SOME ETOL LANGUAGES WHICH ARE NOT DETERMINISTIC +

by

A. Ehrenfeucht *
G. Rozenberg **

* Department of Computer Science
University of Colorado
** Department of Computer Science
State University of New York

TR #CU-CS-018-73 May, 1973

All correspondence to:

G. Rozenberg
Institute of Mathematics
Utrecht University
Utrecht - Uithof
The Netherlands

+ This work supported by National Science Foundation Grant # GJ-660

SOME ETOL LANGUAGES WHICH ARE NOT DETERMINISTIC +

by

A. Ehrenfeucht *
G. Rozenberg **

* Department of Computer Science
University of Colorado
** Department of Computer Science
State University of New York

TR #CU-CS-018-73 May, 1973

All correspondence to:

G. Rozenberg
Institute of Mathematics
Utrecht University
Utrecht - Uithof
The Netherlands

+ This work supported by National Science Foundation Grant # GJ-660

ABSTRACT

It is shown that there exist 0L languages which cannot be generated by deterministic ET0L systems. This solves an open problem posed in Rozenberg (1973).

0. INTRODUCTION

Developmental systems and languages have recently gained considerable attention in both formal language theory and theoretical biology (see, e.g., Lindenmayer (1968), Lindenmayer & Rozenberg (1972), Herman, Lindenmayer & Rozenberg (1973) and their references).

From the biological point of view, developmental systems have provided a useful theoretical framework within which the nature of cellular behavior in development can be discussed, computed and compared. From the formal language theory point of view they have provided us with an alternative to the now standard Chomsky framework (see, e.g., Hopcroft & Ullman, (1969)) for defining languages.

The concept of a tabled 0L system was introduced by Rozenberg (1972) to model the role of environment in developmental processes, and it is currently the subject of active research, see, e.g. Rozenberg (1972, 1973) or Ehrenfeucht & Rozenberg (1973). In this paper we are concerned with a kind of tabled 0L systems called an extended tabled 0L system. We use the abbreviations T0L system for tabled 0L system and ET0L system for extended tabled 0L system.

An ET0L system has the following components:

(i) A finite set of symbols, V, the <u>alphabet</u>.

(ii) A starting string ω, the <u>axiom</u>.

(iii) A finite set of <u>tables</u> of production rules, each of which tells us by what strings in V^* a symbol from V may be replaced. The set of productions that may be applied to a certain symbol depends on the symbol only. In every step of a derivation all symbols in the string must be simultaneously replaced according to the production rules of one single arbitrarily chosen table of the system.

(iv) A subset Σ of the alphabet V, the target alphabet. The language of a system (called an ETOL language) is the set of all strings over the target alphabet which can be derived from the axiom, possibly including the axiom itself.

An ETOL system is called deterministic, abbreviated as an EDTOL system if each of its tables is such that for each letter in the alphabet V the table provides exactly one string by which the letter can be replaced.

The role of the deterministic restriction in developmental systems is an important aspect of the theory from both the biological and formal language theory points of view. This problem in connection with tabled systems was investigated among others in Rozenberg (1972), Rozenberg (1973) and in Ehrenfeucht & Rozenberg (1973).

In particular it was proved in Rozenberg (1973), that every ETOL language can be generated by an ETOL system such that in every table there are no more than two productions provided for each symbol in the alphabet. In the same paper it was left as an open problem whether each ETOL language can be generated by a deterministic ETOL system.

In this paper we prove that the solution to the above problem is negative, it means that there is an ET0L language which cannot be generated by any EDT0L system.

I. DEFINITIONS AND EXAMPLES

In this section we introduce all of the necessary terminology and illustrate it by examples. We use conventional formal language theory notation, see e.g. Hopcroft & Ullman (1969), augmented by the following.

(i) If A is a set, then #A denotes the cardinality of A,

(ii) If x is a word over an alphabet Σ, $x=a_1\dots a_p$ for $p\geq 1, a_1,\dots,a_p \in \Sigma$, then

$Min(x)=\{a \text{ in } \Sigma: a \text{ occurs in } x\}$, $|x|$ denotes the length of x, and if l is a positive integer then

$$Pref_l(x)=\begin{cases} a_1\dots a_l & \text{if } l \leq p, \\ x & \text{otherwise.}\end{cases}$$

x^e will abbreviate the word x catenated with itself e times.

(iii) Λ denotes the empty word.

(iv) N denotes the set of nonnegative integers and $N^+=N-\{0\}$.

Definition 1. An extended tabled 0L system (abbreviated as an ET0L system) is a 4-tuple $G=\langle V,\mathcal{P},\omega,\Sigma\rangle$, where

V is a finite nonempty set ... (the alphabet of G),

ω is an element of V^+ ... (the axiom of G),

Σ is a subset of V ... (the target alphabet of G),

$\mathcal{P}$ is a finite nonempty set, each element of which (called a table) is a finite nonempty binary relation included in $\Sigma\times\Sigma^*$.

$\mathcal{P}$ satisfies the following (completeness) condition:

for every P in $\mathcal{P}$ and for every a in V there exists α in V^* such that $\langle a,\alpha\rangle$ is in P.

If P is a table and $\langle a,\alpha\rangle \in P$ then $\langle a,\alpha\rangle$ is called a <u>production</u> and denoted as $a\to\alpha$. We shall write $a \underset{P}{\to} \alpha$ for "$a\to\alpha$ is in P."

<u>Definition 2.</u> Let $G=\langle V,\mathcal{P},\omega,\Sigma\rangle$ be an ET0L system.

I) Let $x\in V^+$, $y\in V^*$ where $x=a_1\ldots a_p$ for $a_1,\ldots,a_p\in V$.

We say that <u>x directly derives y (in G)</u>, denoted as $x \underset{G}{\Rightarrow} y$, if there exists a table P in $\mathcal{P}$ such that $a_1 \underset{P}{\to} \alpha_1,\ldots,a_p \underset{P}{\to} \alpha_p$ for some $\alpha_1,\ldots,\alpha_p$ such that $y=\alpha_1\ldots\alpha_p$. Also by convention $\Lambda \underset{P}{\Rightarrow} \Lambda$.

II) Let $x,y\in V^*$.

We say that <u>x derives y (in G)</u>, denoted as $x \overset{*}{\underset{G}{\Rightarrow}} y$, if either $x=y$ or there exists a sequence $x_o,\ldots,x_n$ of words in V^* such that $x_o=x$, $x_n=y$ and $x_{i-1} \underset{G}{\Rightarrow} x_i$ for $1\leq i\leq n$.

III) The <u>language of G</u> or the <u>language generated by G</u>, denoted as $L(G)$, is defined as follows

$$L(G)=\{x\in\Sigma^*: \omega \overset{*}{\underset{G}{\Rightarrow}} x\}.$$

(Whenever G is understood we write $x\Rightarrow y$ and $x\overset{*}{\Rightarrow}y$ rather than $x \underset{G}{\Rightarrow} y$ and $x \overset{*}{\underset{G}{\Rightarrow}} y$ respectively).

<u>Definition 3.</u> Let $G=\langle V,\mathcal{P},\omega,\Sigma\rangle$ be an ET0L system.

I) G is called a <u>deterministic ET0L system</u>, abbreviated as an EDT0L system, if for every P in $\mathcal{P}$ and every a in V there exists exactly one α in V^* such that $a \underset{P}{\to} \alpha$.

II) G is called a <u>0L system</u> if $V=\Sigma$ and P contains exactly one table.

<u>Definition 4.</u> Let Σ be a finite alphabet and $L \subseteq \Sigma^*$. L is called a <u>(deterministic) ET0L language</u> or a <u>0L language</u> if there exists a (deterministic) ET0L system or a 0L system G respectively such that $L(G)=L$.

By F(ET0L), F(EDT0L) and F(0L) we shall denote the classes of ET0L languages, EDT0L languages and 0L languages respectively.

<u>Example 1.</u> Let $G=\langle\{a,b\},\{P_1,P_2\},a,\{a,b\}\rangle$ be an ET0L system where $P_1=\{a\to a^2, b\to b\}, P_2=\{a\to a, a\to ab, a\to ba, b\to b\}$.

$L(G)=\{x\in\{a,b\}^+$: the number of occurrences of the letter a in x equals 2^n for some $n\geq 0\}$.

(Note that G is neither an EDT0L nor a 0L system).

<u>Example 2.</u> Let $G=\langle\{a,b\},\{P\},a,\{a,b\}\rangle$ be a 0L system where $P=\{a\to b, a\to a, a\to ab, a\to aa, b\to b\}$. $L(G)=\{a,b\}^+$.

Now, following Herman & Rozenberg (1973), we shall define derivations in ET0L systems.

<u>Definition 5.</u> Let $G=\langle V,P,\omega,\Sigma\rangle$ be an ET0L system. A <u>derivation</u> D in G is a triple $\langle O,f,g\rangle$ where O is a set of ordered pairs of nonnegative integers (the <u>occurrences</u> in D), f is a function from O into V ($f(i,j)$ is the <u>value</u> of D at occurrence $\langle i,j\rangle$), and g is a function from a subset of O into $P\times\bigcup_{P\in P} P$ satisfying the following conditions. There exists a sequence of words $(x_0,x_1,\ldots,x_s)$ in Σ^* (called the <u>trace of D</u>

and denoted as tr(D)) such that

(i) $O=\{\langle i,j\rangle : 0\leq i\leq s,\ 1\leq j\leq |x_i|\}$,

(ii) $f(i,j)$ is the j'th symbol in x_i,

(iii) the domain of g is $\{\langle i,j\rangle : 0\leq i<s,\ 1\leq j\leq |x_i|\}$, (elements of O which are not in the domain of g are called <u>final occurrences</u>),

(iv) for $0\leq i<s$ there exists a P in $\mathcal{P}$ such that for $0\leq j\leq |x_i|$, $g(i,j)=\langle P, f(i,j)\to\alpha_j\rangle$, where $f(i,j)\underset{P}{\to}\alpha_j$ and $\alpha_1\ldots\alpha_{|x_i|}=x_{i+1}$.

In such a case D is said to be a <u>derivation of x_s from x_o.</u> The sequence of tables $T_1,\ldots,T_s$ such that the first element of $g(i,j)$ for $0\leq i\leq s-1, \langle i,j\rangle \in O$ is T_{i+1} is called the <u>control sequence of D</u>. The sequence $T_j,T_{j+1},\ldots,T_m$ where $1\leq j<m\leq s$ is called the <u>control sequence from x_{j-1} to x_m</u>. If $\langle i,j\rangle$ is an occurrence in D then we also say that x_i contains this occurrence.

The set $\{|x_i| : 0\leq i\leq s\}$ is called the set of lengths in D and is denoted by Lg(D).

<u>Example 3</u>. Let G be the ET0L system from Example 1. Let $D_1=\langle O_1,f_1,g_1\rangle$ where

$O_1=\{\langle 0,1\rangle,\langle 1,1\rangle,\langle 1,2\rangle,\langle 2,1\rangle,\langle 2,2\rangle,\langle 2,3\rangle,\langle 3,1\rangle,\langle 3,2\rangle,\langle 3,3\rangle,\langle 3,4\rangle,\langle 3,5\rangle\}$,

$f(0,1)=f(1,1)=f(2,1)=f(2,2)=f(3,2)=f(3,3)=a$,

$f(1,2)=f(2,3)=f(3,1)=f(3,4)=f(3,5)=b$

$g(0,1)=g(2,2)=\langle P_2, a\to ab\rangle$, $g(1,1)=\langle P_1, a\to aa\rangle$, $g(1,2)=\langle P_1, b\to b\rangle$,

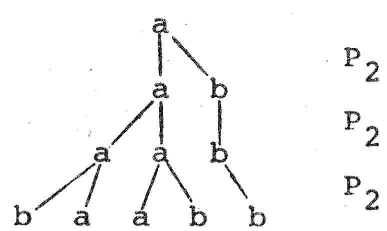
$g(2,1)=\langle P_2, a\to ba\rangle$, $g(2,3)=\langle P_2, b\to b\rangle$

D_1 is a derivation of ba^2b^2 from a. The trace of D_1 is

a, ab, a^2b, ba^2b^2. The control sequence of D_1 is P_2, P_1, P_2. {<3,1>,<3,2>,<3,3>,<3,4>,<3,5>} is the set of final occurences.

Given an ETOL system G and a derivation D in G, there is a natural way of representing D by a graph like structure called the derivation graph of D. Although such a representation can be defined formally, we believe that for the purpose of this paper this and other related notions are best explained by examples.

Example 4. The derivation in Example 3 is represented by the following derivation graph:

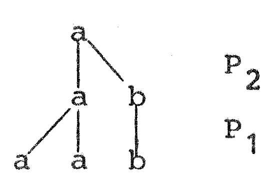


The occurrence <1,2> is a direct ancestor of the occurrence <2,3> and is an ancestor of the occurrence <3,5>. Thus the occurrence <2,3> is the direct descendant of the occurrence <1,2>, and the occurrence <3,5> is a descendant of the occurrence <1,2>. The trace of this derivation is x_0, x_1, x_2, x_3 and in fact we shall also say that <3,5> is a descendant of <1,2> in x_3, and so on. We say that, for example, the occurrence <1,2> contributes two occurrences (<2,1> and <2,2>) on the level x_2 (or simply on the level 2). In fact in this case we can say that the occurrence <1,2> contributes on the level x_2 two occurrences in one step. The number of occurrences contribured on the "final" level (x_3)

by an occurrence <i,j> is called the weight of <i,j>. If the weight of a given occurrence in a derivation D is equal to 0, then this occurrence is called D-improductive (and otherwise it is called D-productive). If x_i is a word in the trace of a derivation D and we 'erase' all improductive occurrences of letters in x_i then the resulting word is denoted by $\mathrm{Prod}(x_i)$. We define $\mathrm{Lg}(\mathrm{Prod}(D))$ to be the set $\{|\mathrm{Prod}(x_i)| : 0 \leq i \leq s\}$.

Note that derivations and derivation graphs are in one-to-one correspondence in the sense that, given a derivation D one can construct the unique derivation graph M which represents D, and conversly, given a derivation graph M, one can uniquely construct the unique derivation D which is represented by M.

By looking at derivation graphs it is easy to see that each occurrence <i,j> in a derivation determines a unique <i,j>-subderivation from f(i,j) into a substring of x_s (where $x_0,\ldots,x_s$ is the trace of the given derivation). We denote such a subderivation by D(i,j). If we are interested in the derivation 'leading' from x_0 to x_e only (where $0 \leq e \leq s$) then we denote it by D^e. (Thus $D^s = D$). For example if e=2 and D is a derivation from Example 4 then D^2 is represented by the following derivation graph



If we consider an EDT0L system then, obviously, the first element of the trace of a given derivation D together with the control sequence of D uniquely determine the derivation.

In such a case if $tr(D)=x_0,\ldots,x_s$ then we write $x_0=\omega=\omega^{[0]}, x_1=\omega^{[1]}, \ldots, x_s=\omega^{[s]}$, where the EDT0L system under consideration has the axiom ω.

II. PRELIMINARY RESULTS

In this section we shall prove a number of preliminary results needed for the proof of the main theorem (Theorem 1).

For the rest of this paper let $L=\{x\varepsilon\{0,1\}^*: |x|=2^n$ for some $n\geq 0\}$.

Lemma 1. For every n in N^+ there exists a word $x(n)$ in L such that $|x(n)|=2^n$ and $x(n)$ does not contain any two non-overlapping occurrences of the same word of length $3n$.

Proof.

Let $n\varepsilon N^+$ and let $y(n)$ be the following word:

$$y(n)=0^n\alpha_1 0^n\alpha_2 \ldots 0^n\alpha_{2^n}0^n$$

where $\alpha_1,\ldots,\alpha_{2^n}$ is a listing (without repetitions) of the set of all words of length n.

Now if we let

$$x(n)=\mathrm{Pref}_{2^n}(y(n))$$

then Lemma 1 obviously holds.

Through a series of lemmata (lemmata 2 through 13) we shall show that the assumption that $L=L(G)$ for some EDT0L system G leads to a contradiction (Lemma 13). In fact we shall show that each EDT0L system which generates $x(n)$ for n "large enough" must also generate a word whose length is not a power of 2.

Thus let us assume that $G=\langle V,P,\omega,\Sigma\rangle$ is an EDT0L system such that $L(G)=L$. Let $\#V=k$, $|\omega|=r$, $m=\max_{P\varepsilon P}\{|\alpha|:a\underset{P}{\rightarrow}\alpha$ for some a in $V\}$.

Let n be a "large enough" positive integer (how large n should be is defined in the proofs of lemmata 7 and 14) and let us consider the word x(n) defined in the proof of Lemma 1. Let D=<O,f,g> be a derivation of x(n) in G and $\omega=\omega_0,\omega_1,\ldots,\omega_s$ be the trace of D.

Let <i,j> be a D-productive occurrence in D.

If <i,j> is not final and the last word in the trace of D(i,j) is of length greater than or equal to 3n, then <i,j> is called a <u>large occurrence (in D)</u>. Otherwise it is called a <u>small occurrence (in D)</u>.

It follows directly from the definitions that

<u>Lemma 2</u>. Each descendant of a small occurrence is a small occurrence.

<u>Lemma 3</u>. Let <i,j>,<i,j'> be in O. If f(i,j)=f(i,j') then <i,j> is a small occurrence if, and only if, <i,j'> is a small occurrence. (Thus we can talk about big and small letters on a given level i).

<u>Lemma 4</u>. Let <i,j> be a big occurrence in D. If j≠j' and <i,j'> is in O, then f(i,j)≠f(i,j').

<u>Proof</u>.

If <i,j> and <i,j'> are in 0 and <i,j> is a big occurrence then (Lemma 3) so is <i,j'>. But G is deterministic and so if j≠j' and f(i,j)=f(i,j'), then x(n) would contain two non-overlapping occurrences of the same subword of length 3n; a contradiction.

Thus as a direct consequence of Lemma 4 we have the following result.

Lemma 5. For every i in N^+ there are at most k big occurrences of the form $\langle i,j \rangle$.

Now, let us estimate the length of $Prod(\omega^{[l]})$ for a given $l \leq s$.

Let $U(l)$ be an arbitrary (but fixed) subset of $\{0,1,\ldots,l\}$ such that for every p in $Lg(Prod(D^l))$, $U(l)$ contains exactly one element v in $\{0,1,\ldots,l\}$ such that $|Prod(\omega^{[v]})|=p$. Let $u_l = \#U(l)$.

Lemma 6. For $l \leq s$, $|Prod(\omega^{[l]})| \leq r \cdot 3 \cdot n + k \cdot m \cdot 3 \cdot n \cdot u_l + k$.

Proof.

(i) The axiom ω contains not more than r small occurrences and each of them contributes not more than 3n small occurrences on a given level l. Hence the component $r \cdot 3 \cdot n$.

(ii) If $p \in U(l)$ then (see Lemma 5) $\omega^{[p]}$ contains at most k big occurrences each of which can introduce in one step not more than m small occurrences. Each of these can contribute not more than $3 \cdot n$ small occurrences on the given level l. But $\#U(l) = u_l$. Hence the component $k \cdot m \cdot 3 \cdot n \cdot u_l$.

(iii) From Lemma 5 it follows that $\omega^{[l]}$ contains at most k big occurrences. Hence the component k.

Lemma 7. $u_s \geq \frac{2^n}{C.n}$ for some positive integer.

Proof.

From Lemma 6 it follows that

$$2^n = |\mathrm{Prod}(\omega^{[s]})| \leq r\cdot 3\cdot n + k\cdot m\cdot 3\cdot n\cdot u_s + k$$

Hence

$$u_s \geq \frac{2^n-(k+r\cdot 3\cdot n)}{k\cdot m\cdot 3n} = \frac{2^n}{k\cdot m\cdot 3n} - \left(\frac{1}{m\cdot 3\cdot n} + \frac{r}{k\cdot m}\right) \geq \frac{2^n}{C\cdot n}$$

for some (large enough) positive integer C, and n "large enough"

so that $\frac{2^n}{k\cdot m\cdot 3n} - \left(\frac{1}{m\cdot 3n} + \frac{r}{k\cdot m}\right) > 0$.

Now for $0 \leq l \leq s$ let us introduce the following notation:

$B(\omega^{[l]})$ denotes the number of big occurrences in $\omega^{[l]}$,

$S(\omega^{[l]})$ denotes the number of small occurrences in $\omega^{[l]}$,

$I(\omega^{[l]})$ denotes the number of D-improductive occurrences in $\omega^{[l]}$,

Using this notation let us introduce the following definition. Two elements l_1, l_2 of $U(s)$ are called compatible if:

(i) $\mathrm{Min}(\omega^{[l_1]}) = \mathrm{Min}(\omega^{[l_2]})$,

(ii) $B(\omega^{[l_1]}) = B(\omega^{[l_2]})$,

(iii) $S(\omega^{[l_1]}) = S(\omega^{[l_2]})$,

(iv) $I(\omega^{[l_1]}) = I(\omega^{[l_2]})$,

(v) If $a \in S(\omega^{[l_1]}) = S(\omega^{[l_2]})$ then a contributes the same number of occurrences in $\omega^{[s]}$ from level l_1 and level l_2.

(We also say that l_1, l_2 are $\langle \mathrm{Min}(\omega^{[l_1]}), B(\omega^{[l_1]}), S(\omega^{[l_1]}), I(\omega^{[l_1]})\rangle$-compatible.

If $\bar{U} \subseteq U(s)$, then U is called compatible if for some subsets A_1,A_2,A_3,A_4 of V, $\bar{U}$ is the set of all pairwise $\langle A_1,A_2,A_3,A_4\rangle$-compatible elements from $U(s)$. (We also say that is $\langle A_1,A_2,A_3,A_4\rangle$-compatible).

Lemma 8. If $\bar{U}$ is a compatible subset of $U(s)$ then

$$\# \bar{U} \geq \frac{2^n}{C_1 \cdot n^{k+1} \cdot 2^k \cdot 3^k}$$

where C_1 is a positive integer constant.

Proof.

Let $\bar{U}$ be an $\langle A_1,A_2,A_3,A_4\rangle$-compatible subset of $U(s)$.

(i) From Lemma 7 it follows that $u_s \geq \frac{2^n}{C \cdot n}$ for some positive integer C.

(ii) There are no more than 2^k choices for A_1.

(iii) For the given A_1 there are no more than 3^k choices of A_2,A_3,A_4 out of A_1.

(iv) For the given choice of A_3 one can choose at most $(3n)^k$ different lengths of contributions into level ω_s.

(v) Thus $\# U \geq \frac{2^n}{C \cdot n \cdot 2^k \cdot 3^k \cdot (3n)^k} = \frac{2^n}{C_1 \cdot n^{k+1} \cdot 2^k \cdot 3^k}$

where $C_1 = C \cdot 3^k$.

So let $\bar{U}$ be a compatible subset of $U(s)$, say $\bar{U} = \{p_1,\ldots,p_t\}$, where $p_1,\ldots,p_t$ are in ascending order.

Let $p_i,p_{i+1} \in \bar{U}$. We say that a small occurrence in $\omega_{p_{i+1}}$ is new if its ancestor in ω_{p_i} is a big occurrence.

Lemma 9. If $p_i, p_{i+1} \varepsilon \bar{U}$ then $\omega_{p_{i+1}}$ contains at least one new occurrence.

Proof.

Let us assume to the contrary, that $\omega_{p_{i+1}}$ does not contain any new occurrence. We shall prove (by induction on the weight of a small occurrence) that under this assumption each small occurrence in ω_{p_i} has exactly one descendant in $\omega_{p_{i+1}}$.

(i) Let e be the maximal weight for a small occurrence in ω_{p_i} and let ω_{p_i} contain v letters with this weight. But $\omega_{p_{i+1}}$ does not contain any new occurrence and so if a small letter in ω_{p_i} with weight e has more than one descendant in $\omega_{p_{i+1}}$, then each of them has a weight smaller than e. Hence (remembering that a letter with weight e can only be introduced by a letter with weight e) $\omega_{p_{i+1}}$ contains at most v-1 letters with weight e which contradicts the compatibility of $\bar{U}$.

(ii) Let us assume that the claim is true for all small letters in ω_{p_i} with weight equal to or greater than w.

(iii) Let a be a small letter in ω_{p_i} with weight w-1 (where $w-1 \geq 1$). By the inductive assumption no small letter in ω_{p_i} with weight equal to or greater than w can introduce a letter with weight (w-1) in $\omega_{p_{i+1}}$. Hence if the letter a has more than one descendant in $\omega_{p_{i+1}}$ then the number of small letters with weight w-1 in ω_{p_i} and in $\omega_{p_{i+1}}$ is different, which contradicts the compatibility of $\bar{U}$.

So we have proved that (under our assumption) each small occurrence in ω_{p_i} has exactly one descendant in $\omega_{p_{i+1}}$. But then

$$|Prod(\omega^{[p_i]})|=|Prod(\omega^{[p_{i+1}]})|$$

which contradicts the fact that $\bar{U} \subseteq U(s)$.

Thus $\omega_{p_{i+1}}$ must contain at least one new small occurrence, and consequently the result holds.

Out of all possible choices of $\bar{U}$ (first the choice of U (s) and then of $\bar{U}$ out of U(s)) let us choose one which will give us a minimal value for $p_{i+1}-p_i$ for some p_{i+1}, p_i in $\bar{U}$ (where $\bar{U} = \{p_1, \ldots, p_t\}$). Let p_{i_o}, p_{i_o+1} be such a minimal choice of consecutive "levels" from the so chosen $\bar{U}$.

<u>Lemma 10</u>. $p_{i_o+1}-p_{i_o} \leq C_2 \cdot n^{k+1}$ for some positive integer C_2.

<u>Proof</u>.

As $|x(n)|=2^n$, $\#Lg(D) \leq 2^n$.

From Lemma 8 it follows that $\# U \geq \dfrac{2^n}{C_1 \cdot n^{k+1} \cdot 2^k \cdot 3^k}$

Thus $p_{i_o+1}-p_{i_o} \leq \dfrac{2^n \cdot C_1 \cdot n^{k+1} \cdot 2^k \cdot 3^k}{2^n} = C_2 \cdot n^{k+1}$, where $C_2=C_1 \cdot 2^k \cdot 3^k$.

<u>Lemma 11</u>. The number of new occurrences of small letters in $\omega_{p_{i_o}+1}$ (denoted as $New(p_{i_o}+1)$) satisfies the inequality

$$New(p_{i_o}+1) \leq C_3 \cdot n^{k+2}$$

for some positive integer C_3.

<u>Proof</u>.

(i) From Lemma 10 it follows that $p_{i_o+1}-p_{i_o} \leq C_2 \cdot n^{k+1}$ for some positive integer C_2.

(ii) According to Lemma 5 there are at most k big occurrences on any "level" in D.

(iii) Each big letter in one step of a derivation can introduce not more than m small occurrences.

(iv) Each small occurrence can contribute at most $3 \cdot n$ occurrences on level $\omega_{p_{i_o}+1}$.

(v) Altogether

$$New(p_{i_o}+1) \leq C_2 \cdot n^{k+1} \cdot k \cdot m \cdot 3 \cdot n = C_3 \cdot n^{k+2}$$

where $C_3 = C_2 \cdot k \cdot m \cdot 3$.

Let $T_1, T_2, \ldots, T_s$ be the control sequence of D.

Let $D(p_i, p_{i+1})$ be the derivation defined by the axiom ω and the control sequence $T_1, T_2, \ldots, T_{p_i-1}, T_{p_{i+1}}, T_{p_{i+1}+1}, \ldots, T_s$.

Let $\omega = z_o, z_1, \ldots, z_h$ be the trace of $D(p_i, p_{i+1})$.

Lemma 12. z_h is in $L(G)$ and $|x(n)|-|z_h| \leq C_4 \cdot n^{k+3}$, where C_4 is a positive integer constant.

Proof.

z_h is in $L(G)$ because G is deterministic and p_i, p_{i+1} are in the same compatible set $\overline{U}$.

(i) $|x(n)|=|\omega_s|= X_1+X_2$, where

X_1 is the number of occurrences in ω_s contributed (in D) by small occurrences in ω_{p_i}, and X_2 is the number of occurrences in ω_s contributed (in D) by big occurrences in ω_{p_i}.

(ii) $|z_h|=Y_1+Y_2$, where

Y_1 is the number of occurrences in z_h contributed in $D(p_i,p_{i+1})$ by small occurrences in $z_{p_i-1}=\omega_{p_i-1}$, and

Y_2 is the number of occurrences in z_h contributed in $D(p_i,p_{i+1})$ by big occurrences in $z_h=\omega_{p_i}$. (The expressions "big occurrences" and "small occurrences" refer here to the classification of occurrences in $z_{p_i}=\omega_{p_i}$ in D and <u>not</u> in $D(p_i,p_{i+1})$).

(iii) Because p_i and p_{i+1} are compatible, each letter a which is small in both ω_{p_i} and $\omega_{p_{i+1}}$ contributes the same number of occurrences in ω_s independently of whether it contributes from ω_{p_i} or from $\omega_{p_{i+1}}$. But the control sequence from $\omega_{p_{i+1}}$ to ω_s in D is identical to the control sequence from z_{p_i} to z_h

in $D(p_i, p_{i+1})$. Hence $X_1=Y_1$.

(iv) $X_2=X_{21}+X_{22}$, where

X_{21} is the number of occurrences in ω_s contributed (in D) by big occurrences in $\omega_{p_{i+1}}$, and

X_{22} is the number of occurrences in ω_s contributed (in D) by new occurrences in $\omega_{p_{i+1}}$.

(v) Because p_i and p_{i+1} are compatible, $B(\omega_{p_i}=z_{p_i})=B(\omega_{p_{i+1}})$. But the control sequence from $\omega_{p_{i+1}}$ to ω_s(in D) is identical to the control sequence from z_{p_i} to z_h in $D(p_i, p_{i+1})$. Thus $X_{21}=Y_2$.

(vi) Thus $|x(n)|-|z_h|\leq 3\cdot n\cdot X_{22}$, and so from Lemma 11 it follows that $|x(n)|-|z_h|\leq 3\cdot n\cdot C_3\cdot n^{k+2}$, where C_3 is the constant defined in Lemma 11. Letting $C_4=3C_3$ we have the result.

Lemma 13. $L(G)$ contains a word (in fact z_h) whose length is not a power of 2.

Proof.

From the proof of Lemma 12 it follows that $|z_h|<|x(n)|=2^n$ and from Lemma 13 it follows that

$$|z_h|\geq|x(n)|-C_4\cdot n^{k+3}.$$

Choosing n large enough we combine these two observations to give

$$2^{n-1}<|z_h|<2^n.$$

III. THE MAIN RESULT.

Theorem 1. There exists a 0L language which cannot be generated by a EDT0L system.

Proof.

Let L be the language defined in the last section.

As the 0L system <{0,1},P,0,{0,1}>, where P consists of the following productions:

$0\to 00, 0\to 01, 0\to 10, 0\to 11,$

$1\to 00, 1\to 01, 1\to 10, 1\to 11,$

generates L, the language L is a 0L language.

But from lemmata 2 through 13 it follows that L is not an EDT0L language, and so Theorem 1 holds.

Corollary 1. $F(\text{EDT0L}) \subsetneq F(\text{ET0L})$.

Proof.

This follows directly from Theorem 1 and from the fact that by definitions 1, 3 and 4 $F(\text{0L}) \subseteq F(\text{ET0L})$ and $F(\text{EDT0L}) \subseteq F(\text{ET0L})$.

IV. DISCUSSION

The reader may easily observe that the proof of Theorem 1 can be done in almost the same way if one chooses instead of L another "similar" language, say

$$K=\{x\varepsilon\{0,1\}^*:|x|=3^n \text{ for some } n\geq 0\}.$$

In fact we can generalize the main result as follows. Let Σ be an alphabet and $x\varepsilon\Sigma^+$. Let us define $\mu(x)$ as the minimal positive integer n such that any two non-overlapping subwords of x are different.

Definition 6. Let M be a language. M is called exponential if there exists a positive integer C larger than 1 such that for every x_1, x_2 in M, if $|x_1|>|x_2|$ then $|x_1|\geq C|x_2|$.

In much the same way as we have proved Theorem 1 one can prove the following result.

Theorem 2. If M is an exponential EDT0L language, then there exists a positive integer constant C_M such that for every x in $M-\{\Lambda\}$ we have $\frac{|x|}{\mu(x)} < C_M$.

Remark. We have presented a detailed proof of Theorem 1 rather than the proof of Theorem 2 since this gives a better illustration of the proof technique.

REFERENCES

Ehrenfeucht, A. & Rozenberg, G. (1973)

Subwords in deterministic T0L languages, submitted for a publication, also available as a technical report of the Computer Science Dept., University of Colorado at Boulder.

Herman, G.T. & Rozenberg, G. (1973)

Developmental systems and languages, to be published by North-Holland Publishing Company.

Herman, G.T. & Lindenmayer, A. & Rozenberg, G. (1973)

Description of developmental systems using recurrence systems, submitted for publication, also available as a technical report of the Computer Science Dept., SUNY at Buffalo.

Hopcroft, J.E. & Ullman, J.D. (1969)

Formal languages and their relation to automata, Addison-Wesley, Reading, Mass.

Lindenmayer, A. (1968)

Mathematical models for cellular interactions in development, Parts I & II, J. Theoret. Biol., 18, 1968, 280-315.

Lindenmayer, A. & Rozenberg, G. (1972)

Developmental system and languages, in Proc. 4 ACM symp. Theory of Comp., 214-221.

Rozenberg, G. (1972)

T0L systems and languages, Information & Control, to appear.

Rozenberg, G. (1973)

Extension of tabled 0L systems and languages, submitted for publication, available as a technical report of Computer Science Dept., SUNY at Buffalo.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. D. Wood and Messrs. K. P. Lee and A. Walker for corrections to the original manuscript.