A RELATIONSHIP BETWEEN ETOL AND EDTOL LANGUAGES

by

A. Ehrenfeucht

G. Rozenberg

S. Skyum

DAIMI PB-40

November 1974

Institute of Mathematics University of Aarhus

DEPARTMENT OF COMPUTER SCIENCE

Ny Munkegade - 8000 Aarhus C - Denmark

Phone 06-12 83 55

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.

I. INTRODUCTION

The theory of L systems originated from the work of A. Lindenmayer (see Lindenmayer [13]). 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 [11], and Rozenberg and Salomaa [16]).

One of the central families of L languages (that is languages generated by L systems) is the family of ET0L languages (see, e.g. Downey [2], Rozenberg [15] and Salomaa [18]). An important research area in the theory of ET0L systems and languages is to provide results which would facilitate proofs that certain languages are not ET0L 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 ET0L languages. In fact it shows how, in certain cases, to reduce this problem to proving that some languages are not deterministic ET0L languages (see Rozenberg [15] and Ehrenfeucht and Rozenberg [6]). This is a great help indeed, because it is easier to investigate the structure of derivations in a deterministic ET0L system, and quite a number of examples of languages that are not deterministic ET0L 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 ET0L 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 [18] and Salomaa [19]).

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:

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

(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 (denotes $x \underset{G}{\Rightarrow} y$) 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 $y = \alpha_1 \ldots \alpha_k$. Also $\Lambda \Rightarrow \Lambda$. We say that x derives y in G (denoted $x \overset{*}{\underset{G}{\Rightarrow}} y$) if and only if either (i) there exists a sequence of words $x_0, x_1, \ldots, x_n$ in V^* (with $n > 1$) such that $x_0 = x$, $x_n = y$ and $x_0 \underset{G}{\Rightarrow} x_1 \underset{G}{\Rightarrow} x_2 \cdots \underset{G}{\Rightarrow} x_n$; 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 $L(G) = \{x \in \Sigma^* : \omega \overset{*}{\underset{G}{\Rightarrow}} x\}$.

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 $\langle a, \alpha \rangle \in P$. It is called a TOL system if $V = \Sigma$. (Thus a DTOL system denotes a deterministic TOL system.)

Definition 5

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

We shall use $\mathcal{L}$(ETOL), $\mathcal{L}$(TOL), $\mathcal{L}$(DTOL), and $\mathcal{L}$(EDTOL) to denote classes of ETOL, TOL, DTOL, and EDTOL languages respectively.

Definition 6

Let $G = \langle V, \mathcal{P}, \omega, \Sigma \rangle$ be an ETOL system. For P in $\mathcal{P}$ let C(P) denote the subset of 2^P defined as follows: for arbitrary T in 2^P, $T \in C(P)$ if and only if for every a in V there exists exactly one α in V^* such that $\langle a, \alpha \rangle \in T$.

Now an ETOL system $H = \langle V, \overline{\mathcal{P}}, \omega, \Sigma \rangle$ is called the combinatorially complete version of G if $\overline{\mathcal{P}} = \{T : T \in C(P) \text{ for some P in } \mathcal{P}\}$.

We have now the following obvious result.

Lemma 1

If H is the combinatorially complete version of G, then $L(H) \subseteq L(G)$.

Notation and Terminology

Let $G = \langle V, \mathcal{P}, \omega, \Sigma \rangle$ be an ET0L system.

(i) If $\langle a, \alpha \rangle$ is an element of some P in $\mathcal{P}$, then we call it a production (for a in P) and write $a \underset{P}{\rightarrow} \alpha$.

(ii) We can talk about words over $\mathcal{P}$ (thus elements from $\mathcal{P}^*$) representing functions from V^* into V^* in the obvious sense. Thus for a word x over V^* and for a word τ over $\mathcal{P}^*$ we use $\tau(x)$ to denote the set of all words that can be obtained from x when applying the sequence of tables τ.

(iii) A coding is a homomorphism which maps a letter into a letter.

Example 1

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

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

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

$P_3 = \{a \rightarrow F,\ b \rightarrow F,\ A \rightarrow a,\ B \rightarrow b,\ C \rightarrow \lambda,\ D \rightarrow \lambda,\ F \rightarrow F\}$.

G_1 is an EDT0L system and $L(G_1) = \{a^n b^m a^n : n \geq 0,\ m \geq n\}$.

Example 2

Let $G_2 = \langle \{a,b,A,A',B,B',C,C',F\}, \mathcal{P}, ABC, \{a,b\} \rangle$ where $\mathcal{P} = \{P\}$ and

$P = \{a \rightarrow F,\ b \rightarrow F,\ c \rightarrow F,\ A \rightarrow A'A,\ A \rightarrow a,\ B \rightarrow B'B,\ B \rightarrow b,\ C \rightarrow C'C,$
$C \rightarrow c,\ A' \rightarrow A',\ A' \rightarrow a,\ B' \rightarrow B',\ B' \rightarrow b,\ C' \rightarrow C',\ C' \rightarrow c,\ F \rightarrow F\}$.

G_2 is an ET0L system (but not an EDT0L system) and $L(G) = \{a^n b^n c^n : n \geq 1\}$.

III. RESULTS

In this section we shall present the main result of this paper (Theorem 2).

First however we need a definition (Definition 7) and an auxiliary result (Theorem 1) which is interesting on its own.

Definition 7

Let $G = \langle V, \mathcal{P}, \omega, V \rangle$ be a T0L system and h a homomorphism from V^* into Σ^*. Let b be in V.

(i) We say that b is a (G, h)-nondeterministic letter if the following conditions hold:

1. There exist words x_1, x_2, x_3 in V^* such that $x_1 b x_2 b x_3$ is in L(G).
2. There exist τ in $\mathcal{P}^*$ and y_1, y_2 in $\tau(b)$ such that $h(y_1) \neq h(y_2)$.

(ii) We say that G is h-deterministic if V does not contain (G, h)-nondeterministic letters.

Theorem 1

Let G be a T0L system over an alphabet V and let h be a homomorphism on V^*. If G is h-deterministic then there exists a DT0L system H such that $h(L(H)) = h(L(G))$.

Proof

Let H be the combinatorially complete version of G. From Lemma 1 it follows that $h(L(H)) \subseteq h(L(G))$.

On the other hand as G is h-deterministic, it is clear that $h(L(G)) \subseteq h(L(H))$ (this can be easily proved, and we leave the proof to the reader). Thus $h(L(H)) = h(L(G))$.

<u>Theorem 2</u>

Let Σ_1, Σ_2 be two disjoint alphabets. Let $K_1 \subseteq \Sigma_1^*$, $K_2 \subseteq \Sigma_2^*$ and let f be a surjective function from K_1 onto K_2. Let $K = \{wf(w) : w \in K_1\}$. (i) If K is an ETOL language then K_2 is an EDTOL language. (ii) Moreover if f is bijective then also K_1 is an EDTOL language.

<u>Proof</u>

Let us assume that K is an ETOL language. It is well known (see Ehrenfeucht and Rozenberg [10]) that each ETOL language is a coding of a TOL language. Thus there exist a TOL system $G = \langle V, \mathcal{P}, \omega, V \rangle$ and a coding h from V^* into Σ^* (where $\Sigma = \Sigma_1 \cup \Sigma_2$) such that $K = h(L(G))$. Let h_2 be a homomorphism from V^* into Σ^* defined as follows:

$$\text{for a in } V,\ h_2(a) = \begin{cases} h(a) \text{ if } h(a) \in \Sigma_2, \\ \Lambda \text{ otherwise}. \end{cases}$$

We shall prove first that G is h_2-deterministic. This will be accomplished once we have shown that for every b in V whenever $x_1bx_2bx_3 \in L(G)$ (for some x_1, x_2, x_3 in V^*) and $\tau \in \mathcal{P}^*$ then for every y_1, y_2 in $\tau(b)$ we have $h_2(y_1) = h_2(y_2)$. To prove this we have to consider 3 cases.

(i) $h(y_1) \in \Sigma_1^+$.

But, for every $\bar{x}_1$ in $\tau(x_1)$, $\bar{x}_2$ in $\tau(x_2)$ and $\bar{x}_3$ in $\tau(x_3)$, $h(\bar{x}_1)h(y_2)h(\bar{x}_2)h(y_1)h(\bar{x}_3) \in K$ and so $h(y_2) \in \Sigma_1^*$.

Thus $h_2(y_1) = h_2(y_2) = \Lambda$.

(ii) $h(y_1) \in \Sigma_2^+$.

But, for every $\bar{x}_1$ in $\tau(x_1)$, $\bar{x}_2$ in $\tau(x_2)$ and $\bar{x}_3$ in $\tau(x_3)$ and for every i, j in $\{1, 2\}$, $h(\bar{x}_1)h(y_i)h(\bar{x}_2)h(y_j)h(\bar{x}_3) \in K$ with $h(\bar{x}_1) = z_1 z_2$ for some z_1 in Σ_1^+ and z_2 in Σ_2^* where $f(z_1) = z_2 h(\bar{x}_1)h(y_i)h(\bar{x}_2)h(y_j)h(\bar{x}_3)$.

Thus $h(y_1) = h(y_2)$ and so $h_2(y_1) = h_2(y_2)$.

(iii) $h(y_1) = \Lambda$.

Note that if we assume now that $h(y_2) \in \Sigma_2^+$ then, almost repeating the reasoning from (ii) we get that $h(y_2) = h(y_1)$, a contradiction. Also it is clear that $h(y_2)$ cannot be in $\Sigma_1^+ \Sigma_2^+$. Thus $h(y_2) \in \Sigma_1^+$ and consequently $h_2(y_2) = \Lambda = h_2(y_1)$.

Now if one notices that $h(y_1)$ cannot be in $\Sigma_1^+ \Sigma_2^+$ then it is clear that the above three cases exhaust all possibilities. But in each of these cases we have $h_2(y_1) = h_2(y_2)$ which proves in fact that G is h_2-deterministic.

(i) Now the proof that K_2 is an EDT0L language goes as follows. The function f is an onto function and so $h_2(L(G)) = h_2(K) = \{f(w) : w \in K_1\} = K_2$. Thus by Theorem 1 there exists a DT0L system H such that $h_2(L(H)) = h_2(L(G)) = K_2$. But it is well known (see Nielsen, Rozenberg, Salomaa and Skyum [14], diagram D7) that if a language is a homomorphic image of a DT0L language then it is an EDT0L language. Consequently K_2 is an EDT0L language which completes the proof of part (i) of the theorem.

(ii) To prove that K_1 is an EDT0L language (if f is bijective) we proceed as follows. (For a word x, x_{mir} denotes the mirror image of x and for a language M, $M_{mir} = \{x_{mir} : x \in M\}$). Let f_{mir} be a function from $K_{1\,mir}$ into $K_{2\,mir}$ defined by $f_{mir}(x) = y$ if and only if $f(x_{mir}) = y_{mir}$. It is clear that f_{mir} is a bijection from $K_{1\,mir}$ onto $K_{2\,mir}$. But $K_{mir} = \{(f(w))_{mir}\, w_{mir} : w_{mir} \in K_1\} = \{x f_{mir}^{-1}(x) : x \in K_{2\,mir}\}$.

Applying Theorem 2 to the language K_{mir} we get that $K_{1\ mir}$ is an EDT0L language. But, obviously, the class of EDT0L languages is closed with respect to the operation of taking the mirror image and so K_1 must be an EDT0L language. Thus (ii) is proved.

IV. APPLICATIONS

We will provide now several examples of languages which are not ET0L languages. As we will do it with the use of Theorem 2, let us first recall some examples of languages that are not in $\mathcal{L}$(EDT0L).

Lemma 2 (Ehrenfeucht and Rozenberg [3] or Ehrenfeucht and Rozenberg [7].)

Let $W_1 = \{x \in \{0,1\}^+ : |x| = 2^n \text{ for some } n \geq 0\}$. $W_1 \notin \mathcal{L}$(EDT0L).

Lemma 3 (Ehrenfeucht and Rozenberg [8])

Let, for each $i \geq 1$, $\Sigma_i = \{[_1, \ldots, [_i,]_i, \ldots,]_1\}$ and let $\mathcal{B}_i$ be the language generated by the context-free grammar $H(\mathcal{B}_i) = \langle \{S_i\}, \Sigma_i, P_i, S \rangle$, where

$P_i = \{S \to [_j SS]_j : 1 \leq j \leq i\} \cup \{S \to [_j S]_j : 1 \leq j \leq i\} \cup \{S \to [_j \,]_j : 1 \leq j \leq i\}$.

For every $i \geq 1$, $\mathcal{B}_i$ is not in $\mathcal{L}$(EDT0L).

For our next result we assume the reader to be familiar with the notion of a Dyck language as defined, e.g., in Salomaa [1], p. 210.

Lemma 4 (Ehrenfeucht and Rozenberg [8])

If K is a Dyck language over an alphabet of at least eight letters then K is not in $\mathcal{L}$(EDT0L).

Now we are ready to prove the following results.

Let τ be an enumeration (possible with repetitions) of all words from W_1 (so τ is a function from positive integers onto W_1). Let $\Sigma_1 = \{a\}$ and let $Z_1 = \{a^n w : w \in W_1 \text{ and } \tau(w) = n\}$.

Proposition 1 $Z_1 \notin \mathcal{L}(ETOL)$.

Proof

This follows directly from Theorem 2 (i) and Lemma 2.

Now let us assume that τ is an enumeration without repetitions of all words from W_1 (so τ is a bijection from positive integers onto W_1). Let $\Sigma_1 = \{a\}$ and let $Z_2 = \{wa^n : w \in W_1 \text{ and } \tau(w) = n\}$.

Proposition 2 $Z_2 \notin \mathcal{L}(ETOL)$.

Proof

As τ^{-1} must be a bijection, if Z_2 is an ETOL language, then (by Theorem 2 (ii)) W_1 must be an EDTOL language which contradicts Lemma 2. Thus Z_2 is not an ETOL language. (Note that Theorem 2 (i) alone was not sufficient for the direct proof of this proposition.)

Finally we can settle a quite important open problem of long standing (see, e.g. Downey [2] and Salomaa [19]) whether or not the class of indexed languages (see Aho [1]). Let $\mathcal{L}(IND)$ denote the class of indexed languages. (Now we assume that the reader is familiar with Aho [1].)

Thoorem 3

Let Σ be a finite alphabet and let $\overline{\Sigma} = \{\overline{a} : a \in \Sigma\}$. Let h be a homomorphism from Σ^* onto $\overline{\Sigma}^*$ defined by $h(a) = \overline{a}$, 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 $\mathcal{L}(IND)$

but is not in $\mathcal{L}$(EDT0L).

Proof

If a language is context-free then it can be generated by a right linear grammar (see Aho [1], Lemma 6.1). Thus, obviously, $M_K \in \mathcal{L}(\text{IND})$. On the other hand Theorem 2 implies that M_K is not in $\mathcal{L}$(EDT0L).

Now, Theorem 3, Lemma 3 and Lemma 4 imply the following results.

Corollary 1

For every $i \geq 1$, $M_{\mathcal{B}_i} \in \mathcal{L}(\text{IND}) - \mathcal{L}(\text{ET0L})$.

Corollary 2

If K is a Dyck language over an alphabet of at least eight letters, then $M_K \in \mathcal{L}(\text{IND}) - \mathcal{L}(\text{ET0L})$.

We end this paper with the following two remarks.

Remark 1

It is shown in Skyum [20], that the result presented in Theorem 2 is quite typical for several familis of parallel languages (in the sense of Salomaa [19]).

Remark 2

A little bit stronger version of Theorem 2 is proved in Ehrenfeucht and Rozenberg [9]. The proof presented there is quite longer than the proof presented in this paper, however, it is using a very different idea.

Acknowledgement

The two first 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.

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 ET0L languages which are not deterministic, University of Colorado, Dept. 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 E0L languages, Information and Control, to appear.

5. A. Ehrenfeucht and G. Rozenberg, A characterization theorem for a subclass of ET0L languages, Acta Informatica, to appear.

6. A. Ehrenfeucht and G. Rozenberg, On structure of derivations in EDT0L systems, University of Colorado, Dept. of Computer Science, Technical Report #CU-CS-046-74, 1974.

7. A. Ehrenfeucht and G. Rozenberg, A pumping theorem for EDT0L languages, University of Colorado, Dept. of Computer Science, Technical Report #CU-CS-047-74, 1974.

8. A. Ehrenfeucht and G. Rozenberg, On some context-free languages which are not EDT0L languages, University of Colorado, Dept. of Computer Science, Technical Report #CU-CS-048-74, 1974.

9. A. Ehrenfeucht and G. Rozenberg, On decomposing some ET0L languages into deterministic ET0L languages, University of Colorado, Dept. of Computer Science, Technical Report #CU-CS-044-74, 1974.

10. A. Ehrenfeucht and G. Rozenberg, Nonterminals versus homomorphisms, in defining languages for some classes of rewriting systems, Acta Informatica, v. 3, 265-283, 1974.

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

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

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

14. M. Nielsen, G. Rozenberg, A. Salomaa and S. Skyum, Nonterminals, homomorphisms and codings in different variations of 0L systems, Parts I and II, Acta Informatica, to appear.

15. G. Rozenberg, Extension of Tables 0L systems and languages, International Journal of Computer and Information Sciences, v. 2, 311-336, 1973.

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

17. A. Salomaa, Formal Languages, Academic Press, 1973.

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

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

20. S. Skyum, Decomposition-theorems for various kinds of languages parallel in nature, Ph.D. thesis, University of Aarhus, Dept. of Computer Science, 1974.

A. Ehrenfeucht
Department of Computer Science
University of Colorado
Boulder, Colorado 80302
USA

G. Rozenberg
Institute of Mathematics
University of Utrecht
Utrecht – De Uithof

or

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

Sven Skyum
Department of Computer Science
University of Aarhus
Ny Munkegade
8000 Aarhus C
Denmark