THREE USEFUL RESULTS CONCERNING L LANGUAGES WITHOUT INTERACTIONS.

A. EHRENFEUCHT
Department of Computer Science,
University of Colorado at Boulder,
Boulder, Colorado, U.S.A.
and
G. ROZENBERG
Institute of Mathematics,
Utrecht University,
Utrecht, Holland.
Department of Mathematics,
Antwerp University, U.I.A.,
Wilrijk, Belgium.

1. INTRODUCTION.

The theory of L systems and languages (see, e.g., [45], [66], [75] and their references) is one of the fast growing areas of formal language theory. Still it is a rather young field (it originated in 1968 from the work of Lindenmayer, (see [59])) and a number of basic problems remain to be solved. One of the open areas within the theory are characterization results for various subclasses of the class of L languages. The kind of results the absence of which feels rather badly in the theory are the characterization results which would allow one to prove that particular languages do not belong to particular subclasses of the class of L languages. So far, almost exclusively, most of such proofs involved combinatorial arguments directed very much at specific properties of the specific language in question (see, e.g., [81] and [91] for proofs of such a kind). This led to the situation that each time it appeared necessary to prove that a given language is not of a particular kind, a whole new proof, mostly ad hoc, must be produced. (The drastic example of this kind is the proof from [35] of the fact that the language $\{x \in \{a,b\}^*$: the number of occurences of the letter a in x is a power of 2$\}$ is not an E0L language. This proof requires from the reader quite an investment of time to follow involved combinatorial arguments, and yet to prove that a slight variation of the above language is not an E0L language could pose a serious problem to the reader).

In this paper we want to present three results, which, although far from resolving the difficulties discussed above, should significantly contribute to this open area. The first of these results provides a partial characterization for a subclass of ET0L languages (see [89]), the second one provides a partial characterization for a subclass of deterministic ET0L languages (see [89]) and the last one provides a partial characterization for a subclass of E0L languages (see, e.g., [35]). The consequences of these results for comparison of various classes of L languages are also discussed.

In this paper we use standard formal language terminology and notation. In particular Λ denotes the empty word, $|x|$ denotes the length of x and $\#A$ denotes the cardinality of A. Also if x is a word over an alphabet Σ and a is in Σ, then $\#_a(x)$ denotes the number of occurrences of a in x; furthermore if $B \subseteq \Sigma$ then $\#_B(x) = \sum_{a \in B} \#_a(x)$. Finally abs(n) denotes the absolute value of n.

2. ET0L SYSTEMS AND LANGUAGES.

The class of ET0L systems and languages was introduced in [89] and is one of the actively investigated topics in the theory of L systems (see, for example, [6], [16], [27], [71] and [72]).

Definition 1. An ET0L system is a construct $G = \langle V,\Sigma,\mathcal{P},\omega\rangle$, where
1) V is a finite set (called the alphabet of G).
2) $\mathcal{P}$ is a finite set (called the set of tables of G), $\mathcal{P} = \{P_1,\ldots,P_f\}$ for some $f \geqslant 1$, each element of which is a finite subset of $V \times V^*$. $\mathcal{P}$ satisfies the following (completeness) condition:

$$(\forall P)_{\mathcal{P}}(\forall a)_V(\exists \alpha)_{V^*}(\langle a,\alpha\rangle \in P).$$

3) $\omega \in V^+$ (called the axiom of G).
4) $\Sigma \subseteq V$ (called the target alphabet of G).
(We assume that V, Σ and each P in $\mathcal{P}$ are nonempty sets.)

Definition 2. An ET0L system $G = \langle V,\Sigma,\mathcal{P},\omega\rangle$ is called:
1) Deterministic if for each P in $\mathcal{P}$ and each a in V there exists exactly one α in V^* such that $\langle a,\alpha\rangle \in P$.
2) An E0L system if $\#\mathcal{P} = 1$.
3) An 0L system if $\#\mathcal{P} = 1$ and $(V-\Sigma) = \phi$.

Definition 3. Let $G = \langle V,\Sigma,\mathcal{P},\omega\rangle$ be an ET0L system. Let $x \in V^+$,

$x = a_1 \ldots a_k$, where each a_j, $1 \leqslant j \leqslant k$, is an element of V, and let $y \in V^*$. We say that x directly derives y in G (denoted as $x \underset{G}{\Rightarrow} y$) if, and only if, there exist P in $\mathscr{P}$ and $p_1, \ldots, p_k$ in P such that $p_1 = \langle a_1, \alpha_1 \rangle$, $p_2 = \langle a_2, \alpha_2 \rangle, \ldots, p_k = \langle a_k, \alpha_k \rangle$ and $y = \alpha_1 \ldots \alpha_k$. Furthermore $\overset{*}{\underset{G}{\Rightarrow}}$ denotes the reflexive and transitive closure of the relation $\underset{G}{\Rightarrow}$.

Definition 4. Let $G = \langle V, \Sigma, \mathscr{P}, \omega \rangle$ be an ET0L system. The language of G, denoted as $L(G)$, is defined as $L(G) = \{x \in \Sigma^* : \omega \overset{*}{\underset{G}{\Rightarrow}} x\}$.

Definition 5. Let L be a language. L is called an ET0L (deterministic ET0L, E0L or 0L) language if, and only if, there exists an ET0L (deterministic ET0L, E0L or 0L) system G such that $L(G) = L$.

Example 1. $G_1 = \langle \{a,b,C,D\}, \{a,b\}, \{P_1,P_2,P_3\}, CD \rangle$, where $P_1 = \{a \to a, b \to b, C \to aCb, D \to Da\}$, $P_2 = \{a \to a, b \to b, C \to Cb, D \to D\}$ and $P_3 = \{a \to a, b \to b, C \to \Lambda, D \to \Lambda\}$, is a deterministic ET0L system such that $L(G) = \{a^n b^m a^n : n \geqslant 0,\ m \geqslant n\}$. (Following usual notation we write $x \to \alpha$ for an element $\langle x, \alpha \rangle$ of a table.)

Example 2. $G_2 = \langle \{A_1,B_1,C_1,a_1,b_1,c_1,F,a,b,c\}, \{a,b,c\}, \{P\}, A_1B_1C_1 \rangle$, where $P = \{A_1 \to A_1a_1, B_1 \to B_1b_1, C_1 \to C_1c_1, A_1 \to a, B_1 \to b, C_1 \to c, a_1 \to a, b_1 \to b, c_1 \to c, a \to F, b \to F, c \to F, F \to F\}$ is an E0L system such that $L(G_2) = \{a^n b^n c^n : n \geqslant 1\}$.

Example 3. $G_3 = \langle \{a\}, \{a\}, p, a \rangle$, where $P = \{a \to a^2\}$ is a 0L system such that $L(G_3) = \{a^{2^n} : n \geqslant 0\}$.

3. ET0L LANGUAGES OVER RARE SUBALPHABETS.

In this section we provide a partial characterization result for a subclass of ET0L languages.

Definition 6. If L is a language over an alphabet Σ and B is a nonempty subset of Σ, then

(1) B is called nonfrequent in L if there exists a constant $C_{B,L}$ such that for every x in L, $\#_B(x) < C_{B,L}$; otherwise B is called frequent in L.

(2) B is called rare in L if for every positive integer k there exists a n_k in N^+ such that for every n larger than n_k, if a word x in L contains n occurrences of letters from B then each two such occurrences are of distance not smaller than k.

Example 4. Let $L = \{(ab^k)^k : k \geqslant 1\}$ and $B = \{a\}$. Then B is frequent in L and also B is rare in L.

Theorem 1. If L is an ET0L language over an alphabet Σ and B is a nonempty subset of Σ which is rare in L, then B is nonfrequent in L.

Here are three examples of applications of Theorem 1.

Corollary 1. Let ψ be a function from positive integers into positive integers such that, for every positive integer n, $\psi(n) \geqslant n$. Then the language $\{(ab^{\psi(n)})^n : n \geqslant 1\}$ is not an ET0L language. Proof: Directly from Theorem 1 and Definition 6.

It is known (see [89], Theorem 19) that the class of ET0L languages is properly included in the class of Λ-free context-free programmed languages (introduced in Rosenkrantz, Programmed grammars and classes of formal languages, Journal of the A.C.M., 16, 107-131). Using Corollary 1 we can provide numerous constructions of Λ-free context-free programmed languages which are not ET0L languages. Thus for instance we have:

Corollary 2. The language $\{(ab^k)^k : k \geqslant 1\}$ is a Λ-free context-free programmed language, but it is not an ET0L language.

Proof. It is not difficult to construct a Λ-free context-free programmed grammar generating $L = \{(ab^k)^k : k \geqslant 1\}$. But $B = \{a\}$ is obviously rare in L whereas it is also frequent in L. Thus by Theorem 1, L is not an ET0L language.

4. DETERMINISTIC ET0L LANGUAGES.

In this section we provide a partial characterization for a subclass of deterministic ET0L languages. First we need a definition.

Definition 7.

(1) Let Σ be an alphabet and $x \in \Sigma^+$. We define $\mu(x)$ as the minimal positive integer n such that any two non-overlapping subwords of x are different.

(2) Let L be a language. L is called exponential if there exists a positive integer C_L larger than 1 such that for every x_1, x_2 in L, if $|x_1| > |x_2|$ then $|x_1| \geqslant C_L|x_2|$.

Example 5. If $\Sigma = \{a,b,c\}$ and $x = abcaba$ then $\mu(x) = 3$. The language $\{x \in \{a,b,c\}^* : |x| = 2^n \text{ for some } n \geq 0\}$ is an exponential language.

Theorem 2. If L is an exponential deterministic ET0L language then there exists a positive integer constant F_L such that, for every x in $L-\{\Lambda\}$, we have $\frac{|x|}{\mu(x)} < F_L$.

As an application of this theorem we can prove now that there exists ET0L languages which are not deterministic ET0L languages. (This was posed as an open problem in [89]). In fact we have even stronger result.

Corollary 4. There exists a 0L language which cannot be generated by an EDT0L system.

Proof. Let $L = \{x \in \{a,b\}^* : |x| = 2^n \text{ for some } n \geq 0\} - \{b\}$. The reader can easily check that L is generated by the 0L system $\langle\{a,b\},\{a,b\},P,a\rangle$ where $P = \{a \to aa, a \to ab, a \to ba, a \to bb, b \to aa, b \to ab, b \to ba, b \to bb\}$ and so L is a 0L language. On the other hand L is exponential, but it does not satisfy the statement of Theorem 2, and so it is not a deterministic ET0L language.

5. E0L LANGUAGES OVER NUMERICALLY DISPERSED SUBALPHABETS.

In this section we provide a partial characterization for a subclass of E0L languages. We start with a definition.

Definition 8. Let L be a language over an alphabet Σ and let B be a nonempty subset of Σ. Let $I_{L,B} = \{n \in N :$ there exists a word ω in L such that $\#_B(\omega) = n\}$.

(1) B is numerically dispersed in L if, and only if, $I_{L,B}$ is infinite and for every positive integer k there exists a positive integer n_k such that, for every u_1, u_2 in $I_{L,B}$, if $u_1 \neq u_2$, $u_1 > n_k$ and $u_2 > n_k$ then $abs(u_1 - u_2) > k$.

(2) B is clustered in L if, and only if, $I_{L,B}$ is infinite and there exist positive integers k_1, k_2 both larger than 1 such that, for every word ω in L, if $\#_B(\omega) \geq k_1$, then ω contains at least two occurrences of symbols from B which are of distance smaller than k_2.

Example 6. Let $L = \{x \in \{a,b\}^* : \#_{\{a\}}(x) = 2^n$ for some $n \geqslant 0\}$ and let $B = \{a\}$. Obviously B is numerically dispersed in L, but B is not clustered in L. However, the language $\{aba\} \cdot L$ is such that B is clustered in L.

Theorem 3. Let L be an E0L language over an alphabet Σ and let B be a nonempty subset of Σ. If B is numerically dispersed in L, then B is clustered in L.

As an example of the application of Theorem 3 we have the following result. (A language L is called a deterministic T0L language if there exists a deterministic ET0L system $G = \langle V,\Sigma, \ ,\omega\}$ such that $L(G) = L$ and $V = \Sigma$.)

Corollary 5. There exist deterministic T0L languages which are not E0L languages.

Proof. Let $L = \{(ab^m)^{2^n} : m,n \geqslant 0\} \cup \{c^{2^n} : n \geqslant 0\}$. L is a deterministic T0L language, because it is the language of the system $\langle\{a,b\},\{a,b\},\{P_1,P_2,P_3\},c\rangle$ where $P_1 = \{a \to a, b \to b, c \to c^2\}$, $P_2 = \{a \to a, b \to b, c \to a\}$ and $P_3 = \{a \to ab, b \to b, c \to c\}$. On the other hand $\{a\}$ is numerically dispersed in L but it is not clustered in L. Consequently, by Theorem 3, L is not an E0L language.