Theoretical Computer Science 18 (1982) 115–148
North-Holland Publishing Company

RECOGNIZABLE FORMAL POWER SERIES ON TREES

J. BERSTEL and C. REUTENAUER
L.I.T.P. and Institut de Programmation, Université Paris 6, Paris, France

Communicated by M. Nivat
Received May 1980
Revised October 1980

1. Introduction

Trees are a very basic object in computer science. They intervene in nearly any domain, and they are studied for their own, or used to represent conveniently a given situation. There are at least three directions where investigations on trees themselves are motivated, and this for different reasons. First, the notion of tree is the basis of algebraic semantics (Nivat [19], Rosen [22], etc.). In this context, the study of special languages of trees (i.e. forests), their classification, and their behaviour under various types of transformations are of great importance (Arnold [1], Dauchet [8], Lilin [16], Mongy [18]). By essence, work in this area is an extension of the algebraic theory of languages; trees and languages are in fact directly related via the derivation trees of an algebraic grammar (Thatcher [25]). A second topic heavily related to trees concerns data structures. Trees, mainly binary trees and its variants, constitute one of the most widely known data structures (see e.g. Knuth [15]). The analysis of the worst-case, expected or average running time behaviour of certain algorithms requires sometimes long and delicate computations (Flajolet [10], Kemp [14], Flajolet and Steyaert [12]). Finally, trees occupy a distinguished place in the enumeration of graphs and maps, both because of the simplicity of their structure and for the relationship between their encodings and algebraic languages. The nature of the enumerating series, and especially the question whether they are algebraic or not, is one of the central problems in this domain (Cori [7], Chottin [4]).

We propose here a theory of formal power series on trees, and present some of their basic properties together with various examples of applications which, as we hope, will show the interest of its development within the framework we just sketched.

A formal power series on trees is a function which associates a number to each tree. Thus we could also have called them 'tree functions', in analogy with the term 'word function' used by several authors (Paz and Salomaa [20], Cobham [5]) as an equivalent denomination for formal power series on words. The main goal of a formal power series is to count, or to represent the result of some computation on trees. Thus one can count the 'multiplicity' in some recognition device, but more concrete examples (height of a tree, evaluation of arithmetic expressions, pattern matching) will be considered. Next, power series are classified according to the amount of difficulty involved to compute them. The so called linear representations of magmas introduced in Section 3 give a formalization of linear computation methods and rely on standard mathematical concepts (multilinear functions on a vector space). Formal power series computed in that way are called recognizable. They are the main object of this paper and two characterizations will be given in Section 6 and 8.

From the point of view of formal language theory, formal power series on trees appear as an extension of the classical theory of formal power series on words (Salomaa and Soittola [23], Eilenberg [9]), developing further the correspondence between properties of (set of) words and trees. They are also a generalization of the notion of forest, the refinement consisting in the introduction of multiplicities. It would have been tedious to present a systematic investigation of all properties of series on words which carry over to formal power series on trees. We focused our attention on three of them. First we prove (Theorem 6.2 and 6.4) the equivalence of the definition by linear representation, and as solutions of systems of linear equations. The first definition corresponds to bottom–up computations, whereas the second is a global, top–down one. Then we investigate the analogue, for formal power series, of the relationship between context-free languages and frontiers of recognizable forest, and prove that the same facts hold in that case (Theorem 8.1 and 8.3). Our third result in this direction is a pumping lemma (Theorem 9.2) for recognizable power series on trees. There are interesting applications of this lemma; thus for instance the set of AVL-trees cannot be the support of a recognizable power series (Example 9.3).

Among the applications, the most intriguing one is perhaps the use of formal power series for the evaluation of arithmetic expressions (Example 4.2, 6.2 and 9.1). It is well known that arithmetic expressions are representable by trees. The function which to an arithmetic expression associates its value is a formal power series, and we verify that it is recognizable when division is forbidden, but is no longer recognizable when division is admitted. Another type of application repeatedly discussed in the sequel concerns enumeration. We show that path length, number of occurrences of a pattern, and others are recognizable power series, but height is not. More important from the methodological point of view is perhaps the fact that the system of linear equation satisfied by some recognizable formal power series on trees yields, without further computation, equations for the enumerating functions counting the property over all trees of a given size. Moreover, the enumerating functions are always algebraic provided the series on trees is recognizable (Proposition 7.2).

Section 2 contains basic definition concerning formal power series on trees. Linear representations of a free magma are introduced in Section 3, and are used to give a first definition of recognizable power series. Section 4 is devoted to a detailed description of several examples. The closure under Hadamard product, shown in

Section 5, has mainly technical interest. In Section 6, we prove the equivalence between recognizable power series and solutions of proper systems of linear equations, and give such systems for some examples. The well-known function 'glove' investigated in Section 7 associates, to a recognizable series on trees, an algebraic power series in several noncommutative variables. As a corollary, we show that the corresponding enumerating series is also algebraic. Two examples are computed. Section 8 contains the proof that the frontier of a recognizable series on trees is algebraic and conversely. In Section 9, we show a pumping lemma and use it to prove that several power series on trees are not recognizable.

The results proved here are, as we already said, generalization of corresponding properties of recognizable forests which are already known for a while (Thatcher [25, 26], Brainerd [3], Thatcher and Wright [27], Maibaum [17], Arnold [1]), even if the proof are generally more complicated. They are of course also related to recognizable power series on words [9, 23].

The definition of power series on trees by means of equations and their relation to enumerating series was already used extensively by Flajolet [10]. A first version of this paper [2] was presented at the 5th C.L.A.A.P.

2. Formal power series on trees

In this section, we give the basic definitions concerning formal power series on trees. Let F be a set of function symbls, that is a graded alphabet

$$F = F_0 \cup F_1 \cup \cdots \cup F_p \cup \cdots.$$

The elements in F_p are the function symbols of *arity* p. We denote by $M(F)$ the *free F-magma* generated by F (see e.g. Cohn [6]). The elements in $M(F)$ are called *trees*. If t is a tree, and $t \notin F_0$, then there exist an integer $p \geq 0$, a symbol function $f \in F_p$, and trees $t_1, \ldots, t_p$ such that

$$t = f(t_1, \ldots, t_p).$$

Some time it is more suggestive to employ, instead of this notation, the more pictorial representation

$$t = \begin{array}{c} f \\ \diagup \cdots \diagdown \\ t_1 \qquad t_p \end{array}$$

We shall use indistinctly one or the other notation.

We assume in the sequel that the set F of function symbols is *finite*. This is not an essential restriction in most of the subsequent developments, but simplifies the exposition without a real loss of generality.

Definition. Let k be a commutative field. A *formal power series on $M(F)$ with coefficients* in k is a mapping

$$S : M(F) \to k.$$

The value of S for a tree $t \in M(F)$ is denoted by $S(t)$ or (S, t) and is called the *coefficient* of t in S. The series S is written in an expanded notation as

$$S = \sum_{t \in M(F)} (S, t)t.$$

The set of all formal power series on $M(F)$ with coefficients in k is denoted by $k\{\{F\}\}$.

Note that we define only formal power series with coefficients in a field, whereas in the 'classical' theory of formal power series on the free monoïd [9, 23] the coefficients are taken in a semiring. We do this in order to simplify the exposition, and also because we are mainly interested in applications where the coefficients have a precise numerical meaning. As already mentioned above, formal power series on trees are to be used in counting processes, and the result of such a process for a given tree t is precisely the value of the function S for t.

Example 2.1. The *height* of a tree t in $M(F)$ is inductively defined by

$$\text{height}(t) = \begin{cases} 0 & \text{if } t \in F_0, \\ 1 + \max\{\text{height}(t_1), \ldots, \text{height}(t_p)\} & \text{if } t = f(t_1, \ldots, t_p). \end{cases}$$

Thus height is a formal power series on $M(F)$ with coefficients in $k = \mathbb{Q}$ (= the field of rational numbers). Other examples will be given below (Sections 4, 6 and 9).

The set $k\{\{F\}\}$ of formal power series is converted into a vector space by the formulas

$$(S_1 + S_2, t) = (S_1, t) + (S_2, t),$$

$$(\alpha S, t) = \alpha(S, t) \quad (\alpha \in k).$$

The set $k\{\{F\}\}$ has also a structure of F-magma. Consider indeed $f \in F_p$, and $S_1, \ldots, S_p$ in $M(F)$. The formal power series

$$S = f(S_1, \ldots, S_p)$$

is defined as follows

$$(S, t) = \begin{cases} (S_1, t_1)(S_2, t_2) \cdots (S_p, t_p) & \text{if } t = f(t_1, \ldots, t_p) \\ 0 & \text{otherwise.} \end{cases}$$

The *support* of a series S is the forest $\text{supp}(S) = \{t \mid (S_1, t) \neq 0\}$. We denote by $k\{F\}$ the subset of all elements in $k\{\{F\}\}$ having finite support. These elements are called *polynomials*. Then $k\{F\}$ is just the *free k-F-algebra* [6]. It is easily seen that $k\{F\}$ is a subvector space and a submagma of $k\{\{F\}\}$.

3. Recognizable formal power series

Before defining the recognizable formal power series, we introduce the notion of linear representation of a free magma. This is a straightforward extension of the corresponding construction for other algebraic structures, and is presumably interesting for itself.

Let V be a finite dimensional vector space over the (commutative) field k. We denote by

$$\mathscr{L}(V^p;\, V)$$

the set of p-linear mappings from V^p into V. In particular, if $p=0$, then V^0 can be identified with k, and $\mathscr{L}(V^0;\, V)$ can be identified with V: indeed, a linear mapping from k into V can be identified with its value on 1. Set

$$\mathscr{L} = \bigcup_{p\geq 0} (V^p;\, V).$$

Then the vector space V is a $\mathscr{L}$-magma with $\mathscr{L}(V^p;\, V)$ as the set of functions with arity p. Thus any mapping $\mu : F \to \mathscr{L}$ which maps F_p into $\mathscr{L}(V^p;\, V)$ converts V into a F-magma.

≈ BOTTOM-UP TREE AUTOMATON

Definition. A *linear representation* of the free magma $M(F)$ is a couple (V, μ), where V is a finite dimensional vector space over k, and where

$$\mu : F \to \mathscr{L}$$

maps F_p into $\mathscr{L}(V^p;\, V)$ for each $p \geq 0$.

Thus for each $f \in F_p$, $\mu(f): V^p \to V$ is p-linear, and since $M(F)$ is free, μ extends uniquely to a morphism

$$\mu : M(F) \to V$$

by the usual formula

$$\mu(t) = \mu(f)(\mu(t_1), \ldots, \mu(t_p))$$

for

$$t = f(t_1, \ldots, t_p).$$

Definition. Let S be a formal power series on $M(F)$. Then S is *recognizable* if there exists a triple (V, μ, λ), where (V, μ) is a linear representation of $M(F)$, and

$$\lambda : V \to k \quad \text{output function}$$

is a linear form, such that

$$(S, t) = \lambda(\mu(t)) \quad \text{for all } t \text{ in } M(F). \tag{3.1}$$

(V, μ, λ) is called a *representation for S*.

There are several remarks on this definition. First, and as we will show in the next section, the notion of representation of S is an 'arithmetization' (in the sense of [23]) of the well-known concept of bottom–up tree-automata (as defined by Thatcher [25]). Moreover a representation analogue to (3.1) has been given by Arnold [1]. He shows that a forest L is recognizable iff it can be written as

$$L = \{t \mid \lambda(\mu(t)) = 1\}$$

where λ, μ are defined as above, but the field k is replaced by the boolean semiring. In a manner completely parallel to the situation for recognizable forests, we will give a top–down definition of recognizable formal power series later, by means of systems of equations.

Next, our definition is an extension of the classical definition of recognizable power series over the free monoïd (Schützenberger [24], Salomaa and Soittola [23]). Let indeed X be an alphabet. Then X can be considered as a F-magma, where $F = F_0 \cup F_1$, and $F_0 = \perp$, $F_1 = X$; to each word w in X is associated the tree $w\perp$. Let now (V, μ, λ) be a representation of a series S over this magma. Then by definition each μx $(x \in X)$ is an endomorphism of V, and $\mu(\perp)$ is an element of V, say γ. Thus we have, for a tree $w\perp$,

$$\mu(w\perp) = \mu(w)\gamma, \qquad (S, w\perp) = \lambda(\mu(w\perp)) = \lambda \cdot \mu w \cdot \gamma.$$

This yields the usual definition of recognizable formal power series on the free monoïd X^*.

Finally, it is sometimes convenient to use the canonical isomorphism between $\mathscr{L}(V^p; V)$ and the space ξ_p of linear mappings of the p-fold tensor product

$$V^{\otimes p} = V \otimes \cdots \otimes V$$

into V. This isomorphism is given by associating, to $f \in \xi_p$, the p-linear mapping from V^p into V defined by

$$(v_1, \ldots, v_p) \mapsto f(v_1 \otimes \cdots \otimes v_p).$$

Setting

$$\xi = \bigcup \xi_p$$

a linear representation can consequently also be given by a mapping $\mu : F \to \xi$ such that $\mu(F_p) \subset \xi_p$.

Proposition 3.1. *The set of recognizable formal power series is a subvector space of* $k\{\{F\}\}$.

Proof. We first show that if S is recognizable and $\alpha \in k$, then αS is recognizable. Let (V, μ, λ) be a representation of S. Then $(V, \mu, \alpha\lambda)$ is a representation of αS, and consequently αS is recognizable.

Assume now that S_1 and S_2 are recognizable, and set $S = S_1 + S_2$. Let (V_i, μ_i, λ_i) be a representation of $S_i (i = 1, 2)$. Set $V = V_1 \times V_2$, and observe that for $\varphi \in \mathscr{L}(V_1^p; V_1)$, $\psi \in \mathscr{L}(V_2^p; V_2)$, the mapping

$$\eta : V^p \to V,$$

$$\eta[(v_1^1, v_1^2), \ldots, (v_p^1, v_p^2)] = (\varphi(v_1^1, \ldots, v_p^1), \psi(v_1^2, \ldots, v_p^2))$$

is p-linear. Consequently, the linear representation (V, μ) of $M(F)$ given by

$$\mu(f)[(v_1^1, v_1^2), \ldots, (v_p^1, v_p^2)] = (\mu_1(f)(v_1^1, \ldots, v_p^1), \mu_2(f)(v_1^2, \ldots, v_p^2))$$

is well defined, and it is immediately verified that

$$\mu(t) = (\mu_1(t), \mu_2(t)), \quad t \in M(F).$$

Thus, if $\lambda : V \to k$ is defined by

$$\lambda(v^1, v^2) = \lambda_1(v^1) + \lambda_2(v^2)$$

we get

$$\lambda(\mu(t)) = \lambda_1(\mu_1(t)) + \lambda_2(\mu_2(t)) = (S_1, t) + (S_2, t) = (S, t),$$

showing that (V, μ, λ) is a representation for S, whence S is recognizable.

The set of recognizable formal power series is also an sub-F-magma. This will be shown later (Proposition 6.5).

4. Examples

Example 4.1 (Counting arguments). The *length* $|t|$ of a tree t is the number of nodes of t, i.e.

$$|t| = \begin{cases} 1 & \text{if } t \in F_0, \\ 1 + |t_1| + \cdots + |t_p| & \text{if } t = f(t_1, \ldots, t_p). \end{cases}$$

The formal power series D defined by

$$(D, t) = |t|$$

is recognizable. We consider the more general case of the formal power series $D_f (f \in F)$ defined by

$$(D_f, t) = \text{number of occurrences of } f \text{ in } t.$$

We show that D_f is recognizable. Since F is finite, and $D = \sum_{f \in F} Df$ it follows by Proposition 3.1 that D is also recognizable.

To show that D_f is recognizable, we define a representation (V, μ, λ) as follows: $V = \mathbb{Q}^2$. Let e_1, e_2 be the canonical basis of V. Then for each g, it suffices to define

$\mu(g)$ on the basis vectors. We define: for $g \in F_q$, $g \neq f$,

$$\mu g(e_{i_1}, \ldots, e_{i_q}) = \begin{cases} e_1 & \text{if } i_1 = \cdots = i_q = 1, \\ e_2 & \text{if there is exactly one } j \text{ with } i_j = 2, \\ 0 & \text{otherwise;} \end{cases}$$

$$\mu f(e_1, \ldots, e_p) = \begin{cases} e_1 + e_2 & \text{if } i_1 = \cdots = i_q = 1, \\ e_2 & \text{if there is exactly one } j \text{ with } i_j = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Note that for $a \in F_0$, this definition implies that

$$a = \begin{cases} e_1 & \text{if } a \neq f, \\ e_1 + e_2 & \text{if } a = f \in F_0. \end{cases}$$

Finally, $\lambda(e_1) = 0$, $\lambda(e_2) = 1$.

We claim that

$$\mu t = e_1 + (D_f, t)e_2 \quad \text{for } t \in M(F). \tag{4.2}$$

This holds indeed if $t \in F_0$; otherwise

$$t = g(t_1, \ldots, t_q)$$

and by linearity

$$\mu(t) = \mu(g)[e_1 + (D_f, t_1)e_2, \ldots, e_1 + (D_f, t_q)e_2]$$

$$= \begin{cases} e_1 + \sum_{i=1}^{q} (D_f, t_i)e_2 & \text{if } g \neq f, \\ e_1 + e_2 + \sum_{i=1}^{p} (D_f, t_i)e_2 & \text{if } g = f \text{ (and } q = p). \end{cases}$$

This proves the claim (4.2). Next the definition of λ gives

$$\lambda(\mu(t)) = (D_f, t)$$

showing that (V, μ, λ) is a representation of D_f which therefore is recognizable.

Note that the finiteness of F is not required to show that D is recognizable. It suffices to define μf by formula (4.1) for each f in F. Note also that (in the case where F is finite) any linear combination of the D_f's is still recognizable. This is interesting when a 'weighted length' is considered.

Example 4.2 (Evaluation of arithmetic expressions). Consider $F = F_0 \cup F_1 \cup F_2$, where $F_1 = \neg$, $F_2 = \{+, \times\}$ ($\neg$ is the function symbol for the unary substraction, that is the mapping $\alpha \mapsto -\alpha$). It is well known that a tree in this free magma represents an arithmetic expression. Such an expression has a well-defined numerical value provided a value is attributed to each of the elements $a \in F_0$ which are leafs of the tree, and provided the symbol functions are interpreted as usual. Thus consider the formal power series

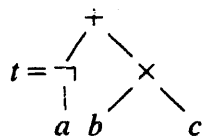
$$\mathrm{eval} : M(F) \to \mathbb{Q}$$

given by a fixed assignment

$$\mathrm{eval}(a) = \bar{a} \in \mathbb{Q} \quad \text{for each } a \in F_0$$

and extended to trees in the natural way. Then $\mathrm{eval}(t)$ is the value of the arithmetic expression represented by t for the given numerical values of the leafs.

For example, if



and $\bar{a} = 7$, $\bar{b} = 2$, $\bar{c} = 3$, then $\mathrm{eval}(t) = -1$.

We claim that eval is a *recognizable formal power series*.

We construct the following representation (V, μ, λ) for the series eval. As before, $V = \mathbb{Q}^2$, $\{e_1, e_2\}$ is its canonical base, and $\mu : F \to \mathscr{L}$ is defined as follows:

$$\mu(a) = e_1 + \bar{a}e_2 \quad \text{for } a \in F_0,$$

$$\mu(\neg)(e_1) = e_1, \qquad \mu(\neg)(e_2) = -e_2,$$

$$\mu(+)(e_1, e_1) = e_1, \qquad \mu(+)(e_1, e_2) = \mu(+)(e_2, e_1) = e_2, \qquad \mu(+)(e_2, e_2) = 0$$

$$\mu(\times)(e_1, e_1) = e_1, \qquad \mu(\times)(e_1, e_2) = \mu(\times)(e_2, e_1) = 0, \qquad \mu(\times)(e_2, e_2) = e_2.$$

$$\lambda(e_1) = 0, \qquad \lambda(e_2) = 1.$$

As above, we verify that

$$\mu(t) = e_1 + \mathrm{eval}(t)e_2, \quad t \in M(F). \tag{4.3}$$

By definition, (4.3) holds if $t \in F_0$. Next, one has

$$\begin{aligned}\mu(\neg t) &= \mu(\neg)(e_1 + \mathrm{eval}(t)e_2) = e_1 - \mathrm{eval}(t)e_2\\ &= e_1 + \mathrm{eval}(\neg t)e_2,\end{aligned}$$

$$\begin{aligned}\mu(+t_1t_2) &= \mu(+)(e_1 + \mathrm{eval}(t_1)e_2, e_1 + \mathrm{eval}(t_2)e_2)\\ &= e_1 + (\mathrm{eval}(t_1) + \mathrm{eval}(t_2))e_2 = e_1 + \mathrm{eval}(+t_1t_2)e_2,\end{aligned}$$

finally,

$$\begin{aligned}\mu(\times t_1t_2) &= \mu(\times)(e_1 + \mathrm{eval}(t_1)e_2, e_1 + \mathrm{eval}(t_2)e_2)\\ &= e_1 + \mathrm{eval}(t_1)\mathrm{eval}(t_2)e_2 = e_1 + \mathrm{eval}(\times t_1t_2)e_2.\end{aligned}$$

By (4.3) we get for all $t \in M(F)$

$$\mathrm{eval}(t) = \lambda(\mu(t))$$

showing that eval is recognizable.

Example 4.3 (Characteristic series). Given a forest L, we define the *characteristic series* $\boldsymbol{L}$ of L by

$$(\boldsymbol{L}, t) = \begin{cases} 1 & \textit{if } t \in L, \\ 0 & \text{otherwise.} \end{cases}$$

If L is a recognizable forest, then $\boldsymbol{L}$ is a recognizable formal power series.

Let indeed $\mathscr{A}$ be a (deterministic) bottom-up tree automaton recognizing L. Then [25] $\mathscr{A}$ is composed of a finite set of states Q, an initial state q_-, a set of final states $Q_+ \subset Q$ and, for each $p \geq 0$, and $f \in F_p$, a function $\hat{f} : Q^p \to Q$. Thus Q, equipped with the $\hat{f}$'s, is a F-magma. A tree is accepted by $\mathscr{A}$ (i.e. t is in L) iff $\hat{t}(q_-) \in Q_+$.

Consider now the k-vector space $V = k^Q$ with the canonical base $(e_q)_{q \in Q}$. For each $f \in F_p$, let

$$\mu(f) : V^p \to V$$

be given by

$$\mu(f)(e_{q_1}, \ldots, e_{q_p}) = e_{\hat{f}(q_1, \ldots, q_p)} \tag{4.4}$$

and define $\lambda : V \to k$ by

$$\lambda(e_q) = \begin{cases} 1 & \text{if } q \in Q_+, \\ 0 & \text{otherwise.} \end{cases}$$

Consider a tree t in $M(F)$, and set $q = \hat{t}(q_-)$. Then it is easily seen that $\mu(t) = e_q$, whence

$$\lambda(\mu(t)) = (\boldsymbol{L}, t).$$

Note that in order to get (4.4), we assumed that the automaton $\mathscr{A}$ is deterministic. In the case of nondeterministic automata, each $\hat{f}(q_1, \ldots, q_p)$ is a subset of Q, and replacing (4.4) by

$$\mu(f)(e_{q_1}, \ldots, e_{q_p}) = \sum_{q \in \hat{f}(q_1, \ldots, q_p)} e_q$$

defines a formal power series which takes into account the multiplicity of acceptance of a tree.

Example 4.4. The 'converse' of the preceding example is not true, that is: the support of a recognizable power series is not necessarily a recognizable forest. Indeed, consider for instance the recognizable series $S_a - S_b$ (with the notations of Example 4.1). Its support is $\{t \mid (S_a, t) \neq (S_b, t)\}$ which is not a recognizable forest.

Example 4.5. The formal power series height, defined in Section 2, is not recognizable. This will be proved in Section 9.

Example 4.6. The *path length* [15] of a tree is defined to be the sum, taken over all nodes, of the lengths of the paths from the root to each node. Denote by $\mathrm{PL}(t)$ the path length of the tree t. Then PL is a recognizable formal power series. We show this in Section 6.

5. Hadamard product

Given two formal power series $S_1, S_2 \in k\{\{F\}\}$, we call Hadamard product of S_1 and S_2, and denote by $S_1 \odot S_2$, the series defined by

$$(S_1 \odot S_2, t) = (S_1, t)(S_2, t).$$

Since k is a field, one has

$$(S_1 \odot S_2, t) \neq 0 \text{ iff } (S_1, t) \neq 0 \text{ and } (S_2, t) \neq 0.$$

Therefore

$$\mathrm{supp}(S_1 \odot S_2) = \mathrm{supp}(S_1) \cap \mathrm{supp}(S_2).$$

This shows the relation between Hadamard product and intersection.

Proposition 5.1. *If S_1 and S_2 are recognizable formal power series on $M(F)$, then $S_1 \odot S_2$ is recognizable.*

Proof. Let (V_i, μ_i, λ_i) be a representation of S_i for $i = 1, 2$. Using the isomorphism mentioned in Section 3, each $\mu_i(f)$, for $f \in F_p$, is a linear mapping from $V_i^{\otimes p}$ into V_i. Now define a representation (V, μ, λ) by

$$V = V_1 \otimes V_2, \qquad \lambda = \lambda_1 \otimes \lambda_2$$

and for $p \geq 0$, $f \in F_p$, $v_1^1, \ldots, v_p^1 \in V_1$, $v_1^2, \ldots, v_p^2 \in V_2$

$$\mu(f)[(v_1^1 \otimes v_1^2) \otimes \cdots \otimes (v_p^1 \otimes v_p^2)] =$$
$$= \mu_1(f)(v_1^1 \otimes \cdots \otimes v_p^1) \otimes \mu_2(f)(v_1^2 \otimes \cdots \otimes v_p^2).$$

$\mu(f)$ is indeed a linear mapping in view of the canonical isomorphism

$$\mathrm{End}(V_1^{\otimes p}, V_1) \otimes \mathrm{End}(V_2^{\otimes p}, V_2) \simeq \mathrm{End}(V^{\otimes p}, V).$$

We claim that for all t in $M(F)$,

$$t = \mu_1 t \otimes \mu_2 t \tag{5.1}$$

This is clear if $t \in F_0$. If $t = f(t_1, \ldots, t_p)$, then

$$\mu(t) = \mu(f)(\mu t_1 \otimes \cdots \otimes \mu t_p)$$
$$= \mu(f)((\mu_1 t_1 \otimes \mu_2 t_1) \otimes \cdots \otimes (\mu_1 t_p \otimes \mu_2 t_2))$$

by the induction hypothesis. Thus

$$\mu(t) = \mu_1(f)(\mu_1 t_1 \otimes \cdots \otimes \mu_1 t_p) \otimes \mu_2(f)(\mu_2(t_1) \otimes \cdots \otimes \mu_2(t_p))$$
$$= \mu_1(t) \otimes \mu_2(t).$$

Thus (5.1) is verified. Finally

$$\lambda\mu(t) = (\lambda_1 \otimes \lambda_2)(\mu_1(t) \otimes \mu_2(t)) = \lambda_1(\mu_1(t))\lambda_2(\mu_2(t))$$
$$= (S_1, t)(S_2, t) = (S_1 \odot S_2, t)$$

showing that (V, μ, λ) is a representation for $S_1 \odot S_2$

6. Systems of linear equations

As already mentioned, a representation can be considered as a bottom–up automaton. Systems of linear equations are then analogue of top–down automata. We show in this section the equivalence of these two definitions.

Let $\Xi = \{\xi_1, \ldots, \xi_n\}$ be a set (of 'variables'). We consider the set

$$F' = F'_0 \cup F'_1 \cup \cdots \cup F'_p \cup \cdots$$

of function symbols derived from F by adjoining Ξ to the set of 0-ary symbols:

$$F'_0 = F_0 \cup \Xi \quad \text{and} \quad F'_p = F_p \quad (p \geq 1).$$

Now consider a sequence $S = (S_1, \ldots, S_n)$ of formal power series in $k\{\{F\}\}$ and define a mapping $\omega = \omega_S$

$$\omega : F'_0 \to k\{\{F\}\},$$
$$\omega(a) = a \quad \forall a \in F_0, \qquad \omega(\xi_i) = S_i, \quad i = 1, \ldots, n.$$

This mapping extends uniquely to a morphism of k-F-algebras, still denoted by ω

$$\omega : k\{F'\} \to k\{\{F\}\}.$$

In particular, for $t = f(t_1, \ldots, t_p)$,

$$\omega(t) = f(\omega(t_1), \ldots, \omega(t_p)).$$

If $p \in k\{F'\}$, we also write $p[S_1, \ldots, S_n]$ for $\omega(p)$. This notation is consistent since $\omega(p)$ is obtained by substituting S_i to ξ_i in p. It follows in particular that for

$t = f(t_1, \ldots, t_p) \in M(F')$

$$t[S_1, \ldots, S_n] = f(t_1[S_1, \ldots, S_n], \ldots, t_p[S_1, \ldots, S_n]).$$

Now let $p_1, \ldots, p_n \in k\{F'\}$. A *solution of the system of linear equations*

$$\xi_i = p_i, \quad i = 1, \ldots, n \tag{6.1}$$

is any n-tuple $(S_1, \ldots, S_n)$ of formal power series satisfying

$$S_i = p_i[S_1, \ldots, S_n], \quad i = 1, \ldots, n. \tag{6.2}$$

Using the morphism ω defined by $S = (S_1, \ldots, S_n)$, (6.2) is equivalent to: $S_i = \omega(p_i)$, $i = 1, \ldots, n$.

The system (6.1) is called *proper* if ← every occurrence of a variable is "guarded"

$$\Xi \cap \mathrm{supp}(p_i) = \emptyset, \quad i = 1, \ldots, n.$$

Just a word to the specification 'linear'. This is because the variables appear only on the leafs of the trees in $\mathrm{supp}(p_i)$, and also for the similarity of the properties of such systems with the 'usual' linear systems.

Proposition 6.1. *A proper system of linear equations has one and only one solution.*

Proof. Assume first that $S = (S_1, \ldots, S_n)$ is a solution of (6.1). Then

$$S_i = \omega(p_i) = \sum_{t \in \mathrm{supp}(p_i)} (p_i, t)\omega(t), \quad i = 1, \ldots, n. \tag{6.3}$$

Since the system is proper, each $t \in \mathrm{supp}(p_i)$ either is in $M(F)$, or has the forms $t = f(t_1, \ldots, t_p)$ for some $f \in F_p$, $p \geq 1 t_1, \ldots, t_p \in M(F')$. Consequently, setting

$$P_i = \mathrm{supp}(p_i) \cap M(F), \qquad Q_i = \mathrm{supp}(p_i) - P_i$$

(6.3) becomes

$$S_i = \sum_{t \in P_i} (p_i, t)t + \sum_{f(t_1, \ldots, t_p) \in Q_i} (p_i, f(t_1, \ldots, t_p)) f(\omega(t_1), \ldots, \omega(t_p))$$

$$\forall i = 1, \ldots, n.$$

Therefore

$$(S_i, a) = (p_i, a) \quad \forall a \in F_0, \quad i = 1, \ldots, n \tag{6.4}$$

and if $s = g(s_1, \ldots, s_q) \in M(F)$, then

$$(S_i, s) = (p_i, s) + \sum_{g(t_1, \ldots, t_q) \in Q_i} (p_i, g(t_1, \ldots, t_q))(\omega(t_1), s_1) \cdots (\omega(t_q), s_q), \tag{6.5}$$

This formula shows that (S_i, s) is uniquely determined by p_i and by the values of the S_j's on trees of length strictly less than the length of s. By (6.4), the S_i's are uniquely determined on trees in F_0. Consequently, the solution is unique. But (6.4), (6.5) can also be considered as defining equations for the formal power series S_i: (6.4) gives the values of the S_i's on elements in F_0, and (6.5) allows to compute the values of the series by induction on the length. This proves the existence of a solution.

We now start the proof of the characterisation of recognizable formal power series through systems of equations.

Theorem 6.2. *Any recognizable formal power series is a component of the solution of some proper system of linear equations.*

Proof. Let S be a recognizable formal power series, and let (V, μ, λ) be a representation of S. Let $\Xi = \{\xi_1, \ldots, \xi_n\}$ be a base of the vector space V and let F' be as above, with $F'_0 = F_0 \cup \Xi$, in order to carry out the proof correctly, we need some additional notations. If $f \in F_p$ and $w = \xi_{i_1}\xi_{i_2} \ldots \xi_{i_p} \in \Xi^p$ we write $f(w)$ instead of $f(\xi_{i_1}, \ldots, \xi_{i_p})$. Thus $f(w) \in M(F')$; for $p = 0$, we set by convention $f(1) = f$.

Next we observe that Ξ^p can be identified to a base of $V^{\otimes p}$, with $w = \xi_{i_1} \ldots \xi_{i_p} \in \Xi^p$ representing the element $\xi_{i_1} \otimes \cdots \otimes \xi_{i_p}$ in $V^{\otimes p}$. Finally we consider the dual base $\{\xi'_1, \ldots, \xi'_n\}$ of Ξ.

Define polynomials $p_i \in k\{F'\}$ by

$$p_i = \sum_{p \geq 0} \sum_{\substack{f \in F_p \\ w \in \Xi^p}} [\xi'_i \circ \mu(f)(w)] f(w), \quad i = 1, \ldots, n. \tag{6.6}$$

Thus in (6.6), the coefficient $(p_i, f(w))$ is the ith component of the vector $\mu(f)(w) \in V$, with respect to the base Ξ. Let S_i be the series defined by the representation (V, μ, ξ'_i). We claim that

$$S_i = p_i[S_1, \ldots, S_n], \quad i = 1, \ldots, n.$$

First if $a \in F_0$, then

$$\begin{aligned}(p_i[S_1, \ldots, S_n], a) &= \sum_p \sum_{f,w} [\xi'_i \circ \mu(f)(w)](f(w)[S_1, \ldots, S_n], a) \\ &= \xi'_i \circ \mu(a)(1) = \xi'_1 \circ \mu a = (S_i, a)\end{aligned}$$

(of course, we have used here the convention made above identifying μa with $\mu a(1)$).

Next if $t = g(t_1, \ldots, t_q)$, then $(f(w)[S_1, \ldots, S_n], t)$ equals 0 except if $f = g$. Consequently (by induction)

$$\begin{aligned}(p_i[S_1, \ldots, S_n], t) &= \sum_{w \in \Xi^q} [\xi'_i \circ \mu g(w)](g(w)[S_1, \ldots, S_n], t) \\ &= \sum_{i \leq i_1, \ldots, i_q \leq n} [\xi'_i \circ \mu g(\xi_{i_1} \otimes \cdots \otimes \xi_{i_q})](g[S_{i_1}, \ldots, S_{i_q}], t) \\ &= \sum [\xi'_i \circ \mu g(\xi_{i_1} \otimes \cdots \otimes \xi_{i_q})](S_{i_1}, t_1) \cdots (S_{i_q}, t_q) \\ &= \sum [\xi'_i \circ \mu g(\xi_{i_1} \otimes \cdots \otimes \xi_{i_q})](\xi'_{i_1} \circ \mu t_1) \cdots (\xi'_{i_q} \circ \mu t_q)\end{aligned}$$

$$= \xi_i' \circ \mu g[\sum (\xi_{i_1}' \circ \mu t_1) \cdots (\xi_{i_q}' \circ \mu t_q) \xi_{i_1} \otimes \cdots \otimes \xi_{i_q}]$$

$$= \xi_i' \circ \mu g\Big[\bigotimes_{1 \leq j \leq q} (\xi_1' \circ \mu t_j)\xi_1 + \cdots + (\xi_n' \circ \mu t_j)\xi_n\Big]$$

$$= \xi_i' \circ \mu g[\mu t_1 \otimes \cdots \otimes \mu t_q]$$

$$= \xi_i' \circ \mu t = (S_i, t).$$

This achieves the verification of the claim.

Now let ξ_0 be a new variable and set

$$p_0 = \lambda(\xi_1)p_1 + \cdots + \lambda(\xi_n)p_n.$$

Then $(S, S_1, \ldots, S_n)$ is the solution of the system

$$p_i = \xi_i, \quad i = 0, 1, \ldots, n$$

since $\lambda = \lambda(\xi_1)\xi_1' + \cdots + \lambda(\xi_n)\xi_n'$.

In order to prove the converse of the theorem we first need a reduction step which shows that any proper system of equations can be 'simulated' by a simple one.

Lemma 6.3. *Let S be a formal power series on $M(F)$ which is a component of the solution of some proper system of linear equations. Then S is also a component of the solution of a proper system of linear equations*

$$\xi_i = p_i, \quad i = 1, \ldots, n$$

with

$$\mathrm{supp}(p_i) \subset F_0 \cup F(\Xi^*)$$

depth 0 or 1 trees on the r.h.s.

where

$$F(\Xi^*) = \{f(\xi_{i_1}, \ldots, \xi_{i_q}) \mid f \in F_q, \xi_{i_1}, \ldots, \xi_{i_q} \in \Xi\}.$$

Proof. The proof is in two steps. Define the height of a polynomial as the maximal height of the trees in its support. We first show the existence of a system of equations with right parts of height at most one. Then we transform this system to one of the desired form.

Let S be the first component of the solution of a proper system of linear equations

$$\xi_i = p_i, \quad i = 1, \ldots, n. \tag{6.7}$$

We proceed by double induction on the maximal height of the p_i's, and on the number of trees in the supports of the p_i's achieving this maximal height. Let t be a tree in the support of say p_i, and assume t is of maximal height among the trees in $\bigcup_{1 \leq i \leq n} \mathrm{supp}(p_i)$. Clearly we may assume that t has height h at least 2. Then

$$t = f(t_1, \ldots, t_p)$$

for some $f \in F_p$, $t_1, \ldots, t_p \in M(F')$. Further there is a nonempty subset $J \subset \{1, \ldots, p\}$

consisting of those indices j such that t_j has height $h-1$. Define a new set of variables $H=\{\eta_j \mid j\in J\}$ and a new system of equations

$$\eta_j = t_j, \quad j\in J,$$

$$\xi_{i'} = p_{i'}, \quad i'\in\{1,\ldots,n\}-\{i\},$$

$$\xi_i = p_i-(p_i,t)t+(p_i,t)t'$$

where $t'=f(s_1,\ldots,s_p)$ with $s_\lambda=\eta_\lambda$ for $\lambda\in J$, and $s_\lambda=t_\lambda$ for $\lambda\notin J$.

Clearly this system is proper since the t_j's and t' are trees of height at least 1. Next there is one tree less of maximal height h in this new system. Finally one checks easily that S is still the component of the solution of this system of equations corresponding to ξ_1.

Now we may assume that (6.7) is a system of equations with each tree in the supports of the p_i's having height 0 or 1. Introduce a new variable η_a for each $a\in F_0$, and define a new system of equations

$$\xi_i = p'_i, \quad i=1,\ldots,n, \qquad \eta_a = a \tag{6.8}$$

where

$$p'_i = \sum_{t\in \operatorname{supp}(p_i)} (p_i,t)t',$$

$$t' = \begin{cases} t & \text{if } t\in F_0, \\ f(z_1,\ldots,z_q) & \text{if } t=f(y_1,\ldots,y_q) \end{cases}$$

and

$$z_j = \begin{cases} y_j & \text{if } y_j\in \Xi, \\ \eta_a & \text{if } y_j = a\in F_0. \end{cases}$$

System (6.8) has the desired form.

Theorem 6.4. *Any formal power series on $M(F)$ which is a component of the solution of a proper system of linear equations is recognizable.*

Proof. Let S be a formal power series on $M(F)$, and suppose that S is the first component of the solution $(S_1,\ldots,S_n)$ of a proper system of linear equations

$$\xi_i = p_i, \quad i=1,\ldots,n. \tag{6.9}$$

We may assume that this system satisfies the conditions of the previous lemma. Consider the vector space $V=k^{\Xi}$, with $\Xi=\xi_1,\ldots,\xi_n$, as base and define a linear representation (V,μ) of $M(F)$ by setting, for $p\geq 0$, $f\in F_p$, $\eta_1,\ldots,\eta_p\in\Xi$

$$\mu(f)(\eta_1,\ldots,\eta_p) = \sum_{1\leq i\leq n} (p_i, f(\eta_1,\ldots,\eta_p))\xi_i.$$

Now define $S'_1, \ldots, S'_n$ to be the formal power series having the representations (V, μ, ξ'_i) $(i = 1, \ldots, n)$. The same computation as in the proof of Theorem 6.2 shows that $(S'_1, \ldots, S'_n)$ is a solution of the system (6.9). In view of Proposition 6.1, we have $S'_i = S_i$ for $i = 1, \ldots, n$. This concludes the proof.

This result has an interesting corollary which completes Proposition 3.1.

Proposition 6.5. *The set of recognizable formal power series on $M(F)$ is a sub-F-magma of $k\{\{F\}\}$.*

Proof. Let $f \in F_p$ and let $S_1, \ldots, S_p \in k\{\{F\}\}$. Then $S_1, \ldots, S_p$ are components of solutions of proper systems of linear equations. Putting these systems together, we may assume that $S_1, \ldots, S_p$ are the first p components of the solution of the system

$$\xi_i = p_i, \quad i = 1, \ldots, n.$$

Then $T = f(S_1, \ldots, S_p)$ is a component of the solution of the system

$$\xi_i = p_i, \quad i = 1, \ldots, n, \qquad \eta = f(\xi_1, \ldots, \xi_p).$$

According to Theorem 6.2 and 6.4, there is a proper system of linear equations for each recognizable formal power series and vice versa. We now give these systems of equations for some of the examples of Section 4.

Example 6.1 (Counting arguments). First, consider the equations

$$\eta = \sum_{a \in F_0} a + \sum_{p \geq 1} \sum_{f \in F_p} f(\eta, \eta, \ldots, \eta).$$

Let B be the solution. Then

$$B = \sum_{a \in F_0} a + \sum_{p \geq 1} \sum_{f \in F_p} f(B, B, \ldots, B).$$

Consequently $(B, t) = 1$ for each tree t in $M(F)$, i.e. B is the characteristic series of $M(F)$.

Now fix a symbol f in some F_p, and consider the equation

$$\xi = f(\eta, \ldots, \eta) + \sum_{q \geq 0} \sum_{g \in F_q} (g(\xi, \eta, \ldots, \eta) + g(\eta, \xi, \eta, \ldots, \eta) + \cdots + g(\eta, \ldots, \eta, \xi)).$$

Then the solution of this equation (together with the previous one) is the power series D_f. Indeed, let S be the power series which solves the equation. Then for a tree $t = g(t_1, \ldots, t_q)$, one has

$$(S, t) = (S, t_1) + (S, t_2) + \cdots + (S, t_q) + \delta$$

where $\delta = 0$ or 1 according to $g \neq f$ or $g = f$. Thus $S = D_f$.

Example 6.2 (Evaluation of arithmetic expression). With the notations of Example 4.2, consider the system of equations

$$
\begin{aligned}
\xi &= \sum_{a\in F_0} \bar{a}a - \begin{array}{c}\neg\\ |\\ \xi\end{array} + \begin{array}{c}+\\ \diagup\ \diagdown\\ \eta\quad\ \xi\end{array} + \begin{array}{c}+\\ \diagup\ \diagdown\\ \xi\quad\ \eta\end{array} + \begin{array}{c}\times\\ \diagup\ \diagdown\\ \xi\quad\ \xi\end{array},\\
\eta &= \sum_{a\in F_0} a + \begin{array}{c}\neg\\ |\\ \eta\end{array} + \begin{array}{c}+\\ \diagup\ \diagdown\\ \eta\quad\ \eta\end{array} + \begin{array}{c}\times\\ \diagup\ \diagdown\\ \eta\quad\ \eta\end{array}.
\end{aligned}
\tag{6.10}
$$

Then the same verification as above shows that the solution of this system is (eval, B), where B is the characteristic series of the magma. For illustration, we carry out the computation of eval(t) where

$$
t = \begin{array}{c} + \\ \diagup\qquad\diagdown \\ \begin{array}{c}\neg\\ |\\ a\end{array}\qquad \begin{array}{c}\times\\ \diagup\ \diagdown\\ b\quad\ c\end{array}\end{array}, \qquad \bar{a}=7,\ \bar{b}=2,\quad \bar{c}=3.
$$

According to (6.10)

$$
\begin{aligned}
\operatorname{eval}(t) &= \operatorname{eval}\left(\begin{array}{c}\neg\\ |\\ a\end{array}\right)\cdot B\left(\begin{array}{c}\times\\ \diagup\ \diagdown\\ b\quad\ c\end{array}\right) + B\left(\begin{array}{c}\neg\\ |\\ a\end{array}\right)\operatorname{eval}\left(\begin{array}{c}\times\\ \diagup\ \diagdown\\ b\quad\ c\end{array}\right)\\
&= \operatorname{eval}\left(\begin{array}{c}\neg\\ |\\ a\end{array}\right) + \operatorname{eval}\left(\begin{array}{c}\times\\ \diagup\ \diagdown\\ b\quad\ c\end{array}\right)\\
&= -\operatorname{eval}(a) + \operatorname{eval}(b)\operatorname{eval}(c)\\
&= -\bar{a} + \bar{b}\bar{c} = -1.
\end{aligned}
$$

Example 6.3 (Path length). For simplicity, we assume the trees binary, i.e. $F_0 = \{\square\}$, $F_2 = \{\bigcirc\}$. Then by definition, the path length of a tree t is defined as

$$
\operatorname{PL}(t) = \begin{cases} 0 & \text{if } t = \square,\\ \operatorname{PL}(t_1) + |t_1| + \operatorname{PL}(t_2) + |t_2| & \text{if } t = \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ t_1\quad\ t_2\end{array}. \end{cases}
$$

Consider the following system of equations:

$$
\begin{aligned}
\zeta &= \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \eta\quad\ \xi\end{array} + \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \xi\quad\ \eta\end{array} + \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \zeta\quad\ \eta\end{array} + \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \eta\quad\ \zeta\end{array},\\
\xi &= \square + \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \xi\quad\ \eta\end{array} + \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \eta\quad\ \xi\end{array} + \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \eta\quad\ \eta\end{array},\\
\eta &= \square + \begin{array}{c}\bigcirc\\ \diagup\ \diagdown\\ \eta\quad\ \eta\end{array}.
\end{aligned}
$$

Then the solution is (PL, D, B), where

$$(B, t) = 1 \quad \text{for each } t \in M(F), \qquad (D, t) = |t| \quad \text{for each } t \in M(F).$$

7. The function 'glove'

We now investigate the relation among formal series on trees, and formal power series on words. The correspondance will be made by means of the function 'glove', which associates, to a given tree, the word obtained by listing the nodes in prefix order. The main result of this section states that the glove of a recognizable tree-series is an algebraic word-series. We believe that the consequences of this theorem for enumeration problems are important. As will be illustrated by examples, we get a tool which automatically delivers the generating function for counting 'recognizable' properties of trees, and furthermore, all these generating functions are algebraic.

The function

$$\mathbf{glove} : M(F) \to F$$

is defined by

$$\mathbf{glove}(a) = a \quad \text{if } a \in F_0,$$

$$\mathbf{glove}(f(t_1, \ldots, t_p)) = f\,\mathbf{glove}(t_1) \cdots \mathbf{glove}(t_p)$$

if $f \in F_p$, $t_1, \ldots, t_p \in M(F)$.

It is folkore that **glove** is injective. The definition is extended to formula power series, i.e.

$$\mathbf{glove} : k\{\{F\}\} \to k\langle\langle F\rangle\rangle$$

(we use $k\langle\langle F\rangle\rangle$ to denote the set of formal power series over the free monoïd F^*, cf. [23]) by setting, for $S \in k\{\{F\}\}$,

$$(\mathbf{glove}(S), w) = 0 \quad \text{if } w \notin \mathbf{glove}(M(F)),$$

$$(\mathbf{glove}(S), \mathbf{glove}(t)) = (S, t), \quad t \in M(F). \tag{7.1}$$

This definition makes sense since *glove* is injective. An equivalent notation for (7.1), using expansion, is

$$\mathbf{glove}(S) = \sum_{t \in M(F)} (S, t)\mathbf{glove}(t).$$

It follows that for $f \in F_p$, $S_1, \ldots, S_p \in k\{\{F\}\}$

$$\mathbf{glove}(f(S_1, \ldots, S_p)) = f\,\mathbf{glove}(S_1) \cdots \mathbf{glove}(S_p). \tag{7.2}$$

Indeed, since

$$f(S_1,\ldots,S_p)=\sum_{t_1,\ldots,t_p\in M(F)}(S_1,t_1)\cdots(S_p,t_p)f(t_1,\ldots,t_p),$$

$$\begin{aligned}\mathbf{glove}(f(S_1,\ldots,S_p)) &= \sum_{t_1,\ldots,t_p\in M(F)}(S_1,t_1)\cdots(S_p,t_p)\mathbf{glove}(f(t_1,\ldots,t_p))\\ &= \sum_{t_1,\ldots,t_p\in M(F)}(S_1,t_1)\cdots(S_p,t_p)f\,\mathbf{glove}(t_1)\cdots\mathbf{glove}(t_1)\cdots\mathbf{glove}(t_p)\\ &= f\Big(\sum_{t_1\in M(F)}(S_1,t_1)\mathbf{glove}(t_1)\Big)\cdots\Big(\sum_{t_p\in M(F)}(S_p,t_p)\mathbf{glove}(t_p)\Big)\\ &= f\,\mathbf{glove}(S_1)\cdots\mathbf{glove}(S_p).\end{aligned}$$

This proves (7.2).

Theorem 7.1. *If S is a recognizable formal power series on $M(F)$, then* $\mathbf{glove}(S)$ *is algebraic. More precisely, if $(S_1,\ldots,S_n)$ is the solution of the proper system of linear equations over $M(F)$:*

$$\xi_i=p_i,$$

then $(\mathbf{glove}(S_1),\ldots,\mathbf{glove}(S_n))$ *is the solution of the proper algebraic system of equations over F^*:*

$$\xi_i=\mathbf{glove}(p_i),\quad i=1,\ldots,n.$$

Proof. Let $(S_1,\ldots,S_n)$ be the solution of the proper system

$$\xi_i=p_i,\quad i=1,\ldots,n$$

where $p_i\in k\{F'\}$, $F'=F\cup\Xi$ and $\Xi=\{\xi_1,\ldots,\xi_n\}$. Then

$$S_i=p_i[S_1,\ldots,S_n]$$

where

$$p_i[S_1,\ldots,S_n]=\omega(p_i)=\sum(p_i,t)\omega(t)$$

and where ω is the morphism defined by

$$\omega(\xi_i)=S_1,\quad i=1,\ldots,n,\qquad \omega(a)=a,\quad a\in F_0.$$

Now set

$$q_i=\mathbf{glove}(p_i),\quad i=1,\ldots,n.$$

The q_i's are in $k\langle F\cup\Xi\rangle$, i.e. are polynomials in the associative variables $x\in F'$. Since $\Xi\cap\mathrm{supp}(p_i)=\emptyset$ and since $1\notin\mathbf{glove}(M(F))$, we have

$$(1\cup\Xi)\cap\mathrm{supp}(q_i)=\emptyset,\quad i=1,\ldots,n,$$

in other words the system

$$\xi_i = q_i, \quad i = 1, \ldots, n$$

is proper. It therefore has a unique proper solution.

Define $T_i = \mathbf{glove}(S_i)$. We have to show that $(T_1, \ldots, T_n)$ is a solution, whence the solution of the above system, i.e.

$$T_i = q_i[T_1, \ldots, T_n]$$

where

$$q_i[T_1, \ldots, T_n] = \hat{\omega}(q_i) = \sum (q_i, w)\hat{\omega}(w)$$

and where $\hat{\omega}$ is the morphism from $k\langle F'\rangle^p$ into $k\langle\langle F\rangle\rangle$ defined by

$$\hat{\omega}(\xi_i) = T_i, \qquad \hat{\omega}(a) = a, \quad a \in F.$$

For this, we will show that for all $t \in M(F')$

$$\mathbf{glove}(\omega(t)) = \hat{\omega}(\mathbf{glove}(t)). \tag{7.3}$$

Assume (7.3) for granted for a moment. Then

$$\begin{aligned} q_i[T_1, \ldots, T_n] &= \hat{\omega}(\mathbf{glove}(p_i)) \\ &= \sum_{t \in M(F)} (p_i, t)\hat{\omega}(\mathbf{glove}(t)) \\ &= \sum_{t \in M(F)} (p_i, t)\mathbf{glove}(\omega(t)) \\ &= \mathbf{glove}(\omega(p_i)) = \mathbf{glove}(S_i) = T_i. \end{aligned}$$

Thus, it suffices to prove (7.3). We do it by induction on the length t of a tree t. If $t = a \in F_0$, then

$$\mathbf{glove}(\omega(a)) = \mathbf{glove}(a) = a = \hat{\omega}(\mathbf{glove}(a)).$$

If $t = \xi_i$, then

$$\mathbf{glove}(\omega(\xi_i)) = \mathbf{glove}(S_i) = T_i = \hat{\omega}(\xi_i) = \hat{\omega}(\mathbf{glove}(\xi_i))$$

Next if $t = f(t_1, \ldots, t_p)$ with $f \in F_p$, $t_1, \ldots, t_p \in M(F)$, then using (7.2)

$$\begin{aligned} \mathbf{glove}(\omega(t)) &= \mathbf{glove}(\omega(f(t_1, \ldots, t_p))) \\ &= \mathbf{glove}(f(\omega(t_1), \ldots, \omega(t_p)) \\ &= f\,\mathbf{glove}(\omega(t_1)) \cdots \mathbf{glove}(\omega(t_p)); \end{aligned}$$

by the induction hypotheses, it follows that

$$\begin{aligned} \mathbf{glove}(\omega(t)) &= f\hat{\omega}(\mathbf{glove}(t_1)) \cdots \hat{\omega}(\mathbf{glove}(t_p)) \\ &= \hat{\omega}(f\,\mathbf{glove}(t_1) \cdots \mathbf{glove}(t_p)) \\ &= \hat{\omega}(\mathbf{glove}(f(t_1, \ldots, t_p))) \\ &= \hat{\omega}(\mathbf{glove}(t)). \end{aligned}$$

Consider now a recognizable formal power series $S \in k\{\{F\}\}$, and let $T = \mathbf{glove}(S) \in K\langle\langle F\rangle\rangle$. The previous theorem shows that T is algebraic. Let z be a new letter, and define a morphism $\alpha : F^* \to z^*$ by $\alpha(f) = z$ for $f \in F$. Thus α maps all letters into z, and any word $w \in F^*$ of length n into z^n. If further w is the glove of some tree t, then the length $|t|$ of t is also n. Thus $\alpha(\mathbf{glove}(t)) = z^n$ for all trees in $M(F)$ of length n.

Next it is well known [23] that α extends uniquely to a morphism

$$\alpha : k\langle\langle F\rangle\rangle \to k[[z]]$$

and that α preserves algebraic formal power series. One even knows more:

If T is a component of the solution of a proper system of equations

$$\xi_i = q_i, \quad i = 1, \ldots, n,$$

then $\alpha(T)$ is the corresponding component of the solution of the system (over the unique letter z)

$$\xi_i = \alpha(q_i), \quad i = 1, \ldots, n$$

when α is extended to $F \cup \Xi$ by $\alpha(\xi) = \xi$ for $\xi \in \Xi$.

Given $S \in k\{\{F\}\}$, we call *enumerating series* of S the formal power series $\alpha(\mathbf{glove}(S))$. If

$$\alpha(\mathbf{glove}(S)) = \sum_{n \geq 1} a_n z^n,$$

then

$$a_n = \sum_{|t|=n} (S, t) \tag{7.4}$$

In the particular case where $S = L$ is the characteristic series of a forest L, we say that $\alpha(\mathbf{glove}(S))$ is the enumerating series of L. Then (7.4) becomes

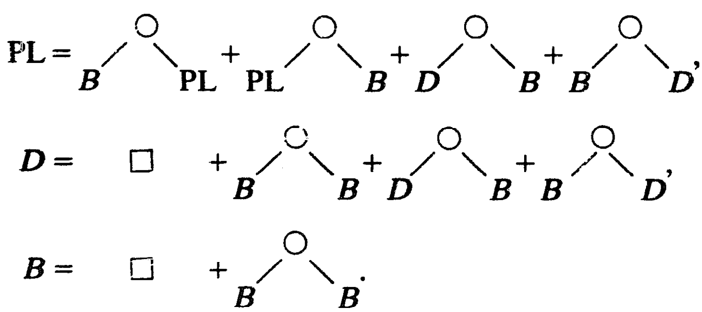
$$a_n = \mathrm{Card}\{t \in L \mid |t| = n\}.$$

According to the preceding discussion, we have

Proposition 7.2. *The enumerating series of a recognizable formal power series on trees (resp. of a recognizable forest) is an algebraic series.*

We now give two examples to show how this result can be applied.

Example 7.1 (Path length). According to Example 6.3, the formal power series PL (path length), D (length), B (characteristic series) satisfy:



Denote by $p(z)$, $d(z)$, $b(z)$ the series $\alpha(\mathbf{glove}(\mathrm{PL}))$, $\alpha(\mathbf{glove}(D))$ and $\alpha(\mathbf{glove}(B))$ respectively. Then by Proposition 7.2,

$$p(z) = 2zp(z)b(z) + 2zd(z)b(z),$$

$$d(z) = z + z(b(z))^2 + 2zd(z)b(z),$$

$$b(z) = z + z(b(z))^2$$

whence

$$d(z) = \frac{b(z)}{1-2zb(z)}, \qquad p(z) = \frac{2zd(z)b(z)}{1-2zb(z)} = \frac{2z(b(z))^2}{(1-2zb(z))^2} = \frac{2(b(z)-z)}{1-4z^2}$$

since $(1-2zb(z))^2 = 1-4z^2$.

Example 7.2 (Pattern matching in trees [12]). Given $M(F)$, the set of function in several variables from $M(F)$ into $M(F)$ is again a F-magma; we consider the sub-F-magma $\hat{M}(F)$ generated by the identical function and by the functions associated to the elements in F. Then $\hat{M}(F)$ is isomorphic to the magma $M(F')$ where $F' = F \cup \{x\}$, with x a new 0-ary symbol. The correspondance is established by associating, to each $m' \in M(F')$ the function $m \in \hat{M}(F)$ which substitutes trees (in $M(F)$) to the leafs of m' with value x.

Let m be an element of $\hat{M}(F)$; then m is a pattern. We say that m *occurs* in a tree t in $M(F)$ iff there exist $t_1, \ldots, t_q \in M(F)$ and $n \in \hat{M}(F)$ such that

$$t = n(m(t_1, \ldots, t_q)). \tag{7.5}$$

Perhaps, the pictorial description of Fig. 1 is useful.

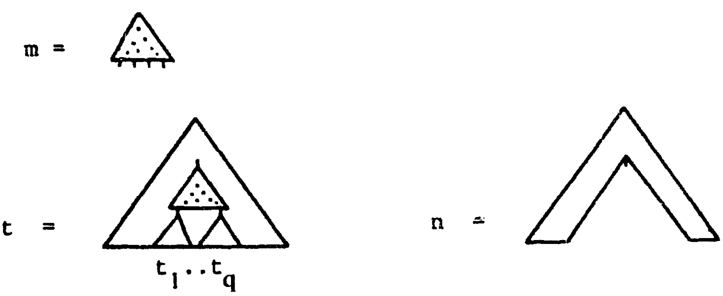


Fig. 1.

Let $D_m : M(F) \to \mathbb{Q}$ denote the formal power series which associates, to each tree t in $M(F)$, the number (D_m, t) of occurrences of m in t, i.e. the number of distinct factorizations (7.5) of t.

Given a tree $t = g(t_1, \ldots, t_p)$, then m occurs at the root of t or in one of the subtrees $t_1, \ldots, t_p$. Consequently, D_m satisfies the following equation:

$$D_m = m(B, B, \ldots, B)$$
$$+ \sum_{p\geq 1} \sum_{g \in F_p} [g(D_m, B, \ldots, B) + g(B, D_m, B, \ldots, B) + \cdots + g(B, \ldots, B, D_m)],$$

$$B = \sum_{p\geq 0} \sum_{g \in F_p} g(B, B, \ldots, B).$$

Let $c_p = \mathrm{Card}\, F_p (p \geq 0)$, and let $k = |m|_F$ be the number of symbols of m in F. Setting $d_m(z) = \alpha(\mathbf{glove}(D_m))$, $b(z) = \alpha(\mathbf{glove}(B))$, Proposition 7.2 gives

$$d_m(z) = z^k (b(z))^q + \sum_{p \geq 1} c_p p z d_m(z) b(z)^{p-1},$$

$$b(z) = \sum_{p \geq 0} c_p z (b(z))^p.$$

Solving the first, and computing the derivative of the second of these two equations gives

$$d_m(z) = \frac{z^k b(z)^q}{\left(1 - \sum_{p \geq 1} c_p p z b(z)^{p-1}\right)}, \qquad b'(z) = b(z) \frac{1}{\left(1 - \sum_{p \geq 1} c_p p z b(z)^{p-1}\right)}$$

whence

$$d_m(z) = z^{k+1} (b(z))^{q-1} b'(z)$$

which is of course the equation given by Flajolet and Steyaert [12].

8. Frontier

There is a well-known relation between algebraic (context-free) languages and recognizable forest; the set of derivation trees of a given algebraic language is a recognizable forest, and conversely the frontiers of the trees in a recognizable forest form an algebraic language [25]. This section is devoted to the generalisation, to formal power series, of these facts. However, there is an inherent difficulty to perform this task: if a word has infinitely many derivation trees, then the corresponding coefficient in the formal power series is infinite. To overcome this obstacle, there are two standard techniques (see e.g. [9]): either one considers complete semirings or one makes the necessary restrictions to avoid this situation. Since we deal with fields, we choose the second alternative.

The mapping 'frontier', denoted by

$$\mathbf{fr} : M(F) \to F_0^*$$

is defined by

$$\mathbf{fr}(a) = a, \quad a \in F_0,$$

$$\mathbf{fr}(f(t_1, \ldots, t_p)) = \mathbf{fr}(t_1)\mathbf{fr}(t_2) \cdots \mathbf{fr}(t_p).$$

In other terms, if $\pi : F^* \to F_0^*$ is the projection erasing all letters not in F_0, then

$$\mathbf{fr} = \pi \circ \mathbf{glove}.$$

If $F_1 = \emptyset$, then it is easily seen that for each $w \in F_0^*$ there are only finitely many trees t in $M(F)$ such that $\mathbf{fr}(t) = w$. Thus, if $S \in k\{\{F\}\}$, one can define the formal power series $\mathbf{fr}(S) \in k\langle\langle F_0 \rangle\rangle$ by

$$(\mathbf{fr}(S), w) = \sum_{\mathbf{fr}(t)=w} (S, t).$$

Theorem 8.1. *Assume that $F_1 = \emptyset$. Then the frontier of a recognizable formal power series on $M(F)$ is algebraic.*

We first give a definition and a lemma. Let G be a subset of F. We denote by $|t|_G$ the number of occurrences of elements of G in t, i.e.

$$|t|_G = \begin{cases} 1 & \text{if } t \in G \cap F_0, \\ 0 & \text{if } t \in F_0 - G, \end{cases}$$

$$|f(t_1, \ldots, t_p)|_G = \delta + |t_1|_G + \cdots + |t_p|_G$$

with $\delta = 0$ or 1 according to $f \notin G$ or $f \in G$.

Lemma 8.2. *Assume $F_1 = \emptyset$. For $t \in M(F)$, one has*

$$|t|_{F_0} \geq \tfrac{1}{2}|t|.$$

In particular, $t \notin F_0$ implies $|t|_{F_0} \geq 2$.

Proof. Straightforward.

Proof of Theorem 8.1. Consider a proper system of linear equations

$$\xi_i = p_i, \quad i = 1, \ldots, n$$

with $\Xi = \{\xi_1, \ldots \xi_n\}$, and $p_i \in k\{F \cup \Xi\}$, $i = 1, \ldots, n$. Let $(S_1, \ldots, S_n)$ be the solution of this system. Let $q_i = \mathbf{glove}(p_i)$ for $i = 1, \ldots, n$. Then $(\mathbf{glove}(S_1), \ldots, \mathbf{glove}(S_n))$ is the solution of the system

$$\xi_i = q_i, \quad i = 1, \ldots, n.$$

Now consider the projection

$$\pi' : (F \cup \Xi)^* \to (F_0 \cup \Xi)^*$$

defined by $\pi'(\xi_i) = \xi_i$, $i = 1, \ldots, n$ and $\pi'(f) = \pi(f)(f \in F)$. Then $r_i = \pi'(q_i)$ is in $k\langle F_0 \cup \Xi \rangle$. Now in view of Lemma 8.2, for each $t \in \mathrm{supp}(p_i)$, either $t \in F_0$ or $|t|_{F_0 \cup \Xi} \geq 2$. Consequently, for each $w \in \mathrm{supp}(q_i)$, one has $w \in F_0$ or $|w|_{F_0 \cup \Xi} \geq 2$. This implies that for each $v \in \mathrm{supp}(r_i)$, either $v \in F_0$ or $|v|_{F_0 \cup \Xi} \geq 2$, and in particular

$$\mathrm{supp}(r_i) \cap \Xi = \emptyset.$$

Consequently the system of equations

$$\xi_i = r_i$$

is proper. On the other hand setting $T_i = \mathbf{glove}(S_i)$, the formal power series

$$\pi(T_i) = \sum_{w \in F^*} (T_i, w)\pi(w) = \sum_{v \in F_0^*} (\pi(T_i), v)v$$

is well defined; indeed

$$(\pi(T_i), v) = \sum_{\substack{w \in \mathrm{supp}(T_i) \\ \pi(w) = v}} (T_i, w)$$

and the set $\{w \in \mathbf{glove}(M(F)) \mid \pi(w) = v\}$ is finite since if w is in this set, then $|v| = |w|_{F_0} \geq \frac{1}{2}|w|$ by Lemma 8.2. Thus

$$\pi(T_i) = \pi(q_i[T_1, \ldots, T_n]) = r_i[\pi(T_1), \ldots, \pi(T_n)], \quad i = 1, \ldots, n$$

or equivalently

$$\mathbf{fr}(S_i) = r_i[\mathbf{fr}(S_1), \ldots, \mathbf{fr}(S_n)], \quad i = 1, \ldots, n.$$

We now prove a converse of Theorem 8.1.

Theorem 8.3. *Let $T \in k\langle\langle X\rangle\rangle$ be an algebraic formal power series. There exist a magma $M(F)$ with $F_0 = X$ and a recognizable formal power series S in $k\{\{F\}\}$ such that $T = \mathbf{fr}(S)$.*

Proof. Let

$$\xi_i = r_i, \quad i = 1, \ldots, n$$

be a proper system of algebraic equations, $r_i \in k\langle X \cup \Xi\rangle$ having the solution $(T_1, \ldots, T_n)$ with $T_1 = T$. Let

$$k = \max_{1 \leq i \leq n} \max_{w \in \mathrm{supp}(r_i)} |w|_{\Xi \cup X}.$$

Define

$$F = F_0 \cup F_1 \cup \cdots \cup F_k$$

where $F_0 = X$, and for each $p \geq q$, $F_p = \{\xi_1^p, \ldots, \xi_n^p\}$ is a copy of Ξ. Let V be the k-vector space with base $X \cup \Xi$. As already done in the proof of Theorem 6.2, we identify $(X \cup \Xi)^p$ with a base of $V^{\otimes p}$ by considering a word $w = \eta_1\eta_2 \ldots \eta_p \in (X \cup \Xi)^p$ as the base element $\eta_1 \otimes \eta_2 \otimes \cdots \otimes \eta_p \in V^{\otimes p}$.

Define a linear representation (V, μ) of $M(F)$ by

$$\mu(a) = a, \quad a \in F_0,$$

for $p \geq 1$, and w in $(X \cup \Xi)^p$

$$\mu(\xi_i^p)(w) = (r_i, w)\xi_i^p.$$

Let $S_1, \ldots, S_n$ be the formal power series on $M(F)$ defined by the representations (V, μ, ξ_i'), $i = 1, \ldots, n$, where $\{\xi_1', \ldots, \xi_n'\} \cup \{x' \mid x \in X\}$ is the dual base of $X \cup \Xi$. According to the proof of Theorem 6.2, the n-tuple $(S_1, \ldots, S_n)$ is the solution of the system of linear equations

$$\xi_i = p_i, \quad i = 1, \ldots, n$$

with

$$\begin{aligned} p_i &= \sum_{p=0}^{k} \sum_{f \in F_p} \sum_{w \in (\Xi \cup X)^p} [\xi_i' \circ \mu f(w)] f(w) \\ &= \sum_{x \in X} [\xi_i' \circ \mu x] + \sum_{p=1}^{k} \sum_{j=1}^{n} \sum_{w \in (\Xi \cup X)^p} [\xi_i' \circ \mu \xi_j^p(w)] \xi_j^p(w) \\ &= \sum_{p} \sum_{j,w} [\xi_i'((r_j, w)\xi_j)] \xi_j^p(w) = \sum_{p} \sum_{w} (r_i, w) \xi_i^p(w). \end{aligned}$$

It follows from the construction of Theorem 8.1 that the n-tuple $(\mathbf{fr}(S_1), \ldots, \mathbf{fr}(S_n))$ is the solution of the system

$$\xi_i = \mathbf{fr}(p_i) = r_i', \quad i = 1, \ldots, n$$

with $r_i' = \sum_p (r_i, w) w = r_i$.

Since the solution of a proper system of algebraic equations is unique [23], one has

$$\mathbf{fr}(S_i) = T_i, \quad i = 1, \ldots, n.$$

9. A pumping lemma

It is well known that there exist pumping lemmas for recognizable forests. These lemmas cannot hold for recognizable formal power series on trees since there are supports of such power series which are shown not to be recognizable forests by using precisely a pumping lemma. Thus the situation is analogue to that encountered when one tries to prove a pumping lemma for recognizable formal power series on words. For these a deep result of Jacob [13], making use of so called pseudoregular matrices, shows the existence of a weakened version of the pumping lemma. We use a slightly sharper statement [21] to prove a pumping lemma for recognizable formal power series on trees, and give then some examples.

Let F be a graded alphabet, x a new symbol, $F' = F \cup \{x\}$, and $F_0' = F_0 \cup \{x\}$. A tree t in $M(F')$ defines a morphism

$$\psi_t : M(F') \to M(F')$$

by setting

$$\psi_t(x) = t, \qquad \psi_t(a) = a \in F_0.$$

Thus $\psi_t(s)$ is the tree obtained by replacing each occurrence of x in the tree s by t. In a dual manner, we obtain a mapping

$$\bar{\varphi}_s : M(F') \to M(F')$$

by setting

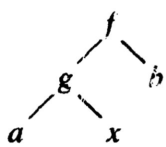
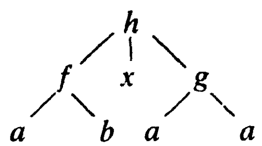
$$\bar{\varphi}_s(t) = \psi_t(s).$$

We denote by φ_s the restriction of $\bar{\varphi}_s$ to $M(F)$. Then φ_s is a mapping into $M(F)$. Next we define

$$A = \{s \in M(F') \mid |s|_x = 1\},$$

$$B = \{s \in M(F') \mid s = f(t_1, \dots, t_p), f \in t_p, \exists i \text{ such that } t_i = x \text{ and } t_j \in M(F) \text{ for all } j \neq i\}.$$

Example:


$\in A$,

$\in B$.

Finally we set

$$\Sigma^* = \{\varphi_s \mid s \in A\}, \qquad \Sigma = \{\varphi_s \mid s \in B\}.$$

The notation is consistent in view of the following

Proposition 9.1. *Σ^* is a free monoïd freely generated by Σ. Its neutral element is φ_x.*

Proof. (1) Σ^* is generated by Σ. Let indeed φ_s be in Σ^* and assume $s \notin B \cup \{x\}$. Then

$$s = f(s_1, \dots, s_p)$$

for some $f \in F_p$, exactly one among $s_1, \dots, s_p$, say s_i, is in A, the other s_j are in $M(F)$. Consequently

$$\sigma = f(s_1, \dots, s_{i-1}, x, s_{i+1}, \dots, s_p)$$

is in B. We claim that

$$\varphi_s = \varphi_\sigma \circ \varphi_{s_i}. \tag{9.1}$$

Indeed, let t be in $M(F)$. Setting

$$u = \varphi_{s_i}(t) = \psi_t(s_i)$$

we have

$$\varphi_\sigma \circ \varphi_{s_i}(t) = \varphi_\sigma(u) = \psi_u(f(s_1, \ldots, s_{i-1}, x, s_{i+1}, \ldots, s_p))$$
$$= f(s_1, \ldots, s_{i-1}, u, s_{i+1}, \ldots, s_p)$$
$$= f(s_1, \ldots, s_{i-1}, \psi_t(s_i), s_{i+1}, \ldots, s_p)$$
$$= \psi_t(f(s_1, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_p) = \varphi_s(t).$$

This proves (9.1). By repeating this argument we get for each s in A a sequence $\sigma_1, \ldots, \sigma_n$ of elements in B such that

$$\varphi_s = \varphi_{\sigma_1} \circ \varphi_{\sigma_2} \circ \cdots \circ \varphi_{\sigma_n}. \tag{9.2}$$

(2) Σ^* is freely generated by Σ. Assume that for some $\varphi_s \in \Sigma^*$, there is a decomposition (9.2), and a decomposition

$$\varphi_s = \varphi_{\tau_1} \circ \cdots \circ \varphi_{\tau_m}$$

for some $\tau_1, \ldots, \tau_m$ in B.

We prove, by induction on $m+n$, the following claim: if for a tree $t \in M(F)$ with height$(t) >$ height$(\tau_1), \ldots,$ height(τ_m), height$(\sigma_1), \ldots,$ height(σ_n) one has

$$\varphi_{\tau_1} \circ \cdots \circ \varphi_{\tau_m}(t) = \varphi_{\sigma_1} \circ \cdots \circ \varphi_{\sigma_n}(t),$$

then $m = n$, $\tau_1 = \sigma_1, \ldots, \tau_m = \sigma_m$.

Indeed, set

$$\sigma_1 = f(s_1, \ldots, s_{i-1}, x, s_{i+1}, \ldots, s_p), \qquad \tau_1 = g(t_1, \ldots, t_{j-1}, x, t_{j+1}, \ldots, t_q)$$

with $s_k \in M(F)$ for $k \neq i$, $t_l \in M(F)$ for $l \neq j$.

Suppose $\varphi_{\tau_1} \circ \cdots \circ \varphi_{\tau_m}(t) = \varphi_{\sigma_1} \circ \cdots \circ \varphi_{\sigma_n}(t)$ for a tree t with the above height property. Set $u = \varphi_{\sigma_2} \circ \cdots \circ \varphi_{\sigma_n}(t)$, $v = \varphi_{\tau_2} \circ \cdots \circ \varphi_{\tau_m}$. Then height$(u)$, height$(v) \geqslant$ height(t).

Next

$$\varphi_{\sigma_1}(u) = \varphi_{\tau_1}(v)$$

that is

$$f(s_1, \ldots, s_{i-1}, u, s_{i+1}, \ldots, s_p) = g(t_1, \ldots, t_{j-1}, v, t_{j+1}, \ldots, t_q).$$

Hence

$$f = g \quad \text{and} \quad (s_1, \ldots, s_{i-1}, u, s_{i+1}, \ldots, s_p) = (t_1, \ldots, t_{j-1}, v, t_j, \ldots, t_q).$$

Because of the inequalities height$(u) >$ height$(\tau_1) >$ height(t_l), $l \neq j$ one has $i = j$ and $s_k = t_k (k \neq l)$ and $u = v$, whence $\sigma_1 = \tau_1$. The claim follows by induction.

This proves (2)

Definition. Let $t \in M(F)$. A *walk* in t is a pair (φ, a) with $\varphi \in \Sigma^*$, $a \in F_0$ such that $t = \varphi(a)$. The *length* of the walk is the length of φ in the free monoid Σ^*.

Theorem 9.2. *Let $S \in k\{\{F\}\}$ be recognizable. There exists a constant N such that for each tree $t \in \mathrm{supp}(S)$, and for any walk (φ, a) in t of length at least N, φ decomposes, in Σ^*, into $\varphi = \varphi_1\varphi_2\varphi_3$ such that $\varphi_1\varphi_2^*\varphi_3(a) \cap \mathrm{supp}(S)$ is infinite.*

Proof. Let (V, μ, λ) be a representation of S. For each $\sigma = f(s_1, \ldots, s_{i-1}, x, s_{i+1}, \ldots, s_p)$ in B, we define an endomorphism

$$\hat{\varphi}_\sigma : V \to V$$

by setting, for $v \in V$,

$$\hat{\varphi}_\sigma(v) = \mu(f)(\mu(s_1), \ldots, \mu(s_{i-1}), v, \mu(s_{i+1}), \ldots, \mu(s_p)). \tag{9.3}$$

The mapping $\varphi_\sigma \mapsto \hat{\varphi}_\sigma$ is extended to Σ^* by composition.

Note that

$$\hat{\varphi}(\mu(t)) = \mu(\varphi(t)) \tag{9.4}$$

for all $\varphi \in \Sigma^*$ and $t \in M(F)$. Indeed, if $\varphi = \varphi_\sigma$, then by (9.3),

$$\begin{aligned}\hat{\varphi}_\sigma(\mu(t)) &= \mu(f)(\mu(s_1), \ldots, \mu(s_{i-1}), \mu(t), \mu(s_{i+1}), \ldots, \mu(s_p))\\ &= \mu(f(s_1, \ldots, s_{i-1}, t, s_{i+1}, \ldots, s_p)) = \mu(\varphi_\sigma(t)).\end{aligned}$$

Next if $\varphi = \varphi' \circ \varphi_\sigma$, then by induction

$$\begin{aligned}\hat{\varphi}(\mu(t)) &= \hat{\varphi}'(\hat{\varphi}_\sigma(\mu(t)) = \hat{\varphi}'(\mu(\varphi_\sigma(t)))\\ &= \mu(\varphi' \circ \varphi_\sigma(t)) = \mu(\varphi(t)).\end{aligned}$$

Now let $N = N(\dim V)$ be the integer of Theorem 3 of [21], let t be a tree in $\mathrm{supp}(S)$ and let (φ, a) be a walk of t of length at least N. According to the theorem quoted, there exists a decomposition

$$\varphi = \varphi_1 \circ \varphi_2 \circ \varphi_3$$

with $\varphi_2 \neq \mathrm{id}$, and φ_2 a pseudo-regular endomorphism. Set

$$u_n = \lambda \circ \hat{\varphi}_1 \circ \hat{\varphi}_2^n \circ \hat{\varphi}_3(\mu a), \quad n \geq 0.$$

Then $u_1 = \lambda\hat{\varphi}(\mu a) = \lambda\mu(\varphi(a)) = \lambda\mu(t) = (S, t) \neq 0$. Consequently, Lemma 1 of [21] asserts that $u_n \neq 0$ for infinitely many n. Since

$$u_n = \lambda \circ \mu(\varphi \circ \varphi_2^n \circ \varphi_3(a)) = (S, \varphi_1\varphi_2^n\varphi_3(a)),$$

this proves the theorem.

During the proof of Theorem 9.2, we also verified the following proposition:

Proposition 9.3. *Let $S \in k\{\{F\}\}$ be a recognizable formal power series, and let (V, μ, λ) be a representation for S. For all $\varphi_1, \varphi_2, \varphi_3 \in \Sigma^*$, $a \in F_0$, the formal power* *series in one variable*

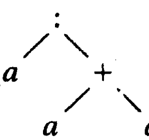
$$u(z)=\sum_{n\geqslant 0}(S,\varphi_1\varphi_2^n\varphi_3(a))z^n$$

is rational, and the sequence $(S,\varphi_1\varphi_2^n\varphi_3(a))$ *satisfies a linear recurrence relation of length at most* $\dim(V)$.

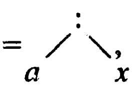
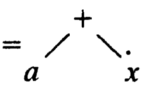
Example 9.1. We show that the formal power series evald evaluating arithmetic expressions *with* division is *not* recognizable (whereas Example 4.2. showed without the division, evaluation is recognizable). It suffice to consider the case where $F=F_0\cup F_2$, with $F_0=\{a\}$, $F_2=\{+,\times,:\}$ (the general case reduces to the present one by Hadamard product (Proposition 5.1)). Define

$$\text{evald}(a)=1.$$

Then for instance

eval() $=\frac{1}{2}$.

Consider the two elements $\varphi_\sigma,\varphi_\tau\in\Sigma$ defined by

$\sigma=$  , $\tau=$  .

The tree $\varphi_\sigma\varphi_\tau^n(a)$ is

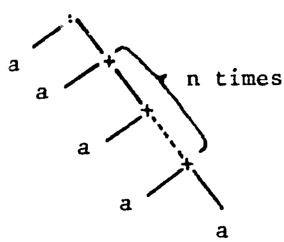


Fig. 2.

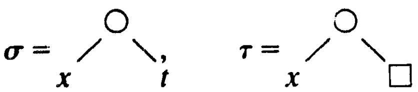
whence $u_n=(\text{evald},\varphi_\sigma\varphi_\tau^n(a))=1/(n+1)$. Thus the series $u(z)=\sum_{n\geqslant 0}u_nz^n$ is not rational, since

$$zu(x)=-\log(1-x).$$

In view of Proposition 9.3, evald is not recognizable.

Example 9.2. The formal power series height is not recognizable provided $F\neq F_0\cup F_1$. We assume for simplicity that $F=F_0\cup F_2$, $F_0=\{\square\}$, $F_2=\{\bigcirc\}$. Suppose that height is recognizable, let (V,μ,λ) be a representation and set $K=\dim(V)$. Define

φ_σ, φ_τ in Σ by



with height$(t) = K$. Then for $n \geq 0$, the tree $t_n = \varphi_\sigma \varphi_\tau^n(\square)$ is

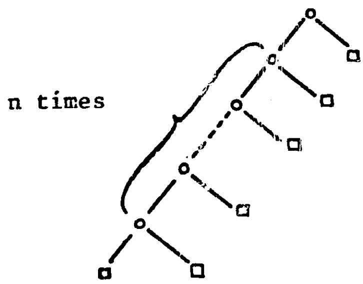


Fig. 3.

and

$$\text{height}(t_n) = \begin{cases} 1+K, & n = 0, \dots, K, \\ 1+n, & n > K. \end{cases}$$

On the other hand, height(t_n) satisfies, according to Proposition 9.3, a linear recurrence relation of length at most K. Since the $K+1$ first elements of the sequence $(\text{height}(t_n))_{n\geq 0}$ are equal to $K+1$, this implies height$(t_n) = K+1$ for all n. Contradiction.

Remark. Flajolet [11] has another proof of the fact that height is not recognizable which runs as follows: Consider the enumerating series $a(x) = \alpha(\textbf{glove}(\text{height}))$. By analyzing the singularities of $a(z)$, Flajolet shows that $a(x)$ is not algebraic. Consequently, height cannot be recognizable in view of Theorem 7.1.

Example 9.3 (AVL-trees). It is easy to show that the set of AVL-trees is not a recognizable forest. We verify that it is even not the support of a recognizable formal power series. (This does not prove that its generating series is non-algebraic, but perhaps explains to some extent why it is not yet known.)

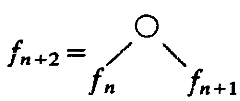
Let $F = F_0 \cup F_2$, with $F_0 = \{\square\}$, $F_2 = \{\bigcirc\}$. A tree t is AVL if either $t = \square$ or $t = \bigcirc(t_1, t_2)$ and the following two conditions are satisfied:

(i) t_1 and t_2 are AVL;

(ii) $|\text{height}(t_1) - \text{height}(t_2)| \leq 1$.

The Fibonnacci trees defined inductively by

$f_0 = f_1 = \square$,



are AVL-trees and height $(f_n) = n - 1$.

Let $\varphi_n = \varphi_{\sigma_n} \in \Sigma$ be defined by

$$\sigma_n = \begin{array}{c} \bigcirc \\ \diagup \quad \diagdown \\ f_n \qquad t \end{array}.$$

Then $\varphi_n(f_{n+1}) = f_{n+2}$, whence

$$f_{n+2} = \varphi_n \circ \varphi_{n-1} \circ \cdots \circ \varphi_1 \circ \varphi_0(\square).$$

Assume now that the set A of AVL-trees is the support of some recognizable formal power series S. Then by Theorem 9.2 for a sufficiently large n,

$$\varphi_n \circ \varphi_{n-1} \circ \cdots \circ \varphi_0 = \varphi' \circ \varphi'' \circ \varphi'''$$

for some φ', φ'', φ''' such that

$$g_k = \varphi' \circ \varphi''^k \circ \varphi'''(\square) \in A$$

for infinitely many k. Now

$$g_k = \varphi_n \circ \cdots \circ \varphi_{q+1} \circ (\varphi_q \circ \cdots \circ \varphi_p)^k \varphi_p \circ \cdots \circ \varphi_0(\square)$$

for some p, q with $0 \leqslant p < q \leqslant n$, and setting $g_k = \varphi_n(g'_k)$, one has

$$g_k = \begin{array}{c} \bigcirc \\ \diagup \quad \diagdown \\ f_n \qquad g'_{k'} \end{array}$$

and

$$\text{height}(g'_k) = n - 1 + (k-1)(q-p) \quad (k \geqslant 1).$$

Consequently no g_k $(k \geqslant 2)$ is in A. Contradiction.

Acknowledgment

We thank the lillois group working on trees for their encouragements and helpful discussions. A. Arnold made pertinent remarks on a first draft. Ph. Flajolet indicated to us the Example 6.3.

References

[1] A. Arnold, Systèmes d'équations dans le magmoïde: Ensembles rationnels et algébriques d'arbres, Thèse d'Etat, Lille (1977).

[2] J. Berstel and C. Reutenauer, Séries formelles reconnaissables d'arbres et applications, *5ème Colloque de Lille: Les Arbres en Algèbre et en Programmation* (1980) 11–21.

[3] W.S. Brainerd, Tree-generating regular systems, *Information and Control* **14** (1969) 217–231.

[4] L. Chottin, Une preuve combinatoire de la formule de Lagrange à deux variables, *Discrete Math.* **13** (1975) 215–224.

[5] A. Cobham, Representation of a word function as the sum of two functions *Math. Systems Theory* **11** (1978) 373–377.
[6] P.M. Cohn, *Universal Algebra* (Wiley, New York, 1965).
[7] R. Cori, Un code pour les graphes planaires et ses applications, *Astérique* **27** (1975).
[8] M. Dauchet, Transductions de forêts: Bimorphismes de magmoïdes, Thèse d'Etat, Lille (1977).
[9] S. Eilenberg, *Automata, Languages and Machines, Vol. A* (Academic Press, New York, 1974).
[10] P. Flajolet, Analyse d'algorithmes de manipulation d'arbres et de fichiers, Thèse d'Etat, Paris IX (1979).
[11] P. Flajolet, Private Communication.
[12] P. Flajolet and J.M. Steyaert, On the analysis of tree-matching algorithms, *5ème Colloque de Lille: Les Arbres en Algèbre et en Programmation* (1980).
[13] G. Jacob, Un théorème de factorisation des produits d'endomorphismes, *J. Algebra* **63** (1980) 389–412.
[14] R. Kemp, On the average stack size of regularly distributed binary trees, in: H. Maurer, Ed., *Proc 6th ICALP*, Lecture Notes in Computer Science **71** (Springer, Berlin, 1979) 340–355.
[15] D.E. Knuth, *The Art of Computer Programming, Vol. 3* (Addison-Wesley, New York, 1973).
[16] E. Lilin, *S*-transducteurs de forêts, *3ème C.L.A.A.P.* (1979) 189–206.
[17] T. Maibaum, Pumping lemmas for term languages, *J. Comput. System Sci.* **17** (1978) 319–330.
[18] J. Mongy, Closure of recognizable tree-languages under intersection and morphisms, *5ème Colloque de Lille: Les Arbres en Algèbres et Programmation* (1980) 135–149.
[19] M. Nivat, On the interpretation of recursive polyadic program schemes, *Symposia Math.* **15** (1975) 255–281.
[20] A. Paz and A. Salomaa, Integral sequential word functions and growth equivalence of Lindenmayer systems, *Information and Control* **23** (1973) 313–343.
[21] C. Reutenauer, An Ogden-like iteration lemma for rational power series, *Acta Informat.* **13** (1980) 189–197.
[22] B. Rosen, Tree-manipulating systems and Church–Rosser theorem, *J. ACM* **20** (1973) 160–187.
[23] A. Salomaa and M. Soittola, *Automata-Theoretic Aspects of Formal Power Series* (Springer, Berlin, 1978).
[24] M.P. Schützenberger, On the definition of a family of automata, *Information and Control* **4** (1961) 245–270.
[25] J.W. Thatcher, Characterizing derivation trees of context-free grammars through a generalization of finite automata theory, *J. Comput. System Sci.* **1** (1967) 317–322.
[26] J.W. Thatcher, Generalized sequential machine maps, *J. Comput System Sci.* **4** (1970) 339–367.
[27] J.W. Thatcher and J.B. Wright, Generalized finite automata theory with an application to a decision problem of second-order logic, *Math. Systems Theory* **2** (1968) 57–81.