Coding by minimal linear grammars

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

July 31, 2019

Abstract

This paper concerns the structure and the properties of a special class of combinatorial systems called minimal linear grammars. The role of unambiguous minimal linear grammars is investigated in the framework of the information transmission and coding problem and some related issues.

Keywords: Coding and Information theory; Uniquely decipherable code; Maximal code; Bernoulli distribution; Context-free linear grammar; Commutative equivalence.

1 Introduction

The aim of this paper is to investigate the structure and the properties of a special class of combinatorial systems called *minimal linear grammars*. A minimal linear grammar is a context-free linear grammar endowed with a unique non-terminal symbol. In the sequel, the class of these objects will be simply denoted MLG. These grammars appear for the first time in the foundational paper by Chomsky and Schützenberger [13], where some problems on the ambiguity of these grammars are formulated and discussed (*cf.* Sec. 4.3, therein). Despite their rather simple structure, the languages generated by such grammars fulfill non trivial properties and are quite different from regular languages in several respects. In particular, as a solution of one of the above mentioned questions raised in [13], Greibach shows that the ambiguity problem is undecidable for

MLG [20]. As another related result, Gross shows the existence of languages that are ambiguous in the class of MLG but are not in the more general class of linear context-free grammars [21].

Minimal linear grammars seem to play an interesting role in *the problem of the information transmission and coding.*

We recall that a (unique decipherable) *code* is a set of words that is useful in the information transmission process. This process is described by the well-known Shannon scheme ([29, 1]): it consists of a source $\mathcal{S}$ with source alphabet Y that sends information to a receiver $\mathcal{R}$. This sending process is realized through a channel endowed with an alphabet A that depends upon its physical structure and therefore is distinct from Y. The latter implies the need of a construction of a coding $\varphi : Y^* \to A^*$ of the messages yielded from $\mathcal{S}$. In particular, if the coding is sequential, then φ is formally represented by an injective morphism between the free monoids spanned by Y and A, respectively. The set of words $C = \varphi(Y)$ that guarantees that φ is injective, is called code. The property of being code is called *unique decipherability* and it is equivalent to the fact that every word of the subsemigroup C^+ generated by C is factorized in a unique way as product of words of C.

It is worth recalling that a well-known theorem of Shannon defines for the *maximal codes* a relevant role in the theory: once a probability distribution is given on the letters of the source alphabet of $\mathcal{S}$, every *optimal code for* $\mathcal{S}$, that is, a code whose average cost is the least possible, must be maximal, provided that the cost of transmission of each letter is uniform. Moreover, as a consequence of a theorem of Kraft and Mc Millan, one can construct an optimal prefix code, thus avoiding the synchronization delay.

A fundamental theorem of Schützenberger and Marcus [28] (see also [4]) then provides an important characterization of maximal codes in term of complete sets and Bernoulli distributions.

In some contexts, however, the specificity of the transmission process makes possible a relaxation of the property of unique decipherability in order to construct more efficient codings. This is the case, for instance, of *multi-decipherable codes* introduced by Lempel in [24]. These codes are such that, given a finite message, every possible parsing of the message into codewords must yield the same multiset of codewords. Some important conjectures on multi-decipherable codes raised in [24] has been solved by Restivo in [27]. An algebraic and more general treatment of multi-decipherable codes has been provided in [7], by using the notion of *variety of codes* introduced by Guzmán in [23]. Another generalization of the notion of code is based upon that of *coding partition* introduced in [6] (see also [2]).

In this paper, we investigate MLG in the information transmission problem. Precisely we prove that unambiguous MLG – denoted UMLG, for short – can be used to realize the coding and show that UMLG generalize codes in a very precise way (*cf.* Sec. 3).

We then study a possible extension to UMLG of the theorem by Schützenberger and Marcus (*cf.* Sec. 4). For this purpose, two notions are introduced: the *maximal grammar* which corresponds to the natural analog of maximal code and the *very dense set* which is a reinforcement of the notion of dense set. In the theoretical setting mentioned above, it is proved that all the logical relations among such concepts and that of Bernoulli set (w.r.t. positive distributions) are very well preserved with one exception: the language generated by a maximal UMLG is always dense but, in general, not very dense.

The attention is then focused on another meaningful relation between maximal UMLG and the measure of the corresponding languages w.r.t. a positive Bernoulli distribution. It is proved that a sufficient condition for the maximality of an UMLG G is that, denoting by L the language generated by G, $\pi(L) = 1$ for some positive Bernoulli distribution π on the terminal alphabet of G. For the reversal implication, a partial answer is obtained w.r.t. the subclass of *proportional* UMLG. A UMLG is called proportional if there exists a rational number q such that for all production $X \to uXv$, one has $|v| = q|u|$. Precisely, for these grammars, it is shown the equivalence among the property of maximality, that of maximality as a proportional UMLG, and the measure $\pi(L) = 1$, where π is the uniform Bernoulli distribution on its terminal alphabet.

In order to describe the last contribution of the paper, it is useful to recall a relevant aspect of codes: *the construction of optimal codes and its relation with the commutative equivalence.* We recall that two words are said to be *commutatively equivalent* if one is obtained from the other by rearranging the letters of the word. Moreover, 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 $f(u)$. A conjecture formulated by Schützenberger in the 50's asked for the existence, for an arbitrary finite code, of a commutatively equivalent prefix one [25, 4]. An affirmative solution to the conjecture would have provided a deep implication in the theory. Precisely, it would have guaranteed, given an arbitrary source in which the cost of transmission of letters was not uniform, the existence of an optimal code, avoiding the synchronization delay. However, in [30], Shor has exhibited a counterexample to the conjecture, that remains still open under the further hypothesis that the code is maximal. The problem has been object of an intense research that yielded remarkable results, notably in connection with Bernoulli sets, and with the related issue of the finite completion of codes [17, 18, 19, 26].

In this framework, UMLG are investigated and compared with codes (*cf.* Sec. 5). The first result shows that, as long as the cost of transmission of letters of the source is uniform, then every UMLG can be replaced by a prefix code with the same average cost per letter, showing that coding by UMLG cannot accelerate the transmission rate obtained by optimal prefix codes.

The situation becomes less clear when the cost of transmission is not uniform. In order to clarify this aspect, a natural adaptation of the notion of commutative equivalence is introduced for MLG. It is then proved a result linking such notion with the search for optimality of UMLG. Such result is akin to a theorem of Carter and Gill [12] concerning the commutative equivalence of codes to prefix codes. As a consequence, a characterization of UMLG that are commutatively equivalent to regular ones is given in terms of assignment of symbol costs and probability distribution on source symbols.

An open question is whether there exists an UMLG which is not commutatively equivalent to any regular one. A negative answer to this question would ensure that, also in the case of non-constant symbol cost, coding by UMLG cannot accelerate the transmission rate obtained by optimal prefix codes.

It is worth noticing that UMLG seem to play an interesting role also in the study of the commutative equivalence of context-free and regular languages. Conditions ensuring that a language is commutatively equivalent with a regular one, have been obtained for the classes of bounded semi-linear languages and codes, context-free languages of finite index and UMLG [9, 10, 11, 14, 15, 16].

The paper is organized as follows. In Section 2, preliminaries on codes and context-free grammars are presented. Section 3 is devoted to the coding process realized by UMLG. In Section 4, the notions of maximality, Bernoulli set, dense and very dense set, are analyzed for UMLG. Section 5 is devoted to the analysis of optimal UMLG and the commutative equivalence of grammars. Section 6 contains concluding remarks and open problems.

2 Preliminaries

We now recall some useful terms and basic properties concerning codes and minimal linear grammars [3, 4].

2.1 Words and codes

Let A be a finite non-empty alphabet and A^* be the free monoid generated by A. The identity of A^* is called the *empty word* and is denoted by ϵ. The set $A^* \setminus \{\epsilon\}$ is denoted by A^+. A subset X of A^* is called a *formal language over* A, or simply a language of A^*. The *length* of a word $w \in A^*$ is the integer $|w|$ inductively defined by $|\epsilon| = 0$, $|wa| = |w| + 1$, $w \in A^*$, $a \in A$. For every $a \in A$, $|w|_a$ denotes the number of occurrences of the letter a in w. If $n \in \mathbb{N}$, then $A^{\leq n}$ (resp., $A^{<n}$) denotes the set of all the words of A^* of length not larger than n (resp., smaller than n).

Let $w \in A^*$. The word u is a *factor* of w if there exist $p, q \in A^*$ such that $w = puq$. If $w = uq$, for some $q \in A^*$ (respectively, $w = pu$, for some $p \in A^*$), then u is called a *prefix* (respectively, a *suffix*) of w. Given a word $w \in A^*$ a *factorization* of w is a sequence of factors of w, $w_1, \ldots, w_k$, such that $w = w_1 \cdots w_k$. For any subset X of A^*, we denote by $\mathrm{Fact}(X)$ the set of all factors of words of X, that is $\mathrm{Fact}(X) = \{u \in \mathrm{Fact}(w) : w \in X\}$.

A subset X of A^+ is a *code (over A)* if every word of X^+ has a unique factorization as a product of words of X. Two important types of codes are given by *prefix* and *suffix* codes. A subset X of A^+ is said to be a *prefix code* if $XA^+ \cap X = \emptyset$ that is, if, for every $u, v \in X$, u is not a proper prefix of v. Symmetrically, a subset X of A^+ is said to be a *suffix code* if $A^+X \cap X = \emptyset$ that is, if, for every $u, v \in X$, u is not a proper suffix of v.

Let $\mathbb{R}_+$ be the set of non-negative real numbers. A *Bernoulli distribution* μ on A is any map

$$\mu : A \to \mathbb{R}_+,$$

such that $\sum_{a \in A} \mu(a) = 1$. A Bernoulli distribution is *positive* if, for all $a \in A$, $\mu(a) > 0$. Any Bernoulli distribution μ over A is extended to a unique morphism (still denoted μ) of A^* into the multiplicative monoid $\mathbb{R}_+$. One then extends μ to the family of subsets of A^* by setting, for every $X \subseteq A^*$, $\mu(X) = \sum_{x \in X} \mu(x)$.

If $A = \{a_1, \ldots, a_t\}$ is an ordered alphabet of t letters, and if $w \in A^*$ is an arbitrary word, then the *Parikh vector* of w is the tuple $\psi(w)$ of $\mathbb{N}^t$ defined as $\psi(w) = (|w|_{a_1}, \ldots, |w|_{a_t})$. The function ψ: $A^* \to \mathbb{N}^t$, mapping w into the Parikh vector of w, is an epi-morphism of the free monoid A^* onto the free commutative additive monoid $\mathbb{N}^t$, called the *Parikh morphism (over A)*. One can introduce in A^* the equivalence relation $\sim$, called *commutative equivalence*, defined as follows: for all $u, v \in A^*$, one has $u \sim v$ if $\psi(u) = \psi(v)$. Thus one has $u \sim v$ if the word v is obtained rearranging the letters of u in a different order. Two languages L and L' are said to be *commutatively equivalent*, and one writes $L \sim L'$, if there exists a bijection $f : L \to L'$ such that, for every $u \in L$, $u \sim f(u)$. A set X of words over an alphabet A is said to be *commutatively prefix* if there exists a prefix set X' such that X is commutively equivalent to X'.

2.2 Minimal linear grammars

We will assume that the reader is familiar with the theory of context-free languages (see [3] for a reference). In the sequel, we will shortly recall some basic concepts in order to fix the corresponding notation.

Let $G = (V, T, P, S)$ be a context-free grammar, where V denotes the vocabulary of G, $N = V \setminus T$ denotes the set of non-terminals of G, T denotes the set of terminals, P denotes the set of productions of G, and $S \in V$ denotes the axiom of G. In order to simplify notation, for an arbitrary non-terminal $X \in V$, the set of all the productions of G of the form $X \to u_i, i = 1, \ldots, k$, having X as the left-side component, will be denoted by $X \to u_1 \mid \cdots \mid u_k$.
For every $\alpha, \beta \in V^*$, we write $\alpha \Rightarrow_G \beta$ if α *directly derives* β in G. As usually, the transitive (resp., transitive and reflexive) closure of the relation $\Rightarrow_G$ will be denoted by $\Rightarrow_G^+$ (resp., $\Rightarrow_G^*$). If no ambiguity arises $\Rightarrow_G$ (resp., $\Rightarrow_G^+$, $\Rightarrow_G^*$) is simply denoted $\Rightarrow$ (resp., $\Rightarrow^+$, $\Rightarrow^*$).

Let $\delta = p_1 p_2 \cdots p_k$ be a finite sequence of productions of G and

$$X = \alpha_0 \Rightarrow \alpha_1 \Rightarrow \cdots \Rightarrow \alpha_k$$

be a derivation where any α_i is obtained from α_{i-1} replacing an occurrence of the left side of p_i by the corresponding right side, $1 \leq i \leq k$. In such a case we write $X \Longrightarrow_\delta \alpha_k$. Moreover, the integer k is said to be the *length* of the derivation. For any non-terminal X and any $\alpha \in V^*$, we write $X \Rightarrow^k \alpha$ if there exists a derivation $X \Longrightarrow_\delta \alpha$ of length k.

We denote by $L(G)$ the language $\{u \in T^* \mid S \Rightarrow_G^* u\}$ of all the words of T^* generated by G. A grammar G is said to be *unambiguous* if every $u \in L(G)$ is generated by exactly one leftmost derivation; otherwise G is said to be *ambiguous*.

The notion of commutative equivalence introduced earlier for words and languages can be suitably extended to grammars. We say that two productions of context-free grammars are *commutatively equivalent* if their left sides are equal and their right sides are commutatively equivalent. Two context free grammars $G = (V, N, P, S)$ and $G' = (V, N, P', S)$ are *commutatively equivalent* if there exists a bijection $f\colon P \to P'$ such that every production $p \in P$ is commutatively equivalent to $f(p)$.

A context-grammar $G = (V, T, P, S)$ is said to be *minimal linear* if $V = \{X\}$, that is, G has a unique non-terminal, say X. In this case, in order to simplify our notation, in the sequel, for an arbitrary minimal linear grammar G, we will use the notation $G = (A, X, P)$ to specify that the terminal alphabet is A, the unique non-terminal symbol is X and the set of productions is P. A production $p \in P$ is said to be *terminating* if p is of the form $X \to w$, with $w \in A^*$, otherwise, it is called *non-terminating*.

In the sequel, we will consider uniquely minimal linear grammars with a unique terminating production. Let

$$P = \{X \to u_1 X v_1,\ X \to u_2 X v_2,\ \ldots,\ X \to u_t X v_t,\ X \to w_T\}. \tag{1}$$

Then $L(G)$ is the set of the words $u_{i_1} \cdots u_{i_n} w_T v_{i_n} \cdots v_{i_1}$ with $i_1, \ldots, i_n \in \{1, \ldots, t\}$. Moreover, it is useful to observe that the grammar G is ambiguous if and only if there exists indexes $i_1, \ldots, i_h, j_1, \ldots j_k \in \{1, \ldots, t\}$ such that

$$u_{i_1} \cdots u_{i_h} = u_{j_1} \cdots u_{j_k}, \quad v_{i_h} \cdots v_{i_1} = v_{j_k} \cdots v_{j_1}, \quad \text{and} \quad i_1 \neq j_1.$$

3 Codes and UMLG

Let $G = (A, X, P)$ be a minimal linear grammar with a unique terminating production p_T and let P_N be the set of its non-terminating productions. We consider an alphabet B such that $\mathrm{Card}(B) = \mathrm{Card}(P_N)$ and a bijection $f\colon B \to P_N$. Such a bijection naturally defines an onto function $c_f\colon B^* \to L(G)$, mapping any word $w = a_1 \cdots a_n$ $(a_1, \ldots a_n \in B)$ to the word $c_f(w) \in L(G)$ such that

$$X \underset{f(a_1)}{\Longrightarrow} \alpha_1 \underset{f(a_2)}{\Longrightarrow} \cdots \underset{f(a_n)}{\Longrightarrow} \alpha_n \underset{p_T}{\Longrightarrow} w.$$

The following statement is trivial.

Proposition 1 *The map $c_f\colon B^* \to L(G)$ is injective if and only if G is an UMLG.*

This result suggests that if G is an UMLG, then one may use the map c_f for ‘encoding’ on the alphabet A the words of B^*.

Thus, UMLG's could be an alternative to the classical encoding via variable length codes. In fact, the following proposition shows that, in a certain sense, UMLG's generalizes codes.

Proposition 2 *Let G be a minimal linear grammar with productions*

$$X \to y_1 X,\ X \to y_2 X,\ \ldots,\ X \to y_t X,\ X \to w_T,$$

$y_1, y_2, \ldots, y_t, w_T \in A^*$. *The grammar G is unambiguous if and only if the set $Y = \{y_1, y_2, \ldots, y_t\}$ is a code. Moreover, $L(G) = Y^*$.*

PROOF We denote by p_i the production $X \to y_i X$, $1 \le i \le t$ and by p_T the terminating production $X \to w_T$.

Suppose that the grammar G is ambiguous. Then, there are two distinct sequences of productions $p_{i_1}, \ldots, p_{i_h}, p_T$ and $p_{j_1}, \ldots, p_{j_k}, p_T$ of G, together with $i_1, \ldots, i_h, j_1, \ldots j_k \in \{1, \ldots, t\}$ generating the same word w. One derives

$$w = y_{i_1} \cdots y_{i_h} w_T = y_{j_1} \cdots y_{j_k} w_T, \tag{2}$$

and, consequently $y_{i_1} \cdots y_{i_h} = y_{j_1} \cdots y_{j_k}$, so that Y is not a code.

Conversely, if Y is not a code, we can find $h, k \ge 1$ and indexes $i_1, \ldots, i_h$, $j_1, \ldots, j_k \in \{1, \ldots, t\}$ such that $y_{i_1} \cdots y_{i_h} = y_{j_1} \cdots y_{j_k}$, with $i_1 \ne j_1$. Thus, (2) is verified by a suitable word w which has, in fact, two distinct derivations in G. We conclude that G is ambiguous. □

4 Maximality and probability

As is well known the notions of maximality and completeness play a fundamental role in the theory of variable length codes. We wish to analyze some analogous conditions for unambiguous minimal linear grammars.

In the sequel we will consider exclusively minimal linear grammars with a unique terminating production, which will be denoted by $p_T : X \to w_T$.

We start with two definitions.

Definition 1 Let G be an UMLG and let P_N be the set of its non-terminating productions. We say that G is *maximal* if there does not exists another UMLG G' on the same terminal alphabet whose set of non-terminating productions properly contains P_N.

Our second definition is a refinement of the notion of dense set, useful when dealing with non-regular sets.

Definition 2 A set of words L on the alphabet A is *very dense* (in A^*) if there exists a finite set $F \subseteq A^*$ such that, for all w in A^*, $FwF \cap L \ne \emptyset$.

We recall that L is *dense* (in A^*) if for all $w \in A^*$ one has $A^* w A^* \cap L \ne \emptyset$. Clearly, every very dense set is dense. Conversely, one can prove that a dense regular set is very dense. This property does not hold, in general, for non-regular sets, as shown by the following example

Example 1 Let L be the language of *binary antipalindromes*, which is generated by the UMLG with productions

$$X \to aXb \mid bXa \mid \epsilon .$$

The set L is dense, but it is not very dense. Indeed, if F is any finite set and n is the maximal length of its words, then $Fa^{n+1}F \cap L = \emptyset$.

We establish a result on the measure of very dense sets which will be useful later.

Lemma 1 *Let L be a very dense subset of A^*. For all positive Bernoulli distribution π on A, one has $\pi(L) = +\infty$.*

Proof As L is very dense, there is a finite set $F = \{x_1, \ldots, x_n\}$ such that, for all w in A^*, $FwF \cap L \neq \emptyset$. Denote by L_{ij}, $i, j = 1, \ldots, n$, the sets

$$L_{ij} = \{w \in A^* \mid x_i w x_j \in L\}.$$

Since $A^* = \bigcup_{i,j=1,\ldots,n} L_{ij}$ and $\pi(A^*) = +\infty$, there is at least one pair (i, j) such that $\pi(L_{ij}) = +\infty$. Consequently, $\pi(x_i L_{ij} x_j) = +\infty$. The conclusion then follows from the fact that $x_i L_{ij} x_j \subseteq L$. $\square$

We recall that, according to a well-known theorem of Schützenberger and Marcus [28], if X is a regular code on the alphabet A and π is a positive Bernoulli distribution on A, the following conditions are equivalent:

1. the code X is maximal,
2. the set X^* is dense,
3. one has $\pi(X) = 1$.

In this context, it is worth recalling that a theorem proven by Böe, de Luca and Restivo [5] (see also [18]) provides another remarkable relation among the concepts above. Indeed it is shown that, for every set X of words of A^+, any two of the following three conditions imply the remaining one: (i) X is a code; (ii) X is a complete set (iii) $\mu(X) = 1$, where μ is a positive Bernoulli distribution.

We would like to establish some similar properties relating maximal UMLG and dense and very dense sets.

Let G be a linear minimal grammar with non-terminating productions

$$X \to u_i X v_i, \quad i = 1 \ldots, t, \tag{3}$$

and π be a Bernoulli distribution on the terminal alphabet. Then we will denote by $\pi(G)$ the number

$$\pi(G) = \sum_{i=1}^{t} \pi(u_i v_i).$$

One can easily verify that $\pi(G) = \pi(L_1)/\pi(w_T)$, where $L_1 = \{w \mid X \Rightarrow^2 w\}$.

The following proposition has been proved in [11]. We report here the proof for the sake of completeness.

Proposition 3 *Let $G = (A, X, P)$ be an UMLG. For all positive Bernoulli distribution π on the terminal alphabet, one has $\pi(G) \leq 1$. Moreover,*

$$\pi(L(G)) = \frac{\pi(w_T)}{1 - \pi(G)},$$

where the right hand side of the above equation has to be meant as $+\infty$ in the case that $\pi(G) = 1$.

Proof We assume that G has the nonterminating productions (3) and the terminating production $X \to w_T$, so that $\pi(G) = \sum_{i=1}^{t} \pi(u_i v_i)$. For all $m \geq 1$, we denote by L_m the set

$$\{w \in A^* \mid X \Rightarrow^{m+1} w\}.$$

As one easily verifies, a word w belongs to L_m if and only if it can be factorized

$$w = u_i s v_i \quad \text{with } s \in L_{m-1},\ 1 \leq i \leq t.$$

Moreover, such a factorization is unique by the unambiguity of G. One derives

$$\pi(L_m) = \sum_{i=1}^{t} \sum_{s \in L_{m-1}} \pi(u_i s v_i) = \pi(G)\pi(L_{m-1}).$$

From the equation above, one obtains

$$\pi(L_m) = p^m q, \quad m \geq 0, \tag{4}$$

where $p = \pi(G)$ and $q = \pi(L_0) = \pi(w_T)$.

Now, let ℓ be the maximal length of the right sides of the productions of G. One easily verifies that, for all $m \geq 0$, the maximal length of the words of L_m is not larger than $(m+1)\ell$. Consequently,

$$\pi(L_m) \leq \sum_{i=0}^{(m+1)\ell} \pi(A^i) = (m+1)\ell, \quad m \geq 0. \tag{5}$$

From Equations (4) and (5) one obtains $p^m q \leq (m+1)\ell$ for all $m \geq 0$. This necessarily implies $p \leq 1$, proving the first part of the statement.

To complete the proof, it is sufficient to observe that, in view of the unambiguity of G, from (4) one derives

$$\pi(L(G)) = \sum_{m=0}^{+\infty} \pi(L_m) = \sum_{m=0}^{+\infty} p^m q = \frac{q}{1-p}.$$

□

Remark 1 Consider the polinomial $Q = \sum_{i=1}^{t} u_i v_i$, where the letters of A are viewed as unknowns. Taking into account that $\pi(G)$ is equal to the value of Q, when each unknown $a \in A$ is replaced by its probability $\pi(a)$, the statement of the proposition above may be extended, by continuity, to all (not necessarily positive) Bernoulli distributions.

As a straightforward consequence of Lemma 1 and Proposition 3 one obtains the following

Proposition 4 *Let G be an UMLG. If $L(G)$ is very dense, then for all Bernoulli distribution π on the terminal alphabet, one has $\pi(G) = 1$.*

Now, we establish a sufficient condition for the maximality of an UMLG.

Proposition 5 *Let G be an UMLG. If there exists a positive Bernoulli distribution π on the terminal alphabet such that $\pi(G) = 1$, then G is maximal.*

PROOF Let P_N be the set of non-terminating productions of G. Clearly, for any minimal linear grammar G' on the same terminal alphabet whose set of non-terminating productions properly contains P_N, one has $\pi(G') > \pi(G) = 1$. Thus, by Proposition 3, G' cannot be an UMLG. We conclude that G is maximal. □

Proposition 6 *Let G be a maximal UMLG. Then $L(G)$ is dense.*

PROOF By contradiction, suppose that $L(G)$ is not dense. Then there is a word s which is not in $\mathrm{Fact}(L(G))$. With no loss of generality, we assume that s is unbordered (if it is not the case, just replace s by $sa^{|s|}$ where a is a letter different from the initial letter of s).

Let G' be the grammar obtained adding to G the production $X \to sXs$. By the maximality of G, there would be two different sequences of productions $p_1, \ldots, p_h$ and $q_1, \ldots, q_k$ of G' generating the same word w. With no loss of generality, we assume $p_1 \neq q_1$.

First suppose $w \in L(G)$. Then $s \notin \mathrm{Fact}(w)$ and, therefore, none of the productions $p_1, \ldots, p_h, q_1, \ldots, q_k$ can be $X \to sXs$. Thus, w has two distinct derivations in G. This contradicts the hypothesis that G is unambiguous.

Now, suppose $w \notin L(G)$. This implies that the production $X \to sXs$ occur in both the sequences $p_1, \ldots, p_h$ and $q_1, \ldots, q_k$. Let i and j be, respectively, the least integers such that $p_i = q_j = (X \to sXs)$. Then one has

$$w = x_1 s w_1 s y_1 = x_2 s w_2 s y_2, \quad x_1 w_T y_1,\ x_2 w_T y_2 \in L(G), \tag{6}$$

where $x_1 X y_1$ and $x_2 X y_2$ are the sentential forms generated respectively by the sequences of productions $p_1, \ldots, p_{i-1}$ and $q_1, \ldots, q_{j-1}$. By the unambiguity of G, one has $x_1 w_T y_1 \neq x_2 w_T y_2$ and therefore, either $x_1 \neq x_2$ or $y_1 \neq y_2$. We assume $x_1 \neq x_2$, as the other case can be symmetrically dealt with. Moreover, with no loss of generality, we assume that $|x_1| < |x_2|$. With such assumptions, from (6) one has

$$x_1 = x_2 z, \quad z s w_1 s y_1 = s w_2 s y_2, \tag{7}$$

for some $z \neq \epsilon$. Notice that z is a factor of the word $x_1 w_T y_1 \in L(G)$, so that s cannot be a prefix of z. Thus, from the latter of (7) one derives that z is a proper prefix of s and s is bordered, which contradicts our assumption. We conclude that $L(G)$ is dense. □

Remark 2 By the previous propositions, any UMLG generating a very dense language is maximal and any maximal UMLG generates a dense language. Conversely, there exist maximal UMLG's generating languages which are not very dense and dense languages generated by non-maximal UMLG's.

For instance, the language of binary antipalindromes considered in Example 1 is dense but it cannot be generated by a maximal UMLG. A maximal UMLG generating a language which is not very dense is given in the next example.

By Proposition 5, a sufficient condition for the maximality of an UMLG G is that $\pi(G) = 1$ for some positive Bernoulli distribution π on the terminal alphabet of G. An open question is whether this condition is also necessary.

Example 2 Let P and S be, respectively, a maximal prefix code and a maximal suffix code on the alphabet A and G be the UMLG with the productions

$$X \to uXv, \quad u \in P,\ v \in S, \qquad X \to \epsilon.$$

One easily verifies that G is an UMLG. One has $L(G) = \bigcup_{n\geq 0} P^n S^n$. If, moreover, P and S are maximal codes, then for all Bernoulli distribution π one has $\pi(P) = \pi(S) = 1$ and, consequently, $\pi(L(G)) = +\infty$. Moreover, both P^* and S^* and, consequently, $L(G)$ are dense sets.

However, in general, $L(G)$ is not very dense. For instance, take $P = \{a, ba, bb\}$ and $S = \{a, ab, bb\}$. Then

$$L(G) = \bigcup_{n\geq 0} \{a, ba, bb\}^n \{a, ab, bb\}^n$$

is not a very dense set. Indeed, one can verify that if F is any finite set and n is the maximal length of its words, then $F(ab)^n aF \cap L(G) = \emptyset$.

The following proposition gives a partial answer to the question settled in Remark 2. We say that an UMLG is *proportional* if there exists a rational number q (called the *ratio* of the grammar) such that for all non-terminating production $X \to uXv$, one has $|v| = q|u|$.

Let G be an UMLG and let P_N be the set of its non-terminating productions. We shall say that G is a *maximal proportional* UMLG if it is proportional and there does not exists another proportional UMLG G' on the same terminal alphabet whose set of non-terminating productions properly contains P_N.

Proposition 7 *Let G be a proportional UMLG and π be the uniform Bernoulli distribution on its terminal alphabet. The following conditions are equivalent:*

1. *the grammar G is a maximal UMLG;*
2. *the grammar G is a maximal proportional UMLG;*
3. $\pi(G) = 1$.

In order to prove Proposition 7, we need some preliminary lemmas. The first one concerns the measure of the set of unbordered words of any length.

Lemma 2 *For all $n \geq 0$, let U_n be the set of unbordered words of length n on a d-letter alphabet A and π be the uniform Bernoulli distribution on A. Then $\pi(U_n) > 1 - d^{-1} - d^{-2}$.*

PROOF The statement is trivial if $d = 1$. Thus, we assume $d \geq 2$.

Let $C(n)$ be the number of unbordered words of length n over A. As is well known (see, e.g., [22]), for all $n \geq 0$ one has

$$C(n) = d^n - \sum_{k=1}^{\lfloor n/2 \rfloor} d^{n-2k} C(k).$$

Taking into account that $\pi(U_n) = d^{-n} C(n)$, this equation can be rewritten as

$$\pi(U_n) = 1 - \sum_{k=1}^{\lfloor n/2 \rfloor} d^{-k} \pi(U_k). \tag{8}$$

In particular, the sequence $\pi(U_n)$ is decreasing, so that for $n \geq 2$ one has

$$\begin{aligned}
\pi(U_n) &\geq 1 - d^{-1}\pi(U_1) - \sum_{k=2}^{\lfloor n/2 \rfloor} d^{-k} \pi(U_2) \\
&= 1 - d^{-1} - (1 - d^{-1}) \sum_{k=2}^{\lfloor n/2 \rfloor} d^{-k} \\
&= 1 - d^{-1} - d^{-2} + d^{-\lfloor n/2 \rfloor - 1} \\
&> 1 - d^{-1} - d^{-2}.
\end{aligned}$$

Since for $n \leq 2$ the statement is trivially true, the conclusion follows. □

Our second lemma establishes a 'density property' of languages generated by maximal proportional UMLG.

Lemma 3 *Let G be a maximal proportional UMLG of ratio q. There exists a finite set F such that for all pair of unbordered words u, v with $|v| = q|u|$, one has*

$$FuFvF \cap L(G) \neq \emptyset. \tag{9}$$

PROOF We set $F = A^{<3\ell}$, where ℓ is the maximal length of the right sides of the productions of G.

Let u, v be as in the statement. If $X \to uXv$ is a production of G, then one has $uw_T v \in L(G)$ and the condition is fulfilled.

Now, suppose that $X \to uXv$ is not a production of G and let G' be the grammar obtained adding such a production to G.

By the maximality of G, the grammar G' is ambiguous. Thus, there are two different sequences of productions $p_1, \ldots, p_h, p_T$ and $p'_1, \ldots, p'_k, p_T$ of G' generating the same word w. We assume that w has minimal length among the words with two distinct derivations.

If p_i is the production $X \to u_i X v_i$ and p'_j is the production $X \to u'_j X v'_j$, $1 \le i \le h$, $1 \le j \le k$, then one has

$$w = u_1 \cdots u_h w_T v_h \cdots v_1 = u'_1 \cdots u'_k w_T v'_k \cdots v'_1. \tag{10}$$

Taking into account that G and, consequently, G' are proportional grammars of ratio q, one has

$$\begin{gathered} |v_h \cdots v_1| = q|u_1 \cdots u_h|, \qquad |v'_k \cdots v'_1| = q|u'_k \cdots u'_1|, \\ |w| = |u_1 \cdots u_h| + |v_h \cdots v_1| + |w_T| = |u'_1 \cdots u'_k| + |v'_k \cdots v'_1| + |w_T|. \end{gathered}$$

One easily derives that

$$|u_1 \cdots u_h| = |u'_1 \cdots u'_k|, \qquad |v_h \cdots v_1| = |v'_k \cdots v'_1|,$$

and therefore, from (10),

$$u_1 \cdots u_h = u'_1 \cdots u'_k, \qquad v_h \cdots v_1 = v'_k \cdots v'_1 \tag{11}$$

Since G is unambiguous at least one of the productions involved in one of the derivations of w is $X \to uXv$. Thus we assume with no loss of generality that $u_t = u$, $v_t = v$ for a suitable t, $1 \le t \le h$. From (11) one derives that

$$u_1 \cdots u_{t-1} = u'_1 \cdots u'_{i-1} x_1, \quad u_{t+1} \cdots u_h = y_1 u'_{j+1} \cdots u'_k, \quad u'_i \cdots u'_j = x_1 u y_1,$$

with $1 \le i \le j \le k$, x_1 a proper prefix of u'_i and y_1 a proper suffix of u'_j. Taking into account that G' is proportional, one has also

$$v_{t-1} \cdots v_1 = x_2 v'_{i-1} \cdots v'_1, \quad v_h \cdots v_{t+1} = v'_k \cdots v'_{j+1} y_2, \quad v'_j \cdots v'_i = y_2 v x_2,$$

with x_2 a proper suffix of v'_i and y_2 a proper prefix of v'_j. Since u is unbordered, one has either $u \ne u_i, \ldots, u_j$ or $u = u_i$ and $i = j$. In the first case, $p'_i, \ldots, p'_j$ are all productions of G so that the word

$$z = u'_i \cdots u'_j w_T v'_j \cdots v'_i = x_1\, u\, y_1 w_T y_2\, v\, x_2$$

belongs to $L(G)$, while $x_1, y_1 w_T y_2, x_2 \in F$. Thus, condition (9) is fulfilled.

In the second case, one has $p_t = p'_i$, $x_1 = x_2 = y_1 = y_2 = \epsilon$, $v = v_i$, $p_t = p'_i$, and the word

$$\begin{aligned} w' &= u_1 \cdots u_{t-1} u_{t+1} \cdots u_h w_T v_h \cdots v_{t+1} v_{t-1} \cdots v_1 \\ &= u'_1 \cdots u'_{i-1} u'_{j+1} \cdots u'_k w_T v'_k \cdots v'_{j+1} v'_{i-1} \cdots v'_1 \end{aligned}$$

has two derivations in G', contradicting the minimality of w. □

Proof (of Proposition 7) The implication $1 \Rightarrow 2$ is trivial and the implication $3 \Rightarrow 1$ is a straightforward consequence of Proposition 5. Thus, it is sufficient to prove the implication $2 \Rightarrow 3$.

Let q be the ratio of G and let F be as in Lemma 3. For all $x_1, x_2, x_3 \in F$ we denote by $D(x_1, x_2, x_3)$ the set

$$\{(u, v) \in A^* \times A^* \mid u, v \text{ unbordered, } |v| = q|u|, \ x_1 u x_2 v x_3 \in L(G)\}.$$

By Lemma 3, for all pair of integers $n, m \geq 0$ such that $m = qn$, each pair $(u, v) \in U_n \times U_m$ belongs to some of the sets $D(x_1, x_2, x_3)$, $x_1, x_2, x_3 \in F$. Thus,

$$\sum_{\substack{x_1, x_2, x_3 \in F \\ (u,v) \in D(x_1, x_2, x_3)}} \pi(uv) \geq \sum_{m=qn} \pi(U_n)\pi(U_m),$$

where the sum in the right hand side is extended to all pairs of integers $m, n \geq 0$ such that $m = qn$. Taking into account that this sum contains infinitely many terms and, by Lemma 2, each of them is larger than $(1 - d^{-1} - d^{-2})^2$, we conclude that it amounts to $+\infty$. Hence, in view of the finiteness of F, there exist $y_1, y_2, y_3 \in F$ such that

$$\sum_{(u,v) \in D(y_1, y_2, y_3)} \pi(uv) = +\infty.$$

Let $K = \{y_1 u y_2 v y_3 \mid (u, v) \in D(y_1, y_2, y_3)\}$. From the previous equation, one obtains $\pi(K) = +\infty$. Since K is a subset of $L(G)$, we conclude that $\pi(L(G)) = +\infty$ and therefore, by Proposition 3, $\pi(G) = 1$. $\square$

5 Optimality

According to the classical model of Shannon [1, 29] a transmitter computes an injective coding function $h \colon B^* \to A^*$ on the messages $w \in B^*$ generated with probability p_w by a source.

Let $\mathbb{R}_+$ denote the additive semigroup of positive real numbers. Any letter $a \in A$ has a *transmission cost* $c(a) \in \mathbb{R}_+$ (ideally, the time necessary for its transmission). We can extend c to a morphism $c \colon A^* \to \mathbb{R}_+$. The *average cost* of the transmission of a message of length n is then given by

$$C_n = \sum_{w \in B^n} p_w c(h(w)).$$

Thus, the limit $C = \lim_{n \to \infty} C_n / n$, if existing, may be interpreted as the *average cost per letter* of the transmission.

In the most common case, one has $p_w = \pi(w)$, $w \in B^*$, for a suitable Bernoulli distribution π on the alphabet B.

The encoding function h may be a monomorphism. In such a case, the set $Y = h(B)$ is a uniquely decipherable code and, as one easily verifies, $C_n = nC_1$, so that

$$C = C_1 = \sum_{y \in Y} p_y c(y), \tag{12}$$

where $p_y = \pi(h^{-1}(y))$.

However, the encoding function h may be the function c_f obtained as described in Section 3 from an UMLG $G = (A, X, P)$ and a bijection $f\colon B \to P_N$. In such a case, setting $B = \{b_1, \ldots, b_t\}$ and $f(b_i) = (X \to u_i X v_i)$, $i = 1, \ldots t$, with easy computations, one obtains

$$\begin{aligned} C_n &= \sum_{i_1,\ldots,i_n \in \{1,\ldots,t\}} \pi(b_{i_1} \cdots b_{i_n}) c(u_{i_1} \cdots u_{i_n} w_T v_{i_n} \cdots v_{i_1}) \\ &= n \sum_{i=1}^{t} \pi(b_i) c(u_i v_i) + c(w_T). \end{aligned}$$

Thus,

$$C = \sum_{i=1}^{t} \pi(b_i) c(u_i v_i). \tag{13}$$

In particular, $C_1 = \sum_{i=1}^{t} \pi(b_i) c(u_i v_i) + c(w_T)$ and $C_0 = c(w_T)$, so that (13) can be rewritten as

$$C = C_1 - C_0 = \sum_{y \in L_1} p_y c(y) - c(w_T),$$

where $L_1 = \{y \mid X \Rightarrow^2 y\}$ and $p_y = \pi(h^{-1}(y))$, $y \in L_1$.

In the simplest case, all letters of the alphabet B have the same cost, say 1, so that $c(v) = |v|$ for all $v \in B^*$. Thus, if h is a monomorphism, then $C = \sum_{y \in Y} p_y |y|$, where $Y = h(A)$, while if h is the function c_f considered above, then one obtains

$$C = \sum_{i=1}^{t} p_i |u_i v_i|,$$

where $p_i = \pi(b_i)$.

From Proposition 3 and the Kraft-McMillan Theorem one derives the following

Proposition 8 *Let G be an UMLG with non-terminating productions of the form $X \to u_i X v_i$, $i = 1, \ldots, t$. There exists a prefix code $Y = \{y_1, \ldots, y_t\}$ such that $|y_i| = |u_i v_i|$, $i = 1, \ldots, t$.*

This proposition shows that when the letter cost is constant, every UMLG can be replaced by a prefix code with the same average cost per letter. Thus, coding by UMLG cannot accelerate the transmission rate obtained by optimal prefix codes.

The situation is less clear when there are letters with different costs.

The following proposition shows a link between the search for optimality and the commutative equivalence of UMLG's. It is analogous to a result of Carter and Gill [12] concerning the commutative equivalence of codes to prefix codes.

Proposition 9 *Two UMLG G and G' are commutatively equivalent if and only if they satisfy the following two conditions.*

1. *The terminating productions of G and G' are commutatively equivalent,*

2. *For every assignment of symbol costs and every probability distribution on source symbols, G and G' have the same average cost per letter.*

Proof It is evident that commutatively equivalent grammars satisfy Conditions 1 and 2.

In order to prove that the conditions are also sufficient, it is sufficient to show that for any non-terminating production $X \to \alpha$ of an UMLG G and any letter a of the terminal alphabet, the number $|\alpha|_a$ is completely determined by the knowledge of the average cost per letter for every assignment of symbol costs and every probability distribution on source symbols.

Let $G = (A, X, P)$ be an UMLG. We set $A = \{a_1, \ldots, a_d\}$ and assume that P is given by (1). Moreover, we consider an alphabet $B = \{b_1, \ldots, b_t\}$ and the bijection $f\colon B \to P_N$ mapping any b_j into the production $X \to u_j X v_j$, $j = 1, \ldots, t$.

Let us verify that for any given cost function $c\colon A^* \to \mathbb{R}_+$, the values of $c(u_i v_i)$, $i = 1, \ldots, t$, are completely determined by the knowledge of the average cost per letter for every probability distribution on source symbols.

Indeed, let π_h, $h = 1, \ldots, t$, be Bernoulli distributions such that the matrix $\big(\pi_h(b_i)\big)_{i,h=1,\ldots,t}$ is non-singular. For instance, one may take $\pi_h(b_i) = (1 + \delta_{ih})/(t+1)$, $i, h = 1, \ldots, t$, where δ_{ih} is the Kronecker delta.

The average cost per letter K_h with respect to the distribution π_h is given by

$$K_h = \sum_{i=1}^{t} \pi_h(b_i) c(u_i v_i).$$

Thus, the numbers $c(u_i v_i)$, $i = 1, \ldots, t$ are solutions of the linear system

$$\sum_{i=1}^{t} \pi_h(b_i) x_i = K_h, \quad h = 1, \ldots, t.$$

As the matrix of this system is non-singular, we conclude that the values of $c(u_i v_i)$ are uniquely determined.

Now let us verify that for all $i = 1, \ldots, t$, the numbers $|u_i X v_i|_{a_j} = |u_i v_i|_{a_j}$, $j = 1, \ldots, d$ are completely determined by the knowledge of $c(u_i v_i)$ for every assignment of symbol costs.

Indeed, let $c^{(k)}\colon A \to \mathbb{R}_+$, $k = 1, \ldots, d$, be d cost functions such that the matrix $\big(c^{(k)}(a_j)\big)_{j,k=1,\ldots,d}$ is non-singular. For instance, one may take $c^{(k)}(a_j) = 1 + \delta_{jk})$, $j, k = 1, \ldots, d$. For $k = 1, \ldots, d$, one has

$$c^{(k)}(u_i v_i) = \sum_{j=1,\ldots,d} c^{(k)}(a_j) |u_i v_i|_{a_j},$$

Thus, the numbers $|u_i v_i|_{a_j}$, $j = 1, \ldots, d$ are solutions of the linear system

$$\sum_{j=1,\ldots,d} c^{(k)}(a_j) x_j = c^{(k)}(u_i v_i), \quad k = 1, \ldots, d.$$

As the matrix of this system is non-singular, we conclude that the values of $c(u_i v_i)$ are uniquely determined.

In conclusion, we have shown that the numbers $|u_i X v_i|_{a_j}$, or equivalently, the commutation classes of the words $u_i X v_i$, are completely determined by the average cost per letter for every assignment of symbol costs and every probability distribution on source symbols. This ensures that two UMLG's G and G' satisfying conditions 1 and 2 are, necessarily, commutatively equivalent. □

As a consequence of the previous proposition we obtain the following

Corollary 1 *An UMLG is commutatively equivalent to a regular one if and only if there exists a code Y such that for every assignment of symbol costs and every probability distribution on source symbols, G and Y have the same average cost per letter.*

Proof Let $G = (A, X, P)$ be as in the proof of Proposition 9 and let $Y = \{y_1, y_2, \ldots, y_t\}$ be a code such that for every assignment of symbol costs and every probability distribution on source symbols, G and Y have the same average cost per letter.

By Proposition 2, the grammar G' with the productions

$$X \to y_1 X,\ X \to y_2 X,\ \ldots,\ X \to y_t X,\ X \to w_T,$$

is an UMLG. Moreover, in view of (12) and (13), for every assignment of symbol costs and every probability distribution on source symbols, G' and Y have the same average cost per letter. Consequently, G and G' satisfy Conditions 1 and 2 of Proposition 9 and, therefore, they are commutatively equivalent.

Conversely, if G is commutatively equivalent to a regular UMLG G', then by Proposition 9, G and G' have the same average cost per letter for every assignment of symbol costs and every probability distribution on source symbols. In view of Proposition 9, (12) and (13), one can find a code Y with the same average cost per letter of G' for every assignment of symbol costs and every probability distribution on source symbols. The conclusion follows. □

An open question is whether there exists an UMLG which is not commutatively equivalent to any regular one. A negative answer to this question would ensure that, also in the case of non-constant symbol cost, coding by UMLG cannot accelerate the transmission rate obtained by optimal prefix codes.

Some sufficient condition ensuring that an UMLG is commutatively equivalent to any regular one has been studied in [9, 11].

6 Further remarks

Ambiguity of minimal linear grammars is undecidable [20]. Thus, it is not possible to devise a "Sardinas-Patterson Algorithm" for UMLG's. Hence, it may be useful to devise some restriction on the form of the productions of a minimal linear grammar which ensure decidability of unambiguity and an efficient parsing.

Code completion has been a challenging subject of research. One may ask whether from any UMLG one can obtain a maximal UMLG by conveniently adding productions.

Example 3 Let G be the UMLG with productions

$$X \to a^5X \mid bX \mid abX \mid ba^2X \mid \epsilon.$$

Since the code $X = \{a^5, b, ab, ba^2\}$ is not included in a finite maximal one (see [4]), the productions of G cannot occur in a regular maximal UMLG. We do not know whether it is possible to obtain a (non-regular) maximal UMLG adding productions to G.

References

[1] R. B. Ash, *Information Theory*, Wiley, New York, (1965).

[2] M.-P. Béal, F. Burderi, A. Restivo, Coding partitions of regular sets, *Internat. J. Algebra Comput.* **19**, 1011–1023, (2009).

[3] J. Berstel, *Transductions and Context-Free Languages* Teubner, Stuttgart, (1979).

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

[5] J. M. Boë, A. de Luca, A. Restivo, Minimal complete sets of words, *Theoret. Comput. Sci.* **12**, 325–332, (1980).

[6] F. Burderi, A. Restivo, Coding partitions, *Discrete Math. Theor. Comput. Sci.* **9**, 227–239, (2007).

[7] F. Burderi, A. Restivo, Varieties of codes and Kraft inequality, *Theory of Computing Systems* **40**, 507–520, (2007).

[8] A. Carpi, A. de Luca, Completions in measure of languages and related combinatorial problems, *Theoret. Comput. Sci.* **330**, 35–57, (2005).

[9] 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, 2019, pp. 169–181.

[10] A. Carpi, F. D'Alessandro, On the Commutative Equivalence of Bounded Semi-linear Codes, In: R. Mercas and D. Reidenbach (Eds.), *WORDS 2019*, Lecture Notes in Computer Science, 11682, Springer, 2019, pp. 1–14.

[11] A. Carpi, F. D'Alessandro, On the commutative equivalence of algebraic formal series and languages, manuscript, (2019).

[12] L. Carter, J. Gill, Conjectures on uniquely decipherable codes, *IRE Trans. Inform. Theory*, **IT-7**, 27–38, (1961).

[13] N. Chomsky, M.-P. Schützenberger, The Algebraic Theory of Context-free Languages, in P. Braffort and D. Hirschberg (eds.), *Computer Programming and Formal Systems*, pp. 118–161, North Holland Publishing Company, Amsterdam, (1963).

[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] A. de Luca, A conjecture on continued fractions, *Theoret. Comput. Sci.* **204**, 75–86 (1998).

[18] A. de Luca, Some combinatorial results on Bernoulli sets and codes, *Theoret. Comput. Sci.* **273**, 143–165 (2002).

[19] A. de Luca, Some decompositions of Bernoulli sets and codes, *ITA* **29**, 161–174 (2005).

[20] S. A. Greibach, The Undecidability of the Ambiguity Problem for Minimal Linear Grammars, *Inf. and Cont.* **6**, 119–125 (1963).

[21] M. Gross, Inherent Ambiguity of Minimal Linear Grammars, *Inf. and Cont.* **7**, 366–368 (1964).

[22] T. Harju, D. Nowotka, Counting bordered and primitive words with a fixed weight, Theoret. Comput. Sci. **340**, 273–279 (2005).

[23] F. Guzmán, Decipherability of codes, *J. Pure Appl. Algebra* **141**, 13–35 (1999).

[24] A. Lempel, On multiset decipherable codes, IEEE Trans. Inform. Theory **32**, 714–716 (1986).

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

[26] A. Restivo, On codes having no finite completions, *Discrete Mathematics* **17**, 309–316 (1977).

[27] A. Restivo, A note on multiset decipherable codes, IEEE Trans. Inform. Theory **35**, 662–663 (1989).

[28] M.-P. Schützenberger, R. S. Marcus, Full decodable code-words sets, *IRE Trans. Inf. Theory*, **IT-5**, 12–15 (1959).

[29] C.E. Shannon, A Mathematical Theory of Communication, *Bell Syst. Tech. J.*, **27**, 379–423, 623–656, (1948).

[30] P. W. Shor, A counterexample to the triangle conjecture, *J. Combin. theory Ser. A.*, **38**, 110–112 (1983).