Theoretical Computer Science 99 (1992) 177–211
Elsevier

Fundamental Study

Alphabetic tree relations

Symeon Bozapalidis
Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece

Communicated by D. Perrin
Received May 1990
Revised November 1990

Abstract

Bozapalidis, S., Alphabetic tree relations, Theoretical Computer Science 99 (1992) 177–211.

We study the class of tree transductions induced by bimorphisms (φ, R, φ') with φ, φ' alphabetic homomorphisms and R a recognizable forest; this class contains many of classical tree-transformations such as union and intersection with a recognizable forest, a-product, a-quotient, top-catenation, branches, subtrees, initial and terminal subtrees, largest common initial subtree, etc.

Furthermore, the considered transductions are closed under composition and inversion and preserve the recognizable and algebraic forests; by applying the last fact to the classical tree transformations cited above, we obtain a series of remarkable results.

We show that Takahashi's relations $A \subseteq T_\Sigma \times T_\Gamma$ can be identified with the squeleton-preserving $\Sigma \nabla \Gamma$-recognizable subsets of $T_\Sigma \times T_\Gamma$.

Finally, we give a classification of some remarkable subclasses of the class of alphabetic transductions.

Contents

Introduction

The importance of rational relations in word language theory is due to the following three fundamental facts:

(i) they are closed under composition and inversion,

(ii) they preserve rational and algebraic languages, and

(iii) they contain almost all the elementary operations on words.

The key to establish most of the results in this domain is Nivat's theorem which states that a relation is rational iff it can be represented by a bimorphism. On the other hand, it is well known that the rational relations can be obtained as the behaviors of a certain type of machines, called rational transducers (see [3, 8]).

At the level of trees it is again very important to impose a class of relations analogous to (i)–(iii) above.

Many authors have studied several classes of tree relations adopting either the bimorphism point of view [7] or the machine point of view (tree transducers [2, 9, 10, 12, 14]). In any case, the considered classes are rarely closed under composition and inversion and do not preserve in general the algebraic forests. On the other hand, Takahashi's relations [13] have properties (i) and (ii) but they cannot describe the elementary tree operations such as top-treecatenation, branches, etc.

Arnold and Dauchet proved in [1] that the "démarquages linéaires" (here called alphabetic homomorphisms) constitute perhaps the largest class of tree homomorphisms reflecting algebraic forests. Consequently, the bimorphisms (φ, R, φ'), with φ, φ' alphabetic, define perhaps the largest class of tree transformations preserving algebraic forests and closed under inversion. The so-defined class $\mathscr{Alph}$ (of alphabetic tree transductions) is shown to be closed under composition and it is incomparable with all the important classes of tree transductions. The following classical tree operations belong to $\mathscr{Alph}$: union and intersection with a recognizable forest, the recognizable constants, top-treecatenation, a-product and a-quotient of forests, branches, subtrees, initial and terminal subtrees, largest common initial subtree, finite unions of products, etc.

In the last section we identify Takahashi's relations with certain squeleton-preserving recognizable subsets of $T_\Sigma \times T_I$. We finally classify remarkable subclasses of $\mathscr{Alph}$.

1. Preliminaries

1.1. Alphabetic homomorphisms

As usual, T_Σ denotes the set of trees over the (finite) ranked alphabet Σ and $T_\Sigma(x_1, \ldots, x_n)$ is the set of all trees indexed by the variables $x_1, \ldots, x_n$.

For $k \geqslant 0$, $m \geqslant 0$, $t \in T_\Sigma(x_1, \ldots, x_k)$ and $t_1, \ldots, t_k \in T_\Sigma(x_1, \ldots, x_m)$, we denote by $t(t_1, \ldots, t_n)$ the result of substituting t_i for x_i in t.

An algebraic tree grammar is a 4-tuple $G=(\Sigma, F, P, S)$, where Σ is a finite ranked alphabet of terminals, F is a finite ranked alphabet of nonterminals or function symbols ($\Sigma \cap F \neq \emptyset$), P is a finite set of rules of the form $\Phi(x_1, \ldots, x_n) \rightarrow t$ with $\Phi \in F_n$ and $t \in T_{\Sigma \cup F}(x_1, \ldots, x_n)$, and $S \in F_0$ is the axiom of G.

Let $n \geqslant 0$ and $t_1, t_2 \in T_{\Sigma \cup F}(x_1, \ldots, x_n)$; we put $t_1 \underset{G}{\Rightarrow} t_2$ iff there is a rule $\Phi(x_1 \ldots x_k) \rightarrow t$, a tree $\eta \in T_{\Sigma \cup F}(x_1, \ldots, x_n, x_{n+1})$ containing exactly one occurrence of x_{n+1}, and trees $\xi_1, \ldots, \xi_k \in T_{\Sigma \cup F}(x_1, \ldots, x_n)$ such that

$$t_1 = \eta(x_1, \ldots, x_n, \Phi(\xi_1, \ldots, \xi_k)),$$

$$t_2 = \eta(x_1, \ldots, x_n, t(\xi_1, \ldots, \xi_k)).$$

$\overset{*}{\underset{G}{\Rightarrow}}$ denotes the reflexive and transitive closure of $\underset{G}{\Rightarrow}$.

The tree language generated by G is

$$L(G) = \{t \in T_\Sigma \mid S \overset{*}{\underset{G}{\Rightarrow}} t\}.$$

Call $L \subseteq T_\Sigma$ algebraic if $L = L(G)$ for some algebraic tree grammar G.

It should be noted for completeness sake that there is another kind of derivation, the so-called IO-derivation (see [11]), leading to another class of tree languages. Explicitly, $t_1 \underset{\mathrm{IO}}{\Rightarrow} t_2$ is defined to have the same meaning as $t_1 \Rightarrow t_2$ except that the ξ_i's are required to be terminal trees, that is $\xi_1, \ldots, \xi_k \in T_\Sigma(x_1, \ldots, x_n)$.

Given now finite ranked alphabets Σ and Γ, a homomorphism from T_Σ to T_Γ is a function φ which to every symbol $\sigma \in \Sigma_n$ corresponds a tree $t(x_{i_1} \ldots x_{i_p}) \in T_\Gamma(x_1, \ldots, x_n)$.

φ is inductively extended to a function $\varphi: T_\Sigma \rightarrow T_\Gamma$ by setting

$$\varphi(\sigma t_1 \ldots t_n) = t(\varphi(t_{i_1}), \ldots, \varphi(t_{i_p})),$$

where $\varphi(\sigma) = t(x_{i_1} \ldots x_{i_p})$.

A homomorphism φ is called linear if for any $\sigma \in \Sigma_n$ the variables $x_1, \ldots, x_n$ appear at most once in the tree $\varphi(\sigma)$. A linear homomorphism $\varphi: T_\Sigma \rightarrow T_\Gamma$ is alphabetic if for each $\sigma \in \Sigma_n$ either

$$\varphi(\sigma) = \gamma(x_{j_1} \ldots x_{j_p}), \quad \gamma \in \Gamma_p \ (p \geqslant 0)$$

or

$$\varphi(\sigma) = x_k, \quad 1 \leqslant k \leqslant n.$$

Finally, we say that $\varphi: T_\Sigma \rightarrow T_\Gamma$ is strictly alphabetic if

$$\varphi(\Sigma_n) \subseteq \Gamma_n, \quad n = 0, 1, 2, \ldots$$

Proposition 1.1 ([1, Theorem 4.1]). *The class of algebraic forests is closed under inverse alphabetic homomorphisms.*

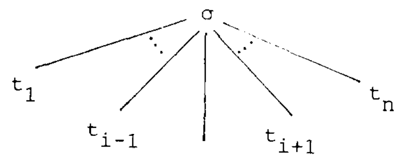

Fig. 1.

This result led us to consider bimorphisms (φ, R, φ') with φ, φ' alphabetic, in order to get a class of tree transductions preserving algebraic forests.

1.2. Monoids

Next we denote by P_Σ the subset of $T_\Sigma(x)$ consisting of all trees with just one occurrence of the variable x. P_Σ becomes a monoid if we define its multiplication to be the substitution at x; this monoid is free, spanned by the elements shown in Fig. 1, with $\sigma \in \Sigma_n$ and $t_j \in T_\Sigma$.

For every $t \in T_\Sigma$ and $\tau \in P_\Sigma$, $t\tau$ is the tree (of T_Σ) obtained by replacing the variable x in τ by t. This operation is actually an action

$$T_\Sigma \times P_\Sigma \to T_\Sigma, \quad (t, \tau) \mapsto t\tau.$$

A Σ-tree automaton $\mathscr{A} = (Q, a, F)$ consists of a finite set Q (the states), a subset $F \subseteq Q$ (the final states) and a Σ-indexed family of functions

$$a_\sigma : Q^n \to Q, \quad \sigma \in \Sigma_n.$$

For $n = 0$, the elements $a_c \in Q$ $(c \in \Sigma_0)$ are the constants of $\mathscr{A}$. There is a function $h_{\mathscr{A}} : T_\Sigma \to Q$ defined inductively by

$$h_{\mathscr{A}}(\sigma t_1 \ldots t_n) = a_\sigma(h_{\mathscr{A}} t_1, \ldots, h_{\mathscr{A}} t_n), \quad \sigma \in \Sigma_n, \ t_j \in T_\Sigma.$$

The behavior of $\mathscr{A}$ is then the tree language $|\mathscr{A}| = h_{\mathscr{A}}^{-1}(F)$.

Call $L \subseteq T_\Sigma$ recognizable if $L = |\mathscr{A}|$ for some tree automaton $\mathscr{A}$. We denote by $\mathrm{Rec}(T_\Sigma)$ the class of all recognizable languages of T_Σ. The monoid P_Σ acts on $\mathscr{A}$ via

$$q \cdot \sigma t_1 \ldots t_{i-1} x t_{i+1} \ldots t_n = a_\sigma(h_{\mathscr{A}} t_1, \ldots, h_{\mathscr{A}} t_{i-1}, q, h_{\mathscr{A}} t_{i+1}, \ldots, h_{\mathscr{A}} t_n)$$

$$q(\tau\pi) = (q\tau)\pi, \ q \in Q, \ \tau, \pi \in P_\Sigma.$$

Later on, we shall need the (free) submonoid L_Σ of P_Σ spanned by the elements shown in Fig. 2, with $\sigma \in \Sigma_n$ and $t_j \in T_\Sigma$.

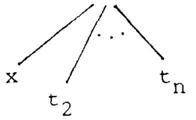

Fig. 2.

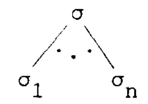

Fig. 3.

The product monoid $L_\Sigma \times L_\Gamma$ acts then on $T_\Sigma \times T_\Gamma$ and in [5] it is shown that the following proposition holds.

Proposition 1.2. *If a relation $A \subseteq T_\Sigma \times T_\Gamma$ can be written in the form*

$$A = \bigcup_{i=1}^{n} B_i \times C_i, \quad B_i \in \mathrm{Rec}(T_\Sigma),\ C_i \in \mathrm{Rec}(T_\Gamma)$$

Then for any $\langle c, d \rangle \in \Sigma_0 \times \Gamma_0$, the set

$$\langle c, d \rangle^{-1} A = \{(\tau, \pi) \in L_\Sigma \times L_\Gamma \mid (c\tau, d\pi) \in A\}$$

is a recognizable subset of $L_\Sigma \times L_\Gamma$.

1.3. Local forests

A transition from a ranked alphabet Σ is a tuple

$$(\sigma, \sigma_1, \ldots, \sigma_n)$$

frequently denoted as shown in Fig. 3, with $\sigma \in \Sigma_n$ and $\sigma_j \in \Sigma$.

We say that the transition $(\sigma, \sigma_1, \ldots, \sigma_n)$ appears inside the tree $t \in T_\Sigma$ if

$$t = (\sigma t_1 \ldots t_n) \cdot \tau,$$

where $\tau \in P_\Sigma$ and for every i $(1 \leqslant i \leqslant n)$, the root of t_i is σ_i. Let

$$E_0 \subseteq \Sigma_0, \qquad E \subseteq \Sigma$$

and let T be a set of transitions from Σ; the local forest generated by the system (E_0, E, T) is the set of all trees $t \in T_\Sigma$ with the following three properties:

(l_1) the root of t belongs to E,

(l_2) all the leaves of t belong to E_0, and

(l_3) all transitions of t belong to T.

Clearly, any local forest is recognizable and every recognizable forest is the projection of a local forest via a strictly alphabetic homomorphism.

Finally, $[k]$ denotes the set $\{1, 2, \ldots, k\}$.

2. The alphabet $\Sigma \vee_k \Gamma$.

Trying to "join" trees with different squeletons (for reasons explained in subsequent sections), we are led to introduce a new operation between ranked alphabets, that consists of concatenating symbols with different ranks; the rank of the formed pairs is the maximum of the ranks of the participating symbols, so that the resulting alphabet has a supremum-like property.

Recall that, for a given finite ranked alphabet Σ, its degree is the biggest natural number N satisfying $\Sigma_N \neq \emptyset$. Let $k \geqslant \deg \Sigma$; from Σ we construct the ranked alphabet $\Sigma^{[k]}$ in the following way:

$$\Sigma_0^{[k]} = \Sigma_0,$$

whereas for $n \geqslant 1$

$$\Sigma_n^{[k]} = \{\sigma_{i_1, \ldots, i_p} \mid \sigma \in \Sigma_p,\ i_1, \ldots, i_p \text{ are distinct elements of } [k] \text{ and } \max\{i_1, \ldots, i_p\} = n\} \cup \{n\}.$$

Example 2.1. Take $\Sigma_0 = \{a\}$, $\Sigma_1 = \{\tau\}$, $\Sigma_2 = \{\sigma\}$ and $\Sigma_n = \emptyset$ $(n \geqslant 3)$; then

$$\begin{aligned}
\Sigma_0^{[4]} &= \{a\},\\
\Sigma_1^{[4]} &= \{\tau_1, 1\},\\
\Sigma_2^{[4]} &= \{\tau_2, \sigma_{12}, \sigma_{21}, 2\},\\
\Sigma_3^{[4]} &= \{\tau_3, \sigma_{13}, \sigma_{31}, \sigma_{23}, \sigma_{32}, 3\},\\
\Sigma_4^{[4]} &= \{\tau_4, \sigma_{14}, \sigma_{41}, \sigma_{24}, \sigma_{42}, \sigma_{34}, \sigma_{43}, 4\},\\
\Sigma_n^{[4]} &= \emptyset \quad \text{for } n > 4,
\end{aligned}$$

and the tree in Fig. 4 is a tree in $T_{\Sigma^{[4]}}$.

Consider, further, two alphabets Σ, Γ and a natural number $k \geqslant \max\{\deg \Sigma, \deg \Gamma\}$. We define their k-supremum $\Sigma \vee_k \Gamma$ to be

$$(\Sigma \vee_k \Gamma)_0 = \Sigma_0 \times \Gamma_0,$$

$$(\Sigma \vee_k \Gamma)_n = \bigcup_{\max(i,j) = n} \Sigma_i^{[k]} \times \Gamma_j^{[k]}.$$

We simply write $\Sigma \vee \Gamma$ in the case

$$k = \max\{\deg \Sigma, \deg \Gamma\}.$$

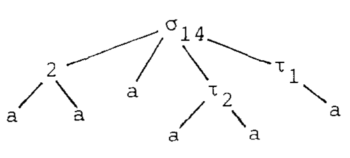

Fig. 4.

There are two (canonical) alphabetic homomorphisms

$$\varphi_\Sigma : T_{\Sigma \vee_k \Gamma} \to T_\Sigma, \qquad \varphi_\Gamma : T_{\Sigma \vee_k \Gamma} \to T_\Gamma$$

with

$$\varphi_\Sigma(\langle \sigma_{i_1 \ldots i_p}, \omega \rangle) = \sigma(x_{i_1} \ldots x_{i_p}), \qquad \varphi_\Sigma(\langle n, \omega \rangle) = x_n,$$

$$\varphi_\Gamma(\langle \omega, \gamma_{j_1 \ldots j_q} \rangle) = \gamma(x_{j_1} \ldots x_{j_q}), \qquad \varphi_\Gamma(\langle \omega, n \rangle) = x_n.$$

Example 2.2. Take

$$\Sigma_0 = \{a\}, \qquad \Sigma_1 = \{\tau\}, \qquad \Sigma_2 = \{\sigma\}, \qquad \Sigma_n = \emptyset,$$

$$\Gamma_0 = \{c\}, \qquad \Gamma_2 = \{\gamma\}, \qquad \Gamma_n = \emptyset \text{ for } n \neq 0, 2.$$

Figure 5 visualizes the action of φ_Σ and φ_Γ.

Proposition 2.3. *For every pair of alphabetic homomorphisms shown in Fig. 6, there exists a unique alphabetic homomorphism*

$$h : T_\Delta \to T_{\Sigma \vee_k \Gamma}, \quad k = \max(\deg \Delta, \deg \Sigma, \deg \Gamma),$$

making commutative the triangle shown in Fig. 7.

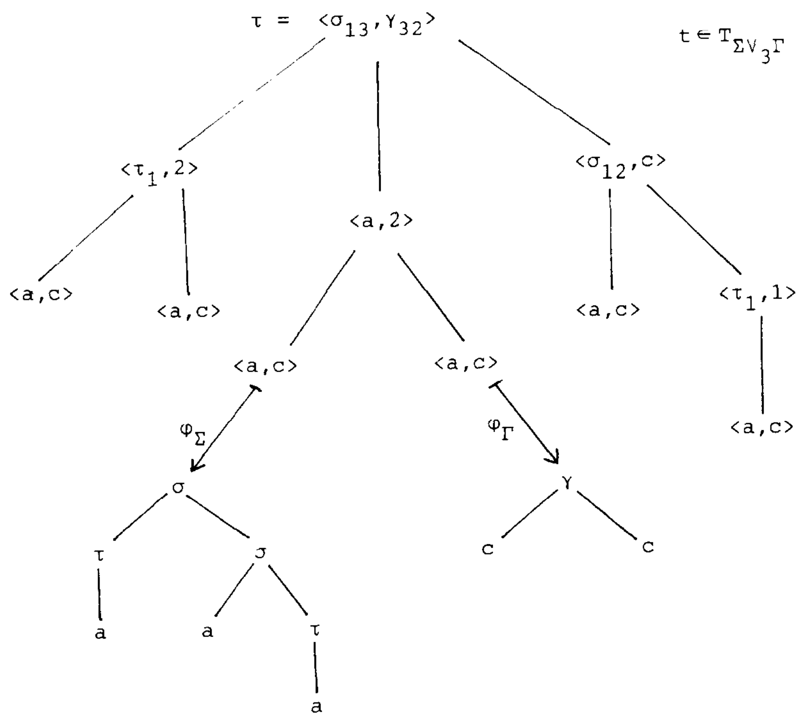

Fig. 5.

$$\begin{tikzcd} & T_\Delta \arrow[dl, "\varphi"'] \arrow[dr, "\varphi'"] & \\ T_\Sigma & & T_\Gamma \end{tikzcd}$$

Fig. 6.

$$\begin{tikzcd} T_\Delta \arrow[rr, "h"] \arrow[dr, "(\varphi,\varphi')"'] & & T_{\Sigma \vee_k \Gamma} \arrow[dl, "(\varphi_\Sigma,\varphi_\Gamma)"] \\ & T_\Sigma \times T_\Gamma & \end{tikzcd}$$

Fig. 7.

Proof. For every $\delta \in \Delta_n$ $(n \geqslant 0)$, suppose that

$$\varphi(\delta) = \sigma(x_{i_1} \dots x_{i_p}) \quad \text{and} \quad \varphi'(\delta) = \gamma(x_{j_1} \dots x_{j_q}).$$

We then put

$$h(\delta) = \langle \sigma_{i_1 \dots i_p}, \gamma_{j_1 \dots j_q} \rangle.$$

In the case

$$\varphi(\delta) = x_k, \qquad \varphi'(\delta) = x_\lambda, \text{ etc.,}$$

we put

$$h(\delta) = \langle k, \lambda \rangle, \text{ etc.}$$

Convention. Frequently, if no confusion is caused, we identify the symbol $\sigma_{12\dots n}$ with σ.

3. Alphabetic relations

A relation $A \subseteq T_\Sigma \times T_\Gamma$ is called alphabetic if there exists a ranked alphabet Δ, a recognizable forest $L \subseteq T_\Delta$ and two alphabetic homomorphisms

$$T_\Sigma \xleftarrow{\varphi} T_\Delta \xrightarrow{\varphi'} T_\Gamma$$

such that

$$A = \{(\varphi t, \varphi' t) \mid t \in L\}.$$

Proposition 3.1. *$A \subseteq T_\Sigma \times T_\Gamma$ is alphabetic iff there exists a recognizable forest*

$$R \subseteq T_{\Sigma \vee_k \Gamma}$$

so that

$$A=\{(\varphi_\Sigma t, \varphi_\Gamma t) \mid t\in R\}.$$

Furthermore, we can choose R to be local.

Proof. Using Proposition 2.3, we get

$$A=\{(\varphi_\Sigma t, \varphi_\Gamma t) \mid t\in h(L)\}$$

with $h(L)\subseteq T_{\Sigma\vee_k\Gamma}$ recognizable. □

A tree transduction $\tau: T_\Sigma \to T_\Gamma$ is alphabetic if its graph (denoted by $\#\tau$) is an alphabetic relation.

Proposition 3.1 can be now restated as follows.

Proposition 3.1′. *$\tau: T_\Sigma\to T_\Gamma$ is alphabetic iff it admits a factorization of the form*

$$T_\Sigma \xrightarrow{\varphi_\Sigma^{-1}} T_{\Sigma\vee_k\Gamma} \xrightarrow{-\cap R} T_{\Sigma\vee_k\Gamma} \xrightarrow{\varphi_\Gamma} T_\Gamma,$$

with $R\in\mathrm{Rec}(T_{\Sigma\vee_k\Gamma})$.

As usual, we do not distinguish a transduction τ from its graph $\#\tau$. $\mathcal{A}\ell ph(\Sigma,\Gamma)$ is the class of all alphabetic transductions from T_Σ to T_Γ.

Proposition 3.2. *$\mathcal{A}\ell ph$ is closed under finite union and inversion (in the sense of relations).*

Proof. Our second assertion comes from the definition. For the first one, let

$$A_i=h_{\Sigma,\Gamma}(L_i),\quad L_i\in\mathrm{Rec}(T_{\Sigma\vee_k\Gamma}),\ i=1,2,$$

where the function $h_{\Sigma,\Gamma}: T_{\Sigma\vee_k\Gamma}\to T_\Sigma\times T_\Gamma$ is defined by

$$h_{\Sigma,\Gamma}(\omega)=(\varphi_\Sigma\omega, \varphi_\Gamma\omega).$$

We then have

$$A_1\cup A_2=h_{\Sigma,\Gamma}(L_1)\cup h_{\Sigma,\Gamma}(L_2)=h_{\Sigma,\Gamma}(L_1\cup L_2),$$

and so $A_1\cup A_2\in\mathcal{A}\ell ph(\Sigma,\Gamma)$. □

Proposition 3.3. *$\mathcal{A}\ell ph(\Sigma,\Gamma)$ contains all relations of the form*

$$\bigcup_{i=1}^{n} B_i\times C_i,$$

with $B_i\in\mathrm{Rec}(T_\Sigma)$ *and* $C_i\in\mathrm{Rec}(T_\Gamma)$ $(i=1,2,\ldots,n)$.

Proof. The result comes from the equality

$$B_i \times C_i = h_{\Sigma,\Gamma}(\varphi_\Sigma^{-1}(B_i) \cap \varphi_\Gamma^{-1}(C_i))$$

and the Proposition 3.2. $\square$

Proposition 3.4. *The following relations are alphabetic:*
(i) *the diagonal* $\Delta \subseteq T_\Sigma \times T_\Sigma$,
(ii) *intersection with a recognizable forest, and*
(iii) *union with a recognizable forest.*

Proof. Taking $\varphi = \varphi' =$ identity in the definition of an alphabetic relation, we get (ii).
(i) is a special case of (ii).
(iii) For $L \in \mathrm{Rec}(T_\Sigma)$, the graph of the transduction $t \mapsto t \cup L$ is

$$\Delta \cup (T_\Sigma \times L) \in \mathcal{A}\ell p h(\Sigma, \Sigma). \quad \square$$

It is well known that the first operation on words, the concatenation, is a rational relation; thus, it is natural to ask if the top-catenation of trees has an analogous property.

We need here to extend the definition of alphabetic relation to n arguments. Precisely, assume that ranked alphabets $\Sigma_1, \ldots, \Sigma_n, \Gamma$ are given; a relation

$$A \subseteq T_{\Sigma_1} \times \cdots \times T_{\Sigma_n} \times T_\Gamma$$

is termed alphabetic if we can determine a ranked alphabet Δ and alphabetic homomorphisms $\varphi_i : T_\Delta \to T_{\Sigma_i}$, $\varphi' : T_\Delta \to T_\Gamma$ in such a manner that

$$A = \{(\varphi_1 t, \ldots, \varphi_n t, \varphi' t) \mid t \in L\},$$

where L is a recognizable forest of T_Δ.

We also need the following auxiliary result.

Lemma 3.5. *Let K be a recognizable forest of T_Σ and $\Omega \subseteq \Sigma$; we denote by $\Omega\langle K \rangle$ the smallest subset of T_Σ containing K and having the property*

$$\omega \in \Omega \quad \text{and} \quad t_1, \ldots, t_n \in \Omega\langle K \rangle \Rightarrow \omega t_1 \ldots t_n \in \Omega\langle K \rangle.$$

Then $\Omega\langle K \rangle$ is recognizable.

Proof. Suppose that K is generated by the regular tree grammar $G = (\Sigma, V, S, R)$; then $\Omega\langle K \rangle$ is generated by the tree grammar G' obtained from G by adding the new rules

$$S \to \omega(S \ldots S), \quad \omega \in \Omega,$$

where S is the axiom of G.

Proposition 3.6. *For a fixed symbol $\tau \in \Sigma_n$, the function*

$$(1) \qquad (t_1, \ldots, t_n) \mapsto \tau t_1 \ldots t_n$$

is alphabetic.

Proof. Consider the alphabet

$$\Sigma \vee \cdots \vee \Sigma \vee \Sigma \quad (n+1) \text{ times}$$

and the recognizable forests

$$\Omega_i \langle K_i \rangle \subseteq T_{\Sigma \vee \cdots \vee \Sigma \vee \Sigma}$$

determined by

$$\Omega_i = \{ \langle 1, \ldots, 1, \underset{\substack{\uparrow \\ i\text{th place}}}{\sigma}, 1, \ldots, 1, \underset{\substack{\uparrow \\ (n+1)\text{th place}}}{\sigma} \rangle \mid \sigma \in \Sigma_n, \; n \geqslant 1 \}$$

and

$$K_i = \{ \langle c, \ldots, c \rangle \mid c \in \Sigma_0 \}.$$

The forest

$$\omega(\Omega_1 \langle K_1 \rangle, \ldots, \Omega_n \langle K_n \rangle) \subseteq T_{\Sigma \vee \cdots \vee \Sigma \vee \Sigma},$$

with

$$\omega = \langle 1, 2, \ldots, n, \tau \rangle,$$

is again recognizable and it is not hard to see that its image under the function

$$T_{\Sigma \vee \cdots \vee \Sigma} \to T_\Sigma \times \cdots \times T_\Sigma, \quad \omega \mapsto (\varphi_\Sigma \omega, \ldots, \varphi_\Sigma \omega)$$

is just the relation

$$\{ (t_1, \ldots, t_n, \tau t_1 \ldots t_n) \mid t_i \in T_\Sigma \},$$

that is, the graph of (1). □

Corollary. *Given $\tau \in \Sigma_n$ and $L_j \in \mathrm{Rec}(T_\Sigma)$ $(j \neq i)$, the transduction*

$$(2) \qquad t \mapsto \tau L_1 \ldots L_{i-1} t L_{i+1} \ldots L_n$$

is alphabetic.

If at least one of the forests L_j is infinite, the transduction (2) can neither be realized by a top-down nor by a bottom-up tree transducer, provided that the alphabets Σ and Γ are finite (see, for instance [2, definitions of Section 1]).

Conversely, arbitrary top-down or bottom-up transductions do not preserve recognizable forests (see [12]) and, thus, by the next proposition these classes are not included in $\mathcal{A}\ell ph$. In conclusion we have the following corollary.

Corollary. *$\mathscr{A}\ell ph$ is incomparable with both the classes of top-down and bottom-up transductions.*

We conclude this section with the following important result that will be of constant use throughout the remainder of this paper.

Proposition 3.7. *The alphabetic transductions preserve recognizable and algebraic forests.*

Proof. Let

$$\tau = T_\Sigma \xrightarrow{\varphi_\Sigma^{-1}} T_{\Sigma\vee\Gamma} \xrightarrow{\cap L} T_{\Sigma\vee\Gamma} \xrightarrow{\varphi_\Gamma} T_\Gamma, \quad L \in \mathrm{Rec}(T_{\Sigma\vee\Gamma})$$

be an alphabetic transduction; since the homomorphism φ_Σ is alphabetic, for every recognizable (algebraic) forest $R \subseteq T_\Sigma$, the forest $\varphi_\Sigma^{-1}(R)$ is again recognizable (algebraic [1]). On the other hand, the forest

$$\varphi^{-1}(R) \cap L$$

is recognizable or algebraic according to that $\varphi^{-1}(R)$ has the same property.

Finally, since linear homomorphisms project recognizable to recognizable forests and algebraic to algebraic forests, from the equality

$$\varphi_\Gamma[\varphi^{-1}(R) \cap L] = \tau(R)$$

we conclude that τ preserves recognizable and algebraic forests, as stated. $\square$

4. Alphabetic substitutions

This section is devoted to the study of a special class of alphabetic transductions, called the alphabetic substitutions, that play the role of rational substitutions in the word case.

Let Σ and Γ be ranked alphabets.

An *alphabetic substitution* from T_Σ to T_Γ is a pair (f_0, f_n), where

(α) f_0 is a function from Σ_0 to $\mathrm{Rec}(T_\Gamma)$, and

(β) for every $\sigma \in \Sigma_n$, $f_n(\sigma)$ is a finite set of indexed trees of the form

$$\gamma(x_{j_1} \ldots x_{j_q}), \quad \gamma \in \Gamma_q \ (q \geqslant 0),$$

where $j_1, \ldots, j_q$ are distinct elements of $[n]$.

Such a substitution (f_0, f_n) induces a function

$$f : T_\Sigma \to P(T_\Gamma) \quad (= \text{subsets of } T_\Gamma)$$

defined inductively as

$$f(c)=f_0(c), \quad c\in\Sigma_0,$$

$$f(\sigma t_1\ldots t_n)=\{\gamma(s_{j_1}\ldots s_{j_q})\mid\gamma(x_{j_1}\ldots x_{j_q})\in f_n(\sigma) \text{ and } s_j\in f(t_j)\}.$$

Proposition 4.1. *Every alphabetic substitution $f=(f_0,f_n)$ from T_Σ to T_Γ is an alphabetic transduction.*

Proof. Consider the diagram

$$T_\Sigma \xleftarrow{\varphi_\Sigma} T_{\Sigma\vee\Gamma} \xrightarrow{\varphi_\Gamma} T_\Gamma,$$

the recognizable forest

$$K=\bigcup_{c\in\Sigma_0}\varphi_\Sigma^{-1}(c)\cap\varphi_\Gamma^{-1}(f_0(c))\subseteq T_{\Sigma\vee\Gamma}$$

and the following subset of symbols of $\Sigma\vee\Gamma$:

$$\Omega=\{\langle\sigma,\gamma_{j_1\ldots j_q}\rangle\mid\sigma\in\Sigma_n \text{ and } \gamma(x_{j_1}\ldots x_{j_q})\in f_n(\sigma)\}.$$

By virtue of Lemma 3.5, the forest $\Omega\langle K\rangle\subseteq T_{\Sigma\vee\Gamma}$ is recognizable and

$$\#f=\{(\varphi_\Sigma t,\varphi_\Gamma t)\mid t\in\Omega\langle K\rangle\},$$

that is, $f\in\mathcal{Alph}(\Sigma,\Gamma)$, as desired. □

We mention in the sequel some interesting examples of alphabetic substitutions.

(1) *a-product and a-quotient*

In order to develop a regularity theory for trees, Thatcher and Wright [15] have introduced the a-product of forests

$$V\,\dot{\underset{a}{\cdot}}\,U,\quad a\in\Sigma_0,\quad V,U\subseteq T_\Sigma$$

in the following manner:

$$V\,\dot{\underset{a}{\cdot}}\,U=\bigcup_{t\in V}t\,\dot{\underset{a}{\cdot}}\,U,$$

where the forest $t\,\dot{\underset{a}{\cdot}}\,U$ is inductively defined by

$$t\,\dot{\underset{a}{\cdot}}\,U=\begin{cases}U & \text{if } t=a,\\ c & \text{if } t=c\in\Sigma_0-\{a\},\\ \sigma(t_1\,\dot{\underset{a}{\cdot}}\,U)\cdots(t_n\,\dot{\underset{a}{\cdot}}\,U) & \text{if } t=\sigma t_1\ldots t_n.\end{cases}$$

The a-quotient $V/_aU$ is then

$$V/_aU=\{t\mid(t\,\dot{\underset{a}{\cdot}}\,U)\cap V\neq\emptyset\}.$$

Proposition 4.2. *The transductions*

$$t \mapsto t \mathbin{;}_a U \quad \text{and} \quad t \mapsto t \cdot_a U, \qquad U \in \mathrm{Rec}(T_\Sigma)$$

are both alphabetic. Consequently, if the forest $V \subseteq T_\Sigma$ *is recognizable or algebraic, the same is true for* $V \mathbin{;}_a U$ *and* $V \cdot_a U$, *respectively.*

Proof. $t \mapsto t \mathbin{;}_a U$ can be described by the next alphabetic substitution f from T_Σ to itself:

$$f(a) = U,$$

$$f(\sigma) = \{\sigma\}, \quad \sigma \in \Sigma - \{a\}.$$

By Proposition 4.1, $t \mapsto t \mathbin{;}_a U$ is an alphabetic transduction and therefore so is its inverse

$$s \mapsto \{t \mid s \in t \mathbin{;}_a U\}.$$

By applying Proposition 3.7 to the above transductions, we get the stated results. □

Let now Ω be a distinguished leaf of the ranked alphabet Σ ($\Omega \in \Sigma_0$) and let "$\sqsubseteq$" be the usual order relation on T_Σ defined by

$$\Omega \sqsubseteq t \qquad \text{for all } t \in T_\Sigma,$$

$$t_i \sqsubseteq t_i' \quad \text{for } i = 1, \ldots, n,$$

and $\sigma \in \Sigma_n$ implies

$$\sigma t_1 \ldots t_n \sqsubseteq \sigma t_1' \ldots t_n'.$$

If $t \sqsubseteq t'$, we say t is an initial subtree of t'.

For every $t \in T_\Sigma$ and $U \subseteq T_\Sigma$ we put

$$t^{\sqsubseteq} = \{s \mid s \sqsubseteq t\}, \qquad U^{\sqsubseteq} = \bigcup_{t \in U} t^{\sqsubseteq}.$$

Clearly,

$$t^{\sqsubseteq} = t \cdot_\Omega T_\Sigma,$$

so that the transduction $t \mapsto t^{\sqsubseteq}$ is alphabetic; we, therefore, can state the following corollary.

Corollary. *The initial subtrees of the trees of an algebraic forest constitute an algebraic forest, too.*

(2) *Branches*

To any ranked alphabet Σ we associate a monadic alphabet $\Gamma(\Sigma)$ by setting

$$\Gamma(\Sigma)_0 = \{\langle c, 0\rangle \mid c \in \Sigma_0\},$$

$$\Gamma(\Sigma)_1 = \{\langle \sigma, i\rangle \mid \sigma \in \Sigma_n,\ n \geqslant 1 \text{ and } 1 \leqslant i \leqslant n\}.$$

The transduction "branches"

$$\mathrm{br}: T_\Sigma \to T_{\Gamma(\Sigma)}$$

is now given by

$$\mathrm{br}(c) = \{\langle c, 0\rangle\}, \quad c \in \Sigma_0,$$

$$\mathrm{br}(\sigma t_1 \ldots t_n) = \langle \sigma, 1\rangle \,\mathrm{br}(t_1) \cup \cdots \cup \langle \sigma, n\rangle \,\mathrm{br}(t_n)$$

for $\sigma \in \Sigma_n$, $n \geqslant 1$, $t_i \in T_\Sigma$.

Obviously, br coincides with the alphabetic substitution f, where

$$f(\sigma) = \begin{cases} \{\langle \sigma, 0\rangle\} & \text{if } \sigma \in \Sigma_0 \\ \{\langle \sigma, 1\rangle(x_1), \ldots, \langle \sigma, n\rangle(x_n)\}, & \text{if } \sigma \in \Sigma_n,\ n \geqslant 1. \end{cases}$$

Hence, the following proposition.

Proposition 4.3 (Courcelle [6]). *The branches of the trees of a recognizable (algebraic) forest form a recognizable (algebraic) monadic forest.*

5. Composition of alphabetic transductions

The good behavior of the studied transductions is confirmed by the main result of this section that states that the class $\mathcal{A}lph$ is closed under composition; a number of applications follow.

We start with the following lemma.

Lemma 5.1. *For any pair of alphabetic homomorphisms*

$$T_\Sigma \xrightarrow{\varphi} T_\Delta \xleftarrow{\psi} T_\Gamma$$

we can construct a ranked alphabet Θ, a local forest $L \subseteq T_\Theta$ and two alphabetic homomorphisms

$$T_\Sigma \xleftarrow{\alpha} T_\Theta \xrightarrow{\beta} T_\Gamma$$

so that

$$\#(\psi^{-1} \circ \varphi) = \{(\alpha t, \beta t) \mid t \in L\}.$$

Proof. Θ is the subalphabet of $\Sigma \vee \Gamma$ consisting of the following symbols:
- $\langle \sigma, \kappa\rangle$, $\sigma \in \Sigma_n$ and $1 \leqslant \kappa \leqslant N = \max(\deg \Sigma, \deg \Gamma)$,
- $\langle \lambda, \gamma\rangle$, $\gamma \in \Sigma_n$, $1 \leqslant \lambda \leqslant N$,
- $\langle \sigma, \gamma\rangle$, $\sigma \in \Sigma_n$, $\gamma \in \Gamma_m$ $(m, n \geqslant 1)$,
- $\Theta_0 = \Sigma_0 \times \Gamma_0$.

α and β are the restrictions of φ_Σ and φ_Γ on T_Θ, respectively.

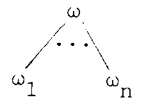

Fig. 8.

Finally, the local forest $L \subseteq T_\Theta$ is generated by the system (E_0, E, T), where

$$E_0 = \Sigma_0 \times \Gamma_0,$$

$$E = \{\langle \sigma, \kappa\rangle \mid \varphi(\sigma) = x_\kappa\} \cup \{\langle \lambda, \gamma\rangle \mid \psi(\gamma) = x_\lambda\} \cup \{\langle \sigma, \gamma\rangle \mid \varphi(\sigma) = \psi(\lambda)\},$$

and T contains all transitions (Fig. 8) of the following four types:

(α) $\omega = \langle \sigma, \kappa\rangle \in E$ and $\omega_\kappa \in E$,

(β) $\omega = \langle \lambda, \gamma\rangle \in E$ and $\omega_\lambda \in E$,

(γ) $\omega = \langle \sigma, \gamma\rangle \in E$ with $\varphi(\sigma) = \delta(x_{i_1} \ldots x_{i_p}) = \psi(\gamma)$ and $\omega_{i_1}, \ldots, \omega_{i_p} \in E$,

(δ) $\omega \in \Theta - E$ and $\omega_1, \ldots, \omega_n \in \Theta$.

From now on, we assume that at least one of the (finite) alphabets Σ, Γ has degree $\geqslant 2$ (in the case where both are monadic, we work with the alphabet $\Sigma^{[2]} \vee \Gamma$ instead of $\Sigma \vee \Gamma$).

Now we claim that for every tree $s \in T_\Sigma$ we can (exclusively using transitions of type δ) build up a tree $u \in T_\Theta$ such that $\alpha(u) = s$. Indeed, u has the same squeleton as s, and if σ is the label of a (nonleaf) node of s, then at the corresponding node of u we put a label of the form $\langle \sigma, \kappa\rangle$, where in the case $\varphi(\sigma) = x_{\kappa'}$, we take care of the inequality $\kappa \neq \kappa'$ (this choice is always possible because of the hypothesis made on the degrees of Σ and Γ); if a is a leaf of s, then at the corresponding place of u we put any one of the leaves $\langle a, c\rangle$, $c \in \Gamma_0$.

By construction u is projected by α on s. Working similarly, for every $t \in T_\Gamma$, we can determine a tree $v \in T_\Theta$ so that $\beta(v) = t$ and v is built up using only transitions δ).

To illustrate the situation, take the following example:

$$\Sigma_0 = \{a, a'\}, \qquad \Sigma_1 = \{\tau\}, \qquad \Sigma_2 = \{\sigma, \sigma'\}, \qquad \Sigma_n = \emptyset \text{ for } n > 2,$$

$$\Gamma_0 = \{c, c'\}, \qquad \Gamma_2 = \{\zeta\}, \qquad \Gamma_3 = \{\gamma\}, \qquad \Gamma_n = \emptyset \text{ for } n \neq 0, 2, 3,$$

$$\Delta_0 = \{d\}, \qquad \Delta_2 = \{\delta\}, \qquad \Delta_n = \emptyset \text{ for } n \neq 0, 2,$$

and

$$\varphi(\sigma) = \psi(\gamma) = \delta(x_2 x_1), \qquad \varphi(\sigma') = x_2, \qquad \psi(\zeta) = x_1,$$

$$\varphi(a) = c = \psi(c) = \varphi(\tau).$$

Let s given by Fig. 9 be a tree in T_Σ. Following the construction made above, we see that all transitions of the tree u given by Fig. 10 belong to type δ) and moreover $\alpha(u) = s$. Similarly, if the tree shown in Fig. 11 is a tree of T_Γ, then the tree shown in Fig. 12 is projected by β on t and every transition of V is of type δ).

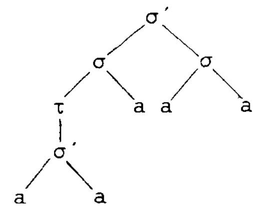

Fig. 9.

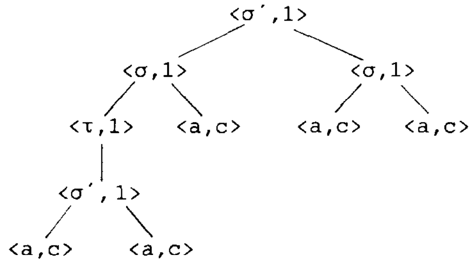

Fig. 10.

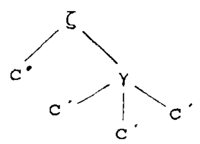

Fig. 11.

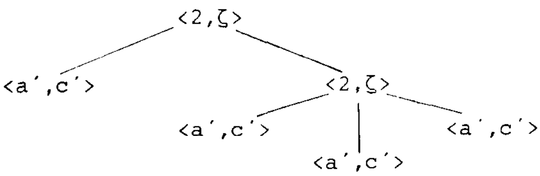

Fig. 12.

Next we observe that the action of β on the tree u follows the unique path P_u (see Fig. 13) suggested by the sequence of the second coordinates of the labels appearing in this path. Analogously, the action of α on v follows the unique path P_v (see Fig. 14).

We can now join the trees u and v as follows: at the place of the leaf $\langle a, c\rangle$ met in (P_u) we put the root $\langle 2, \zeta\rangle$ of v and at the place of the leaf $\langle a', c'\rangle$ met in (P_v) we put the leaf

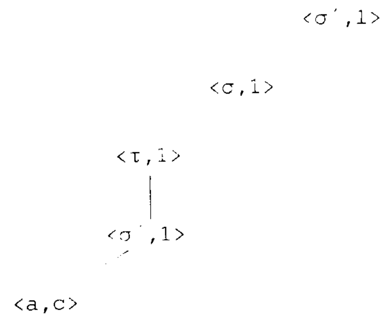

Fig. 13.

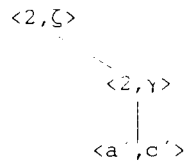

Fig. 14.

$\langle a, c' \rangle$: the so obtained tree shown in Fig. 15 has the property that

$$\alpha(w) = s \quad \text{and} \quad \beta(w) = t,$$

and is built up using transitions of type δ).

The reader will have no difficulty in giving a proof of the above fact for the general case. We summarize our result in the form of the following claim.

Claim (i). *Given trees $s \in T_\Sigma$ and $t \in T_\Gamma$, we can determine trees $u, v, w \in T_\Theta$ constructed by transitions of type δ) and having the following properties:*

$$\alpha(u) = s, \qquad \beta(v) = t$$

$$\alpha(w) = s, \qquad \beta(w) = t.$$

After this preliminary discussion, we are ready to establish the proposed equality. For this purpose consider a pair

$$(s, t) \in \#(\psi^{-1} \circ \varphi),$$

that is, $\varphi(s) = \psi(\tau)$.

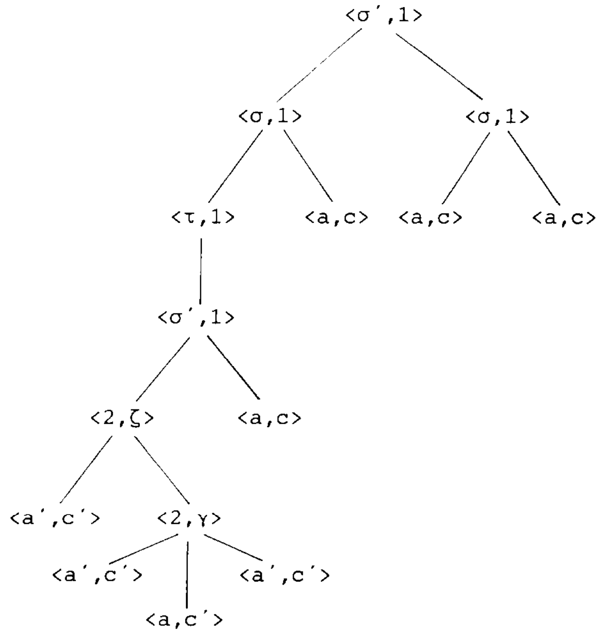

Fig. 15.

Every tree $s \in T_\Sigma$ and $t \in T_\Gamma$ admits decompositions of the forms shown in Figs. 16 and 17, respectively, and described by the following expressions:

Tree s: $s_i \in T_\Sigma$, $\varphi(\tau_1) = x_k$

$$z_i \in T_\Sigma,\ n \geqslant 0, \quad \varphi(\tau_n) = x_{k'},$$

$$t_j^\sigma \in T_\Sigma, \quad \varphi(\sigma) \neq x_k \ \forall k,$$

$$s_i^\sigma \in T_\Sigma, \quad \varphi(\tau_1^\sigma) = x_\lambda,$$

$$z_i^\sigma \in T_\Sigma,\ m \geqslant 0, \quad \varphi(\tau_m^\sigma) = x_{\lambda'},$$

$$t_j^\rho \in T_\Sigma, \quad \varphi(p) \neq x_k \ \forall k.$$

Tree t: $t_j \in T_\Gamma$, $\psi(\pi_1) = x_\mu$

$$u_i \in T_\Sigma,\ n' \geqslant 0, \quad \psi(\pi_{n'}) = x_{\mu'},$$

$$s_j^\gamma \in T_\Gamma, \quad \psi(\gamma) \neq x_k \ \forall k,$$

$$t_j^\zeta \in T_\Sigma, \quad \psi(\zeta) \neq x_\kappa \ \forall k.$$

Therefore, the equality $\varphi(s) = \psi(t)$ implies that

$$\varphi(\sigma) = \psi(\gamma), \qquad \varphi(\rho) = \psi(\zeta), \text{ etc.}$$

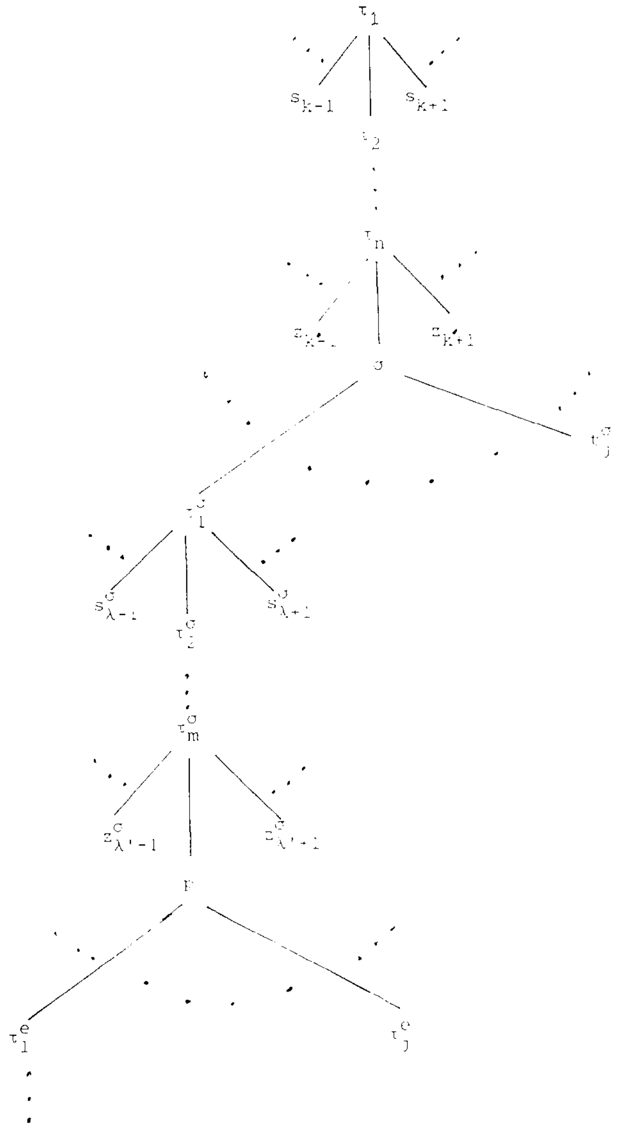

Fig. 16.

The next tree p (see Fig. 18) lies in L because all its transitions belong to T and it is projected via α and β on s and t, respectively:

Tree p

$w_i \in T_\Theta$, $\alpha(w_i) = s_i$ [Claim (i)],

$v_i \in T_\Theta$, $\alpha(v_i) = z_i$ [Claim (i)],

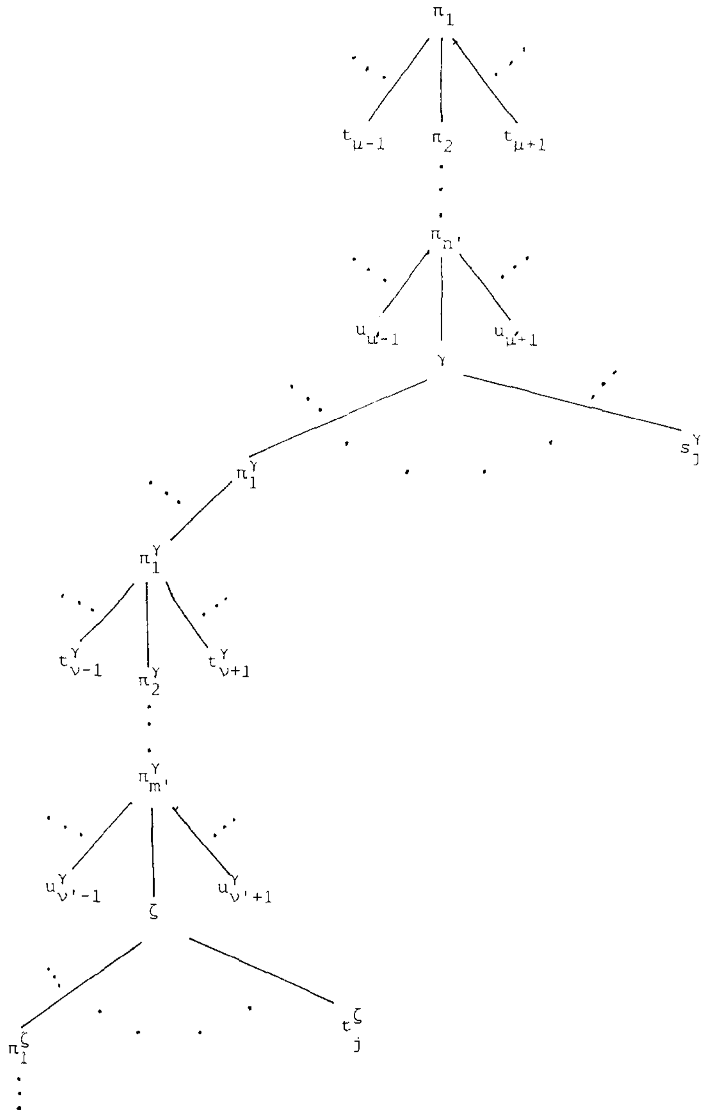

Fig. 17.

$$y_i \in T_\Theta, \quad \beta(y_i) = t_i \qquad \text{[Claim (i)]},$$

$$\varepsilon_i \in T_\Theta, \quad \beta(\varepsilon_j) = u_i \qquad \text{[Claim (i)]},$$

$$r_j \in T_\Theta, \quad \alpha(r_j) = s_j^{\sigma}, \ \beta(t_j) = t_j^{\gamma} \quad \text{[Claim (i)]}.$$

We, therefore, get $(s,t) \in \{(\alpha p, \beta p) \mid p \in L\}$, that is,

$$\#(\psi^{-1} \circ \varphi) \subseteq \{(\alpha p, \beta p) \mid p \in L\}.$$

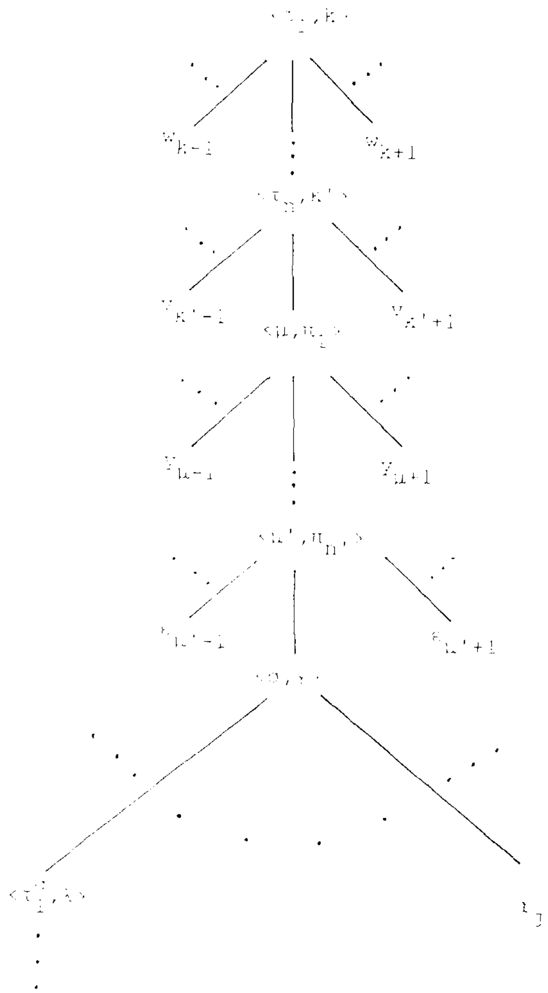

Fig. 18.

The opposite inclusion is proved using similar arguments.

Now we are in a position to prove the following theorem.

Theorem 5.2. *The class $\mathscr{A}lph$ is closed under composition.*

Proof. Let

$$\tau = T_{\Sigma_1} \xrightarrow{\varphi_1^{-1}} T_\Sigma \xrightarrow{\cap R} T_\Sigma \xrightarrow{\varphi} T_\Delta,$$

$$\pi = T_\Delta \xrightarrow{\psi^{-1}} T_\Gamma \xrightarrow{\cap K} T_\Gamma \xrightarrow{\psi_1} T_{\Gamma_1}$$

be two alphabetic transductions.

$$\begin{tikzcd} & T_{\Theta} \arrow[dl, "(\varphi_1 \circ \alpha)^{-1}"'] \arrow[rr, "- \cap \varphi^{-1}(R) \cap \psi^{-1}(K) \cap L"] & & T_{\Theta} \arrow[dr, "\psi_1 \circ \beta"] & \\ T_{\Sigma_1} & & & & T_{\Gamma_1} \end{tikzcd}$$

Fig. 19.

By virtue of Lemma 5.1, there exists an alphabet Θ, a local forest $L \subseteq T_\Theta$ and two alphabetic homomorphisms

$$T_\Sigma \xleftarrow{\alpha} T_\Theta \xrightarrow{\beta} T_\Gamma$$

so that

$$\psi^{-1} \circ \varphi = \beta^{-1} \circ (- \cap L) \circ \alpha.$$

The composition $\pi \circ \tau$ is then equal to the alphabetic transduction shown in Fig. 19.

5.1. *Applications*

(1) *Branches.* Since the transduction "branches" $\mathrm{br}: T_\Sigma \to T_{\Gamma(\Sigma)}$ and its inverse $\mathrm{br}^{-1}: T_{\Gamma(\Sigma)} \to T_\Sigma$ are alphabetic, so is their composition $\mathrm{br}^{-1} \circ \mathrm{br}$. Consequently, for a given algebraic forest $L \subseteq T_\Sigma$, the forest

$$\tilde{L} = \{t \mid \mathrm{br}(t) \cap \mathrm{br}(L) \neq \emptyset\}$$

is also algebraic.

We can state a similar result for the transduction $t \mapsto t^{\sqsubseteq}$. On the other hand, if $\varphi: T_\Sigma \to T_\Gamma$ is an alphabetic homomorphism, its kernel $\varphi^{-1} \circ \varphi$ is an alphabetic (equivalence) relation; therefore, the saturation by $\varphi^{-1} \circ \varphi$ of an algebraic forest $L \subseteq T_{\Sigma_1}$, remains still algebraic.

(2) *Subtrees.* Recall that for a given word

$$w = x_1 \ldots x_n$$

a nonempty subword of w is a word of the form

$$x_{j_1} \ldots x_{j_k} \quad \text{with} \quad 1 \leqslant j_1 < \cdots < j_k \leqslant n.$$

The transduction "subwords" is known to be rational ([3, 8]).

We shall describe the tree analogue of the above. For this, let us consider a ranked alphabet Σ of degree N and denote still by N the ranked alphabet:

$$N_\kappa = \{\kappa\}, \quad \kappa = 0, 1, 2, \ldots, N.$$

We define the alphabetic substitution $f=(f_0, f_n)$ from T_Σ to $T_{N\times\Sigma}$ by putting

$$f_0(c)=\{(0,c)\},\quad c\in\Sigma_0,$$

$$f_n(\sigma)=\{(0,\sigma),(n,1),(n,2),\ldots,(n,n)\}.$$

Then the composition

$$SubTr: T_\Sigma \xrightarrow{f} T_{N\times\Sigma} \xrightarrow{\varphi_\Sigma} T_\Sigma \xrightarrow{``\sqsubseteq"} T_\Sigma$$

is an alphabetic transduction and for every $t\in T_\Sigma$, $SubTr(t)$ is the set of all subtrees of t (recall that "$\sqsubseteq$" is the initial subtree transduction). Figure 20 visualizes the above operation. The uppermost part represents the tree t and the lowermost, the $SubTr(t)$.

(3) *Terminal subtrees.* We say that s is a terminal subtree of $t\in T_\Sigma$ if there exists an indexed tree $\tau(x)\in T_\Sigma(x)$ such that $t=\tau(s)$.
$TSubTr(t)$ denotes the set of all terminal subtrees of t.

Proposition 5.3. *The transduction*

$$TSubTr: T_\Sigma \to T_\Sigma$$

is alphabetic.

Proof. Let L be the local forest of $T_{N\times\Sigma}$ (we conserve the previous notations) generated by the system (E_0, E, T), where

$$E_0=\{0\}\times\Sigma_0,$$

$$E=\{\langle 0,\sigma\rangle, \langle n,j\rangle \mid \sigma\in\Sigma \text{ and } 1\leqslant j\leqslant n\leqslant N\},$$

while T contains transitions of the three types shown in Fig. 21. A straightforward calculation shows that $TSubTr$ is equal to the composition

$$T_\Sigma \xrightarrow{f} T_{N\times\Sigma} \xrightarrow{\cap L} T_{N\times\Sigma} \xrightarrow{\varphi_\Sigma} T_\Sigma$$

and this proves the stated result. $\square$

Corollary. *The terminal subtrees of the trees belonging to an algebraic forest constitute an algebraic forest, too.*

(4) *Largest common initial subtree.* Given a tree $t\in T_\Sigma$, its length is the total number of symbols of Σ that occur in t. For $t, t'\in T_\Sigma$, $t\wedge t'$ denotes the common initial subtree of t, t' of maximum length.

We additively extend to forests the above operation:

$$R\wedge R'=\{t\wedge t' \mid t\in R,\ t'\in R'\}.$$

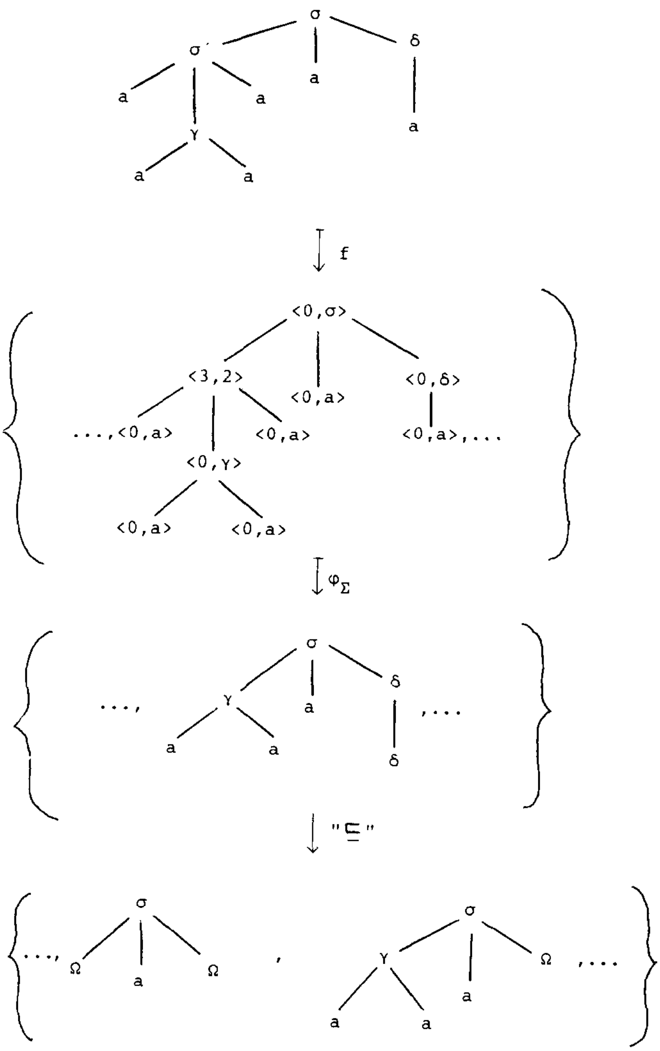

Fig. 20.

Proposition 5.4. *The transduction*

$$t \mapsto t \wedge R, \quad R \in \mathrm{Rec}(T_\Sigma)$$

is alphabetic.

Proof. We need some preliminaries: for each $t \in T_\Sigma$, let Ωt be the tree of $T_{\Sigma \vee \Sigma}$ obtained by putting the leaf Ω on the left side of any symbol of t (see Fig. 22).

We similarly define $t\Omega$ and for $K \subseteq T_\Sigma$ we set

$$\Omega K = \{\Omega t \mid t \in K\} \quad \text{and} \quad K\Omega = \{t\Omega \mid t \in K\}.$$

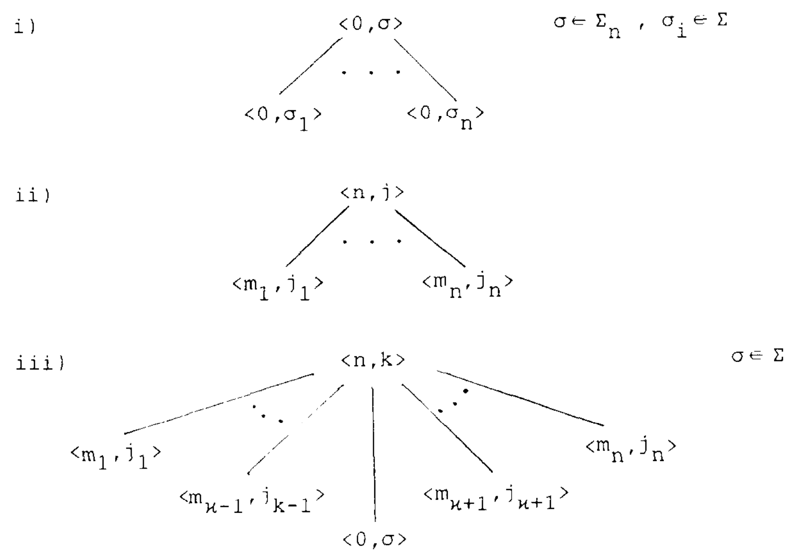

Fig. 21.

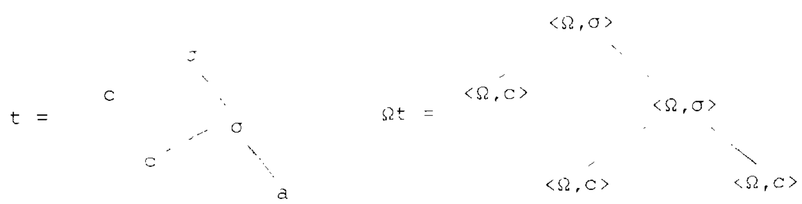

Fig. 22.

If K is recognizable, both ΩK and $K\Omega$ are recognizable (use tree grammars).

Consider now the (alphabetic) substitution $f: T_\Sigma \to T_{\Sigma \vee \Sigma}$ with

$$f(\Omega) = L \quad \text{and} \quad f(\sigma) = \{\langle \sigma, \sigma \rangle\}, \qquad \sigma \in \Sigma - \{\Omega\},$$

where L is the following recognizable forest of $T_{\Sigma \vee \Sigma}$:

$$L = \Omega T_\Sigma \cup T_\Sigma \Omega \cup \left(\bigcup_{\substack{\sigma, \sigma' \in \Sigma \\ \sigma \neq \sigma'}} \langle \sigma, \sigma' \rangle (T_{\Sigma \vee \Sigma}, \ldots, T_{\Sigma \vee \Sigma}) \right).$$

The composition

$$\tau = T_\Sigma \xrightarrow{f} T_{\Sigma \vee \Sigma} \xrightarrow{\varphi_{2,\Sigma}^{-1}(R)} T_{\Sigma \vee \Sigma} \xrightarrow{\varphi_{1,\Sigma}} T_\Sigma$$

is then an alphabetic transduction and we can see that

$$\tau^{-1}(t) = t \wedge R.$$

The result, therefore, comes from the second part of Proposition 3.2. □

Corollary. (i) *If the forests R, R' are recognizable, then so is $R \wedge R'$; consequently, we can decide if $R \wedge R'$ is empty, finite or infinite.*

(ii) *If R is recognizable and L is algebraic, then $R \wedge L$ is algebraic and thus we can decide if $R \wedge L = \emptyset$ or not.*

6. $\Sigma \bigtriangledown \Gamma$-Recognizable relations

Given two ranked alphabets Σ and Γ, $\Sigma \bigtriangledown \Gamma$ is the subalphabet of $\Sigma \vee \Gamma$ defined by

$$(\Sigma \bigtriangledown \Gamma)_n = \bigcup_{\max(\kappa, \lambda) = n} \Sigma_\kappa \times \Gamma_\lambda.$$

$T_\Sigma \times T_\Gamma$ can now be converted into a $\Sigma \bigtriangledown \Gamma$-algebra by setting

$$\langle \sigma, \gamma \rangle ((s_1, t_1), \ldots, (s_n, t_n)) = (\sigma s_1 \ldots s_\kappa, \gamma t_1 \ldots t_\lambda)$$

with $n = \max(\kappa, \lambda)$.

The scope of this section is to study the class $\mathrm{Rec}_{\Sigma \bigtriangledown \Gamma}$ of recognizable subsets of this algebra. By definition, the above class is closed under the boolean operations: union, intersection and complement.

The uniquely existing $\Sigma \bigtriangledown \Gamma$-homomorphism

$$h: T_{\Sigma \bigtriangledown \Gamma} \to T_\Sigma \times T_\Gamma$$

is given by

$$h(\omega) = (\varphi_1 \omega, \varphi_2 \omega),$$

where φ_1, φ_2 are the restrictions of $\varphi_\Sigma, \varphi_\Gamma$, respectively on $T_{\Sigma \bigtriangledown \Gamma}$.

We observe that h is a surjective function.

Arguing as in [4] we can prove the following proposition.

Proposition 6.1. *A is a recognizable subset of the $\Sigma \bigtriangledown \Gamma$-algebra $T_\Sigma \times T_\Gamma$ iff there exists a recognizable forest $K \subseteq T_{\Sigma \bigtriangledown \Gamma}$ such that*

$$A = h(K) \quad \text{and} \quad h^{-1}(A) = K.$$

Consequently,

$$\mathrm{Rec}_{\Sigma \bigtriangledown \Gamma} \subseteq \mathscr{Alph}(\Sigma, \Gamma)$$

and $\mathrm{Rec}_{\Sigma \bigtriangledown \Gamma}$ is closed under inversion in the sense of relations.

Proposition 6.2. (i) *$\mathrm{Rec}_{\Sigma \bigtriangledown \Gamma}$ contains all finite unions of products $B \times C$, with $B \in \mathrm{Rec}(T_\Sigma)$ and $C \in \mathrm{Rec}(T_\Gamma)$.*

(ii) *The diagonal $\Delta \subseteq T_\Sigma \times T_\Sigma$ is $\Sigma \bigtriangledown \Sigma$-recognizable.*

(iii) *For $L \in \mathrm{Rec}(T_\Sigma)$, the transductions*

$$t \mapsto t \cap L \quad \text{and} \quad t \mapsto t \cup L$$

are $\Sigma \bigtriangledown \Sigma$-recognizable.

Proof. (i) As in [4].

(ii) Results from Theorem 6.3 and the fact that Δ is a Takahashi relation.

(iii) Since $\Delta, L \times L \in \mathrm{Rec}_{\Sigma \times \Sigma}$, their intersection

$$\Delta \cap (L \times L) = - \cap L$$

lies also in $\mathrm{Rec}_{\Sigma \times \Sigma}$. On the other hand,

$$- \cup L = \Delta \cup (T_\Sigma \times L) \in \mathrm{Rec}_{\Sigma \times \Sigma}. \qquad \square$$

Let us remind that $A \subseteq T_\Sigma \times T_\Gamma$ is a Takahashi relation if there exists a ranked alphabet Δ, a recognizable forest $K \subseteq T_\Delta$ and a pair of strictly alphabetic homomorphisms

$$T_\Sigma \xleftarrow{\alpha} T_\Delta \xrightarrow{\beta} T_\Gamma$$

so that

$$A = \{(\alpha t, \beta t) \mid t \in K\}.$$

The last definition: let N be the degree of the alphabet Σ and $\mathrm{sq}: T_\Sigma \to T_N$ the homomorphism

$$\mathrm{sq}(\sigma) = \{n\} \quad \text{for } \sigma \in \Sigma_n.$$

For $t \in T_\Sigma$, $\mathrm{sq}(t)$ is the squeleton of t.

We say that the relation $A \subseteq T_\Sigma \times T_\Gamma$ is squeleton-preserving if $(s, t) \in A$ implies $\mathrm{sq}(s) = \mathrm{sq}(t)$.

The main result of this section is the following theorem.

Theorem 6.3. *For any $A \subseteq T_\Sigma \times T_\Gamma$ the following conditions are equivalent*:

(i) *A is a Takahashi relation,*

(ii) *there is a recognizable forest $R \subseteq T_{\Sigma \times \Gamma}$ such that*

$$A = \{(p_\Sigma t, p_\Gamma t) \mid t \in R\},$$

where $\Sigma \times \Gamma$ is the product alphabet

$$(\Sigma \times \Gamma)_n = \Sigma_n \times \Gamma_n, \quad n = 0, 1, 2, \ldots$$

and $P_\Sigma : T_{\Sigma \times \Gamma} \to T_\Sigma$, $P_\Gamma : T_{\Sigma \times \Gamma} \to T_\Gamma$ are the canonical projections

$$P_\Sigma \langle \sigma, \gamma \rangle = \sigma, \qquad P_\Gamma \langle \sigma, \gamma \rangle = \gamma.$$

(iii) *A is a $\Sigma \nabla \Gamma$-recognizable and squeleton-preserving relation.*

$$\begin{tikzcd} T_\Delta \arrow[r, "\mu"] \arrow[dr, "(\alpha,\beta)"'] & T_{\Sigma\times\Gamma} \arrow[d, "(p_\Sigma, p_\Gamma)"] \\ & T_\Sigma\times T_\Gamma \end{tikzcd}$$

Fig. 23.

Proof. (ii)$\Rightarrow$(i): Obvious.

(i)$\Rightarrow$(ii): Let

$$T_\Sigma \xleftarrow{\alpha} T_\Delta \xrightarrow{\beta} T_\Gamma \quad L\in \mathrm{Rec}(T_\Delta)$$

be the bimorphism defining the Takahashi relation A; there results an alphabetic homomorphism

$$\mu: T_\Delta \to T_{\Sigma\times\Gamma}, \quad \mu(\delta)=\langle \alpha(\delta), \beta(\delta)\rangle, \; \delta\in\Delta,$$

making commutative the triangle shown in Fig. 23.
Whence,

$$A=\{(p_\Sigma t, p_\Gamma t) \mid t\in\mu(L)\},$$

with $\mu(L)\in\mathrm{Rec}(T_{\Sigma\times\Gamma})$.

(iii)$\Rightarrow$(ii): Since A is $\Sigma\bigtriangledown\Gamma$-recognizable, there exists a finite $\Sigma\bigtriangledown\Gamma$-algebra $\mathscr{A}$, a $\Sigma\bigtriangledown\Gamma$-homomorphism $\gamma: T_\Sigma\times T_\Gamma \to \mathscr{A}$ and a subset $R\subseteq\mathscr{A}$ so that

$$A=\gamma^{-1}(R).$$

Consider the canonical homomorphism

$$\varphi: T_{\Sigma\times\Gamma}\to T_{\Sigma\bigtriangledown\Gamma}, \quad \varphi\langle\sigma,\gamma\rangle=\langle\sigma,\gamma\rangle;$$

then every $\Sigma\bigtriangledown\Gamma$-algebra (homomorphism) can be viewed as a $\Sigma\times\Gamma$-algebra (homomorphism). Therefore, if we put

$$g: T_{\Sigma\times\Gamma}\to T_\Sigma\times T_\Gamma, \quad g(\omega)=(p_\Sigma\omega, p_\Gamma\omega),$$

then the composition

$$T_{\Sigma\times\Gamma} \xrightarrow{g} T_\Sigma\times T_\Gamma \xrightarrow{\gamma} \mathscr{A}$$

is the uniquely existing $\Sigma\bigtriangledown\Gamma$-homomorphism from $T_{\Sigma\times\Gamma}$ to $\mathscr{A}$. Consequently, the forest

$$K=(\gamma\circ g)^{-1}(R)=g^{-1}(\gamma^{-1}(R))=g^{-1}(A)$$

is recognizable; since by assumption A preserves squeletons, we have

$$A\subseteq g(T_{\Sigma\times\Gamma})$$

$$\begin{tikzcd} T_{\Sigma\times\Gamma} \arrow[rr, "\varphi"] \arrow[dr, "g"'] & & T_{\Sigma\nabla\Gamma} \arrow[dl, "h"] \\ & T_{\Sigma}\times T_{\Gamma} & \end{tikzcd}$$

Fig. 24.

and so

$$\{(p_\Sigma t, p_\Gamma t) \mid t \in K\} = g(K) = g(g^{-1}(A)) = A.$$

(ii)$\Rightarrow$(iii): In order to show that

$$A = g(K), \quad K \in \mathrm{Rec}(T_{\Sigma\times\Gamma})$$

is a $\Sigma\nabla\Gamma$-recognizable relation, we shall use the Proposition 6.1. From the commutative diagram (Fig. 24) we get

$$A = h(\varphi(K)) \text{ with } \varphi(K) \in \mathrm{Rec}(T_{\Sigma\nabla\Gamma}).$$

So, we must establish the equality

$$h^{-1}(h(\varphi(K))) = \varphi(K)$$

which, by the injectivity of φ, is equivalent to

$$\varphi^{-1}(h^{-1}(h(\varphi(K)))) = \varphi^{-1}(\varphi(K)),$$

that is, to

$$g^{-1}(g(K)) = K$$

which is obvious because g is injective, too.

That A preserves squeletons is immediate. $\square$

Remark. In (iii) above we cannot replace the alphabet $\Sigma\nabla\Gamma$ by $\Sigma\times\Gamma$ because in this case the Proposition 6.1 does not hold.

Corollary. *The equality of Takahashi relations is decidable.*

Proof. This comes from the condition (iii) of Theorem 6.3 and the Proposition 6.2. In fact, we have

$$A_1 = A_2 \text{ iff } h^{-1}(A_1) = h^{-1}(A_2)$$

and the last equality is decidable because $h^{-1}(A_1)$ and $h^{-1}(A_2)$ are recognizable forests of $T_{\Sigma\nabla\Gamma}$. $\square$

Corollary. *The class of Takahashi relations is strictly contained in the class* $\mathrm{Rec}_{\Sigma\nabla\Gamma}$ *and it is incomparable with the class of all relations of the form*

$$\cup B \times C, \quad B \in \mathrm{Rec}(T_\Sigma), \; C \in \mathrm{Rec}(T_\Gamma).$$

Proof. It is easy to see that the transduction

$$t \mapsto t \cup R, \quad R \text{ recognizable}$$

is not squeleton-preserving and thus its graph is not a Takahashi relation.

In order to prove our second assertion we observe that the intersection of the classes

$$(\cup B \times C) \cap (\text{Takahashi})$$

is just all finite relations $\{(s_1, t_1), \ldots, (s_n, t_n)\}$ with $\mathrm{sq}(s_i) = \mathrm{sq}(t_i)$ for every i. Consequently, the diagonal Δ cannot belong to the class $(\cup B \times C)$; on the other hand, for any $L \in \mathrm{Rec}(T_\Sigma)$ the transduction

$$\tau_L : T_\Sigma \to T_\Sigma, \quad \tau_L(t) = L \quad \text{for all } t \in T_\Sigma$$

does not preserve squeletons and $\# \tau_L = T_\Sigma \times L$. □

Proposition 6.4. *The class* $\mathrm{Rec}_{\Sigma \nabla \Gamma}$ *properly contains the class* $(\cup B \times C)$.

Proof. We assert that the transduction $t \mapsto t \cup R$ with graph $\Delta \cup (T_\Sigma \times R)$ does not belong to $(\cup B \times C)$. Assume the contrary and take $R = \{c\}$, $c \in \Sigma_0$. Then, by Proposition 1.2, the set

$$\langle a, a \rangle^{-1} \{\Delta \cup (T_\Sigma \times \{c\})\} = \Delta \subseteq L_\Sigma \times L_\Sigma \quad (a \in \Sigma_0 - \{c\})$$

is recognizable subset of $L_\Sigma \times L_\Sigma$; a contradiction (see [5]). □

To completely justify the Fig. 25 it remains to show the following proposition.

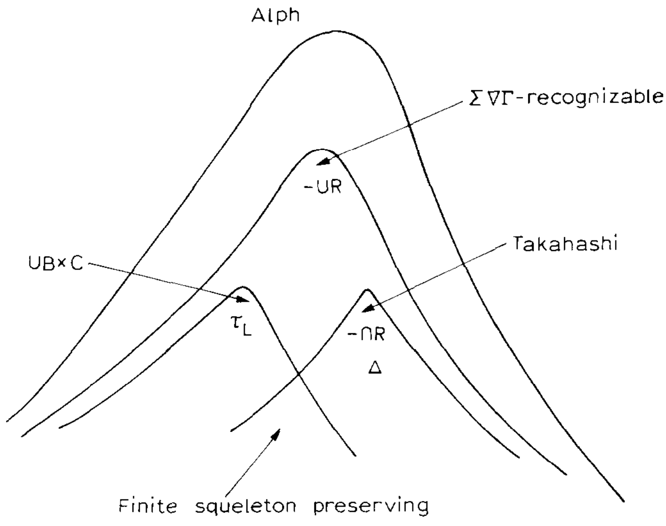

Fig. 25.

Proposition 6.5. *There are alphabetic transductions that are not recognizable.*

The proof requires a preliminary discussion.

Let Σ be a ranked alphabet and consider a branch w (Fig. 26) of a tree $t \in T_\Gamma$ (see Section 4); for any $t' \in T_\Sigma$ and $\tau \in P_\Sigma$ we say that

$$t = t'\tau$$

is a factorization of t along w if there exists an index m ($\leqslant n$) such that the branch shown in Fig. 27 is a branch of τ and that shown in Fig. 28 is a branch of t'.

$$\begin{array}{c} (\sigma_n, i_n) \\ | \\ \vdots \\ | \\ (\sigma_1, i_1) \\ | \\ (c, 0) \end{array} \qquad \sigma_j \in \Sigma - \Sigma_0\ , \quad c \in \Sigma_0$$

Fig. 26.

$$\begin{array}{c} (\sigma_n, i_n) \\ | \\ \vdots \\ | \\ (\sigma_m, i_m) \\ | \\ (x, 0) \end{array}$$

Fig. 27.

$$\begin{array}{c} (\sigma_{m-1}, i_{m-1}) \\ | \\ \vdots \\ | \\ (\sigma_1, i_1) \\ | \\ (c, 0) \end{array}$$

Fig. 28.

Lemma 6.6. *For any recognizable forest $K \subseteq T_\Sigma$, there exists a natural number $N \geqslant 1$ such that each tree $t \in K$ having depth $\geqslant N$ can be factorized along any of its maximal branches as*

$$t = t'\tau\pi, \quad t' \in T_\Sigma, \quad \tau, \pi \in P_\Sigma$$

with $\tau \neq x$, $\mathrm{dp}(t'\tau) \leqslant N$ *and*

$$t'\tau^\kappa\pi \in K \text{ for } \kappa = 0, 1, 2, \ldots$$

Proof. Recall that the function "depth"

$$\mathrm{dp}: T_\Sigma \to \mathbb{N} \quad (= \text{natural numbers})$$

is inductively defined by

$$\mathrm{dp}(c) = 0, \quad c \in \Sigma_0$$

$$\mathrm{dp}(\sigma t_1 \ldots t_n) = 1 + \max\{\mathrm{dp}(t_i) \mid 1 \leqslant i \leqslant n\}.$$

Now, let $\mathscr{A} = (Q, F, a)$ be a finite tree automaton accepting K and set $N = \operatorname{card} Q$. We choose a maximal branch of the tree $t \in K$, say w, as shown in Fig. 29. Then t is uniquely factorized along w as follows:

$$t = c\tau_1 \ldots \tau_n, \quad |\tau_i| = 1 \ (1 \leqslant i \leqslant n),$$

where $|\tau|$ denotes the length of τ viewed as a word of the free monoid P_Σ.

Let us set

$$q_0 = q_c, \qquad q_{\kappa+1} = q_\kappa \cdot \tau_{k+1} \quad 0 \leqslant \kappa \leqslant n-1.$$

Since $\mathrm{dp}(t) \geqslant N$, we have $n \geqslant N$ and therefore it must hold

(3) $\quad q_\kappa = q_\lambda = q$ for $\kappa < \lambda$.

We choose κ to be the smallest index for which (3) holds and λ to be such that

$$q \notin \{q_{\kappa+1}, \ldots, q_{\lambda-1}\}.$$

$$\begin{array}{c} (\sigma_n, i_n) \\ | \\ \vdots \\ | \\ (\sigma_1, i_1) \\ | \\ (c, 0) \end{array}$$

Fig. 29.

Then

$$t' = c\tau_1 \ldots \tau_\kappa, \qquad \tau = \tau_{\kappa+1} \ldots \tau_\lambda, \qquad \pi = \tau_{\lambda+1} \ldots \tau_n$$

have all the desired properties.

Now let us return to the proof of the Proposition 6.5:
Set

$$\Sigma_2 = \{\sigma, \tau\}, \qquad \Sigma_0 = \{a, b\}, \qquad \Sigma_n = 0, \qquad n \neq 0, 2$$

and consider the top-catenation given by Fig. 30 and assume it to be $\Sigma \times \Sigma$-recognizable. Let $N \geqslant 1$ be the number associated by Lemma 6.6 with the recognizable forest $K = h^{-1}(A) \subseteq T_{\Sigma \times \Sigma}$, where A is the graph of the top-catenation shown in Fig. 30. Take t as shown in Fig. 31, $\mathrm{dp}(t) = 2N$.

By virtue of Proposition 6.1 the tree ω shown in Fig. 32 belongs to $K \subseteq T_{\Sigma \times \Sigma}$. Next, any terminal subtree of ω having depth $\leqslant N$ is not in K and this leads to a contradiction.

By similar arguments we can show that the transductions "subtrees", "initial subtrees", "terminal subtrees", etc., are not recognizable, too. □

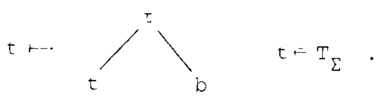

Fig. 30.

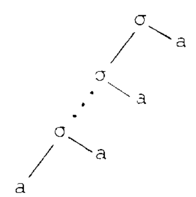

Fig. 31.

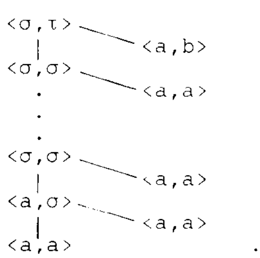

Fig. 32.

References

[1] A. Arnold and M. Dauchet, Forêts algèbriques et homomorphisme inverses, *Inform. and Control* **37** (2) (1978) 182–196.

[2] Baker, Tree transducers and tree languages, *Inform. and Control* **37** (3) (1978) 241–266.

[3] J. Berstel, *Transductions and Context-Free Languages* (Teubner, Stuttgart, 1979).

[4] S. Bozapalidis, Sur une classe des relations d'arbres, *RAIRO Inform. Théor. Appl.* **22** (1) (1988) 43–47.

[5] S. Bozapalidis and S. Ioulidis, Actions, finite state sets and applications to trees, *Internat. J. Computer and Math.* **31** (1989) 11–17.

[6] B. Courcelle, A representation of trees by languages II, *Theoret. Comput. Sci.* **7** (1978) 25–58.

[7] M. Dauchet, Transductions de forêts. Bimorphismes de magmoides, Thèse d'Etat, Université de Lille I, 1977.

[8] S. Eilenberg, *Automata, Languages and Machines* (Academic Press, New York, 1974).

[9] J. Engelfriet, Bottom-up, top-down tree transformations: a comparison, *Math. Systems Theory* **9** (1975) 198–231.

[10] J. Engelfriet, A hierarchy of tree transducers, in: *Proc. 3me Colloque de Lille sur les arbres en Algèbre et en Programmation* (1978) 103–106.

[11] J. Engelfriet and E.-M. Schmidt, IO and OI, I, *J. Comput. System Sci.* **15** 3 (1977) 328–353.

[12] W.C. Rounds, Mappings and grammars on trees, *Math. Systems Theory* **4** (1970) 257–287.

[13] A. Takahashi, Primitive transformations on regular sets and recognizable sets in: Nivat, ed. *Automata, Languages and Programming* (North-Holland, Amsterdam, 1973) 475–480.

[14] J.W. Thatcher, Generalized^2 sequential machine maps, *J. Comput. System Sci.* **1** (1970) 339–367.

[15] J.W. Thatcher and Wright, Generalized finite automata theory with an application to a decision problem of second-order logic, *Math. Systems Theory* **2** (1) (1968) 57–81.