RAIRO-Theor. Inf. Appl. 44 (2010) 257–279
DOI: 10.1051/ita/2010014 Available online at:
www.rairo-ita.org

ON THE SYNTACTIC COMPLEXITY OF TREE SERIES

Symeon Bozapalidis[1] and Antonios Kalampakas[2,3]

Abstract. We display a complexity notion based on the syntax of a tree series which yields two distinct hierarchies, one within the class of recognizable tree series and another one in the class of non-recognizable tree series.

Mathematics Subject Classification. 68Q01, 68Q15.

1. Introduction

The notion of *syntactic complexity* of a recognizable graph language has originated in [5]. It is a structural complexity measure giving rise to a syntactic classification inside the class of recognizable graph languages. The syntactic complexity of a recognizable graph language L is given by a function mapping any couple of natural numbers, representing the type of a graph, to the number of the distinct syntactic classes at this type. We have displayed graph languages with various types of complexities and showed that the set of connected graphs has Bellian complexity. Furthermore, the language of Eulerian graphs is syntactically more complicated than that of connected graphs (*cf.* [8]).

This notion has been also investigated in the setup of tree languages. More precisely, in [6], the syntactic complexity of a tree language is defined to be the number of the distinct syntactic classes of all trees with a fixed yield length. This allows a classification of tree languages according to their structural properties and several interesting results are obtained. The class of recognizable tree languages is properly contained in that of languages with bounded complexity. The Dyck tree

Keywords and phrases. Tree series, syntactic complexity, recognizability.

[1] Aristotle University of Thessaloniki, Department of Mathematics, 54124 Thessaloniki, Greece; bozapali@math.auth.gr

[2] Democritus University of Thrace, Department of Production Engineering and Management, 67100 Xanti, Greece; akalamp@math.auth.gr

[3] Technical Institute of Kavala, Department of Exact Sciences, 65404 Kavala, Greece.

Article published by EDP Sciences

language of order k

$$D_k = \{t \mid t \in T_\Gamma,\ |t|_{f_1} = \cdots = |t|_{f_k},\ \mathrm{rank}(f_i) = 2\},$$

has polynomial syntactic complexity of degree $k-1$, and the diagonal language

$$L_d = \{f(t,t) \mid t \in T_\Gamma\}, \qquad f \text{ a fixed binary symbol,}$$

has exponential syntactic complexity.

In the present paper we develop a structural complexity theory for tree series. Let T_Γ be the set of all trees over a ranked alphabet Γ and P_Γ the set of all trees with just one occurrence of the variable x. Then P_Γ becomes a free monoid with operation the substitution at x. Moreover, P_Γ acts on the set T_Γ by the same way. A formal power series on trees with coefficients in a field K, is just a function $S : T_\Gamma \to K$. The *right derivative of* S *at* t and the *left derivative of* S *at* τ are defined respectively by

$$St^{-1} : P_\Gamma \to K, \qquad (St^{-1}, \tau') = (S, \tau' t) \text{ for every } \tau' \in P_\Gamma,$$
$$\tau^{-1}S : T_\Gamma \to K, \qquad (\tau^{-1}S, t') = (S, \tau t') \text{ for every } t' \in T_\Gamma.$$

We denote by $V(S)$ (resp. $F(S)$) the subspace of K^{P_Γ} (resp. K^{T_Γ}) generated by the right (resp. left derivatives) of S.

Tree series are interpreted by $K-\Gamma$-algebras $\mathcal{A} = (A, \alpha)$ where A is a K-vector space and α is a family of multilinear functions

$$\alpha_f : A^m \to A, \quad f \in \Gamma_m,\ m \geq 0.$$

We can endow $V(S)$, in a canonical way, with the structure of a $K-\Gamma$-algebra, called the *syntactic* $K-\Gamma$*-algebra of* S.

A tree series $S : T_\Gamma \to K$ is said to be *recognizable* if it can be realized by a pair $(\mathcal{A}, \varphi)$ where $\mathcal{A}$ is a finitely dimensional $K-\Gamma$-algebra and $\varphi : A \to K$ a linear form. This means that $S = \varphi \circ h_{\mathcal{A}}$, where $h_{\mathcal{A}}$ is the unique $K-\Gamma$-algebra morphism from T_Γ to $\mathcal{A}$. We have the following characterization result from [4]: a tree series $S : T_\Gamma \to K$ is recognizable if and only if $\dim F(S) < \infty$ if and only if $\dim V(S) < \infty$. In this case we have $\dim F(S) = \dim V(S)$.

Syntactic complexity of tree series is introduced in Section 4, it is described by a function $SC_S : \mathbb{N} \to \mathbb{N}$ which sends every natural number n to the maximum number of linearly independent right derivatives St^{-1} where t runs over all trees whose width (*i.e.*, yield length) is n. It is proved that for every tree language L, the syntactic complexity of its characteristic series does not exceed that of L. We say that a tree series $S : T_\Gamma \to K$ has bounded, polynomial or exponential syntactic complexity according to the explicit formula defining the function SC_S.

Although every recognizable series has bounded syntactic complexity this does not characterize recognizability. Indeed, the series that sends every tree t to $\frac{1}{|t|}$ (where $|t|$ is the size of t) has bounded complexity but is not recognizable. Moreover, we display a tree series (the characteristic series of D_k) with polynomial syntactic complexity, and show that from every series S with non zero coefficients we can construct a series S_d with exponential complexity $SC_{S_d}(n) = C_{n-1}$, where

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

is the nth Catalan number.

Berstel and Reutenauer demonstrated in [1], by using a pumping lemma technique, that the series height sending every tree to its height is not recognizable. In Section 5 we calculate its syntactic complexity

$$SC_{\text{height}}(n) = \lfloor n - \log_2 n \rfloor,$$

where $\lfloor r \rfloor$ denotes the greatest integer not exceeding r. On the other hand, it is rather surprising that the series height_p that sends every tree to its height modulo p (p prime) is recognizable. Moreover, we present a tree language, over the alphabet Γ consisting of a binary symbol f and a unary symbol α, that has the same syntactic complexity with *height*, namely the set L_{avl} of all AVL trees *i.e.*, all trees t such that either $t = \alpha$ or $t = f(t_1, t_2)$, with $t_1, t_2 \in L_{avl}$ and

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

Since all recognizable tree series have bounded syntactic complexity this notion is only appropriate for a classification of non-recognizable tree series. In the last section we provide a complexity measure suitable for the class of recognizable tree series. It is based on a refined context namely $P_\Gamma^{(n)}$ which is the set of all trees in $T_\Gamma(X)$ where the variables $x_1, \ldots, x_n$ occur in the yield of the tree (in this order from left to right) exactly once. Two dual notions of derivatives are obtained for all $\tau \in P_\Gamma^{(n)}$ and $t_1, \ldots, t_n \in T_\Gamma$

$$\tau^{-1}S : T_\Gamma^n \to K, \quad (\tau^{-1}S, (t_1, \ldots, t_n)) = (S, \tau[t_1, \ldots, t_n]),$$

and

$$S(t_1, \ldots, t_n)^{-1} : P_\Gamma^{(n)} \to K, \quad (S(t_1, \ldots, t_n)^{-1}, \tau) = (S, \tau[t_1, \ldots, t_n]),$$

and their corresponding sets are

$$F(S)^{(n)} = \langle \tau^{-1}S \mid \tau \in P_\Gamma^{(n)} \rangle \text{ and } V(S)^{(n)} = \langle S(t_1, \ldots, t_n)^{-1} \mid t_1, \ldots, t_n \in T_\Gamma \rangle.$$

We prove that a series $S : T_\Gamma \to K$ is recognizable if and only if $dimF(S)^{(n)} < \infty$, for all n, if and only if $\text{dim}V(S)^{(n)} < \infty$, for all n. In this case it holds $\text{dim}F(S)^{(n)} = \text{dim}V(S)^{(n)}$. The refined syntactic complexity RSC_S of a recognizable tree series S is described by the function sending every natural number n to the dimension of $F(S)^{(n)}$. Since for any *non* recognizable tree series S it holds

$$RSC_S(n) = \infty, \text{ for all } n$$

the above notion of complexity is suitable to classify only recognizable tree series. The recognizable series *tree size*, *branch enumeration*, and *branch length* have the same linear complexity,

$$RSC_S(n) = n + 1.$$

Moreover, the recognizable tree series height_p has exponential refined syntactic complexity namely,

$$RSC_{\text{height}_\text{p}}(n) = p^n.$$

The presented syntactic complexities yield the following hierarchy within the class of tree series

$$\begin{aligned} RSCBOUND \subset RSCPOL \subset RSCEXP \subset REC \subset \\ \subset SCBOUND \subset SCPOL \subset SCEXP, \end{aligned}$$

where REC is the class of recognizable tree series and $UBOUND$, $UPOL$, $UEXP$ denote respectively the classes of tree series with bounded, polynomial and exponential U-complexity ($U = SC, RSC$).

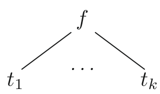
2. Tree languages and their syntactic complexity

To construct trees we need a (finite) ranked alphabet $\Gamma = \bigcup_{k\geq 0} \Gamma_k$ and a set X of variables. We pose $X_n = \{x_1, \ldots, x_n\}$, for $n \geq 1$.

The set of trees over Γ and X is the smallest set $T_\Gamma(X)$ inductively defined by the items

- $\Gamma_0 \cup X \subseteq T_\Gamma(X)$;
- $t_1, \ldots, t_k \in T_\Gamma(X)$ and $f \in \Gamma_k$ implies $f(t_1, \ldots, t_k) \in T_\Gamma(X)$.

Often the tree $f(t_1, \ldots, t_k)$ is depicted as



hence the *denomination tree*. We denote T_Γ the set of all trees over Γ without variables (*i.e.*, $T_\Gamma = T_\Gamma(\emptyset)$). Subsets of $T_\Gamma(X)$ are refereed to as *tree languages*.

The height of a tree $t \in T_\Gamma(X)$ is the length of its longest branch. Formally the function height : $T_\Gamma(X) \to \mathbb{N}$ is inductively defined by

- $\text{height}(\alpha) = 0$, for $\alpha \in \Gamma_0 \cup X$;
- $\text{height}(f(t_1, \ldots, t_k)) = 1 + \max\{\text{height}(t_1), \ldots, \text{height}(t_k)\}$, $f \in \Gamma_k$ and $t_1, \ldots, t_k \in T_\Gamma(X)$.

The *yield* of a tree $t \in T_\Gamma$ is the word $y(t)$ obtained by concatenating from left to right the leaves (*i.e.*, the symbols of Γ_0 occurring in t). Formally, $y : T_\Gamma \to \Gamma_0^*$, is inductively defined by

$$y(c) = c, \ \ (c \in \Gamma_0), \quad y(f(t_1, \dots, t_k)) = y(t_1) \dots y(t_k), \ \ (f \in \Gamma_k, t_i \in T_\Gamma).$$

The width of $t \in T_\Gamma(X)$ is the length of its yield: $|y(t)| = \text{width}(t)$.

The basic operation on trees is substitution. Given $t, t_1, \dots, t_n \in T_\Gamma(X_n)$, we denote by $t[t_1, \dots, t_n]$ the result of substituting t_i at the place of x_i, inside t, $1 \leq i \leq n$. Denote by P_Γ the subset of $T_\Gamma(x)$ consisting of all trees with just one occurrence of the variable x. P_Γ becomes a monoid with the substitution at x as operation: for $\tau, \pi \in P_\Gamma$, $\tau \cdot \pi = \tau[\pi]$. This monoid is *free* over the set of trees of the form

$$f(t_1, \dots, t_{i-1}, x, t_{i+1}, \dots, t_k), \quad f \in \Gamma_k, k \geq 1, t_j \in T_\Gamma. \tag{1}$$

This means that every $\tau \in P_\Gamma$, $\tau \neq x$, is uniquely decomposed as a product of trees of the form (1) and the number of these trees is denoted by $||\tau||$, actually $||\tau||$ is the length of the unique path starting from the root of τ and ending to its unique leaf labeled by x.

The monoid P_Γ acts, again by substitution at x, on the set T_Γ:

$$P_\Gamma \times T_\Gamma \to T_\Gamma, \quad (\tau, t) \mapsto \tau \cdot t = \tau[t].$$

Let $L \subseteq T_\Gamma$ be a tree language, $t \in T_\Gamma$, and $\tau \in P_\Gamma$. The *right derivative of* L *at* t and the *left derivative of* L *at* τ are

$$Lt^{-1} = \{\tau \mid \tau \in P_\Gamma, \tau \cdot t \in L\}, \quad \tau^{-1}L = \{t \mid t \in T_\Gamma, \tau \cdot t \in L\},$$

respectively. The equivalence relation $\sim_L$ on (the algebra) T_Γ

$$t \sim_L t' \text{ if } Lt^{-1} = Lt'^{-1}$$

is actually a congruence, *i.e.*,

$$t_1 \sim_L t'_1, \dots, t_k \sim_L t'_k \text{ and } f \in \Gamma_k, \quad \text{imply} \quad f(t_1, \dots, t_k) \sim_L f(t'_1, \dots, t'_k).$$

It is called the syntactic congruence of the language L.

Proposition 2.1 (*cf.* [7]). *The following conditions are equivalent for a language* $L \subseteq T_\Gamma$

- (*i*) *L is recognized by a finite tree automaton.*
- (*ii*) $\text{card}\{Lt^{-1} \mid t \in T_\Gamma\} < \infty$.
- (*iii*) $\text{card}\{\tau^{-1}L \mid \tau \in P_\Gamma\} < \infty$.
- (*iv*) *The syntactic congruence $\sim_L$ has finite index (i.e., a finite number of classes).*

In [6] we develop a complexity theory in order to investigate and classify tree languages according to their syntactic structure. Syntactic complexity is a way to measure the complexity of the syntax of a tree language. It counts the number of distinct syntactic classes of trees with a fixed width. Formally the *syntactic complexity* of a tree language $L \subseteq T_\Gamma$ is the function

$$SC_L : \mathbb{N} \to \mathbb{N}_\infty, \quad SC_L(n) = \operatorname{card}\{\bar{t} \mid t \in T_\Gamma, \operatorname{width}(t) = n\}, \ \ n \in \mathbb{N},$$

where $\bar{t}$ stands for the $\sim_L$-class of t. Alternatively we have

$$SC_L(n) = \operatorname{card}\{Lt^{-1} \mid t \in T_\Gamma, \operatorname{width}(t) = n\}, \ \ n \in \mathbb{N}.$$

It should be noticed that $SC_L(n) < \infty$, for all $n \in \mathbb{N}$, whenever the alphabet Γ has no unary symbols, because in this case the set of trees with *width* $= n$ is finite.

A tree language $L \subseteq T_\Gamma$ is said to have

- *bounded syntactic complexity*, whenever there is a constant $c \in \mathbb{N}$ such that

$$SC_L(n) \leq c, \text{ for all } n;$$

- *polynomial syntactic complexity* whenever there is a polynomial

$$p(n) = \sum_{k=0}^{m} a_k n^k, \qquad a_k \in \mathbb{N}$$

  such that

$$SC_L(n) \leq p(n), \text{ for all } n;$$

- *exponential syntactic complexity* whenever there is a constant $c \in \mathbb{N}$ such that

$$SC_L(n) \leq c^n, \text{ for all } n.$$

By Proposition 2.1 every recognizable tree language has bounded syntactic complexity, but the converse is not true. The language

$$L_{bal} = \{t_k \mid k \geq 0\}, \text{ with } t_0 = \alpha,\ t_{k+1} = f(t_k, t_k),\ k \geq 0,$$

where $\operatorname{rank}(\alpha) = 0$, $\operatorname{rank}(f) = 2$, is an instance of a non-recognizable tree language with bounded syntactic complexity: $SC_{L_{bal}}(n) \leq 2$, for all n.

Proposition 2.2 ([6], Prop. 5)**.** *Given the ranked alphabet* $\Gamma = \{f_1, \ldots, f_k, \alpha\}$, $\operatorname{rank}(f_i) = 2$, $1 \leq i \leq k$, $\operatorname{rank}(\alpha) = 0$*, the Dyck tree language of order* k

$$D_k = \{t \mid t \in T_\Gamma,\ |t|_{f_1} = \cdots = |t|_{f_k}\}$$

has polynomial syntactic complexity of degree $k-1$*, namely*

$$SC_{D_k}(n) = \frac{1}{(k-1)!}\, n(n+1)\cdots(n+k-2).$$

Throughout this paper we will often use the alphabet $\hat{\Gamma}$ consisting of a binary symbol f and a nullary symbol α, $\hat{\Gamma} = \{f, \alpha\}$.

Proposition 2.3 ([6], Prop. 6)**.** *The diagonal language*

$$L_d = \{f(t,t) \mid t \in T_{\hat{\Gamma}}\},$$

has exponential syntactic complexity, namely,

$$SC_{L_d}(n+1) = \frac{1}{n+1}\binom{2n}{n} \simeq \frac{4^n}{n^{3/2}\sqrt{\pi}}.$$

3. Formal power series on trees

A *formal power series* on trees (*tree series* for sort), with coefficients in a field K, is just a function $S : T_\Gamma \to K$. The support of S is the set $\{t \in T_\Gamma \mid (S,t) \neq 0\}$. For $t \in T_\Gamma$ the element (S,t) is referred to as the coefficient of S at t. In the set $K\langle\langle T_\Gamma\rangle\rangle$ of all such tree series, addition, scalar multiplication and Hadamard product are defined pointwise

$$(S_1 + S_2, t) = (S_1, t) + (S_2, t), \quad (\lambda S, t) = \lambda(S, t), \quad (S_1 \odot S_2, t) = (S_1, t)(S_2, t),$$

for $S_1, S_2, S \in K\langle\langle T_\Gamma\rangle\rangle$, $\lambda \in K$ and $t \in T_\Gamma$.

We denote by $K\langle T_\Gamma\rangle$ the set of all tree series $S : T_\Gamma \to K$ with finite support, called *polynomials*. Thus any polynomial $p \in K\langle T_\Gamma\rangle$ can be written as a finite formal sum

$$p = \sum_{j=1}^{n} \lambda_j t_j$$

for some $n \geq 1$, $\lambda_j \in K$, and $t_j \in T_\Gamma$.

The *derivatives* of a tree series $S : T_\Gamma \to K$ at $t \in T_\Gamma$ and $\tau \in P_\Gamma$, are defined by

$$St^{-1} : P_\Gamma \to K, \qquad (St^{-1}, \tau') = (S, \tau' t) \text{ for every } \tau' \in P_\Gamma,$$
$$\tau^{-1}S : T_\Gamma \to K, \qquad (\tau^{-1}S, t') = (S, \tau t') \text{ for every } t' \in T_\Gamma,$$

which belong to the K-spaces K^{P_Γ} and K^{T_Γ} of all functions $P_\Gamma \to K$ and $T_\Gamma \to K$ respectively with the pointwise addition and scalar multiplication.

We denote by $V(S)$ (resp. $F(S)$) the subspace K^{P_Γ} (resp. K^{T_Γ}) generated by all the right (resp. left derivatives) of S:

$$V(S) = \langle St^{-1} \mid t \in T_\Gamma\rangle \quad (\text{resp. } F(S) = \langle \tau^{-1}S \mid \tau \in P_\Gamma\rangle).$$

Tree series are interpreted *via* multilinear algebras. A $K - \Gamma$-algebra $\mathcal{A} = (A, \alpha)$ is a K-vector space A together with a family of multilinear functions

$$\alpha_f : A^m \to A, \quad f \in \Gamma_m, m \geq 0.$$

A morphism from $\mathcal{A} = (A, \alpha)$ to $\mathcal{B} = (B, \beta)$ is a linear function $h : A \to B$ compatible with Γ-operations

$$h(\alpha_f(q_1, \ldots, q_m)) = \beta_f(h(q_1), \ldots, h(q_m)), \qquad f \in \Gamma_m, m \geq 0, q_i \in A.$$

The space $K\langle T_\Gamma\rangle$ is converted, in a natural way, into a $K-\Gamma$-algebra with the property: for any $K-\Gamma$-algebra $\mathcal{A}=(A,\alpha)$ there is a function $h_\mathcal{A}: T_\Gamma \to A$ inductively defined by

$$h_\mathcal{A}(f(t_1,\dots,t_m)) = \alpha_f(h_\mathcal{A}(t_1),\dots,h_\mathcal{A}(t_m)), \quad f\in\Gamma_m, m\geq 0, t_i\in T_\Gamma.$$

Its linear extension

$$\bar{h}_\mathcal{A}: K\langle T_\Gamma\rangle \to A, \quad \bar{h}_\mathcal{A}\left(\sum_i \lambda_i t_i\right) = \sum_i \lambda_i h_\mathcal{A}(t_i)$$

is the unique existing $K-\Gamma$-algebra morphism from $K\langle T_\Gamma\rangle$ to $\mathcal{A}$. Call $\mathcal{A}$ *reachable*, whenever $\bar{h}_\mathcal{A}$ is a surjective function. The monoid P_Γ acts on each $K-\Gamma$-algebra $\mathcal{A}=(A,\alpha)$

$$P_\Gamma\times A\to A, \quad (\tau,q)\mapsto \tau\cdot q$$

by induction on $||\tau||$ as follows

- $x\cdot q = q$ for all $q\in A$;
- if $\tau = f(t_1,\dots,t_{i-1},x,t_{i+1},\dots,t_k)$, $f\in\Gamma_k$, $t_j\in T_\Gamma, q\in A$,

$$\tau\cdot q = \alpha_f(h_\mathcal{A}(t_1),\dots,h_\mathcal{A}(t_{i-1}),q,h_\mathcal{A}(t_{i+1}),\dots,h_\mathcal{A}(t_k));$$

- if $\tau=\tau_1\cdot\tau_2$, $\tau_1\neq x\neq\tau_2, q\in A$,

$$\tau\cdot q = \tau_1\cdot(\tau_2\cdot q).$$

Lemma 3.1. *For any $\mathcal{A}=(A,\alpha)$, it holds*

$$h_\mathcal{A}(\tau t) = \tau\cdot h_\mathcal{A}(t), \quad \textit{for all } \tau\in P_\Gamma, t\in T_\Gamma.$$

We can endow $V(S)$ with the structure of a $K-\Gamma$-algebra if for each $f\in\Gamma_k$ we define the operation $(V(S))_f: V(S)^k\to V(S)$ by setting

$$(V(S))_f(St_1^{-1},\dots,St_k^{-1}) = Sf(t_1,\dots,t_k)^{-1}, \quad \text{for all } t_1,\dots,t_k\in T_\Gamma.$$

The so obtained pair $\mathcal{V}_S=(V(S),V(S))$ is called the *syntactic* $K-\Gamma$-*algebra of* S.

Clearly, $\mathcal{V}_S$ is reachable since the function

$$\bar{h}_{\mathcal{V}_S}: K\langle T_\Gamma\rangle\to V(S), \quad \bar{h}_{\mathcal{V}_S}\left(\sum_j \lambda_j t_j\right) = \sum_j \lambda_j St_j^{-1}$$

is surjective.

A tree series $S: T_\Gamma\to K$ is said to be *recognizable* if there exists a pair $(\mathcal{A}=(A,\alpha),\varphi)$, called a *realization* of S, with $\mathcal{A}$ a finitely dimensional $K-\Gamma$-algebra and $\varphi: A\to K$ a linear form such that $S=\varphi\circ h_\mathcal{A}$. We have the next characterization result.

Theorem 3.1 (*cf.* [4])**.** *A tree series* $S : T_\Gamma \to K$ *is recognizable iff* $\dim F(S) < \infty$ *iff* $\dim V(S) < \infty$. *In this case we have* $\dim F(S) = \dim V(S)$.

The *syntactic representation* of $S : T_\Gamma \to K$ is $(\mathcal{V}_S = (V(S), V(S)), \varphi_S)$ where φ_S is the linear form

$$\varphi_S : V(S) \to K, \quad \varphi_S(St^{-1}) = (S, t).$$

This representation is universal in the following sense.

A realization $(\mathcal{A} = (A, \alpha), \varphi)$ of S is reachable if the underline $K - \Gamma$-algebra (A, α) is reachable.

Proposition 3.1 (*cf.* [3])**.** *For every reachable realization* $(\mathcal{A} = (A, \alpha), \varphi)$ *of* S *there is a unique surjective morphism of* $K - \Gamma$*-algebras* $h : \mathcal{A} \to \mathcal{V}_S$ *such that* $h_{\mathcal{V}_S} = h \circ h_{\mathcal{A}}$ *and* $\varphi_S \circ h = \varphi$. *The function* h *is defined as follows: if* $(h_{\mathcal{A}}(t_i))_{i \in I}$ *is a basis for* A, *then*

$$h(h_{\mathcal{A}}(t_i)) = St_i^{-1}, \quad i \in I.$$

A characterization of tree series recognizability through matrix representations is presented in [3] whereas an effective construction of $\mathcal{V}_S$ is given in [2].

Given a language $L \subseteq T_\Gamma$ its characteristic series $\mathrm{ch}(L) : T_\Gamma \to K$, has as coefficients

$$\begin{aligned}(\mathrm{ch}(L), t) &= 1 \quad \text{if} \quad t \in L,\\ &= 0, \ \text{otherwise.}\end{aligned}$$

Proposition 3.2. *If* $L \subseteq T_\Gamma$ *is recognizable, then so is its characteristic series* $\mathrm{ch}(L)$.

Proof. The identity

$$\tau^{-1}\mathrm{ch}(L) = \mathrm{ch}(\tau^{-1}L),$$

holds for every $\tau \in P_\Gamma$. Indeed, for all $t \in T_\Gamma$ we have

$$\begin{aligned}(\tau^{-1}\mathrm{ch}(L), t) = 1 &\Leftrightarrow (\mathrm{ch}(L), \tau t) = 1 \Leftrightarrow \tau t \in L\\ &\Leftrightarrow t \in \tau^{-1}L \Leftrightarrow (\mathrm{ch}(\tau^{-1}L), t) = 1.\end{aligned}$$

Thus $\mathrm{card}\{\tau^{-1}(\mathrm{ch}(L)) \mid \tau \in P_\Gamma\} \leq \mathrm{card}\{\tau^{-1}L \mid \tau \in P_\Gamma\} < \infty$, and so $\dim(\mathrm{ch}(L)) < \infty$, *i.e.*, $\mathrm{ch}(L)$ is recognizable (Thm. 3.1). □

Proposition 3.3. *If the series* $S, S' : T_\Gamma \to K$ *are recognizable, then so are the series* $S + S', \lambda S$ $(\lambda \in K)$, $S \odot S'$, $\tau^{-1}S$ $(\tau \in P_\Gamma)$.

Proof. We only exhibit the case $S \odot S'$. Consider the syntactic realizations $(V(S), V(S), \varphi_S)$, $(V_{S'}, v_{S'}, \varphi_{S'})$ of S, S' respectively. Then the triple

$$(V(S) \otimes V_{S'}, V(S) \otimes v_{S'}, \varphi_S \otimes \varphi_{S'})$$

is a realization of $S \odot S'$ (*cf.* [1]). Here $V(S) \otimes V_{S'}$ denotes the tensor product of the spaces $V(S)$ and $V_{S'}$,

$$(V(S) \otimes v_{S'})_f : (St^{-1})^k \to V(S) \otimes V_{S'}, \quad f \in \Gamma_k, k \geq 0,$$

is given by

$$(V(S) \otimes v_{S'})_f(St_{i_1}^{-1} \otimes S't_{j_1}'^{-1}, \ldots, St_{i_k}^{-1} \otimes S't_{j_k}'^{-1}) = Sf(t_{i_1}, \ldots, t_{i_k})^{-1} \otimes S'f(t'_{j_1}, \ldots, t'_{j_k})^{-1}$$

whereas the linear form $\varphi_S \otimes \varphi_{S'} : V(S) \otimes V_{S'} \to k$ is given by

$$(\varphi_S \otimes \varphi_{S'})(St^{-1} \otimes S't'^{-1}) = (S, t)(S', t').$$

The result follows from the next inequality

$$\dim V_{S \odot S'} \leq \dim(V(S) \otimes V_{S'}) \leq \dim V(S) \cdot \dim V_{S'} < \infty. \qquad \square$$

4. Syntactic complexity of tree series

Syntactic complexity is a tool to study the syntax of a tree series $S : T_\Gamma \to K$. It is described by a function $SC_S : \mathbb{N} \to \mathbb{N}_\infty$ which sends every natural number n to the maximum number of linearly independent right derivatives St^{-1} where t runs over all trees whose width is n. Thus, by setting

$$V_n(S) = \langle St^{-1} \mid t \in T_\Gamma, \text{width}(t) = n \rangle,$$

we get

$$SC_S(n) = \dim V_n(S).$$

Notice that in the case Γ deprives unary symbols ($\Gamma_1 = \emptyset$), then $\dim V_n(S) < \infty$, *i.e.*, SC_S ranges over $\mathbb{N}$. The bounded, polynomial or exponential syntactic complexity classes for tree series $S : T_\Gamma \to K$ are defined analogously to the case of tree languages. The corresponding classes of tree series are denoted $SCBOUND(\Gamma)$, $SCPOL(\Gamma)$, $SCEXP(\Gamma)$.

By Proposition 3.2, it is clear that for every tree language $L \subseteq T_\Gamma$, the syntactic complexity of ch(L) does not exceed that of L.

Proposition 4.1. *For every $L \subseteq T_\Gamma$ it holds $SC_{\text{ch}(L)} \leq SC_L$.*

According to Theorem 3.1 every recognizable series $S : T_\Gamma \to K$ has bounded syntactic complexity, but this fact does not characterize recognizability.

Theorem 4.1. *There is a non-recognizable tree series which has bounded syntactic complexity.*

Proof. Consider the series

$$S : T_{\hat{\Gamma}} \to \mathbb{Q}, \quad (S, t) = \frac{1}{|t|},$$

where $|t|$ is the number of symbols of $\hat{\Gamma} = \{f, \alpha\}$ occurring in t. We construct inductively the sequence of trees

$$t_0 = \alpha, \quad t_{k+1} = f(t_k, \alpha),$$

and observe that $|t_k| = 2k + 1$, $k \geq 0$. Now, the derivatives $St_0^{-1}, \ldots, St_{k-1}^{-1}$, are linearly independent. Indeed, let

$$\lambda_0 St_0^{-1} + \lambda_1 St_1^{-1} + \cdots + \lambda_{k-1} St_{k-1}^{-1} = 0,$$

that is

$$\tfrac{\lambda_0}{|\tau|+|t_0|} + \tfrac{\lambda_1}{|\tau|+|t_1|} + \cdots + \tfrac{\lambda_{k-1}}{|\tau|+|t_{k-1}|} = 0, \quad \text{for all } \tau \in P_{\hat{\Gamma}},$$

where $|\tau|$ is the number of symbols of $\hat{\Gamma}$ occurring in τ. As τ ranges over $P_{\hat{\Gamma}}$ we may view $|\tau|$ as a variable x so that the previous relation can be written as

$$\tfrac{\lambda_0}{x+1} + \tfrac{\lambda_1}{x+3} + \cdots + \tfrac{\lambda_{k-1}}{x+2k-1} = 0, \quad \text{for all } x.$$

By differentiating the above relation p times ($p = 0, 1, \ldots, k-1$) we get the system

$$\frac{\lambda_0}{(x+1)^p} + \frac{\lambda_1}{(x+3)^p} + \cdots + \frac{\lambda_{k-1}}{(x+2k-1)^p} = 0, \quad \text{for all } x.$$

Taking $x = 0$ above, we finally obtain the Vandermonde system

$$\frac{\lambda_0}{1^p} + \frac{\lambda_1}{3^p} + \cdots + \frac{\lambda_{k-1}}{(2k-1)^p} = 0, \quad p = 0, 1, \ldots, k-1,$$

from which we get $\lambda_0 = \lambda_1 = \cdots = \lambda_{k-1} = 0$ as wanted. Consequently, by virtue of Theorem 3.1 the series S is not recognizable.

Let us calculate its syntactic complexity. For every n all trees t with $width(t) = n$ satisfy $|t| = 2n - 1$. It turns out that all derivatives St^{-1}, $width(t) = n$ are equal to each other and so the subspace they generate has dimension 1, that is $SC_S(n) = 1$ for all n. Hence S has bounded syntactic complexity. □

The previous result states that $SCBOUND(\Gamma)$ strictly contains the class $REC(\Gamma)$ of recognizable tree series over Γ. Hence, we get the hierarchy

$$REC(\Gamma) \subset SCBOUND(\Gamma) \subset SCPOL(\Gamma) \subset SCEXP(\Gamma).$$

Lemma 4.1. *Let* $(\mathcal{A} = (A, \alpha), \varphi)$ *be a realization of the series* $S : T_\Gamma \to K$ *and put*

$$V_n(\mathcal{A}) = \langle h_{\mathcal{A}}(t) \mid t \in T_\Gamma, \mathrm{width}(t) = n \rangle.$$

There exists a linear surjective function $h_n : V_n(\mathcal{A}) \to V_n(S)$ *defined as follows: if* $(h_{\mathcal{A}}(t_i))_{1 \leq i \leq k}$ *is a basis of the space* $V_n(\mathcal{A})$*, then*

$$h_n(h_{\mathcal{A}}(t_i)) = St_i^{-1}, \quad 1 \leq i \leq k.$$

Consequently, $\mathrm{dim} V_n(S) \leq \mathrm{dim} V_n(\mathcal{A})$.

Proof. Immediate from Proposition 3.1. □

Proposition 4.2. *Let $S, S' : T_\Gamma \to K$, $\lambda \in K$ and $\tau \in P_\Gamma$, then*

$$SC_{S+S'} \leq SC_S + SC_{S'}, \qquad SC_{\lambda S} = SC_S \ (\lambda \neq 0),$$
$$SC_{S\odot S'} \leq SC_S \cdot SC_{S'}, \qquad SC_{\tau^{-1}S} = SC_S.$$

Therefore, the classes $SCBOUND(\Gamma)$, $SCPOL(\Gamma)$, $SCEXP(\Gamma)$ are closed under sum, scalar product, Hadamard product and left derivatives.

Proof. We only treat the series $S \odot S'$. Applying Lemma 4.1 for $\mathcal{A} = V(S) \otimes V_{S'}$ and taking into account the proof of Proposition 3.3, we get a surjective function $h_n : V_n(V(S) \otimes V_{S'}) \to V_n(S \odot S')$ which maps the vector

$$h_{V(S)\otimes V_{S'}}(t) = h_{V(S)}(t) \otimes h_{V_{S'}}(t)$$

to the vector $(S \odot S')t^{-1}$. Therefore,

$$\begin{aligned}\dim V_n(S \odot S') &\leq \dim V_n(V(S) \otimes V_{S'}) \\ &\leq \dim(V_n(S) \otimes V_n(S')) = \dim V_n(S) \cdot \dim V_n(S')\end{aligned}$$

that is $SC_{S\odot S'} \leq SC_S \cdot SC_{S'}$ as wanted. □

It is not hard to see that the characteristic series of the tree languages D_k, introduced in Section 3, has the same syntactic complexity as D_k which is polynomial of degree $k-1$ (see Prop. 2.2).

A series $S : T_\Gamma \to K$ will be called *syntactically hard* if it has the highest possible complexity. This means that

$$\dim V_n(S) = \mathrm{card}\{t \mid t \in T_\Gamma, \mathrm{width}(t) = n\}.$$

From any series $S : T_\Gamma \to K$ such that $(S, t) \neq 0$ for all $t \in T_\Gamma$, a syntactically hard series $S_d : T_\Gamma \to K$ can be derived by setting

$$\begin{aligned}(S_d, s) &= (S, t)^2, \text{ if } s = f(t, t), \\ &= 0, \text{ otherwise,}\end{aligned}$$

provided Γ has a binary symbol f. For the proof of the fact that S_d is syntactically hard, we have to show that, for all n, the list of derivatives $S_d t^{-1}$, $t \in T_\Gamma$, $\mathrm{width}(t) = n$, is linearly independent. Indeed, assume that

$$\sum_{t\in T_\Gamma, \mathrm{width}(t)=n} \lambda_t S_d t^{-1} = 0$$

or that

$$\sum_{t\in T_\Gamma, \mathrm{width}(t)=n} \lambda_t (S_d, \tau t) = 0, \text{ for all } \tau \in P_\Gamma.$$

Setting above $\tau = f(t,x)$ for some $t \in T_\Gamma$, we find

$$\lambda_t(S_d, f(t,t)) = 0$$

or $\lambda_t(S,t)^2 = 0$, or

$$\lambda_t = 0, \quad \text{for all } t \in T_\Gamma, \text{ width(t)} = n,$$

as wanted.

In the case that we deal with the simple alphabet $\hat{\Gamma} = \{f, \alpha\}$, then it is well known from Combinatorics that $card\{t \mid width(t) = n\}$ is the $(n-1)$th Catalan number C_{n-1}, where

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Hence,

Proposition 4.3. *For every $S : T_{\hat{\Gamma}} \to K$, with $(S,t) \neq 0$ for all $t \in T_{\hat{\Gamma}}$, the series $S_d : T_{\hat{\Gamma}} \to K$ defined above is syntactically hard, i.e., $SC_{S_d}(n) = C_{n-1}$.*

5. Tree height

Berstel and Reutenauer, using a pumping lemma technique, demonstrated that the series sending every tree to its height is not recognizable (*cf.* [1]). Our task in the sequel will be to compute the syntactic complexity of height : $T_\Gamma \to \mathbb{Q}$.

For the sake of simplicity, in this section, we assume that Γ is the ranked alphabet $\hat{\Gamma}$ consisting of a binary symbol f and a nullary symbol α. We denote by $\lfloor r \rfloor$ the the greatest integer less than or equal to r and by $\lceil r \rceil$ the least integer greater than or equal to r.

Theorem 5.1. *The syntactic complexity of height is given by the formula*

$$SC_{\text{height}}(n) = \lfloor n - \log_2 n \rfloor, \quad \text{for all } n.$$

Proof. First of all we observe that, for all $t \in T_\Gamma$, it holds

$$\text{height}(t) + 1 \leq \text{width}(t) \leq 2^{\text{height}(t)}$$

from which we get that

$$\log_2(\text{width}(t)) \leq \text{height}(t) \leq \text{width}(t) - 1. \tag{2}$$

Our next step will be to compute the dimension of the subspace

$$V_n(\text{height}) = \langle \text{height}\, t^{-1} \mid t \in T_\Gamma, \text{width}(t) = n \rangle, \quad \text{for all } n.$$

According to (2) if width$(t) = n$, then

$$\log_2 n \leq \text{height}(t) \leq n - 1. \tag{3}$$

Let $t_1, \dots, t_k$ be trees with different heights, $\text{height}(t_i) = m + i$, $1 \le i \le k$, satisfying (3) ($m = \lceil \log_2 n - 1 \rceil$ and $k = \lfloor n - \log_2 n \rfloor$). We are going to show that the derivatives

$$\text{height}\, t_1^{-1}, \dots, \text{height}\, t_k^{-1}$$

form a basis of $V_n(\text{height})$. Assume that

$$\lambda_1(\text{height}\, t_1^{-1}) + \cdots + \lambda_k(\text{height}\, t_k^{-1}) = 0$$

or that

$$\lambda_1 \text{height}(\tau t_1) + \cdots + \lambda_k \text{height}(\tau t_k) = 0$$

for all $\tau \in P_\Gamma$. Choose $\tau_1, \dots, \tau_k \in P_\Gamma$ such that $\text{height}(\tau_i) = m + k$ and $||\tau_i|| = i$. By evaluating the above equation for $\tau = \tau_i$, $1 \le i \le k$, we obtain the linear homogeneous system

$$\lambda_1 \text{height}(\tau_i t_1) + \cdots + \lambda_k \text{height}(\tau_i t_k) = 0 \tag{4}$$

for $i = 1, \dots, k$.

Now we need to calculate $\text{height}(\tau_i t_j)$ for every i and j. Let $i = 1$, this means that the path from the root of τ to the unique leaf labeled by the variable x has length 1 while the height of τ is $m + k$. Hence, the height of the tree $\tau_1 t_j$ will remain $m+k$ when the height of t_j is smaller than $m+k$ (*i.e.*, for $j = 1, \dots, k-1$) and only for $j = k$, which means that $\text{height}(t_k) = m + k$, the height of $\tau_1 t_k$ will become $1 + m + k$. Thus, we get that

$$\text{height}(\tau_1 t_j) = \begin{cases} m + k, & \text{for } j = 1, \dots, k-1; \\ m + k + 1, & \text{for } j = k. \end{cases}$$

Using the same arguments we get that

$$\text{height}(\tau_2 t_j) = \begin{cases} m + k, & \text{for } j = 1, \dots, k-2; \\ m + k + 1, & \text{for } j = k-1; \\ m + k + 2, & \text{for } j = k, \end{cases}$$

and similarly for $i > 2$. Finally for $i = k$ we get

$$\text{height}(\tau_k t_j) = m + k + j, \qquad \text{for } j = 1, \dots, k.$$

Therefore, the system (4) becomes

$$\begin{aligned} \lambda_1(m+k) + \lambda_2(m+k) + \cdots + \lambda_{k-1}(m+k) + \lambda_k(m+k+1) &= 0 \\ \lambda_1(m+k) + \lambda_2(m+k) + \cdots + \lambda_{k-1}(m+k+1) + \lambda_k(m+k+2) &= 0 \\ \cdots &= 0 \\ \lambda_1(m+k+1) + \lambda_2(m+k+2) + \cdots + \lambda_{k-1}(m+2k-1) + \lambda_k(m+2k) &= 0 \end{aligned}$$

whose determinant is non vanishing and so $\lambda_1 = \cdots = \lambda_k = 0$. This shows that

$$\text{height}\, t_1^{-1}, \ldots, \text{height}\, t_k^{-1}$$

are linearly independent. On the other hand we observe that if $t, t' \in T_\Gamma$ are such that $\text{height}(t) = \text{height}(t')$, then $\text{height}\, t^{-1} = \text{height}\, t'^{-1}$ and so $\text{height}\, t_1^{-1}, \ldots,$ $\text{height}\, t_k^{-1}$ generates the space $V_n(\text{height})$. We conclude that

$$\dim V_n(\text{height}) = \lfloor n - \log_2 n \rfloor, \quad \text{for all } n$$

hence the proposed formula for SC_{height}. □

Remark. In the case that Γ is an arbitrary finite ranked alphabet with no unary symbols then the inequality (2) becomes

$$\log_{k_2} \text{width}(t) \leq \text{height}(t) \leq \frac{\text{width}(t) - 1}{k_1 - 1}$$

where k_1 (resp. k_2) is the smallest (resp. greatest) positive rank such that $\Gamma_{k_1} \neq \emptyset$ (resp $\Gamma_{k_2} \neq \emptyset$). However, the computed syntactic complexity is in the same class as above.

In the sequel we display a tree language having the same syntactic complexity with the series height. The language L_{avl} of AVL-trees consists of all trees $t \in T_\Gamma$ such that either $t = \alpha$ or $t = f(t_1, t_2)$ with $t_1, t_2 \in L_{avl}$ and

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

If $t, t' \in L_{avl}$ and $\text{height}(t) = \text{height}(t')$, then $L_{avl}t^{-1} = L_{avl}t'^{-1}$. Moreover, choosing a sequence of AVL-trees t_n with $\text{height}(t_n) = n$, $(n \geq 0)$, it is easy to see that the corresponding derivatives $L_{avl}t^{-1}$ are distinct. Hence L_{avl} is not recognizable.

For a fixed n, we have

$$\text{card}\{L_{avl}t^{-1} \mid \text{width}(t) = n\} = \text{card}\{\text{height}(t) \mid \text{width}(t) = n\}$$

and the last number, as we have previously seen, is $\lfloor n - \log_2 n \rfloor$. We conclude that

Proposition 5.1. *It holds* $SC_{L_{avl}} = SC_{\text{height}}$.

Recall that for p prime the ring $\mathbb{Z}_p$ of modulo p integers is a field. We notice the rather surprising result that the series sending every tree t to its height modulo p (p prime) is recognizable. Indeed, we have to show that the space generated by the right derivatives of the series

$$\text{height}_p : T_\Gamma \to \mathbb{Z}_p, \quad \text{height}_p(t) = \text{height}(t) (\text{mod}\, p)$$

has finite dimension. Actually we shall show that

$$\dim V_{\text{height}_p} = p.$$

As always $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ stands for the field of *mod* p integers.

Let us choose trees t_i, with $\text{height}(t_i) = i$, $0 \leq i \leq p-1$ and suppose that

$$\lambda_0 \text{height}_p t_0^{-1} + \lambda_1 \text{height}_p t_1^{-1} + \cdots + \lambda_{p-1} \text{height}_p t_{p-1}^{-1} = 0$$

or that

$$\lambda_0 \text{height}_p(\tau t_0) + \lambda_1 \text{height}_p(\tau t_1) + \cdots + \lambda_{p-1} \text{height}_p(\tau t_{p-1}) = 0$$

for all $\tau \in P_\Gamma$. Choose $\tau_1, \ldots, \tau_{p-1} \in P_\Gamma$ such that $\text{height}(\tau_i) = p-1$ and $||\tau_i|| = i$. Evaluating the previous equality for $\tau = \tau_i$, $0 \leq i \leq p-1$ we find a linear homogeneous system whose determinant

$$\begin{vmatrix} p-1 & p-1 & \cdots & p-1 & p-1 & p-1 \\ p-1 & p-1 & \cdots & p-1 & p-1 & 0 \\ p-1 & p-1 & \cdots & p-1 & 0 & 1 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots \\ p-1 & p-1 & 0 & 1 & \cdots & p-3 \\ p-1 & 0 & 1 & \cdots & p-3 & p-2 \end{vmatrix} = \begin{vmatrix} 0 & \cdots & 0 & p-1 \\ \vdots & \cdots & p-1 & \vdots \\ 0 & \cdots & \cdots & \vdots \\ p-1 & \cdots & \cdots & \vdots \end{vmatrix}$$

is $(-1)^p(p-1)^p$. Since p is prime, p does not divide $(p-1)^p$, that is the above determinant is non vanishing modulo p. Thus $\lambda_i = 0$, for $i = 0, 1, \ldots, p-1$ and so the considered list of right derivatives is linearly independent. Moreover, it generates the space V_{height_p} since for every $t \in T_\Gamma$ we have

$$\text{height}_p t^{-1} = \text{height}_p t_k^{-1}, \quad \text{with } k = \text{height}_p t.$$

We conclude:

Proposition 5.2. *For every prime number* p *the tree series* height_p *is recognizable. Moreover,* $\dim V_{\text{height}_p} = p$ *and so* $\text{card} V_{\text{height}_p} = p^p$.

6. The complexity of recognizable tree series

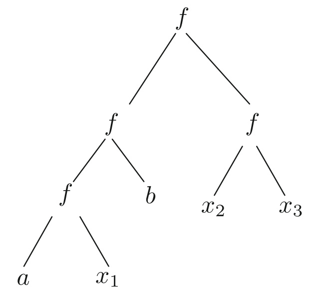
The hierarchy formed by the previously introduced syntactic complexity locates all recognizable tree series into its first level, the class of bounded tree series. Our intention in the present section is to built a hierarchy within the class of recognizable series by providing an efficient complexity measure for these series.

As we have seen the function SC_S introduced in the previous section gives no information for recognizable tree series. Our intention in the present section is to provide an efficient complexity measure for these series.

Let us denote by $P_\Gamma^{(n)}$ the subset of $T_\Gamma(X_n)$ formed by all trees where $x_1, \ldots, x_n$ occur in the yield of the tree (in this order from left to right) exactly once.

For instance the tree

$\tau =$



$\in P_\Gamma^{(3)}$.

For every $n \geq 1$ there is a function

$$P_\Gamma^{(n)} \times T_\Gamma^n \to T_\Gamma, \quad (\tau, t_1, \ldots, t_n) \mapsto \tau[t_1, \ldots, t_n].$$

With respect to $S : T_\Gamma \to K$, two dual notions of derivatives can be defined:

$$\tau^{-1}S : T_\Gamma^n \to K, \quad (\tau^{-1}S, (t_1, \ldots, t_n)) = (S, \tau[t_1, \ldots, t_n]),$$

and

$$S(t_1, \ldots, t_n)^{-1} : P_\Gamma^{(n)} \to K, \quad (S(t_1, \ldots, t_n)^{-1}, \tau) = (S, \tau[t_1, \ldots, t_n]),$$

for all $\tau \in P_\Gamma^{(n)}$ and $t_1, \ldots, t_n \in T_\Gamma$. For every n we set

$$F(S)^{(n)} = \langle \tau^{-1}S \mid \tau \in P_\Gamma^{(n)} \rangle \text{ and } V(S)^{(n)} = \langle S(t_1, \ldots, t_n)^{-1} \mid t_1, \ldots, t_n \in T_\Gamma \rangle.$$

In particular $F(S)^{(1)} = F(S)$ and $V(S)^{(1)} = V(S)$.

We need some additional notation. Given a $K-\Gamma$-algebra $\mathcal{A} = (A, \alpha)$, for every $t \in T_\Gamma(X_n)$ and every $q_1, \ldots, q_n \in A$, the element $t[q_1, \ldots, q_n] \in A$ is inductively defined as follows

- for $t = x_i$, $x_i[q_1, \ldots, q_n] = q_i$, $1 \leq i \leq n$;
- for $t = c \in \Gamma_0$, $c[q_1, \ldots, q_n] = \alpha_c$;
- for $t = f(t_1, \ldots, t_k)$, $f \in \Gamma_k$, $k \geq 1$, $t_i \in T_\Gamma(X_n)$

$$f(t_1, \ldots, t_k)[q_1, \ldots, q_n] = \alpha_f(t_1[q_1, \ldots, q_n], \ldots, t_k[q_1, \ldots, q_n]).$$

Lemma 6.1. *For any $K - \Gamma$-algebra $\mathcal{A} = (A, \alpha)$ it holds:*

$$h_\mathcal{A}(\tau[t_1, \ldots, t_n]) = \tau[h_\mathcal{A}(t_1), \ldots, h_\mathcal{A}(t_n)]$$

for all $\tau \in P_\Gamma^{(n)}$ and $t_1, \ldots, t_n \in T_\Gamma$.

Proof. Straightforward. □

Lemma 6.2. *Let $\mathcal{A} = (A, \alpha)$ be a finitely dimensional $K - \Gamma$-algebra. Then, for every $t \in T_\Gamma$,* height$(t) >$ dim$A$*, the element $h_\mathcal{A}(t)$ can be written as a linear combination of elements of the form $h_\mathcal{A}(s)$, for $s \in T_\Gamma$ with* height$(s) \leq$ dim$A$.

Proof. Let us decompose t as follows

$$t = \tau_p \tau_{p-1} \cdots \tau_1 c, \qquad c \in \Gamma_0$$

where all τ_i are of the form

$$\tau_i = f_i(t_{i,1}, \ldots, t_{i,k_i-1}, x, t_{i,k_i+1}, \ldots, t_{i,\lambda_i}), \quad f_i \in \Gamma_{\lambda_i}, t_{i,j} \in T_\Gamma$$

and $p > \dim A$. Since the list

$$h_{\mathcal{A}}(c), h_{\mathcal{A}}(\tau_1 c), \ldots, h_{\mathcal{A}}(\tau_{p-1} \cdots \tau_1 c), h_{\mathcal{A}}(\tau_p \tau_{p-1} \cdots \tau_1 c)$$

contains more than *domA* entries, some of them will be linear combination of its previous elements in the above list. Hence there is an i with $1 \leq i \leq p$ and:

$$h_{\mathcal{A}}(\tau_i \tau_{i-1} \cdots \tau_1 c) = \sum_{j=1}^{i-1} \lambda_j h_{\mathcal{A}}(\tau_j \tau_{j-1} \cdots \tau_1 c)$$

and so

$$\begin{aligned} h_{\mathcal{A}}(t) = h_{\mathcal{A}}(\tau_p \tau_{p-1} \cdots \tau_1 c) &= \tau_p \tau_{p-1} \cdots \tau_{i+1} h_{\mathcal{A}}(\tau_i \tau_{i-1} \cdots \tau_1 c) \\ &= \sum_{j=1}^{i-1} \lambda_j \tau_p \tau_{p-1} \cdots \tau_{i+1} h_{\mathcal{A}}(\tau_j \tau_{j-1} \cdots \tau_1 c) \\ &= \sum_{j=1}^{i-1} \lambda_j h_{\mathcal{A}}(\tau_p \tau_{p-1} \cdots \tau_{i+1} \tau_j \tau_{j-1} \cdots \tau_1 c) \\ &= \sum_{j=1}^{i-1} \lambda_j h_{\mathcal{A}}(t_j) \end{aligned}$$

with $t_j = \tau_p \tau_{p-1} \ldots \tau_{i+1} \tau_j \tau_{j-1} \ldots \tau_1 c$ and hence $\mathrm{height}(t_j) \leq \mathrm{height}(t)$ for all j.

By repeating this process if necessary we arrive to the desired decomposition. □

Theorem 6.1. *For a tree series $S : T_\Gamma \to K$ the next conditions are equivalent.*

(*i*) *S is recognizable;*
(*ii*) $\dim F(S)^{(n)} < \infty$, *for all n;*
(*iii*) $\dim V(S)^{(n)} < \infty$, *for all n.*

In this case we have

$$\dim F(S)^{(n)} = \dim V(S)^{(n)}.$$

Proof. $(i) \Rightarrow (ii)$. Assume that $(\mathcal{A} = (A, \alpha), \varphi)$ is a finitely dimensional realization of S. We shall demonstrate that for every $t_1, \ldots, t_n \in T_\Gamma$

$$S(t_1, \ldots, t_n)^{-1} \in \langle S(s_1, \ldots, s_n)^{-1} \mid s_i \in T_\Gamma, \mathrm{height}(s_i) \leq \dim A,\ 1 \leq i \leq n \rangle.$$

Indeed, for all $\tau \in P_\Gamma^{(n)}$ we have

$$(S(t_1,\dots,t_n)^{-1},\tau) = (S,\tau[t_1,\dots,t_n]) = \varphi(h_{\mathcal{A}}(\tau[t_1,\dots,t_n])) \\ = \varphi(\tau[h_{\mathcal{A}}(t_1),\dots,h_{\mathcal{A}}(t_n)]) \quad \text{(by Lem. 6.1)}.$$

By Lemma 6.2 we have that there are s_j with $\mathrm{height}(s_j) \le \dim A$ and

$$h_{\mathcal{A}}(t_i) = \sum_{j=1}^{q} \lambda_{ij} h_{\mathcal{A}}(s_j), \qquad q = \dim\langle h_{\mathcal{A}}(s) \mid \mathrm{height}(s) \le \dim A\rangle$$

and so by continuing the above string of equalities we get

$$\begin{aligned}
&= \sum_{j_1,\dots,j_n=1}^{q} \lambda_{1j_1}\cdots\lambda_{nj_n}\varphi(\tau[h_{\mathcal{A}}(s_{j_1}),\dots,h_{\mathcal{A}}(s_{j_n})]) \\
&= \sum_{j_1,\dots,j_n=1}^{q} \lambda_{1j_1}\cdots\lambda_{nj_n}\varphi(h_{\mathcal{A}}(\tau[s_{j_1},\dots s_{j_n}])) \\
&= \sum_{j_1,\dots,j_n=1}^{q} \lambda_{1j_1}\cdots\lambda_{nj_n}(S,\tau[s_{j_1},\dots s_{j_n}]) \\
&= \sum_{j_1,\dots,j_n=1}^{q} \lambda_{1j_1}\cdots\lambda_{nj_n}(S(s_{j_1},\dots s_{j_n})^{-1},\tau) \\
&= \left(\sum_{j_1,\dots,j_n=1}^{q} \lambda_{1j_1}\cdots\lambda_{nj_n}S(s_{j_1},\dots s_{j_n})^{-1},\tau\right).
\end{aligned}$$

Consequently, $S(t_1,\dots t_n)^{-1} = \sum_{j_1,\dots,j_n} \lambda_{1j_1}\cdots\lambda_{nj_n}S(s_{j_1},\dots s_{j_n})^{-1}$ as asserted.

$(ii) \Rightarrow (i)$. Follows from Theorem 3.1 and the fact that $P_\Gamma^{(1)} = P_\Gamma$.

$(iii) \Rightarrow (ii)$. Assume that $\dim V(S)^{(n)} = k$ and choose a basis of $V(S)^{(n)}$

$$S(t_{11},\dots t_{1n})^{-1},\dots,S(t_{k1},\dots t_{kn})^{-1}. \tag{5}$$

The function $\Phi : F(S)^{(n)} \to K^k$ defined by the formula

$$\Phi(\tau^{-1}S) = ((S,\tau[t_{11},\dots,t_{1n}]),\dots,(S,\tau[t_{k1},\dots,t_{kn}]))$$

is well defined and linear. In addition it is injective that is its kernel collapses to the zero vector

$$\Phi(p) = (0,\dots,0) \text{ implies } p = 0.$$

Indeed, let

$$p = \sum_j \lambda_j(\tau_j^{-1}S),$$

the condition $\Phi(p) = (0, \dots, 0)$ means that

$$\sum_j \lambda_j (S, \tau_j[t_{i1}, \dots, t_{in}]) = 0 \text{ for all } i = 1, \dots, k. \tag{6}$$

Then for every $t_1, \dots, t_n \in T_\Gamma$ we have

$$\begin{aligned}\left(\sum_j \lambda_j (\tau_j^{-1} S), (t_1, \dots, t_n) \right) &= \sum_j \lambda_j (S, \tau_j[t_1, \dots, t_n]) \\ &= \sum_j \lambda_j (S(t_1, \dots, t_n)^{-1}, \tau_j).\end{aligned}$$

By decomposing $S(t_1, \dots, t_n)^{-1}$ along the basis (5),

$$S(t_1, \dots, t_n)^{-1} = \sum_i \mu_i S(t_{i1}, \dots, t_{in})^{-1}$$

and replacing it in the last member of the above equalities we get

$$\begin{aligned}&= \sum_j \lambda_j \left(\sum_i \mu_i (S(t_{i1}, \dots, t_{in})^{-1}, \tau_j) \right) = \sum_i \mu_i \sum_j \lambda_j (S, \tau_j[t_{i1}, \dots, t_{in}]) \\ &\overset{(6)}{=} \sum_i \mu_i \cdot 0 = 0.\end{aligned}$$

Therefore, $p = 0$ as wanted. From the injectivity of Φ we obtain

$$\dim F(S)^{(n)} \leq \dim K^k = k = \dim V(S)^{(n)}.$$

$(ii) \Rightarrow (iii)$. We only have to dualize the previous argument to get the inequality

$$\dim V(S)^{(n)} \leq \dim F(S)^{(n)}. \qquad \square$$

The *refined syntactic complexity* of a tree series $S : T_\Gamma \to K$ is the function

$$RSC_S : \mathbb{N} \to \mathbb{N}_\infty, \qquad RSC_S(n) = \dim F(S)^{(n)}, \text{ for all } n.$$

According to the previous theorem, for any recognizable series S, the function RSC_S cannot reach ∞.

Example 6.1. All the tree series

- tree size: $T_{size} : T_\Gamma \to \mathbb{Q}$, $(T_{size}, t) = |t|$;
- branch enumeration: $Ben : T_\Gamma \to \mathbb{Q}$, (Ben, t) = number of branches of t;
- branch length: $BL : T_\Gamma \to \mathbb{Q}$, (BL, t) = sum of lengths of all branches of t,

are recognizable, and they have the same linear refined syntactic complexity, namely

$$RSC_S(n) = n + 1, \quad \text{for all } n.$$

The complexity notion we have introduced is only appropriate for recognizable tree series as confirms the following result.

Proposition 6.1. *Assume that the alphabet* Γ *has no unary symbols and* $S : T_\Gamma \to K$ *is a non-recognizable tree series. Then*

$$RSC_S(n) = \infty, \text{ for all } n.$$

Proof. First we establish the logical implication

$$\dim F(S)^{(n+1)} < \infty \Rightarrow \dim F(S)^{(n)} < \infty, \text{ for all } n.$$

Suppose that

$$\tau_1^{-1}S, \ldots, \tau_p^{-1}S, \qquad (\tau_1, \ldots, \tau_p \in P_\Gamma^{(n+1)}) \tag{7}$$

is a basis of $F(S)^{(n+1)}$. For every $\tau \in P_\Gamma^{(n+1)}$, $i \in \{1, 2, \ldots, n+1\}$ and $c \in \Gamma_0$ we introduce the tree

$$\tau^{(i,c)} = \tau[x_1, \ldots, x_{i-1}, c, x_i, \ldots, x_n] \in P_\Gamma^n.$$

We denote by A_n the set of all trees $\pi \in P_\Gamma^{(n)}$ whose yield is $x_1 x_2 \cdots x_n$. The hypothesis $\Gamma_1 = \emptyset$ ensures that A_n is a finite set. We shall show that the finite set of series

$$\{\pi^{-1}S \mid \pi \in A_n\} \cup \{(\tau_j^{(i,c)})^{-1}S \mid j = 1, \ldots, p,\ i = 1, \ldots, n+1,\ c \in \Gamma_0\} \tag{8}$$

generates the subspace $F(S)^{(n)}$ and so its dimension cannot exceed the cardinality of this list. For this it suffices to show that for every $\pi \in P_\Gamma^{(n)}$, the series $\pi^{-1}S$ is a linear combination of the list (8). If $y(\pi) = x_1 x_2 \cdots x_n$, then $\pi \in A_n$ and we have nothing to show. Assume that $y(\pi) = w_0 x_1 w_1 \cdots x_{i-1} w_{i-1} \cdots x_n w_n$ with at least one of the words $w_0, w_1, \ldots, w_n$ non empty. For instance let $w_{i-1} = u_{i-1}c$, $c \in \Gamma_0$ and take

$$\bar{\pi} = \pi[x_1/x_1, \ldots, x_{i-1}/x_{i-1}, x_i/c, x_{i+1}/x_i, \ldots, x_{n+1}/x_n] \in P_\Gamma^{(n+1)}$$

then $\bar{\pi}^{-1}S$ is written as a linear combination of the basis (7)

$$\bar{\pi}^{-1}S = \sum_{j=1}^{p} \lambda_j \tau_j^{-1}S \qquad (\lambda_j \in K).$$

We have

$$\pi^{-1}S = (\bar{\pi}^{(i,c)})^{-1}S = \sum_{j=1}^{p} \lambda_j (\tau_j^{(i,c)})^{-1}S,$$

as wanted. It follows that

$$\dim F(S)^{(n)} = \infty \text{ implies } \dim F(S)^{(n+1)} = \infty, \text{ for all } n. \tag{9}$$

Since $P_\Gamma^{(1)} = P_\Gamma$ and $F_S^{(1)} = F_S$ and S is non-recognizable, it follows from Theorem 1 that $\dim F_S^{(1)} = \infty$. Thus, by (9) we have that $\dim F_S^{(n)} = \infty$ for all $n \geq 1$. □

As we have seen in Section 5, the series $\text{height}_p : T_\Gamma \to \mathbb{Z}_p$ is recognizable for any prime number p. Its refined syntactic complexity is exponential, namely $RSC_{\text{height}_p}(n) = p^n$, for all n. Here we need some notation, for $\tau \in P_\Gamma^{(n)}$, $||\tau||_k$ is the length of the path starting from the root of τ and ending to the variable x_k, $1 \leq k \leq n$. Next, for $i_1, \ldots, i_n \in \{0, 1, \ldots, p-1\}$, consider the trees $\tau_{i_1,\ldots,i_n} \in P_\Gamma^{(n)}$ so that

$$\text{height}(\tau_{i_1,\ldots,i_n}) = p - 1 \text{ and } ||\tau_{i_1,\ldots,i_n}||_k = i_k,\ 1 \leq k \leq n.$$

The derivatives

$$\tau_{i_1,\ldots,i_n}^{-1}\text{height}_p : T_\Gamma^n \to \mathbb{Z}_p, \quad 0 \leq i_1, \ldots, i_n \leq p - 1$$

are linearly independent *i.e.*, the equation

$$\sum_{0 \leq i_1,\ldots,i_n \leq p-1} \lambda_{i_1,\ldots,i_n} \tau_{i_1,\ldots,i_n}^{-1} \text{height}_p = 0 \tag{10}$$

implies that all coefficients $\lambda_{i_1,\ldots,i_n}$ are vanishing. For this let us choose the tree $t_j \in T_\Gamma$, $\text{height}(t_j) = j$, $0 \leq j \leq p - 1$ and evaluate the equation (10) at all the n-tuples $(t_{j_1}, \ldots, t_{j_n})$, $0 \leq j_1, \ldots, j_n \leq p - 1$. There results the homogeneous system

$$\sum_{0 \leq i_1,\ldots,i_n \leq p-1} \lambda_{i_1,\ldots,i_n} \text{height}_p(\tau_{i_1,\ldots,i_n} t_{j_1,\ldots,j_n}) = 0, \tag{11}$$

for all $0 \leq j_1, \ldots, j_n \leq p - 1$, where $\text{height}_p(\tau_{i_1,\ldots,i_n} t_{j_1,\ldots,j_n})$ is

- $p - 1$, if $\max(i_1 + j_1, \ldots, i_n + j_n) \leq p - 1$,
- $\max(i_1 + j_1, \ldots, i_n + j_n) - p$, if $\max(i_1 + j_1, \ldots, i_n + j_n) \geq p$.

At this point we proceed as in the proof of Proposition 5.2: the determinant of the system (11) is equal to $(-1)^{p^n}(p - 1)^{p^n}$ which is non zero modulo p. We summarize:

Proposition 6.2. *The modulo p height function* $\text{height}_p : T_\Gamma \to \mathbb{Z}_p$ *has exponential refined syntactic complexity:*

$$RSC_{\text{height}_p}(n) = p^n, \qquad \textit{for all } n.$$

By denoting $RSCBOUND(\Gamma)$, $RSCPOL(\Gamma)$, $RSCEXP(\Gamma)$ the classes of recognizable series with bounded, polynomial and exponential complexity respectively, we obtain the hierarchy

$$RSCBOUND(\Gamma) \subset RSCPOL(\Gamma) \subset RSCEXP(\Gamma)$$

inside the class $REC(\Gamma)$.

Acknowledgements. We wish to thank the referees for their corrections and remarks which have clearly improved this paper.

References

[1] J. Berstel and C. Reutenauer, Recognizable formal power series on trees. *Theoret. Comput. Sci.* **18** (1982) 115–148.
[2] S. Bozapalidis, Effective construction of the syntactic algebra of a recognizable series on trees. *Acta Informatica* **28** (1991) 351–363.
[3] S. Bozapalidis and A. Alexandrakis, Représentations Matricielles des Séries d'Arbres Reconnaissables. *Theor. Inf. Appl.* **23** (1989) 449–459.
[4] S. Bozapalidis and O.L. Bozapalidou, The rank of a formal tree power series. *Theoret. Comput. Sci.* **27** (1983) 211–215.
[5] S. Bozapalidis and A. Kalampakas, Recognizability of graph and pattern languages. *Acta Informatica* **42** (2006) 553–581.
[6] S. Bozapalidis and A. Kalampakas, *On the Complexity of the Syntax of Tree Languages*, Proceedings of the 3rd International Conference of Algebraic Informatics, CAI09. *Lect. Notes Comput. Sci.* **5725** (2009) 189–203.
[7] F. Gécseg and M. Steinby, *Tree Automata*. Akademiai Kiado, Budapest (1984).
[8] A. Kalampakas, *The Syntactic Complexity of Eulerian Graphs*, Proceedings of the 2nd International Conference of Algebraic Informatics, CAI07. *Lect. Notes Comput. Sci.* **4728** (2007) 208–217.

Communicated by J. Berstel.
Received March 17, 2009. Accepted December 16, 2009.