies Conjecture 1.*

Clearly, our proof of Corollary 5 does not require the validity of the (weaker) Conjecture 2 for all languages. Its validity for the family of HDTOL languages would be sufficient, however not just for regular sets.

REFERENCES

[1] Albert, M.H., and Lawrence, J., A Proof of Ehrenfeucht's Conjecture, manuscript (1985).

[2] Albert, J., Culik II, K., and Karhumäki, J., Test Sets for Context-Free Languages and Algebraic Systems of Equations in Free Monoids, *Inform. Control 52* (1982), 172-186.

[3] Culik II, K., Homomorphisms: Decidability Equality and Test Sets, in: R. Book (ed.), *Formal Language Theory, Perspectives and Open Problems,* Academic Press, New York, 1980.

[4] Culik II, K., A Purely Homomorphic Characterization of Recursively Enumerable Sets, *J. Assoc. Comput. Mach. 26* (1979), 345-350.

[5] Culik II, K., and Fris, I., The Decidability of the Equivalence Problem for DOL Systems, *Inform. Control 35* (1977), 20-39.

[6] Culik II, K., and Karhumäki, J., Systems of Equations over a Free Monoid and Ehrenfeucht's Conjecture *Discrete Math. 43* (1983), 139-153.

[7] Culik II, K., and Salomaa, A., On the Decidability of Homomorphism Equivalence for Languages, *J. Comput. System Sci. 17* (1978), 163-175.

[8] Ehrenfeucht, A., and Rozenberg, G., Elementary Homomorphisms and a Solution to the DOL Sequence Equivalence Problem, *Theoret. Comput. Sci. 7* (1978), 169-183.

[9] Engelfriet, J., and Rozenberg, G., Fixed Point Languages, Equality Languages and Representation of Recursively Enumerable Languages, *J. Assoc. Comput. Mach. 27* (1980), 499-518.

[10] Harrison, M., *Introduction to Formal Language Theory,* Addison-Wesley, Reading, MA, 1978.

[11] Karhumäki, J., On the Equivalence Problem for Binary DOL Systems, *Inform. Control 50* (1981), 276-284.

[12] Karhumäki, J., The Ehrenfeucht Conjecture: A Compactness Claim for Finitely Generated Free Monoids, *Theoret. Comput. Sci. 29* (1984), 285-308.

[13] Karhumäki, J., and Kleijn, H.C.M., On the Equivalence of Compositions of Morphisms and Inverse Morphisms on Regular languages, *R.A.I.R.O.*

Theoret. Inform., to appear.

[14] Karhumäki, J., and Wood, D., Inverse Morphic Equivalence on Languages, *Inform. Process. Letters 19* (1984), 213-218.

[15] Makanin, G.S., The Problem of Solvability of Equations in a Free Semigroup, *Mat. Sb. 103* (1977), 147-236 (English Transl. in: *Math USSR Sb. 32* (1977), 129-198).

[16] Maon, Y., On the Equivalence of Some Transductions Involving Letter to Letter Morphisms on Regular Languages, manuscript (1984).

[17] Restivo, A., and Reutenauer, C., On Cancellation Properties of Languages which are Supports of Rational Formal Power Series, *J. Comput. System Sci. 29* (1984), 153-159.

[18] Rozenberg, G., The Equivalence Problem for Deterministic TOL-Systems is Undecidable, *Inform. Process. Letters 1* (1972), 201-204.

[19] Ruohonen, K., Test Sets for Iterated Morphisms, manuscript (1984).

[20] Rozenberg, G., and Salomaa, A., *The Mathematical Theory of L Systems*, Academic Press, New York, 1980.

[21] Salomaa, A., *Jewels of Formal Language Theory*, Computer Science Press, Rockville, MD, 1981.

[22] Salomaa, A., and Soittola, M., *Automata-Theoretic Aspects of Formal Power Series*, Springer, Berlin, 1978.