All correspondence to the second author

# ON ETOL SYSTEMS WITH RANK

by

A. Ehrenfeucht*

and

G. Rozenberg**

#CU-CS-117-77 September, 1977

*A. Ehrenfeucht, Department of Computer Science, University of Colorado at Boulder, Boulder, Colorado, 80309 USA

**G. Rozenberg, Department of Mathematics, University of Antwerp, U.I.A., Wilrijk, Belgium

WORKING PAPER

ABSTRACT

The notion of an ETOL system with rank is defined. It extends naturally studied already notions of an DOL system with rank and of an ETOL system of finite index. It turns out that in this way one gets an infinite hierarchy of classes of languages (each one being a full AFL) within the class of ETOL languages. This hierarchy starts with the class of ETOL languages of finite index and it fills in the class of nonexpansive ETOL languages. Some other properties of the class of ETOL systems with rank are also studied.

## 0. INTRODUCTION

ET0L systems and languages constitute certainly a central class among various classes of L systems and languages.

Recently some work has been done investigating the effect of the classical finite index restriction applied to ET0L systems. The results obtained so far (see, e.g., [4], [5], [6] and [7]) indicate that this class of systems is quite basic for the theory of L systems as well as for our understanding of differences between sequential and parallel rewriting systems.

In a sense this paper continues the research on ET0L systems of finite index. As a matter of fact it extends this notion so as to provide an insight into larger subfamily of the family of ET0L systems and languages. The starting point is an observation that one can extend the notion of rank of a D0L system (as introduced in [2]) to ET0L systems. In this way one gains a structural approach to ET0L systems which as it turns out is a generalization of ET0L systems in the following sense. While increasing rank in ET0L systems one obtains an infinite hierarchy of classes of languages which starts with the class of ET0L systems of finite index and which fills in the class of nonexpansive ET0L languages. In this sense the ET0L systems with rank play the same role as the context free grammars of finite index play in the theory of context-free languages!! It is instructive to observe that at the same time this hierarchy properly extends the hierarchy of D0L systems with rank in the sense that for each k the class of D0L languages of rank k contains languages that cannot be generated by ET0L systems of rank smaller than k.

The paper is organized as follows.

In Section I we introduce some basic definitions and notations. In Section II the notion of rank of an ET0L system is defined and a normal form theorem is proved. It is also shown that (unlike in the finite index case) the deterministic restriction on ET0L systems with rank is a proper one. In Section III we show that the notion of rank gives rise to an infinite hierarchy of classes of languages. Section IV provides characterization of both the class of ET0L languages with rank and the class of ET0L languages of rank 1. Finally we also examine some closure properties of the classes of languages considered.

NOTATION AND PRELIMINARIES

We assume the reader to be familiar with the rudiments of formal language theory, e.g. in the scope of [8] and with the basic notions of L systems, see, e.g.[3]. Now we will systematically list some basic definitions, notations and results to be used in the sequel.

(0) First of all, we do not distinguish between a singleton and its element. Thus the set $\{a\}$ will often be denoted as a.

(1) For a word x, we denote by $\underline{\min}\, x$ the set of symbols that occur in x, $|x|$ denotes the length of x.

(2) For an alphabet $\Delta$ and a word x, $\#_{\Delta}(x)$ denotes the number of occurrences of symbols from $\Delta$ in x.

(3) For an alphabet $\Sigma$ and a language L,
$\underline{\mathrm{Length}}_{\Sigma}(L)$ denotes the set$\{\#_{\Sigma}(x) : x \in L\}$,
$\underline{\mathrm{Length}}\ (L)$ is defined by $\underline{\mathrm{Length}}(L) = \{|x| : x \in L\}$.

(4) Let $\Delta$ be a subset of an alphabet V. Then the homomorphisms $\underline{\mathrm{Pres}}_{\Delta,V}$ and $\underline{\mathrm{Er}}_{\Delta,V}$ (denoted $\underline{\mathrm{Pres}}_{\Delta}$ and $\underline{\mathrm{Er}}_{\Delta}$ if V is understood) are defined as follows:

$$\underline{\mathrm{Pres}}_{\Delta,V}(A) = \begin{cases} A \text{ iff } A \in \Delta, \\ \Lambda \text{ iff } A \in V \setminus \Delta, \end{cases}$$

and

$$\underline{\mathrm{Er}}_{\Delta,V}(A) = \begin{cases} A \text{ iff } A \in V \setminus \Delta, \\ \Lambda \text{ iff } A \in \Delta. \end{cases}$$

(5) Let $G = \langle V, \mathcal{P}, S, \Sigma \rangle$ be an ETOL system.

(5.1) A symbol A in V is called <u>active</u> if there is a production in G of the form $A \to \alpha$ with $\alpha \neq A$. The set of active symbols in G is denoted by A(G) and the set $V \setminus A(G)$ of nonactive symbols in G is denoted by NA(G). G is said to be in <u>Active Normal Form</u> (abbreviated ANF) if $\Sigma \cap A(G) = \emptyset$.

(5.2) Let $\rho = T_1 \ldots T_m$ be a word in $\mathcal{P}^+$ and let $x \in V^*$. Then $\rho(x)$ denotes the set

$\{y : x \overset{T_1}{\Longrightarrow} x_1 \overset{T_2}{\Longrightarrow} x_2 \overset{T_3}{\Longrightarrow} \ldots \overset{T_n}{\Longrightarrow} y\}$; we also write $x \overset{\rho}{\Longrightarrow} y$ for every y in $\rho(x)$. For a language K, $\rho(K)$ is difined by $\rho(K) = \bigcup_{x \in K} \rho(x)$.

(5.3) <u>The deterministic version</u> $(G)_D$ of G is the unique EDTOL system $(G)_D = \langle V, \overline{\mathcal{P}}, S, \Sigma \rangle$ where $P \in \overline{\mathcal{P}}$ if and only if P is a homomorphism and $P \subseteq T$ for some T in $\mathcal{P}$. Let $\Delta$ be a subset of V. We say that G is <u>deterministic in</u> $\Delta$ if $\#T(a) = 1$ for every T in $\mathcal{P}$ and every a in $\Delta$.

(6) An ETOL system (context free grammar) G is of <u>index</u> k if for any word in the language of G, denoted L(G), there exists a derivation such that no intermediate word in this derivation contains more than k active symbols (nonterminals). We say that G is of <u>uncontrolled</u> <u>index k</u> if, for every word in L(G), every derivation of it is such that no intermediate word in this derivation contains more than k active (nonterminal) symbols. We say that G is of (<u>uncontrolled</u>) <u>finite index</u> if it is of (uncon-

trolled) index k for some k. We will use $L(\mathrm{ET0L})_{\mathrm{FIN}}(L(\mathrm{CF})_{\mathrm{FIN}})$ to denote the class of languages generated by ET0L systems (context free grammars) of finite index.

(7) An ET0L system $G = \langle V,P,S,\Sigma \rangle$ is in Finite Index Normal Form (abbreviated as FINF) if and only if it has the following properties:

(i) G is deterministic,

(ii) G is propagating,

(iii) G is in ANF, and

(iv) G is of uncontrolled finite index.

The following result was proved in [4]

<u>Theorem</u> 1. There exists an algorithm which, given an ET0L system G of finite index, produces an equivalent ET0L system H such that H is in FINF.

(8) Let $G = \langle V,P,S,\Sigma \rangle$ be an ET0L system.

(8.1) The relation $\underset{G}{<}$ (or $<$ if G is understood) on V is defined as follows. For any two symbols a and b from V, $a \underset{G}{<} b$ if $a \underset{G}{\Rightarrow} \alpha_1 b \alpha_2$ for some $\alpha_1, \alpha_2 \in V^*$.

(8.2) The relations $\underset{G}{\leqslant}$, $\underset{G}{\overset{*}{<}}$, $\underset{G}{\overset{*}{\leqslant}}$ (denoted $\leqslant$, $\overset{*}{<}$, $\overset{*}{\leqslant}$ if G is understood) are defined as the reflexive, transitive and reflexive and transitive closure of $\underset{G}{<}$ respectively.

(8.3) For an element a from V, [a] denotes the equivalence class of a with respect to $\underset{G}{\overset{*}{\leqslant}}$, i.e. $[a] = \{b \in V : a \underset{G}{\overset{*}{\leqslant}} b \underset{G}{\overset{*}{\leqslant}} a\}$. We use [V] to denote the set of all such classes.

(8.4) A letter a from V is called <u>recursive</u> if $a \overset{*}{\underset{G}{<}} a$.

(8.5) A letter a from V is called <u>useful</u> if $S \overset{*}{\underset{G}{<}} a$.

(9) Let $G = \langle V,\mathcal{P},S,\Sigma\rangle$ be an ET0L system.

(9.1) The <u>top-down level</u> is a function from [V] into nonnegative integers denoted by tl and defined inductively as follows:

(0) Let $A \in [V]$. Then $tl(A) = 0$ if $b \underset{G}{<} a$ for some $a \in A$ implies that $b \in A$.

(i+1). Let $Y_0 = [V]$. Let $Y_{i+1} = Y_i \setminus tl^{-1}(i)$. Then $tl(A) = i+1$ if $A \in Y_{i+1}$ and there exists a symbol b with $tl([b]) = i$ such that $b < a$ for some $a \in A$.

(9.2) The <u>bottom-up level</u> is a function from [V] into nonnegative integers denoted by bl and defined inductively as follows.

(0). Let $A \in [V]$ Then $bl(A) = 0$ if $a \underset{G}{<} b$ for some $a \in A$ implies that $b \in A$.

(i+1). Let $Y_0 = [V]$. Let $Y_{i+1} = Y_i \setminus bl^{-1}(i)$. Then $bl(A) = i+1$ if $A \in Y_{i+1}$ and $a \underset{G}{<} b$ for some $a \in A, b \notin A$ implies that $bl([b]) \leq i$.

(9.3) The top-down (bottom-up) level of a symbol from V is defined as the top-down (bottom-up) level of its class, i.e. $tl(a) = tl([a])$ and $bl(a) = bl([a])$ for all a in V.

(10). Let $G = \langle V,\mathcal{P},S,\Sigma\rangle$ be an ET0L system.

(10.1) The <u>success-language</u> of G, denoted <u>Succ</u>(G), is defined by <u>Succ</u>$(G) = \{x \in V^* : (\exists y)_{\Sigma^*}(x \overset{*}{\underset{G}{\Rightarrow}} y)\}$.

(10.1) For a word $\alpha$ from $V^*$ and a subset Z of V, we define the sets $\underline{SUCC}_{G,Z}(\alpha)$ and $\underline{NSUCC}_{G,Z}(\alpha)$, also denoted $\underline{SUCC}_{Z}(\alpha)$ and $\underline{NSUCC}_{Z}(\alpha)$ if G is understood, by

$\underline{SUCC}_{G,Z}(\alpha) = \{\underline{Pres}_Z(x) : \alpha \overset{*}{\underset{G}{\Rightarrow}} x \text{ and } x \in \underline{Succ}(G)\}$ and

$\underline{NSUCC}_{G,Z}(\alpha) = \{|\omega| : \omega \in \underline{SUCC}_{G,Z}(\alpha)\}$.

Hence the success-language of an ET0L system G is the set of all strings that can derive a terminal word. For an ET0L system G, an alphabet Z and a word $\alpha$, the set $\underline{SUCC}_{G,Z}(\alpha)$ is obtained from the set of all words in the success-language of G that can be derived from $\alpha$, by erasing all occurrences of symbols not belonging to Z.

## II. ET0L SYSTEMS WITH RANK

In this section we define the concept of rank of an ET0L system. This goes through the notion of the rank of a symbol in an ET0L system. Roughly speaking, a symbol is of rank 0 in some ET0L system G, if it can derive only a finite number of words which are in the success language of G. A symbol X is of rank 1 in G if only a finite number of words is obtained by erasing all occurrences of symbols of rank 0 in all the words in the success language of G which can be derived from X. This definition can then be extended inductively to any positive integer k. If, in this way we can assign a rank to every letter in the alphabet of G, then G is said to be of rank k where k is the rank of the axiom.

Formally it is defined as follows.

Definiton 1. Let $G = \langle V, P, S, \Sigma \rangle$ be an ET0L system.

(I) We define $\underline{rank}_G$ (or $\underline{rank}$ if G is understood) to be a (partial) function from V into the set of nonnegative integers defined inductively as follows.

(0). Let $Z_0 = V$. Then for a in V, $\underline{rank}_G(a) = 0$ if and only if $\underline{SUCC}_{G,Z_0}(a)$ is a finite set.

(i+1). Let $Z_{i+1} = V \setminus \{a \in V : \underline{rank}(a) \leqslant i\}$. Then for a in $Z_{i+1}$, $\underline{rank}_G(a) = i+1$ if and only if $\underline{SUCC}_{G,Z_{i+1}}(a)$ is a finite set.

(II). We say that G is an ET0L system with rank if and only if $\underline{rank}_G$ is a total function on V. Moreover we say that G is of rank m, denoted <u>rank</u> G = m, if every letter V is of rank not larger than m and at least one letter from V is of rank m.

We will use $R_i(G)$ (or $R_i$ if G is understood) to denote the set of all letters which have rank i in G. Also we will use $(\mathrm{ET0L})_{\mathrm{RAN}(i)}$ (respectively $(\mathrm{ET0L}_{\mathrm{RAN}})$ to denote the class of ET0L systems of rank not larger than i (respectively of ET0L systems with rank). As usual, $\mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}(i)})$ and $\mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}}$ denote the corresponding classes of languages.

It is useful to note here that this definition of rank, when restricted to the case of D0L systems, coincides with the corresponding definition of the rank of a D0L system from [2].

Our next result will provide a normal form for ET0L systems with rank which will be useful in the sequel. First we need a definiton.

<u>Definition</u> 2. Let $G = \langle V,P,S,\Sigma \rangle$ be an ET0L system of rank $i (i \geq 0)$. We say that G is in Rank Normal Form, abbreviated as RNF, if the following holds:

(i) <u>Succ</u>(G) = $V^*$,

(ii) G is in ANF,

(iii) G is propagating,

(iv) G is deterministic in $R_i(G)$, and

(v) $S \leqslant a$ for all $a \in V$.

Theorem 2. There exists an algorithm which, given an ET0L system G which is of rank i, will produce an equivalent ET0L system H of rank not bigger than i which is in RNF.

Proof.

Let G = V,P,S,Σ> be an ET0L system of rank i.

(1) First we note that the standard algorithm which, given an arbitrary ET0L system, produces an equivalent EPT0L system (see [3]) does not increase the rank of any symbol from V. Hence we can assume that G is propagating.

(2) Since the construction from Lemma 1 in [4] which produces an equivalent ET0L system in ANF does not increase the rank, we can assume that G is in ANF and thus, $\Sigma \subseteq R_0(G)$.

(3) One observes that the number of occurrences of symbols from $R_i(G)$ in an intermediate word of a successful derivation is bounded by some constant k. Hence a similar construction as the one used in the proof of Lemma 2, in [4] will yield an equivalent ET0L system $G' = \langle V',P',S',\Sigma\rangle$ which is deterministic in $R_i(G)$. Again it is clear that this construction does not increase the rank of G and also G' is again propagating and in ANF.

(4) Consider the set $U(G') = \{\underline{\min}\ x : S' \overset{*}{\underset{G'}{\Rightarrow}} x \overset{*}{\underset{G'}{\Rightarrow}} \omega \in \Sigma^*\}$ of useful alphabets of G'. (Note that, as it is shown in [7], this set can be effectively constructed.)

Let $V_H = \{X_\Delta : X \in \Delta$ for some $\Delta \in U(G') \cup \Sigma$ be a new alphabet. For every pair Δ,Δ' of useful alphabets and for every table T which is such that $T(a) \cap \Delta'^* \neq \emptyset$ for every $a \in \Delta$, we define a new table $T_{\Delta,\Delta'}$ as follows:

$T_{\Delta,\Delta'} = \{a_\Delta \to \alpha_{\Delta'} : a \overset{T}{\Rightarrow} \alpha,\ a \in \Delta,\ \alpha \in \Delta'^*\} \cup$
$\cup \{a_x \to a_x : x \neq \Delta\} \cup \{a \to a : a \in \Sigma\}.$

Also for every $\Delta \in U(G')$ which is such that $\Delta \subseteq \Sigma$, we define a table $T_\Delta = \{a_\Delta \to a : a \in \Delta\} \cup$
$\cup \{a_x \to a_x : x \neq \Delta\} \cup \{a \to a : a \in \Sigma\}.$

Let $P_H$ be the set of all new tables that can be defined in this manner.

Consider the ETOL system $H = \langle V_H, P_H, S_S, \Sigma \rangle$. It should be clear to the reader that H is equivalent to G' and also that $\underline{\text{Succ}}(H) = V_H^*$. Furthermore it is easily seen that H is both propagating and in ANF because the rank of H is not greater than the rank of G' and that H is deterministic in $R_i(H)$. Hence the theorem holds.

The following is an immediate corollary of the definition of Rank Normal Form.

Corollary 1. Let $G = \langle V, P, S, \Sigma \rangle$ be an ETOL system of rank m which is in RNF. Then $a \underset{G}{\leqslant} b$ implies that, $\underline{\text{rank}}_G(a) \geqslant \underline{\text{rank}}_G(b)$ and, consequently, $\underline{\text{rank}}(G) = \underline{\text{rank}}_G(S)$.

When looking at the definition of Rank Normal Form one notices that a system G of rank i which is in RNF, is deterministic in $R_i(G)$. An obvious question to ask then is whether this can be strengthened to complete determinism, i.e. is $L(\text{ETOL})_{\text{RAN}} = L(\text{EDTOL})_{\text{RAN}}$? The next theorem shows that the answer is negative.

<u>Theorem</u> 3. There exists a 0L language of rank 2 which is not an EDT0L language

<u>Proof</u>.

Let G = <{a,b,0,1},P,a> be a 0L system where $P = \{a \to ab, b \to b0,\ b \to b1, 0 \to 0, 1 \to 1\}$. Clearly G is of rank 2 and

$L(G) = \{a, ab\} \cup$

$\cup \{abbx_1bx_2bx_3\ldots bx_n : n \geq 1 \text{ and, for } 1 \leq i \leq m, x_i \in \{0,1\}^i\}$.

We will show that L(G) is not an EDT0L language.

(i) Let f be a function on positive integers defined by $f(x) = 4\log_2 x+1$. It was proved in [1] that $\{0,1\}^+$ contains infinitely many f-random words. Now if each f-random word x in $\{0,1\}^+$ is used to build a word $\bar{x}$ in L(G) in such a way that $x = x_1x_2\ldots x_n$ and $\bar{x}$ is the longest prefix of $abbx_1bx_2\ldots bx_n$ that is in L(G), then we say that L(G) contains infinitely many words that are $4\log_2$- random.

(ii) Let us assume to the contrary that L(G) is an EDT0L language. Let g be a function on the positive integers defined by $g(x) = 4\log_2(x)$. Then by Theorem from [1] it follows that there exist positive integers s and t such that for every g-random word y from L(G) that is larger than s there exist words $y_0, \ldots y_t, \sigma_0, \ldots \sigma_t$ with $\sigma_0 \ldots \sigma_t \neq \Lambda$ such that $y = y_1 \ldots y_t$ and for every positive integer m, $\sigma_0^m y_1 \sigma_1^m\ y_2 \ldots y_t \sigma_t^m$ is in L(G).

By (i) L(G) contains g-randon words longer than s; let us consider such a word y. Let $\sigma_i$ be a nonempty word satisfying the quoted above theorem. There are three cases to consider.

(1) $a \varepsilon \underline{\min}\, \sigma_i$. Then L(G) would contain words with more than one occurrence of a; a contradiction.

(2) $b \varepsilon \underline{\min}\, \sigma_i$. Then L(G) would contain a word with a subword $b_1 z_1 b z_2 b$ where $z_1 z_2 \varepsilon \{0,1\}^*$ and $|z_1| = |z_2|$; a contradiction.

(3) $\sigma_i \varepsilon \{0,1\}^+$. Then L(G) would contain words with n occurrences of b and more than $\sum_{i=1}^{m-1} i$ occurrences of letters from $\{0,1\}$; a contradiction.

As each of the three possible cases yields a contradiction, L(G) cannot be an EDTOL language.

## III. AN INFINITE HIERARCHY IN $L(\text{ET0L})_{\text{RAN}}$

In this section we will show that ET0L systems with rank do not exhaust all ET0L systems, i.e. $L(\text{ET0L})_{\text{RAN}} \subsetneq L(\text{ET0L})$. Furthermore, we will also prove that $L(\text{ET0L})_{\text{RAN}(i)} \subsetneq L(\text{ET0L})_{\text{RAN}(i+1)}$ for every $i \geqslant 0$, thus establishing an infinite hierarchy in between $L\text{FIN} = L(\text{ET0L})_{\text{RAN}(1)}$ and $L(\text{ET0L})_{\text{RAN}}$.

The proof of this goes through a sequence of lemmas containing some results on D0L systems which are interesting on their own.

First we need some definitions.

Definition 3. Given a homomorphism $\delta$ from an alphabet V into V* we define $\mu_\delta$ to be a function from V into $2^V$ defined by $\mu_\delta = \underline{\min}\,\delta(a)$. The function $\mu_\delta$ is extended to the function $\bar{\mu}_\delta$ from $2^V$ into $2^V$ by

$$\bar{\mu}_\delta(Z) = \bigcup_{a \in Z} \mu_\delta(a).$$

As usual, to avoid cumbersome notation we will use the same symbol $\mu_\delta$ to denote both $\mu_\delta$ and its extension $\bar{\mu}_\delta$. Since we often identify in notation a singleton set and its element, then we get that, for example, $\mu_\delta(a)$ denotes also $\bar{\mu}_\delta(\{a\})$.

Definition 4. Let $G = \langle V,\delta,\omega\rangle$ be a D0L system with $E(G) = \omega_0,\omega_1,\ldots$ We say that G is instant if

1. $\underline{\min}(\omega_i) = \underline{\min}(\omega_j)$ for all $i,j \geqslant 0$, and
2. $\mu_\delta = \mu_{\delta^2}$.

Lemma 1. Let G be an instant D0L system. Then $\underset{G}{<} = \underset{G}{\overset{*}{<}}$.

Proof.

Let $G = \langle V, \delta, \omega_0 \rangle$ be as in the statement of the lemma. Let a,b and c $\varepsilon$ V be such that $a < b < c$. Hence $b \ \varepsilon\ \mu_\delta(a)$ and $c \ \varepsilon\ \mu_{\delta^2}(a)$. Then, by definition

$c \ \varepsilon\ \mu_\delta(a) = \mu_{\delta^2}(a)$ and thus $a < c$. Hence $\underset{G}{<}$ is transitive.

From Lemma 1 we can conclude immediately the following.

Lemma 2. Let G = $V, \delta, \omega_0\rangle$ be an instant D0L system. Then $tl(a) \leqslant 1$ for every a in V.

The following lemma shows that for an instant D0L system G and a given symbol a, the function $f_a : \mathbb{N}_0 \to \mathbb{N}$, which associates with n number of occurrences of a in the (n+1)'th word of the sequence of G, is either strictly increasing or constant.

Lemma 3. Let $G = \langle V, \delta, \omega_0 \rangle$ be an instant D0L system. then, for every a in V

either $\#_a(\omega_m) < \#_a(\omega_{m+1})$ for all $m \geqslant 1$,

or $\#_a(\omega_m) = \#_a(\omega_{m+1})$ for all $m \geqslant 1$.

Proof.

The lemma is trivial if a is not useful. Hence, in the sequel, we can always assume that every letter is useful. By Lemma 2 we know that $tl(a) \leqslant 1$ for every a in V. First we show that the lemma holds for letters a of top-down level 0.

(1) Let $a \in V$ be such that $tl(a) = 0$, hence $a \in \underline{min}\ \omega_0$. We clain that a is recursive. Indeed, if a would not be recursive then $\underline{min}\ \omega_0$ must contain a letter $b \neq a$ such that $a \in \underline{min}\ \delta(b)$, or $b < a$ and, since $a \nless b$, $tl(a) \neq 0$, a contradiction.

Since the $tl(a) = 0$, we know that

$$\#_a(\omega_{n+1}) = \sum_{b \in [a]} \#_b(\omega_n) \cdot \#_a(\delta(b)) \text{ for every } n \geqslant 0.$$

Since a is recursive, this implies that the function $f_a(m) = \#_a(\omega_n)$, $m > 1$ is nondecreasing, moreover $f_a$ is either strictly increasing or constant with the latter happening only if $\#[a] = 1$ and $\#_a\delta(a) = 1$. Hence the lemma is true for letters of top-down level 0.

(2) To show that the lemma holds for letters of top-down level 1, let us divide the set $\{a \in V : tl(a) = 1\}$ into subclasses $C_0, C_1, \ldots$ as follows, Let $C_0$ be the set of all letters a of top-down level 1 which are such that $b < a$ and $b \notin [a]$ implies that $tl(b) = 0$. For every $i \geqslant 0$, let $C_{i+1}$ be the set of all letters a in $tl^{-1}(1) \setminus \bigcup_{j=0}^{i} C_i$ which are such that $b < a$ and $b \notin [a]$ implies that $b \in \bigcup_{j=0}^{i} C_j$.

Next we show, by induction on i, that the lemma is true for all letters in $\bigcup_{j=0}^{i} C_j$.

(2.7) Let a be a letter in $C_0$. Then for every $m \geqslant 1$, $\#_a(\omega_{m+1}) = \text{Term}_1 + \text{Term}_2$ where

$$\text{Term}_1 = \sum_{b \in [a]} \#_b(\omega_m) \cdot \#_a \delta(b) \text{ and}$$

$$\text{Term}_2 = \sum_{b \in A} \#_b(\omega_m) \cdot \#_a \delta(b),$$

where A is the set of all letters b in $tl^{-1}(0)$ which introduce an occurrence of a.

We consider two cases.

(i) a is recursive. Then $f_a(m) = \#_a(\omega_m)$ is a non-increasing function and either it is strictly growing or constant with the latter happening only if $\text{Term}_2 = 0$, $\#[a] = 1$ and $\#_a \delta(a) = 1$.

(ii) a is not recursive. Then $[a] = \{a\}$ and consequently, $\text{Term}_1 = 0$. Since the lemma holds for letters in $tl^{-1}(0)$, it follows that $\text{Term}_2$ yields either a strictly growing of a constant function.

(2.2) Assume that the lemma holds for letters in $\bigcup_{j=0} C_i$.

(2.3) Let a be a letter in $C_{i+1}$. Then, for $m \geqslant 1$, $\#_a(\omega_{m+1}) = \text{Term}_1 + \text{Term}_2$ where

$$\text{Term}_1 = \sum_{b \in [a]} \#_b(\omega_m) \cdot \#_a \delta(b) \text{ and}$$

$$\text{Term}_2 = \sum_{b \in A} \#_b(\omega_m) \cdot \#_a \delta(b)$$

where A is the set of all letters in $\bigcup_{j=0}^{i} C_j$ which introduce an occurrence of a. By an argument as in (2.1) we can show that the lemma holds for a.

This completes the induction and hence the lemma holds.

We will refer to the letters satisfying the first condition from the statement of the above lemma as <u>dynamic</u> letters and to the letters satisfying the second condition as <u>static</u> letters.

<u>Lemma</u> 4. Let $G = \langle V, \delta, \omega \rangle$ be an instant D0L system of rank k. Then for every a in V, there exists a polynomial $g_a$ of degree not larger than k such that, for every positive integer n, $\#_a(\omega_n) = g_a(n)$.

<u>Proof</u>.

The proof goes by induction on the rank of G.

(1) If <u>rank</u> G = 0 then the lemma follows immediately from Lemma 3.

(2) Let us assume that the lemma holds if <u>rank</u> G $\leqslant$ k-1

(3) Let <u>rank</u> G = k.

Let us reduce the rank of G by erasing all letters (and productions for them) of rank 0. Let $G_0$ be the so obtained system. By the inductive assumption the lemma holds for all the letters from $G_0$. Thus to complete the proof we have to compute the number of occurrences of the omitted letters.

First we divide the letters from $R_0(G)$ into categories as follows.

Let a be in $R_0(G)$.

a is of <u>category</u> 0 if $\delta(a) = \Lambda$.

a is of <u>category</u> 1 if $\delta(a) = \alpha_1 a \alpha_i$ where $\delta(\alpha_1 \alpha_2) = \Lambda$.

For $i \geqslant 1$, a is of <u>category</u> (i+1) if $\mu_\delta(a)$ contains a letter of category i and every other letter in $\mu_\delta(a)$ is either a itself or a letter of category not larger than i. Let $\underline{Cat}_i(G)$ denote the set of letters from $R_0(G)$ of category i. We observe the following.

(i) If $a \in R_0(G)$ then a is recursive if and only if $a \in \underline{Cat}_1(G)$. This is proved as follows. Obviously, if $a \in \underline{Cat}_1(G)$ then a is recursive. On the other hand, if a is recursive then $\delta(a) = a_1 a \alpha_2$ and because $a \in R_0(G)$ and G is instant, for every b in $\underline{min}\ \alpha_1\alpha_2$, $\delta(b) = \Lambda$.

(ii) For $i \geqslant 3$, $\underline{Cat}_i(G) = \emptyset$. This is proved as follows. Assume, to the contrary, that $a \in \underline{Cat}_3(G)$. Then by (i) a is not recursive and $\mu_\delta(a)$ contains a letter, say b, of category 2. Again by (i), $\mu_{\delta^2}(a)$ contains only letters from $\underline{Cat}_1(G) \cup \underline{Cat}_0(G)$ and so $b \notin \mu_{\delta^2}(a)$. But then $\mu_\delta(a) \neq \mu_{\delta^2}(a)$ which contradicts the fact that G is instant.

It is instructive to note here that (ii) cannot be strengthened since, for example, for the instant D0L system

$G = \langle\{a,b,c,d\}, \delta, abcd\rangle$ with $\delta(d) = dabc$, $\delta(c) = \Lambda$, $\delta(b) = bc$ and $\delta(a) = bc$ we have $R_0(G) = \{a,b,c\}$, $\underline{Cat}_0(G) = \{c\}$, $\underline{Cat}_1(G) = \{b\}$ and $\underline{Cat}_2(G) = \{a\}$.

(iii) If $a \in \underline{Cat}_1(G)$ and $b < a$ for some $b \neq a$, $b \in R_0(G)$, then $b \in \underline{Cat}_2(G)$.

This follows immediately from (i) and the fact that b is in $R_0(G)$.

Now let us proceed to compute the number of occurrences of letters from $R_0(G)$ in $\omega_1, \omega_2, \ldots$

(3.1) Let $a \in \underline{Cat}_2(G)$.

Thus, by (i), a is not recursive and so, for $n \geq 1$,

$$\#_a(\omega_{n+1}) = \sum_{\substack{b \leq a \\ G}} \#_b(\omega_n) \cdot \#_a\delta(b).$$

By the inductive assumption, the above expression yields a sum of polynomials of degree not larger than k-1 and so it yields a polynomial of degree not larger than k-1.

(3.2) Let $a \in \underline{Cat}_1(G)$.

Thus, because of (i), a is recursive and so, for $n \geqslant 1$,

$$\#_a(\omega_{n+1}) = \#_a(\omega_n) + \sum_{\substack{b \leq a \\ G \\ b \neq a}} \#_b(\omega_n) \cdot \#_a\delta(b).$$

By the inductive assumptions the above recursive expression yields a polynomial of degree not larger than k.

(3.3) Let $a \in \underline{Cat}_0(G)$.

Again we get a formula as for letters in $\underline{Cat}_2(G)$:

$$\#_a(\omega_{n+1}) = \sum_{\substack{b \leq a \\ G}} \#_b(\omega_n) \cdot \#_a\delta(b),$$

where now the summation may involve polynomials of degree k (letters b from $\underline{Cat}_1(G)$ and so it yields a polynomial of degree not larger than k.

This ends the proof of the lemma.

Lemma 5. Let $G = \langle V, \delta, \omega \rangle$ be an instant D0L system of rank k and let $\psi$ be a homomorphism on V. Then $\underline{Length}\,\psi(L(G))$ is the range of a polynomial of degree not larger than k. Moreover, $\underline{Length}\,\psi(L(G))$ is infinite if and only if V contains a dynamic letter b such that $\psi(b) \neq \Lambda$.

Proof.

For each $a \in V$, let $g_a$ be the polynomial from Lemma 4 associated with a. Then, for $n \geq 1$,

$$|\psi(\omega_n)| = \sum_{a \in V} |\psi(a)| \cdot g_a(n).$$

The second part of this lemma follows from Lemma 3.

Theorem 4. Let $K \subseteq \Sigma^*$ be an ETOL language of rank m. Let $\Sigma_1 \subseteq \Sigma$ be such that $\underline{\text{Length}}_{\Sigma_1}(K)$ is infinite. Then there exists a strictly growing polynomial f of degree not larger than m, such that $\underline{\text{Range}}\, f \subseteq \underline{\text{Length}}_{\Sigma_1}(K)$.

Proof

Let K and $\Sigma_1$ be as in the statement of the Theorem. Obviously, $L(\text{ETOL})_{\text{RAN}(m)}$ is closed under homomorphism and thus it follows that $L = \underline{\text{Pres}}_{\Sigma_1}(K)$ is an infinite ETOL language of rank not larger than m. Let G be an ETOL system of rank i, $i \leq m$, generating L. Be Theorem 2, we can assume that G is in RNF. Let $(G)_D = \langle V, P, S, \Sigma_1 \rangle$ be the deterministic version of G. Clearly $(G)_D$ is also in RNF and, since L is infinite, $L((G)_D)$ is infinite. Hence there exists a derivation D of a word $\omega \in L$ such that there are intermediate words x and y in D with $\underline{\min}\, x = \underline{\min}\, y = \Delta$ and $|y| > |x|$.

Thus we have that $S \overset{\mu}{\underset{(G)_D}{\Longrightarrow}} x \overset{\rho}{\underset{(G)_D}{\Longrightarrow}} y \overset{\nu}{\underset{(G)_D}{\Longrightarrow}} \omega$ for some control words $\mu, \rho$ and $\nu$ in $P^*$.

It is then not difficult to see that for every a in $\Delta$ there exist positive integers $r_a$ and $p_a$ such that

$\underline{\min} \rho^{r_a}(a) = \underline{\min} \rho^{r_a + p_a}(a)$. Let $k_a$ be the smallest integer such that $n_a = k_a p_a > r_a$. Thus $\underline{\min} \rho^{n_a}(a) = \underline{\min} \rho^{2n_a}(a)$.

Let $n = \prod_{a \in \Delta} n_a$ and define a homomorphism $\delta$ on $\Delta^*$ by $\delta(a) = \rho^n(a)$ for every a in $\Delta$. It should then be clear to the reader that $H = \langle \Delta, \delta, x \rangle$ is an instant D0L system which is obvioulsy of rank not larger than m. Also $\Delta$ contains dynamic letters (in H) since $|\delta(x)| > |x|$. Let $\psi$ be the non-erasing homomorphism on $\Delta^*$ defined by $\psi(a) = \nu(a)$ for every a in $\Delta$.

The theorem then follows from lemma 5 and the fact that $\underline{\text{Length}}\, \psi(L(H)) \subseteq \underline{\text{Length}}\, L = \underline{\text{Length}}_{\Sigma_1}(K)$.

From the previous theorem, the main results of this section follow now easily.

Corollary 2. $K = \{a^{2^n} : n \geq 0\} \notin \mathcal{L}(\text{ET0L})_{\text{RAN}}$.

Proof.

By Theorem 4, K cannot be generated by an ET0L system with rank.

Corollary 3. For every $i \geq 0$,

$\mathcal{L}(\text{ET0L})_{\text{RAN}(i)} \subsetneq \mathcal{L}(\text{ET0L})_{\text{RAN}(i+1)}$.

Proof.

Let i be a nonnegative integer. Define an alphabet

$V = \{a_1, \ldots a_{i+2}\}$ and a homomorphism $\delta$ on V by $\delta(a_j) = a_j a_{j+1}$ for all $1 \leq j \leq i+1$ and $\delta(a_{i+2}) = a_{i+2}$. Consider the D0L system $G = \langle V, \delta, a_1 \rangle$.

Clearly, $K = L(G) \in \mathcal{L}(\text{ET0L})_{\text{RAN}(i+1)}$ but Length K is infinite and Length K = Range f where f is a polynomial of degree (i+1). Hence by Theorem 4, $K \notin \mathcal{L}(\text{ET0L})_{\text{RAN}(i)}$.

# IV. SOME CHARACTERIZATION RESULTS AND CLOSURE PROPERTIES

We start this section by proving two characterization results : the first one characterizes the class $L(\mathrm{ET0L})_{\mathrm{RAN}(1)}$ of ET0L languages of rank 1, the other provides a (partial) 'structural' characterization of the class $L(\mathrm{ET0L})_{\mathrm{RAN}}$ of ET0L languages with rank.

As it is noted in the introduction, the idea of rank can be regarded as an extension of the concept of finite index. Indeed, one observes that, given an ET0L system G, the set $R_0(G)$ of symbols which have rank 0 in G contains the set NA(G) of nonactive symbols in G. Moreover if follows from the definition of $R_0(G)$ that even if a symbol is of rank 0 but also active, it is still "not very active" since it can change in only a finite number of words during a successful derivation. Hence the notion of a symbol of rank 0 can be regarded as an extension of the notion of a nonactive symbol.

On the other hand one could give an equivalent definition of an ET0L system of (uncontrolled) finite index as follows.

An ET0L system G is of uncontrolled finite index if the set obtained by erasing all occurrences of nonactive symbols from all the words in the success language of G that can be derived from the axiom is finite.

This definition closely resembles the definition of a symbol of rank 1 since one must only replace "nonactive symbols" by "symbols of rank 0" and "axiom" by "symbol". (As a matter of fact one could say informally that a symbol X is of rank i+1 if its "associated" system, take X as the new axiom, is of uncontrolled finite index "in the set of all symbols with rank not smaller than i+1". (That is we consider all other symbols as "not (very) active").

The following theorem shows that as far as classes of languages are concerned, this resemblance is translated as identity.

Lemma 6. $\mathcal{L}(\mathrm{ET0L})_{\mathrm{FIN}} \subseteq \mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}(1)}$.

Proof.

Let $K \in \mathcal{L}(\mathrm{ET0L})_{\mathrm{FIN}(k)}$ for some $k \geqslant 0$. Then by Theorem 1 there exists an ET0L system G in FINF such that L(G) = K. Since G is in Active Normal Form, every terminal in G is rewritten as itself only and so all terminals are of rank 0 in G. Since G is of uncontrolled finite index, if a is a nonterminal and $u \in \underline{\mathrm{SUCC}}_{G,Z_0}(a)$, then $|u| \leqslant k$. Consequently every nonterminal is of rank at most 1. Thus G is of rank 1 and the lemma holds.

Lemma 7. $(\mathrm{ET0L})_{\mathrm{RAN}(1)} \subseteq (\mathrm{ET0L})_{\mathrm{FIN}}$

Proof

Let $K \in \mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}(1)}$ and let $G = \langle V,\mathcal{P},S,\Sigma\rangle$ be an ET0L system of rank nor larger than 1 generating K. Since the lemma trivially holds if G has rank 0, we can assume that the rank of G is 1. By Theorem 1 we can also assume that G is in RNF. First we notice the following

(i) First of all, because G is in RNF, every element of $\Sigma$ is of rank 0.

(ii) There exists a positive integer k such that if $S \underset{G}{\Rightarrow} \omega_1, \underset{G}{\Rightarrow} \cdots \underset{G}{\Rightarrow} \omega_t$ is a useful derivation in G, then for $1 \leq i \leq t$, $\#_{R_1}(\omega_i) \leqslant k$. (This follows immediately from the definition of the rank of a letter.)

We shall now prove the lemma by constructing an ET0L system H of index k+1 which is equivalent to G.

Let K be the set $K = \{\underline{SUCC}_{G,V}(x) : x \in R_0(G)^*, |x| \leqslant \underline{maxr}\ G\}$ for every x in K, let $L_x$ be defined by $L_x = \{y \in \Sigma^* : x \overset{*}{\underset{G}{\Rightarrow}} y\}$ and let TRACK be the set of all subsets of $(K \cup \{¢\}) \times \bigcup_{x \in K} L_x$, where ¢ is a new symbol.

(Note that it follows from the assumptions that TRACK is a finite set.)

Let $\psi$ be a finite substitution from V defined by

$$\psi(a) = \begin{cases} \{a\} & \text{if } a \in R_1(G), \text{ and} \\ L_a & \text{if } a \in R_0(G) \end{cases}$$

Also, let $\underline{MARK} = \{[x,X] : x \in \bigcup_{t \leqslant k} V^t \cup \{¢\}, X \in \underline{TRACK}\}$ be a new alphabet and let \$ be a new symbol.

For every $x = A_1 \ldots A_t \in R_1(G)^* (t \in k)$ and $T \in P$ such that for $1 \leqslant i \leqslant t$, $T(A_i) = \alpha_{i,0} B_{i,1} \alpha_{i,1} \ldots B_{i,n_i} \alpha_{i,n_i}$ where

$\alpha_{i,j} \in R_0(G)^* (0 \leqslant i \leqslant t, 0 \leqslant j \leqslant n_i)$,

$B_{i,j} \in R_1(G) (1 \leqslant i \leqslant n, 1 \leqslant j \leqslant n_i)$ for some $n_1, \ldots n_t$

such that $\sum_{i=1}^{t} n_i \leqslant k$ and for every set

$\Omega = \{<\alpha_{i,j},\beta_{i,j}> : 0 \leqslant i \leqslant t, 0 \leqslant j \leqslant m_i\}$ which is such that $\beta \in \psi(\alpha)$ for every $<\alpha,\beta>$ in $\Omega$, we construct a new table $T_{(x;\Omega)}$ as follows.

(1) $A_i \to \beta_{i,0} B_{i,1} \beta_{i,1} \ldots B_{i,n_i} \alpha_{i,n_i}$ is in $T_{(x;\Omega)}$ for every $1 \leqslant i \leqslant t$.

(2) $a \to a$ is in $T_{(x;\Omega)}$ for all a in $\Sigma$.

(3) Let y $\in$ <u>TRACK</u>.

Then $[x,Y] \to [\underline{Pres}_{R_1(G)}(T(x)),Z]$ is in $T_{(x;\Omega)}$ for every Z in <u>TRACK</u> which is such that:

(3.1) $\Omega \subseteq Z$, and

(3.2) for every $<\alpha,\beta> \in Y$, there exists a word $\alpha' \in (T(\alpha) \cap \underline{SUCC}_{G,V}(\alpha)) \cup \{¢\}$ such that $<\alpha',\beta> \in Z$.

(4) $X \to ¢$ is in $T_{(x;\Omega)}$ for every X in <u>MARK</u> $\cup (V \setminus \Sigma) \cup \{\$,¢\}$.

Finally, we construct a special final table $T_{fin}$ as follows:

(1) $[\Lambda,Y] \to \Lambda$ is in $T_{fin}$ for every Y from <u>TRACK</u> which is such that $<\alpha,\beta> \in Y$ implies $\alpha = \beta$.

(2) $X \to ¢$ is in $T_{fin}$ for every X in <u>MARK</u> $\cup (V \setminus \Sigma) \cup \{¢\}$

(3) $\$ \to S[S,\emptyset]$ is in $T_{fin}$.

(4) $a \to a$ is in $T_{fin}$ for every a in $\Sigma$.

Let $P'$ be the set of all newly defined letters.

Consider the ET0L system $H = <V \cup \{¢,\$\} \cup \underline{MARK}, P',\$,\Sigma>$.

It follows from the construction that H is of (uncontrolled) index k.

H simulates G as follows.

Let $D : S = x_0 \overset{T_1}{\underset{G}{\Longrightarrow}} x_1 \overset{T_2}{\underset{G}{\Longrightarrow}} \cdots \overset{T_n}{\underset{G}{\Longrightarrow}} x_n \in \Sigma^*$ be a successful derivation in G.

Then D will be simulated by a successful derivation

$D' : \$ \overset{T_{fin}}{\underset{H}{\Longrightarrow}} S[S,\emptyset] = y_0 \underset{H}{\Longrightarrow} \cdots \underset{H}{\Longrightarrow} y_n \overset{T_{fin}}{\underset{H}{\Longrightarrow}} x_n$ which is such that $\underline{Pres}_{R_1(G)}(x_i) = \underline{Pres}_{R_1(G)}(y_i) = Z_i$ for every $1 \leqslant i \leqslant n$.

Moreover, a step $x_{i-1} \overset{T_i}{\underset{G}{\Longrightarrow}} x_i$ from D will be simulated in the step $y_{i-1} \overset{(T_i)(z_{i-1},\Omega_i)}{\underset{H}{\Longrightarrow}} y_i$ where $y_i$ is such that all newly introduced occurrences (in $x_i$) of symbols of rank 0 are immediately replaced by their descendent-words in $x_n$ ($\Omega_i$ contains this particular set of "replacements"). Also the marker (at the right hand-side of $y_{i-1}$ and $y_i$) is changed in such a way that it contains these new replacements (3.1) from the definition of $T_{(x;\Omega)}$) and it also keeps track of the "evolution" of the earlier "replacement-guesses" ((3.2) in the definition of $T_{(x;\Omega)}$). Note that the marker disappears only if all the "guessed" replacements turned out to be correct ((1) in the definition of $T_{fin}$).

We leave to the reader the formal proof of the fact that $L(H) = L(G)$. Hence the lemma holds.

The following theorem is an immediate consequence of Lemma 6 and Lemma 7.

Theorem 5. $L(\mathrm{ET0L})_{\mathrm{RAN}(1)} = L(\mathrm{ET0L})_{\mathrm{FIN}}$.

Our next result gives a structural characterization of the class of all languages generated by ET0L systems with rank.

First we need a definition.

Definition 5. Let $G = \langle V, P, S, \Sigma \rangle$ be an ET0L system We call G expansive if $a \overset{*}{\underset{G}{\Rightarrow}} x_0 a x_1 a x_2$ for some $a \in V$, $x_0$, $x_1$ and $x_2$ in $V^*$. If G is not expansive then it is called nonexpansive.

The class of nonexpansive ET0L systems will be denoted by $(\mathrm{ET0L})_{\mathrm{NE}}$; as usual $L(\mathrm{ET0L})_{\mathrm{NE}}$ denotes the corresponding class of languages.

Theorem 6. Every nonexpansive ET0L system has a rank.

Proof.

Let $G = \langle V, P, S, \Sigma \rangle$ be a nonexpansive ET0L system. Let us consider the relation $\underset{G}{\leqslant}$. We will prove by induction on the bottom-up level of an equivalence class of $\underset{G}{\leqslant}$ that every letter a in V has a rank not larger than bl(a).

(1) Let A be an equivalence class of bottom-up level 0. We consider two cases.

(1.1) A contains only one element a and a is always rewritten in G as $\Lambda$. Then clearly rank (a) = 0.

(1.2) A contains at least one recursive letter (and hence all elements of A are recursive letters.)

Let $a \in A$. Whenever $a \underset{G}{\Rightarrow} z_1 \underset{G}{\Rightarrow} \dots \underset{G}{\Rightarrow} z_t$ is a successful derivation from a, then, for $1 \leqslant i \leqslant t, \#_A(z_i) \leqslant \#A$ (otherwise for b in A we would have $a \overset{*}{\underset{G}{\Rightarrow}} b\beta b\gamma$ for some $\alpha, \beta, \gamma \in V^*$ and consequently $a \overset{*}{\underset{G}{\Rightarrow}} \bar{\alpha} a \bar{\beta} a \bar{\gamma}$ for some $\bar{\alpha}, \bar{\beta}, \bar{\gamma} \in V^*$; a contradiction).

Thus <u>rank</u> (a) = 0.

(2) Let us assume that every letter from an equivalence class of bottom-up level i has a rank which is not larger than i.

(3) Let A be an equivalence class of bottom-up level i+1. We consider two cases.

(3.1) A contains only one element a and a is always rewritten as a word consisting only of letters of bottom-up level not larger than i. It then follows from the inductive assumption that a has a rank not larger than i.

(3.2) A contains at least one recursive letter (hence all elements of A are recursive).

Let $a \in A$. Again (as in 1.2), whenever $a \underset{G}{\Rightarrow} z_1 \underset{G}{\Rightarrow} \dots \underset{G}{\Rightarrow} z_t$ is a successful derivation from a, then, for $1 \leqslant i \leqslant t, \#_A(z_i) \leqslant \#A$. From this and the inductive assumption it follows that $\underline{SUCC}_{Z_{i+1}}(a)$ is a finite (or empty) set (where $Z_{i+1}$ is as in Definition 1) and consequently <u>rank</u> (a) $\leqslant$ i+1.

Thus every letter from V has a rank, which implies that G has a rank and so the theorem holds.

It is instructive to note that, with the present definition of nonexpansiveness, the above theorem cannot be strengthened into an if and only if result. For example $G = <\{a,b,c\},\{a \to aab, b \to c, c \to c\}, a, \{a,b\}>$ is an expansive E0L system of rank 0. However if one adopts the definition from [7] of nonexpansiveness, then Theorem 6 can be made into if and only if, using a similar proof as in the next theorem, which shows that even with the present definition of nonexpansiveness, the classes $L(\mathrm{ET0L})_{\mathrm{NE}}$ and $L(\mathrm{ET0L})_{\mathrm{RAN}}$ are the same.

Theorem 7. $L(\mathrm{ET0L})_{\mathrm{RAN}} = L(\mathrm{ET0L})_{\mathrm{NE}}$.

Proof.

It follows from Theorem 6 that $L(\mathrm{ET0L})_{\mathrm{NE}} \subseteq L(\mathrm{ET0L})_{\mathrm{RAN}}$.

To show the other inclusion, let L be in $L(\mathrm{ET0L})_{\mathrm{RAN}(i)}$ for some $i \geqslant 0$ and let $G = <V, P, S, \Sigma>$ be an ET0L system of rank $k (k \leqslant i)$ such that $L(G) = L$. By Theorem 1, we can assume that G is in RNF. We will show that G is non-expansive.

Assume the contrary, i.e. $a \stackrel{*}{\Rightarrow} x_0 a x_1 a x_2$ for some $a \in V$, $x_0, x_1$ and $x_2$ in $V^*$. It follows that $\underline{\mathrm{Pres}}_a\{x : a \stackrel{*}{\Rightarrow} x\}$ is an infinite language. Together with the fact that $\underline{\mathrm{Succ}}(G) = V^*$, this implies that a has no rank; a contradiction. Hence the theorem holds.

Next we show that our new classes of languages $L(\mathrm{ET0L})_{\mathrm{RAN}}$ and $L(\mathrm{ET0L})_{\mathrm{RAN}(i)}$, $0 \leqslant i \leqslant n$, also have nice algebraic properties: $L(\mathrm{ET0L}_{\mathrm{RAN}}$ is a substitution closed ful AFL and for every $0 \leqslant i \leqslant m$ $(\mathrm{ET0L})_{\mathrm{RAN}(i)}$ is a full

AFL which is closed under substitution with ET0L languages of finite index.

Theorem 8. For every positive integer i, $L(\text{ET0L})_{\text{RAN}(i)}$ is a full AFL which is closed under substitution with $L(\text{ET0L})_{\text{RAN}(1)}$ languages.

Proof.

Let i be a positive integer.

(1) Closure of $L(\text{ET0L})_{\text{RAN}(i)}$ under union, product and intersection with regular sets can be established using the standard techniques from the corresponding proofs for $L(\text{ET0L})$ (see [3]) since these do not increase the rank.

(2) Let L be a language in $L(\text{ET0L})_{\text{RAN}(i)}$ and let $G = \langle V, P, S, \Sigma \rangle$ be an ET0L system of rank $k, k \leq i$, generating L. By Theorem 1 we can assume that G is in RNF. Let Z and ¢ be new symbols and define a new table $T_0 = \{Z \rightarrow ZS\} \cup \{X \rightarrow ¢ : X \in V \setminus \Sigma\} \cup \{a \rightarrow a : a \in \Sigma\}$. Consider the ET0L system

$H = \langle V \cup \{Z, ¢\}, \{T_0\} \cup \{P \cup \{Z \rightarrow Z, ¢ \rightarrow ¢\} : P \in P\}, Z, \Sigma \rangle$.

Clearly the rank of H does not exceed the rank of G and also $L(H) = L^+$. Thus $L(\text{ET0L})_{\text{RAN}(i)}$ is closed under the + operator. The closure of $L(\text{ET0L})_{\text{RAN}(i)}$ under the * operator can be established in a similar way.

(3) Let $L \subseteq \Sigma^*$ be in $L(\text{ET0L})_{\text{RAN}(i)}$ and let $\tau$ be a substitution from $\Sigma$ into $\Sigma_1^*$ such that $\tau(a) \in L(\text{ET0L})_{\text{RAN}(1)}$ for every a in $\Sigma$.

Let $G = \langle V, P, S, \Sigma \rangle$ be an ET0L system of rank not larger than i such that $L(G) = L$. Obviously we can assume that

G is in RNF. For every a in $\Sigma$, let $G_a = \langle V_a, P_a, S_a, \Sigma_1 \rangle$ be an ET0L system of rank not larger than 1 such that $L(G_a) = \tau(a)$. By Theorem 1, we can assume that $G_a$ is in FINF for all a in $\Sigma$. Clearly we can assume that all of the alphabets V and $V_a \setminus \Sigma_1$, $a \in \Sigma$, and $\Sigma_1$ are mutually disjoint. Let ¢ be a new symbol and define

$$\bar{V} = \bigcup_{a \in \Sigma} V_a \cup (V \setminus \Sigma).$$

Let $\phi$ be a homomorphism on V defined by $\phi(a) = a$ if $a \in V \setminus \Sigma$, and $\phi(a) = S_a$ if $a \in \Sigma$.

(1) For every table T from $P$ we define a new table

$$T' = \{a \to \phi(\alpha) : a \in V \setminus \Sigma, a \overset{T}{\Rightarrow} \alpha\} \cup$$

$$\cup \{X \to ¢ : X \in \bar{V} \setminus (V \cup \Sigma_1) \cup \{a \to a : a \in \Sigma_1\}$$

(2) For every $a \in \Sigma$ and for every table T from $P_a$ we define a new table $T' = T \cup \{X \to X : X \in \bar{V} \setminus V_a\}$.

Consider the ET0L system $H = \langle \bar{V}, \{T' : T \in P \cup \bigcup_{a \in \Sigma} P_a\}, S, \Sigma_1 \rangle$.

Obviously, $L(H) = \tau(L)$. Now we observe the following.

(i) If $a \in \bigcup_{a \in \Sigma} V_a$ then $\underline{\text{rank}}_H(a) \leqslant 1$.

(ii) If $a \in V$ is such that $\underline{\text{rank}}_G(a) \leqslant 1$, then $\underline{\text{rank}}_H(a) < 1$.

Indeed, it is not difficult to see that, in this case, if $a \overset{*}{\Rightarrow} x \in \underline{\text{Succ}}(H)$ then the number of occurrences of symbols of rank 1 in x is not greater than $k + k\,\underline{\text{maxr}}(H) \cdot \ell$ where k is the maximal number of occurrences of symbols of rank 1 in any word of $\underline{\text{SUCC}}_{G,V}(a)$ and where $\ell$ is the maximal of all $G_a$ (for $a \in \Sigma$).

From (i) and (ii) it easily follows that H is of rank not larger than i and thus $\mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}(i)}$ is closed under substitution with $\mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}(1)}$ languages.

From (1), (2), (3), the fact that $\mathcal{L}\mathrm{REG} \subseteq \mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}(1)}$ and from this well-known result which states that a class of languages which is closed under intersection with regular sets and substitution with regular sets is also closed under inverse homomorphism, the theorem follows.

<u>Theorem</u> 9. $\mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}}$ is a full substitution closed AFL.

<u>Proof</u>.

If we use the identity $\mathcal{L}(\mathrm{ET0L})_{\mathrm{RAN}} = \mathcal{L}(\mathrm{ET0L})_{\mathrm{NE}}$ then the usual constructions (see,e.g.,[3]) are easily seen to be applicable.

REFERENCES

[1] Ehrenfeucht, A. and Rozenberg, G., 1975. A pumping theorem for deterministic ET0L languages. Revue Francaise d'Automatique d'Informatique et de Recherche Operationelle, R-2, 9 : 13-23.

[2] Ehrenfeucht, A. and Rozenberg, G., 1976. On the structure of polynomially bounded D0L systems, Fundamenta Informatica, to appear.

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

[4] Rozenberg, G. and Vermeir, D., 1976. On ET0L systems of finite index, Information and Control, to appear.

[5] Rozenberg, G. and Vermeir, D., 1976. On the finite index restriction on several families of grammars. Dept. of Math., University of Antwerp, U.I.A. Technical Report No 76-19.

[6] Rozenberg, G. and Vermeir, D., 1977. On the finite index restriction on several families of grammars, Part two: context dependent systems and grammars, Dept. of Math., University of Antwerp, U.I.A., Technical Report No 77-04.

[7] Rozenberg, G. and Vermeir, D., 1977. On recursion in ET0L systems, Dept. of Math., University of Antwerp, U.I.A., Technical Report No 77-08.

[8] Salomaa, A., Formal languages, 1973. Academic Press New York.

ACKNOWLEDGMENTS

The authors greatfully acknowledge the financial support by the NSF Grant Number MCS75-09972 and by IBM Belgium. They are also very much indebted to L. Fosdick for making the work on this and other papers possible.

All correspondence to the second author

# ON ET0L SYSTEMS WITH RANK

by

A. Ehrenfeucht
Department of Computer Science
University of Colorado at Boulder,
Boulder, Colorado, USA

G. Rozenberg
Department of Mathematics
University of Antwerp, U.I.A.,
Wilrijk, Belgium

D. Vermeir
Department of Economics
University of Leuven, K.U.L.,
Leuven, Belgium

WORKING PAPER