Rational representation of forests by tree automata

By G. MARÓTI

1. Introduction

In this paper we give a new representation of forests which is more powerful than the usual one in the following sense: for this representation there exists a proper variety which is complete, i.e., every regular forest can be represented (in this new sense) by a tree automaton built on a finite algebra belonging to this variety (Theorem 5). This representation is a generalization of the rational one developed by F. Gécseg in [1]. Moreover our Theorem 5 yields immediately the result of F. Gécseg and G. Horváth [2]: there exists a proper variety over the type $G=\{g, h\}$, where the arities of g and h are 2 and zero, respectively, such that every context-free language can be recognized by a finite tree-automaton belonging to this variety.

2. *Fr*-homomorphism and *Fr*-embedding

Let F be a nonvoid set and r a mapping of F into the set N of all nonnegative integers. We call the ordered pair $\langle F, r\rangle$ a *type*. The elements of F are the *operational symbols*. If $f\in F$ and $r(f)=n$ $(n\in N)$ then we say that the *arity* of f is n (or f is an *n-ary* operational symbol). We will refer to the type $\langle F, r\rangle$ simply by F. The subset of all 0-ary operational symbols will be denoted by F^0.

Take the set $X=\{x_0, x_1, \ldots\}$ and a type F. The set $T_{F,n}$ of the *n-ary polynomial symbols over* F is defined by

1) $x_0, \ldots, x_{n-1}\in T_{F,n}$,

2) if $p_0, \ldots, p_{m-1}\in T_{F,n}$ and $f\in F$ is an m-ary operational symbol $(m\geqq 0)$ then $f(p_0, \ldots, p_{m-1})\in T_{F,n}$,

3) $T_{F,n}$ is the smallest set satisfying 1) and 2).

The set T_F of all polynomial symbols over F is defined as the union of all $T_{F,n}$

$$T_F = \bigcup_{n=0}^{\infty} T_{F,n}.$$

Every polynomial symbol $p\in T_F$ can be represented by a tree P (by a loop-free connected graph) whose nodes are labelled by the elements of the set $F\cup X$ in such a way that if a node has the label $f\in F$ then there are exactly $r(f)$ edges leaving it. We use the terminology that P is the tree belonging to the polynomial symbol p.

Consider the polynomial symbols $p\in T_{F,m}$ and $p_0, \dots, p_{m-1}\in T_{F,n}$. Then $p(p_0, \dots, p_{m-1})$ denotes the following n-ary polynomial symbol over F:

1) if $p=x_i$ $(0\leqq i\leqq m-1)$ then $p(p_0, \dots, p_{m-1})=p_i$,

2) if $p=f(q_0, \dots, q_{k-1})$, where $f\in F$ and $r(f)=k$, then

$$p(p_0, \dots, p_{m-1})=f(q_0(p_0, \dots, p_{m-1}), \dots, q_{k-1}(p_0, \dots, p_{m-1})).$$

Next we define the *mapping fr (frontier)*: *fr* is a mapping of T_F into the free monoid generated by the set X satisfying the following conditions:

1) $fr(x_i)=x_i$ $(i=0, 1, \dots)$,

2) if $h\in F^0$, then $fr(h)=\varepsilon$ (ε denotes the empty word),

3) if $p=f(p_0, \dots, p_{m-1})$, where $f\in F$ and $r(f)=m$, then

$$fr(p)=fr(p_0)\dots fr(p_{m-1}).$$

Let us consider now two types F and G. The mapping $\alpha: T_F\to T_G$ is called an *fr-homomorphism (frontier-homomorphism)* if it satisfies the following conditions:

(i) $\alpha(x_i)=x_i$ $(i=0, 1, \dots)$

(ii) $fr(\alpha(f(x_0, \dots, x_{n-1})))=fr(f(x_0, \dots, x_{n-1}))$, where $f\in F$ and $r(f)=n$ $(n\geqq 0)$,

(iii) $\alpha(f(p_0, \dots, p_{n-1}))=\alpha(f)(\alpha(p_0), \dots, \alpha(p_{n-1}))$.

Corollary 1. For every polynomial symbol $p\in T_F$ and for every *fr*-homomorphism $\alpha: T_F\to T_G$ we have

$$fr(\alpha(p))=fr(p).$$

Let $d(p)$ denote the *depht* of the polynomial symbol p, i.e., if p is equal to x_i or a 0-ary operational symbol then $d(p)=0$, and if p is of the form $p=f(p_0, \dots, p_{m-1})$ then $d(p)=\max_{i=0,\dots,m-1}\{d(p_i)\}+1$.

Corollary 2. Let $\alpha: T_F\to T_G$ be an *fr*-homomorphism, and assume that for every $f\in F$, $d(\alpha(f(x_0, \dots, x_{r(f)-1})))\geqq 1$. Then for each $p\in T_F$

$$d(\alpha(p))\geqq d(p)$$

holds.

Proof. Let $p\in T_F$. If $d(p)=0$ then the assertion is trivial. Assume that Corollary 2 is true for all polynomial symbols whose depht is less than that of $p=f(p_0, \dots, p_{m-1})$. Then

$$d(\alpha(p))=d(\alpha(f(p_0, \dots, p_{m-1})))=d(\alpha(f)(\alpha(p_0), \dots, \alpha(p_{m-1}))\geqq$$
$$\geqq 1+\max_{i=0,\dots,m-1}\{d(\alpha(p_i))\}\geqq 1+\max_{i=0,\dots,m-1}\{d(p_i)\}=d(p).$$

If the *fr*-homomorphism $\alpha: T_F\to T_G$ is one-to-one then it is called *fr-embedding*.

Let us denote by $T_F[1]$ the set of all polynomial symbols from F with depth less than or equal to 1:

$$T_F[1]=\{p \mid p\in T_F \text{ and } d(p)\leqq 1\}.$$

For every mapping $\varphi: T_F[1] \to T_G$ satisfying condition (ii) there exists exactly one *fr*-homomorphism $\alpha: T_F \to T_G$ such that $\alpha \restriction T_F[1] = \varphi$, where $\alpha \restriction T_F[1]$ denotes the restriction of α to $T_F[1]$.

If we take two types F and G then an *fr*-homomorphism not necessarily exists between T_F and T_G. For example if G consists of unary operational symbols only and in F there exists an operational symbol with arity greater than or equal to 2 then, obviously, there is no *fr*-homomorphism between T_F and T_G.

Consider the type F and denote by $S(F)$ the following set of nonnegative integers

$$S(F) = \{n \mid \exists f \in F \text{ with } r(f) = n\}.$$

The set $\{0, \ldots, m-1\}$ will be denoted by $\tilde{m}$ for all natural number m.

Theorem 1. Let F and G be types, $S(F) = \{n_0, \ldots, n_{r-1}\}$ and $S(G) = \{m_0, \ldots, m_{s-1}\}$. If there exists an *fr*-homomorphism α: $T_F \to T_G$, then for a suitable mapping φ: $\tilde{r} \to \tilde{s}$ we have

$$(m_0-1, \ldots, m_{s-1}-1) | (n_0 - m_{0\varphi}, \ldots, n_{r-1} - m_{(r-1)\varphi}). \tag{1}$$

Proof. Let α: $T_F \to T_G$ be an *fr*-homomorphism. If G has an operational symbol with arity zero, then (1) holds for every mapping φ: $\tilde{r} \to \tilde{s}$ because of

$$(m_0-1, \ldots, -1, \ldots, m_{s-1}-1) = 1.$$

In the opposite case take a $p \in T_F$ of depth 1, and let $q = \alpha(p)$. Then

$$|fr(q)| = |fr(p)| = n_k \tag{2}$$

for some $k \in \tilde{r}$. Now consider the tree Q belonging to q. In consequence of (2) Q has n_k leaves. Delete in Q all leaves belonging to a given subtree of Q with depth 1. We get a tree with $n_k - (m_{i_1} - 1)$ leaves, where $i_1 \in \tilde{s}$. Continue the deletion of the leaves of the subtrees from Q with depth 1 as long as we get a tree of depth 1. At each step the number of leaves of the current tree was reduced by $(m_{i_v} - 1)$ for some $i_v \in \tilde{s}$. At the end of the process, the tree of depth 1 must have m_j leaves, where $j \in \tilde{s}$. In this way for suitable nonnegative integers $l_0, \ldots, l_{s-1}$ we have

$$n_k - l_0(m_0-1) - \ldots - l_{s-1}(m_{s-1}-1) = m_j. \tag{3}$$

Hence

$$n_k - m_j = l_0(m_0-1) + \ldots + l_{s-1}(m_{s-1}-1). \tag{4}$$

Let d be the greatest common divisor of $m_0-1, \ldots, m_{s-1}-1$. Then d divides the right side of (4). Therefore d divides $n_k - m_j$ as well.

Take the correspondence $k \to j$, and denote it by ψ

$$k\psi = j. \tag{5}$$

Since $p \in T_F[1]$ was arbitrary, while it runs over the set $T_F[1]$ in (5), thus k must run (not necessarily once) over the numbers $0, \ldots, r-1$ and, meanwhile, for every $k \in \tilde{r}$, $k\psi$ assigns a subset of $\tilde{s}$. Let φ be a choice function of the system of sets $\{k\psi \mid k \in \tilde{r}\}$. Because of (4), $n_k - m_{k\varphi}$ can be divided by d for every $k \in \tilde{r}$. Therefore, d divides their greatest common divisor, as we stated. □

Unfortunately, condition (1) is not sufficient. Indeed, let F consist of a single unary operational symbol and let $G=\{g\}$ with $r(g)=2$. It is clear that condition (1) holds, but that in T_G there is no tree with a single leaf.

Theorem 2. Using the notations of the previous theorem, the necessary and sufficient condition of the existence of an fr-homomorphism between T_F and T_G is the validity of the following equalities

$$n_k = m_{k\varphi}+l_0(m_0-1)+\ldots+l_{s-1}(m_{s-1}-1) \quad (k = 0, \ldots, r-1), \tag{6}$$

where φ is a mapping of $\tilde{r}$ into $\tilde{s}$ and l_i are nonnegative integers for $i=0, \ldots, s-1$.

Proof. The necessity of conditions is trivial by the proof of the previous theorem.

Before we are going to prove the sufficiency let us note, that if a natural number n is of the form

$$n = m_i+y_0(m_0-1)+\ldots+y_{s-1}(m_{s-1}-1),$$

where $i\in\tilde{s}$ and $y_0, \ldots, y_{s-1}$ are nonnegative integers, then there exists a q in T_G such that

$$fr(q) = x_0\ldots x_{n-1}.$$

We proof this statement by induction on s. For $s=1$,

$$n = m_0+y_0(m_0-1).$$

If $g\in G$ with $r(g)=m_0$, then the polynomial symbol

$$g(\ldots g(g(x_0, \ldots, x_{m_0-1}), x_{m_0}, \ldots, x_{2m_0-1}), \ldots, x_{n-1})$$

is appropriate. Remark, that this choice is possible since $n>0$ implies $m_0>0$.

Now assume that our statement has been shown for $s=v$, i.e. for each natural number n' of the form

$$n' = m_i+y_0(m_0-1)+\ldots+y_{v-1}(m_{v-1}-1)$$

there exists the desired q' in $T_{G'}$ and let

$$n = m_i+y_0(m_0-1)+\ldots+y_v(m_v-1) = n'+y_v(m_v-1).$$

We distinguish three cases. If $m_v>1$, then we can choose for q the polynomial symbol

$$g(\ldots g(q', x_{n'}, \ldots, x_{n'+m_v-1})\ldots x_{n-1}),$$

where $g\in G$ and $r(g)=m_v$. If $m_v=1$ then $n=n'$ and, therefore, q' itself is suitable. Finally, if $m_v=0$ and h is 0-ary operational symbol in G, then let q be the polynomial symbol which can be obtained from q' by replacing the variables $x_n, x_{n+1}, \ldots, x_{n'-1}$ by h.

Now assume that conditions (6) hold for the types F and G. In order to show the sufficiency of our conditions it is enough to define a mapping α: $T_F[1]\to T_G$ with $fr(\alpha(p))=fr(p)$ for every $p\in T_F[1]$. If in F there is no 0-ary operational symbol then for $f(x_{j_0}, \ldots, x_{j_{n_k-1}})$ let

$$\alpha(f(x_{j_0}, \ldots, x_{j_{n_k-1}})) = q(x_{j_0}, \ldots, x_{j_{n_k-1}}),$$

where $q \in T_G$ the polynomial symbol with $fr(q) = x_0 \dots x_{n_k-1}$, whose existence was shown above. In the opposite case in G there must be a 0-ary operational symbol as well, say h. For every $f \in F^0$ let $\alpha(f) = h$. Furthermore if $f \in F \setminus F^0$ is of arity n_k $(k \in \tilde{r})$ and $q \in T_G$ is the polynomial symbol with $fr(q) = x_0 \dots x_{n_k-1}$, then for $f(y_{j_0}, \dots, y_{j_{n_k-1}}) \in T_F[1]$ let

$$\alpha(f(y_{j_0}, \dots, y_{j_{n_k-1}})) = q(z_{j_0}, \dots, z_{j_{n_k-1}}),$$

where

$$z_{j_i} = \begin{cases} y_{j_i} & \text{if} \quad y_{j_i} \in X, \\ h & \text{if} \quad y_{j_i} \in F^0 \quad (i = 0, \dots, n_k - 1). \end{cases}$$

Theorem 2 provides two necessary and sufficient conditions for the existence of fr-embedding $\alpha: T_F \to T_G$ for every type F.

Theorem 3. The following three conditions are equivalent:
1) for every type F there exists an fr-homomorphism of T_F into T_G,
2) in G there exist a 0-ary and an at least binary operational symbols,
3) for every type F there exists an fr-embedding T_F into T_G.

Proof. Because of the previous theorem, 1) is equivalent to 2), and it is clear that 3) implies 2). Therefore, it is enough to prove the implication 2)⇒3).

For this let $g, h \in G$ with $r(h) = 0$ and $r(g) \geqq 2$. Consider an arbitrary type F and take a one-to-one mapping γ of F into T_G, for which

$$|fr(\gamma(f))| = r(f)$$

holds for every $f \in F$. Now we define the mapping β: $T_F[1] \to T_G$ in the following manner:

1) $\beta(x_i) = x_i$
2) $\beta(f) = g(h, \dots, h, \gamma(f))$ if $f \in F^0$,
3) $\beta(f(y_{i_0}, \dots, y_{i_{n-1}})) = g(h, \dots, h, \gamma(f)(\beta(y_{i_0}), \dots, \beta(y_{i_{n-1}})))$, where $y_{i_j} \in X \cup F^0$ $(j = 0, \dots, n-1)$ and $f \in F \setminus F^0$.

Obviously β is one-to-one. Moreover, for every $p \in T_F[1]$ we have

$$fr(\beta(p)) = fr(p).$$

Assume that $F = \{f_0, \dots, f_{k-1}\}$ and take the following unary polynomial symbols from T_G

$$q_0 = g(x_0, h, \dots, h)$$
$$q_j = g(q_{j-1}, h, \dots, h) \quad (j = 0, \dots, k-1).$$

Finally, let us denote by α' the mapping of $T_F[1]$ into T_G for which

$$\alpha'(p) = q_j(\beta(p)),$$

where $p = f_j(p_0, \dots, p_{n-1}) \in T_F[1]$. Obviously, α' can be extended to an fr-homomorphism $\alpha: T_F \to T_G$. We claim that α is an fr-embedding. Indeed, assume that for the polynomial symbols p and q in T_F, $\alpha(p) = \alpha(q)$. We proceed by induction on the depth of p.

If $p = x_i$ then $\alpha(p) = x_i$. Moreover,

$$0 = d(x_i) = d(\alpha(p)) = d(\alpha(q)) \geqq d(q) \geqq 0$$

implies $d(q) = 0$, which yields $q = x_i$. If $p = f_j \in F^0$ then $d(p) = \alpha(f_j) = q_j(g(h, \ldots, h, \gamma(f_j)))$. Assume that q has the form $f_k(t_0, \ldots, t_{m-1})$. Then

$$\alpha(q) = \alpha(f_k(t_0, \ldots, t_{m-1})) = q_k(g(h, \ldots, h, \gamma(f_k)(\alpha(t_0), \ldots, \alpha(t_{m-1})))).$$

This and the assumption $\alpha(p)=\alpha(q)$ jointly imply

$$q_j(g(h, \ldots, h, \gamma(f_j))) = q_k(g(h, \ldots, h, \gamma(f_k)(\alpha(t_0), \ldots, \alpha(t_{m-1})))).$$

But this yields $j=k$.

Finally, assume that $p=f_j(p_0, \ldots, p_{n-1})$ and that the statement has been shown for every p' with $d(p')<d(p)$. Let $q=f_k(q_0, \ldots, q_{m-1})$. Then $\alpha(p)=\alpha(q)$ implies

$$q_j(g(h, \ldots, h, \gamma(f_j)(\alpha(p_0), \ldots, \alpha(p_{n-1})))) = \\ = q_k(g(h, \ldots, h, \gamma(f_k)(\alpha(q_0), \ldots, \alpha(q_{m-1})))). \quad (7)$$

But this holds only if $q_j=q_k$, which is equivalent to $j=k$. Thus (7) yields that $\alpha(p_i)=\alpha(q_i)$ $(i=0, \ldots, k-1)$, which makes the proof complete.

3. *Fr*-representation

Let F be a finite type and $\mathfrak{A}=\langle A, F\rangle$ a finite F-algebra (for terminology, see [3] and/or [1]). The triple $\overline{\mathfrak{A}}=(\mathfrak{A}, \underline{a}, A')$ is called an *n-ary tree automaton over F*, or shortly *n-ary F-automaton*, where $A'\subseteq A$ is the *set of final states* and $\underline{a}\in A^n$ is the *initial vector*.

According to the terminology used in the theory of tree automata the polynomial symbols over F and the subsets of T_F will be called *F-trees* and *F-forests*, respectively.

Consider the n-ary F-automaton $\overline{\mathfrak{A}}=(\mathfrak{A}, \underline{a}, A')$ and let us denote by $T(\overline{\mathfrak{A}})$ the following subset of $T_{F,n}$

$$T(\overline{\mathfrak{A}}) = \{p | p\in T_{F,n} \text{ and } p_{\mathfrak{A}}(\underline{a})\in A'\}.$$

We say that the forest $T\subseteq T_{F,n}$ *can be recognized* by $\overline{\mathfrak{A}}$ (or $\overline{\mathfrak{A}}$ *represents the forest* T) if $T=T(\overline{\mathfrak{A}})$.

Let $T_1, T_2\subseteq T_{F,n}$ and $0\leq i\leq n-1$. The *x_i-product* of T_1 and T_2 is the forest which can be obtained by replacing every occurence of x_i of some tree from T_2 by a tree in T_1. We denote the x_i-product of T_1 and T_2 by $T_1x_iT_2$. Let $T^{0,i}=\{x_i\}$ and $T^{k,i}=T^{k-1,i}\cup T^{k-1,i}x_iT$ $(k=1, 2, \ldots)$. Finally, let us denote by $T^{*,i}$ the union of all forests $T^{k,i}$:

$$T^{*,i} = \bigcup_{k=0}^{\infty} T^{k,i}.$$

$T^{*,i}$ is called the *x_i-iteration* of the forest T.

We say that the forest $T\subseteq T_{F,n}$ is *m-regular* if it can be obtained from finitely many trees of $T_{F,m}$ by finitely many applications of union, x_i-product and x_i-iteration. A forest T is called regular if it is m-regular for some m.

It is well known that a forest is regular if and only if it can be recognized by a tree automaton [1].

Take a forest $T \subseteq T_{F,n}$ and an n-ary G-automaton $\overline{\mathfrak{A}}=(\mathfrak{A}, \underline{a}, A')$. We say that $\overline{\mathfrak{A}}$ *fr-represents* the forest T (or T *can be fr-recognized* by $\overline{\mathfrak{A}}$) if there exists an *fr*-embedding $\alpha\colon T_F \to T_G$ such that $\alpha(T)=T(\overline{\mathfrak{A}})$.

Theorem 4. A forest is regular if and only if it can be *fr*-recognized by a tree automaton.

Proof. We shall show that the image and the complete inverse image of a regular forest under an *fr*-homomorphism are regular as well. This yields for us the sufficience of our conditions. The necessity is trivial.

Let $\alpha\colon T_F \to T_G$ be *fr*-homomorphism. From the definition of union and x_i-product of forests immediately follows that for each $T_1, T_2 \subseteq T_{F,n}$ we have

$$\alpha(T_1 \cup T_2) = \alpha(T_1) \cup \alpha(T_2), \tag{8}$$

$$\alpha(T_1 x_i T_2) = \alpha(T_1) x_i \alpha(T_2). \tag{9}$$

After this by induction on k it is easy to show that

$$\alpha(T_1^{k,i}) = \alpha(T_1)^{k,i} \quad (k = 0, 1, \ldots).$$

From this we get

$$\alpha(T_1^{*,i}) = \alpha\left(\bigcup_{k=0}^{\infty} T_1^{k,i}\right) = \bigcup_{k=0}^{\infty} \alpha(T_1^{k,i}) = \bigcup_{k=0}^{\infty} \alpha(T_1)^{k,i} = \alpha(T_1)^{*,i}. \tag{10}$$

Consider now the regular forest $T \subseteq T_F$, and assume that it can be obtained from the trees $p_0, \ldots, p_{k-1} \in T_F$ by finitely many application of regular operations (union, x_i-product, x_i-iteration). Because of (8)—(10), $\alpha(T)$ must be obtained from $\alpha(p_0), \ldots, \alpha(p_{k-1})$ by finitely many applications of the regular operations, namely in exactly such a manner as T is built up from $p_0, \ldots, p_{k-1}$. Therefore, $\alpha(T)$ is regular as well.

Now take two forests $T \subseteq T_{G,n}$ and $T' \subseteq T_{F,n}$, and assume that $T' = \alpha^{-1}(T)$ and that T is regular. Then for some n-ary G-automaton $\overline{\mathfrak{A}}$, $T = T(\overline{\mathfrak{A}})$. Take the F-algebra $\mathfrak{B} = \langle B, F\rangle$ such that $B = A$ and for every $f \in F$, $f_{\mathfrak{B}} = \alpha(f)_{\mathfrak{A}}$. Moreover consider the n-ary F-automaton $\overline{\mathfrak{B}} = (\mathfrak{B}, \underline{a}, A')$. We claim that $T(\overline{\mathfrak{B}}) = T'$. Indeed for every $p \in T_{F,m}$, $p \in T(\overline{\mathfrak{B}})$ if and only if $p_{\mathfrak{B}}(\underline{a}) \in A'$. But $p_{\mathfrak{B}}(\underline{a}) = \alpha(p)_{\mathfrak{A}}\ (\underline{a}) \in A'$ is equivalent to $\alpha(p) \in T (= T(\overline{\mathfrak{A}}))$. Finally, $\alpha(p) \in T$ if and only if $p \in \alpha^{-1}(T)\ (= T')$. The proof is complete.

Let K be a class of G-algebras. We say that K is *fr-complete*, if for every regular forest T (not necessarily over the type G) there exists a finite algebra $\mathfrak{A} = \langle A, F\rangle$ in K, an $\underline{a} \in A^n$ and $A' \subseteq A$ such that the tree automaton $\overline{\mathfrak{A}} = (\mathfrak{A}, \underline{a}, A')$ *fr*-represents the forest T.

Our aim is to prove the existence of a nontrivial *fr*-complete variety. In order to show this, take the type G in which there exist two operational symbols g and h with $r(g) \geqq 2$ and $r(h) = 0$. Furthermore consider the equation

$$g(h, \ldots, h, g(h, \ldots, h)) = g(h, \ldots, h, g(h, \ldots, h), g(h, \ldots, h)). \tag{11}$$

Theorem 5. The variety defined by the equation (11) is *fr*-complete.

Proof. Let $\alpha: T_G \to T_G$ be an *fr*-homomorphism such that:

1) $\alpha(h) = g(g(h, \ldots, h), h, \ldots, h)$,

2) $\alpha(g(x_{i_0}, \ldots, x_{i_{m-1}})) = g(g(x_{i_0}, \ldots, x_{i_{m-1}}), h, \ldots, h)$,

3) on the set of all other polynomial symbols of T_G with depth less than or equal to 1 α is the identity mapping.

We claim that α is *fr*-embedding. Indeed, let $\alpha(p)=\alpha(q)$. If $p=x_i$ then obviously q must be equal to x_i. If $p=h$ then because of $\alpha(h)=g(g(h, \ldots, h), h, \ldots, h)$, q does not contain any operational symbols different from g and h. Therefore, if $d(q) \geqq 1$, then q must have the form $g(p_0, \ldots, p_{m-1})$. In this way from

$$g(g(h, \ldots, h), h, \ldots, h) = g(g(\alpha(p_0), \ldots, \alpha(p_{m-1})), h, \ldots, h)$$

it follows that $h=\alpha(p_0)$ which is a contradiction. Therefore, $d(q)=0$ and thus q must be equal to h. Finally, if p is 0-ary operational symbol different from h then $p=q$ obviously holds.

Now assume that $d(p) \geqq 1$ and that our statement has been shown for every polynomial symbol with depth less than that of p. Moreover, let $p=g_1(p_0, \ldots, p_{k-1})$ and $q=g_2(q_0, \ldots, q_{l-1})$. Then

$$\alpha(g_1)(\alpha(p_0), \ldots, \alpha(p_{k-1})) = \alpha(g_2)(\alpha(q_0), \ldots, \alpha(q_{l-1})) \tag{12}$$

yields that $\alpha(g_1)$ and $\alpha(g_2)$ must begin with the same operational symbol, but this is possible only if $g_1=g_2$. Therefore, from (12) we get that $k=l$ and $\alpha(p_i)=\alpha(q_i)$ $(i=0, \ldots, k-1)$. According to our induction hypothesis, this yields that $p=q$.

Now take an arbitrary type F and an *fr*-embedding $\beta: T_F \to T_G$. Then $\gamma=\alpha\beta$ is an *fr*-embedding of T_F into T_G as well. For the sake of simplicity introduce the notations

$$t_1 = g(h, \ldots, h, g(h, \ldots, h))$$

and

$$t_2 = g(h, \ldots, h, g(h, \ldots, h), g(h, \ldots, h)).$$

Then

$$\mathrm{sub}\,(t_i) \cap \gamma(T_F) = \emptyset \quad (i = 1, 2). \tag{13}$$

Moreover, for every $p \in \gamma(T_F)$

$$t_i \notin \mathrm{sub}\,(p) \quad (i = 1, 2). \tag{14}$$

Let $T \subseteqq T_{F,n}$ be a regular forest which can be obtained from the trees $p_0, \ldots, p_{k-1} \in T_{F,m}$ by finitely many applications of regular operations. According to (14), $\gamma(p_0), \ldots, \gamma(p_{k-1})$ can be represented by the m-ary G-automata $\mathfrak{A}_0, \ldots, \mathfrak{A}_{k-1}$ such that on the algebras $\mathfrak{A}_0, \ldots, \mathfrak{A}_{k-1}$ the equation $t_1=t_2$ holds ([1] lemma 2).

Note that the power set of $\gamma(T_F)$ is closed under the regular operations, that is if $T_1, T_2 \subseteqq (T_F)$ then $T_1 \cup T_2$, $T_1 x_i T_2$ and $T_1^{*,i} \subseteqq (T_F)$ as well. Indeed,

$$T_1 \cup T_2 = \gamma(\bar{\gamma}^1(T_1) \cup \bar{\gamma}^1(T_2)) \subseteqq \gamma(T_F), \tag{15}$$

$$T_1 x_i T_2 = \gamma(\bar{\gamma}^1(T_1) x_i \bar{\gamma}^1(T_2)) \subseteqq \gamma(T_F), \tag{16}$$

$$T_1^{*,i} = \gamma(\bar{\gamma}^1(T_1)^{*,i}) \subseteqq \gamma(T_F). \tag{17}$$

Therefore, for every forest $T' \subseteq T_G$ which can be obtained from $\gamma(p_0), \ldots, \gamma(p_{k-1})$ by finitely many applications of regular operations we have

$$\operatorname{sub}(t_i) \cap T' = \emptyset \quad (i = 1, 2).$$

By lemmas 3, 4 and 5 of [1] if for the forests T_1 and T_2

1) $\operatorname{sub}(t_i) \cap T_j = \emptyset$ $(i, j = 1, 2)$, and

2) T_1 and T_2 can be recognized by the tree automata $\overline{\mathfrak{A}}_1$ and $\overline{\mathfrak{A}}_2$, respectively, such that on the algebras $\mathfrak{A}_1$ and $\mathfrak{A}_2$ $t_1 = t_2$ holds,

then the forests $T_1 \cup T_2$, $T_1 x_i T_2$ and $T_1^{*,i}$ can be represented by the tree automata $\mathfrak{B}_1$, $\mathfrak{B}_2$ and $\mathfrak{B}_3$, respectively, such that on the algebras $\mathfrak{B}_i$ $(i = 1, 2, 3)$ $t_1 = t_2$ holds as well.

From this and from statements (14)—(17) we get, that every forest which can be obtained from $\gamma(p_0), \ldots, \gamma(p_{k-1})$ by finitely many applications of regular operations (among them $\gamma(T)$) can be represented by a G-automaton belonging to the variety defined by the equation (11). This ends the proof of our theorem.

From the above theorem we can see that the existence of a 0-ary and an at least binary operational symbols in the type G is sufficient for the existence of a proper fr-complete variety. But, by Theorem 3 it is necessary as well. Therefore, the simplest types over which there exist fr-complete varieties are those which consist of exatly one 0-ary and one at least binary operational symbols.

By the languages over the alphabet $X = \{x_0, \ldots, x_{n-1}\}$ accepted by an n-ary F-automaton $\overline{\mathfrak{A}}$ we mean

$$L(\overline{\mathfrak{A}}) = \{fr(p) \mid p \in T(\overline{\mathfrak{A}})\}.$$

In [2] it was shown by F. Gécseg and G. Horváth that there exists a proper variety over the type $G = \{g, h\}$ with $r(g) = 2$ and $r(h) = 0$ such that every context-free language can be accepted by a finite tree automaton belonging to this variety. This result directly follows from Theorem 5.

4. *Fr*-equivalence of tree automata

In [1] F. Gécseg introduced the concept of *rational equivalence* of tree automata. Namely, two tree automata $\overline{\mathfrak{A}}$ and $\overline{\mathfrak{B}}$ (not necessarily of the same type) are called rationally equivalent if for every forest T, T can be rationally represented by $\overline{\mathfrak{A}}$ if and only if T can be rationally represented by $\overline{\mathfrak{B}}$. Now we define the analogous concept for fr-representation. We call two tree automata $\overline{\mathfrak{A}}$ and $\overline{\mathfrak{B}}$ *fr-equivalent* if the class of forests fr-representable by $\overline{\mathfrak{A}}$ is equal to the class of all those forests, which can be fr-represented by $\overline{\mathfrak{B}}$.

One can naturally raise the following questions:

1) Is the rational equivalence of tree automata decidable? In other words, does there exist an algorithm to decide for arbitrary two tree automata whether they are rationally equivalent or not?

2) Is the fr-equivalence of tree automata decidable?

In this section we give positive answers to each of these questions.

We shall need the following two simple lemmas.

Lemma 1. Let $\alpha\colon T_F \to T_F$ be an *fr*-embedding and assume that there exists a forest $T \subseteq T_F$ such that $\alpha(T)=T$. Then for each $p \in T$ we have

$$d(\alpha(p)) = d(p).$$

Proof. For every natural number n let

$$T_n = \{p \mid p \in T \text{ and } d(p) = n\}.$$

We shaw that for every n, $\alpha(T_n)=T_n$. Indeed, let n_0 be the least natural number with $T_{n_0} \neq \emptyset$. If $q \in T_{n_0}$ then $\alpha^{-1}(q) \in T$ and $d(\alpha^{-1}(q)) \leqq n_0$ which implies that $\alpha^{-1}(q) \in T_{n_0}$. Therefore, $\alpha^{-1}(T_{n_0}) \subseteq T_{n_0}$. But α^{-1} is one-to-one and T_{n_0} is finite. Thus the restriction of α^{-1} to T_{n_0} is onto, i.e., $\alpha^{-1}(T_{n_0})=T_{n_0}$. Hence $\alpha(T_{n_0})=T_{n_0}$. Now take an arbitrary natural number n such that $T_n \neq \emptyset$ and assume that for every $m<n$, $\alpha(T_m)=T_m$. For each $q \in T_n$ we have $d(\alpha^{-1}(q)) \leqq n$. If $d(\alpha^{-1}(q))<n$ then $\alpha^{-1}(q) \in T_m$ for some $m<n$ implying $q \in T_m$, which is impossible. Therefore, $d(\alpha^{-1}(q))=n$, or equivalently $\alpha^{-1}(q) \in T_n$. Finally, again from the finiteness of T_n we get that $\alpha(T_n)=T_n$. □

Consider the types F and G. We call the mapping γ of F onto G a *projection* if γ preserves arity. If we have an *fr*-homomorphism $\alpha\colon T_F \to T_G$ such that

1) for every $f \in F$, $d(\alpha(f))=1$,

2) for every $f \in F$, $\alpha(f)$ has exatly $r(f)$ leaves,

3) for every $g \in G$, $g(x_0, \ldots, x_{r(g)-1}) \in \alpha(T_F)$, then we can take the projection $\gamma\colon F \to G$ for which $\gamma(f)=g$ if and only if $\alpha(f(x_0, \ldots, x_{r(f)-1}))=g(x_0, \ldots, x_{r(f)-1})$. For this we use the notation $\gamma=\alpha\restriction F$.

The next result is obvious.

Lemma 2. Take three *fr*-embeddings $\alpha\colon T_F \to T_G$, $\beta\colon T_G \to T_H$ and $\gamma\colon T_F \to T_H$ such that $\gamma=\beta\alpha$. Then $\gamma\restriction F$ is a projection if and only if $\alpha\restriction F$ and $\beta\restriction G$ are projections as well.

Consider an F-automaton $\mathfrak{A}$ and a G-automaton $\mathfrak{B}$. We say that $\mathfrak{A}$ and $\mathfrak{B}$ are *equivalent up to the notation of their operational symbols* if there exists a one-to-one projection γ of F onto G such that $\gamma(T(\mathfrak{A}))=T(\mathfrak{B})$. Moreover, we use the terminology that *F is reduced for $\mathfrak{A}$* if for every $f \in F$ there is a tree p in $T(\mathfrak{A})$ such that f occurs in p.

Theorem 6. Take an F-automaton $\mathfrak{A}$ and a G-automaton $\mathfrak{B}$ such that F and G are reduced for $\mathfrak{A}$ and $\mathfrak{B}$, respectively. Then the following three conditions are equivalent:

1) $\mathfrak{A}$ and $\mathfrak{B}$ are rationally equivalent,

2) $\mathfrak{A}$ and $\mathfrak{B}$ are *fr*-equivalent,

3) $\mathfrak{A}$ and $\mathfrak{B}$ are equivalent up to the notation of their operational symbols.

Proof. The equivalence of 1) and 3) was proved in [1]. Furthermore, it is obvious that 3) implies 2). Thus it is enough to show that 3) follows form 2).

First we prove, that if for an *fr*-embedding $\alpha\colon T_F \to T_F$ there exists a $q \in T_F$ such that $\alpha(q)=q$, than for every operational symbol f occuring in q we have $\alpha(f)=f$. Indeed, if $d(q) \leqq 1$ then this statement is trivial. Now let $q=$

$=f(q_0, \ldots, q_{k-1})$ and assume that for every tree q' with $d(q')<d(q)$ our statement is true. From $\alpha(q)=q$ we get

$$\alpha(f)(\alpha(q_0), \ldots, \alpha(q_{k-1})) = f(q_0, \ldots, q_{k-1}).$$

But this yields that $\alpha(f)=f$ and that $\alpha(q_i)=q_i$ $(i=0, \ldots, k-1)$.

Now take an F-automaton $\mathfrak{A}$ and a G-automaton $\mathfrak{B}$ such that F and G are reduced for $\mathfrak{A}$ and $\mathfrak{B}$, respectively. Assume that $\mathfrak{A}$ and $\mathfrak{B}$ are *fr*-equivalent. Then there exist two *fr*-embeddings $\alpha: T_F \to T_G$ and $\beta: T_G \to T_F$ such that $\alpha(T(\mathfrak{A}))= =T(\mathfrak{B})$ and $\beta(T(\mathfrak{B}))=T(\mathfrak{A})$. Therefore, for the *fr*-embedding $\gamma=\beta\alpha$ we have $\gamma(T(\mathfrak{A}))=T(\mathfrak{A})$. Thus, by Lemma 1, γ preserves the depth of trees in $T(\mathfrak{A})$. For the sake of simplicity let us denote $T(\mathfrak{A})$ by T.

Consider the trees $p_0, \ldots, p_{m-1} \in T$ such that for every $f \in F$ there exists a $j \in \tilde{m}$ for which f occurs in p_j. Let $d(p_0)=n_0, \ldots, d(p_{m-1})=n_{m-1}$. Therefore, $p_j \in T_{n_j}$ $(j=0, \ldots, m-1)$. (We recall that T_{n_j} is the set of all trees from T whose depth is n_j.) Let

$$\gamma_j = \gamma \upharpoonright T_{n_j} \quad (j = 0, \ldots, m-1).$$

Since T_{n_j} is finite and γ_j is one-to-one thus there exist natural numbers $k_0, \ldots, k_{m-1}$ such that

$$\gamma_j^{k_j} = \mathrm{id}_{T_{n_j}} \quad (j = 0, \ldots, m-1). \tag{18}$$

Take $d=k_0 \ldots k_{m-1}$. From (18) it follows that

$$\gamma^d \upharpoonright (T_{n_0} \cup \ldots \cup T_{n_{m-1}}) = \mathrm{id}_{T_{n_0} \cup \ldots \cup T_{n_{m-1}}}.$$

Therefore, for the *fr*-embedding $\gamma^d: T_F \to T_F$ we have

$$\gamma^d(p_j) = p_j \quad (j = 0, \ldots, m-1).$$

Because of the choice of the trees $p_0, \ldots, p_{m-1}$ the first assertion of this proof yields that $\gamma^d \upharpoonright F = \mathrm{id}_F$. Thus $\gamma^d \upharpoonright F$ is a one-to-one projection of F onto F, but by Lemma 2 this is true if and only if $\gamma \upharpoonright F$ is a projection of F onto F as well. Then Lemma 2, $\gamma=\beta\alpha$ and the fact that $\gamma \upharpoonright F$ is a projection jointly imply that $\alpha \upharpoonright F$ is a projection of F onto G. The proof is complete.

According to the above theorem in order to decide the rational equivalence (*fr*-equivalence) of arbitrary two tree automata $\mathfrak{A}$ and $\mathfrak{B}$ it is enough to check whether there exists a one-to-one projection γ between the types of $\mathfrak{A}$ and $\mathfrak{B}$ such that $\gamma(T(\mathfrak{A}))=T(\mathfrak{B})$. But the set of all one-to-one projections between two finite types is finite, and for a given one-to-one projection γ the equality $\gamma(T(\mathfrak{A}))=T(\mathfrak{B})$ is decidable by taking the minimal tree automata recognizing $\gamma(T(\mathfrak{A}))$ and $T(\mathfrak{B})$. Thus we have

Theorem 7. The rational equivalence and the *fr*-equivalence of tree automata are decidable.

DEPT. OF COMPUTER SCIENCE
A. JÓZSEF UNIVERSITY
H-6720 SZEGED, HUNGARY
SOMOGYI U. 7.

References

[1] GÉCSEG, F., Universal algebras and tree automata, *Proceedings of the Conference on Fundamentals of Computation Theory*, Poznan-Kórnik v. 56, 1977, p. 89.
[2] GÉCSEG, F. and GY. HORVÁTH, On representation of trees of context-free languages by tree automata, *Foundations of Control Engineering*, v. 1, 1976, pp. 161—168.
[3] GRÄTZER, G., *Universal Algebras*, D. Van Nostrand Company, Princeton, New Yersey, 1968.

(Received Nov. 22, 1977)