ON DECOMPOSING SOME ETOL LANGUAGES INTO DETERMINISTIC ETOL LANGUAGES*

by

A. Ehrenfeucht**

and

G. Rozenberg***

Report #CU-CS-045-74 July 1974

* This work supported by NSF Grant #GJ-660

** Department of Computer Science
University of Colorado
Boulder, Colorado 80302 U.S.A.

*** Institute of Mathematics
Utrecht University
Utrecht-De Uithof HOLLAND
and
Department of Mathematics
University of Antwerp, U.I.A.
BELGIUM

All correspondence to G. Rozenberg
Institute of Mathematics
Utrecht University
Utrecht-De Uithof HOLLAND

ABSTRACT

This paper provides a method of decomposing a subclass of ETOL languages into deterministic ETOL languages. This allows one to use every known example of a language which is not a deterministic ETOL language to produce languages which are not ETOL languages.

ACKNOWLEDGMENT

The authors would like to thank Professor Lloyd D. Fosdick for arranging a very enjoyable stay for the second author at the Computer Science Department of the University of Colorado at Boulder. Thanks to this arrangement the work on this and other papers could be continued.

I. INTRODUCTION

The theory of L systems originated from the work of A. Lindenmayer (see Lindenmayer [ii]). Although initially proposed as a theory for the development of filamentous organisms, in the last four years it turned out to be useful and interesting from both the biological and formal points of view (see, e.g., Herman and Rozenberg [g], and Rozenberg and Salomaa [13]).

One of the central families of L languages (that is languages generated by L systems) is the family of ETOL languages (see, e.g. Downey [2], Rozenberg [12] and Salomaa [15]). An important research area in the theory of ETOL systems and languages is to provide results which would facilitate proofs that certain languages are not ETOL languages. Although some such results are already available (see, e.g., Ehrenfeucht and Rozenberg [4], and Ehrenfeucht and Rozenberg [5]), a lot of work in this direction remains to be done.

This paper provides a criterion for proving that some languages are not ETOL languages. In fact it shows how, in certain cases, to reduce this problem to proving that some languages are not deterministic ETOL languages (see Rozenberg [12] and Ehrenfeucht and Rozenberg [6]). This is a great help indeed, because it is easier to investigate the structure of derivations in a deterministic ETOL system, and quite a number of examples of languages that are not deterministic ETOL languages are already available (see, e.g., Ehrenfeucht and Rozenberg [7] and Ehrenfeucht and Rozenberg [8]).

As a corollary of our results we get that the family of ETOL languages is strictly included in the family of index languages of Aho (see, Aho [1]). This was quite an important open problem of a rather long standing (see, e.g., Downey [2], Salomaa [15] and Salomaa [16]).

We assume the reader to be familiar with rudiments of formal language theory, e.g. in the scope of the first four chapters of Hopcroft and Ullman [10].

II. DEFINITIONS

In this section we provide definitions and examples of systems and languages used in this paper.

Definition 1. An extended table L system without interactions, abbreviated as an ETOL system, is defined as a four-tuple $G = \langle V, \mathcal{P}, \omega, \Sigma\rangle$ such that:

(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 $f \geq 1$, 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. Let $G = \langle V, \mathcal{P}, \omega, \Sigma\rangle$ be an ETOL system. Let $x \in V^+$, $x = a_1 \ldots a_k$, where each a_j, $1 \leq j \leq k$, is an element of V, and let $y \in V^*$. We say that x directly derives y in G (denoted $x \underset{G}{\Longrightarrow} y$) if and only if there exist P in $\mathcal{P}$ and $p_1, \ldots, p_k$ in P such that $p_1 = \langle a_1, \alpha_1\rangle$, $p_2 = \langle a_2, \alpha_2\rangle, \ldots,$ $p_k = \langle a_k, \alpha_k\rangle$ (for some $\alpha_1, \ldots, \alpha_k \in V^*$) and $y = \alpha_1 \ldots \alpha_k$. We say that x derives y in G (denoted $x \overset{*}{\underset{G}{\Longrightarrow}} y$) if and only if either (i) there exists a sequence of words $x_o, x_1, \ldots, x_n$ in V^* (with $n > 1$) such that $x_o = x$, $x_n = y$ and $x_o \underset{G}{\Longrightarrow} x_1 \underset{G}{\Longrightarrow} x_2 \ldots \underset{G}{\Longrightarrow} x_n$; or (ii) $x = y$.

Definition 3. Let $G = \langle V, \mathcal{P}, \omega, \Sigma\rangle$ be an ETOL system. The language of G (denoted as L(G)) is defined as $L(G) = \{x \in \Sigma^* : \omega \overset{*}{\underset{G}{\Longrightarrow}} x\}$.

Definition 4. An ETOL system $G = \langle V, \mathcal{P}, \omega, \Sigma\rangle$ is called deterministic (abbreviated EDTOL system) if for each P in $\mathcal{P}$ and each a in V there exists exactly one α in V^* such that $\langle a, \alpha\rangle \in P$.

Definition 5. Let Σ be a finite alphabet and $K \subseteq \Sigma^*$. K is called an ETOL (EDTOL) language if and only if there exists an ETOL (EDTOL) system G such that $L(G) = K$.

We shall use L(ETOL) and L(EDTOL) to denote the class of ETOL languages and the class of EDTOL languages, respectively.

Definition 6. An ETOL system $G = \langle V, P, \omega, \Sigma \rangle$ is called synchronized if for every x, y such that $x \in V^* \Sigma V^*$, $y \in V^+$ and x derives y in at least one step, then $y \in V^*(V - \Sigma) V^*$.

Definition 7. Let $G = \langle V, P, \omega, \Sigma \rangle$ be an ETOL system with $P = \{P_1, \ldots, P_f\}$. Let $Z \subseteq V$. For P in P a P(Z) table is a set of ordered pairs $\langle a, \alpha \rangle$, a in V and α in V^*, such that, for each a in $(V - Z)$ each element $\langle a, \alpha \rangle$ from P is in P(Z), for each a in Z, P(Z) contains exactly one element $\langle a, \alpha \rangle$ from P and P(Z) contains nothing else. An ETOL system $H = \langle V, \overline{P}, \omega, \Sigma \rangle$ is called the Z-combinatorially complete version of G if $\overline{P} = \{T$: for some P in P, T is a P(Z) table$\}$; If $Z = V$, then we say that it is the combinatorially complete version of G.

Notation. Let $G = \langle V, P, \omega, \Sigma \rangle$ be an ETOL system. If $\langle a, \alpha \rangle$ is an element of some P in P, then we call it a production (for a in P) and write $a \underset{P}{\longrightarrow} \alpha$. A derivation in G is a sequence of words $(x_0, x_1, \ldots, x_n)$ such that $x_0 = \omega$ and $x_j \underset{G}{\Longrightarrow} x_{j+1}$, for $0 \leq j \leq n-1$. (We also say that it is a derivation of x_n in G). Sometimes by a derivation we shall mean a sequence $(x_0, \ldots, x_n)$ together with the precise set of productions used in each derivation step but this will always be clear from the context and should not lead to confusion.

Example 1. Let $G_1 = \langle \{a, b, A, B, C, D, F\}, P, CD, \{a, b\} \rangle$ where $P = \{P_1, P_2, P_3\}$ and

$P_1 = \{a \longrightarrow F,\ b \longrightarrow F,\ A \longrightarrow A,\ B \longrightarrow B,\ C \longrightarrow ACB,\ D \longrightarrow DA\}$,

$P_2 = \{a \longrightarrow F,\ b \longrightarrow F,\ A \longrightarrow A,\ B \longrightarrow B,\ C \longrightarrow CB,\ D \longrightarrow D\}$,

$P_3 = \{a \longrightarrow F,\ b \longrightarrow F,\ A \longrightarrow a,\ B \longrightarrow b,\ C \longrightarrow \Lambda,\ D \longrightarrow \Lambda\}$.

G_1 is a synchronized EDTOL system and

$L(G_1) = \{a^n\ b^m\ a^n : n \geq 0,\ m \geq n\}$.

Example 2. Let $G_2 = <\{a, b, A, A', B, B', C, C', F\}, \mathcal{P}, A\ B\ C, \{a, b\}>$ where $\mathcal{P} = \{P\}$ and

$P = \{a \longrightarrow F,\ b \longrightarrow F,\ c \longrightarrow F,\ A \longrightarrow A'A,\ A \longrightarrow a,\ B \longrightarrow B'B,\ B \longrightarrow b,\ C \longrightarrow C'C,$

$C \longrightarrow c,\ A' \longrightarrow A',\ A' \longrightarrow a,\ B' \longrightarrow B',\ B' \longrightarrow b,\ C' \longrightarrow C',\ C' \longrightarrow c,\ F \longrightarrow F\}$.

G_2 is a synchronized (but not deterministic) ETOL system and $L(G) = \{a^n\ b^n\ c^n : n \geq 1\}$.

III. RESULTS

Theorem 1. Let Σ_1, Σ_2 be two finite disjoint alphabets and let $K_1 \subseteq \Sigma^*$, $K_2 \subseteq \Sigma^*$. Let f be a bijective function from K_1 onto K_2. Let $K = \{wf(w) : w \in K_1\}$. If $K \in \mathcal{L}(\text{ETOL})$ then $K, K_1, K_2 \in \mathcal{L}(\text{EDTOL})$.

Proof

The idea of our proof is to start with an arbitrary ETOL system generating K and then to construct EDTOL systems generating K, K_1 and K_2 respectively. As the construction is quite involved we have split it up into several steps.

Let K satisfy the statement of the theorem.

Let $G = \langle V, \mathcal{P}, \omega, \Sigma\rangle$, where $\Sigma = \Sigma_1 \cup \Sigma_2$, be a synchronized ETOL system generating K. By Theorems 3 and 4 in Rozenberg [12], we may assume that $\omega = S$, where $S \in V - \Sigma$, and, for every a in V and every P in $\mathcal{P}$, if $a \underset{P}{\longrightarrow} \alpha$ then either $\alpha \in \Sigma^*$ or $\alpha \in (V-\Sigma)^+$.

Now we present, in several steps, our construction.

STEP 1.

Let $V^{(1)} = \{a^{(1)} : a \in V - \Sigma_2\}$, $V^{(2)} = \{a^{(2)} : a \in V - \Sigma_1\}$ and $V^{(m)} = \{a^{(m)} : a \in V - \Sigma\}$. If $j \in \{1,2,m\}$, and $\alpha = b_1 \ldots b_k$ with $k \geq 1$ and $b_1, \ldots, b_k \in V$, then $\alpha^{(j)} = b_1^{(j)} \ldots b_k^{(j)}$. Also for $j \in \{1,2,m\}$, $\Lambda^{(j)} = \Lambda$. Let $\Sigma_1^{(1)} = \{a^{(1)} : a \in \Sigma_1\}$ and $\Sigma_2^{(2)} = \{a^{(2)} : a \in \Sigma_2\}$. Let F be a new symbol ($F \notin V^{(1)} \cup V^{(2)} \cup V^{(m)} \cup \Sigma$). Let $V_1 = V^{(1)} \cup V^{(2)} \cup V^{(m)} \cup \Sigma \cup \{F\}$.

For each table P in $\mathcal{P}$ we construct a new table $\overline{P}$ as follows:

(i) if $a \underset{P}{\longrightarrow} \alpha$ with $\alpha \in \Sigma_1^*$, then $a^{(1)} \underset{\overline{P}}{\longrightarrow} \alpha^{(1)}$,

(ii) if $a \underset{P}{\longrightarrow} \alpha$ with $\alpha \in \Sigma_2^*$, then $a^{(2)} \underset{\overline{P}}{\longrightarrow} \alpha^{(2)}$,

(iii) if $a \underset{P}{\longrightarrow} \alpha$ with $\alpha \in (V-\Sigma)^+$, say $\alpha = b_1 \ldots b_k$ for some $k \geq 1$ and $b_1, \ldots, b_k \in (V - \Sigma)$, then

$a^{(1)} \underset{\overline{P}}{\longrightarrow} \alpha^{(1)}$,

$a^{(2)} \underset{\overline{P}}{\longrightarrow} \alpha^{(2)}$,

$a^{(m)} \xrightarrow[\overline{P}]{} \alpha^{(1)}$,

$a^{(m)} \xrightarrow[\overline{P}]{} b_1^{(1)} \ldots b_{\ell-1}^{(1)} b_\ell^{(m)} b_{\ell+1}^{(2)} \ldots b_k^{(2)}$ for every ℓ in $\{1, \ldots, k\}$,

$a^{(m)} \xrightarrow[\overline{P}]{} b^{(1)} \ldots b_{\ell-1}^{(1)} b_\ell^{(2)} \ldots b_k^{(2)}$ for every ℓ in $\{1, \ldots, k\}$,

(iv) $F \xrightarrow[\overline{P}]{} F$,

(v) for every j in {1,2,m} and every a in V, $a^{(j)} \xrightarrow[\overline{P}]{} F$,

(this is the easiest way to have $\overline{P}$ satisfying the completeness condition),

(vi) only productions obtained from (i) through (v) are in $\overline{P}$.

Let P_t be a new table such that
$P_t = \{a^{(1)} \longrightarrow a : a^{(1)} \in \Sigma_1^{(1)}\} \cup \{a^{(2)} \longrightarrow a : a^{(2)} \longrightarrow \Sigma_2^{(2)}\} \cup$
$\{a \longrightarrow F : a \in V - (\Sigma_1^{(1)} \cup \Sigma_2^{(2)})\}$.

Let $\mathcal{P}_1 = \{P_t\} \cup \{\overline{P} : P \in \mathcal{P}\}$.

Finally let $G_1 = \langle V_1, \mathcal{P}_1, S^{(m)}, \Sigma\rangle$.

STEP 2.

Let $G_2 = \langle V_2, \mathcal{P}_2, S^{(m)}, \Sigma\rangle$ be the $V^{(m)}$- combinatorially complete version of G_1. (Note that $V_2 = V_1$).

STEP 3.

Let $V^{(m:1)} = \{a^{(m:1)} : a^{(m)} \in V^{(m)}\}$ and $V^{(m:2)} = \{a^{(m:2)} : a^{(m)} \in V^{(m)}\}$.
Let $V_3 = (V - V^{(m)}) \cup V^{(m:1)} \cup V^{(m:2)}$.

For each table P in $\mathcal{P}$ we construct a new table $\hat{P}$ as follows:

(i) if $a \in V - V^{(m)}$ and $a \xrightarrow[P]{} \alpha$, then $a \xrightarrow[\hat{P}]{} \alpha$,

(ii) if $a^{(m)} \in V^{(m)}$ and $a^{(m)} \xrightarrow[P]{} \alpha^{(1)} b^{(m)} \alpha^{(2)}$ for some $b^{(m)}$ in $V^{(m)}$, then $a^{(m:1)} \xrightarrow[\hat{P}]{} \alpha^{(1)} b^{(m:1)}$ and $a^{(m:2)} \xrightarrow[\hat{P}]{} b^{(m:2)} \alpha^{(2)}$,

(iii) if $a^{(m)} \in V^{(m)}$ and $a^{(m)} \xrightarrow[P]{} \alpha^{(1)}$, then
$a^{(m:1)} \xrightarrow[\hat{P}]{} \alpha^{(1)}$ and $a^{(m:2)} \xrightarrow[\hat{P}]{} \alpha^{(2)}$,

(iv) if $a^{(m)} \in V^{(m)}$ and $a^{(m)} \underset{P}{\longrightarrow} \alpha^{(2)}$, then

$a^{(m:1)} \underset{\hat{P}}{\longrightarrow} \Lambda$ and $a^{(m:2)} \underset{\hat{P}}{\longrightarrow} \alpha^{(2)}$

(v) if $a^{(m)} \in V^{(m)}$ and $a^{(m)} \underset{P}{\longrightarrow} \alpha^{(1)} \alpha^{(2)}$, then

$a^{(m:1)} \underset{\hat{P}}{\longrightarrow} \alpha^{(1)}$ and $a^{(m:2)} \underset{\hat{P}}{\longrightarrow} \alpha^{(2)}$.

(vi) only productions obtained from (i) through (v) are in $\hat{P}$.

Let $\mathcal{P}_3 = \{\hat{P} : P \in \mathcal{P}_2\}$.

Finally let $G_3 = \langle V_3, \mathcal{P}_3, S^{(m:1)} S^{(m:2)}, \Sigma\rangle$

STEP 4.

Let $G_4 = \langle V_4, \mathcal{P}_4, S^{(m:1)} S^{(m:2)}, \Sigma\rangle$ be the combinatorially complete version of G_3.

Let $\overline{P}_t = \{a^{(1)} \longrightarrow a : a^{(1)} \in \Sigma_1^{(1)}\} \cup \{a^{(2)} \longrightarrow \Lambda : a^{(2)} \in \Sigma_2^{(2)}\} \cup$
$\cup \{a \longrightarrow F : a \in V_4 - (\Sigma_1^{(1)} \cup \Sigma_1^{(2)})\}$.

Let $\overline{\overline{P}}_t = \{a^{(1)} \longrightarrow \Lambda : a^{(1)} \in \Sigma^{(1)}\} \cup \{a^{(2)} \longrightarrow a : a^{(2)} \in \Sigma_2^{(2)}\} \cup$
$\cup \{a \longrightarrow F : a \in V_4 - (\Sigma_1^{(1)} \cup \Sigma_2^{(2)})\}$.

Let $\overline{\mathcal{P}}_4 = (\mathcal{P}_4 - \{P_t\}) \cup \{P_t\}$ and $\overline{\overline{\mathcal{P}}}_4 = (\mathcal{P}_4 - \{P_t\}) \cup \{\overline{\overline{P}}_t\}$.

Let $G_4^{(1)} = \langle V_4, \overline{\mathcal{P}}_4, S^{(m:1)} S^{(m:2)}, \Sigma\rangle$ and $G_4^{(2)} = \langle V_4, \overline{\overline{\mathcal{P}}}_4, S^{(m:1)} S^{(m:2)}, \Sigma\rangle$.

To complete the proof of Theorem 1 we have to show that $L(G_4) = K$, $L(G_4^{(1)}) = K_1$ and $L(G_4^{(2)}) = K_2$.

Let us first prove that $L(G_4) = K$.

We do this by proving that the sequence of ETOL systems G, G_1, G_2, G_3, G_4 has this property that they all generate the same language.

I) $L(G) = L(G_1)$.

We shall present now the main idea behind the proof of this equality, leaving to the reader the formal proof.

$L(G) \subseteq L(G_1)$.

If one takes a derivation tree in G of a word x of the form $x_1\, x_2$ with $x_1 \varepsilon \Sigma_1^*$ and $x_2 \varepsilon \Sigma_2^*$, then (in the bottom-up fashion) one can classify all the nodes in this tree into three categories: those which contribute to x_1, those which contribute to x_2 and those which contribute to both x_1 and x_2. If a node belongs to the first category and its label is a then we change it to $a^{(1)}$, if a node belongs to the second category and its label is a then we change it to $a^{(2)}$, and if a node belongs to the third category and its label is a then we change it to $a^{(m)}$. But it clearly follows from the construction of G that such a derivation tree with one extra level added (corresponding to the application of P_t from $\mathcal{P}_1$) corresponds to a derivation in G_1 and consequently x is in $L(G_1)$. Thus $L(G) \subseteq L(G_1)$.

$L(G_1) \subseteq L(G)$.

If one takes a derivation of a word x in $L(G_1)$ and then omits all superscripts in the letters of the form $a^{(i)}$ with i in $\{1,2,m\}$ and also omits the last level (corresponding to the application of table P_t from $\mathcal{P}_1$), then, clearly, one gets a valid derivation in L(G). Consequently x is in L(G). Thus $L(G) \subseteq L(G)$.

II). $L(G_1) = L(G_2)$.

This equality follows immediately by observing that in each derivation in G_1 each intermediate word contains at most one occurrence of a letter of the form $a^{(m)}$.

III). $L(G_2) = L(G_3)$.

What G_3 does is simply split each derivation tree of a word x in $L(G_2)$ into two trees (glued together): one corresponding to the derivation of the prefix x_1 (in Σ_1^*) of x and the second corresponding to the derivation of the suffix x_2(in Σ_2^*) of x.

IV). $L(G_3) = L(G_4)$.

Clearly $L(G_4) \subseteq L(G_3)$.

Now let $D = (y_0 = S^{(m:1)}\, S^{(m:2)}, y_1, y_2, \ldots, y_n)$ be a derivation in G_3 of the word y_n which is of the form $\beta_1\beta_2$ where $\beta_1 \in \Sigma_1^*$ and $\beta_2 \in \Sigma_2^*$. Let $T_1, \ldots, T_n$ be the sequence of tables from P_3 that was applied in this particular derivation.

Let us change derivation D (in a top-down fashion) to derivation $\overline{D} = (\overline{y}_0 = S^{(m:1)} S^{(m:2)}, \overline{y}_1, \overline{y}_2, \ldots, \overline{y}_n = \overline{\beta}_1\beta_2)$, with $\overline{\beta}_1$ in Σ_1^*, as follows:

1) for every j in $\{1, \ldots, n\}$ and every a in $V^{(1)} \cup V^{(m:1)}$ we rewrite each occurrence of a in $\overline{y}_{j-1}$ by the same production from T_j, but it must be a production used in rewriting an occurrence of a in y_{j-1},

2) for every j in $\{1, \ldots, n\}$ and every a in $V - (V^{(1)} \cup V^{(m:1)})$ we rewrite each occurrence of a in $\overline{y}_{j-1}$ in exactly the same way it was rewritten in y_{j-1}.

We can note now that, since f is a bijective function, $\overline{\beta}_1 = \beta_1$.

Let us change derivation D (in a top-down fashion) to derivation $\overline{\overline{D}} = (\overline{\overline{y}}_0 = S^{(m:1)} S^{(m:2)}, \overline{\overline{y}}_1, \ldots, \overline{\overline{y}}_n = \beta_1\overline{\overline{\beta}}_2)$, with $\overline{\overline{\beta}}_2$ in Σ_2^* as follows:

3) for every j in $\{1, \ldots, n\}$ and every a in $V^{(2)} \cup V^{(m:2)}$ we rewrite each occurrence of a in $\overline{\overline{y}}_{j-1}$ by the same production from T_j, but it must be a production used in rewriting an occurrence of a in y_{j-1},

4) for every j in $\{1, \ldots, n\}$ and every a in $V - (V^{(2)} \cup V^{(m:2)})$ we rewrite each occurrence of a in $\overline{\overline{y}}_{j-1}$ in exactly the same way it was rewritten in y_{j-2}.

We can note now that, since f is a bijective function, $\overline{\overline{\beta}}_2 = \beta_2$.

It follows immediately from our conclusions about $\overline{D}$ and $\overline{\overline{D}}$ that there exists a derivation of y_n in G_4. Thus $L(G_3) \subseteq L(G_4)$.

From I) through IV) we get that $L(G) = L(G_4)$.

But G_4 is an EDTOL system and consequently K is in $\mathcal{L}$(EDTOL).

We leave to the reader the obvious proofs that $K_1 = L(G_4^{(1)})$ and $K_2 = L(G_4^{(2)})$. Both of these equalities follow easily from the observation that G_4 is a synchronized EDTOL system and in every successful derivation in G_4 the last table applied must be P_t. But $G_4^{(1)}$ and $G_4^{(2)}$ are EDTOL systems and consequently both K_1 and K_2 are in $\mathcal{L}$(EDTOL).

This completes the proof of Theorem 1.

IV. APPLICATIONS

First of all, using Theorem 1 and Theorem 2 we can provide examples of languages which are not in L(ETOL).

Let $W_1 = \{x \in \{0, 1\}^+ : |x| = 2^n$ for some $n \geq 0\}$.

Let us recall the following result proved in Ehrenfeucht and Rozenberg [3] and Ehrenfeucht and Rozenberg [7].

Lemma 1. $W_1 \in L$(EDTOL).

Let f_1 be a function from $\{0, 1\}^+$ into $\{c, d\}^+$ defined as follows. For $k \geq 1$ and $b_1, \ldots, b_k \in \{0, 1\}$, $f_1(b_1 \ldots b_k) = x_1 x_2 \ldots x_k$ where for every i in $\{1, \ldots, k\}$

$$x_i = \begin{cases} c & \text{if } b_i = 0, \\ d & \text{if } b_i = 1. \end{cases}$$

Let $W_2 = \{x\, f_1(x) : x \in W_1\}$.

Proposition 1. $W_2 \notin L$(ETOL).

Proof.

If we set $K_1 = W_1$, f equal to f_1 restricted to W_1 and K_2 equal to the range of function f then we have $W_2 = \{x\, f(x) : x \in K_1\}$ where f is a bijective function onto K_2. Thus if $W_2 \in L$(ETOL) then from Theorem 1 it follows that K_1 is an EDTOL language which contradicts Lemma 1. Consequently $W_2 \notin L$(ETOL).

Let f_2 be a homomorphism from $\{a, b\}^+$ onto $\{0, 1\}^+$ defined by $f_2(a) = 0$ and $f_2(b) = 1$. Let

$W_3 = \{x\, f_2(x) : x \in \{a, b\}^+$ and $|x| = 2^n$ for some $n \geq 0\}$.

Proposition 2. $W_3 \notin L$(ETOL).

Proof.

If we set $K_1 = \{x \in \{a, b\}^+ : |x| = 2^n$ for some $n \geq 0\}$, f equal to f_2 restricted to K_1 and $K_2 = W_1$ then we have $W_3 = \{x\, f(x) : x \in K_1\}$ where f is a bijective function onto K_2. Thus if $W_2 \in L$(ETOL) then from Theorem 1 it follows that K_2 is an EDTOL language which contradicts Lemma 1. Consequently $W_3 \notin L(\text{ETOL})$.

Let f_3 be a homomorphism from $\{a, b, c\}^+$ onto $\{0, 1\}^+$ defined by $f_3(a) = 0$, $f_3(b) = 1$ and $f_3(c) = \Lambda$. Let f_c be a homomorphism from $\{a, b, c\}^+$ onto $\{0, 1\}^+$ defined by $f_c(a) = a$, $f_c(b) = b$ and $f_c(c) = \Lambda$. Let
$W_4 = \{x\ f_3(x) : x \in \{a, b, c\}^+ \text{ and } |f_c(x)| = 2^n \text{ for some } n \geq 0\}$.

Proposition 3. $W_4 \notin L(\text{ETOL})$.

Proof.

If we set $K_1 = \{x\ e\ \{a, b, c\}^+ : |f_c(x)| = 2^n \text{ for some } n \geq 0\}$, f equal to f_3 restricted to K_1 and $K_2 = W_1$ then we have $W_4 = \{x\ f(x) : x \in K_1\}$ where f is a function onto K_2. Thus if $W_2 \in L(\text{ETOL})$ then from Theorem 2 it follows that K_2 is an EDTOL language which contradicts Lemma 1. Consequently $W_3 \in L(\text{ETOL})$.

Note that the function f as defined in the proof of Proposition 2 is not a bijective function, hence it was necessary to apply Theorem 2 rather than Theorem 1.

Finally we can settle a quite important open problem of long standing (see, e.g., Downey [2] and Salomaa [16]) whether or not the class of ETOL languages is contained in the class of indexed languages (see Aho [1]). Let $L(\text{IND})$ denote the class of indexed languages. (Now we assume that the reader is familiar with Aho [1]).

Theorem 3. Let Σ be a finite alphabet and let $\bar{\Sigma} = \{\bar{a} : a \in \Sigma\}$. Let h be a homomorphism from Σ^* onto $\bar{\Sigma}^*$ defined by $h(a) = \bar{a}$, for every a in Σ. Let K be a context-free language over Σ such that K is not an EDTOL language. Then the language† $M_K = \{w(h(w))^{mir} : w \in K\}$ is in $L(\text{IND})$ but it is not in $L(\text{EDTOL})$.

Proof.

If a language is context-free then it can be generated by a right linear indexed right linear grammer (see Aho [1], Lemma 6.1). Thus, obviously, $M_K \in L(\text{IND})$.

On the other hand from Theorem 1 it follows that M_K is not in $L(\text{EDTOL})$.

---

†For a word x, x^{mir} denotes the mirror image of x.

Now we turn to our next theorem.

Theorem 2. Let Σ_1, Σ_2 be two disjoint alphabets and let $K_1 \subseteq \Sigma_1^*$, $K_2 \subseteq \Sigma_2^*$. Let f be a surjective function from K_1 onto K_2. Let $K = \{wf(w) : w \in K_1\}$. If K is in $\mathcal{L}$(ETOL) then

(i) $K_2 \in \mathcal{L}$(EDTOL), and

(ii) There exists $\overline{K}_1$ such that $\overline{K}_1 \subseteq K_1$, $f(\overline{K}_1) = K_2$, $\overline{K}_1$ is in $\mathcal{L}$(EDTOL) and $\{wf(w) : w \in \overline{K}_1\}$ is in $\mathcal{L}$(EDTOL).

Proof.

Most of the proof of this theorem was done already in the proof of Theorem 1. Let us note that in showing (in the proof of Theorem 1) that $L(G) = L(G_1) = L(G_2) = L(G_3) = L(G_4)$ the particular property of the function f (its bijectiveness) was used only in proving that $L(G_3) = L(G_4)$.

Thus let G, G_1, G_2, G_3, G_4, $G_4^{(1)}$ and $G_4^{(2)}$ be defined as in the proof of Theorem 1.

As we still require that f is a surjective (but not necessarily bijective) function one can clearly see that $L(G_4) = \{\gamma_1\gamma_2 : \gamma_1 \in \Sigma_1^*, \text{ and } \gamma_2 \in \Sigma_2^*\}$ where for every $\beta_1\beta_2$ in K, where $\beta_1 \in \Sigma_1^*$ and $\beta_2 \in \Sigma_2^*$, there exists γ_1 in Σ_1^* such that $\gamma_1\beta_2$ is in $L(G_4)$. Also it is clear that $L(G_4) \subseteq L(G_3)$ where $L(G_3) = L(G) = K$. Consequently

$\{\gamma_1 \in \Sigma_1^* :$ there exists γ_2 in Σ_2 such that $\gamma_1\gamma_2$ is in $L(G_4)\} \subseteq K_1$.

But then the theorem follows from the equalities:

$K_2 = L(G_4^{(2)})$, $\overline{K}_1 = L(G_4^{(1)})$ and $L(G_4) = \{wf(w) : w \in \overline{K}_1\}$.

Hence Theorem 3 follows.

For each $i \geq 1$, let $\Sigma_i = \{[_1, \ldots, [_i,]_i, \ldots,]_1\}$ and let $\mathcal{B}_i$ be the language generated by the context-free grammar $H(\mathcal{B}_i) = \langle \{S\}, \Sigma_i, P_i, S \rangle$, where $P_i = \{S \longrightarrow [_j SS]_j : 1 \leq j \leq i\} \cup \{S \longrightarrow [_j S]_j : 1 \leq j \leq i\} \cup \{S \longrightarrow [_j\ ]_j : 1 \leq j \leq i\}$.

Let us recall now two results from Ehrenfeucht and Rozenberg [8]. (We assume the reader to be familiar with the notion of a Dyck language, see, e.g., Salomaa [14], p. 210).

<u>Lemma 2.</u> For every $i \geq 1$, $\mathcal{B}_i$ is not in $\mathcal{L}$(EDTOL).

<u>Lemma 3.</u> If K is a Dyck language over an alphabet of at least eight letters then K is not in $\mathcal{L}$(EDTOL).

Now from Theorem 3, Lemma 2 and Lemma 3 we have the following results.

<u>Corollary 1.</u> For every $i \geq 1$, $M_{\mathcal{B}_i} \in \mathcal{L}(\text{IND}) - \mathcal{L}(\text{ETOL})$.

<u>Corollary 2.</u> If K is a Dyck language over an alphabet of at least eight letters, then $M_K \in \mathcal{L}(\text{IND}) - \mathcal{L}(\text{ETOL})$.

REFERENCES

1. A. V. Aho, Indexed grammars - an extension of context-free grammars, Journal of the Assoc. for Comp. Mach., v. 15, 647-671, 1968.

2. P. J. Downey, Developmental systems and recursion schemes, Proceedings of the Conference on Biologically Motivated Automata Theory, McLean, Virginia, 1974.

3. A. Ehrenfeucht and G. Rozenberg, Some ETOL languages which are not deterministic, University of Colorado, Department of Computer Science, Technical Report #CU-CS-018-73, 1973.

4. A. Ehrenfeucht and G. Rozenberg, The number of occurrences of letters versus their distribution in some EOL languages, Information and Control, to appear.

5. A. Ehrenfeucht and G. Rozenberg, A characterization theorem for a subclass of ETOL languages, Acta Informatica, to appear.

6. A. Ehrenfeucht and G. Rozenberg, On structure of derivations in EDTOL systems, University of Colorado, Department of Computer Science, Technical Report #CU-CS-046-74, 1974.

7. A. Ehrenfeucht and G. Rozenberg, A pumping theorem for EDTOL languages, University of Colorado, Department of Computer Science, Technical Report #CU-CS-047-74, 1974.

8. A. Ehrenfeucht and G. Rozenberg, On some context free languages which are not EDTOL languages, University of Colorado, Department of Computer Science, Technical Report #CU-CS-047-74, 1974.

9. G. T. Herman and G. Rozenberg, Developmental systems and languages, North-Holland Publishing company, to appear, 1974.

10. J. Hopcroft and J. Ullman, Formal languages and their relation to automata, Addison Wesley, 1968.

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

12. G. Rozenberg, Extension of Tabled OL systems and languages, International Journal of Computer and Information Sciences, v. 2, 311-336, 1973.

13. G. Rozenberg and A. Salomaa (eds), L systems, Lecture Notes in Computer Science No. 15, Springer-Verlag, 1974.

14. A. Salomaa, Formal languages, Academic Press, 1973.

15. A. Salomaa, Recent results on L systems, Proceedings of the Conference on Biologically Motivated Automata Theory, McLean, Virginia, 1974.

16. A. Salomaa, Parallelism in rewriting systems, Proceedings of the Second Colloquium on Automata, Languages and Programming, Saarbrücken, July, 1974.