L SYSTEMS OF FINITE INDEX
(extended abstract)

G.Rozenberg   D.Vermeir
Department of Mathematics   Department of Applied Economics
University of Antwerp, U.I.A.   University of Leuven, K.U.L.
Wilrijk, Belgium   Leuven, Belgium

1. INTRODUCTION

L systems theory (see e.g., [1] , [5] , [7]) forms today one of the most vigorously investigated areas of formal language theory. It brought with it a whole new range of problems (and techniques for solving them) and it shed new light on quite a number of fundamental problems and problem areas of formal language theory. One such interesting problem area accentuated by the theory of L systems is the difference between sequential and parallel rewriting in language generating devices (see e.g. [10]). The classical Chomsky grammars are totally sequential (in one rewriting step only one symbol is rewritten) while L systems are totally parallel (in one rewriting step all symbols in a string must be rewritten). A natural step towards understanding the nature of the difference between sequential and parallel rewriting is to look for devices that lie somewhere in-between these two extremes.

In our paper we present such an approach by putting certain restrictions on L systems. In particular we require that the L systems we consider are such that each word in the language can be derived in such a fashion that each intermediate string contains no more than an a priori bounded number of symbols that can be rewritten by something else than itself. L systems satisfying this restriction are said to be of finite index. Indeed they constitute a good example of systems where the rewriting mechanism represents a "compromise" between totally sequential and totally parallel rewriting: although the number of "rewriting activities" in a string is constrained to a limited number of letters, all of these can be rewritten in parallel.

In addition to the above argument the following points indicate our motivation for the research presented in this paper.
1) The notion of finite index we consider here for the case of L systems is a close analog of the classical notion of finite index in formal language theory (see, e.g. [9]). Hence it is natural to investigate this topic also in the framework of L systems.
2) The theory of L systems originated from the biological considerations of Lindenmayer (see [3] or [4]). As a matter of fact a lot of examples considered in the biologically oriented part of L systems theory are indeed examples of development satisfying the finite index restriction. This kind of behaviour is quite common in biological development, so much so that it is very probable that, even if the notion of finite index was not already introduced in formal language theory, it would be

introduced in the framework of L systems theory.

3) It turns out (see Section 6) that the class of languages generated by ETOL systems under the finite index restriction is a very central class in the sense that an amazingly large number of classes of language generating devices under the finite index restriction generate precisely this class of languages. However the advantage of using ETOL systems to investigate this class stems from the fact that they form a very neat algebraic structure (by Theorem 2 it suffices to consider a finitely generated semigroup of homomorphisms), whereas in almost all other cases the underlying generating structure is really complicated (obscure?). Thus from this point of view L systems can be considered as a very useful tool in investigating quite a large number of language generating systems.

We assume the reader is familiar with the rudiments of formal language theory (see, e.g., [9]) and the rudiments of L systems theory (see, e.g., [1]). We use standard terminology and notation. Perhaps only the following requires an explanation : $\underline{\min}(x)$ denotes the set of all letters appearing in a word x and $\#_B(x)$ denotes the number of occurrences of letters from B in x.

2. BASIC DEFINITIONS

A feature that distinguishes nonterminal symbols from terminal symbols in Chomsky grammars is that the nonterminals are symbols which are subject to further transformation (rewriting). In ETOL systems the situation is more complicated: both terminal and nonterminal symbols can be rewritten. Hence the need for the notion of an <u>active</u> symbol which is defined as follows.

Definition 1. Let $G = <V, \mathcal{R}, S, \Sigma>$ be an ETOL system. A letter a from V is called <u>active (in</u> G) if there exist a table P in $\mathcal{R}$ and a word α in V^* such that $a \underset{P}{\to} \alpha$ is in P and $\alpha \neq a$. Then $A(G) = \{a \in V : a$ is active in $G\}$.

If we want to introduce the notion of a finite index ETOL system in the way that is usually done in formal language theory then it is clear that one should count active rather than nonterminal symbols.

Definition 2. Let G be an ETOL system.

1) Let k be a positive integer. We say that G <u>is of index k</u> if for every word x in L(G) there exists a derivation of x in G with the trace $x_1, \ldots, x_n$ such that, for $1 \leq j \leq n$, $\#_{A(G)}(x_j) \leq k$.

2) We say that G is of <u>finite index</u> if G is of index k for some $k \geq 1$.

Definition 3. Let K be an ETOL language.

1) Let k be a positive integer. We say that K is of index k if there exists an ETOL system G of index k such that L(G) = K.

2) We say that K is of finite index if K is of index k for some $k \geq 1$.

Among ETOL systems of finite index one can naturally distinguish these in which _every_ succesful derivation satisfies a finite index restriction. These systems are formally defined now. Note the analogous situation in the case of context free grammars of finite index versus ultralinear grammars (see, e.g. [9]) or in the case of ETOL systems with fragmentation with outside and inside control (see, [8]).

Definition 4. Let G be an ETOL system.

1) Let k be a positive integer. We say that G _is of uncontrolled index k_, if for every word x in L(G) whenever $x_1,...,x_n$ is the trace of a derivation of x in L(G) then, for $1 \leqslant j \leqslant n$, $\#_{A(G)}(x_j) \leqslant k$.

2) We say that G is of _uncontrolled finite index_ if G is of uncontrolled index k for some $k \geqslant 1$.

Definition 5. Let K be an ETOL language.

1) Let k be a positive integer. We say that K is of _uncontrolled index_ k if there exists an ETOL system G of uncontrolled index k such that L(G) = K.

2) We say that K is of _uncontrolled finite index_ if K is of uncontrolled index k for some $k \geqslant 1$.

If X denotes the class of language generating devices, then we use $\mathcal{L}(X)_{FIN}$ and $\mathcal{L}(X)_{FINU}$ to denote the class of languages generated by elements of X under the finite index and uncontrolled finite index restrictions respectively. If we want to fix a particular index k then we use $\mathcal{L}(X)_{FIN(k)}$ or $\mathcal{L}(X)_{FINU(k)}$.

3. NORMAL FORM AND DECIDABILITY RESULTS

In this section we establish a number of technical results that facilitate our further investigation.

First we show that one can always "organize" an ETOL system in such a way that the nonterminal and active symbols coincide.

Theorem 1. There exists an algorithm which given an ETOL (EDTOL) system G produces an equivalent ETOL (EDTOL) system $H = <V, \mathcal{P}, S, \Sigma>$ such that $A(H) = V \setminus \Sigma$.

If an ETOL system satisfies the conditions required of H in the above theorem then we say that it is in Active Normal Form.
The following result is the basic technical tool in investigating ETOL languages of finite index.

Theorem 2. There exists an algorithm which given an arbitrary ETOL system G of index k produces an equivalent EPDTOL system H which is of uncontrolled index k and is in Active Normal Form.

It is very instructive to compare Theorem 2 with the corresponding result in the case of sequential rewriting : the class of context free languages of finite index and the class of ultralinear languages do not coincide!!

If an ETOL system satisfies the conditions required of the system H in the previous theorem then we say that it is in <u>Finite Index Normal Form</u> (abbreviated as FINF).

A natural question is whether we can get a normal form result which would allow us to consider only EOL systems of finite index. The answer is negative but we do show that we can restrict ourselves to ETOL systems of finite index with two tables only.

Theorem 3. 1) There exists an algorithm which for every ETOL system G of (uncontrolled) index k produces an equivalent EPDTOL system $H = <V, \mathcal{P}, S, \Sigma>$ of (uncontrolled) index k such that $\# \mathcal{P} = 2$.

2) There exists an ETOL language K of finite index such that for every ETOL system $G = <V, \mathcal{P}, S, \Sigma>$ which generates K we have $\# \mathcal{P} \geqslant 2$.

Here is another normal form result that turned out to be quite useful in our considerations.

Definition 6. An ETOL system $G = <V, \mathcal{P}, S, \Sigma>$ is said to be of <u>fixed index</u> k if for every word x in L(G) and every derivation of x with a trace $S, x_0, \dots, x_{n+1}$ is such that for every i in $\{0, \dots, n\}$, $\#_{A(G)}(x_i) = k$.

The notion of an ETOL system of <u>uncontrolled fixed index</u> k is defined in an analogous way.

Theorem 4. There exists an algorithm which for every ETOL system G of index k produces an equivalent EPDTOL system of uncontrolled fixed index k.

Given an ETOL system $G = <V, \mathcal{P}, S, \Sigma>$ and a language $C \subseteq \mathcal{P}^{\star}$ we define $L_C(G)$ to be the set of all these words in L(G) that can be derived in G in such a way that the sequence of tables used in the derivation belongs to C.

Theorem 5. Let Σ be a finite alphabet. For every positive integer k there exists an EPDTOL system G_k of uncontrolled index k such that $\{L_C(G_k) : C$ is a regular set$\}$ is identical with the family of ETOL languages of index k over Σ.

As far as basic decision problems are concerned we would like to report the following two results.

Theorem 6. 1) There is no algorithm which given an arbitrary ETOL system G and a positive integer k decides whether or not G is of index k.

2) There is no algorithm which given an arbitrary ETOL system G decides whether or not G is of finite index.

The above result is a "negative" one; but we are already used to living with such results in formal language theory. However the situation in our case is much better than usual. We know already (see Theorem 2) that we can restrict ourselves to ETOL systems of uncontrolled finite index, and in this case we have the following result.

Theorem 7. 1) There is an algorithm which given an arbitrary ETOL system G and a positive integer k decides whether or not G is of uncontrolled index k. 2) There is an algorithm which given an arbitrary ETOL system G decides whether or not G is of uncontrolled finite index.

4. PROPERTIES OF $\mathcal{L}(ETOL)_{FIN}$

One of the useful research directions in investigating a class of languages is to aim at results that allow one to see "the structure" of a single language in the class and that allow one to prove (constructively) that some languages are not in the class. We next present several examples of this kind.

Theorem 8. Let K be an ETOL language of index k. There exist positive integers n_o, m_o such that for every word w in K longer than n_o, $w = w_o \alpha_1 w_1 \cdots \alpha_{2k} w_{2k}$ with $\alpha_1 \cdots \alpha_{2k} \neq \Lambda$, $|\alpha_i| \leq m_o$ for $1 \leq i \leq k$ and for every non-negative integer t, $w_o \alpha_1^t w_1 \alpha_2^t \cdots \alpha_{2k}^t w_{2k} \in K$.

As a corollary to this theorem we get the following result.

Theorem 9.

$$\mathcal{L}(ETOL)_{FIN(1)} \subsetneqq \mathcal{L}(ETOL)_{FIN(2)} \subsetneqq \cdots \subsetneqq \mathcal{L}(ETOL)_{FIN(i)} \subsetneqq \cdots \subset \mathcal{L}(ETOL)_{FIN}.$$

For any class of languages Y, let PAR(Y) denote the family of sets of Parikh vectors of each language in Y. Let $\mathcal{L}(REG)$ denotes the family of regular languages.

Theorem 10. $\underline{PAR}(\mathcal{L}(ETOL)_{FIN}) = \underline{PAR}(\mathcal{L}(REG))$.

Finally it turns out that $\mathcal{L}(ETOL)_{FIN}$ has a nice algebraic structure.

Theorem 11. (i) For every $k \geq 1$, $\mathcal{L}(ETOL)_{FIN(k)}$ is a full semi AFL (which is full principal).
(ii) $\mathcal{L}(ETOL)_{FIN}$ is a substitution closed full AFL (which is not full principal).

5. METALINEAR ETOL SYSTEMS

One way of studying the structure of ETOL systems of finite index is to study productions available for active symbols. For example one can have an ETOL system of fixed index k where, in one step, one active letter can produce k active letters while all other active letters produce only nonactive letters. In general we can talk about active letters of type j, where $1 \leq j \leq k$, in the sense that such a letter can introduce in one step j active letters and never more. Thus on the one hand we can have "complicated" letters of type k and on the other hand "very simple" letters of type 1 (or even "trivial" letters of type 0). In this section we will study ETOL systems of this simple type: by analogy with context-free grammar

theory we call them <u>metalinear</u>.

<u>Definition 7</u>. An ETOL system $G = <V,\mathcal{P},S,\Sigma>$ is called <u>metalinear</u>
if
1) S does not appear at the right-hand side of any production in any table of $\mathcal{P}$,
2) for every letter X in $A(G) \setminus \{S\}$ and every table P in $\mathcal{P}$, if $X \rightarrow \alpha$ is in P then
$\#_{A(G)}(\alpha) \leqslant 1$.
If $\max\{\#_{A(G)}(\alpha) : (\exists P)_{\mathcal{P}} [S \rightarrow \alpha \text{ is in P}]\} \leqslant k$ then we say that G is k-<u>metalinear</u>.

<u>Definition 8</u>. A language K is called a (k-) <u>metalinear ETOL</u>
<u>language</u> if there exists a (k-) metalinear ETOL system G such that $K = L(G)$.

Again if X denotes the class of ETOL systems or any of its sub-
classes then we use $\mathcal{L}(X)_{m\ell}$ and $\mathcal{L}(X)_{m\ell(k)}$ to denote the class of languages generated
by elements of X under the metalinear and k-metalinear restriction respectively.

As we have indicated already, one reason to study metalinear ETOL
systems is that they represent the simplest nontrivial structure among ETOL systems
of finite index. Still another reason is that in some cases they can be used to
yield simple representation theorems as for example the following one.

<u>Theorem 12</u>. If $K \in \mathcal{L}(HDOL)_{FIN}$ then there exist $K_1, \dots, K_n$ in
$\mathcal{L}(HDOL)_{m\ell}$ such that $K = \bigcup_{i=1}^{n} K_i$.

Metalinear ETOL languages obey the following "pumping" theorem
which is very useful in proving that certain languages are not in $\mathcal{L}(ETOL)_{m\ell}$.

<u>Theorem 13</u>. Let K be a k-metalinear ETOL language. There exist
positive integer constants N,M,R such that for every word w in K which is longer than
N, $w = \alpha_1 \dots \alpha_{2p}$ for some $p \leqslant k$ where for $1 \leqslant i \leqslant 2p$ if $\alpha_i = \bar{\alpha}_i \beta_i \bar{\bar{\alpha}}_i$ with $|\beta_i| \geqslant M$ then
$\beta_i = \bar{\beta}_i \gamma_i \bar{\bar{\beta}}_i$ with $|\gamma_i| \leqslant R$ and for every nonnegative integer t, $\varphi_t \bar{\alpha}_i \bar{\beta}_i \gamma_i^t \bar{\bar{\beta}}_i \bar{\bar{\alpha}}_i \psi_t \in K$
for some φ_t, ψ_t.

As a corollary to this theorem we get the following result.
<u>Theorem 14</u>.

$\mathcal{L}(ETOL)_{m\ell(1)} \subsetneqq \mathcal{L}(ETOL)_{m\ell(2)} \subsetneqq \dots \subsetneqq \mathcal{L}(ETOL)_{m\ell(i)} \subsetneqq \dots \subsetneqq \mathcal{L}(ETOL)_{m\ell}$.

The class $\mathcal{L}(ETOL)_{m\ell}$ is not as stable under AFL operations as the
whole class $\mathcal{L}(ETOL)_{FIN}$.

<u>Theorem 15</u>. (i) For every $k \geqslant 1$, $\mathcal{L}(ETOL)_{m\ell(k)}$ is a full principal
semi AFL.
(ii) $\mathcal{L}(ETOL)_{m\ell}$ is a full semi AFL which is not principal.

6. ACTIVE RECURSION IN ETOL SYSTEMS

Another way of looking at the structure of ETOL systems of finite index is to look at the "character of recursion" present in a system. This for example was succesfully done in the case of context-free grammars (see, e.g., [9]) where it was proved that the class of languages generated by nonexpansive context-free grammars coincides with the class of languages generated by context-free grammars of finite index. However, this direct approach does not work in the case of ETOL systems. (We call an ETOL system G nonexpansive if it does not contain a letter X with the property that $X \overset{*}{\underset{G}{\Rightarrow}} \alpha X \beta X \gamma$ for some α,β,γ).

Theorem 16. The class of languages generated by nonexpansive ETOL systems strictly contains the class of languages generated by ETOL systems of finite index.

Thus to characterize ETOL languages of finite index by recursion one has to look for a stronger condition than expansiveness. A natural property is self embedding which in the case of context-free grammars separates regular from context-free languages. This however does not work because in an ETOL system the effect of the recursiveness of a symbol in a string is dependent also on how other symbols in the string are rewritten. This leads us to the notion of an actively recursive symbol.

Definition 9. Let $G = <V,\mathcal{P},S,\Sigma>$ be an ETOL system. A symbol X from V is called actively recursive (abbreviated a-recursive) if there exist x_1,x_2 in V^*, z in Σ^* and ρ in $\mathcal{P}^*$ such that

1) $S \overset{*}{\Rightarrow} x_1 X x_2 \overset{*}{\Rightarrow} z$,

2) $X \overset{\rho}{\Rightarrow} \alpha X \beta$, $x_1 \overset{\rho}{\Rightarrow} \bar{x}_1$, $x_2 \overset{\rho}{\Rightarrow} \bar{x}_2$ where $\#_{A(G)}(\alpha\beta) \geqslant 1$ and $\underline{\min}(\bar{x}_1 \alpha X \beta \bar{x}_2) \subseteq \underline{\min}(x_1 X x_2)$.

Definition 10. An ETOL system is called non actively recursive (abbreviated na-recursive) if it does not contain a-recursive letters.

It turns out that this kind of recursion allows one to characterize metalinear ETOL languages.

Theorem 17. A language is a metalinear ETOL language if and only if it is generated by a na-recursive ETOL system.

In order to characterize ETOL languages of finite index we have to find a weaker restriction than "na-recursive" or in other words (since we are using the negation of a notion of recursive) we have to find a stronger notion of the recursiveness of a symbol.

Definition 11. Let $G = <V,\mathcal{P},S,\Sigma>$ be an ETOL system, A symbol X from V is called lasting actively recursive (abbreviated la-recursive) if there exist x_1,x_2 in V^* , z in Σ^* and ρ in $\mathcal{P}^*$ such that

1) $S \overset{*}{\Rightarrow} x_1 X x_2 \overset{*}{\Rightarrow} z$,

2) $X \overset{\rho}{\Rightarrow} \alpha X \beta$, $x_1 \overset{\rho}{\Rightarrow} \overline{x}_1$, $x_2 \overset{\rho}{\Rightarrow} \overline{x}_2$ where $\#_{A(G)} (\alpha\beta) \geqslant 1$, $\underline{\min}(\overline{x}_1 \alpha X \beta \overline{x}_2) \subseteq \underline{\min}(x_1 X x_2)$, and moreover

3) there exists an active symbol Y such that $Y \in \underline{\min}(\alpha\beta)$ and $Y \overset{\rho}{\Rightarrow} \gamma Y \gamma_2$ where $\underline{\min}(\gamma_1 \gamma_2) \subseteq \underline{\min}(x_1 X x_2)$.

<u>Definition 12</u>. An ETOL system is called <u>not lasting actively recursive</u> (abbreviated nla-<u>recursive</u>) if it does not contain la-recursive letters.

<u>Theorem 18</u>. A language is an ETOL language of finite index if and only if it is generated by a nla-recursive ETOL system.

7. ON THE EFFECT OF FINITE INDEX RESTRICTION ON VARIOUS CLASSES OF LANGUAGE GENERATING DEVICES

In this section we investigate the effect of the finite index restriction on various classes of language generating devices. First of all we show that the class $\mathcal{L}(ETOL)_{FIN}$, the basic topic of our investigation in this paper, forms quite a central class of languages : one arrives at it from an amazingly large number of classes of language generating devices under the finite index restriction. In the following theorem we quote a number of better known grammars with this property.

<u>Theorem 19</u>. Under the finite index restriction the following classes of language generating systems produce exactly the class of ETOL languages of finite index :

1) the class of scattered grammars,
2) the class of context-free programmed grammars,
3) the class of unconditional transfer context-free programmed grammars,
4) the class of matrix grammars,
5) the class of ordered grammars,
6) the class of context-free grammars with regular control,
7) the class of state grammars, see [2] ,
8) the class of random context grammars, see [11] ,
9) the class of forbidding grammars, see [6] ,
10) the class of permitting grammars, see [11] ,
11) the class of ETIL systems.

Thus ETOL systems of finite index form quite a strong bridge between classical formal language theory and the theory of L systems. This is demonstrated even further by the fact that our technique of proving Theorem 19 allows us to carry on several results obtained for ETOL systems of finite index into other classes of language generating devices. For example we have the following results.

<u>Definition 13</u>. Let X be a class of language generating devices. We call X fi-<u>stable</u> if $\mathcal{L}(X)_{FIN} = \mathcal{L}(X)_{FINU}$.

Theorem 20. Each class of grammars listed in the statement of Theorem 19 is fi-stable.

Definition 14. Let X be a class of language generating devices. We call X pi-indifferent if $\mathcal{L}(X)_{FIN} = \mathcal{L}(PX)_{FIN}$, where PX is the subclass of X consisting of these devices in X that do not use erasing.

Theorem 21. Each class of grammars listed in the statement of Theorem 19 is pi-indifferent.

In our last result we compare the language generating power of the class of ETOL systems with that of several other classes of grammars under the finite index restriction. (In what follows IP denotes the class of Indian parallel grammars, see, e.g. [10], and CF denotes the class of context-free grammars).

Theorem 22. The following diagram holds :
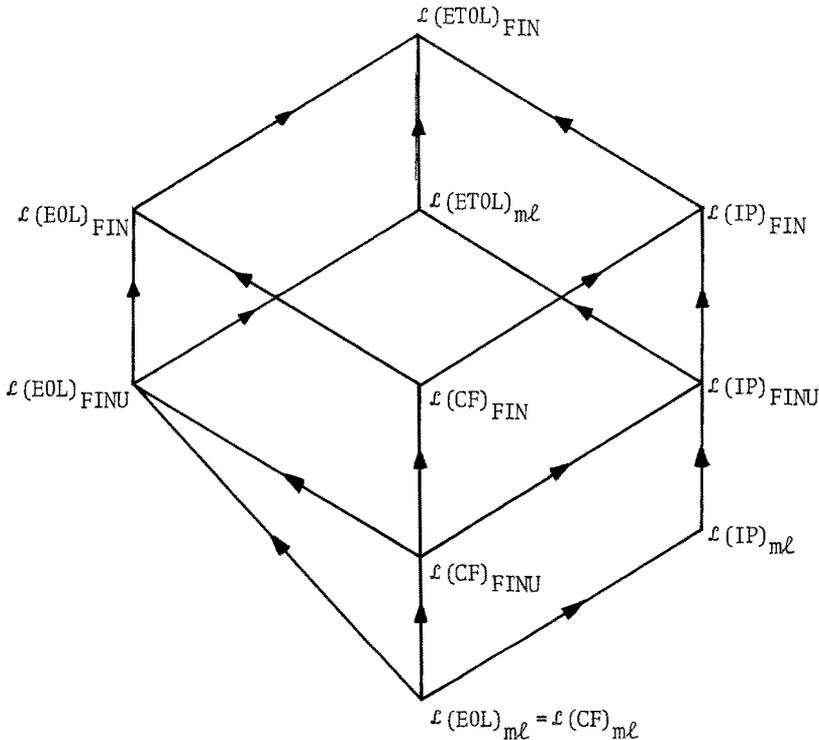
where if there is a directed chain of edges in the diagram leading from a class X to a class Y then $X \subsetneqq Y$; otherwise the classes X and Y are incomparable but not disjoint.

We hope that this paper convinces the reader that the theory of L systems of finite index is an interesting one, well motivated and that it constitutes

an important bridge between the classical formal language theory and the theory of L systems.

REFERENCES

[1] Herman, G.T. and Rozenberg, G., 1975. Developmental Systems and Languages, North Holland Publishing Company, Amsterdam.

[2] Kasai, T., 1970. An hierarchy between context-free and context-sensitive languages, Journal of Computer and System Science, 4, 492-508.

[3] Lindenmayer, A., 1968. Mathematical models for cellular interactions in development, Parts I and II, Journal of Theoretical Biology, 18 : 280-315.

[4] Lindenmayer, A., 1975. Developmental systems and Languages in their biological context, contribution to [1].

[5] Lindenmayer, A. and Rozenberg, G., 1976 (eds.). Automata, Languages, Development; at the crossroads of biology, mathematics and computer science, North Holland Publishing Company, Amsterdam.

[6] Penttonen, M., 1975. ETOL grammars and N grammars, Information Processing Letters, 4, 11-13.

[7] Rozenberg, G. and Salomaa, A., 1976. The mathematical theory of L systems, In : J.Tou (ed.), Advances in Information Systems Science, Plenum Press, New York, London, 161-206.

[8] Rozenberg, G., Ruohonen, K. and Salomaa, A., 1976. Developmental systems with fragmentation, International Journal of Computer Mathematics,

[9] Salomaa, A., 1973. Formal Languages, Academic Press, New York, London.

[10] Salomaa, A., 1974. Parallelism in rewriting systems, Lecture Notes in Computer Science, Springer Verlag, Heidelberg, No. 14 : 523-533.

[11] van der Walt, Random context languages, 1971. Proceedings of the IFIP Congress 1971, North Holland Publishing Company, Amsterdam, 66-68.