RAIRO. INFORMATIQUE THÉORIQUE

J. KARHUMÄKI
H. C. M. KLEIJN

On the equivalence of compositions of morphisms and inverse morphisms on regular languages

RAIRO. Informatique théorique, tome 19, nº 3 (1985), p. 203-211
<http://www.numdam.org/item?id=ITA_1985__19_3_203_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 19, n° 3, 1985, p. 203 à 211)

ON THE EQUIVALENCE OF COMPOSITIONS OF MORPHISMS AND INVERSE MORPHISMS ON REGULAR LANGUAGES (*)

by J. KARHUMÄKI [1] and H. C. M. KLEIJN [2]

Communicated by J. BERSTEL

Abstract. — *We establish as our main results the following two theorems on compositions of morphisms and inverse morphisms. It is undecidable whether or not two transductions of the form* $h_2 h_1^{-1}$, *where* h_1 *and* h_2 *are morphisms, are equivalent (word by word) on a given regular language, while the same problem for transductions of the form* $h_1^{-1} h_2$ *is decidable. Consequently, a sharp borderline between decidable and undecidable problems is found.*

Résumé. — *Nos résultats principaux sont les deux théorèmes suivants sur la composition de morphismes et de morphismes inverses. Il est indécidable si deux transductions de la forme* $h_2 h_1^{-1}$, *où* h_1 *et* h_2 *sont des morphismes, sont équivalents (mot à mot) sur un langage rationnel donné, alors que le même problème est décidable pour des transductions de la forme* $h_1^{-1} h_2$. *Par conséquent, ceci constitue une ligne de démarquation précise entre des problèmes décidables et indécidables.*

1. INTRODUCTION

Among the most natural problems in formal language theory are different kinds of equivalence problems. A typical example is the question of whether or not two transductions of a certain type are equivalent on their domain, *cf.* [1]. We consider this problem in a very simple set-up, namely assuming that the transductions are compositions of morphisms and inverse morphisms and that they are restricted to regular languages.

It was proved in [6] that the equivalence problem for 1-free nondeterministic sequential mappings is undecidable. Consequently, the equivalence problem for rational transductions is also undecidable, *cf.* [1]. On the other hand, this problem becomes decidable when the single-valued rational transductions are considered, *cf.* [2].

(*) Received in March 1984, revised in November 1984.
(1) Department of Mathematics, University of Turku, Turku, Finland.
(2) Institute of Applied Mathematics and Computer Science, University of Leiden, P.O. Box 9512, The Netherlands.

The problem of whether or not two morphisms are equivalent (word by word) on a given language of certain type was raised in [4], where the problem was also shown to be decidable for context-free languages. The topic of this paper, i.e., to study the equivalence of more complicated mappings on languages of certain type, was suggested in [9].

As we saw the problem of whether or not two morphisms are equivalent on a regular language is decidable. On the other hand, recent characterization results of rational transductions, *cf.* [8], [10] or [11], imply that for suitable compositions of morphisms and inverse morphisms the problem of whether or not such compositions are equivalent on a regular language becomes undecidable. That leads one to look for the borderline between the decidability and the undecidability.

The purpose of this note is to point out this borderline. We show, using the previously mentioned result of Griffiths and a recent result of Turakainen, *cf.* [12], that it is undecidable whether or not two transductions of the form $h_2 h_1^{-1}$ are equivalent on a regular language. Furthermore, using the Cross Section Theorem of Eilenberg, *cf.* [5] or [1], we prove that the same problem for transductions of the form $h_1^{-1} h_2$ is decidable. Consequently, we have found a " well-defined " borderline between decidability and undecidability for this particular problem setting.

To emphasize that the above undecidability result is not due to the fact that our family of languages is too complicated but rather because of the properties of morphisms, we also show that this problem remains undecidable if regular languages are replaced by languages of the form F^*, where F is finite. Hence, it is also undecidable whether or not two transductions of the form $h_3 h_2^{-1} h_1$ are equivalent on Σ^*.

Finally, using a result from [2], we conclude that for arbitrary compositions of morphisms and inverse morphisms, such that either all morphisms or all inverse morphisms in at least one of the compositions are injective, their equivalence on a regular language can be decided.

2. DEFINITIONS AND BASIC RESULTS

In this note we adopt the terminology of [1] and we use it also as general reference on basic results on formal languages. We now recall the notions and results needed later on.

Let Σ^* be the free monoid generated by a finite alphabet Σ. The identity of Σ^* is denoted by 1 and $\Sigma^+ = \Sigma^* - \{ 1 \}$. A *transduction* $\tau : \Sigma^* \to \Delta^*$ is a mapping from Σ^* into the set of subsets of Δ^*. For two transductions τ and τ' their composition (if defined) is denoted by $\tau' \circ \tau$, or simply $\tau' \tau$. The domain of a transduction τ is denoted by dom (τ). The transduction determined by the inverse of a transduction τ is denoted by τ^{-1}. Transductions $\tau, \tau' : \Sigma^* \to \Delta^*$ are said to be *equivalent on a language* $L \subseteq \Sigma^*$, in symbols $\tau \overset{L}{\equiv} \tau'$ if $\tau(x) = \tau'(x)$ for all x in L. They are said to be *equivalent* if they are equivalent on Σ^*.

Let $\mathscr{L}$ be a family of languages and θ a family of transductions (defined on suitable alphabets). We denote by

$$EP_{\forall}(\theta, \mathscr{L})$$

the problem of deciding whether or not two given transductions from θ are equivalent on a given language of $\mathscr{L}$. We shall use the notations $\mathscr{H}$ and $\mathscr{H}^{-1}$ for the families of all morphisms and inverse morphisms, respectively. By $\mathscr{H} \circ \mathscr{H}^{-1}$, for instance, we mean the family of transductions of the form $h_2 h_1^{-1}$, where h_1 and h_2 are morphisms. The family of all regular languages is denoted by Reg.

In the next few lines we recall some results and terminology concerning rational transductions. If necessary the reader may consult [1]. A transduction $\tau : \Sigma^* \to \Delta^*$ is *rational* if and only if it is " realized " by a transducer, i.e., by a sixtuple $(\Sigma, \Delta, Q, q_0, F, E)$, where Σ is an input alphabet, Δ is an output alphabet, Q is a set of states, q_0 is the initial state, F is a set of final states, and

$$E \subseteq Q \times \Sigma^* \times \Delta^* \times Q$$

is a set of transitions of T.

A transducer is called *1-free* if $E \subseteq Q \times \Sigma^* \times \Delta^+ \times Q$ and *simple* if $F = \{ q_0 \}$. Further we call a transducer *1-output* if $E \subseteq Q \times \Sigma^* \times \Delta \times Q$. By a *nondeterministic sequential* transducer we mean a transducer satisfying $F = Q$ and $E \subseteq Q \times \Sigma \times \Delta^* \times Q$. Of course, a rational transduction is called *1-free*, *simple*, *1-output* or *nondeterministic sequential* if it is realized by such a transducer. Finally, a transduction $\tau : \Sigma^* \to \Delta^*$ is called *single-valued* if, for each x in Σ^*, $\tau(x)$ contains at most one element, i.e., τ defines a partial function from Σ^* into Δ^*.

The following characterization result for rational transductions is given in [8] and [11] (*cf.* also [10] and [3]).

PROPOSITION 1 : *Each rational transduction* $\tau : \Sigma^* \to \Delta^*$ *admits a factorization*

$$\Sigma^* \xrightarrow{\cdot \$} (\Sigma \cup \{ \$ \})^* \xleftarrow{h_1} \Gamma_1^* \xrightarrow{h_2} \Gamma_2^* \xleftarrow{h_3} \Gamma_3^* \xrightarrow{h_4} \Delta^*$$

where each h_i *is a morphism and* $\cdot \$$ *denotes the marking, i.e., the mapping which associates with each word* x *a new word* $x\$$, *where* $\$$ *is a new symbol not in* Σ.

For simple transductions the above result was recently generalized by Turakainen, *cf.* [12]. For our purposes his result can be stated as follows :

PROPOSITION 2 : *Each* 1-*output simple rational transduction* $\tau : \Sigma^* \to \Delta^*$ *admits a factorization*

$$\Sigma^* \xleftarrow{h_1} \Gamma_1^* \xrightarrow{h_2} \Gamma_2^* \xleftarrow{h_3} \Delta^*$$

where, for any large enough m, the morphisms h_i can be chosen such that, for all $a \in \Delta$, h_3 is defined by $h_3(a) = a\$^m$, where \$ is a new symbol not in Δ.

We shall need not only Proposition 2 but also properties of the construction needed to prove the proposition. Hence, for the sake of completeness, we repeat this construction (implicitly presented already in [10]). If τ is realized by 1-output simple transducer $T = (\Sigma, \Delta, \{ s_0, \ldots, s_n \}, s_0, \{ s_0 \}, E)$, then the morphisms are defined as follows :

$$\begin{aligned} h_1 &: E^* \to \Sigma^* , & h_1(s_i, u, v, s_j) &= u , \\ h_2 &: E^* \to (\Delta \cup \{ \$ \})^* , & h_2(s_i, u, v, s_j) &= S^i v\$^{m-j} , \\ h_3 &: \Delta^* \to (\Delta \cup \{ \$ \})^* , & h_3(a) &= a\$^m , \end{aligned}$$

where m is any natural number $\geqq n$, and \$ is a new symbol not in Δ.

The following remarks on the proof of Proposition 2 will be useful in our later considerations. Firstly, h_3 is injective and can be chosen to be the same for arbitrary two 1-output simple rational transductions. Secondly, the morphisms of the proposition satisfy

$$h_1 h_2^{-1} h_3 h_3^{-1} h_2 h_1^{-1} (\operatorname{dom}(\tau)) \subseteq \operatorname{dom}(\tau) . \tag{1}$$

Next we state two more known results used in our later considerations. The first one is due to Griffiths, *cf.* [6], and it is, in our terms, as follows :

PROPOSITION 3 : *It is undecidable whether or not two* 1-*free nondeterministic sequential transductions are equivalent.*

Our last proposition, due to Blattner and Head, *cf.* [2], is as follows :

PROPOSITION 4 : *It is decidable whether or not a rational transduction is single-valued. Moreover, the equivalence of two single-valued rational transductions is decidable.*

Observe that Proposition 4 implies that the equivalence of two rational transductions, one of which is single-valued, is decidable, too.

3. RESULTS

By Propositions 1 and 3 we conclude that it is undecidable whether or not two given rational transductions of the form $h_4\, h_3^{-1}\, h_2\, h_1^{-1}$ are equivalent on $\Sigma^*\,\$$, where $\$$ is a new symbol not in Σ. This observation can be strengthened as follows :

THEOREM 1 : *$EP_\forall(\mathscr{H}^{-1} \circ \mathscr{H} \circ \mathscr{H}^{-1}$, Reg) is undecidable.*

Proof : By Proposition 2, for 1-output simple rational transductions τ and τ' from Σ^* into Δ^* there exist morphisms h_1, h_2, h_3, g_1, g_2 and g_3 (in suitable alphabets) such that $\tau = h_3^{-1}\, h_2\, h_1^{-1}$ and $\tau' = g_3^{-1}\, g_2\, g_1^{-1}$. Hence, Theorem 1 follows from Proposition 3 and the following lemma.

LEMMA 1 : *For each 1-free nondeterministic sequential rational transduction $\tau_1 : \Sigma^* \to \Delta^*$ there exists a 1-output simple rational transduction*

$$\tau_s : (\Sigma \cup \{ \$ \})^* \to (\Delta \cup \{ \$ \})^*,$$

with $\$ \notin \Sigma \cup \Delta$, such that $\tau_1(x)\,\$ = \tau_s(x\$)$ for all x in Σ^*.

Proof of Lemma 1 : Let τ_1 be realized by a 1-free transducer $T_1 = (\Sigma, \Delta, Q, q_0, Q, E)$. Then we define a 1-output simple transducer $T'_s = (\Sigma \cup \{ \$ \}, \Delta \cup \{ \$ \}, Q', q_0, \{ q_0 \}, E')$, where $\$$ is a new symbol not in $\Sigma \cup \Delta$ and the transitions (and states) of T'_s are defined as follows. Let $e = (q, u, v, q')$, with $v = b_1 \dots b_n$, $n \geqq 2$ and each b_i in Δ, be a transition in E. Then define a set $\psi(e)$ of transitions as follows :

$$\psi(e) = \{ (q, u, b_1, q_1), (q_{n-1}, 1, b_n, q') \} \cup \{ (q_i, 1, b_{i+1}, q_{i+1}) \mid i = 1, \dots, n-2 \},$$

where the states $q_2, \dots, q_{n-1}$ are new not in Q. We also require that the sets of new states obtained from different transitions as above are mutually disjoint. Further, for a transition $e = (q, u, v, q')$, with $v \in \Delta$, in E let $\psi(e) = \{ e \}$. Then we set $E' = \bigcup_{e \in E} \psi(e) \cup \{ (q, \$, \$, q_0) \mid q \in Q \}$.

Then, clearly, the statement of Lemma 1 and hence also Theorem 1 follows. □

Observe that in Theorem 1 the language on which the equivalence of transductions is considered can be assumed to be of the form $\Sigma^*\,\$$, or, as is easy to see, of the form $\mathrm{dom}(h_3^{-1}\, h_2\, h_1^{-1})$.

In the next theorem we still strengthen the result of Theorem 1.

THEOREM 2 : *$EP_\forall(\mathscr{H} \circ \mathscr{H}^{-1}$, Reg) is undecidable.*

Proof : Let $(h_3^{-1}\, h_2\, h_1^{-1}, g_3^{-1}\, g_2\, g_1^{-1}, L)$ be an instance of the problem of

Theorem 1. Without affecting the undecidability we may assume, as is straightforward to see, that this instance satisfies the following conditions. Firstly, $L = \operatorname{dom}(h_3^{-1} h_2 h_1^{-1}) = \operatorname{dom}(g_3^{-1} g_2 g_1^{-1})$. Secondly, $h_3 = g_3$ and moreover h_3 is injective. Thirdly, the morphisms satisfy

$$\begin{cases} h_1 h_2^{-1} h_3 h_3^{-1} h_2 h_1^{-1}(L) \subseteq L \quad \text{and} \\ g_1 g_2^{-1} h_3 h_3^{-1} g_2 g_1^{-1}(L) \subseteq L \end{cases} \tag{2}$$

The last two sentences follow from remarks after Proposition 2.

We shall show

$$\begin{cases} h_3^{-1} h_2 h_1^{-1} \overset{L}{\equiv} h_3^{-1} g_2 g_1^{-1} \quad \text{if and only if} \\ h_2 h_1^{-1}(L) = g_2 g_1^{-1}(L) \quad \text{and} \quad h_1 h_2^{-1} \overset{L'}{\equiv} g_1 g_2^{-1} \\ \text{where} \quad L' = h_3 h_3^{-1} h_2 h_1^{-1}(L)\,. \end{cases} \tag{3}$$

The equivalence (3) together with the known properties of regular languages imply Theorem 2.

To prove (3) we first observe, by (2), that $h_3^{-1} h_2 h_1^{-1} \overset{L}{\equiv} h_3^{-1} g_2 g_1^{-1}$ if and only if $L_1 = h_3^{-1} h_2 h_1^{-1}(L) = h_3^{-1} g_2 g_1^{-1}(L)$ and $h_1 h_2^{-1} h_3 \overset{L_1}{\equiv} g_1 g_2^{-1} h_3$. Hence, the injectiveness of h_3 yields (3). □

Let $\mathscr{C}_w$ denote the family of morphisms h satisfying $|h(a)| \leqq 1$ for all letters a and $\mathscr{H}_1$ a family of nonerasing morphisms. Then, by a careful analysis of our proof of Theorem 2, one can see that the family $\mathscr{H} \circ \mathscr{H}^{-1}$ of transductions can actually be replaced by the family $\mathscr{C}_w \circ \mathscr{H}_1^{-1}$. Indeed, in the proof of Lemma 1 the transitions of T_s are in $Q \times (\Sigma \cup \{1\}) \times \Delta \times Q$ and for such transductions the morphisms h_1 and h_2 in the proof of Proposition 2 are in $\mathscr{C}_w$ and $\mathscr{H}_1$, respectively.

To emphasize that the above undecidability results are mainly due to powerful properties of morphisms and not because the equivalence is restricted to complicated enough languages, we still strengthen our result slightly. In order to be able to do this let $\mathscr{F}$ denote the family of languages of the form F^* where F is finite.

THEOREM 3 : *$EP_\forall(\mathscr{H} \circ \mathscr{H}^{-1}, \mathscr{F})$ is undecidable.*

Proof : According to the proof of Theorem 2, $EP_\forall(\mathscr{H} \circ \mathscr{H}^{-1}, \text{Reg})$ remains undecidable even if only the regular star languages, i.e., regular languages of the form L_1^*, are considered. Indeed, $\operatorname{dom}(h_3^{-1} h_2 h_1^{-1})$ is always a star language. Further by a result in [10] (*cf.* also [12]) for each regular star language

L_* there exist an *injective* morphism h and a finite language F such that $L_* = h^{-1}(F^*)$. Hence, the result follows from Theorem 2. $\square$

The remark following Theorem 2 applies to Theorem 3, too. As another remark we state the following interesting corollary of Theorem 3.

THEOREM 4 : *It is undecidable whether or not two transductions from* $\mathscr{H} \circ \mathscr{H}^{-1} \circ \mathscr{H}$ *are equivalent.* $\square$

Our next result shows that the order of the morphisms and the inverse morphisms in our previous results is crucial.

THEOREM 5 : $EP_\forall(\mathscr{H}^{-1} \circ \mathscr{H}, \text{Reg})$ *is decidable.*

Proof : Let $(h_2^{-1}\, h_1, g_2^{-1}\, g_1, L)$ be an instance of the problem and let $L' = h_2^{-1}\, h_1(L)$. If $L' \neq g_2^{-1}\, g_1(L)$, then the transductions $h_2^{-1}\, h_1$ and $g_2^{-1}\, g_1$ are not equivalent on L. Since L' is effectively regular this can be decided. So we may assume that $h_2^{-1}\, h_1(L) = g_2^{-1}\, g_1(L) = L'$. By similar arguments we may also assume that $L \subseteq \operatorname{dom}(h_2\, h_1^{-1}) = \operatorname{dom}(g_2^{-1}\, g_1)$.

We define a partition of L induced by h_1, in symbols $\sim_{L,h_1}$, as follows : $x \sim_{L,h_1} x'$ if and only if $h_1(x) = h_1(x')$. Similarly, we define partitions $\sim_{L',h_2}$, $\sim_{L,g_1}$ and $\sim_{L',g_2}$. Furthermore, let L_{cs} be a cross section of L with respect to h_1, i.e., L_{cs} is a regular subset of L such that h_1 maps L_{cs} bijectively onto $h_1(L)$. Such a cross section can be effectively found, *cf.* [5] or [1]. Similarly, let L'_{cs} be a cross section of L' with respect to h_2.

With the above notation we establish the following diagram :

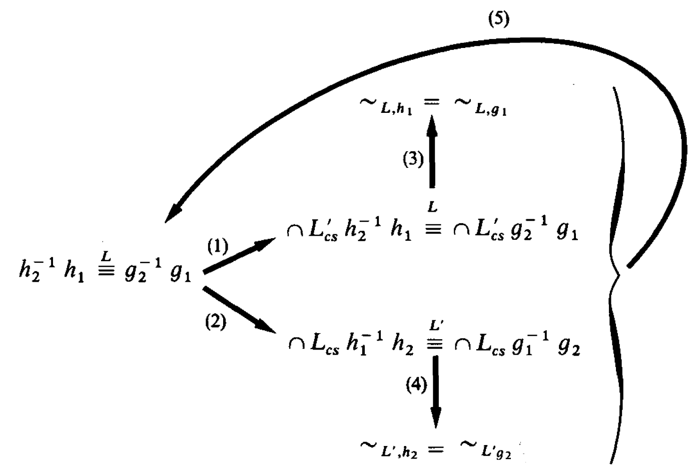

Diagram 1.

Before proving Diagram 1 we point out how Theorem 5 follows from it. By definition, rational transductions $\cap L'_{cs} h_2^{-1} h_1 \cap L$ and $\cap L_{cs} h_1^{-1} h_2 \cap L'$ are single-valued. Consequently, by Proposition 4, it is decidable whether or not the transductions $\cap L'_{cs} h_2^{-1} h_1$ and $\cap L'_{cs} g_2^{-1} g_1$ are equivalent on L, as well as whether or not the transductions $\cap L_{cs} h_1^{-1} h_2$ and $\cap L_{cs} g_1^{-1} g_2$ are equivalent on L'. Hence Diagram 1 implies Theorem 5.

Now, we turn to prove the implications in Diagram 1. Implication (1) is trivial, and also (2) is easy to see. To prove (3) let $\cap L'_{cs} h_2^{-1} h_1 \overset{L}{\equiv} \cap L'_{cs} g_2^{-1} g_1$ and $x \sim_{L,h_1} x'$. Then $h_1(x) = h_1(x')$ and so $\cap L'_{cs} g_2^{-1} g_1(x) = \cap L'_{cs} g_2^{-1} g_1(x')$. This together with our assumption $L \subseteq \operatorname{dom}(h_2 h_1^{-1}) = \operatorname{dom}(g_2 g_1^{-1})$ implies that $g_1(x) = g_1(x')$, i.e., $x \sim_{L,g_1} x'$. Therefore (3) follows by symmetry. Implication (4) can be proved in the same way as (3). Finally, implication (5) follows from the fact that conditions $h_2^{-1} h_1 \overset{L}{\equiv} g_2^{-1} g_1$ and

$$\cap L'_{cs} h_2^{-1} h_1 \overset{L}{\equiv} \cap L'_{cs} g_2^{-1} g_1$$

are equivalent under the assumptions $\sim_{L,h_1} = \sim_{L,g_1}$ and $\sim_{L',h_2} = \sim_{L',g_2}$. Hence, we have proved Diagram 1 and so also Theorem 5. □

It is worth noting that if we consider a weaker equivalence i.e., the existential equivalence of [9], of transductions then the problem of Theorem 5 becomes undecidable. We have even the following stronger result : It is undecidable whether or not, for a given triple (h^{-1}, g^{-1}, L), where h and g are morphisms and L a regular language, the relation $h^{-1}(x) \cap g^{-1}(x) \neq \emptyset$ holds for all x in L. The details can be found in [7].

We conclude this note with another decidability result.

THEOREM 6 : *It is decidable whether or not transductions τ and τ' are equivalent on L, where τ and τ' are compositions of morphisms and inverse morphisms such that either all the morphisms of τ or all the inverse morphisms of τ are injective and L is a regular language.*

Proof : Firstly, we assume that all the inverse morphisms of τ are injective. Then τ is single-valued and therefore the equivalence of τ and τ' on L can be decided by Proposition 4.

Secondly, we assume that all the morphisms of τ are injective. Then $\tau^{-1} \tau(L) \subseteq L$ and therefore τ and τ' are equivalent on L if and only if τ^{-1} and $\cap L\tau'^{-1}$ are equivalent on $\tau(L)$. But now τ^{-1} is single-valued and therefore the result follows from Proposition 4 and the fact that $\tau(L)$ is effectively regular. □

As a consequence of Theorem 6 we note that the undecidability problem of Theorem 2 becomes decidable if one assumes that at least one of the four morphisms is injective.

ACKNOWLEDGEMENT

The authors are grateful to Y. Maon, G. Rozenberg and A. Salomaa for useful comments and to the Academy of Finland for financial support.

REFERENCES

1. J. Berstel, *Transductions and Context-Free Languages* (B. G. Teubner, Stuttgard, 1979).
2. M. Blattner and T. Head, *Single-valued a-transducers,* J. Comput. System Sci. 15 (1977) 310-327.
3. K. Culik II, F. E. Fich and A. Salomaa, *A homomorphic characterization of regular languages,* Discrete Appl. Math. 4 (1982) 149-152.
4. K. Culik II and A. Salomaa, *On the decidability of homomorphism equivalence for languages,* J. Comput. System Sci. 17 (1978) 163-175.
5. S. Eilenberg, *Automata, Languages, and Machines,* Vol. A (Academic Press, New York, 1974).
6. T. V. Griffiths, *The unsolvability of the equivalence problem for λ-free nondeterministic generalized machines,* J. Assoc. Comput. Mach. 15 (1968) 409-413.
7. J. Karhumäki and Y. Maon, *A simple undecidable problem : Existential agreement of inverse of two morphisms on a regular language,* submitted for publication.
8. J. Karhumäki and M. Linna, *A note on morphic characterization of languages,* Discrete Appl. Math. 5 (1983) 243-246.
9. J. Karhumäki and D. Wood, *Inverse morphic equivalence on languages,* manuscript (1983).
10. M. Latteux and J. Leguy, *On the composition of morphisms and inverse morphisms,* Lecture Notes in Computer Science 154 (1983) 420-432.
11. P. Turakainen, *On homomorphic characterization of principal semi AFL's without using intersection with regular sets,* Inform. Sci. 27 (1982) 141-149.
12. P. Turakainen, *A machine-oriented approach to compositions of morphisms and inverse morphisms,* EATCS Bull. 20 (1983) 162-166.