Theoretical Informatics and Applications

Theoret. Informatics Appl. **36** (2002) 5–28 DOI: 10.1051/ita:2002002

POLYPODIC CODES*

Symeon Bozapalidis¹ and Olympia Louscou–Bozapalidou²

Abstract. Word and tree codes are studied in a common framework, that of polypodes which are sets endowed with a substitution like operation. Many examples are given and basic properties are examined. The code decomposition theorem is valid in this general setup.

Mathematics Subject Classification. 68R05, 05C90.

1. INTRODUCTION

The objective of code theory is to study the properties of decompositions of a pattern into smaller patterns taken from a set. Encoding actually means embedding a free object into another of the same category. Thus, the study of codes is reduced to that of subobjects of a free object.

It is well known, that in the category of Γ -algebras (groups) every sub-algebra (sub-group) of a free algebra (free group) is free itself. Therefore, in the above categories every subobject of a free object is generated by a code. However, this is true neither in the case of monoids used in classical code theory, nor in the case of polypodes used here for tree codes.

A polypode is an algebraic structure, very convenient for studying trees, graphs, words, etc. They model the operation of substitution frequently used in Computer Science.

More precisely, an $\,n-polypode \ (n\geq 1)\,$ is a set M equipped with an operation of the form

 $M \times M^n \to M, \quad (m, m_1, ..., m_n) \longmapsto m [m_1, ..., m_n] \quad (fixed n)$ (1)

which is associative and admits a unit n-tuple.

O EDP Sciences 2002

Keywords and phrases: Code, polypode, trees.

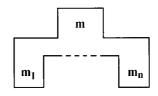
^{*} This work is partially supported by the EU project IST-2000-26016 IMCOMP.

 $^{^{1}}$ Department of Mathematics, Aristotle University of Thessaloniki, 54006 Thessaloniki, Greece.

 $^{^2}$ Posidonos 7, Nea Krini, 55132 Thessaloniki, Greece.

These structures were introduced by Menger who used them in the context of Logic [7]. Give'on [5] considered certain Menger algebras, which he called m-ary monoids, as transition monoids of tree automata.

A main reason we have changed Give'on's terminology is that the above structure is not only an extension of the monoid concept but also of the Γ -algebra concept (the carrier set is merged with the operator domain). In addition the term "polypode" better depicts the functioning of (1): single elements are combined with an *n*-tuple to produce single elements.



Additive polypodes have already been used to define equationally and study context-free formal power series on trees (cf. [1]).

The paper is divided into seven sections.

In Section 2 we introduce the algebra of polypodes and indicate that many interesting classes (of trees, words, etc.) are polypodes and many well-known functions connecting these classes such as yield, tree homomorphism, Parikh's function etc., are actually morphisms of polypodes.

The notion of a polypodic code is introduced in Section 3: a subset C of an n-polypode M is a *code* whenever the canonical encoding polypode morphism $T_{\Gamma_C}(X_n) \to M$ is injective, where Γ_C is an alphabet of n-ranked symbols in a bijection with the elements of C and $T_{\Gamma_C}(X_n)$ denotes the set of trees constructed from Γ_C .

Examples of tree and word codes in the above sense are given.

Section 4 is devoted to discussing some properties of polypodic codes.

The first one is that if C is a polypodic code lying within the image of a polypode morphism $h: M \to N$, *i.e.* $C \subseteq Im(h)$, then choosing elements in the C-fibres of h,

$$m_c \in h^{-1}(c)$$
 for all $c \in C$,

we get a polypodic code in M. For instance if $C \subseteq X_n^*$ is a polypodic word code, then choosing trees t_c with yield (frontier) $c \in C$, we obtain a tree code.

If C is a code in the n-polypode M, then C^n is an ordinary code in the monoid M^n whose multiplication is

$$(m_1, ..., m_n)(m'_1, ..., m'_n) = (m_1[m'_1, ..., m'_n], ..., m_n[m'_1, ..., m'_n]).$$

This result is used to establish that any recognizable (in the sense of [4,6]) tree code C is thin which means that there is an n-tuple of trees $t_1..., t_n \in T_{\Gamma}(X_n)$ such that for all $s, s_1, ..., s_n \in T_{\Gamma}(X_n)$,

$$s[t_1[s_1,...,s_n],...,t_n[s_1,...,s_n]] \notin C.$$

In Section 5 we study the relationship between tree subpolypodes of $T_{\Gamma}(X_n)$ and tree codes. The main result is that if M is a free subpolypode of $T_{\Gamma}(X_n)$, then its minimal normalized generating set is a tree code. Conversely, if $C \subseteq T_{\Gamma}(X_n)$ is a tree code, then pol(C), the subpolypode generated by C, is free and its minimal normalized set of generators coincides with C.

As a consequence we get that if M is a maximal free subpolypode of $T_{\Gamma}(X_n)$ then its minimal normalized generating set is a maximal tree code.

Formal power series on trees and polypodic codes are closely related. A set $C \subseteq T_{\Gamma}(X_n)$ is a tree code if and only if the polypodic star of the characteristic series of C is equal to the characteristic series of the subpolypode generated by C:

$$pol(char(C)) = char(pol(C)).$$

The operation of code composition (cf. [2]) can be extended to the polypodic case and an analogue to the important decomposition theorem is also achieved: each finite tree code can be factorized into indecomposable tree codes (Sect. 7).

It should be pointed out that another tree code notion has been presented by Nivat (cf. [9]) and has been further developed by Restivo and his students (cf. [8]).

2. Polypodes

An *n*-polypode $(n \ge 1)$ is a set M endowed with an operation of the form

$$M \times M^n \to M, \quad (m, m_1, ..., m_n) \longmapsto m[m_1, ..., m_n]$$

which is associative in the sense that

 $m[m_1, ..., m_n] [m'_1, ..., m'_n] = m[m_1[m'_1, ..., m'_n], ..., m_n[m'_1, ..., m'_n]],$

and unitary, *i.e.* there is an *n*-tuple $(e_1, ..., e_n) \in M^n$ such that

$$m[e_1, ..., e_n] = m$$
 and $e_i[m_1, ..., m_n] = m_i$

for all $m, m_i, m'_i \in M$ (i = 1, ..., n).

The algebra of polypodes is defined in the obvious way. Let us only describe subpolypode generation.

Assume a subset A of an n-polypode M is given; its successive polypodic powers $pol_k(A)$ are defined by

- $pol_o(A) = \{e_1, ..., e_n\}$, where $(e_1, ..., e_n)$ is the unit of M, and
- $pol_k(A) = A [pol_{k-1}(A), ..., pol_{k-1}(A)].$

The *polypodic star* of A is the union

$$pol(A) = \bigcup_{k \ge 0} pol_k(A)$$

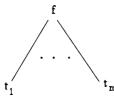
and it is the least subpolypode of M containing A.

In the sequel we discuss free polypodes.

Let $\Gamma = (\Gamma_m)_{m \ge 0}$ be a (not necessarily finite) ranked alphabet and $X_n = \{x_1, ..., x_n\}$ be a set of variables. The set of Γ -trees indexed by X_n , denoted by $T_{\Gamma}(X_n)$, is the smallest set such that:

 $-\Gamma_o \cup X_n \subseteq T_{\Gamma}(X_n)$, and

 $-f \in \Gamma_m, m \ge 1$ and $t_1, ..., t_m \in T_{\Gamma}(X_n)$ imply $f(t_1, ..., t_m) \in T_{\Gamma}(X_n)$. A term $f(t_1, ..., t_m)$ is often depicted by



which justifies the denomination "tree".

In case Γ is *n*-ranked, *i.e.* $\Gamma_k = \emptyset$ for $k \neq n$, the elements of $T_{\Gamma}(X_n)$ are called *n*-ary trees.

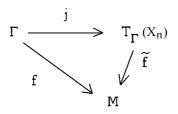
Now, given trees $t, t_1, ..., t_n \in T_{\Gamma}(X_n)$, we use the notation $t[t_1, ..., t_n]$ for the result of substituting t_i for all occurrences of x_i in t $(1 \le i \le n)$.

This operation converts $T_{\Gamma}(X_n)$ into an *n*-polypode whose unit is $(x_1, ..., x_n)$.

Theorem 1. If Γ is n-ranked, $T_{\Gamma}(X_n)$ is the free n-polypode generated by Γ . This means that the function

$$j: \Gamma \to T_{\Gamma}(X_n), \quad \gamma \longmapsto \gamma(x_1, ..., x_n)$$

has the following universal property: for each function $f: \Gamma \to M$ (M an n-polypode) there exists a unique morphism of polypodes $\tilde{f}: T_{\Gamma}(X_n) \to M$ such that the triangle



commutes. The morphism \tilde{f} is inductively defined by

$$-\tilde{f}(x_i) = e_i, \text{ where } (e_1, ..., e_n) \text{ is the unit of } M, \text{ and } -\tilde{f}(\sigma(t_1, ..., t_n)) = f(\sigma) \left[\tilde{f}(t_1), ..., \tilde{f}(t_n)\right].$$

Example 1. (Yield functions) Let Σ be an ordinary alphabet. The set $(\Sigma \cup X_n)^*$ of all words over $\Sigma \cup X_n$ can be structured into an n-polypode via word substitution: for $w, w_1, ..., w_n \in (\Sigma \cup X_n)^*$, $w[w_1, ..., w_n]$ is the result of replacing w_i at all occurences of x_i in w $(1 \le i \le n)$.

Now, for a given n-ranked alphabet Γ , the function $\Gamma \to X_n^*$, $\gamma \longmapsto x_1...x_n$ is uniquely extended into a polypode morphism $y : T_{\Gamma}(X_n) \to X_n^*$ which is the well-known yield function:

 $\begin{array}{ll} - & y(x_i) = x_i, & 1 \leq i \leq n \\ - & y\left(f(t_1, ..., t_n)\right) = y(t_1) ... y(t_n), & f \in \Gamma, & t_j \in T_{\Gamma}(X_n). \end{array}$

Example 2. (Variables)

On the set $\mathcal{P}(X_n)$ of all subsets of $X_n = \{x_1, ..., x_n\}$ a polypodic operation can be defined as follows: for any $A, A_1, ..., A_n \subseteq X_n$

$$A[A_1, \dots, A_n] = A_{i_1} \cup \dots \cup A_{i_k}$$

where $A = \{x_{i_1}, ..., x_{i_k}\}$. The function $\Gamma \to \mathcal{P}(X_n)$ (Γ n-ranked)

$$\gamma \longmapsto \{x_1, ..., x_n\}$$

is extended into a morphism of polypodes

$$Var: T_{\Gamma}(X_n) \to \mathcal{P}(X_n)$$

which to any tree t assigns its set of variables.

 $t \in T_{\Gamma}(X_n)$ is non-deleting if $Var(t) = X_n$.

In order to get information about the variable occurences in a tree, we use the *n*-polypode \mathbb{N}^n (\mathbb{N} the natural numbers). Its operation is given by

$$(\alpha_1, ..., \alpha_n) \left[\left(\alpha_1^{(1)}, ..., \alpha_n^{(1)} \right), ..., \left(\alpha_1^{(n)}, ..., \alpha_n^{(n)} \right) \right]$$
$$= \left(\sum_{\kappa=1}^n \alpha_k \cdot \alpha_1^{(\kappa)} ..., \sum_{\lambda=1}^n \alpha_\lambda \cdot \alpha_1^{(\lambda)} \right).$$

Its unit is

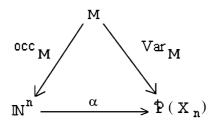
Denote by $|t|_{x_i}$ the number of occurrences of x_i in $t \in T_{\Gamma}(X_n)$. Then the function

$$t\longmapsto\left(\left|t\right|_{x_{1}},...,\left|t\right|_{x_{n}}\right)$$

from $T_{\Gamma}(X_n)$ to \mathbb{N}^n is just the unique polypode morphism extending the function

$$\Gamma \to \mathbb{N}^n, \quad \gamma \longmapsto (1, ..., 1), \quad \gamma \in \Gamma.$$

In general, any pair of polypode morphisms $Var_M : M \to \mathcal{P}(X_n), occ_M : M \to \mathbb{N}^n$ rendering the diagram



commutative, can be interpreted as a variable parametrization and an occurrence function of M respectively; the above function α associates with any $p \in \mathbf{N}^n$ a set of variables $A_p \subseteq X_n$ such that $x_k \in A_p$ if and only if the k-th component of p is non-zero. For instance, in $(\Sigma \cup X_n)^*$ there is an obvious such pair.

Example 3. (Tree homomorphisms)

Assume two ranked alphabets Γ and Δ are given. Any sequence of functions

$$h_m: T_m \to T_\Delta(\xi_1, ..., \xi_m), \quad m = 0, 1, ...$$

can be inductively organized into a single function $h: T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$ by setting inductively

 $- h(x_i) = x_i, \quad 1 \le i \le n$

 $\begin{array}{ll} & - & h\left(f(t_1,...,t_m)\right) = h_n(f)\left[h(t_1),...,h(t_m)\right]\\ which \ is \ called \ a \ tree \ homomorphism \ from \ \Gamma \ to \ \Delta.\\ For \ all \ t,t_1,...,t_n \in T_{\Gamma}(X_n) \ it \ holds \end{array}$

$$h(t[t_1,...,t_n]) = h(t)[h(t_1),...,h(t_n)],$$

that is to say, h is a polypode morphism.

Conversely, any polypode morphism $T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$ (Γ n-ranked) is uniquely determined by a function $\Gamma \to T_{\Delta}(X_n)$.

POLYPODIC CODES

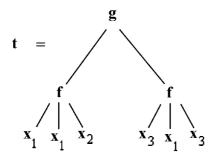
3. Polypodic codes

We begin by introducing some notation.

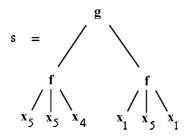
We say that a tree $t \in T_{\Gamma}(X_n)$ is in *normal form* if the next two conditions are satisfied:

- i) in its yield y(t) the leftmost letter is x_1 , the next one is either x_1 or x_2 , the letter after is either x_1 or x_2 or x_3 , etc.;
- ii) if $w = x_1...x_j$ is an initial segment of y(t), then all the variables $x_2, ..., x_{j-1}$ occur in w.

Example 4. The tree



is manifestly in normal form, while the tree



is not. However, s can be uniquely reduced to normal form t through the substitution

$$x_5 \rightarrow x_1$$
, $x_4 \rightarrow x_2$, $x_1 \rightarrow x_3$.

In general, each $t \in T_{\Gamma}(X_n)$ can take its normal normal form n(t) by a unique renaming of its variables.

If M is a subpolypode of $T_{\Gamma}(X_n)$, then $t \in M$ implies $n(t) \in M$. This comes from the equality

$$n(t) = t [x_{i_1}, ..., x_{i_p}], i_1, ..., i_p \in \{1, ..., n\},$$

and the fact that $t, x_1, ..., x_n \in M$.

A tree language $F \subseteq T_{\Gamma}(X_n)$ is normalized if all its trees are in normal form.

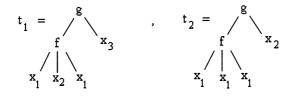
Let $t \in T_{\Gamma}(X_n)$ be in normal form and assume that the variables $x_{i_1}, ..., x_{i_k}$ appear in it $(i_1 < ... < i_k, k \ge 2)$. Performing in t the substitution $x_{i_k} \to x_{i_1}, ..., x_{i_2} \to x_{i_1}$ we get a tree whose normal form is called the $(x_{i_1}, ..., x_{i_k})$ -merging of t. Let VM(t) stand for all trees obtained by merging some set of variables of t. Clearly $t \notin VM(t)$.

For $F \subseteq T_{\Gamma}(X_n)$, we set

$$norm(F) = n(F) - VM(n(F))$$

with $n(F) = \{n(t)/t \in F\}.$

Example 5. Take $F = \{t_1, t_2\}$ with



We have

$$VM(t_1) = \left\{ \begin{array}{cccc} t_2 & f_1 & f_2 \\ f_1 & f_1 & f_1 \\ f_1 & f_2 & x_1 \\ x_1 & x_2 & x_1 \\ x_1 & x_2 & x_1 \\ \end{array} \right\}$$

Therefore, norm $(F) = \{t_1\}.$

Now, let C be a subset of $T_{\Gamma}(X_n)$ and consider the n-ranked alphabet Γ_C with $\Gamma_C = \{\gamma_c / c \in C\}$. C is said to be a *polypodic tree code* (or shortly a *tree code*) whenever the canonical tree homomorphism (= polypode morphism)

$$\varphi_C : T_{\Gamma_C}(X_n) \to T_{\Gamma}(X_n), \quad \varphi_C(\gamma_c) = c \quad (\forall c \in C),$$

is injective. This means that φ_C is actually a polypode isomorphism from $T_{\Gamma_C}(X_n)$ to pol(C), *i.e.* each tree $t \in T_{\Gamma}(X_n)$ admits at most one decomposition by trees in C.

12

Fact 1. If C is a polypodic tree code, then $C \cap X_n = \emptyset$. Indeed, if $x_i \in C$, then

$$\varphi_C(x_i) = x_i = \varphi_C(\gamma_{x_i}),$$

i.e. φ_C is not injective.

Fact 2. If C is a polypodic tree code, then each tree $c \in C$ is non-deleting, i.e. the variables $x_1, ..., x_n$ occur at least once in c.

Indeed, if for instance x_1 does not occur in $c \in C$, then there exist the following two distinct factorizations of the same tree:

$$c[t, t_2, ..., t_n] = c[t', t_2, ..., t_n]$$
, for any $t \neq t'$ and $t, t', t_2, ..., t_n \in pol(C)$.

It turns out that for a tree code C, we have $n(C) \cap VM(n(C)) = \emptyset$.

Fact 3. Since

$$pol(C) = pol \{n(c) / c \in C\}$$

we only have to deal with polypodic tree codes all trees of which are in normal form.

More generally, let M be an n-polypode and $C \subseteq M$. C is a polypodic M-code if the canonical polypode morphism

$$\varphi_C : T_{\Gamma_C}(X_n) \to M, \quad \varphi_C(\gamma_c) = c \quad (c \in C),$$

is injective. Again $C \cap \{e_1, ..., e_n\} = \emptyset$, with $(e_1, ..., e_n)$ denoting the unit of M.

Classical codes can be obtained as instances of tree codes.

Each ordinary alphabet Σ can be viewed as an 1-ranked alphabet; then $T_{\Sigma}(x_1)$ is nothing but a copy of Σ^* . A tree code in this case is just a code in the usual sense. Of course, codes can be defined in an arbitrary monoid A. More precisely, $L \subseteq A$ is a *code* if the canonical monoid morphism

$$h_L: \Sigma_L^* \to A, \quad \Sigma_L = \{\overline{m} \mid m \in L\}, \quad h_L(\overline{m}) = m$$

is injective.

For instance, the set of primes P is a code in the multiplicative monoid $\mathbb R$ of real numbers.

Example 6. The set

$$\Gamma = \{\gamma(x_1, ..., x_n) / \gamma \in \Gamma\} \subseteq T_{\Gamma}(X_n)$$

is clearly a tree code.

Example 7. Given an n-polypode M, the uniform powers of a subset $L \subseteq M$ are inductively defined by

 $- L^{\langle 1 \rangle} = L;$

 $- L^{\langle \kappa+1\rangle} = L\left[L^{\langle \kappa\rangle}, ..., L^{\langle \kappa\rangle}\right].$

If $L = \{t\}$, then $t^{\langle \kappa \rangle}$ is the uniform κ -th power of $t \in M$.

Now, if Γ is an n-ranked alphabet, its uniform powers $\Gamma^{\langle \kappa \rangle}$, $\kappa \geq 1$, constitute tree codes, as well.

Example 8. If Σ is an ordinary alphabet, then the subset

$$C = \{\sigma x_1 x_2 \ / \ \sigma \in \Sigma\}$$

of $(\Sigma \cup X_2)^*$ is a polypodic code.

Proposition 1. If $h: M \to N$ is an injective morphism of polypodes and C is a polypodic code in M, then h(C) is a polypodic code in N.

Proof. This follows directly from the definition of a tree code.

Example 9. Consider a polypodic code $C \subseteq M$. For all $\kappa \geq 1$, $C^{\langle \kappa \rangle}$ is a polypodic code, as well. Indeed, consider the canonical morphism

$$\varphi_C : T_{\Gamma_C}(X_n) \to M, \quad \varphi_C(\gamma_c) = c \qquad c \in C.$$

Then

$$\varphi_C\left(\Gamma_C^{\langle \kappa\rangle}\right) = C^{\langle \kappa\rangle}$$

and the result comes by observing that $\Gamma^{\langle \kappa \rangle}$ is a polypodic code and φ_C is injective.

Example 10. (Viewing trees as words)

Let Γ be an n-ranked alphabet and consider the function

$$e: T_{\Gamma}(X_n) \to (\Gamma \cup X_n)^*$$

inductively defined by

- $e(x_i) = x_i, \quad 1 \le i \le n;$ - $e(\gamma(t_1, ..., t_n)) = \gamma e(t_1) \dots e(t_n) \text{ for } \gamma \in \Gamma \text{ and } t_i \in T_{\Gamma}(X_n)$ that is e converts terms into Polish form.

It can be seen that e is an injective morphism of n-polypodes which transforms trees to words by erasing parentheses and commas

$$e(t[t_1,...,t_n]) = e(t_i)[e(t_1),...,e(t_n)], \quad t_i \in T_{\Gamma}(X_n).$$

Consequently, the sets $e(\Gamma) = \{\gamma x_1 \dots x_n \mid \gamma \in \Gamma\}$ and $e(\Gamma^{\langle 2 \rangle}) = \{\gamma \gamma_1 x_1 \dots x_n \dots \gamma_n x_1 \dots x_n \mid \gamma, \gamma_i \in \Gamma\}$ are polypodic codes in $(\Gamma \cup X_n)^*$.

14

POLYPODIC CODES

4. Properties of polypodic codes

Inverse polypode morphisms preserve in some sense polypodic codes.

Proposition 2. Let us consider a morphism of n-polypodes $h: M \to N$ and let $C \subseteq N$ be a polypodic code such that $C \subseteq Im(h) = \{h(m) \mid m \in M\}$. Choose for each $c \in C$ an element $\tilde{c} \in M$ such that $h(\tilde{c}) = c$. Then $\tilde{C} = \{\tilde{c} \mid c \in C\}$ is a polypodic M-code.

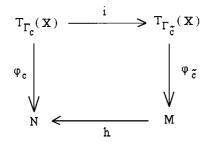
Proof. Let us consider the encoding morphisms

$$\varphi_C: T_{\Gamma_C}(X_n) \to N, \quad \varphi_{\widetilde{C}}: T_{\Gamma_{\widetilde{C}}}(X_n) \to M$$

and the canonical polypode isomorphism

$$i: T_{\Gamma_C}(X_n) \xrightarrow{\sim} T_{\Gamma_{\widetilde{C}}}(X_n)$$

sending γ_c to $\gamma_{\tilde{c}}$ ($c \in C$). From the commutativity of the diagram



and the injectivity of φ_C , we get the injectivity of $\varphi_{\widetilde{C}}$. Hence \widetilde{C} is an *M*-code as wanted.

Let us apply the above result to the morphism yield $y: T_{\Gamma}(X_n) \to X_n^*$ (see Sect. 2). For any polypodic code C in X_n^* such that $C \subseteq y(T_{\Gamma}(X_n))$ (*i.e.* all words of C are yields of some trees in $T_{\Gamma}(X_n)$), we can obtain a tree code by choosing trees t_c with the property

$$y(t_c) = c$$
, for every $c \in C$

A monoid can be associated with any *n*-polypode M, namely M^n with multiplication given by

 $(m_1, ..., m_n)(m'_1, ..., m'_n) = (m_1[m'_1, ..., m'_n], ..., m_n[m'_1, ..., m'_n])$

and with the unit of M as its unit element.

Proposition 3. If a subset $C \subseteq M$ is a polypodic code then

$$C^n = \{(c_1, ..., c_n) / c_i \in C, i = 1, ..., n\}$$

is a code in the monoid M^n .

Proof. Let us have

$$\left(c_{1}^{(1)},...,c_{n}^{(1)}\right)...\left(c_{1}^{(m)},...,c_{n}^{(m)}\right) = \left(d_{1}^{(1)},...,d_{n}^{(1)}\right)...\left(d_{1}^{(p)},...,d_{n}^{(p)}\right)$$

with $c_j^{(i)}, d_j^{(i)} \in C$ for all i, j. Then, for all $i \ (1 \le i \le m)$ we have

$$c_{i}^{(1)}\left[c_{1}^{(2)},...,c_{n}^{(2)}\right]...\left[c_{1}^{(m)},...,c_{n}^{(m)}\right] = d_{i}^{(1)}\left[d_{1}^{(2)},...,d_{n}^{(2)}\right]...\left[d_{1}^{(p)},...,d_{n}^{(p)}\right]$$

or

$$\varphi_{C}(\gamma_{c_{i}^{(1)}})\left[\varphi_{C}(\gamma_{c_{1}^{(2)}}),...,\varphi_{C}(\gamma_{c_{n}^{(2)}})\right]...\left[\varphi_{C}(\gamma_{c_{1}^{(m)}}),...,\varphi_{C}(\gamma_{c_{n}^{(m)}})\right]$$

$$=\varphi_{C}(\gamma_{d_{i}^{(1)}})\left[\varphi_{C}(\gamma_{d_{1}^{(2)}}),...,\varphi_{C}(\gamma_{d_{n}^{(2)}})\right]...\left[\varphi_{C}(\gamma_{d_{1}^{(p)}}),...,\varphi_{C}(\gamma_{d_{n}^{(p)}})\right]$$

where $\varphi_C: T_{\Gamma_C}(X_c) \to M$ is the encoding morphism. Hence

$$\begin{split} \varphi_{C} \left(\gamma_{c_{i}^{(1)}} \left[\gamma_{c_{1}^{(2)}}, ..., \gamma_{c_{n}^{(2)}} \right] ... \left[\gamma_{c_{1}^{(m)}}, ..., \gamma_{c_{n}^{(m)}} \right] \right) \\ = \varphi_{C} \left(\gamma_{d_{i}^{(1)}} \left[\gamma_{d_{1}^{(2)}}, ..., \gamma_{d_{n}^{(2)}} \right] ... \left[\gamma_{d_{1}^{(p)}}, ..., \gamma_{d_{n}^{(p)}} \right] \right) \end{split}$$

or, finally

$$\gamma_{c_i^{(1)}} \left[\gamma_{c_1^{(2)}}, ..., \gamma_{c_n^{(2)}} \right] \dots \left[\gamma_{c_1^{(m)}}, ..., \gamma_{c_n^{(m)}} \right] = \gamma_{d_i^{(1)}} \left[\gamma_{d_1^{(2)}}, ..., \gamma_{d_n^{(2)}} \right] \dots \left[\gamma_{d_1^{(p)}}, ..., \gamma_{d_n^{(p)}} \right].$$

The last equality being in $T_{\Gamma_C}(X_n)$, we deduce that m = p and $\gamma_{c_j^{(i)}} = \gamma_{d_j^{(i)}}$ for all i = 1, ..., m and j = 1, ..., n. Therefore, $c_j^{(i)} = d_j^{(i)}$ for all i, j, and finally

$$\left(c_{1}^{(i)},...,c_{n}^{(i)}\right) = \left(d_{1}^{(i)},...,d_{n}^{(i)}\right) \qquad (i = 1,...,n)$$

proving that C^n is a code in the monoid M^n .

Traditionally, a set L of a monoid A is said to be *thin* if there exists an element $a \in A$ such that

$$A \ a \ A \cap L = \emptyset.$$

16

It is well known (cf. [2]) that every recognizable code of Σ^* is thin. A similar result holds for tree codes. We say that a set of trees $F \subseteq T_{\Gamma}(X_n)$ is thin if there is an *n*-tuple of trees $(t_1, ..., t_n) \in T_{\Gamma}(X_n)^n$ so that for all $s, s_1, ..., s_n \in T_{\Gamma}(X_n)$

 $s[t_1[s_1, ..., s_n], ..., t_n[s_1, ..., s_n]] \notin F.$

Proposition 4. Any recognizable tree code $C \subseteq T_{\Gamma}(X_n)$ is thin.

Proof. We first establish the

Claim. If $F \subseteq T_{\Gamma}(X_n)$ is recognizable, then F^n is a recognizable subset of the monoid $T_{\Gamma}(X_n)^n$.

It is not hard to show that if $F \subseteq T_{\Gamma}(X_n)$ is recognizable, then

$$card\left\{m^{-}F \ / \ m \in T_{\Gamma}(X_{n})\right\} < \infty \tag{(r)}$$

where $m^-F = \{(t_1, ..., t_n) / m[t_1, ..., t_n] \in F\}.$

Now, let us determine an arbitrary left derivative of F^n in $T_{\Gamma}(X_n)^n$. For $(m_1, ..., m_n) \in T_{\Gamma}(X_n)^n$ we have

$$(m_1, ..., m_n)^{-1} F^n = \{ (t_1, ..., t_n) / (m_1, ..., m_n)(t_1, ..., t_n) \in F^n \}$$

= $\{ (t_1, ..., t_n) / m_i[t_1, ..., t_n] \in F , 1 \le i \le n \}$
= $\bigcap_{i=1}^n m_i^- F.$

Taking into account (r) we get

card {
$$(m_1, ..., m_n)^{-1} F^n / (m_1, ..., m_n) \in T_{\Gamma}(X_n)^n$$
} < ∞ ,

in other words, F^n has finitely many distinct left derivatives in $T_{\Gamma}(X_n)^n$, so it is recognizable, as claimed.

Now, let us prove our claim that if $C \subseteq T_{\Gamma}(X_n)$ is a recognizable tree code, then C^n is a recognizable code of the monoid $T_{\Gamma}(X_n)^n$. Then, following the argument of the corresponding proposition of [2], we can show that F^n is thin, *i.e.* there is an *n*-tuple $(a_1, ..., a_n)$ of elements of $T_{\Gamma}(X_n)$ such that for all $(s_1, ..., s_n)$, $(w_1, ..., w_n) \in T_{\Gamma_n}(X_n)^n$,

$$(s_1, ..., s_n) (a_1, ..., a_n) (w_1, ..., w_n) \notin F^n.$$

Therefore, a fortiori, we have $(s, ..., s)(a_1, ..., a_n)(w_1, ..., w_n) \notin F^n$, *i.e.*,

$$s[a_1[w_1, ..., w_n], ..., a_n[w_1, ..., w_n]] \notin F$$

as wanted.

5. Free subpolypodes and polypodic codes

We start with a definition. Let M be an n-polypode and $L, F \subseteq M$. The \Box -product of L by F, denoted $L \Box F$, is given by

$$L \Box F = \left(\bigcup_{i=1}^{n} L[e_1, ..., e_{i-1}, F, e_{i+1}, ..., e_n]\right)$$
$$\cup \left(\bigcup_{i < j} L[..., e_{i-1}, F, e_{i+1}, ..., e_{j-1}, F, e_{j+1}, ...]\right)$$

$$\cup \dots L[F, \dots, F].$$

Proposition 5. Let M be a subpolypode of $T_{\Gamma}(X_n)$. We put

$$G = M - X_n$$
 and $C = G - G \square G$.

Then, norm(C) is the minimal normalized set of generators of M.

Proof. We are going to show first that norm(C) generates M, *i.e.* that

$$pol(norm(C)) = M.$$

We only establish that $M \subseteq pol(norm(C))$ since the opposite inclusion is obvious.

We proceed by induction on the complexity of $t \in M$. If $t = x_i$ $(1 \le i \le n)$, we have nothing to show.

Assume now that $t \neq x_i$, for all $i(1 \leq i \leq n)$. If t does not belong to $G \square G$, then $t \in C$. Otherwise

$$t = s[t_1, ..., t_n]$$
 with $s \in G$ and $t_1, ..., t_n \in G \cup \{e_1, ..., e_n\}$.

Thus all s, t_j have height < height(t) and the induction assumption guarantees that

$$s, t_j \in pol(norm(C))$$
.

We conclude that $t \in pol(norm(C))$ as wanted. Furthermore, let D be another normalized generating set of M : M = pol(D). Without loss of generality we may suppose that $X_n \cap D = \emptyset$. We shall show that $norm(C) \subseteq D$. Let $t \in norm(C)$. Since $t \in pol(D)$ we can write

$$t = s[s_1, \dots, s_n]$$

with $s \in D$ and $s_j \in pol(D)$, j = 1, ..., n. But since $t \notin G \square G$ and t, s are in normal form,

$$s_1 = x_1, \dots, \quad s_n = x_n$$

and therefore t = s, *i.e.* $t \in D$ as desired.

By definition a subpolypode M of $T_{\Gamma}(X_n)$ is *free* if there exists an isomorphism of polypodes

$$h: T_{\Delta}(X_n) \to M$$

for some *n*-ranked alphabet Δ .

Theorem 2. If the subpolypode M of $T_{\Gamma}(X_n)$ is free, then its minimal normalized set of generators is a polypodic code.

Conversely, if C is a polypodic code, then pol(C) is a free subpolypode of $T_{\Gamma}(X_n)$ whose minimal normalized generating set is C.

Proof. If $h : T_{\Delta}(X_n) \to M$ is a polypode isomorphism, h can be viewed as a monomorphism $T_{\Delta}(X_n) \to T_{\Gamma}(X_n)$.

Claim. If t is a non deleting tree in $T_{\Delta}(X_n)$, then h(t) is a non-deleting tree in $T_{\Gamma}(X_n)$. In fact, if, for instance, the variable x_1 is missing in h(t), then choosing two different trees $s, s' \in T_{\Delta}(X_n)$, we get

$$t[s, s_2, ..., s_n] \neq t[s', s_2, ..., s_n]$$

while

$$h(t[s, s_2, ..., s_n]) = h(t[s', s_2, ..., s_n])$$

contradicting the injectivity of h.

Obviously, the set $K = \{n(h(\delta)) / \delta \in \Delta\}$ is a polypodic code in $T_{\Gamma}(X_n)$ and

$$M = h\left(\left(T_{\Delta}(X_n)\right) = pol(K)\right)$$

i.e. K generates M. Furthermore, we set $\Omega = T_{\Delta}(X_n) - X_n$ and

$$\Delta = \{\delta(x_{i_1}, ..., x_{i_n}) / i_1, ..., i_n \in \{1, 2, ..., n\}\}.$$

Then $h(\Omega) = M - X_n = G$ and $\widetilde{\Delta} = \Omega - \Omega \Box \Omega$.

The injectivity of h gives

$$h(\Delta) = h(\Omega - \Omega \Box \Omega) = h(\Omega) - h(\Omega) \Box h(\Omega) = G - G \Box G = C.$$

Thus $K = norm(h(\widetilde{\Delta})) = norm(C)$ proving that K coincides with the minimal normalized set of generators of M.

Conversely, assume that ${\cal C}$ is a polypodic code. The canonical injective morphism

$$\varphi_C: T_{\Gamma_C}(X_n) \to T_{\Delta}(X_n), \quad \varphi_C(\gamma_c) = c \quad (c \in C),$$

can be viewed as an isomorphism of $T_{\Gamma_C}(X_n)$ on pol(C), *i.e.* pol(C) is a free subpolypode of $T_{\Gamma}(X_n)$.

It remains to show that the minimal normalized set of generators of pol(C) is equal to C. For this we put

$$A = T_{\Gamma_C}(X_n) - X_n, \quad G = pol(C) - X_n$$

$$\widetilde{\Gamma}_C = \{ \gamma_c(x_{i_1}, ..., x_{i_n}) \ / \ c \in C \quad i_1, ..., i_n \in \{1, ..., n\} \} \cdot$$

Then $\widetilde{\Gamma}_C = A - A \Box A$. Using now the injectivity of φ_C we get

$$\begin{split} \varphi_C(\widetilde{\Gamma}_C) &= \varphi_C(A - A \Box A) \\ &= \varphi_C(A) - \varphi_C(A \Box A) \\ &= G - G \Box G \\ &= \{c(x_{i_1}, ..., x_{i_n}) \ / \ c \in C, \ i_1, ..., i_n \in \{1, ..., n\} \} \end{split}$$

Finally, since $VM(C) \cap C = \emptyset$, we get $norm(\varphi_C(\widetilde{\Gamma}_C)) = C$ and this completes our proof.

As a consequence we get the following important result:

Corollary 1. If $C, C' \subseteq T_{\Gamma}(X_n)$ are both polypodic codes such that pol(C) = pol(C'), then C = C'.

A free subpolypode M of $T_{\Gamma}(X_n)$ is termed *maximal* if there is no free subpolypode N of $T_{\Gamma}(X_n)$ such that

$$M \subsetneq N \not\subseteq T_{\Gamma}(X_n).$$

Proposition 6. If the free subpolypode M is maximal, then its minimal normalized set of generators C is a maximal polypodic code.

Proof. Let D be a polypodic code such that $C \subsetneq D$. Then

$$pol(C) \subseteq pol(D)$$
 and $pol(C) \neq pol(D)$

because of the previous corollary.

Taking into account that M is maximal we shall have $pol(D) = T_{\Gamma}(X_n) = pol(\Gamma)$. Hence $D = \Gamma$ and therefore $C \not\subseteq \Gamma$. Choose $\gamma \in \Gamma - C$ and put $E = C \cup \gamma^{\langle \kappa \rangle}$, where $\gamma^{\langle \kappa \rangle}$ is the uniform κ -th power of γ .

Then E is a polypodic code and

$$M \subsetneq pol(E) \subsetneq pol(\Gamma) = T_{\Gamma}(X_n)$$

because $\gamma^{\langle \kappa \rangle} \notin M$ and $\gamma(x_1, ..., x_n) \notin pol(E)$.

6. Formal series on trees and tree codes

In this section we discuss the interconnection between series on trees and codes. Let Γ be an *n*-ranked alphabet, $X_n = \{x_1, ..., x_n\}$ and K a commutative semiring. A *formal series on trees* is a function

$$S: T_{\Gamma}(X_n) \to K.$$

The value of S at $t \in T_{\Gamma}(X_n)$ is denoted by (S, t) and referred to as the *coefficient* of S in t. In expansion form S can be written

$$S = \sum_{t \in T_{\Gamma}(X_n)} (S, t) \ t.$$

Example 11. For each tree language $F \subseteq T_{\Gamma}(X_n)$ its characteristic series

$$char(F): T_{\Gamma}(X_n) \to \mathbb{N} \quad (naturals)$$

is given by

$$(char(F),t) = 1$$
 if $t \in F$ and 0 else.

The sum and scaler product of formal series is defined pointwise: for $s, s': T_{\Gamma}(X_n) \to K$ and $\lambda \in L$ we set

$$(S+S',t) = (S,t) + (S',t) , (\lambda S,t) = \lambda \cdot (S,t) , t \in T_{\Gamma}(X_n).$$

The basic operation on series on trees is substitution. For $S, S_1, ..., S_n : T_{\Gamma}(X_n) \to K$ we define $S[S_1, ..., S_n] : T_{\Gamma}(X_n) \to K$ by setting

$$(S [S_1, ..., S_n] , s) = \sum (S, t)(S_1, t_1) ... (S_n, t_n)$$

the sum ranging over all decompositions

$$s = t[t_1, \dots, t_n] \qquad s, t, t_j \in T_{\Gamma}(X_n).$$

Proposition 7. Series substitution is multilinear in all positions, i.e.

$$(\lambda S + \lambda' S')[S_1, ..., S_n] = \lambda S [S_1, ..., S_n] + \lambda' S' [S_1, ..., S_n]$$

$$S[\ldots, \lambda S_i + \lambda S'_i, \ldots] = \lambda S[\ldots, S_i, \ldots] + \lambda' S[\ldots, S'_i, \ldots].$$

Proof. Straightforward.

The successive polypodic powers of $S: T_{\Gamma}(X_n) \to K$ are the formal series

$$pol_{\kappa}(S): T_{\Gamma}(X_n) \to K \qquad \kappa = 1, 2, \dots$$

inductively defined by

 $\begin{array}{l} - \operatorname{pol}_1(S) = S \\ - \operatorname{pol}_2(S) = S[S, ..., S] \\ \vdots \\ - \operatorname{pol}_{\kappa+1}(S) = S \left[\operatorname{pol}_{\kappa}(S) , \ldots , \operatorname{pol}_{\kappa}(S) \right]. \end{array}$ Suppose now S satisfies the condition

$$(S, X_n) = 0,$$

that is to say, all coefficients of S of the variables $x_1, ..., x_n$ vanish. Then the family $(pol_{\kappa}(S))_{\kappa \geq 0}$ is locally finite (*i.e.* for each tree $t \in T_{\Gamma}(X_n)$ there are only finitely many non zero coefficients $(pol_{\kappa}(S), t)$). The polypodic star of S is then defined by

$$pol(S) = \sum_{\kappa \ge 1} pol_{\kappa}(S).$$

Proposition 8. If $C \subseteq T_{\Gamma}(X_n) - X_n$, then the coefficient of the series

 $pol(char(C)): T_{\Gamma}(X_n) \to \mathbb{N}$ (the natural numbers)

at $s \in T_{\Gamma}(X_n)$ equals the number of distinct decompositions of s by trees in C, i.e. it is equal to $\operatorname{card} \varphi_C^{-1}(s)$, with $\varphi_C : T_{\Gamma_C}(X_n) \to T_{\Gamma}(X_n)$ standing for the canonical tree homomorphism.

Proof. Straightforward.

Since C is a polypodic code iff for each $s \in T_{\Gamma}(X_n)$,

$$pol(char(C), s) = 1$$
 or $0,$

we get the following characterization result:

Proposition 9. $C \subseteq T_{\Gamma}(X_n)$ is a polypodic code if and only if

pol(char(C)) = char(pol(C)).

7. The composition operation

The code composition operation is extended to polypodic codes and a decomposition theorem is obtained.

Let $F \subseteq T_{\Gamma}(X_n)$ and $G \subseteq T_{\Delta}(X_n)$ be tree languages and assume that $h : T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$ is a tree homomorphism with the property $h(\Gamma) = G$.

The image $h(F) \subseteq T_{\Delta}(X_n)$ is by definition the *composition* of F by G through h; it is denoted by $F \circ_h G$.

Fact 1. For all $F \subseteq T_{\Gamma}(X_n)$ it holds $F \circ_i \Gamma = F$ where, $i: T_{\Gamma}(X_n) \to T_{\Gamma}(X_n)$ is the identity function.

Fact 2. Let $F \subseteq T_{\Gamma}(X_n)$ and consider the *n*-ranked alphabet $\Gamma_F = \{\gamma_f \mid f \in F\}$ and the canonical homomorphism

$$\varphi_F : T_{\Gamma_F}(X_n) \to T_{\Gamma}(X_n), \quad \varphi_F(\gamma_f) = f, \quad (f \in F).$$

Then it holds $\Gamma_F \circ_{\varphi_F} F = F$. Furthermore, tree language compositions are associative whenever defined. This can be stated formally as follows:

Proposition 10. Consider tree homomorphisms

$$h: T_{\Gamma}(X_n) \to T_{\Delta}(X_n), \ g: T_{\Delta}(X_n) \to T_{\Sigma}(X_n)$$

and tree languages

$$F \subseteq T_{\Gamma}(X_n), \quad G \subseteq T_{\Delta}(X_n), \quad H \subseteq T_{\Sigma}(X_n)$$

such that

$$h(\Gamma) = G, \ g(\Delta) = H.$$

Then

$$(F \circ_h G) \circ_g H = F \circ_g \circ_h (G \circ_g H).$$

Proof. Consider the canonical tree homomorphism $\varphi_F : T_{\Gamma_F}(X_n) \to T_{\Gamma}(X_n)$ as defined previously. Then

$$g \circ (h \circ \varphi_F) = (g \circ h) \circ \varphi_F$$

and therefore

$$[g \circ (h \circ \varphi_F)] (\Gamma_F) = g ((h \circ \varphi_F) (\Gamma_F)) = (F \circ_h G) \circ_g H$$

and

$$[(g \circ h) \circ \varphi_F] (\Gamma_F) = (g \circ h) (\varphi_F (\Gamma_F)) = F \circ_{g \circ h} (G \circ_h H)$$

as wanted.

Proposition 11. Let us consider any tree languages $C, G \subseteq T_{\Delta}(X_n)$. In order to have

$$C = F \circ G$$
 for some F ,

it is necessary and sufficient that $C \subseteq pol(G)$.

Proof. Assume that $C = F \circ_h G$, for some tree homomorphism $h: T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$ with $h(\Gamma) = G$ and some $F \subseteq T_{\Gamma}(X_n)$. Then obviously $C = h(F) \subseteq pol(C)$.

Conversely, assume that $C \subseteq pol(G)$ and let $\varphi_G : T_{\Gamma_G}(X_n) \to T_{\Delta}(X_n)$ be the canonical tree homomorphism. Then by construction

$$C = \varphi_G(F) = F \circ_{\varphi_G} G$$

for $F = \varphi_G^{-1}(C)$, and the proof is complete.

For the composition of two polypodic codes $F \subseteq T_{\Gamma}(X_n)$ and $G \subseteq T_{\Delta}(X_n)$ through $h: T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$, we make the supplementary hypotheses that

- h is injective on the set Γ and

- $\Gamma = alph(F)$.

The last condition means that in the construction of the trees of F we use all the symbols of Γ .

Proposition 12. Under the above assumptions, $C = F \circ_h G$ is a polypodic code.

Proof. It comes by observing that C is the image of the polypodic code F under h which is an injective function because G is a polypodic code.

Proposition 11 takes the following form:

Proposition 13. For any polypodic codes $C, G \subseteq T_{\Delta}(X_n)$,

$$C = F \circ G \ iff \ C \subseteq pol(G) \ and \ alph_G(C) = G.$$

The notion of completeness for ordinary codes can be carried over to the trees. We say that $C \subseteq T_{\Gamma}(X_n)$ is a *complete* polypodic code whenever for each *n*-tuple of trees $(t_1, ..., t_n) \in T_{\Gamma}(X_n)^n$, we can find trees $s, s_1, ..., s_n \in T_{\Gamma}(X_n)$ such that

$$s[t_1[s_1,...,s_n],...,t_n[s_1,...,s_n]] \in pol(C).$$

This means that arbitrary patterns can be extended to elements of the code C.

Proposition 14. Let $F \subseteq T_{\Gamma}(X_n)$ and $G \subseteq T_{\Delta}(X_n)$ be polypodic codes composable through $h: T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$. If both F and G are complete, then so is $C = F \circ G$. Moreover, if C is complete, then so is G.

Proof. Since the code $G \subseteq T_{\Delta}(X_n)$ is complete, for all $w_1, ..., w_n \in T_{\Delta}(X_n)$ we can find $u, v_1, ..., v_n \in T_{\Delta}(X_n)$ such that

$$t = u \ [w_1 \ [v_1, ..., v_n], ..., w_n \ [v_1, ..., v_n] \] \in pol \ (G).$$

Choose $\overline{w} \in T_{\Gamma}(X_n)$ in such a way that $h(\overline{w}) = t$.

Since F is complete we can find trees $\overline{u}, \overline{v}_1, ..., \overline{v}_n \in T_{\Gamma}(X_n)$ such that

$$s = \overline{u} \left[\overline{w} \left[\overline{v}_1, ..., \overline{v}_n \right], ..., \overline{w} \left[\overline{v}_1, ..., \overline{v}_n \right] \right] \in pol(F).$$

Thus $h(s) \in pol(C)$, *i.e.*

$$h(\overline{u}) [t [h(\overline{v}_1), ..., h(\overline{v}_n)] , ..., t [h(\overline{v}_1), ..., h(\overline{v}_n)] \in pol(C)].$$

We set

$$u' = h(\overline{u})[u,...,u] \quad \text{and} \quad v'_i = v_i \left[h(\overline{v}_1),...,\,h(\overline{v}_n)\right], \ i = 1,...,n.$$

Then

$$u'[w_1[v'_1, ..., v'_n], ..., w_n[v'_1, ..., v'_n]] = h(s) \in pol(C)$$

proving that C is a complete polypodic code.

From the inclusion $pol(C) \subseteq pol(G)$ and the completeness of C, we get the completeness of G.

Proposition 15. Assume that $F \subseteq T_{\Gamma}(X_n)$, $G \subseteq T_{\Delta}(X_n)$ and $C = F \circ_h G$, with $h: T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$. If C is amaximal polypodic code, then so are both F and G.

Proof. Assume that F is not maximal. Then there is a tree $t \in T_{\Gamma}(X_n)$ such that $F \cup t$ is a polypodic code. Therefore by the injectivity of h, we get that

$$h(F \cup t) = h(F) \cup h(t) = C \cup h(t)$$

is a polypodic code properly containing C, a contradiction.

If G is not maximal, then there is a tree $s \in T_{\Delta}(X_n)$ such that $G \cup s$ is a polypodic code. Add to Γ the new *n*-ranked symbol γ_s and define

$$h_s: T_{\Gamma \cup \gamma_s}(X_n) \to T_{\Gamma}(X_n)$$

by setting $h_s(\gamma) = h(\gamma)$, for each $\gamma \in \Gamma$, and $h_s(\gamma_s) = s$.

Then h_s is injective since $G \cup s$ is a polypodic code. As $F \cup \gamma_s$ is also a polypodic code (over $\Gamma \cup \gamma_s$), we get that

$$h_s(F \cup \gamma_s) = h(F) \cup h(\gamma_s) = C \cup s$$

is a polypodic code properly containing C (since $s \notin C$), a contradiction.

Next result is very useful in practice.

Proposition 16. Given a maximal polypodic code $C \subseteq T_{\Delta}(X_n)$, for any polypodic code $G \subseteq T_{\Delta}(X_n)$,

C decomposes in G iff $pol(C) \subseteq pol(G)$.

In particular, C is indecomposable iff pol(C) is a maximal subpolypode of $T_{\Delta}(X_n)$.

Proof. In one direction we have nothing to show. Assume next that $pol(C) \subseteq pol(G)$ and let $\overline{G} = alph_G(C)$; according to the previous proposition we must prove that $\overline{G} = G$. Since $pol(C) \subseteq pol(\overline{G})$, we deduce that $C = F \circ \overline{G}$. By virtue of Proposition 15 \overline{G} is maximal and $\overline{G} \subseteq G$. Thus $\overline{G} = G$ as wanted.

Now, we are in a position to state next main result.

8. Decomposition theorem

Theorem 3. Each finite polypodic code $C \subseteq T_{\Gamma}(X_n)$ can be written as a composition

$$C = G_1 \circ G_2 \circ \ldots \circ G_{\kappa}$$

of some indecomposable polypodic codes $G_1, G_2, ..., G_{\kappa}$.

We are going to extend the known argument for word codes to the tree case. First an auxiliary result.

Lemma 1. For a finite polypodic code $C \subseteq T_{\Delta}(X_n)$, we set

$$\alpha(C) \ = \ \sum_{t \ \in \ C} \left(size(t) - 1 \right) \ = \ \sum_{t \ \in \ C} \ size(t) \ - \ card(C),$$

where size(t) denotes the number of symbols of Γ occurring in the tree t. If $C = F \circ G$, then

$$\alpha(C) \geqq \alpha(F) + \alpha(G).$$

Proof. If $h : T_{\Gamma}(X_n) \to T_{\Delta}(X_n)$ is injective and $h(\Gamma) = G$, then card(C) = card(F) and

$$\alpha(C) - \alpha(F) = \sum_{t \in C} size(t) - \sum_{s \in F} size(s)$$

$$= \sum_{s \in F} \left(size(h(s)) - size(s) \right).$$

Denoting by $size_{\gamma}(t)$ the multiplicity of $\gamma \in \Gamma$ in t, we have

$$size(h(s)) = \sum_{\gamma \in \Gamma} size(h(\gamma)) \cdot size_{\gamma}(s).$$

Thus continuing the previous string of equalities we get

$$\begin{aligned} \alpha(C) - \alpha(F) &= \sum_{s \in F} \left(\sum_{\gamma \in \Gamma} \left(size(h(\gamma)) \cdot size_{\gamma}(s) - size_{\gamma}(s) \right) \right) \\ &= \left(\sum_{\gamma \in \Gamma} \left(size(h(\gamma)) - 1 \right) \right) \cdot \left(\sum_{s \in F} size_{\gamma}(s) \right). \end{aligned}$$

Obviously $\sum_{s \in F} size_{\gamma}(s) \ge 1$ for all $\gamma \in \Gamma$. Thus

$$\begin{aligned} \alpha(C) - \alpha(F) & \geqq \sum_{\gamma \in \Gamma} (size(h(\gamma)) - 1) \\ &= \sum_{t \in G} (size(t) - 1) \\ &= \alpha(G). \end{aligned}$$

We now return to the proof of the decomposition theorem.

If $\alpha(C) = 0$, then $C \subseteq \Delta$ and we have nothing to show. Assume that $\alpha(C) > 0$. If C is indecomposable, we are done. Otherwise

$$C = F \circ G.$$

Since $0 < \alpha(F) < \alpha(C)$ and $0 < \alpha(G) < \alpha(C)$, the induction implies that

$$F = F_1 \circ \dots \circ F_{\lambda}, \quad G = G_1 \circ \dots \circ G_{\lambda}$$

with F_i , G_j indecomposable. The result follows.

Acknowledgements. We would like to thank the anonymous referee for his fruitful suggestions and comments.

References

- S. Bozapalidis, An Introduction to Polypodic Structures. J. Universal Comput. Sci. 5 (1999) 508-520.
- [2] J. Berstel and D. Perrin, Theory of Codes. Academic Press (1985).
- [3] B. Courcelle, Graph rewriting: An Algebraic and Logic Approach, edited by J. van Leeuwen. Elsevier, Amsterdam, Handb. Theoret. Comput. Sci. B (1990) 193-242.
- [4] F. Gécseg and M. Steinby, Tree Languages, edited by G. Rozenberg and A. Salomaa. Springer-Verlag, New York, Handb. Formal Lang. 3, pp. 1-68.
- [5] V. Give'on, Algebraic Theory of m-automata, edited by Z. Kohavi and A. Paz. Academic Press, New York, Theory of Machines and Computation (1971) 275-286.

- [6] J. Engelfriet, Tree Automata and tree Grammars. DAIMI FN-10 (1975).
- [7] K. Menger, Super Associative Systems and Logical Functions. Math. Ann. 157 (1964) 278-295.
- [8] S. Mantaci and A. Restivo, Tree Codes and Equations, in Proc. of the 3rd International Conference DLT'97, edited by S. Bozapalidis. Thessaloniki (1998) 119-132.
- M. Nivat, Binary Tree Codes. Tree Automata and Languages. Elsevier Science Publishers B.V. North Holland (1992) 1-19.

Communicated by J.-E. Pin. Received February 23, 2001. Accepted May 27, 2002.

> To access this journal online: www.edpsciences.org