Volume 4, Issue 5, May 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Comparison results of 2D Monogenic Pure Context Free Picture Grammars with other relating models

P. Julia Grace
Department of Computer Science,
JBAS College for Women (Autonomous), Chennai,India

Abstract: Picture languages generated by array grammars or accepted by array automata have been studied by researchers and various models have been proposed in the literature, motivated by problems arising in the framework of syntactic methods of pattern recognition and image processing. The notion of Pure 2D Picture Grammars, generating rectangular picture arrays of symbols is already there in the literature. In this paper, we introduce another notion called 2D Monogenic Pure Context-free Picture Grammars and Compare its generative power with other related models. We have also obtained certain closure properties and shown that this class is polynomial-time identifiable in the limit from positive data.
***Keywords** : Grammars, Languages, Sironmoney Matrix Grammar, Pure 2D context Free Grammar, 2D Monogenic CFG.*

1. INTRODUCTION

In formal language theory, a picture language is a set of *pictures*, where a picture is a 2D array of characters over some alphabet. Picture languages generated by array grammars or accepted by array automata have been studied by researchers and various models have been proposed in the literature, motivated by problems arising in the framework of syntactic methods of pattern recognition and image processing. The basic class of generative array grammars, due to Rosenfeld [2, 3, 4, 7], directly extends the notions of string grammars and string languages to array grammars and array languages. This is done by introducing array rewriting rules corresponding to the Chomskian hierarchy. The Siromoney models were introduced in [9, 10], for the description of two-dimensional digital pictures viewed as rectangular arrays of terminal symbols.

In this two – dimensional model, called as Siromoney matrix grammar, generation of rectangular arrays takes place in two phases with a sequential mode of rewriting in the first phase generating strings of intermediate symbols and a parallel mode of rewriting these strings in the second phase to yield rectangular picture patterns. Recently there has been a renewed interest in the study of Siromoney matrix grammars [13, 14].

Another very general rectangular array generating model, called extended controlled tabled L array system (ECTLAS) was proposed by Siromoney and Siromoney [11], incorporating into arrays the developmental type of generation used in the well-known biologically motivated L-systems. Here the symbols either on the left, right, up or down borders of a rectangular array are rewritten simultaneously by equal length strings to generate rectangular picture arrays.

Pure context-free grammars [1] which make use of only terminal symbols, unlike the Chomskian grammars, have been investigated in formal string language theory for their language generating power and other properties. It has been shown [16] that the whole class of pure context-free languages is not inferable from positive data only. Hence, Tanida and Yokomori [16] have introduced monogenic pure context-free (mono-PCF) grammars in which each string generated is uniquely determined by its predecessor in a derivation. They have shown that this subclass of pure context-free languages is identifiable in the limit from positive data. In [15] a new two-dimensional grammar, called Pure 2D context – free grammar, for picture array generation based on pure context-free rules has been introduced and its generative power with those in [10,11,12] have been compared. Unlike the models in [10,11], rewriting any column or any row of the rectangular array are rewritten and there is no priority of rewriting columns and rows as in [10] in which the second phase of generation can take place only after the first phase is over. In this paper, we define 2D Monogenic Pure Context-free Picture Grammars (2DMonoPCPG) and compare its generative power with other related models and obtain certain closure properties also. We have show that this class is polynomial-time identifiable in the limit from positive data.

2. PRELIMINARIES

Let $\sum$ be a finite alphabet. A word or string $w = a_1 a_2..a_n$ $(n \geq 1)$ over $\sum$ is a sequence of symbols from $\sum$. The length of a word w is denoted by |w|. The set of all words over $\sum$, including the empty word λ with no symbols, is denoted by $\sum^*$. We call words of $\sum^*$ as horizontal words. For any word $w = a_1a_2..a_n$, we denote by w^T the vertical word.

$$\begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_n. \end{matrix}$$

We also define $(w^T)^T = w$. We set λ^t as λ itself. A rectangular $m \times n$ array M over $\sum$ is of the form

$$M = \begin{matrix} a_{11} \cdots a_{1n} \\ \cdots \\ \cdots \\ \cdots \\ a_{m1} \cdots a_{mn} \end{matrix}$$

where each $a_{ij} \in \sum$, $1 \leq i \leq m$, $1 \leq j \leq n$. The set of all rectangular arrays over $\sum$ is denoted by $\sum^{**}$, which includes the empty array λ. $\sum^{**} - \{\lambda\} = \sum^{++}$. We denote respectively by © and ® the column concatenation and row concatenation of arrays in $\sum^{**}$. In contrast to the case of strings, these operations are partially defined, namely, for any X, Y $\in \sum^{**}$, X © Y is defined if and only if X and Y have the same number of rows. Similarly X ® Y is defined if and only if X and Y have the same number of columns. We refer to [5, 6] for array grammars. For notions of formal language theory we refer to [8]. We briefly recall pure context-free grammars [1] and the rectangular picture generating models in [10, 11, 12, 13].

Definition 2.1: A **Siromoney matrix grammar** is a 2-tuple (G_1, G_2) where $G_1 = (H_1, I_1, P_1, S)$ is a regular or context free grammar , H_1 is a finite set of horizontal non terminals, $I_1 = \{S_1, S_2, \ldots. S_k\}$, a finite set of intermediates, $H_1 \cap I_1$ = NULL, P_1 is a finite set of production rules called horizontal production rules , S is a start symbol, $S \in H_1$. $G_2 = (G_{21}, G_{22}, .. G_{2k})$ where $G_{2i} = (V_{2i}, T, P_{2i}, S_i)$, $1 \leq i \leq k$ are regular grammars , V_{2i} is a finite set of vertical non terminals, $V_{2i} \cap V_{2j}$ = null, i not = j, T is a finite set of terminals, P2i is a finite set right linear production rules of the form $X \rightarrow aY$ or $X \rightarrow a$ where X, Y $\in V_{2i}$, $a \in T$, $S_i \in V_{2i}$ is the start symbol of G_{2i}.
The set L(G) of picture arrays generated by G consists of all $m \times n$ arrays $[a_{ij}]$ such that $1 \leq i \leq m$, $1 \leq j \leq n$ and $S \Rightarrow^* G_1\ S_{i1} S_{i2} \ldots S_{in} \Rightarrow^* G_2\ [a_{ij}]$. We denote the picture language classes of regular , CF Siromoney matrix grammars by RML, CFML respectively.
The regular / CFSMG specify a finite set of tables of rules in the second phase of generation with each tables having either right linear non terminal rules or right linear terminal rule. The resulting families of picture array languages are denoted by TRML and TCFML and are known to properly include RML and CFML respectively.

Definition 2.2 [15] :A **Pure 2D Context-free grammar** (P2DCFG) is a 4-tuple $G = (\Sigma, P_c, P_r, M)$ where
- Σ is a finite set of symbols;
- $P_c = \{t_{ci} / 1 \leq i \leq m\}$, $P_r = \{t_{rj} / 1 \leq j \leq n\}$;

Each t_{ci}, $(1 \leq i \leq m)$, called a column table, is a set of context-free rules of the form $a \rightarrow \alpha$, $a \in \Sigma$, $a \in \Sigma^*$ such that for any two rules $a \rightarrow \alpha$, $b \rightarrow \beta$ in t_{ci}, , we have $|\alpha| = |\beta|$ where $|\alpha|$ denotes the length of $|\alpha|$; Each t_{rj}, $(1 \leq j \leq n)$, called a row table, is a set of context-free rules of the form $c \rightarrow \gamma^T$, $c \in \Sigma$, and $\gamma \in \Sigma^*$ such that for any two rules $c \rightarrow \gamma^T$, $d \rightarrow \delta^T$ in t_{rj}, we have $|\gamma| = |\delta|$
- $M_1 \subseteq \Sigma^{**} — \{\lambda\}$ is a finite set of axiom arrays.

Derivations are defined as follows: For any two arrays M_1,M_2, we write $M_1 \Rightarrow M_2$ if M2 is obtained from M_1 by either rewriting a column of $M_1$by rules of some column table t_{Ci} in P_c or a row of M_1 by rules of some row table t_{rj} in P_r. =>* is the reflexive transitive closure of => .The picture array language L(G] generated by G is the set of rectangular picture arrays $\{M^{|} / M_0 \Rightarrow^* M^{|} \in \Sigma^{**}$, for some $M_O \in M$ }. The family of picture array languages generated by Pure 2D Context-free grammars is denoted by P2DCFL.

Definition 2.3: A **2D pure context free picture grammar** (2DMonoPCFG) grammar $G = (\Sigma, P_c, P_r, M_A)$ is said to be **monogenic** if and only if it holds that

1) whenever $M \Rightarrow^R M'$, $M, M' \in \Sigma^{**}$, there exists unique arrays W_1 and W_2 such that

$$M = \begin{matrix} W_1 \\ x \\ W_2 \end{matrix} \text{ and } M' = \begin{matrix} W_1 \\ y \\ W_2 \end{matrix}, \quad W_1, W_2 \in \Sigma^{**}$$ and $\exists$ exactly one table $t_{rj} \in P_r$ which has rules for all the symbols of x such that $x \Rightarrow^R y$, where x is a row matrix of size 1 x n and $y \in \Sigma^{**}$ is a m x n matrix, $m, n \geq 1$ and moreover, there is no array M'' such that $M'' \neq M'$ and $M \Rightarrow^R M''$.

2) Whenever $M \Rightarrow^C M'$, $M\ M' \in \Sigma^{**}$ there exists unique arrays W_1 and W_2 such that $M = W_1\, x\, W_2$ and $M' = W_1\, y\, W_2$, $W_1, W_2 \in \Sigma^{**}$ and $\exists$ exactly one table $t_{cj} \in P_c$ which has rules for all the symbols of x such that $x \Rightarrow^C y$, where x is a column matrix of size m x 1 and $y \in \Sigma^{**}$ is a matrix of size m x n, m, n ≥ 1 and moreover , there is no array M'' such that $M'' \neq M'$ and $M \Rightarrow^C M''$.

3) M_A is singleton with a 3 × 3 array as axiom.

Example 1.

Let $S = \begin{matrix} x & c & y \\ a & p & b \\ x & e & y \end{matrix}$, $P_r = \{tr_1\}$, $P_c = \{tc_1\}$

$P_c = \{ c \rightarrow xcy,\ p \rightarrow aqb,\ e \rightarrow xey\}$

$$P_r = \left\{ a \rightarrow \begin{matrix} x \\ a \\ x \end{matrix},\ q \rightarrow \begin{matrix} c \\ p \\ e \end{matrix},\ b \rightarrow \begin{matrix} y \\ b \\ y \end{matrix} \right\}$$

$$\begin{matrix} x & c & y \\ a & p & b \\ x & e & y \end{matrix} \overset{tc_i \in P_c}{\Rightarrow} \begin{matrix} x & x & c & y & y \\ a & a & q & b & b \\ x & x & e & y & y \end{matrix} \overset{tr_j \in P_r}{\Rightarrow} \begin{matrix} x & x & c & y & y \\ x & x & c & y & y \\ a & a & p & b & b \\ x & x & e & y & y \\ x & x & e & y & y \end{matrix} \overset{tc_i \in P_c}{\Rightarrow}$$

$$\begin{matrix} x & x & x & c & y & y & y \\ x & x & x & c & y & y & y \\ a & a & a & q & b & b & b \\ x & x & x & e & y & y & y \\ x & x & x & e & y & y & y \end{matrix} \overset{tr_j \in P_r}{\Rightarrow} \begin{matrix} x & x & x & \mathbf{c} & y & y & y \\ x & x & x & \mathbf{c} & y & y & y \\ x & x & x & \mathbf{c} & y & y & y \\ \mathbf{a} & \mathbf{a} & \mathbf{a} & \mathbf{p} & \mathbf{b} & \mathbf{b} & \mathbf{b} \\ x & x & x & \mathbf{e} & y & y & y \\ x & x & x & \mathbf{e} & y & y & y \\ x & x & x & \mathbf{e} & y & y & y \end{matrix}$$

(figure 1)

3. COMPARISON RESULTS:

We can compare the new 2D grammar model introduced here with those in [6,10,11,19,20]

Theorem 3.1: The family of 2DMonogenicPCFL(2DMonoPCL) is incomparable with the families of Regular Siromoney Matrix Language (RSML) and Context Free Sironmoney Matrix Model (CFSML) but not disjoint with these families.

Proof: The picture language consisting of rectangular arrays over a single symbol a of all sides m x n (m,n ≥1) is generated by a regular Siromoney matrix grammar G. in fact the language of horizontal words generated in the first phase of G1 is $\{S_1^n / n \geq 1\}$ where S_1 is an intermediate symbol and the language of vertical words generated by S_1 in the second phase is $\{(a^n)^T / n \geq 1\}$. But this cannot be generated by a 2DMonoPCFL as the array cannot be partitioned as W_1 or $W_1 x W_2$, where W_1, W_2

x

W2

€∑** where the elements of x are unique. The 2DMonoPCFG in Example 2 cannot be generated by any Regular Siromoney matrix grammar since each of the generated pictures has an equal number of columns to the left and right of the middle column $(c^n\, p\, e^n)^T$, where n > 1 and an equal number of rows above and below the middle row $a^m\, p\, b^m$ where m>1.

The picture language generating the letter H can be generated by a regular Siromoney matrix grammar $G = (G_h, G_v)$ and $G_v = (G_1, G_2)$ where $G_1 = \{S_1 \rightarrow xS_1, S \rightarrow x\}$, $G_2 = \{S_2 \rightarrow bS2, S_2 \rightarrow xA, A \rightarrow bA, A \text{ -> } b\}$.The language of horizontal words generated in the first phase of G_1 is $\{S_1 S_2^n S_1 / n >= 1\}$ where S_1 ,S_2 are intermediate symbols and the language of words generated by S_1 in the second phase is $\{(x^n)^T / n \geq 1\}$ and the language of words generated by S_2 in the second phase is $\{(b^n\, x\, b^n)^T / n \geq 1$, b denotes the blank}

The incomparability with CFML is due to the fact that the picture language generating token T, by the CF-SMG using the horizontal language $L = \{S_1^n S_2^n S_1^n / n \geq 1\}$ in the 1st phase and the languages $\{ a^m / m \geq 1\}$ by S_2 and $\{a(b)^m / m \geq 1\}$ by S_1 in the second phase can also be generated by using a 2DMonoPCF in Example 3.

On the other hand the picture language consisting of rectangular arrays of the form $M_1 \circ M_2$ where M_1, M_2 are rectangular arrays over the symbol a, b respectively with equal number of columns can be generated by a context free Siromoney matrix grammar with the language of horizontal words $S_1^n\, S_2^n$ (S_1,S_2 are intermediate symbols) in the first phase and S_1,S_2 generating vertical words over a, b respectively. But this picture language, cannot be generated by any 2D Monogenic PCF grammar since the string language $\{ a^n\, b^n / n >= 1\}$is not a pure CFL[4] and an argument similar to this can be done in 2D case. Also, the 2DMonoPCF given in example 2, cannot be generated by any context free Siromoney matrix grammar.

Example 2:

Let $S = \begin{matrix} r & c & t \\ b & p & b \\ b & e & b \end{matrix}$, $P_c = \{c \rightarrow xcy,\ p \rightarrow bqb,\ c \rightarrow bcb\}$, $P_r = \left\{ b \rightarrow \begin{matrix} b \\ b \\ b \end{matrix},\ q \rightarrow \begin{matrix} l \\ p \\ m \end{matrix} \right\}$.

```
                                          r x  c  y t
                           r x c y  t     b b  l  b b
r c  t      r  r  c y t     b b l  b b     b b  l  b b
b p b  =>    b b q b b  =>  b b p  b b  => b  b q  b b  =>  . . .  This gives the letter T.
b e b       b b e b b      b b m b b     b  b m b b
                           b  b e b b     b  b  e  b b
```

(figure 2)

Theorem 3.2: **The family of 2DMonoPCFL is incomparable with the families of TRML and TCFML but not disjoint with these families.**

Proof: In view of the proper inclusions RML $\subseteq$ TRML, CFML $\subseteq$ TCFML, and incomparability (Theorem 1) of 2DMonoPCFL with RSML and CFSML, in figure 1 can neither belong to TRML nor to TCFML, in view of the fact that in the picture arrays in Figure 1 each has an equal number of rows above and below the middle row $a^n\, p\, b^n$ it is enough to note that the picture array language of example 2 generating picture arrays as shown.

Theorem 3.3: **The family of 2DMonoPCFL is not closed under union, column catenation, row catenation but is closed under projection and transposition.**

Proof: Let the alphabet be {a,b,c,x,y}. Non-closure under union follows by the fact that $L1 = \{X_1 © (c^n)^T © Y_1 / X_1 \in \{a\}^{++}, Y_1 \in \{b\}^{++}, |X_1|_c = |Y_1|_c\}$where $|X|_c$ stands for the number of columns of X, is generated by a P2DMonogenicCFG with a column table consisting of a rule $c \rightarrow acb$ and a row table with rules $a \rightarrow a$, $b \rightarrow b$,

$c \rightarrow c$. Likewise $L_2 = \{X_2 © (c^n)^T © Y_2 / X_2 \in \{x\}^{++}, Y_2 \in \{y\}^{++}, |X_2|_c = |Y_2|_c \}$

c

is also generated , by a similar 2DMonoPCFG. It can be seen that L_1 U L_2 cannot be generated by any 2DMonoPCFG, since such a grammar will require a column table with rules of the forms c → adb and c → xdy. But then this will yield arrays not in the union.

Non-closure under column catenation of arrays can be seen by considering L_1© L_2 and noting that any 2DMonoPCFG generating L_1 © L_2 will again require a column table with rules c → adb and c → xdy, which is not possible. Non-closure under row catenation can be seen in the same manner.

If L is a picture array language generated by a 2DMonoPCFG G and L^T is a transposition of L, then the 2DMonoPCFG $G^'$ to generate L^T is formed by taking the column tables of G as row tables as column tables but for a rule a →α in the column table of G, the rule a → a^T (α є Σ^{**}) is added in a row table of G, the rule b → β is added in the corresponding column table of $G^'$.

Proposition 3.1. Every 2D-Mono Pure Context-free Picture Language L is a finite union of languages of the form (X © $(c^m)^T$ © Y) ® ((a^n) q (b^n)) ® (X © $(d^m)^T$ © Y) where X is an array $(x)_{m\times n}$ and Y is an array $(y)_{m\times n}$.

4. POLYNOMIAL-TIME IDENTIFICATION IN THE LIMIT

Let ***L*** be a class of languages over a fixed alphabet ∑. Further, let ***G*** be the class of grammars such that for all L ∈ ***L***, there exists a G ∈ ***G*** such that L = L(G). Let G be a grammar in G representing a given L. A positive presentation of language L is any infinite sequence of strings such that every string w ∈ L occurs at least once in the sequence, and no other strings not in L(G) occur in the sequence.

An algorithm ***A*** is said to identify a language L in the limit from positive data using ***G*** if and only if for any positive presentation of L the infinite sequence of G_i's in ***G*** produced by ***A*** satisfies the property that there exist G in ***G*** such that for all sufficiently large i, the i-th conjecture G_i is identical to G and L(G) = L. A class of languages ***L*** is identifiable in the limit from positive data using G if and only if i) there exists an algorithm ***A*** that, given an L in ***L***, identifies L in the limit from positive data using G and ii) the inference algorithm ***A*** for ***L*** satisfies the property that there exists a polynomial p such that for any L in ***L*** and for any positive presentation of L, the time used by ***A*** between receiving the i-th example w_i and outputting the i-th conjectured grammar G_i is at most $p(l_1 + \ldots + l_i)$ where $l_j = lg(w_j)$

Proposition 4.1. The class of P2DCFL is not identifiable from positive data.
Proposition 4.2. The class of 2DMonoPCFL is identifiable from positive data.

5. ACKNOWDEDGEMENT

The author thank *Dr. J.D.Emerald, Asst Professor, Dept of Mathematics, Govt. Arts College, Wallajapet, Vellore* and *Dr. K.G. Subramanian*, School of Mathematical Sciences, Universiti Sains Malaysia, for their valuable suggestions and guidance in preparing this paper.

REFERENCES

[1] Maurer H.A. and Salomaa A., Wood D., Pure Grammars, Information and control 44, 47-72 (1980).
[2] Milgram D.L. and Rosenfeld A., Array automata and array grammars, In proceedings of IFIP Congress 1971, Booklet TA-2, North Holland, Amsterdam, 1972.
[3] Rosenfeld A., Array grammars normal forms, Information Control, 23 (1975).
[4] Rosenfeld A., Some notes on finite-state picture languages, Information Control, 31 (1976).
[5] Rosenfeld A., Picture Languages – Formal models for picture recognition, Academic press, New York (1979).
[6] Rosenfeld A. and Siromoney, R., Picture Languages – a survey, Languages of design 1, 229-245 (1993).
[7] Rosenfeld A., Isotonic grammars, parallel grammars and picture grammars, Machine Intelligence, 6 (1971).
[8] Salomaa A., Formal Languages, Academic Press, London (1973).
[9] Siromoney G., Siromoney R., Krithivasan K., Picture Grammars with array rewriting rules, Information Control, 22 (1973).

[10] Siromoney G., Siromoney R., Krithivasan K. Abstract families of matrices and picture languages, Computer Graphics and Image Processing 1, 234-307 (1972).
[11] Siromoney G. and Siromoney R., Extended Controlled Tabled L-arrays, Information and control, 35(2), 119-138 (1977).
[12] Siromoney R., Subramanian K.G. and Rangarajan K., Parallel/Sequential rectangular arrays with tables, International Journal of Computer Mathematics, 143-158 (1977).
[13] Stiebe R., Picture generation using Matrix systems, Journal of Information Processing and Cybernetics 28, 311-327 (1992).
[14] Stiebe R., Slender Siromoney Matrix Languages, In Proceedings of the 1st International Conference on Language and Automata, Theory and Applications, Taragona, Spain (2007).
[15] Subramanian K.G., Nagar A.K. and Geethalakshmi M., Pure 2D Picture Grammars (P2DPG) and P2DPG with Regular Control, Lecture Notes in Computer Science 4958, Springer-Verlag, Berlin Heidelberg, 330–341 (2008).
[16] Tanida N. and Yokomori T., Inductive Inference of Monogenic Pure Context-Free Languages, IEICE Trans. Inf. and Systems, Vol. E79-D, No. 11, 1503-1510 (1996).
[17] Wang P.S.P., Array Grammars, Patterns and recognizers, World Scientific, Singapore (1989).
[18] Giammarresi D, Restivo, : Two dimensional languages. Ln: Rozenberg, G., Salomaa A., (eds) Handbook of formal languages, Vol 3, pp, 215-267. Springer (1997)
[19] Giammarresi D, Restivo : Recognizable Picture Languages. International Journal of Pattern Recognition and Artificial Intelligence 31-46 (1992)
[20] Rani Siromoney, K.G.Subramanian, Lisa Mathew, "Learning of Pattern and Picture Languages" International Journal of Pattern Recognition and Artificial Intelligence , 275-284, (1992)

Author Profile

Dr. (Mrs.) P. **Julia grace** received her B.Sc., Computer Science in 2000 and MCA in 2003, both from Manonmaniam Sundaranar University, Tirunelveli. She got the University FIRST Rank in MCA. She obtained her M.Phil., from Madurai Kamaraj University in 2005 and Ph.D from Mother Teresa Women's University, Kodaikanal in 2013. At present she is working as an Assistant Professor in Computer Science, JBAS College for Women, Chennai. So far, she has published 7 research articles in reputed International Journals and presented more than 10 papers in National, International Conferences. Her areas of interest are Artificial Intelligence, Computer Graphics, Machine Learning and Theoretical Computer Science.

