A Note on Unambiguity, Finite Ambiguity and Complementation in Recognizable Two-Dimensional Languages*

Marcella Anselmo[1] and Maria Madonia[2]

[1] Dipartimento di Informatica ed Applicazioni,
Università di Salerno I-84084 Fisciano (SA) Italy
anselmo@dia.unisa.it

[2] Dip. Matematica e Informatica, Università di Catania,
Viale Andrea Doria 6/a, 95125 Catania, Italy
madonia@dmi.unict.it

Abstract. The paper deals with some open questions related to unambiguity, finite ambiguity and complementation of two-dimensional recognizable languages. We give partial answers based on the introduction of special classes of languages of "high complexity", in a sense specified in the paper and motivated by some necessary conditions holding for recognizable and unambiguous languages. In the last part of the paper we also show a new necessary condition for recognizable two-dimensional languages on unary alphabet.

Keywords: Automata and Formal Languages, Unambiguity, Complement, Two-dimensional languages.

1 Introduction

The theory of formal languages of strings is well founded and the research in this framework continues from many decades. The increasing interest for pattern recognition and image processing has more recently motivated the research on languages of pictures or two-dimensional languages, and nowadays this is a research field of great interest. Since the sixties, many approaches have been presented in the literature in order to find in two dimensions a counterpart of notions and results of the one-dimensional languages theory: finite automata, grammars, logics and regular expressions. In 1991, an unifying point of view was presented by A. Restivo and D. Giammarresi who defined the family REC of *recognizable picture languages* (see [11] and [12]), as an equivalent of the class of recognizable (or regular) string languages. This definition takes as starting point a characterization of recognizable string languages in terms of local languages and projections (cf. [10]): the pair of a local picture language and a projection

* This work was partially supported by MIUR Project *"Automi e Linguaggi Formali: aspetti matematici e applicativi"* (2005), by ESF Project *"AutoMathA"* (2005-2010), by 60% Projects of University of Catania and Salerno (2007, 2008).

S. Bozapalidis and G. Rahonis (Eds.): CAI 2009, LNCS 5725, pp. 147–159, 2009.

is called *tiling system*. Tiling systems have also analogies with the tiling of the infinite plane.

REC family inherits several properties from the class of regular string languages. A crucial difference lies in the fact that REC family is not closed under complementation: there are languages in REC whose complement is not in REC [12]. It is then important to take into account also the class co-REC of languages whose complement is in REC. The strict inclusion REC$\subset$ (REC$\cup$ co-REC) holds even in the unary case [19] and it fits the fact that the definition of recognizability by tiling systems is intrinsically non-deterministic. The notion of determinism on tiling systems is discussed in [2].

The non-closure of REC under complementation motivated the definition of *unambiguous two-dimensional languages*, whose family is denoted UREC [11]. Informally, a picture language belongs to UREC when it admits an unambiguous tiling system, that is if every picture has a unique pre-image in its corresponding local language. In [3], the proper inclusion of UREC in REC is proved; it holds true in the unary case too (see [6]). In other words there exist in REC inherently ambiguous languages. An open question is whether UREC is closed under complementation or not. Its answer depends on the following open problem.

Question 1. Does $L \in$ REC and $\overline{L} \notin$ REC imply that $L \notin$ UREC?

Question 1 was firstly stated in [22]. The converse is actually an open question too: Does $L \in$ REC $\setminus$ UREC imply that $\overline{L} \notin$ REC? Note that positive answers to both Question 1 and the converse mean that UREC=REC $\cap$ co-REC and that UREC is the largest subset of REC closed under complementation. Also note that such question is related to some difficult problems on complexity classes [7].

All the inherently ambiguous languages known in the literature are indeed *infinitely ambiguous*, in the sense that it is not possible to recognize them by a tiling system, in such a way that each picture has a fixed number of pre-images at most (see Section 2 for more details). The question whether this is always the case or not is open. Let us state it as follows.

Question 2. Does there exist a language $L \in$ REC $\setminus$ UREC such that L is finitely ambiguous?

In this paper we will answer Questions 1 and 2 in some particular cases, where languages involved have "high complexity", as specified in the following. We will introduce a class *HP* $\subseteq$ co-REC $\setminus$ REC of not-recognizable languages and a class *HK* $\subseteq$ REC $\setminus$ UREC of ambiguous languages, whose languages are "hard" with respect to some complexity functions. We will show that:

1. If $L \in$ REC and $\overline{L} \in$ HP then $L \notin$ UREC.
2. If $L \in$ REC and $L \in$ HK then L is infinitely ambiguous.

Let us emphasize that at present it is not known whether the inclusions HP$\subseteq$ co-REC $\setminus$ REC and HK$\subseteq$ REC $\setminus$ UREC are strict or not. No example (nor a candidate) exists showing the inclusions are strict. Hence in the case HP= co-REC $\setminus$ REC and/or HK=REC $\setminus$ UREC, our results would be a positive answer to Questions 1 and/or 2, in their general setting.

The introduction of classes HP and HK is motivated by some necessary conditions for languages in REC and in UREC, respectively, stating that: if $L \in$REC then the size of some permutation matrices associated to L cannot grow so quickly; and if $L \in$UREC then the rank of some matrices associated to L cannot grow so quickly. Indeed here "HP" stands for "High Permutation matrix" and "HK" stands for "High ranK". All the examples of languages that witness the strict inclusions UREC$\subset$ REC$\subset$ (REC $\cup$ co-REC) have been provided applying the necessary conditions we have just mentioned. The main difficulty in this framework is that there are no characterizations of REC and UREC, that could be easily and fruitfully applied, while we do not know whether the mentioned necessary conditions are also sufficient. A main question is thus the following.

Question 3. Find characterizations of meaningful classes of two-dimensional languages.

An intermediate step in view of solving Question 3 is to look for necessary conditions as tight as possible. In the last part of the paper we will introduce a new necessary condition for the belonging of a picture language to REC, in the case when the alphabet is unary. We will then compare in an example the obtained bound with the other ones known in the literature. Remark that the case when the alphabet has a single letter means studying the shapes of pictures before their contents. This is not a simpler subcase: all separation results known for the general case also holds in the unary case. See [1,5,6,7,18] for recent papers on unary two-dimensional languages.

Let us give some more details on the ideas on which the mentioned necessary conditions are based, since our results will be basically related to them.

In 1998 O. Matz [18] isolated a technique for showing that a language is not recognizable. It consists in considering for any recognizable picture language L and integer m the string language $L(m)$ of all pictures in L of fixed height m. Then if $L \in$ REC it is possible to associate to any tiling system recognizing L a family $\{A_m\}$, where each A_m is an automaton accepting $L(m)$ with c^m states at most, for some constant c. Using some known lower bound on the size of an automaton, he proved a necessary condition for the belonging of a picture language to REC (based on the cardinality of a set of pairs of pictures).

In [3] Matz's technique was firstly used together with some lower bound on the size of unambiguous string automata based on the Hankel matrices of the string languages $L(m)$. Recently in [14] the idea of finding necessary conditions for picture languages by studying the Hankel matrices of $L(m)$ has been considered by rephrasing Matz's necessary condition (for belonging to REC) and Cervelle's necessary condition (for belonging to REC$\cup$co-REC) (see [8]) in terms of parameters of the Hankel matrices.

The paper is organized as follows. After giving the basic definitions and results on two-dimensional languages in Section 2, in Section 3 we recall some necessary conditions for two-dimensional languages and introduce the classes HP and HK. Section 4 contains the main results concerning Questions 1 and 2, while the new necessary condition for the unary case is in Section 5.

2 Preliminaries

In this section we recall some definitions about two-dimensional recognizable languages. More details can be mainly found in [12].

A *two-dimensional string* (or a *picture*) over a finite alphabet Σ is a two-dimensional rectangular array of elements of Σ. The set of all pictures over Σ is denoted by Σ^{**} and a *two-dimensional language* over Σ is a subset of Σ^{**}.

Given a picture $p \in \Sigma^{**}$, let $p_{(i,j)}$ denote the symbol in p with coordinates (i,j), $\ell_1(p) = m$, the number of rows and $\ell_2(p) = n$ the number of columns; the pair (m,n) is the *size* of p. Note that when a one-letter alphabet is concerned, a picture p is totally defined by its size (m,n), and we will write $p = (m,n)$. Remark that the set Σ^{**} includes also all the empty pictures, i.e. all pictures of size $(m,0)$ or $(0,n)$ for all $m,n \geq 0$. It will be needed to identify the symbols on the boundary of a given picture: for any picture p of size (m,n), we consider the *bordered picture* $\widehat{p}$ of size $(m+2,n+2)$ obtained by surrounding p with a special *boundary symbol* $\# \notin \Sigma$.

A *tile* is a picture of size $(2,2)$ and $B_{2,2}(p)$ is the set of all sub-blocks of size $(2,2)$ of a picture p. Given an alphabet Γ, a two-dimensional language $L \subseteq \Gamma^{**}$ is *local* if there exists a finite set Θ of tiles over $\Gamma \cup \{\#\}$ such that $L = \{p \in \Gamma^{**} | B_{2,2}(\widehat{p}) \subseteq \Theta\}$ and we will write $L = L(\Theta)$.

A *tiling system* is a quadruple $(\Sigma, \Gamma, \Theta, \pi)$ where Σ and Γ are finite alphabets, Θ is a finite set of tiles over $\Gamma \cup \{\#\}$ and $\pi : \Gamma \rightarrow \Sigma$ is a projection. A two-dimensional language $L \subseteq \Sigma^{**}$ is *tiling recognizable* if there exists a tiling system $(\Sigma, \Gamma, \Theta, \pi)$ such that $L = \pi(L(\Theta))$ (extending π in the usual way). For any $p \in L$, a local picture $p' \in L(\Theta)$, such that $p = \pi(p')$, is called a *pre-image* of p. We denote by *REC* the family of all *tiling recognizable* picture languages.

The family REC is closed with respect to different types of operations. The *column concatenation* of p and q (denoted by $p \obar q$) and the *row concatenation* of p and q (denoted by $p \ominus q$) are partial operations, defined only if $\ell_1(p) = \ell_1(q)$ and if $\ell_2(p) = \ell_2(q)$, respectively and are given by:

$$p \obar q = \begin{array}{|c|c|}\hline p & q \\ \hline\end{array} \qquad\qquad p \ominus q = \begin{array}{|c|}\hline p \\ \hline q \\ \hline\end{array}.$$

REC family is closed under row and column concatenation and their closures, under union, intersection and under rotation (see [12] for all the proofs).

Let us give some examples to which we will refer later.

Example 1. Let $L_{fc=lc}$ be the language of pictures over $\Sigma = \{a,b\}$, with more than one column, whose first column is equal to the last one. Language $L_{fc=lc} \in$ REC. Informally we can define a local language where information about first column symbols of a picture p is brought along horizontal direction, by means of subscripts, to match the last column of p. Tiles are defined to have always same subscripts within a row while, in left and right border tiles, subscripts and main symbols should match. Below it is an example of a picture $p \in L_{fc=lc}$ together with a pre-image p' of p.

$$p = \begin{array}{|c|c|c|c|c|}\hline b & b & a & b & b \\ \hline a & a & b & a & a \\ \hline b & a & a & a & b \\ \hline a & b & b & b & a \\ \hline \end{array} \quad p' = \begin{array}{|c|c|c|c|c|}\hline b_b & b_b & a_b & b_b & b_b \\ \hline a_a & a_a & b_a & a_a & a_a \\ \hline b_b & a_b & a_b & a_b & b_b \\ \hline a_a & b_a & b_a & b_a & a_a \\ \hline \end{array}.$$

Let $L_{fc=c'}$ be the language of pictures such that the first column is equal to some i-th column, $i \neq 1$. Note that $L_{fc=c'} = L_{fc=lc} \obar \Sigma^{**}$ and thus $L_{fc=c'} \in$ REC. Similarly we can show that the languages $L_{c'=lc} = \Sigma^{**} \obar L_{fc=lc}$, and $L_{c=c'} = \Sigma^{**} \obar L_{fc=lc} \obar \Sigma^{**}$ are in REC.

Example 2. Consider the language $CORNERS$ of all pictures p over $\Sigma = \{a, b\}$ such that whenever $p_{(i,j)} = p_{(i',j)} = p_{(i,j')} = b$ then also $p_{(i',j')} = b$. Intuitively, whenever three corners of a rectangle carry a b, then also the fourth one does. In [18], it is shown that $CORNERS \notin REC$. Consider now, the language $L = \overline{CORNERS}$. We have $L \in REC$; indeed, we can set

$L_1 = \Sigma^{**} \ominus (\Sigma^{**} \obar (\boxed{b} \ominus \Sigma^{**} \ominus \boxed{b}) \obar \Sigma^{**} \obar (\boxed{b} \ominus \Sigma^{**} \ominus \boxed{a}) \obar \Sigma^{**}) \ominus \Sigma^{**}$, and then L is equal to the union of L_1 with the languages obtained by its 90°, 180° and 270° rotations.

A recognizable two-dimensional language $L \subseteq \Sigma^{**}$ is *unambiguous* if and only if it admits an unambiguous tiling system $\mathcal{T}$; a tiling system $\mathcal{T} = (\Sigma, \Gamma, \Theta, \pi)$ is *unambiguous* for L if and only if any picture $p \in L$ has an unique pre-image in the local language $L(\Theta)$ (see [11]). The family of all unambiguous recognizable two-dimensional languages is denoted by *UREC*. In [3] it is proved that the inclusion of UREC in REC is strict and in [6] that this strict inclusion holds even if the alphabet is unary. Therefore in REC there exist languages that are inherently ambiguous.

Let us now recall the definitions of k-ambiguity, finite and infinite ambiguity given in [4] for languages in REC. Note that a similar definition of k-ambiguity is contained in [21]. A tiling system $\mathcal{T} = (\Sigma, \Gamma, \Theta, \pi)$ recognizing L is said to be *k-ambiguous* if every picture $p \in L$ has at most k pre-images. A recognizable language L is said *k-ambiguous* if $k = \min\{s \mid \mathcal{T}$ is s-ambiguous tiling system and $\mathcal{T}$ recognizes $L\}$. A language L is *finitely ambiguous* if it is k-ambiguous for some k whereas a language L is *infinitely-ambiguous* if it is not finitely ambiguous.

3 Classes HP and HK

In this section we introduce the definitions of the classes HP and HK of picture languages motivated by some necessary conditions we recall as well.

Let $L \subseteq \Sigma^{**}$ be a picture language. For any $m \geq 1$, we can consider the subset $L(m) \subseteq L$ containing all pictures in L with exactly m rows. Note that the language $L(m)$ can be viewed as a string language over the alphabet of the columns of height m. If L is in REC then it is possible to associate to any tiling system recognizing L a family $\{A_m\}$, where each A_m is an automaton accepting $L(m)$ with a number of states that is at most c^m for some constant c (see [18]).

Moreover, for any string language L, one can define the infinite boolean Hankel matrix $M_L = \|a_{\alpha\beta}\|_{\alpha\in\Sigma^*,\beta\in\Sigma^*}$ where $a_{\alpha\beta} = 1$ if and only if $\alpha\beta \in L$ (see [15]). Observe that, when L is a regular language, the number of different rows of M_L is finite (Myhill-Nerode Theorem). A sub-matrix $M_{(U,V)}$ of an Hankel matrix M_L is a matrix specified by a pair of languages (U, V), with $U, V \subseteq \Sigma^{**}$, that is obtained by intersecting all rows and all columns of M_L that are indexed by the strings in U and V, respectively. Moreover, given a matrix M, we denote by $Rank_Q(M)$, the rank of M over the field of rational numbers Q. A *permutation matrix* is a boolean matrix that has exactly one 1 in each row and in each column.

Definition 1. *[13] Let L be a picture language.*

i) The row complexity function $R_L(m)$ gives the number of distinct rows of the matrix $M_{L(m)}$
ii) The permutation complexity function $P_L(m)$ gives the size of the maximal permutation matrix that is a sub-matrix of $M_{L(m)}$
iii) The rank complexity function $K_L(m)$ gives the rank of the matrix $M_{L(m)}$.

The following theorem collects some necessary conditions for picture languages.

Theorem 1. *Let $L \subseteq \Sigma^{**}$.*

1. *If $L \in REC \cup co\text{-}REC$ then there is a $c \in \mathbb{N}$ such that, for all $m \geq 1$, $R_L(m) \leq 2^{c^m}$.*
2. *If $L \in REC$ then there is a $c \in \mathbb{N}$ such that, for all $m \geq 1$, $P_L(m) \leq c^m$.*
3. *If $L \in UREC$ then there is a $c \in \mathbb{N}$ such that, for all $m \geq 1$, $K_L(m) \leq c^m$.*
4. *If $L \in REC \setminus UREC$ and L is k-ambiguous then there is a $c \in \mathbb{N}$ such that, for all $m \geq 1$, $K_L(m) \leq c^m$.*

Proof. Item 1 is essentially due to J. Cervelle [8] and item 2 to O. Matz [18]; both of them are rephrased in the matrix framework as in [14]. Item 3 is proved in [3]. Item 4 can be found in [21], for a bit different definition of k-ambiguity, but it holds even for the definition presented in this paper. Indeed if L is k-ambiguous then there exists a constant $c \in \mathbb{N}$ such that, for any $m \geq 1$, there is a k-ambiguous automaton A_m that recognizes language $L(m)$ and has c^m states at most: $|A_m| \leq c^m$. Then we can apply a lower bound on the number of states of k-ambiguous automata in [15] that guarantees that $|A_m| \geq Rank_Q(M_L)^{1/k} - 1$. Therefore $K_L(m)^{1/k} \leq d^m$, for some constant $d \in \mathbb{N}$, and finally $K_L(m) \leq (d^k)^m$. □

Note that in [2], some subclasses of UREC have been introduced and other necessary conditions founded on $R_L(m)$ have been proved.

Definition 2. *HP is the class of all picture languages $L \in co\text{-}REC$ for which there does not exist a constant c such that $P_L(m) \leq c^m$, for all $m \geq 1$.*

HK is the class of all picture languages $L \in REC$ for which there does not exist a constant c such that $K_L(m) \leq c^m$, for all $m \geq 1$.

From previous results, if $L \in$ HP then $L \notin$ REC and if $L \in$ HK then $L \notin$ UREC. Let us now show some examples of languages in HP and in HK. For this, we will use a result, proved in the following Lemma, concerning the rank of some special boolean matrices. In the following, for any matrix $A = \|a_{ij}\|$ with $i = 1, \cdots, m$, $j = 1, \cdots, n$, A_{ij} will denote the (i, j) minor of A.

Lemma 1. *Let $A = \|a_{ij}\|$ be a boolean square matrix of size k such that, for any $1 \leq i, j \leq k$, $a_{ij} = 0$ if and only if $i + j = k + 1$. Then $Rank_Q(A) = k$.*

Proof. It suffices to prove that $det(A) \neq 0$. Remark that A is a square matrix with 0 in all counter-diagonal positions and 1 elsewhere. Let us evaluate $det(A)$ along its first row: $det(A) = \sum_{i=1}^{k}(-1)^{1+i}a_{1i}det(A_{1i}) = det(A_{11}) + (-1)det(A_{12}) + \ldots + (-1)^k det(A_{1n-1}) + 0\ det(A_{1n})$.

Since, for any $i = 2, \ldots, k-1$, the matrix A_{1i} can be obtained from the matrix A_{1i-1} by swapping its $(k-i+1)$-th row with its $(k-i)$-th one, we can say that $det(A) = (k-1)det(A_{11})$. Therefore, in order to prove that $det(A) \neq 0$, it suffices to show that $det(A)_{11} \neq 0$. Note that A_{11} is a square matrix, of size $k-1$, that has 0 in all the positions immediately above the counter-diagonal and 1 elsewhere. Let us denote by B^h the square matrix of size h of this form and let us show that, for any h, B^h has a non-null determinant. The proof is by induction on h. The basis, $h = 2$, is obvious. Suppose that it is true for B^{h-1} and consider the matrix $B^h = \|b_{ij}\|$. If we evaluate $det(B^h)$ along its first column, we have $det(B^h) = \sum_{i=1}^{h}(-1)^{1+i}b_{i1}det(B^h_{i1})$. Remark that the first $(h-2)$ terms of the sum are equal to 0 (every matrix B^h_{i1} has two identical rows, the last one and the second-last one, and therefore it has a null determinant) and the $(h-1)$-th term is equal to 0 too (note that $b_{(h-1)1} = 0$). So we have $det(B^h) = (-1)^{h+1}det(B^{h-1})$ and, therefore, by inductive hypothesis, $det(B^h) \neq 0$. □

Now, let us fix some notation: we denote by ε the empty string and, for $\Sigma = \{a\}$ and $n \in N$, by a^n the string over Σ^* of length n. Moreover, for $n_1, n_2, \ldots, n_m \in N$, we denote by $lcm(n_1, n_2, \ldots, n_m)$ the lowest common multiple of $n_1, n_2, \ldots, n_m$.

Example 3. Consider, for any $m \geq 0$, the function $f(m) = lcm(2^m+1, \ldots, 2^{m+1})$ and the language L_M over the unary alphabet $\Sigma = \{a\}$, $L_M = \{(m, n) \mid n$ is not a multiple of $f(m)\}$. It was shown that $L_M \in$ REC (see [19,20]).

Now, we will show that $L_M \in$ HK. Indeed, for any $m > 1$, consider languages $L_M(m)$ as defined above and the corresponding boolean matrix $M = M_{L_M(m)}$. Let us denote by c the picture over the alphabet Σ with m rows and one column and consider the set S of $f(m)$ rows of M indexed by c, c^2, …, $c^{f(m)}$. They are all distinct (the rows indexed by c^i and c^{i+1} differ in the position corresponding to the column indexed by $c^{f(m)-i}$) and, moreover, any other row in M is equal to one of the rows in S. So $R_L(m) = f(m)$. Consider now, in M, the finite sub-matrix M_c composed by the $f(m)$ rows indexed by c, c^2, …, $c^{f(m)}$, in this order, and the $f(m)$ columns indexed by ε, c, c^2, …, $c^{f(m)-1}$, in this order, as in the following figure.

	ε	c	c^2	$\cdots$	$c^{f(m)-2}$	$c^{f(m)-1}$	$c^{f(m)}$
c	1	1	1	$\cdots$	1	1	0
c^2	1	1	1	$\cdots$	1	0	1
c^3	1	1	1	$\cdots$	0	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c^{f(m)-3}$	1	1	0	$\cdots$	1	1	1
$c^{f(m)-2}$	1	0	1	$\cdots$	1	1	1
$c^{f(m)-1}$	0	1	1	$\cdots$	1	1	1

Then, the i-th row of M_c will have symbol 1 in all its entries except the $f(m)+1-i$ position that will have symbol 0. For this, matrix M_c satisfies the hypothesis of Lemma 1, and, therefore, $Rank_Q(M_c) = f(m)$. But $f(m) = Rank_Q(M_c) \leq K_L(m) \leq R_L(m) = f(m)$, so we have $K_L(m) = f(m)$. Since $f(m) = 2^{\theta(2^m)}$ (see [18,19]), then $K_{L_M}(m)$ cannot be bounded by k^m where k is a constant, and therefore $L_M \in$ HK.

At last, it is easy to see that, for any $m > 1$, $P_L(m) = 2$.

Consider now the language $\overline{L}_M$ and, for any $m > 1$, languages $\overline{L}_M(m)$. The finite sub-matrix of $M_{\overline{L}_M(m)}$, with same rows and columns indexes as M_c, is a square matrix of size $f(m)$ with 1 in all counter-diagonal positions and 0 elsewhere. It is easy to show that, for any m, $P_{\overline{L}_M}(m) = R_{\overline{L}_M}(m) = f(m)$ and, therefore, $\overline{L}_M \in$ HP. Furthermore we have $K_{\overline{L}_M}(m) = f(m)$ too.

Example 4. Let CORNERS be the language defined in Example 2. We are going to show that CORNERS$\in$ HP, following the proof that CORNERS$\notin$ REC in [18].

Consider for any $n \geq 1$ a partition $\mathcal{P}$ of $\{1, 2, \cdots, 2n\}$ into two-element sets and fix a bijection $\alpha_{\mathcal{P}} : \mathcal{P} \to \{1, 2, \cdots, n\}$. Then define picture $P_{\mathcal{P}}$ over $\{a, b\}$ as the picture of size $(2n, n)$ such that the position (i, j) in $P_{\mathcal{P}}$ carries a b if and only if $j = \alpha_{\mathcal{P}}(\{i, i'\})$ and $\{i, i'\} \in \mathcal{P}$. As an example let $n = 3$, $\mathcal{P}= \{(1, 2), (3, 4), (5, 6)\}$ and $\mathcal{P}'= \{(1, 3), (2, 4), (5, 6)\}$; then fix $\alpha_{\mathcal{P}}((1, 2)) = 1$, $\alpha_{\mathcal{P}}((3, 4)) = 2$, and $\alpha_{\mathcal{P}}((5, 6)) = 3$; $\alpha_{\mathcal{P}'}((1, 3)) = 1$, $\alpha_{\mathcal{P}'}((2, 4)) = 2$, and $\alpha_{\mathcal{P}'}((5, 6)) = 3$. Pictures $P_{\mathcal{P}}$ and $P_{\mathcal{P}'}$ are as follows:

$$P_{\mathcal{P}} = \begin{array}{|c|c|c|}\hline b & a & a \\ \hline b & a & a \\ \hline a & b & a \\ \hline a & b & a \\ \hline a & a & b \\ \hline a & a & b \\ \hline\end{array} \qquad P_{\mathcal{P}'} = \begin{array}{|c|c|c|}\hline b & a & a \\ \hline a & b & a \\ \hline b & a & a \\ \hline a & b & a \\ \hline a & a & b \\ \hline a & a & b \\ \hline\end{array}$$

Let $M_{L(2n)}$ be the Hankel matrix of the language $L(2n)$ of pictures in CORNERS of fixed height $2n$, and $M_{(U,V)}$ its sub-matrix specified by the pair of languages (U, V) with $U = V = \{P_{\mathcal{P}} \mid \mathcal{P}$ is a partition of $\{1, 2, \cdots, 2n\}$ into two-element sets$\}$. We have that $M_{(U,V)}$ is a permutation matrix. Indeed the entry $(P_{\mathcal{P}}, P_{\mathcal{P}'})$ of $M_{(U,V)}$ is 1 iff $\mathcal{P} = \mathcal{P}'$.

Furthermore the size of matrix $M_{(U,V)}$ is equal to the number A_n of partitions of $\{1, 2, \cdots, 2n\}$ into two-element sets. And it can be shown that $A_n \geq n!$ and then there does not exist $c \in \mathbb{N}$ such that $A_n \leq c^n$.

Let us mention that another language in HK is $L_{c=c'}$ as introduced in Example 1 (see [2]), while its complement is in HP (see [13]).

4 Some Results on Classes HP and HK

In this section we give partial answers to Questions 1 and 2 as stated in the Introduction, in the case the involved languages belong to classes HP and HK introduced in Section 3. Firstly let us compare the values of the complexity functions $R_L(m)$, $P_L(m)$ and $K_L(m)$ introduced in Section 3, for a language L and its complement, in the case L is in REC $\cup$ co-REC (and therefore functions $R_L(m)$, $P_L(m)$ and $K_L(m)$ have finite values.

Proposition 1. *Let $L \in$ REC $\cup$ co-REC.*

1. $R_{\overline{L}}(m) = R_L(m)$.
2. $P_L(m) + P_{\overline{L}}(m) \leq R_L(m) + 2$ *and the bound is tight.*
3. $K_L(m) + K_{\overline{L}}(m) \leq 2R_L(m)$ *and the bound is tight.*
4. $K_{\overline{L}}(m) \geq P_L(m)$ *and the bound is tight.*

Proof

1. The Hankel matrices for $\overline{L}$ can be obtained by exchanging entries 0 with entries 1 in the Hankel matrices for L.
2. Let $m \geq 1$, $M_{(U,V)}$ be a permutation matrix of maximal size that is a sub-matrix of the Hankel matrix $M_{L(m)}$ for $L(m)$, and let $\overline{M}_{(U',V')}$ be a permutation matrix of maximal size that is a sub-matrix of the Hankel matrix $M_{\overline{L}(m)}$ for the complement of $L(m)$. We claim that $|U \cap U'|$, $|V \cap V'| \leq 2$. In other words the sub-matrices of $M_{L(m)}$ specified by (U, V) and (U', V'), respectively, cannot overlap more than on a square matrix of size 2. Consider indeed a column of $M_{L(m)}$ indexed by a string in $|V \cap V'|$. The set of entries on such column indexed by strings in U are all 0's except for one 1 and then they cannot share more than two elements (a 0 and a 1) with the set of entries indexed by strings in U' (that are all 1's except for one 0).

   The bound is tight for language L_M in Example 3: $P_{L_M}(m) = 2$ and $P_{\overline{L}_M}(m) = R_{L_M}(m) = f(m)$ and, therefore, $P_{L_M}(m) + P_{\overline{L}_M}(m) = R_{L_M}(m) + 2$.
3. The inequality follows from Item 1 and from the remark $K_L(m) \leq R_L(m)$. The bound is tight for language L_M in Example 3: $K_{L_M}(m) = K_{\overline{L}_M}(m) = R_{L_M}(m) = f(m)$ and, therefore, $K_{L_M}(m) + K_{\overline{L}_M}(m) = 2R_{L_M}(m)$.
4. Let P be a maximal permutation matrix of $H_L(m)$. Clearly, P is a boolean square matrix of size $P_L(m)$ and we can assume, without loss of generality, that P has 1 in all counter-diagonal positions and 0 elsewhere. Now, consider $H_{\overline{L}}(m)$ and its submatrix of size $P_L(m)$, say $\overline{P}$, that corresponds to the permutation matrix P of $H_L(m)$. Remark that $\overline{P}$ is a square matrix of size

$P_L(m)$ with 0 in all counter-diagonal positions and 1 elsewhere. Therefore, the matrix $\overline{P}$ satisfies the hypothesis of Lemma 1 and we have $Rank_Q(\overline{P}) = P_L(m)$ that implies $K_{\overline{L}}(m) \geq P_L(m)$. The bound is tight for language $\overline{L}_M$, that is the complement of language L_M in Example 3. □

Corollary 1. *If $L \in HP$ then $\overline{L} \in HK$.*

The following proposition is a positive answer to Question 1 (see Section 1) in the case where $\overline{L} \notin$ REC since $\overline{L} \in$ HP. Recall that if a language is in HP then it is necessarily not in REC; and that we do not know at present whether this is also a sufficient condition. Note that if this condition were also sufficient then HP=co-REC\REC. Vice versa, if HP = co-REC\REC then the condition would be also sufficient for languages in co-REC.

Proposition 2. *If $L \in REC$ and $\overline{L} \in HP$ then $L \notin UREC$.*

Proof. If $\overline{L} \in$ HP then, from Corollary 1, $L \in$ HK and therefore, from the definition of HK and Item 3 of Theorem 1, $L \notin$ UREC. □

As an application, consider the following example.

Example 5. In Example 4 we showed that $CORNERS \in$ HP. Applying Corollary 1 we have that $\overline{CORNERS} \in$ HK and finally $\overline{CORNERS} \notin$ UREC.

The following proposition is a negative answer to Question 2 (see Section 1) in the case where $L \notin$ UREC since $L \in$ HK. Recall that if a language is in HK then it is necessarily inherently ambiguous; and that we do not know at present whether this is also a sufficient condition. If this condition were also sufficient then HK=REC\UREC. Vice versa, if HK = REC\UREC then the condition would be also sufficient for languages in REC.

Proposition 3. *Any language $L \in REC \setminus UREC$ such that $L \in HK$ is infinitely ambiguous.*

Proof. The proof follows from item 4 in Theorem 1. □

Example 6. In Example 5 we showed that $\overline{CORNERS} \in$ HK; hence from Proposition 3 we have that $\overline{CORNERS}$ is infinitely ambiguous.

5 Necessary Conditions in the Unary Case

In this section we will introduce a new necessary condition for the belonging of a picture language to REC, in the case when the alphabet is unary. We will then compare the obtained bound with the other ones known in the literature (see Theorem 1). All along this section $|\Sigma| = 1$.

Let us recall some classical results on unary regular string languages (see [9,10]). A unary language is regular if and only if it is ultimately periodic. The size of a deterministic finite automaton (dfa) for an unary language is (λ, μ), where λ is the number of states in the cycle and μ is the number of states not in the cycle. A language is said properly ultimately λ-cyclic when it is accepted by a dfa of size (λ, μ) and by no dfa of size (λ', μ') with $\lambda' < \lambda$; λ is said the period of the language. Obviously any regular unary language is properly ultimately λ-cyclic for some $\lambda \in I\!N$.

Suppose $\lambda \in I\!N$ factorizes in prime powers as $\lambda = p_1^{k_1} p_2^{k_2} \cdots p_s^{k_s}$; we will denote by $spp(\lambda)$ the sum of its prime powers, that is $spp(\lambda) = p_1^{k_1} + p_2^{k_2} + \cdots + p_s^{k_s}$. The following lower bound was proved for unary non-deterministic finite automata (nfa) in [16]: Each nfa accepting a properly ultimately λ-cyclic language has at least $spp(\lambda)$ states in its cycles. We will apply this lower bound to obtain a new necessary condition for unary picture languages in REC. Recall the definition of $L(m)$ in Section 3.

Proposition 4. *Let L be a recognizable language over a unary alphabet and for any $m \geq 1$, λ_m be the period of $L(m)$. Then there exists a constant $c \in I\!N$ such that for any $m \geq 1$, $spp(\lambda_m) \leq c^m$.*

Proof. If $L \in$ REC then there exist nfa's A_m that recognize languages $L(m)$ and have c^m states at most, for some constant $c \in I\!N$ (see Section 3): $|A_m| \leq c^m$ for any $m \geq 1$. In the case $|\Sigma| = 1$, we can apply the above mentioned lower bound of [16] to automata A_m's, and obtain that for any $m \geq 1$: $spp(\lambda_m) \leq |A_m| \leq c^m$. □

When the alphabet is unary, Proposition 4, with items 1 and 2 of Theorem 1, provides three different conditions for the belonging of a language to REC. Let us summarize the three conditions.

Proposition 5. *Let L be a recognizable language over a unary alphabet and for any $m \geq 1$, λ_m be the period of $L(m)$.*

Then there exists a constant $c \in I\!N$ such that for any $m \geq 1$:

1. $\log R_L(m) \leq c^m$
2. $P_L(m) \leq c^m$
3. $spp(\lambda_m) \leq c^m$.

In the following example we compare the three bounds of Proposition 5.

Example 7. Consider language L_M as defined in Example 3: $L_M = \{(m, n) \mid n$ is not a multiple of $f(m)\}$ where $f(m) = lcm(2^m + 1, \ldots, 2^{m+1})$, for all $m \geq 0$. Recall that $L_M \in$ REC. The Hankel matrices associated to languages $L(m)$ are described in Example 3. We show that $P_{L_M}(m) \leq \log R_{L_M}(m) \leq spp(\lambda_m)$.

We have $\lambda_m = R_{L_M}(m) = f(m)$ and $P_{L_M}(m) = 2$. Suppose that for any $m \geq 0$, λ_m factorizes in prime powers as $\lambda_m = \prod p_{m,i}^{k_{m,i}}$, then $spp(\lambda_m) = \sum p_{m,i}^{k_{m,i}}$. Hence $2 = P_L(m) < \log R_{L_M}(m) = \log(\prod p_{m,i}^{k_{m,i}}) = \sum \log(p_{m,i}^{k_{m,i}}) < \sum p_{m,i}^{k_{m,i}} = spp(\lambda_m)$.

6 Conclusions and Open Questions

In the paper we afforded some open problems on unambiguity, finite ambiguity and complementation (Questions 1, 2 and 3 in the Introduction) and gave some partial answers. A complete answer to Question 1 seems far to be found, also due to its interpretation in the computational complexity framework.

With Question 2, we considered the possibility that in REC there exist finitely ambiguous languages, and showed that this is not true for a class of languages in REC. Note that in a bit different framework (see [4]), when the recognition is accomplished without border symbols (tiles with # are not allowed), it is shown that there is an infinite hierarchy of finitely ambiguous languages. Therefore the border symbols have to play a major role, in order to show that in REC there do not exist finitely ambiguous languages. We figure that when a language is recognized by a tiling system with finite ambiguity, then it is possible to obtain an unambiguous tiling system for the language by paying special attention to border tiles.

Finally the problem of finding characterizations of meaningful classes of recognizable languages (Question 3) deserves some more investigation.

References

1. Anselmo, M., Giammarresi, D., Madonia, M.: New Operators and Regular Expressions for two-dimensional languages over one-letter alphabet. Theoretical Computer Science 340(2), 408–431 (2005)
2. Anselmo, M., Giammarresi, D., Madonia, M.: From determinism to non-determinism in recognizable two-dimensional languages. In: Harju, T., Karhumäki, J., Lepistö, A. (eds.) DLT 2007. LNCS, vol. 4588, pp. 36–47. Springer, Heidelberg (2007)
3. Anselmo, M., Giammarresi, D., Madonia, M., Restivo, A.: Unambiguous Recognizable Two-dimensional Languages. RAIRO: Theoretical Informatics and Applications 40(2), 227–294 (2006)
4. Anselmo, M., Jonoska, N., Madonia, M.: Framed Versus Unframed Two-dimensional Languages. In: Nielsen, M., Kucera, A., Miltersen, P.B., Palamidessi, C., Tuma, P., Valencia, F.D. (eds.) SOFSEM 2009. LNCS, vol. 5404, pp. 79–92. Springer, Heidelberg (2009)
5. Anselmo, M., Madonia, M.: Deterministic two-dimensional languages over one-letter alphabet. In: Bozapalidis, S., Rahonis, G. (eds.) CAI 2007. LNCS, vol. 4728, pp. 147–159. Springer, Heidelberg (2007)
6. Anselmo, M., Madonia, M.: Deterministic and unambiguous two-dimensional languages over one-letter alphabet. Theoretical Computer Science 410, 1477–1485 (2009)
7. Bertoni, A., Goldwurm, M., Lonati, V.: On the complexity of unary tiling-recognizable picture languages. In: Thomas, W., Weil, P. (eds.) STACS 2007. LNCS, vol. 4393, pp. 381–392. Springer, Heidelberg (2007)
8. Cervelle, J.: Langages de figures. Rapport de stage, ENS de Lyon (1997)
9. Chrobak, M.: Finite automata and unary languages. Theoret. Comput. Sci. 47, 149–158 (1986)
10. Eilenberg, S.: Automata, Languages and Machines, vol. A. Academic Press, London (1974)
11. Giammarresi, D., Restivo, A.: Recognizable picture languages. Int. Journal Pattern Recognition and Artificial Intelligence 6(2&3), 241–256 (1992)
12. Giammarresi, D., Restivo, A.: Two-dimensional languages. In: Rozenberg, G., et al. (eds.) Handbook of Formal Languages, vol. III, pp. 215–268. Springer, Heidelberg (1997)

13. Giammarresi, D., Restivo, A.: Matrix based complexity functions and recognizable picture languages. In: Grader, E., Flum, J., Wilke, T. (eds.) Logic and Automata: History and Perspectives. Texts in Logic and Games 2, pp. 315–337. Amsterdam University Press (2007)
14. Giammarresi, D., Restivo, A.: Ambiguity and complementation in recognizable two-dimensional languages. In: Ausiello, G., Karhumäki, J., Mauri, G., Ong, L. (eds.) Procs. Intern. Conf. on Theoret. Compu. Sci. IFIP, vol. 273, pp. 5–20. Springer, Boston (2008)
15. Hromkovic, J., Karumäki, J., Klauck, H., Schnitger, G., Seibert, S.: Communication Complexity Method for Measuring Nondeterminism in Finite Automata. Information and Computation 172, 202–217 (2002)
16. Jiang, T., McDowell, E., Ravikumar, B.: structure and complexity of minimal nfa's over a unary alphabet. Intern. J. Found. Comput. Sci. 2, 163–182 (1991)
17. Lindgren, K., Moore, C., Nordahl, M.: Complexity of two-dimensional patterns. Journal of Statistical Physics 91(5-6), 909–951 (1998)
18. Matz, O.: On piecewise testable, starfree, and recognizable picture languages. In: Nivat, M. (ed.) FOSSACS 1998. LNCS, vol. 1378, p. 203. Springer, Heidelberg (1998)
19. Matz, O.: Dot-depth and monadic quantifier alternation over pictures. Ph.D. thesis Technical Report 99-08, RWTH Aachen (1999)
20. Matz, O.: Dot-depth, monadic quantifier alternation, and first-order closure over grids and pictures. Theoretical Computer Science 270(1-2), 1–70 (2002)
21. Mäurer, I.: Characterizations of Recognizable Picture Series, Ph.D Thesis Universität Leipzig, Institut für Informatik, Abteilung Automaten und Sprachen (2007)
22. Potthoff, A., Seibert, S., Thomas, W.: Nondeterminism versus determinism of finite automata over directed acyclic graphs. Bull. Belgian Math. Soc. 1, 285–298 (1994)