On Bounded Linear Codes and the Commutative Equivalence

Arturo Carpi
Dipartimento di Matematica e Informatica,
Università degli Studi di Perugia,
via Vanvitelli 1, 06123 Perugia, Italy
e-mail: arturo.carpi@unipg.it

Flavio D'Alessandro *
Dipartimento di Matematica,
Università di Roma "La Sapienza"
Piazzale Aldo Moro 2, 00185 Roma, Italy
e-mail: dalessan@mat.uniroma1.it
and
Department of Mathematics,
Boğaziçi University
34342 Bebek, Istanbul, Turkey

Abstract

The problem of the commutative equivalence of semigroups generated by semi-linear languages is studied. In particular conditions ensuring that the Kleene closure of a bounded semi-linear code is commutatively equivalent to a regular language are investigated.

Keywords: Commutative equivalence, Bounded semi-linear language, Uniquely decipherable code, Kleene closure, Exponential growth

1 Introduction

In this paper, we study the commutative equivalence of context-free and regular languages. Two words are said to be *commutatively equivalent* if one is obtained from the other by rearranging the letters of the word. Two languages L_1 and L_2 are said to be *commutatively equivalent* if there exists a bijection $f\colon L_1 \to L_2$ such that every word $u \in L_1$ is commutatively equivalent to $f(u)$. This notion plays an important role in the study of several problems of Theoretical Computer Science such as, for instance, in the Theory of Codes, where it is involved in the

*The research of F. D'Alessandro was partially supported by Project 2221 of TUBITAK (The Scientific and Technological Research Council of Turkey).

celebrated Schützenberger conjecture about the commutative equivalence of a maximal finite code with a prefix code (see e.g. [3, 36]). The question of our interest can be formulated as follows:

Commutative Equivalence Problem. *Given a context-free language L_1, does there exist a regular language L_2 which is commutatively equivalent to L_1?*

In the sequel, for short, we refer to it as *CE Problem.* A language which is commutatively equivalent to a regular one will be called *commutatively regular.*

It is worth noticing that commutatively equivalent languages share the same alphabet and their generating series are equal. In particular, the generating series of a commutatively regular language must be rational. This remark leads us to recall that a conceptually related study was conducted by Béal and Perrin in [2], where the generating series of regular languages on alphabets of prescribed size are studied. Béal and Perrin provided a characterization of such series and this remarkable contribution thus defines the theoretical setting in which the CE Problem can be naturally fitted in.

For our discussion, the following notions are useful. Given a language L, the *growth function* g_L returns, for any non-negative integer n, the number of the words of L whose length is less than or equal to n. A language L is called *sparse* if its growth function is polynomially upper bounded. A language L is said to be *of exponential growth* if there exists a real number $k > 1$ such that $g_L(n) > k^n$ for all sufficiently large n. Two results are relevant in this context. The first proved in [5, 32, 38] states that every context-free language is either sparse or of exponential growth. The second, proved in [28, 34], states that the class of sparse context-free languages coincides with that of bounded context-free languages. We recall that a language L is termed *bounded* if there exist k words $u_1, \ldots, u_k$ such that $L \subseteq u_1^* \cdots u_k^*$. Bounded context-free languages play a meaningful role in Computer Science and in Mathematics and have been widely investigated in the past so that their structure has been characterized by several theorems [4,9,10,12,13,17–19,22,23,25–31,34,35,37]. A characterization of regular bounded sets, based upon a combinatorial property of the factors of the words of the language, has been obtained by Restivo [37] and, subsequently, extended to context-free languages by Boasson and Restivo [4].

Very recently, results on the counting functions of context-free languages, based upon the notion of strongly counting-regularity, have been obtained in [29]. An excellent survey on the relationships between bounded languages and semigroups has been given by de Luca and Varricchio in [19].

Another theorem that is central in this setting has been proved by Ginsburg and Spanier [22] (see also [23]). This theorem allows one to represent, in a canonical way, bounded context-free languages by means of sets of vectors. For this purpose, let us first introduce a notion. Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded language where, for every $i = 1, \ldots, k$, u_i is a word over the alphabet A. Let $\varphi\colon \mathbb{N}^k \to u_1^* \cdots u_k^*$ be the map defined as: for every tuple $(\ell_1, \ldots, \ell_k) \in \mathbb{N}^k$,

$$\varphi(\ell_1, \ldots, \ell_k) = u_1^{\ell_1} \cdots u_k^{\ell_k}.$$

The map φ is called the *Ginsburg map*. Ginsburg and Spanier proved that L is context-free if and only if the subset $\varphi^{-1}(L)$ of $\mathbb{N}^k$ is a finite union of linear sets, each having a stratified set of periods. Roughly speaking, a stratified set of periods corresponds to a system of well-formed parentheses.

In view of Ginsburg and Spanier's theorem, bounded context-free languages are special instances of a broader class of languages called *bounded semi-linear*. A language $L \subseteq u_1^* \cdots u_k^*$ is called bounded semi-linear if $L = \varphi(B)$, where B is a semi-linear set, that is, a finite union:

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

of *linear subsets* B_i of $\mathbb{N}^k$, $1 \leq i \leq n$, of dimension $m_i \geq 0$:

$$B_i = \{\mathbf{b}_{i0} + x_{i1}\mathbf{b}_{i1} + \cdots + x_{im_i}\mathbf{b}_{im_i} \mid x_1, \ldots, x_{m_i} \in \mathbb{N}\}, \tag{2}$$

where $\mathbf{b}_{ij} \in \mathbb{N}^k$, $1 \leq i \leq n,\ 0 \leq j \leq m_i$.

In [14–16] the solution (in the affirmative) of the CE Problem was given for sparse languages: Every bounded context-free language L_1 is commutatively equivalent to a regular language L_2. Moreover the language L_2 can be effectively constructed starting from an effective presentation of L_1 (e.g. a context-free grammar generating L_1). It is also shown that the CE Problem can be solved in the affirmative for the wider class of bounded semi-linear languages.

In contrast with the results mentioned above, the CE Problem remains open for the class of context-free languages of exponential growth.

A relevant fact in this context is that the generating series of a commutatively regular language L is always rational. This implies that the answer to the CE Problem is not affirmative in general. Indeed, there exist context-free languages whose generating series are algebraic but not rational. It is worth noting that Flajolet even provided remarkable examples of *linear unambiguous context-free languages* with a transcendental generating series [21, Theorem 3].

The study of the CE Problem has been further investigated by the authors in connection with languages of exponential growth generated by unambiguous non-expansive grammars [1] and unambiguous minimal linear grammars [11, 24], respectively. In particular conditions ensuring that such languages are commutatively regular have been provided [6].

It is worth noting that unambiguous minimal linear grammars give a generalization of the concept of unique-factorization code since they inherit several properties of this structure [8].

As a continuation of this work, we investigate the CE Problem with respect to the Kleene closure of languages. This operation is of interest for this study since it preserves the property of context-freeness of languages, while it does not preserve the property of boundedness: finite sets of non-commuting words are the simplest example of bounded languages whose Kleene closure is not bounded. It is also interesting to note that Dyck and semi-Dyck languages provide another natural class of monoids that are not commutatively regular. In view of the classical theorem by Chomsky and Schützenberger for the representation of context-free languages, such monoids can be considered very general. It is useful to observe that a Dyck monoid is a free monoid whose minimal set of generators, i.e. the set of Dyck prime words, is a context-free bifix code.

In contrast with the previous situation, the main result of this paper shows, up to a technical restriction, that the monoid generated by a bounded semi-linear code is commutatively regular. Precisely, the main contribution of the paper is the following.

Theorem 1 *Let L be a bounded semi-linear language and $L = \varphi(B)$ where B is the semi-linear set of Eq. (1) associated with L.*

Suppose that, for every $i = 1, \dots, n$ and every $j = 0, \dots, m_i$, the vector $\mathbf{b}_{ij}$ of Eq. (2) is such that its corresponding word $\varphi(\mathbf{b}_{ij})$ contains two distinct letters.

If L is a code, then there exists a regular code L' which is commutatively equivalent to L. Consequently, L^* is commutatively equivalent to $(L')^*$. Moreover, L' can be effectively constructed starting from an effective presentation of L.

It is worth noticing that the validity of Theorem 1 by no means depends upon the choice of the representation of the semi-linear set associated with the language. Indeed Proposition 1 will show that the restriction imposed on the image, under the Ginsburg map, of the generators of the semi-linear set defining the language, is equivalent to the existence of a real number $\rho < 1$ such that, for every $z \in L$ and all $a \in A$, one has $|z|_a < \rho|z|$, where $|z|_a$ is the number of occurrences of the letter a in z and $|z|$ is the length of z.

In order to prove Theorem 1 we use two arguments: the first concerns codes and equations of words and makes it possible to separate the languages that represent the simple sets B_i of the decomposition of B. The second one is a technique of an algebro-geometrical nature for the decomposition of semilinear sets into disjoint parallelepipeds.

As a surprising application of the tools mentioned above, we also prove that a bounded semi-linear language satisfying the hypotheses of Theorem 1 is commutatively equivalent to the union of a finite set and a regular code. The latter result still holds even if the language is not a code (Proposition 5).

The paper is organized as follows. In Section 2, preliminaries on bounded semi-linear languages and codes are presented. In Section 3, some results of combinatorial flavour used in the proof of Theorem 1, are provided. Section 4 is devoted to the proof of our main result, while Section 5 presents some concluding remarks.

Some results of this paper have been presented at WORDS 2019 [7].

2 Preliminaries

The aim of this section is to introduce some preliminary results on semi-linear sets and bounded context-free languages. We assume that the reader is familiar with the basic notions of context-free languages (see [3, 23] for a reference).

The free abelian monoid on k generators is identified with $\langle \mathbb{N}^k, + \rangle$ with the usual additive structure.

Definition 1 Let B be a subset of $\mathbb{N}^k$. The following definitions hold:

1. B is *linear* if there are $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m \in \mathbb{N}^k$ such that
$$B = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m \mid x_i \in \mathbb{N},\ 1 \leq i \leq m\}. \tag{3}$$
2. B is *simple* if there are $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m \in \mathbb{N}^k$ such that (3) holds true and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are linearly independent in $\mathbb{Q}$.
3. B is *semi-linear* if B is a finite union of linear sets in $\mathbb{N}^k$.
4. B is *semi-simple* if B is a finite disjoint union of simple sets in $\mathbb{N}^k$.

It is well known that a simple set B has a unique representation in the form (3) with $\mathbf{b}_1, \ldots, \mathbf{b}_m$ linearly independent in $\mathbb{Q}$. It will be called the *unambiguous representation* (or, briefly, *the representation*) of B. The vector $\mathbf{b}_0$ is called *constant* and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are called *generators* of the representation, respectively. The number m is called the *dimension* of B. Obviously one has $m \leq k$ and the dimension of a singleton is 0.

The following is an important characterization of semi-linear sets.

Theorem 2 (Eilenberg and Schützenberger [20]) *Let B be a subset of $\mathbb{N}^k$. Then B is semi-linear if and only if B is semi-simple.*

It is worth to remark that Theorem 2 was proved independently by Ito in [33].

Let $A = \{a_1, \ldots, a_t\}$ be an alphabet of t letters and let A^* be the free monoid generated by A. The empty word of A^* is denoted by ε and the set $A^* \setminus \{\varepsilon\}$ is denoted by A^+. The length of every word u is denoted by $|u|$.

For every $a \in A$ and $u \in A^*$, the number of occurrences of a in u will be denoted by $|u|_a$. We let $\psi \colon A^* \to \mathbb{N}^t$ denote the *Parikh map* over A, defined, for each $u \in A^*$, as
$$\psi(u) = (|u|_{a_1}, |u|_{a_2}, \ldots, |u|_{a_t}).$$
The map ψ is a monoid epi-morphism from A^* onto $\mathbb{N}^t$. The kernel congruence of ψ is called the commutative equivalence of A^*.

Given words $u, v \in A^*$, u is called a *factor of* v if $v = pus$, for some $p, s \in A^*$. In particular, if in the latter, $p = \varepsilon$ (resp., $s = \varepsilon$), then u is called a *prefix of* v (resp., *a suffix of* v). The set of factors of u is denoted $\mathrm{Fact}(u)$. For an arbitrary subset L of words of A^*, $\mathrm{Fact}(L) = \bigcup_{u \in L} \mathrm{Fact}(u)$.

A subset L of words of A^+ is said to be a *code (over A)* if every word of L^+ admits a unique factorization in term of words of L. A set L over the alphabet A is said to be a *prefix set* if $LA^+ \cap L = \emptyset$.

Let L_1, L_2 be two languages over A. We say that L_1 is *commutatively equivalent to* L_2 if there is a bijection $f \colon L_1 \to L_2$ such that, for every $u \in L_1$, one has $\psi(u) = \psi(f(u))$. In the sequel, by simplicity, if L_1 and L_2 are so, we write $L_1 \sim L_2$. If $u_1, \ldots, u_k$ are k words of A^+, the *Ginsburg map*

$$\varphi \colon \mathbb{N}^k \to u_1^* \cdots u_k^*, \tag{4}$$

is the map defined, for each tuple $(\ell_1, \ldots, \ell_k) \in \mathbb{N}^k$, as:

$$\varphi(\ell_1, \ldots, \ell_k) = u_1^{\ell_1} \cdots u_k^{\ell_k}.$$

By a remarkable observation of [26, Lemma 2.1], the function obtained as the restriction of φ on the semi-linear set of Eq. (1) is injective. Moreover a previously mentioned theorem by Ginsburg and Spanier [22] provides a fundamental tool to represent, in terms of semi-linear sets, the bounded context-free languages. For our purposes, these two results can be stated as follows.

Theorem 3 *Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded semi-linear language. Then there exists a semi-simple set B of $\mathbb{N}^k$ such that $\varphi(B) = L$ and φ is injective on B. Moreover, B can be effectively constructed.*

In particular, the condition above holds for bounded context-free languages.

We end this section by stating two properties that will be used in the next sections, even without any explicit mention. Both are simple corollaries of the corresponding definitions introduced above.

Lemma 1 *For all $a \in A$, the function mapping any $\mathbf{v} \in \mathbb{N}^k$ into $|\varphi(\mathbf{v})|_a$ is a linear function of $\mathbb{N}^k$ into $\mathbb{N}$, as well as the function mapping any $\mathbf{v} \in \mathbb{N}^k$ into $|\varphi(\mathbf{v})|$. Moreover, composing the Ginsburg map and the Parikh map, one obtains a linear function of $\mathbb{N}^k$ into $\mathbb{N}^t$.*

Proof Let $\mathbf{v} = (\ell_1, \ldots, \ell_k)$ be an element of $\mathbb{N}^k$ and $a \in A$. By definition, one has $\varphi(\mathbf{v}) = u_1^{\ell_1} \cdots u_k^{\ell_k}$ and, therefore,

$$|\varphi(\mathbf{v})|_a = |u_1^{\ell_1} \cdots u_k^{\ell_k}|_a = \sum_{i=1}^{k} \ell_i |u_i|_a.$$

Hence, $|\varphi(\mathbf{v})|_a$ is the scalar product of the vector $\mathbf{v}$ by $(|u_1|_a, \ldots, |u_k|_a)$. We conclude that the function mapping any $\mathbf{v} \in \mathbb{N}^k$ into $|\varphi(\mathbf{v})|_a$ is linear.

The linearity of the other two functions considered in the statement follows from the result above and the equations

$$|\varphi(\mathbf{v})| = \sum_{i=1}^{t} |\varphi(\mathbf{v})|_{a_i}, \quad \psi(\varphi(\mathbf{v})) = (|\varphi(\mathbf{v})|_{a_1}, \ldots, |\varphi(\mathbf{v})|_{a_t}), \quad \mathbf{v} \in \mathbb{N}^k.$$

This completes the proof. □

Lemma 2 *Let $(Y_i)_{i \in I}$ and $(Z_i)_{i \in I}$ be partitions of the languages Y and Z, respectively. If one has $Y_i \sim Z_i$ for all $i \in I$, then the languages Y and Z are commutatively equivalent.*

Proof The bijections $f_i \colon Y_i \to Z_i$, $i \in I$, preserving the Parikh vectors piecewise define a bijection $f \colon Y \to Z$ with the same property. □

3 Some results of combinatorics on words

In this section, we provide properties of the combinatorial structure of the bounded semi-linear languages satisfying the hypotheses of Theorem 1. These properties will be used to construct the solution of our main problem.

3.1 On some combinatorial properties of bounded semi-linear languages

We start to establish a useful characterization of the languages satisfying the hypotheses of Theorem 1.

Proposition 1 *Let L be a bounded semi-linear language and*

$$B = \bigcup_{i=1}^{n} \{\mathbf{b}_{i0} + x_1\mathbf{b}_{i1} + \cdots + x_{m_i}\mathbf{b}_{im_i} \mid x_i \in \mathbb{N},\ 1 \leq i \leq m_i\} \tag{5}$$

be a semi-linear set such that $L = \varphi(B)$, $\mathbf{b}_{ij} \in \mathbb{N}^k$, $1 \leq i \leq n$, $0 \leq j \leq m_i$, and $\mathbf{b}_{ij} \neq \mathbf{0}$, $1 \leq i \leq n$, $0 \leq j \leq m_i$. The following two conditions are equivalent:

(i) for every $i = 1, \ldots, n$ and $j = 0, \ldots, m_i$, the word $\varphi(\mathbf{b}_{ij})$ contains two distinct letters,

(ii) there exists a real number $\rho < 1$ such that, for every $z \in L$ and all $a \in A$, one has

$$|z|_a < \rho|z|.$$

Proof Let us verify that (i) implies (ii). If (i) holds true, then for all $a \in A$, $i = 1, \ldots n$, and $j = 0, \ldots, m_i$ one has $|\varphi(\mathbf{b}_{ij})|_a/|\varphi(\mathbf{b}_{ij})| < 1$. Thus, there is $\rho < 1$ such that

$$\frac{|\varphi(\mathbf{b}_{ij})|_a}{|\varphi(\mathbf{b}_{ij})|} < \rho,$$

for $a \in A$, $1 \leq i \leq n$, $0 \leq j \leq m_i$. A word $z \in L$ can be written as

$$z = \varphi(\mathbf{b}_{i0} + x_1\mathbf{b}_{i1} + \cdots + x_{m_i}\mathbf{b}_{im_i})$$

with $1 \leq i \leq n$, $x_1, \ldots, x_{m_i} \in \mathbb{N}$. Thus, in view of Lemma 1, for all $a \in A$,

$$|z|_a = |\varphi(\mathbf{b}_{i0})|_a + \sum_{j=1}^{m_i} x_j|\varphi(\mathbf{b}_{ij})|_a < \rho|\varphi(\mathbf{b}_{i0})| + \sum_{j=1}^{m_i} x_j\rho|\varphi(\mathbf{b}_{ij})| = \rho|z|.$$

Now, let us verify that (ii) implies (i). Taking $z = \varphi(\mathbf{b}_{i0})$, $1 \leq i \leq n$, Condition (ii) gives $|\varphi(\mathbf{b}_{i0})|_a < \rho|\varphi(\mathbf{b}_{i0})| < |\varphi(\mathbf{b}_{i0})|$. This inequality shows that $\varphi(\mathbf{b}_{i0})$ is not a power of a.

Now, take $z = \varphi(\mathbf{b}_{i0} + q\mathbf{b}_{ij})$, $1 \leq i \leq n$, $1 \leq j \leq m_i$, $q > 0$. Then, by linearity from Lemma 1, Condition (ii) becomes

$$|\varphi(\mathbf{b}_{i0})|_a + q|\varphi(\mathbf{b}_{ij})|_a < \rho|\varphi(\mathbf{b}_{i0})| + q\rho|\varphi(\mathbf{b}_{ij})|.$$

Dividing both sides of this equation by q and letting q tend to $+\infty$, one obtains $|\varphi(\mathbf{b}_{ij})|_a \leq \rho|\varphi(\mathbf{b}_{ij})| < |\varphi(\mathbf{b}_{ij})|$, so that $\varphi(\mathbf{b}_{ij})$ is not a power of a. By the arbitrariness of a, we conclude that all words $\varphi(\mathbf{b}_{ij})$ contain at least two distinct letters. $\square$

Remark 1 Clearly, any semi-linear set has several representations in the form (5). However, Proposition 1 has the following useful consequence: the property that for all the vectors $\mathbf{b}_{ij}$ involved in (5), the word $\varphi(\mathbf{b}_{ij})$ contains at least two distinct letters does not depend upon the choice of such a representation.

We conclude this section with the following lemma. It implies that, in order to prove Theorem 1, it suffices to prove that every bounded semi-linear language satisfying the Condition (i) of the above proposition is commutatively equivalent to a regular code.

Lemma 3 *Let Y and Z be two commutatively equivalent codes. Then Y^* and Z^* are commutatively equivalent.*

Proof Let $f\colon Y \to Z$ be the bijection such that $\psi(f(y)) = \psi(y)$ for all $y \in Y$. We shall prove that this map can be uniquely extended to an isomorphism $g\colon Y^* \to Z^*$, preserving Parikh vectors. Indeed, it is sufficient to set

$$g(u) = \begin{cases} \varepsilon & \text{if } u = \varepsilon \\ f(y_1)\cdots f(y_n) & \text{if } u = y_1 \cdots y_n,\ y_i \in Y,\ 1 \leq i \leq n. \end{cases} \tag{6}$$

As Y is a code, the definition is consistent since any word $u \in Y^+$ admits a unique factorization $u = y_1 \cdots y_n$, with $y_i \in Y$, $1 \leq i \leq n$. Moreover, taking into account that f is a bijection and Z is a code, any word $v \in Z^+$ admits a unique factorization $v = f(y_1) \cdots f(y_n)$, with $y_i \in Y$, $1 \leq i \leq n$. This ensure that g is a bijection. Notice that if $u = y_1 \cdots y_n$, with $y_i \in Y$, $1 \leq i \leq n$, then

$$\psi(f(u)) = \psi(f(y_1)\cdots f(y_n)) = \sum_{i=1}^{n} \psi(f(y_i)) = \sum_{i=1}^{n} \psi(y_i) = \psi(u).$$

We conclude that $Y^* \sim Z^*$. $\square$

3.2 On bounded languages and codes

We will now prove some results relating bounded languages with codes.

In the sequel of this section, we assume that L is a bounded language on the alphabet A satisfying Condition (ii) of Proposition 1:

> *There exists a real number $\rho < 1$ such that for all $v \in L$, and all $a \in A$, $|v|_a < \rho|v|$.*

As L is bounded, one has $L \subseteq u_1^* \cdots u_k^*$ where k is a fixed positive integer and $u_1, \ldots, u_k \in A^*$. Our first result shows how to obtain some words which are not factors of L^*.

Lemma 4 *There exists a positive integer γ such that for every pair of distinct letters $a, b \in A$,*

$$(a^\gamma b)^{2k+1} \notin \mathrm{Fact}(L^*). \tag{7}$$

Proof We take an integer γ such that $\gamma \geq (2k+1)/(1-\rho)$ and $\gamma \geq |u_i|$, $1 \leq i \leq m$. By an argument introduced in [14, Lemma 9] one can verify that $(a^\gamma b)^k \notin \mathrm{Fact}(L)$.

Now, by contradiction, suppose $(a^\gamma b)^{2k+1} \in \mathrm{Fact}(L^*)$. Taking into account that $(a^\gamma b)^k$ cannot be a factor of L, we can factorize

$$(a^\gamma b)^{2k+1} = s_1 v s_2,$$

with $v \in L^*$, a suffix s_1 and a prefix s_2 of some words of L and, moreover, $|s_i| < |(a^\gamma b)^k|$, $i = 1, 2$. Hence,

$$|v| > |(a^\gamma b)^{2k+1}| - 2|(a^\gamma b)^k| > \gamma. \tag{8}$$

On the other side, one has $|v|_a < \rho|v|$ and $|v|_b \leq 2k+1$, so that $|v| < \rho|v| + 2k+1$. From this inequality, one easily derives $|v| < (2k+1)/(1-\rho) \leq \gamma$, which contradicts (8). □

Let γ be as in the previous lemma. We set

$$H = \{a(a^\gamma b)^{2k+2} v \mid a, b \in A,\ a \neq b,\ v \in A^*\}. \tag{9}$$

Notice that H is a right ideal of A^*, i.e., $HA^* = H$. A useful combinatorial property of the set H is that the product L^*H is unambiguous as shown by the following.

Lemma 5 *Let $x, x' \in L^*$ and $s, s' \in H$. If $xs = x's'$, then $x = x'$ and $s = s'$.*

Proof By the definition of H, there are letters $a, b, c, d \in A$ and words $v, v' \in A^*$ such that $s = a(a^\gamma b)^{2k+2} v$, $s' = c(c^\gamma d)^{2k+2} v'$, $a \neq b$ and $c \neq d$. Thus, the equation $xs = x's'$ can be rewritten as

$$xa(a^\gamma b)^{2k+2} v = x'c(c^\gamma d)^{2k+2} v'.$$

By contradiction suppose $|x| > |x'|$. Taking into account that by Lemma 4 the word $(c^\gamma d)^{2k+1}$ cannot be a factor of x, one obtains

$$x = x'c(c^\gamma d)^p c^q, \quad a^{\gamma+1} = c^{\gamma-q} d c^q, \quad b(a^\gamma b)^{2k+1} v = c^{\gamma-q} d (c^\gamma d)^{2k-p} v',$$

with some $0 \leq p \leq 2k$, $0 \leq q \leq \gamma$. From the second of the equations above, one derives that $c = a$ and $d = a$, which yields a contradiction, as $c \neq d$.

A similar contradiction is obtained if $|x'| > |x|$. We conclude that $|x| = |x'|$ and, consequently, $x = x'$ and $s = s'$. □

Consider a partition

$$L = \bigcup_{i=1}^{n} L_i, \tag{10}$$

of the set L and a collection of prefix codes

$$W_i \subseteq H, \tag{11}$$

$i = 1, \ldots, n$.

Proposition 2 *If L is a code, then $Y = \bigcup_{i=1}^{n} L_i W_i^*$ is a code.*

Proof By contradiction, suppose that Y is not a code. Then, there is a relation

$$y_1 y_2 \cdots y_m = y_1' y_2' \cdots y_{m'}' \tag{12}$$

with $y_1, \ldots y_m, y_1' \ldots y_{m'}' \in Y$ and $y_1 \neq y_1'$. Let us verify that $y_1, y_1' \notin L$ and there are an index i and words z, w, w' such that

$$y_1 = zw, \quad y_1' = zw', \quad z \in L_i, \quad w, w' \in W_i^+. \tag{13}$$

We assume that in each side of (12) appears some element of $Y \backslash L = \bigcup_{i=1}^{n} L_i W_i^+$. Indeed, if it is not the case, it is sufficient to append such a word to both sides of the equation.

Thus, let p, p' be the least indices such that $y_p, y_{p'}' \in Y \setminus L$. Then there are $i, i' \in \{1, \ldots, n\}$, $z \in L_i$, $w \in W_i^+$, $z' \in L_{i'}$ and $w' \in W_{i'}^+$ such that $y_p = zw$ and $y_{p'}' = z'w'$. In view of (12) one has

$$y_1 \cdots y_{p-1} z \; w y_{p+1} \cdots y_m = y_1' \cdots y_{p'-1}' z' \; w' y_{p'+1}' \cdots y_{m'}'.$$

Notice that, by the minimality of p and p' one has $y_1 \cdots y_{p-1} z,\ y_1' \cdots y_{p'-1}' z' \in L^*$. On the other side, since $W_i, W_{i'} \subseteq H$ and H is a right ideal, one has $w y_{p+1} \cdots y_m, w' y_{p'+1}' \cdots y_{m'}' \in H^+ Y^* = H$. Thus, by Lemma 5 one gets

$$y_1 \cdots y_{p-1} z = y_1' \cdots y_{p'-1}' z'.$$

Taking into account that L is a code and $y_1 \neq y_1'$, one easily derives that $p = p' = 1$ and $z = z'$. The former identity shows that $y_1, y_1' \notin L$ and the latter implies, in particular, $i = i'$, so that $w, w' \in W_i^+$. Thus, (13) is established.

With no loss of generality, we assume that $|y_1| \geq |y_1'|$. Thus, from (12) one has $y_1 = y_1' s$, $s y_2 \cdots y_m = y_2' \cdots y_{m'}'$, for some $s \in A^*$, and also

$$z s y_2 \cdots y_m = z y_2' \cdots y_{m'}'. \tag{14}$$

Moreover, $s \neq \varepsilon$, as $y_1 \neq y_1'$. Let us verify that $zs \in Y$. Indeed, from (13) one easily derives that $w = w's$. Taking into account that $w, w' \in W_i^+$ and W_i is a prefix code, one obtains $s \in W_i^+$ and, therefore, $zs \in L_i W_i^+ \subseteq Y$.

Thus, (14) is formally equal to (12), with $y_1 = zs$ and $y_1' = z$. However, in this case, one has $y_1' = z \in L$. In view of this contradiction, we conclude that Y is necessarily a code. $\square$

Clearly, if one removes the hypothesis that L is a code, the previous proposition does not hold true, since $L \subseteq Y$. However, the next proposition shows that the set $Y \setminus L$ is a code also in this case.

Proposition 3 *The set $Z = \bigcup_{i=1}^{n} L_i W_i^+$ is a code.*

Proof The proof is very similar to that of Proposition 2. Thus, we limit ourselves to outline it, focusing on the main differences.

By contradiction, suppose that Z is not a code. Then, there are $y_1, \ldots, y_m$, $y'_1, \ldots, y'_{m'} \in Z$ such that $y_1 \neq y'_1$ and the relation (12) is satisfied. With no loss of generality, we assume that $|y_1| \geq |y'_1|$. Actually, one has $|y_1| > |y'_1|$ as, in the case that $|y_1| = |y'_1|$, from (12) one would derive $y_1 = y'_1$. Since $y_1, y'_1 \in Z$, there are $i, i' \in \{1, \ldots, n\}$, $z \in L_i$, $w \in W_i^+$, $z' \in L_{i'}$ and $w' \in W_{i'}^+$ such that $y_1 = zw$ and $y'_1 = z'w'$. In view of (12) one has

$$zwy_2 \cdots y_m = z'w'y'_2 \cdots y'_{m'}.$$

Proceeding as in the proof of Proposition 2, one obtains that

$$z = z', \quad i = i', \quad w = w's, \quad sy_2 \cdots y_m = y'_2 \cdots y'_{m'},$$

with $s \in W_i^+$. Since $y'_2 \in Z$, there are $i'' \in \{1, \ldots, n\}$, $z'' \in L_{i''}$ and $w'' \in W_{i''}^+$ such that $y'_2 = z''w''$. Thus, in view of the previous equations, one has

$$sy_2 \cdots y_m = z''w''y'_3 \cdots y'_{m'}.$$

One has $z'' \in L$ and $sy_2 \cdots y_m, w''y'_3 \cdots y'_{m'} \in H$, so that by Lemma 5 one derives $z'' = \varepsilon$, and, consequently, $\varepsilon \in L$. This yields a contradiction, because $|u|_a < \rho|u|$ for all $u \in L$ and $a \in A$, while $|\varepsilon|_a = \rho|\varepsilon| = 0$. □

The following lemma, which will be useful in the sequel, shows that the code Y considered in Proposition 2 is partitioned by the sets $L_i W_i^*$ and, similarly, the set Z considered in Proposition 3 is partitioned by the sets $L_i W_i^+$.

Lemma 6 *The sets $L_i W_i^*$, $1 \leq i \leq n$, are pairwise disjoint.*

Proof A word $y \in L_i W_i^* \cap L_j W_j^*$, $1 \leq i, j \leq n$ can be factorized as

$$y = z_i w_i = z_j w_j, \quad \text{with } z_i \in L_i,\ w_i \in W_i^*,\ z_j \in L_j,\ w_j \in W_j^*.$$

Let us verify that $z_i = z_j$. Indeed, this is trivial if both w_i and w_j are empty and is a straightforward consequence of Lemma 5 if both w_i and w_j are non-empty, as $W_i^+, W_j^+ \subseteq H$. Finally, if only one of the words w_i and w_j, say w_i, is non-empty, then one has $yw_i = z_i w_i w_i = z_j w_i$ and the conclusion follows again from Lemma 5. Thus, in any case $z_i = z_j \in L_i \cap L_j$ and, since the sets L_i form a partition, one has $i = j$. □

The following lemma provides a tool to construct prefix codes included in H. We notice that it holds in the general hypothesis that H is of the form (9) with k and γ any positive integers.

Lemma 7 *Let $z_1, \dots, z_m$ be words (not necessarily distinct) and suppose that each of them contains at least two distinct letters. For all $N \geq (2k+2)\gamma + m$, there exist m distinct words $w_1, \dots, w_m \in H$ such that $W = \{w_1, \dots, w_m\}$ is a prefix code and*

$$\psi(w_i) = N\psi(z_i), \quad i = 1, \dots, m. \tag{15}$$

PROOF We first define the required words and then prove that they fulfill the desired property. For every $i = 1, \dots, m$, we define the word w_i as follows. Let a and b be two distinct letters occurring in z_i. Then one has

$$|z_i^N|_a, |z_i^N|_b \geq N \geq (2k+2)\gamma + i \geq 2k+3.$$

Thus, we can consider a word $\widehat{z}_i$ obtained from z_i^N by deleting $(2k+2)\gamma + i$ occurrences of a and $2k+3$ occurrences of b. We set

$$w_i = a(a^\gamma b)^{2k+2} a^{i-1} b\, \widehat{z}_i.$$

By construction, one has $w_i \in H$ and $\psi(w_i) = \psi(z_i^N) = N\psi(z_i)$, $i = 1, \dots, m$.

Let us verify that $W = \{w_1, \dots, w_m\}$ is a prefix code. Let $1 \leq i < j \leq m$. Then, there are letters $a, b, c, d \in A$ such that $a(a^\gamma b)^{2k+2}a^{i-1}b$ is a prefix of w_i, $c(c^\gamma d)^{2k+2}c^{j-1}d$ is a prefix of w_j, with $a \neq b$ and $c \neq d$. Clearly, the word $a(a^\gamma b)^{2k+2}a^{i-1}b$ cannot be a prefix of $c(c^\gamma d)^{2k+2}c^{j-1}d$. Indeed, otherwise, one would have $a = c$, $b = d$ and $i = j$. We derive that neither w_i is a prefix of w_j, nor w_j is a prefix of w_i. Thus the words w_i, $i = 1, \dots, m$ are pairwise distinct and W is a prefix code. $\square$

4 The construction of the regular language

Partitions of semi-linear sets have been one of the main tools for the solution of the CE problem for sparse context-free languages [14]. Here, we present a very simplified version of the decomposition considered in [14], which, however, in this case is sufficient for our purpose.

Proposition 4 *Let $B \subseteq \mathbb{N}^k$ be a semi-simple set and N be a positive integer. Then B is a disjoint union of finitely many simple sets of the form*

$$\Big\{\mathbf{b} + N\sum_{i=1}^{m} x_i \mathbf{b}_i \;\Big|\; x_i \in \mathbb{N},\ 1 \leq i \leq m\Big\}, \tag{16}$$

with $0 \leq m \leq k$, $\mathbf{b}, \mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{N}^k$ and $\mathbf{b}_1, \dots, \mathbf{b}_m$ linearly independent.

PROOF Since, by Definition 1, any semi-simple set is a finite disjoint union of simple sets, we may reduce ourselves, with no loss of generality, to the case that B is a simple set. In such a case, one has

$$B = \Big\{\mathbf{b}_0 + \sum_{i=1}^{m} x_i \mathbf{b}_i \;\Big|\; x_i \in \mathbb{N},\ 1 \leq i \leq m\Big\}$$

with $0 \le m \le k$, $\mathbf{b}_0, \ldots, \mathbf{b}_m \in \mathbb{N}^k$ and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ linearly independent. Thus, any vector $\mathbf{v} \in B$ can be uniquely written as

$$\mathbf{v} = \mathbf{b}_0 + \sum_{i=1}^{m} (r_i + N x_i)\mathbf{b}_i = \mathbf{b}_0 + \sum_{i=1}^{m} r_i \mathbf{b}_i + N \sum_{i=1}^{m} x_i \mathbf{b}_i, \tag{17}$$

with $0 \le r_i < N$ and $x_i \ge 0$, $i = 1, \ldots, m$. In other terms, any vector of B can be uniquely written as $\mathbf{b} + N \sum_{i=1}^{m} x_i \mathbf{b}_i$ with

$$\mathbf{b} \in R = \Big\{ \mathbf{b}_0 + \sum_{i=1}^{m} r_i \mathbf{b}_i \;\Big|\; 0 \le r_i < N,\ i = 1, \ldots, m \Big\}.$$

This ensures that the set B is the disjoint union of the sets (16) with $\mathbf{b} \in R$. Since R is finite, the statement is proved. □

In the sequel of this section, we make the following assumption:

> *The set $L \subseteq u_1^* \cdots u_k^*$ is a bounded semi-linear language satisfying the hypotheses of Theorem 1.*

Thus, by Proposition 1, all results of Section 3.2 apply to L. In particular, by Lemma 4, there is an integer $\gamma > 0$ such that $(a^\gamma b)^{2k+1} \notin \mathrm{Fact}(L^*)$ for all $a, b \in A$ and one can consider the set H defined by (9).

In view of Theorem 3, one has $L = \varphi(B)$ for some semi-simple set $B \subseteq \mathbb{N}^k$ and φ is injective on B. Let $N \ge (2k+2)\gamma + k$. In view of Proposition 4, we assume that

$$B = \bigcup_{j=1}^{n} B_j,$$

where the sets $B_1, \ldots, B_n$ are pairwise disjoint simple sets of the form (16). We denote $L_j = \varphi(B_j)$, $j = 1, \ldots, n$.

Now, we construct regular languages which are commutatively equivalent to the languages L_j.

Lemma 8 *For every $j = 1, \ldots, n$, there are a regular language L'_j and a prefix code $W_j \subseteq H$ such that $L_j \sim L'_j$ and $L'_j \subseteq L_j W_j^*$.*

PROOF The set B_j has the form (16) that is,

$$B_j = \Big\{ \mathbf{b} + N \sum_{i=1}^{m} x_i \mathbf{b}_i \;\Big|\; x_i \in \mathbb{N},\ 1 \le i \le m \Big\},$$

with $0 \le m \le k$, $\mathbf{b}, \mathbf{b}_1, \ldots, \mathbf{b}_m \in \mathbb{N}^k$ and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ linearly independent. By Proposition 1 (see also Remark 1) the words $\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_m)$ contain two distinct letters and, therefore, by Lemma 7, there exist m distinct words $w_1, \ldots, w_m \in H$ such that $W_j = \{w_1, \ldots, w_m\}$ is a prefix code and

$$\psi(w_i) = N\psi(\varphi(\mathbf{b}_i)), \quad i = 1, \ldots, m. \tag{18}$$

We set $L'_j = \varphi(\mathbf{b})\, w_1^* \cdots w_m^*$. Thus, L'_j is regular and $L'_j \subseteq L_j W_j^*$. Hence, in order to complete the proof, it suffices to verify that $L_j \sim L'_j$.

Since φ is injective on B and the set B_j is simple, any word of $L_j = \varphi(B_j)$ can be uniquely written as $\varphi\big(\mathbf{b} + N\sum_{i=1}^m x_i \mathbf{b}_i\big)$ with $x_1, \dots, x_m \in \mathbb{N}$. On the other side, as W_j is a code, any word of $L'_j = \varphi(\mathbf{b})\, w_1^* \cdots w_m^*$ can be uniquely written as $\varphi(\mathbf{b})\, w_1^{x_1} \cdots w_m^{x_m}$ with $x_1, \dots, x_m \in \mathbb{N}$. Thus, there is a bijection $f\colon L_j \to L'_j$ mapping the generic word $\varphi\big(\mathbf{b} + N\sum_{i=1}^m x_i \mathbf{b}_i\big)$ of L_j onto $\varphi(\mathbf{b})\, w_1^{x_1} \cdots w_m^{x_m}$. Moreover, by Lemma 1, one has

$$\psi\Big(\varphi\Big(\mathbf{b} + N\sum_{i=1}^m x_i \mathbf{b}_i\Big)\Big) = \psi(\varphi(\mathbf{b})) + \sum_{i=1}^m x_i N \psi(\varphi(\mathbf{b}_i)),$$

while

$$\psi(\varphi(\mathbf{b})\, w_1^{x_1} \cdots w_m^{x_m}) = \psi(\varphi(\mathbf{b})) + \sum_{i=1}^m x_i \psi(w_i).$$

Hence, in view of (18), the left-hand sides of the two previous equations are equal. This proves that the bijection f preserves the Parikh vectors and, therefore, L_j and L'_j are commutatively equivalent. Thus, the proof is completed. □

We are now ready to prove our main theorem.

Proof of Theorem 1 Since B is the disjoint union of the sets B_j, $1 \le j \le n$, and φ is injective on B, L has the partition

$$L = \bigcup_{j=1}^n L_j.$$

Let L'_j and W_j, $1 \le j \le n$, be the languages determined by the previous lemma. By Proposition 1 and Lemma 6, the languages $L_j W_j^*$ and, consequently, their subsets L'_j, are pairwise disjoint. Thus, taking into account that $L_j \sim L'_j$, $1 \le j \le n$, by Lemma 2, one obtains that L is commutatively equivalent to the language

$$L' = \bigcup_{j=1}^n L'_j,$$

which is regular since all the languages L'_j are regular.

Let us verify that if L is a code, then L' is a code, too. Indeed, since $L'_j \subseteq L_j W_j^*$, $1 \le j \le n$, one has $L' \subseteq \bigcup_{j=1}^n L_j W_j^*$. By Proposition 1 and Proposition 2, this latter set is a code as well as its subset L'.

In view of Lemma 3, we conclude that L^* and L'^* are commutatively equivalent. This shows the second assertion in Theorem 1.

Finally we observe that, since all the steps of the proof of Theorem 1 are effective, we get an explicit construction of the language L'. This completes the proof. □

The following example clarifies some of the basic ideas underlying the proof of Theorem 1.

Example 1 Let L be the bounded semi-linear language over the alphabet $A = \{a, b\}$ defined as

$$L = \{a^m b^n a^n b^h \mid n > m > h > 0\}.$$

The Ginsburg map

$$\varphi : \mathbb{N}^4 \to a^* b^* a^* b^*$$

is defined as

$$\varphi(x_1, x_2, x_3, x_4) = a^{x_1} b^{x_2} a^{x_3} b^{x_4}, \quad x_1, x_2, x_3, x_4 \in \mathbb{N}.$$

We have $L = \varphi(B)$, where B is the simple subset of $\mathbb{N}^4$

$$B = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + x_3\mathbf{b}_3 \mid x_1, x_2, x_3 \geq 0\},$$

with $\mathbf{b}_0 = (2, 3, 3, 1)$, $\mathbf{b}_1 = (1, 1, 1, 1)$, $\mathbf{b}_2 = (1, 1, 1, 0)$, $\mathbf{b}_3 = (0, 1, 1, 0)$. It is easily verified that L is a code. Moreover $\varphi(\mathbf{b}_0) = a^2b^3a^3b, \varphi(\mathbf{b}_1) = abab, \varphi(\mathbf{b}_2) = aba, \varphi(\mathbf{b}_3) = ba$, so that the hypotheses of Theorem 1 are satisfied.

The words $(ab)^2$ and $(ba)^2$ are not factors of L^*. Hence, we may take

$$H = \{a(ab)^3, b(ba)^3\}A^*.$$

Moreover, in order to carry out the construction in the proof of Lemma 7, we need $N \geq |a(ab)^3aab|_a, |a(ab)^3aab|_b$, so that we may take $N = 6$.

With such a choice for N, the set B will be partitioned in the simple sets

$$\{\mathbf{b} + 6x_1\mathbf{b}_1 + 6x_2\mathbf{b}_2 + 6x_3\mathbf{b}_3 \mid x_1, x_2, x_3 \geq 0\} \tag{19}$$

with $\mathbf{b} = \mathbf{b}_0 + r_1\mathbf{b}_1 + r_2\mathbf{b}_2 + r_3\mathbf{b}_3$, $r_1, r_2, r_3 = 0, \ldots, 5$. Set $h_0 = 1 + r_1$, $m_0 = 2 + r_1 + r_2$, $n_0 = 3 + r_1 + r_2 + r_3$, $h = x_1$, $m = x_1 + x_2$, $n = x_1 + x_2 + x_3$. Then the sets (19) can be rewritten as

$$\{(m_0 + 6m, n_0 + 6n, n_0 + 6n, h_0 + 6h) \mid n \geq m \geq h \geq 0\},$$

with $h_0 = 1, \ldots, 6$, $m_0 = h_0 + 1, \ldots, h_0 + 6$, $n_0 = m_0 + 1, \ldots, m_0 + 6$. Consequently the language L is partitioned in the sets

$$\{a^{m_0+6m} b^{n_0+6n} a^{n_0+6n} b^{h_0+6h} \mid n \geq m \geq h \geq 0\}. \tag{20}$$

The words

$$w_1 = a(ab)^3ba^8b^8, \quad w_2 = a(ab)^3aba^7b^2, \quad w_3 = a(ab)^3aabb^2$$

belong to H, satisfy the identities $\psi(w_i) = 6\psi(\varphi(\mathbf{b}_i))$, $i = 1, 2, 3$, and the set $W = \{w_1, w_2, w_3\}$ is a prefix code. Thus, the set (20) is commutatively equivalent to the regular code $a^{m_0}b^{n_0}a^{n_0}b^{h_0}w_1^*w_2^*w_3^*$ and, therefore, L is commutatively equivalent to the regular code

$$\bigcup_{\substack{h_0=1,\ldots,6 \\ m_0=h_0+1,\ldots,h_0+6 \\ n_0=m_0+1,\ldots,m_0+6}} a^{m_0}b^{n_0}a^{n_0}b^{h_0}\, w_1^*\, w_2^*\, w_3^*.$$

We conclude this section establishing a surprising application of Proposition 3.

Proposition 5 *Let L be a bounded semi-linear language which satisfies the equivalent conditions (i) and (ii) of Proposition 1. Then L is commutatively equivalent to the union of a finite set and a regular code.*

Proof We have shown that L is commutatively equivalent to the union L' of the sets L'_j determined by Lemma 8. By inspecting the proof of that lemma, one can see that these sets have the form

$$L'_j = u_j w_1^* \cdots w_m^*$$

where $W_j = \{w_1, \ldots, w_m\}$ is a code and $u_j \in L_j$. Hence, $L'_j \setminus \{u_j\} \subseteq L_j W_j^+$ and therefore

$$L' \setminus \{u_1, \ldots, u_n\} \subseteq \bigcup_{j=1}^{n} L_j W_j^+.$$

The right hand side of the previous equation is a code by Proposition 3. Thus, L is commutatively equivalent to the union of the finite set $\{u_1, \ldots, u_n\}$ and the regular code $L' \setminus \{u_1, \ldots, u_n\}$. $\square$

5 Concluding Remarks

In this paper, we started the study of the Kleene closure of bounded semi-linear languages in connection with the CE Problem. We have given a positive answer to this problem in the case that the bounded semi-linear language is a code satisfying some restriction on the number of letters of the words that generate the language.

A continuation of this research could naturally concern Theorem 1. It would be interesting to extend the theorem by dropping the combinatorial restriction on the words of the language, or by relaxing the property of being a code.

The interest of the last question is related to the following aspects. As mentioned in the Introduction, the generating series of a commutatively regular language is always rational. Thus, in the search for criteria to characterize commutatively regular context-free languages, it would be interesting to investigate whether such languages coincide with those whose generating series are rational.

If the languages generated by all non-terminals of an unambiguous context-free grammar have a rational generating series, then the grammar belongs to the class of non-expansive context-free grammars [1]. For this class of grammars, the CE Problem admits a positive answer in several instances [6, 8]. Also, bounded context-free languages are generated by non-expansive grammars [35].

On the other side, it is interesting to observe that, among the rational operations on languages, the Kleene closure is the sole operation not preserving the property of boundedness. Thus, in view of the last remarks, the study of the conditions that guarantee that a monoid generated by a bounded semi-linear language is commutatively regular appears as a natural issue.

References

[1] G. Baron, W. Kuich, The characterization of nonexpansive grammars by rational power series, Inform. Control **48**, 109–118 (1981).

[2] M.-P. Béal, D. Perrin, On the generating sequences of regular languages on k symbols, J. ACM **50**, 955–980 (2003).

[3] J. Berstel, D. Perrin, C. Reutenauer, *Codes and Automata*, Encyclopedia of Mathematics and its Applications No. 129, Cambridge University Press, Cambridge (2009).

[4] L. Boasson, A. Restivo, Une caractérisation des langages algébriques bornés, RAIRO Theor. Inform. Appl. **11**, 203–205 (1977).

[5] M. R. Bridson, R. H. Gilman, Context-free languages of sub-exponential growth, J. Comput. System Sci. **64**, 308–310 (1999).

[6] A. Carpi, F. D'Alessandro, On the commutative equivalence of context-free languages, In: M. Hoshi, S. Seki (Eds), *Developments in Language Theory 2018*, Lecture Notes in Computer Science, 11088, Springer, Berlin, (2018), pp. 169–181.

[7] A. Carpi, F. D'Alessandro, On the commutative equivalence of bounded semi-linear codes, In: R. Mercas, D. Reidenbach (Eds), *Combinatorics on Words 12th International Conference WORDS 2019*, Lecture Notes in Computer Science, 11682, Springer, Berlin, (2019), pp. 119-132.

[8] A. Carpi, F. D'Alessandro, Coding by minimal linear grammars, Theoret. Comput. Sci., **834**, 14–25 (2020).

[9] A. Carpi, F. D'Alessandro, O. H. Ibarra, I. McQuillan, Relationships Between Bounded Languages, Counter Machines, Finite-Index Grammars, Ambiguity, and Commutative Regularity, Theoret. Comput. Sci., in press, available on line, https://doi.org/10.1016/j.tcs.2020.10.006 (2020).

[10] C. Choffrut, F. D'Alessandro, S. Varricchio, On bounded rational trace languages, Theory Comput. Syst. **46**, 351–369 (2010).

[11] N. Chomsky, M.-P. Schützenberger, The algebraic theory of context-free languages, In P. Braffort and D. Hirschberg (Eds), *Computer Programming and Formal Systems*, North Holland, Amsterdam, (1963), pp. 118–161.

[12] F. D'Alessandro, B. Intrigila, S. Varricchio, The Parikh counting functions of sparse context-free languages are quasi-polynomials, Theoret. Comput. Sci. **410**, 5158–5181 (2009).

[13] F. D'Alessandro, B. Intrigila, S. Varricchio, Quasi-polynomials, linear Diophantine equations and semi-linear sets, Theoret. Comput. Sci. **416**, 1–16 (2012).

[14] F. D'Alessandro, B. Intrigila, On the commutative equivalence of bounded context-free and regular languages: the code case, Theoret. Comput. Sci. **562**, 304–319 (2015).

[15] F. D'Alessandro, B. Intrigila, On the commutative equivalence of semi-linear sets of $\mathbb{N}^k$, Theoret. Comput. Sci. **562**, 476–495 (2015).

[16] F. D'Alessandro, B. Intrigila, On the commutative equivalence of bounded context-free and regular languages: the semi-linear case, Theoret. Comput. Sci. **572**, 1–24 (2015).

[17] F. D'Alessandro, O. H. Ibarra, I. McQuillan, On finite-index indexed grammars and their restrictions, In: F. Drewes, B. Truthe, and C. Martin Vides (Eds), *Language and Automata Theory and Applications 2017*, Lecture Notes in Computer Science, 10168, Springer, Heidelberg, (2017), pp. 287–298.

[18] F. D'Alessandro, O. H. Ibarra, I. McQuillan, On finite-index indexed grammars and their restrictions, Information and Computation, in press, available on line, https://doi.org/10.1016/j.ic.2020.104613, (2020).

[19] A. de Luca, S. Varricchio, *Finiteness and Regularity in Semigroups and Formal Languages*, Springer, Berlin (1999).

[20] S. Eilenberg, M.-P. Schützenberger, Rational sets in commutative monoids, J. Algebra **13**, 173–191 (1969).

[21] P. Flajolet, Analytic models and ambiguity of context-free languages, Theoret. Comput. Sci. **49**, 283–309 (1987).

[22] S. Ginsburg, E. H. Spanier, Semigroups, Presburger formulas, and languages, Pacific J. Math., **16**, 285–296 (1966).

[23] S. Ginsburg, *The mathematical theory of context-free languages*, Mc Graw-Hill, New York (1966).

[24] M. Gross, Inherent Ambiguity of Minimal Linear Grammars, Inform. Control, **7**, 366–368 (1964).

[25] J. Honkala, A decision method for Parikh slenderness of context-free languages, Discrete Appl. Math. **73**, 1–4 (1997).

[26] J. Honkala, Decision problems concerning thinness and slenderness of formal languages, Acta Inf. **35**, 625–636 (1998).

[27] J. Honkala, On Parikh slender context-free languages, Theoret. Comput. Sci. **255**, 667–677 (2001).

[28] O. H. Ibarra, B. Ravikumar, On sparseness, ambiguity and other decision problems for acceptors and transducers, In: B. Monien, G. Vidal-Naquet (Eds), *3rd Annual Symposium on Theoretical Aspects of Computer Science*, Lecture Notes in Computer Science, 210, Springer, Berlin, (1986), pp. 171–179.

[29] O. H. Ibarra, I. McQuillan, B. Ravikumar, On Counting Functions of Languages, In: M. Hoshi, S. Seki (Eds), *Developments in Language Theory 2018*, Lecture Notes in Computer Science, 11088, Springer, Berlin, (2018), pp. 429–440.

[30] L. Ilie, On generalized slenderness of context-free languages, *Words, semigroups, & transductions,* World Sci. Publ., River Edge, NJ, pp. 189–202 (2001).

[31] L. Ilie, G. Rozenberg, A. Salomaa, A characterization of poly-slender context-free languages, RAIRO Theor. Inform. Appl. **34**, 77–86 (2000).

[32] R. Incitti, The growth function of context-free languages, Theoret. Comput. Sci. **255**, 601–605 (2001).

[33] R. Ito, Every semi-linear set is a finite union of disjoint linear sets, J. Comput. Syst. Sci. **3**, 295–342 (1969).

[34] M. Latteux, G. Thierrin, On bounded context-free languages, Elektron. Inform. Verarb. u. Kybern. **20**, 3–8 (1984).

[35] M. Luker, A family of languages having only finite-index grammars. Inform. Control. **39**, 14–18 (1978).

[36] D. Perrin, M.-P. Schützenberger, Un problème élémentaire de la théorie de l'information, in *Théorie de l'Information*, Colloq. Internat. CNRS No. 276, (1977), pp. 249-260.

[37] A. Restivo, A characterization of bounded regular sets, In: H. Brakhage (Ed), *Automata Theory and Formal Languages*, *Lecture Notes in Computer Science*, 33, Springer, Berlin, (1975), pp. 239–244.

[38] V. I. Trofimov, Growth functions of some classes of languages, Kibernetika (Kiev), no. 6, i, 9–12, 149 (1981).