# Theoretical Informatics and Applications 

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

## POLYPODIC CODES *

## Symeon Bozapalidis ${ }^{1}$ and Olympia Louscou-Bozapalidou ${ }^{2}$


#### 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 $\Gamma$-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

$$
\begin{equation*}
M \times M^{n} \rightarrow M, \quad\left(m, m_{1}, \ldots, m_{n}\right) \longmapsto m\left[m_{1}, \ldots, m_{n}\right] \quad(\text { fixed } \quad n) \tag{1}
\end{equation*}
$$

which is associative and admits a unit $n$-tuple.

[^0]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 $\Gamma$-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 $T_{\Gamma_{C}}\left(X_{n}\right) \rightarrow M$ is injective, where $\Gamma_{C}$ is an alphabet of $n$-ranked symbols in a bijection with the elements of $C$ and $T_{\Gamma_{C}}\left(X_{n}\right)$ denotes the set of trees constructed from $\Gamma_{C}$.

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 $h: M \rightarrow N$, i.e. $C \subseteq \operatorname{Im}(h)$, then choosing elements in the $C$-fibres of $h$,

$$
m_{c} \in h^{-1}(c) \quad \text { for all } c \in C
$$

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

$$
\left(m_{1}, \ldots, m_{n}\right)\left(m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right)=\left(m_{1}\left[m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right], \ldots, m_{n}\left[m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right]\right)
$$

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 $t_{1} \ldots, t_{n} \in T_{\Gamma}\left(X_{n}\right)$ such
that for all $s, s_{1}, \ldots, s_{n} \in T_{\Gamma}\left(X_{n}\right)$,

$$
s\left[t_{1}\left[s_{1}, \ldots, s_{n}\right], \ldots, t_{n}\left[s_{1}, \ldots, s_{n}\right]\right] \notin C .
$$

In Section 5 we study the relationship between tree subpolypodes of $T_{\Gamma}\left(X_{n}\right)$ and tree codes. The main result is that if $M$ is a free subpolypode of $T_{\Gamma}\left(X_{n}\right)$, then its minimal normalized generating set is a tree code. Conversely, if $C \subseteq T_{\Gamma}\left(X_{n}\right)$ is a tree code, then $\operatorname{pol}(C)$, 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 $T_{\Gamma}\left(X_{n}\right)$ then its minimal normalized generating set is a maximal tree code.

Formal power series on trees and polypodic codes are closely related. A set $C \subseteq T_{\Gamma}\left(X_{n}\right)$ 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 $C$ :

$$
\operatorname{pol}(\operatorname{char}(C))=\operatorname{char}(\operatorname{pol}(C)) .
$$

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 $(n \geq 1)$ is a set $M$ endowed with an operation of the form

$$
M \times M^{n} \rightarrow M, \quad\left(m, m_{1}, \ldots, m_{n}\right) \longmapsto m\left[m_{1}, \ldots, m_{n}\right]
$$

which is associative in the sense that

$$
m\left[m_{1}, \ldots, m_{n}\right]\left[m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right]=m\left[m_{1}\left[m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right], \ldots, m_{n}\left[m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right]\right]
$$

and unitary, i.e. there is an $n$-tuple $\left(e_{1}, \ldots, e_{n}\right) \in M^{n}$ such that

$$
m\left[e_{1}, \ldots, e_{n}\right]=m \quad \text { and } \quad e_{i}\left[m_{1}, \ldots, m_{n}\right]=m_{i}
$$

for all $m, m_{i}, m_{i}^{\prime} \in M \quad(i=1, \ldots, n)$.
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 $\operatorname{pol}_{k}(A)$ are defined by

- $\operatorname{pol}_{o}(A)=\left\{e_{1}, \ldots, e_{n}\right\}, \quad$ where $\left(e_{1}, \ldots, e_{n}\right)$ is the unit of $M$, and
$-\operatorname{pol}_{k}(A)=A\left[\operatorname{pol}_{k-1}(A), \ldots, \operatorname{pol}_{k-1}(A)\right]$.

The polypodic star of $A$ is the union

$$
\operatorname{pol}(A)=\bigcup_{k \geq 0} \operatorname{pol}_{k}(A)
$$

and it is the least subpolypode of $M$ containing $A$.
In the sequel we discuss free polypodes.
Let $\Gamma=\left(\Gamma_{m}\right)_{m \geq 0}$ be a (not necessarily finite) ranked alphabet and $X_{n}=$ $\left\{x_{1}, \ldots, x_{n}\right\}$ be a set of variables. The set of $\Gamma$-trees indexed by $X_{n}$, denoted by $T_{\Gamma}\left(X_{n}\right)$, is the smallest set such that:
$-\Gamma_{o} \cup X_{n} \subseteq T_{\Gamma}\left(X_{n}\right)$, and
$-f \in \Gamma_{m}, m \geq 1 \quad$ and $\quad t_{1}, \ldots, t_{m} \in T_{\Gamma}\left(X_{n}\right) \quad$ imply $\quad f\left(t_{1}, \ldots, t_{m}\right) \in T_{\Gamma}\left(X_{n}\right)$.
A term $f\left(t_{1}, \ldots, t_{m}\right)$ is often depicted by

which justifies the denomination "tree".
In case $\Gamma$ is $n$-ranked, i.e. $\Gamma_{k}=\emptyset$ for $k \neq n$, the elements of $T_{\Gamma}\left(X_{n}\right)$ are called $n$-ary trees.

Now, given trees $t, t_{1}, \ldots, t_{n} \in T_{\Gamma}\left(X_{n}\right)$, we use the notation $t\left[t_{1}, \ldots, t_{n}\right]$ for the result of substituting $t_{i}$ for all occurrences of $x_{i}$ in $t \quad(1 \leq i \leq n)$.

This operation converts $T_{\Gamma}\left(X_{n}\right)$ into an $n$-polypode whose unit is $\left(x_{1}, \ldots, x_{n}\right)$.
Theorem 1. If $\Gamma$ is n-ranked, $T_{\Gamma}\left(X_{n}\right)$ is the free $n$-polypode generated by $\Gamma$.
This means that the function

$$
j: \Gamma \rightarrow T_{\Gamma}\left(X_{n}\right), \quad \gamma \longmapsto \gamma\left(x_{1}, \ldots, x_{n}\right)
$$

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

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

- $\tilde{f}\left(x_{i}\right)=e_{i}$, where $\left(e_{1}, \ldots, e_{n}\right)$ is the unit of $M$, and
$-\tilde{f}\left(\sigma\left(t_{1}, \ldots, t_{n}\right)\right)=f(\sigma)\left[\tilde{f}\left(t_{1}\right), \ldots, \tilde{f}\left(t_{n}\right)\right]$.
Example 1. (Yield functions) Let $\Sigma$ be an ordinary alphabet. The set $\left(\Sigma \cup X_{n}\right)^{*}$ of all words over $\Sigma \cup X_{n}$ can be structured into an n-polypode via word substitution: for $w, w_{1}, \ldots, w_{n} \in\left(\Sigma \cup X_{n}\right)^{*}, w\left[w_{1}, \ldots, w_{n}\right]$ is the result of replacing $w_{i}$ at all occurences of $x_{i}$ in $w(1 \leq i \leq n)$.

Now, for a given $n$-ranked alphabet $\Gamma$, the function $\Gamma \rightarrow X_{n}^{*}, \gamma \longmapsto x_{1} \ldots x_{n}$ is uniquely extended into a polypode morphism $y: T_{\Gamma}\left(X_{n}\right) \rightarrow X_{n}^{*}$ which is the well-known yield function:

- $y\left(x_{i}\right)=x_{i}, \quad 1 \leq i \leq n$
$-\quad y\left(f\left(t_{1}, \ldots, t_{n}\right)\right)=y\left(t_{1}\right) \ldots y\left(t_{n}\right), \quad f \in \Gamma, \quad t_{j} \in T_{\Gamma}\left(X_{n}\right)$.


## Example 2. (Variables)

On the set $\mathcal{P}\left(X_{n}\right)$ of all subsets of $X_{n}=\left\{x_{1}, \ldots, x_{n}\right\}$ a polypodic operation can be defined as follows: for any $A, A_{1}, \ldots, A_{n} \subseteq X_{n}$

$$
A\left[A_{1}, \ldots, A_{n}\right]=A_{i_{1}} \cup \ldots \cup A_{i_{k}}
$$

where $A=\left\{x_{i_{1}}, \ldots, x_{i_{k}}\right\}$. The function $\Gamma \rightarrow \mathcal{P}\left(X_{n}\right) \quad$ ( $\Gamma$ n-ranked)

$$
\gamma \longmapsto\left\{x_{1}, \ldots, x_{n}\right\}
$$

is extended into a morphism of polypodes

$$
\operatorname{Var}: T_{\Gamma}\left(X_{n}\right) \rightarrow \mathcal{P}\left(X_{n}\right)
$$

which to any tree $t$ assigns its set of variables.
$t \in T_{\Gamma}\left(X_{n}\right)$ is non-deleting if $\operatorname{Var}(t)=X_{n}$.
In order to get information about the variable occurences in a tree, we use the $n$-polypode $\mathbb{N}^{n}$ (N the natural numbers). Its operation is given by

$$
\begin{aligned}
\left(\alpha_{1}, \ldots, \alpha_{n}\right)\left[\left(\alpha_{1}^{(1)}, \ldots, \alpha_{n}^{(1)}\right), \ldots,\left(\alpha_{1}^{(n)}, \ldots, \alpha_{n}^{(n)}\right)\right. & ] \\
& =\left(\sum_{\kappa=1}^{n} \alpha_{k} \cdot \alpha_{1}^{(\kappa)} \ldots, \sum_{\lambda=1}^{n} \alpha_{\lambda} \cdot \alpha_{1}^{(\lambda)}\right) .
\end{aligned}
$$

Its unit is

$$
((1,0, \ldots, 0), \ldots,(0, \ldots, 0,1)) .
$$

Denote by $|t|_{x_{i}}$ the number of occurences of $x_{i}$ in $t \in T_{\Gamma}\left(X_{n}\right)$. Then the function

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

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

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

In general, any pair of polypode morphisms $\operatorname{Var}_{M}: M \rightarrow \mathcal{P}\left(X_{n}\right)$, occ $c_{M}: M \rightarrow \mathbf{N}^{n}$ rendering the diagram

commutative, can be interpreted as a variable parametrization and an occurrence function of $M$ respectively; the above function $\alpha$ 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 $\left(\Sigma \cup X_{n}\right)^{*}$ there is an obvious such pair.

Example 3. (Tree homomorphisms)
Assume two ranked alphabets $\Gamma$ and $\Delta$ are given. Any sequence of functions

$$
h_{m}: T_{m} \rightarrow T_{\Delta}\left(\xi_{1}, \ldots, \xi_{m}\right), \quad m=0,1, \ldots
$$

can be inductively organized into a single function $h: T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right)$ by setting inductively

- $h\left(x_{i}\right)=x_{i}, \quad 1 \leq i \leq n$
$-h\left(f\left(t_{1}, \ldots, t_{m}\right)\right)=h_{n}(f)\left[h\left(t_{1}\right), \ldots, h\left(t_{m}\right)\right]$
which is called a tree homomorphism from $\Gamma$ to $\Delta$.
For all $t, t_{1}, \ldots, t_{n} \in T_{\Gamma}\left(X_{n}\right)$ it holds

$$
h\left(t\left[t_{1}, \ldots, t_{n}\right]\right)=h(t)\left[h\left(t_{1}\right), \ldots, h\left(t_{n}\right)\right]
$$

that is to say, $h$ is a polypode morphism.
Conversely, any polypode morphism $T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right)$ ( $\Gamma$ n-ranked) is uniquely determined by a function $\Gamma \rightarrow T_{\Delta}\left(X_{n}\right)$.

## 3. Polypodic codes

We begin by introducing some notation.
We say that a tree $t \in T_{\Gamma}\left(X_{n}\right)$ 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 $w=x_{1} \ldots x_{j}$ is an initial segment of $y(t)$, then all the variables $x_{2}, \ldots, x_{j-1}$ 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 $t \in T_{\Gamma}\left(X_{n}\right)$ can take its normal normal form $n(t)$ by a unique renaming of its variables.

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

$$
n(t)=t\left[x_{i_{1}}, \ldots, x_{i_{p}}\right], i_{1}, \ldots, i_{p} \in\{1, \ldots, n\}
$$

and the fact that $t, x_{1}, \ldots, x_{n} \in M$.
A tree language $F \subseteq T_{\Gamma}\left(X_{n}\right)$ is normalized if all its trees are in normal form.
Let $t \in T_{\Gamma}\left(X_{n}\right)$ be in normal form and assume that the variables $x_{i_{1}}, \ldots, x_{i_{k}}$ appear in it ( $i_{1}<\ldots<i_{k}, k \geq 2$ ). Performing in $t$ the substitution $x_{i_{k}} \rightarrow x_{i_{1}}, \ldots$, $x_{i_{2}} \rightarrow x_{i_{1}}$ we get a tree whose normal form is called the $\left(x_{i_{1}}, \ldots, x_{i_{k}}\right)$-merging of $t$. Let $V M(t)$ stand for all trees obtained by merging some set of variables of $t$. Clearly $t \notin V M(t)$.

For $F \subseteq T_{\Gamma}\left(X_{n}\right)$, we set

$$
\operatorname{norm}(F)=n(F)-V M(n(F))
$$

with $n(F)=\{n(t) / t \in F\}$.
Example 5. Take $F=\left\{t_{1}, t_{2}\right\}$ with


We have


Therefore, norm $(F)=\left\{t_{1}\right\}$.
Now, let $C$ be a subset of $T_{\Gamma}\left(X_{n}\right)$ and consider the $n$-ranked alphabet $\Gamma_{C}$ with $\Gamma_{C}=\left\{\gamma_{c} / c \in C\right\} . C$ is said to be a polypodic tree code (or shortly a tree code) whenever the canonical tree homomorphism (= polypode morphism)

$$
\varphi_{C}: T_{\Gamma_{C}}\left(X_{n}\right) \rightarrow T_{\Gamma}\left(X_{n}\right), \quad \varphi_{C}\left(\gamma_{c}\right)=c \quad(\forall c \in C)
$$

is injective. This means that $\varphi_{C}$ is actually a polypode isomorphism from $T_{\Gamma_{C}}\left(X_{n}\right)$ to $\operatorname{pol}(C)$, i.e. each tree $t \in T_{\Gamma}\left(X_{n}\right)$ admits at most one decomposition by trees in $C$.

Fact 1. If $C$ is a polypodic tree code, then $C \cap X_{n}=\emptyset$. Indeed, if $x_{i} \in C$, then

$$
\varphi_{C}\left(x_{i}\right)=x_{i}=\varphi_{C}\left(\gamma_{x_{i}}\right)
$$

i.e. $\varphi_{C}$ 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 $x_{1}, \ldots, x_{n}$ 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:

$$
c\left[t, t_{2}, \ldots, t_{n}\right]=c\left[t^{\prime}, t_{2}, \ldots, t_{n}\right], \text { for any } t \neq t^{\prime} \text { and } t, t^{\prime}, t_{2}, \ldots, t_{n} \in \operatorname{pol}(C)
$$

It turns out that for a tree code $C$, we have $\mathrm{n}(C) \cap V M(n(C))=\emptyset$.
Fact 3. Since

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

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

$$
\varphi_{C}: T_{\Gamma_{C}}\left(X_{n}\right) \rightarrow M, \quad \varphi_{C}\left(\gamma_{c}\right)=c \quad(c \in C)
$$

is injective. Again $C \cap\left\{e_{1}, \ldots, e_{n}\right\}=\emptyset$, with $\left(e_{1}, \ldots, e_{n}\right)$ denoting the unit of $M$.
Classical codes can be obtained as instances of tree codes.
Each ordinary alphabet $\Sigma$ can be viewed as an 1-ranked alphabet; then $T_{\Sigma}\left(x_{1}\right)$ is nothing but a copy of $\Sigma^{*}$. 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

$$
h_{L}: \Sigma_{L}^{*} \rightarrow A, \quad \Sigma_{L}=\{\bar{m} / m \in L\}, \quad h_{L}(\bar{m})=m
$$

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

$$
\Gamma=\left\{\gamma\left(x_{1}, \ldots, x_{n}\right) / \gamma \in \Gamma\right\} \subseteq T_{\Gamma}\left(X_{n}\right)
$$

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

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

If $L=\{t\}$, then $t^{\langle\kappa\rangle}$ is the uniform $\kappa$-th power of $t \in M$.
Now, if $\Gamma$ is an n-ranked alphabet, its uniform powers $\Gamma^{\langle\kappa\rangle}, \kappa \geq 1$, constitute tree codes, as well.

Example 8. If $\Sigma$ is an ordinary alphabet, then the subset

$$
C=\left\{\sigma x_{1} x_{2} / \sigma \in \Sigma\right\}
$$

of $\left(\Sigma \cup X_{2}\right)^{*}$ is a polypodic code.
Proposition 1. If $h: M \rightarrow N$ 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

$$
\varphi_{C}: T_{\Gamma_{C}}\left(X_{n}\right) \rightarrow M, \quad \varphi_{C}\left(\gamma_{c}\right)=c \quad c \in C
$$

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 $\varphi_{C}$ is injective.

Example 10. (Viewing trees as words)
Let $\Gamma$ be an n-ranked alphabet and consider the function

$$
e: T_{\Gamma}\left(X_{n}\right) \rightarrow\left(\Gamma \cup X_{n}\right)^{*}
$$

inductively defined by

- $e\left(x_{i}\right)=x_{i}, \quad 1 \leq i \leq n ;$
$-\quad e\left(\gamma\left(t_{1}, \ldots, t_{n}\right)\right)=\gamma e\left(t_{1}\right) \ldots e\left(t_{n}\right)$ for $\gamma \in \Gamma$ and $t_{i} \in T_{\Gamma}\left(X_{n}\right)$
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

$$
e\left(t\left[t_{1}, \ldots, t_{n}\right]\right)=e\left(t_{i}\right)\left[e\left(t_{1}\right), \ldots, e\left(t_{n}\right)\right], \quad t_{i} \in T_{\Gamma}\left(X_{n}\right)
$$

Consequently, the sets $e(\Gamma)=\left\{\gamma x_{1} \ldots x_{n} / \gamma \in \Gamma\right\}$ and $e\left(\Gamma^{\langle 2\rangle}\right)=$ $\left\{\gamma \gamma_{1} x_{1} \ldots x_{n} \ldots \gamma_{n} x_{1} \ldots x_{n} / \gamma, \gamma_{i} \in \Gamma\right\}$ are polypodic codes in $\left(\Gamma \cup X_{n}\right)^{*}$.

## 4. Properties of polypodic codes

Inverse polypode morphisms preserve in some sense polypodic codes.
Proposition 2. Let us consider a morphism of n-polypodes $h: M \rightarrow N$ and let $C \subseteq N$ be a polypodic code such that $C \subseteq \operatorname{Im}(h)=\{h(m) / m \in M\}$. Choose for each $c \in C$ an element $\widetilde{c} \in M$ such that $h(\widetilde{c})=c$. Then $\widetilde{C}=\{\widetilde{c} / c \in C\}$ is a polypodic $M$-code.

Proof. Let us consider the encoding morphisms

$$
\varphi_{C}: T_{\Gamma_{C}}\left(X_{n}\right) \rightarrow N, \quad \varphi_{\widetilde{C}}: T_{\Gamma_{\widetilde{C}}}\left(X_{n}\right) \rightarrow M
$$

and the canonical polypode isomorphism

$$
i: T_{\Gamma_{C}}\left(X_{n}\right) \widetilde{\rightarrow} T_{\Gamma_{\tilde{C}}}\left(X_{n}\right)
$$

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

and the injectivity of $\varphi_{C}$, 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 $y: T_{\Gamma}\left(X_{n}\right) \rightarrow X_{n}^{*}$ (see Sect. 2). For any polypodic code $C$ in $X_{n}^{*}$ such that $C \subseteq y\left(T_{\Gamma}\left(X_{n}\right)\right)$ (i.e. all words of $C$ are yields of some trees in $T_{\Gamma}\left(X_{n}\right)$ ), we can obtain a tree code by choosing trees $t_{c}$ with the property

$$
y\left(t_{c}\right)=c, \quad \text { for every } c \in C
$$

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

$$
\left(m_{1}, \ldots, m_{n}\right)\left(m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right)=\left(m_{1}\left[m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right], \ldots, m_{n}\left[m_{1}^{\prime}, \ldots, m_{n}^{\prime}\right]\right)
$$

and with the unit of $M$ as its unit element.

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

$$
C^{n}=\left\{\left(c_{1}, \ldots, c_{n}\right) / c_{i} \in C, \quad i=1, \ldots, n\right\}
$$

is a code in the monoid $M^{n}$.
Proof. Let us have

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

with $c_{j}^{(i)}, d_{j}^{(i)} \in C$ for all $i, j$.
Then, for all $i(1 \leq i \leq m)$ we have

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

or

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

where $\varphi_{C}: T_{\Gamma_{C}}\left(X_{c}\right) \rightarrow M$ is the encoding morphism. Hence

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

or, finally

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

The last equality being in $T_{\Gamma_{C}}\left(X_{n}\right)$, we deduce that $m=p$ and $\gamma_{c_{j}^{(i)}}=\gamma_{d_{j}^{(i)}}$ for all $i=1, \ldots, m$ and $j=1, \ldots, n$. Therefore, $c_{j}^{(i)}=d_{j}^{(i)}$ for all $i, j$, and finally

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

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

$$
A \text { a } A \cap L=\emptyset .
$$

It is well known (cf. [2]) that every recognizable code of $\Sigma^{*}$ is thin. A similar result holds for tree codes. We say that a set of trees $F \subseteq T_{\Gamma}\left(X_{n}\right)$ is thin if there is an $n$-tuple of trees $\left(t_{1}, \ldots, t_{n}\right) \in T_{\Gamma}\left(X_{n}\right)^{n}$ so that for all $s, s_{1}, \ldots, s_{n} \in T_{\Gamma}\left(X_{n}\right)$

$$
s\left[t_{1}\left[s_{1}, \ldots, s_{n}\right], \ldots, t_{n}\left[s_{1}, \ldots, s_{n}\right]\right] \notin F
$$

Proposition 4. Any recognizable tree code $C \subseteq T_{\Gamma}\left(X_{n}\right)$ is thin.
Proof. We first establish the

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

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

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

where $m^{-} F=\left\{\left(t_{1}, \ldots, t_{n}\right) / m\left[t_{1}, \ldots, t_{n}\right] \in F\right\}$.
Now, let us determine an arbitrary left derivative of $F^{n}$ in $T_{\Gamma}\left(X_{n}\right)^{n}$. For $\left(m_{1}, \ldots, m_{n}\right) \in T_{\Gamma}\left(X_{n}\right)^{n}$ we have

$$
\begin{aligned}
\left(m_{1}, \ldots, m_{n}\right)^{-1} F^{n} & =\left\{\left(t_{1}, \ldots, t_{n}\right) /\left(m_{1}, \ldots, m_{n}\right)\left(t_{1}, \ldots, t_{n}\right) \in F^{n}\right\} \\
& =\left\{\left(t_{1}, \ldots, t_{n}\right) / m_{i}\left[t_{1}, \ldots, t_{n}\right] \in F, 1 \leq i \leq n\right\} \\
& =\bigcap_{i=1}^{n} m_{i}^{-} F
\end{aligned}
$$

Taking into account (r) we get

$$
\operatorname{card}\left\{\left(m_{1}, \ldots, m_{n}\right)^{-1} F^{n} /\left(m_{1}, \ldots, m_{n}\right) \in T_{\Gamma}\left(X_{n}\right)^{n}\right\}<\infty
$$

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

Now, let us prove our claim that if $C \subseteq T_{\Gamma}\left(X_{n}\right)$ is a recognizable tree code, then $C^{n}$ is a recognizable code of the monoid $T_{\Gamma}\left(X_{n}\right)^{n}$. 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 $\left(a_{1}, \ldots, a_{n}\right)$ of elements of $T_{\Gamma}\left(X_{n}\right)$ such that for all $\left(s_{1}, \ldots, s_{n}\right)$, $\left(w_{1}, \ldots, w_{n}\right) \in T_{\Gamma_{n}}\left(X_{n}\right)^{n}$,

$$
\left(s_{1}, \ldots, s_{n}\right)\left(a_{1}, \ldots, a_{n}\right)\left(w_{1}, \ldots, w_{n}\right) \notin F^{n}
$$

Therefore, a fortiori, we have $(s, \ldots, s)\left(a_{1}, \ldots, a_{n}\right)\left(w_{1}, \ldots, w_{n}\right) \notin F^{n}$, i.e.,

$$
s\left[a_{1}\left[w_{1}, \ldots, w_{n}\right], \ldots, a_{n}\left[w_{1}, \ldots, w_{n}\right]\right] \notin F
$$

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-product of $L$ by $F$, denoted $L$$F$, is given by

$$
\begin{gathered}
L \square F=\left(\bigcup_{i=1}^{n} L\left[e_{1}, \ldots, e_{i-1}, F, e_{i+1}, \ldots, e_{n}\right]\right) \\
\cup\left(\bigcup_{i<j} L\left[\ldots, e_{i-1}, F, e_{i+1}, \ldots, e_{j-1}, F, e_{j+1}, \ldots\right]\right) \\
\cup \ldots L[F, \ldots, F]
\end{gathered}
$$

Proposition 5. Let $M$ be a subpolypode of $T_{\Gamma}\left(X_{n}\right)$. We put

$$
G=M-X_{n} \quad \text { and } \quad C=G-G \square G .
$$

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

$$
\operatorname{pol}(\operatorname{norm}(C))=M .
$$

We only establish that $M \subseteq \operatorname{pol}(\operatorname{norm}(C))$ since the opposite inclusion is obvious.
We proceed by induction on the complexity of $t \in M$. If $t=x_{i}(1 \leq i \leq n)$, 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

$$
t=s\left[t_{1}, \ldots, t_{n}\right] \text { with } s \in G \text { and } t_{1}, \ldots, t_{n} \in G \cup\left\{e_{1}, \ldots, e_{n}\right\} .
$$

Thus all $s, t_{j}$ have height $<\operatorname{height}(t)$ and the induction assumption guarantees that

$$
s, t_{j} \in \operatorname{pol}(\operatorname{norm}(C)) .
$$

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

$$
t=s\left[s_{1}, \ldots, s_{n}\right]
$$

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

$$
s_{1}=x_{1}, \ldots, \quad s_{n}=x_{n}
$$

and therefore $t=s$, i.e. $t \in D$ as desired.
By definition a subpolypode $M$ of $T_{\Gamma}\left(X_{n}\right)$ is free if there exists an isomorphism of polypodes

$$
h: T_{\Delta}\left(X_{n}\right) \rightarrow M
$$

for some $n$-ranked alphabet $\Delta$.
Theorem 2. If the subpolypode $M$ of $T_{\Gamma}\left(X_{n}\right)$ 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 $T_{\Gamma}\left(X_{n}\right)$ whose minimal normalized generating set is $C$.
Proof. If $h: T_{\Delta}\left(X_{n}\right) \rightarrow M$ is a polypode isomorphism, $h$ can be viewed as a monomorphism $T_{\Delta}\left(X_{n}\right) \rightarrow T_{\Gamma}\left(X_{n}\right)$.
Claim. If $t$ is a non deleting tree in $T_{\Delta}\left(X_{n}\right)$, then $h(t)$ is a non-deleting tree in $T_{\Gamma}\left(X_{n}\right)$. In fact, if, for instance, the variable $x_{1}$ is missing in $h(t)$, then choosing two different trees $s, s^{\prime} \in T_{\Delta}\left(X_{n}\right)$, we get

$$
t\left[s, s_{2}, \ldots, s_{n}\right] \neq t\left[s^{\prime}, s_{2}, \ldots, s_{n}\right]
$$

while

$$
h\left(t\left[s, s_{2}, \ldots, s_{n}\right]\right)=h\left(t\left[s^{\prime}, s_{2}, \ldots, s_{n}\right]\right)
$$

contradicting the injectivity of $h$.
Obviously, the set $K=\{n(h(\delta)) / \delta \in \Delta\}$ is a polypodic code in $T_{\Gamma}\left(X_{n}\right)$ and

$$
M=h\left(\left(T_{\Delta}\left(X_{n}\right)\right)=\operatorname{pol}(K)\right.
$$

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

$$
\widetilde{\Delta}=\left\{\delta\left(x_{i_{1}}, \ldots, x_{i_{n}}\right) / i_{1}, \ldots, i_{n} \in\{1,2, \ldots, n\}\right\}
$$

Then $h(\Omega)=M-X_{n}=G$ and $\widetilde{\Delta}=\Omega-\Omega$$\Omega$.
The injectivity of $h$ gives

$$
h(\widetilde{\Delta})=h(\Omega-\Omega \square \Omega)=h(\Omega)-h(\Omega) \square h(\Omega)=G-G \square G=C
$$

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

Conversely, assume that $C$ is a polypodic code. The canonical injective morphism

$$
\varphi_{C}: T_{\Gamma_{C}}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right), \quad \varphi_{C}\left(\gamma_{c}\right)=c \quad(c \in C)
$$

can be viewed as an isomorphism of $T_{\Gamma_{C}}\left(X_{n}\right)$ on $\operatorname{pol}(C)$, i.e. $\operatorname{pol}(C)$ is a free subpolypode of $T_{\Gamma}\left(X_{n}\right)$.

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

$$
\begin{gathered}
A=T_{\Gamma_{C}}\left(X_{n}\right)-X_{n}, \quad G=\operatorname{pol}(C)-X_{n} \\
\widetilde{\Gamma}_{C}=\left\{\gamma_{c}\left(x_{i_{1}}, \ldots, x_{i_{n}}\right) / c \in C \quad i_{1}, \ldots, i_{n} \in\{1, \ldots, n\}\right\}
\end{gathered}
$$

Then $\widetilde{\Gamma}_{C}=A-A \square A$. Using now the injectivity of $\varphi_{C}$ we get

$$
\begin{aligned}
\varphi_{C}\left(\widetilde{\Gamma}_{C}\right) & =\varphi_{C}(A-A \square A) \\
& =\varphi_{C}(A)-\varphi_{C}(A \square A) \\
& =G-G \square G \\
& =\left\{c\left(x_{i_{1}}, \ldots, x_{i_{n}}\right) / c \in C, i_{1}, \ldots, i_{n} \in\{1, \ldots, n\}\right\}
\end{aligned}
$$

Finally, since $V M(C) \cap C=\emptyset$, we get $\operatorname{norm}\left(\varphi_{C}\left(\widetilde{\Gamma}_{C}\right)\right)=C$ and this completes our proof.

As a consequence we get the following important result:
Corollary 1. If $C, C^{\prime} \subseteq T_{\Gamma}\left(X_{n}\right)$ are both polypodic codes such that pol $(C)=$ $\operatorname{pol}\left(C^{\prime}\right)$, then $C=C^{\prime}$.

A free subpolypode $M$ of $T_{\Gamma}\left(X_{n}\right)$ is termed maximal if there is no free subpolypode $N$ of $T_{\Gamma}\left(X_{n}\right)$ such that

$$
M \subsetneq N \nsubseteq T_{\Gamma}\left(X_{n}\right)
$$

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

$$
\operatorname{pol}(C) \subseteq \operatorname{pol}(D) \quad \text { and } \quad \operatorname{pol}(C) \neq \operatorname{pol}(D)
$$

because of the previous corollary.
Taking into account that $M$ is maximal we shall have $\operatorname{pol}(D)=T_{\Gamma}\left(X_{n}\right)=$ $\operatorname{pol}(\Gamma)$. Hence $D=\Gamma$ and therefore $C \nsubseteq \Gamma$. Choose $\gamma \in \Gamma-C$ and put $E=C \cup \gamma^{\langle\kappa\rangle}$, where $\gamma^{\langle\kappa\rangle}$ is the uniform $\kappa$-th power of $\gamma$.

Then $E$ is a polypodic code and

$$
M \subsetneq \operatorname{pol}(E) \subsetneq \operatorname{pol}(\Gamma)=T_{\Gamma}\left(X_{n}\right)
$$

because $\gamma^{\langle\kappa\rangle} \notin M$ and $\quad \gamma\left(x_{1}, \ldots, x_{n}\right) \notin \operatorname{pol}(E)$.

## 6. Formal series on trees and tree codes

In this section we discuss the interconnection between series on trees and codes.
Let $\Gamma$ be an $n$-ranked alphabet, $X_{n}=\left\{x_{1}, \ldots, x_{n}\right\}$ and $K$ a commutative semiring. A formal series on trees is a function

$$
S: T_{\Gamma}\left(X_{n}\right) \rightarrow K
$$

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

$$
S=\sum_{t \in T_{\Gamma}\left(X_{n}\right)}(S, t) t
$$

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

$$
\operatorname{char}(F): T_{\Gamma}\left(X_{n}\right) \rightarrow \mathbb{N} \quad(\text { naturals })
$$

is given by

$$
(\operatorname{char}(F), t)=1 \quad \text { if } \quad t \in F \quad \text { and } \quad 0 \quad \text { else. }
$$

The sum and scaler product of formal series is defined pointwise: for $s, s^{\prime}: T_{\Gamma}\left(X_{n}\right) \rightarrow$ $K$ and $\lambda \in L$ we set

$$
\left(S+S^{\prime}, t\right)=(S, t)+\left(S^{\prime}, t\right),(\lambda S, t)=\lambda \cdot(S, t), t \in T_{\Gamma}\left(X_{n}\right)
$$

The basic operation on series on trees is substitution. For $S, S_{1}, \ldots, S_{n}: T_{\Gamma}\left(X_{n}\right) \rightarrow$ $K$ we define $S\left[S_{1}, \ldots, S_{n}\right]: T_{\Gamma}\left(X_{n}\right) \rightarrow K$ by setting

$$
\left(S\left[S_{1}, \ldots, S_{n}\right], s\right)=\sum(S, t)\left(S_{1}, t_{1}\right) \ldots\left(S_{n}, t_{n}\right)
$$

the sum ranging over all decompositions

$$
s=t\left[t_{1}, \ldots, t_{n}\right] \quad s, t, t_{j} \in T_{\Gamma}\left(X_{n}\right) .
$$

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

$$
\begin{gathered}
\left(\lambda S+\lambda^{\prime} S^{\prime}\right)\left[S_{1}, \ldots, S_{n}\right]=\lambda S\left[S_{1}, \ldots, S_{n}\right]+\lambda^{\prime} S^{\prime}\left[S_{1}, \ldots, S_{n}\right] \\
S\left[\ldots, \lambda S_{i}+\lambda S_{i}^{\prime}, \ldots\right]=\lambda S\left[\ldots, S_{i}, \ldots\right]+\lambda^{\prime} S\left[\ldots, S_{i}^{\prime}, \ldots\right] .
\end{gathered}
$$

Proof. Straightforward.
The successive polypodic powers of $S: T_{\Gamma}\left(X_{n}\right) \rightarrow K$ are the formal series

$$
\operatorname{pol}_{\kappa}(S): T_{\Gamma}\left(X_{n}\right) \rightarrow K \quad \kappa=1,2, \ldots
$$

inductively defined by

$$
-\operatorname{pol}_{1}(S)=S
$$

$-\operatorname{pol}_{2}(S)=S[S, \ldots, S]$
$-\operatorname{pol}_{\kappa+1}(S)=S\left[\operatorname{pol}_{\kappa}(S), \ldots, \operatorname{pol}_{\kappa}(S)\right]$.
Suppose now $S$ satisfies the condition

$$
\left(S, X_{n}\right)=0
$$

that is to say, all coefficients of $S$ of the variables $x_{1}, \ldots, x_{n}$ vanish. Then the family $\left(\operatorname{pol}_{\kappa}(S)\right)_{\kappa \geq 0}$ is locally finite (i.e. for each tree $t \in T_{\Gamma}\left(X_{n}\right)$ there are only finitely many non zero coefficients $\left(\operatorname{pol}_{\kappa}(S), t\right)$ ). The polypodic star of $S$ is then defined by

$$
\operatorname{pol}(S)=\sum_{\kappa \geq 1} \operatorname{pol}_{\kappa}(S)
$$

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

$$
\operatorname{pol}(\operatorname{char}(C)): T_{\Gamma}\left(X_{n}\right) \rightarrow \mathbb{N} \quad(\text { the natural numbers })
$$

at $s \in T_{\Gamma}\left(X_{n}\right)$ equals the number of distinct decompositions of $s$ by trees in $C$, i.e. it is equal to card $\varphi_{C}^{-1}(s)$, with $\varphi_{C}: T_{\Gamma_{C}}\left(X_{n}\right) \rightarrow T_{\Gamma}\left(X_{n}\right)$ standing for the canonical tree homomorphism.

Proof. Straightforward.
Since $C$ is a polypodic code iff for each $s \in T_{\Gamma}\left(X_{n}\right)$,

$$
\operatorname{pol}(\operatorname{char}(C), s)=1 \quad \text { or } \quad 0
$$

we get the following characterization result:
Proposition 9. $C \subseteq T_{\Gamma}\left(X_{n}\right)$ is a polypodic code if and only if

$$
\operatorname{pol}(\operatorname{char}(C))=\operatorname{char}(\operatorname{pol}(C)) .
$$

## 7. The composition operation

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

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

The image $h(F) \subseteq T_{\Delta}\left(X_{n}\right)$ is by definition the composition of $F$ by $G$ through $h$; it is denoted by $F \circ_{h} G$.
Fact 1. For all $F \subseteq T_{\Gamma}\left(X_{n}\right)$ it holds $F \circ_{i} \Gamma=F$ where, $i: T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Gamma}\left(X_{n}\right)$ is the identity function.

Fact 2. Let $F \subseteq T_{\Gamma}\left(X_{n}\right)$ and consider the $n$-ranked alphabet $\Gamma_{F}=\left\{\gamma_{f} / f \in F\right\}$ and the canonical homomorphism

$$
\varphi_{F}: T_{\Gamma_{F}}\left(X_{n}\right) \rightarrow T_{\Gamma}\left(X_{n}\right), \quad \varphi_{F}\left(\gamma_{f}\right)=f, \quad(f \in F)
$$

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

$$
h: T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right), g: T_{\Delta}\left(X_{n}\right) \rightarrow T_{\Sigma}\left(X_{n}\right)
$$

and tree languages

$$
F \subseteq T_{\Gamma}\left(X_{n}\right), \quad G \subseteq T_{\Delta}\left(X_{n}\right), \quad H \subseteq T_{\Sigma}\left(X_{n}\right)
$$

such that

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

Then

$$
\left(F \circ_{h} G\right) \circ_{g} H=F \circ_{g \circ h}\left(G \circ_{g} H\right) .
$$

Proof. Consider the canonical tree homomorphism $\varphi_{F}: T_{\Gamma_{F}}\left(X_{n}\right) \rightarrow T_{\Gamma}\left(X_{n}\right)$ as defined previously. Then

$$
g \circ\left(h \circ \varphi_{F}\right)=(g \circ h) \circ \varphi_{F}
$$

and therefore

$$
\left[g \circ\left(h \circ \varphi_{F}\right)\right]\left(\Gamma_{F}\right)=g\left(\left(h \circ \varphi_{F}\right)\left(\Gamma_{F}\right)\right)=\left(F \circ_{h} G\right) \circ_{g} H
$$

and

$$
\left[(g \circ h) \circ \varphi_{F}\right]\left(\Gamma_{F}\right)=(g \circ h)\left(\varphi_{F}\left(\Gamma_{F}\right)\right)=F \circ_{g \circ h}\left(G \circ_{h} H\right)
$$

as wanted.

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

$$
C=F \circ G \quad \text { for some } \quad F
$$

it is necessary and sufficient that $C \subseteq \operatorname{pol}(G)$.
Proof. Assume that $C=F \circ_{h} G$, for some tree homomorphism $h: T_{\Gamma}\left(X_{n}\right) \rightarrow$ $T_{\Delta}\left(X_{n}\right)$ with $h(\Gamma)=G$ and some $F \subseteq T_{\Gamma}\left(X_{n}\right)$. Then obviously $C=h(F) \subseteq$ $\operatorname{pol}(C)$.

Conversely, assume that $C \subseteq \operatorname{pol}(G)$ and let $\varphi_{G}: T_{\Gamma_{G}}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right)$ 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 $F \subseteq T_{\Gamma}\left(X_{n}\right)$ and $G \subseteq T_{\Delta}\left(X_{n}\right)$ through $h: T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right)$, we make the supplementary hypotheses that

- $h$ is injective on the set $\Gamma$ and
- $\Gamma=\operatorname{alph}(F)$.

The last condition means that in the construction of the trees of $F$ we use all the symbols of $\Gamma$.
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 $C, G \subseteq T_{\Delta}\left(X_{n}\right)$,

$$
C=F \circ G \text { iff } C \subseteq \operatorname{pol}(G) \text { and } \operatorname{alph}_{G}(C)=G
$$

The notion of completeness for ordinary codes can be carried over to the trees. We say that $C \subseteq T_{\Gamma}\left(X_{n}\right)$ is a complete polypodic code whenever for each $n$-tuple of trees $\left(t_{1}, \ldots, t_{n}\right) \in T_{\Gamma}\left(X_{n}\right)^{n}$, we can find trees $s, s_{1}, \ldots, s_{n} \in T_{\Gamma}\left(X_{n}\right)$ such that

$$
s\left[t_{1}\left[s_{1}, \ldots, s_{n}\right], \ldots, t_{n}\left[s_{1}, \ldots, s_{n}\right]\right] \in \operatorname{pol}(C)
$$

This means that arbitrary patterns can be extended to elements of the code $C$.
Proposition 14. Let $F \subseteq T_{\Gamma}\left(X_{n}\right)$ and $G \subseteq T_{\Delta}\left(X_{n}\right)$ be polypodic codes composable through $h: T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right)$. 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 $G \subseteq T_{\Delta}\left(X_{n}\right)$ is complete, for all $w_{1}, \ldots, w_{n} \in T_{\Delta}\left(X_{n}\right)$ we can find $u, v_{1}, \ldots, v_{n} \in T_{\Delta}\left(X_{n}\right)$ such that

$$
t=u\left[w_{1}\left[v_{1}, \ldots, v_{n}\right], \ldots, w_{n}\left[v_{1}, \ldots, v_{n}\right]\right] \in \operatorname{pol}(G)
$$

Choose $\bar{w} \in T_{\Gamma}\left(X_{n}\right)$ in such a way that $h(\bar{w})=t$.
Since $F$ is complete we can find trees $\bar{u}, \bar{v}_{1}, \ldots, \bar{v}_{n} \in T_{\Gamma}\left(X_{n}\right)$ such that

$$
s=\bar{u}\left[\bar{w}\left[\bar{v}_{1}, \ldots, \bar{v}_{n}\right], \ldots, \bar{w}\left[\bar{v}_{1}, \ldots, \bar{v}_{n}\right]\right] \in \operatorname{pol}(F)
$$

Thus $h(s) \in \operatorname{pol}(C)$, i.e.

$$
h(\bar{u})\left[t\left[h\left(\bar{v}_{1}\right), \ldots, h\left(\bar{v}_{n}\right)\right], \ldots, t\left[h\left(\bar{v}_{1}\right), \ldots, h\left(\bar{v}_{n}\right)\right] \in \operatorname{pol}(C)\right] .
$$

We set

$$
u^{\prime}=h(\bar{u})[u, \ldots, u] \quad \text { and } \quad v_{i}^{\prime}=v_{i}\left[h\left(\bar{v}_{1}\right), \ldots, h\left(\bar{v}_{n}\right)\right], i=1, \ldots, n .
$$

Then

$$
u^{\prime}\left[w_{1}\left[v_{1}^{\prime}, \ldots, v_{n}^{\prime}\right], \ldots, w_{n}\left[v_{1}^{\prime}, \ldots, v_{n}^{\prime}\right]\right]=h(s) \in \operatorname{pol}(C)
$$

proving that $C$ is a completepolypodic code.
From the inclusion $\operatorname{pol}(C) \subseteq \operatorname{pol}(G)$ and the completeness of $C$, we get the completeness of $G$.

Proposition 15. Assume that $F \subseteq T_{\Gamma}\left(X_{n}\right), G \subseteq T_{\Delta}\left(X_{n}\right)$ and $C=F \circ_{h} G$, with $h: T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right)$. 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 $t \in T_{\Gamma}\left(X_{n}\right)$ 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 $s \in T_{\Delta}\left(X_{n}\right)$ such that $G \cup s$ is a polypodic code. Add to $\Gamma$ the new $n$-ranked symbol $\gamma_{s}$ and define

$$
h_{s}: T_{\Gamma \cup \gamma_{s}}\left(X_{n}\right) \rightarrow T_{\Gamma}\left(X_{n}\right)
$$

by setting $h_{s}(\gamma)=h(\gamma)$, for each $\gamma \in \Gamma$, and $h_{s}\left(\gamma_{s}\right)=s$.
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

$$
h_{s}\left(F \cup \gamma_{s}\right)=h(F) \cup h\left(\gamma_{s}\right)=C \cup s
$$

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 $C \subseteq T_{\Delta}\left(X_{n}\right)$, for any polypodic code $G \subseteq T_{\Delta}\left(X_{n}\right)$,

$$
C \quad \text { decomposes in } G \quad i f f \quad \operatorname{pol}(C) \subseteq \operatorname{pol}(G) .
$$

In particular, $C$ isindecomposable iff pol $(C)$ is a maximal subpolypode of $T_{\Delta}\left(X_{n}\right)$.
Proof. In one direction we have nothing to show. Assume next that $\operatorname{pol}(C) \subseteq$ $\operatorname{pol}(G)$ and let $\bar{G}=a l p h_{G}(C)$; according to the previous proposition we must prove that $\bar{G}=G$. Since $\operatorname{pol}(C) \subseteq \operatorname{pol}(\bar{G})$, we deduce that $C=F \circ \bar{G}$. By virtue of Proposition $15 \bar{G}$ is maximal and $\bar{G} \subseteq G$. Thus $\bar{G}=G$ as wanted.

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

## 8. Decomposition theorem

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

$$
C=G_{1} \circ G_{2} \circ \ldots \circ G_{\kappa}
$$

of some indecomposable polypodic codes $G_{1}, G_{2}, \ldots, G_{\kappa}$.
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 $C \subseteq T_{\Delta}\left(X_{n}\right)$, we set

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

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

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

Proof. If $h: T_{\Gamma}\left(X_{n}\right) \rightarrow T_{\Delta}\left(X_{n}\right)$ is injective and $h(\Gamma)=G$, then $\operatorname{card}(C)=$ $\operatorname{card}(F)$ and

$$
\begin{aligned}
\alpha(C)-\alpha(F) & =\sum_{t \in C} \operatorname{size}(t)-\sum_{s \in F} \operatorname{size}(s) \\
& =\sum_{s \in F}(\operatorname{size}(h(s))-\operatorname{size}(s)) .
\end{aligned}
$$

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

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

Thus continuing the previous string of equalities we get

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

Obviously $\sum_{s \in F} \operatorname{size}_{\gamma}(s) \geqq 1$ for all $\gamma \in \Gamma$. Thus

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

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 aredone. Otherwise

$$
C=F \circ G .
$$

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

$$
F=F_{1} \circ \ldots \circ F_{\lambda}, \quad G=G_{1} \circ \ldots \circ G_{\lambda}
$$

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

[1] 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. SpringerVerlag, 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) 278295.
[8] S. Mantaci and A. Restivo, Tree Codes and Equations, in Proc. of the $3^{\text {rd }}$ International Conference DLT'97, edited by S. Bozapalidis. Thessaloniki (1998) 119-132.
[9] 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.

[^1]
[^0]:    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.

[^1]:    To access this journal online:
    www.edpsciences.org

