On Θ-determined EOL Languages

by

A. Ehrenfeucht
and
G. Rozenberg[+]

Working paper

[+] Department of Mathematics, Utrecht University, Utrecht, Holland and

Department of Mathematics, Antwerp University UIA, Wilrijk, Belgium.

ABSTRACT

A new subclass of EOL languages is introduced. Several
properties of languages in this class are proved and then applied to
provide elegant proofs that some languages are not EOL languages.

INTRODUCTION

The mathematical theory of L-systems constitutes today a vigorously investigated fragment of formal language theory (see, e.g., [Herman and Rozenberg, 1975], [Rozenberg and Salomaa, 1974] and [Rozenberg and Salomaa, 1975]).

One of the more important research areas in the theory of L systems is to provide results characterizing the structure of languages from the given family of L languages (as opposed to, e.g., results characterizing closure or decidability properties of various families of L languages). Such results are very much needed when proving that certain languages do not belong to certain families of L languages.

This paper continues research in this direction. We introduce the subclass of the class of EOL languages (the so called Θ-deterministic EOL languages) and then we prove that if a language belongs to this subclass then this fact bears quite strong consequences as far as properties of this language are concerned. Then we provide applications of our results for a rather difficult in general task of proving that certain languages are not EOL languages. (For example we can easily prove that $\{a^n b^m a^n : m \geq n \geq 1\}$ as well as $\{www : w \in \{a,b\}^+\}$ are not EOL languages).

We shall use the usual formal language theoretic terminology and notation. Perhaps only the following points require an explanation:

1) For a word x, $|x|$ denotes its length. If b is a letter then $\#_b x$ denotes the number of occurrences of b in x and if B is a set of letters then $\#_B x$ denotes $\sum_{b \in B} \#_b x$.

2) Let Σ be a finite alphabet and let $B \subseteq \Sigma$. Then $h_{B,\Sigma}$ denotes the homomorphism from Σ into Σ^* defined by

$$h_{B,\Sigma}(a) = \begin{cases} a & \text{if } a \varepsilon B \\ \Lambda & \text{if } a \varepsilon \Sigma - B. \end{cases}$$

Whenever the Σ is clear from context we write simply h_B rather than $h_{B,\Sigma}$. Also, in order not to burden the notation too much, we shall use the same symbol in denoting the extension of the function h_B to languages (thus we have $h_B(L) = \bigcup_{x \varepsilon L} h_B(x)$).

3) For a word x, Min(x) denotes the set of letters that occur in x. For a language L, MIN(L) = {Min(x) : $x \varepsilon L$}.

I.   EOL SYSTEMS AND LANGUAGES

In this section we recall the notions of an E(P)OL system and of an E(P)OL language. Then we prove a normal form theorem for EPOL systems which will turn out to be useful in our further considerations.

Definition 1. An EOL system is a construct $G = <\Sigma, P, \omega, \Delta>$ where

Σ is a finite nonempty alphabet, referred to as the alphabet of G,

Δ is a subset of Σ, its elements are referred to as terminals of G, whereas elements of $\Sigma-\Delta$ are referred to as nonterminals of G,

ω is an element of Σ^*, referred to as the axiom of G,

P   is a finite nonempty relation, $P \subseteq \Sigma x \Sigma^*$, its elements are referred as productions of G.

It is required that $(\forall a)_\Sigma (\exists \alpha)_{\Sigma^*} [<a,\alpha> \epsilon P]$.

Remark

1)   If in the definition above $P \subseteq \Sigma x \Sigma^+$, then we call G a propagating EOL system and abbreviate it as an EPOL system.

2)   Productions $<a,\alpha>$ from P are usually written in the form $a \to \alpha$, and we write $a \to \alpha$ as an abbreviation for "$a \to \alpha$ is in P".
$\quad\quad\quad\quad\quad P$

3)   If $\Delta=\Sigma$ and for every a in Σ there exists exactly one α in Σ^* such that $<a,\alpha> \epsilon P$ then G is called a DOL system. If additionally G is propagating then G is called a PDOL system.

<u>Definition 2</u>. Let $G = \langle\Sigma, P, \omega, \Delta\rangle$ be an EOL system.

1)   Let $x, y \in \Sigma^*$, $x = a_1 \ldots a_n$ with $a_1, \ldots, a_n$ in Σ.

1.1) We say that x <u>directly derives</u> y <u>in</u> G, denoted as $x \underset{G}{\Longrightarrow} y$, if

there exist $\alpha_1, \ldots, \alpha_n$ in Σ^* such that $a_1 \underset{P}{\to} \alpha_1, \ldots, a_n \underset{P}{\to} \alpha_n$

and $y = \alpha_1 \ldots \alpha_n$.

1.2) For a positive integer k, we say that x <u>derives</u> y <u>in</u> G <u>in</u> k

<u>steps</u>, denoted as $x \underset{G}{\overset{k}{\Longrightarrow}} y$, if there exist $x_1, \ldots, x_k$ in Σ^*

such that $x \underset{G}{\Longrightarrow} x_1 \underset{G}{\Longrightarrow} x_2 \ldots \underset{G}{\Longrightarrow} x_k = m$. If there exists a k

such that $x \underset{G}{\overset{k}{\Longrightarrow}} y$ then we write $x \underset{G}{\overset{+}{\Longrightarrow}} y$.

1.3) $\overset{*}{\underset{G}{\Longrightarrow}}$ denoted the transitive and the reflexive closure of

the relation $\underset{G}{\Longrightarrow}$. If $x \overset{*}{\underset{G}{\Longrightarrow}} y$ then we say that x <u>derives</u> y <u>in</u>

G.

2)   The <u>set of sentential forms of</u> G, denoted by SENT(G), is

defined by SENT(G) = $\{x \epsilon \Sigma^* : \omega \overset{*}{\underset{G}{\Longrightarrow}} x\}$.

3)   The language of G, denoted by L(G), is defined by

$L(G) = \{x \epsilon \Delta^* : \omega \overset{*}{\underset{G}{\Longrightarrow}} x\}$.

<u>Remark</u>

1)   For x in Σ^* and k in N^+, we write L(G,x,k) to denote the set

$\{y \epsilon \Sigma^* : x \overset{k}{\underset{G}{\Longrightarrow}} y\}$.

2)   If $x \overset{+}{\underset{G}{\Longrightarrow}} y$ then there is a <u>derivation D of</u> y <u>from</u> x <u>in</u> G meaning

the sequence of words x_0, x_1, ..., x_n such that $x_0 = x$, $x_n = y$ and

$x_i \underset{G}{\Longrightarrow} x_{i+1}$, for $0 \leq i \leq n-1$, together with a precise description of

how a single derivation step $x_i \underset{G}{\Longrightarrow} x_{i+1}$ is done. However, very

often the sequence $x_0, x_1, ..., x_n$ called the <u>trace of</u> D and

denoted by Trace (D) provides sufficient information about D

and we will use it in the sequel.

<u>Definition 3</u>. A language K is called an EOL <u>language</u> (EPOL

<u>language</u>) if there exists an EOL system (EPOL system) G such that L(G) = K.

It is well-known (see, e.g., [Herman and Rozenberg, 1975] p. 184)

that for every EOL language K there exists an EPOL system G such that
L(G) = K-{ Λ }. For this reason in this paper we will restrict ourselves
to EPOL systems.

Also, because of the type of problems we are concerned with
in this paper we consider, unless stated otherwise, only infinite EOL
languages and EOL systems which generate infinite languages.

Let us recall (see, e.g., [Herman and Rozenberg, 1975]) that
an EPOL system G = <Σ, P, ω, Δ>is called synchronized if for every a in
Δ and every α in Σ*, if $a \xRightarrow[G]{+} \alpha$ then α is not in Δ*. It is well known,
see, e.g., Theorem 4.4 in [Herman and Rozenberg, 1975] that for every
EPOL system there exists (effectively) an equivalent synchronized
EPOL system. From the proof of this theorem in [Herman and Rozenberg,
1975] it is clear that for every synchronized EPOL system G = <Σ, P, ω, Δ>
we can assume the existence of a unique symbol F such that if a∈Δ then
a → F is the only production for a in P and F → F is the only production
for F in P. As a matter of fact we reserve the symbol F to denote in
the sequel this unique "synchronization symbol".

If G = <Σ, P, ω,Δ> is a synchronized EPOL system such that
ω∈Σ-Δ then we will use W(G) to denote Σ- (Δ∪{F,ω}).

We define now a stronger version of a synchronized EPOL system.

Definition 4. An EPOL system G = <Σ, P, ω,Δ> is called neatly
synchronized if the following conditions hold:
1)   ω∈(Σ-Δ) and ω does not appear at the right-hand side of any
        production in P.

2)   G is synchronized.

3) $(\forall a)_{W(G)-\{\omega\}} \quad (\exists x)_{\Delta^+} \quad [a \xRightarrow[G]{+} x]$.

4) $(\forall a)_{\Sigma-\{\omega\}} \ (\forall k,k')_{N^+}[MIN(L(G,a,k)) = MIN(L(G,a,k'))].$

5) $(\forall a)_{\Sigma} \ [\underline{if} \ a \xrightarrow{P} \alpha \ \underline{then} \ \alpha\varepsilon(W(G))^+ \ \underline{or} \ \alpha\varepsilon\Delta^+ \ \underline{or} \ \alpha=F].$

We will prove now that each EPOL language can be generated by a neatly synchronized EPOL system.

<u>Theorem 1.</u>   There is an algorithm which for every EPOL system G produces a neatly synchronized EPOL system $\bar{G}$ such that $L(G) = L(\bar{G})$.

<u>Proof.</u>

Let $G = \langle\Sigma, P, \omega,\Delta\rangle$ be an EPOL system.

It is well known (see, e.g., Theorem 4.4 and its proof in [Herman and Rozenberg, 1975]) that one can effectively produce an EPOL system G^1 satisfying conditions 1) and 2) of Definition 4. Now if a symbol a in $W(G^1) - \{\omega\}$ does not satisfy condition 3) from Definition 4 then, clearly, it can be removed from the alphabet of G^1 (together with all productions in G^1 involving this symbol) and it will not change the language of G^1.

Now, by Corollary 1 and its proof in [Rozenberg, 1975] one can use G^1 to produce G^2 which will be such that $L(G^2) = L(G^1)$ and G^2 will satisfy conditions 1) through 4).

Finally it is clear that if $G^2 = \langle\bar{\Sigma}, \bar{P}, \bar{\omega}, \Delta\rangle$ contains a production of the form $a \to \alpha$ where $\alpha\notin(W(G^2))^+\cup\Delta^+\cup\{F\}$ then one can remove it without changing the language of the system.

Hence the theorem holds.

<u>Remark</u>

In the sequel we shall always deal with EPOL systems in which

the axiom is a nonterminal which do not occur at the right-hand side of any production. We reserve the symbol S to denote the axiom of an EPOL system.

Following [Rozenberg, 1975] we recall now the notion of a decomposition of an EPOL system.

Definition 5. Let $G = <\Sigma, P, S, \Delta>$ be an EPOL system and let k be a positive integer. A k-decomposition of G, denoted as Dec (G,k), is the EPOL system Dec $(G,k) = <\Sigma, P_k, S, \Delta>$ where P_k is defined as follows:

1)   For a in $\Sigma-\{S\}$, $a \to \alpha$ iff $a \xRightarrow[P]{k} \alpha$,
        P_k

2) $S \to \alpha$ iff $S \xRightarrow[P]{m} \alpha$ for some m in $\{1,\ldots,k\}$.
      P_k

It should be clear to the reader that for every EPOL system G:

1)   For every k in N^+, $L(\text{Dec }(G,k)) = L(G)$.

2)   If G is neatly synchronized, then, for every k in N^+, $L(\text{Dec }(G,k))$ satisfies conditions 1) through 4) of Definition 4, but it does not have to satisfy the condition 5) of this definition. However it suffices to remove productions which do not satisfy this condition to obtain a neatly synchronized system defining the same language. We shall use $\widetilde{\text{Dec}}$ (G,k) to denote such a modified system.

II. Θ-DETERMINED EPOL SYSTEMS

In this section we introduce a new subclass of the class of EPOL systems. Then we prove two results on the structure of derivations in these systems.

Definition 6. Let $G = <\Sigma, P, S, \Delta>$ be a neatly synchronized EPOL system and let Θ be a nonempty subset of Δ.

1) For a in $W(G)$, we say that a is Θ-<u>determined</u> (<u>in</u> G) if
$(\forall k)_{N^+} [\#h_\Theta ((L(G,a,k) \cap \Delta^*) = 1]$; otherwise we say that a is Θ-<u>undetermined</u>.

2) We say that G is Θ-<u>determined</u> if every a in $W(G)$ is Θ-<u>determined</u>.

Remark.

We will use $\underline{length}_\Theta(G,a,k)$ to denote the length of a unique word in $h_\Theta(L(G,a,k) \cap \Delta^*)$

We leave to the reader the easy proof of the following.

Lemma 1. If G is Θ-determined EPOL system, then, for every k in N^+, $\widetilde{Dec}(G,k)$ is Θ-determined.

We are ready to prove a result on the possibility of a very suitable slicing of a Θ-determined EPOL system.

Lemma 2. Let G be a Θ-determined EPOL system. There exists $\ell \geq 1$ such that $\widetilde{Dec}(G,\ell)$ satisfies the following:

$$(\forall a)_{\widetilde{W(\underline{Dec}\ (G,\ell))}}\ [\ ((\exists c_a)_{N^+}(\forall k)_{\geq 1}\ [\underline{length}_\Theta\ \widetilde{(\underline{Dec}\ (G,\ \ell)},a,k) < c_a])$$

$$\underline{or}\ ((\forall k)_{N^+}[\underline{length}_\Theta\ \widetilde{(\underline{Dec}\ (G,\ell)},a,k) > k])].$$

Proof.

Let $G = \langle \Sigma, P, S, \Delta \rangle$.

Let $RED(G) = \{a \in W(G) : (\exists \alpha)_{(\Delta-\Theta)^+}[a \underset{P}{\to} \alpha]\}$ and let

$\bar{G} = \langle \bar{\Sigma}, \bar{P}, S, \Delta \rangle$ where $\bar{\Sigma} = \Sigma - RED(G)$ and

$\bar{P} = \{a \to \bar{\alpha} : (a \in \bar{\Sigma})\ \underline{and}\ ((\exists \alpha)_{\Sigma^+}[\bar{\alpha} = h_{\bar{\Sigma}}(\alpha)\ \underline{and}\ a \underset{P}{\to} \alpha]\}$.

Since G is Θ-determined, if $a \in RED(G)$, then there is no β in $\Delta^* \Theta \Delta^*$

such that $a \underset{G}{\to} \beta$. Thus (since G is neatly synchronized):

1) if a is in $RED(G)$, then for every k in N^+, $L(G,a,k)$ contains no

   words in $\Delta^* \Theta \Delta^*$; consequently, for every k in N^+, $\underline{length}_\Theta (G,a,k)=0$,

2) if a is in $\bar{\Sigma}$, then for every k in N^+,

   $h_\Theta (L(G,a,k) \cap \Delta^*) = h_\Theta (L(\bar{G},a,k) \cap \Delta^*)$.

Now let $\bar{\bar{G}} = \langle \bar{\Sigma}, \bar{\bar{P}}, S, \Delta \rangle$, where $\bar{\bar{P}}$ is defined as follows:

⁻ for a in $\bar{\Sigma} - W(\bar{G})$, productions in $\bar{\bar{P}}$ are precisely these from $\bar{P}$,

⁻ for each a in $W(\bar{G})$ we choose one arbitrary production from $\bar{P}$ of the

form $a \to \alpha$ with α in $(W(\bar{G}))^+$ (we denote the right-hand side of this

production by $\underline{nont}(a)$) and one arbitrary production from $\bar{P}$ of the form

$a \to \beta$ with β in Δ^+ (we denote the right-hand side of this production

by $\underline{term}(a)$). Now the productions for a in $\bar{\bar{P}}$ are $a \to \underline{term}(a)$ and

$a \to \underline{nont}(a)$.

Obviously we have

3) $\quad (\forall a)_{W(\bar{G})} \, (\forall k)_{\geq 1} \quad [h_\Theta(L(\bar{G},a,k) \cap \Delta^*) = h_\Theta(L(\bar{\bar{G}},a,k) \cap \Delta^*)].$

Now we are in position to prove the following

Claim 2.1 There exist a finite set U, a finite number of PDOL systems $H_1,\ldots,H_f$, $f \geq 1$, and a Λ-free homomorphism ϕ such that

$$L(\bar{\bar{G}}) = U \cup \bigcup_{i=1}^{f} \phi \,(L(H_i)).$$

Proof of Claim 2.1

Let U be the set of all terminal words that can be derived from S in $\bar{\bar{G}}$ in one step. Let Y be the set of all words over $W(\bar{\bar{G}})$ which can be derived from S in $\bar{\bar{G}}$ in one step, say $Y = \{w_1,\ldots,w_f\}$.

For $1 \leq i \leq f$, let H_i be the PDOL system defined by $H_i = \langle V,R,w_i,V \rangle$ where $V = W(\bar{\bar{G}})$ and $R = \{a \to \underline{\text{nont}}\,(a):a\varepsilon V\}$.

Finally let ϕ be the Λ-free homomorphism from V into Δ^+ defined by $\phi(a) = \underline{\text{term}}\,(a)$.

The reader can easily see that, indeed,
$$L(\bar{\bar{G}}) = U \cup \bigcup_{i=1}^{f} \phi(L(H_i))$$
and so the claim holds.

Let us recall now the following result from the theory of DOL sequences (see, e.g., [Lee and Rozenberg, 1974] or [Nielsen, 1974]):

- if H is a DOL system, then there exists a constant C such that

   if a word w is derived in $\underline{\text{Dec}}$ (H,C) in k steps then $|w| > k$.

Clearly this result together with 1) and Claim 2.1 proves the lemma.

Now we will show how the number of occurrences of elements from Θ in words generated by a Θ-determined EPOL system are bounding the total length of these strings.

Lemma 3. Let G be a Θ-determined EPOL system. There exist positive integer constants C, D such that, for every x in L(G), if $\#_\Theta(x) > C$, then $|x| < D^{\#_\Theta(x)}$.

Proof.

Let $G = \langle\Sigma, P, S, \Delta\rangle$. By Lemma 2 we can assume that

$$(\forall a)_{W(G)} [((\exists C_a)_{N^+}(\forall k)_{\geq 1} \underline{[\text{length}}_\Theta(G,a,k) < C_a])$$

$$\underline{\text{or}} \quad (\forall k)_{\geq 1} \underline{[\text{length}}_\Theta(G,a,k) < C_a])].$$

Let $BOUND(G) = \{a\varepsilon W(G): (\exists C_a)_{N^+}(\forall k)_{\geq 1} \underline{[\text{length}}_\Theta(G,a,k) < C_a]\}$

and let for every a in BOUND(G), $\bar{C}_a$ be the smallest positive integer constant m satisfying the statement $(\forall k)_{\geq 1} \underline{[\text{length}}_\Theta(G,a,k) < m]$.

Let $ONE(G) = \{x\varepsilon SENT(G): S \underset{G}{\Longrightarrow} x\}$ and $g_0 = \max\{\#_\Theta(x) : x\varepsilon ONE(G)\cap L(G)\}$.

Let $\overline{ONE}(G) = ONE(G) \cap (BOUND(G))^+$ and let

$g_1 = (\max\{|y| : y\varepsilon\overline{ONE}(G)) \cdot (\max\{\bar{C}_a : a\varepsilon BOUND(G)\})$.

Let $C = \max\{g_0, g_1\} + 1$.

Let $D = \max\{|\alpha| : \alpha$ is the right-hand side of a production in P$\}$.

$$\cdot \max\{|y| : y\varepsilon ONE(G)\}.$$

Now let us assume that $x\varepsilon L(G)$.

(i)   If $x\varepsilon ONE(G)$, then $\#_\Theta(x) \leq g_0 < C$, and so the lemma trivially holds.

(ii)   If $x\notin ONE(G)$, then let $S, y_1, \ldots, y_\ell = x$, with $\ell \geq 2$, be the trace of a derivation in G.

(ii.1) If $y_1 \in \overline{ONE}(G)$, then $\#_\Theta(x) \leq g_1 < C_\Theta$ and so the lemma trivially holds.

(ii.2) If $y_1 \notin \overline{ONE}(G)$ then it contains an occurrence of an element b from $W(G)$ such that $(\forall k)_{\geq 1} [\underline{length}_\Theta (G,k,b) > k]$. Thus $\#_\Theta(x) \geq \ell-2$. But by the definition of D, $|x| < D^{\ell-2}$ and so

$$|x| < D^{\#_\Theta(x)} \; .$$

This ends the proof of the lemma.

III.   Θ-DETERMINED EOL LANGUAGES

In this section we introduce the notion of a Θ-determined language. We show that the class of   Θ-determined EOL languages is precisely the class of languages generated by Θ-determined EPOL systems. Then we show how, using the results of the last section, one can easily prove that some particular languages are not EOL languages.

Definition 7. Let K be a language over an alphabet Σ and let Θ be a nonempty subset of Σ. We say that K is a Θ-determined language if

$$(\forall k)_{N^+}(\exists n_k)_{N^+}(\forall x,y)_K$$

$$[\underline{if}\ |x|,\ |y| > n_k\ \underline{and}\ x = x_1 u x_2\ \underline{and}\ y = x_1 v x_2\ \underline{and}\ |u|,|v| < k$$

$$\text{then } h_\Theta(u) = h_\Theta(v)]$$

Example

1)   Take $K = \{a^n b^m a^n : m \geq n \geq 1\}$ and let   Θ = {a}. Then K is Θ-determined.

2)   Take $K = \{a^n b^m a^n : n,m \geq 1\}$ and let   Θ = {a}. Then K is not Θ-determined.

Lemma 4. Let K be a   Θ-determined EOL language. Then there exists a   Θ-determined EPOL system G such that L(G) = K.

Proof.

Let $H = \langle \Sigma, P, S, \Delta \rangle$ be an EPOL system such that $L(H) = K$. According to Theorem 1 we can assume that H is neatly synchronized.

Let $a \in W(H)$. We call a slim in H if there exists a positive integer p_a such that if a occurs in a word w over $W(H)$ which is in $SENT(H)$ then $|w| < p_a$.

First we prove the following.

Claim 4.1 Let $a \in W(H)$. If a is Θ-undetermined, then it is slim.

Proof of Claim 4.1

Since a is Θ-undetermined, there exist d in N^+ and x_1, x_2 in Δ^+ such that $a \overset{d}{\underset{H}{\Longrightarrow}} x_1$, $a \overset{d}{\underset{H}{\Longrightarrow}} x_2$, and $h_\Theta(x_1) \neq h_\Theta(x_2)$. If we assume now that a is not slim, then for every t in N^+ there exists a word $z_1 a z_2$ in $SENT(H)$ with $|z_1 a z_2| > t$. Consequently there exists a positive integer constant p (take $p = \max\{|x_1|, |x_2|\}$) such that, for every t in $N^+, L(H)$ contains words of the form $w_1 x_1 w_2$ and $w_1 x_2 w_2$ where

$|w_1 x_1 w_2| > t$, $|w_1 x_2 w_2| > t$, $|x_1| \leq p$, $|x_2| \leq p$ but $h_\Theta(x_1) \neq h_\Theta(x_2)$. Since $L(H)$ is Θ-deterministic, this is a contradiction.

Thus Claim 4.1 holds.

For w in $L(H)$ let $D_{H,w}$ denote a fixed derivation of w in H which is such that no other derivation of w in H takes less steps than $D_{H,w}$.

Claim 4.2 There exists a constant B_H in N^+ such that for every w in $L(H)$, if Trace $(D_{H,w}) = S, y_1, \ldots, y_m = w$ and y_i contains an occurrence of a Θ-undetermined letter, then $i < B_H$.

Proof of Claim 4.2

This follows from Claim 4.1 and from the fact that in Trace $(D_{H,w})$ the number of words of the same length, say ℓ, is limited by $(\#\Sigma)^{\ell}$.

Now let k be a positive integer such that $k > B_H$ where B_H is the constant from Claim 4.2. Let G_k be obtained from $\widetilde{\underline{Dec}}$ (H,k) in such a way that from the productions (in $\widetilde{\underline{Dec}}$ (H,k)) for every symbol with the exception of S one removes these which contain occurrences of Θ-undetermined letters on their right-hand side.

Obviously $L(G_k) \subseteq L(\widetilde{\underline{Dec}}$ (H,k)) = L(H)

To see that $L(H) = L(\widetilde{\underline{Dec}}$ (H,k)) $\subseteq L(G_k)$ we proceed as follows.

Let $x \in L(H)$.

If x can be derived in H in no more than k steps, then x is derived in G_k in one step.

If every derivation of x in H takes more than k steps, then let us consider $D_{H,x}$. Let Trace $(D_{H,x}) = S, y_1, \ldots, y_k, y_{k+1}, \ldots, y_m = x$. Note that, for $1 \leq i \leq k$, $S \underset{G_k}{\Longrightarrow} y_i$, and (see Claim 4.2), for

$k + 1 \leq i \leq m - 1$, y_i does not contain Θ-undetermined letters. Thus there exists an integer u, $1 \leq u \leq k$, such that

$S, y_u, y_{u+k}, \ldots, y_{m-2k}, y_{m-k}, y_m = x$

is the trace of a derivation in G_k. Consequently $x \in L(G_k)$.

But, see Lemma 1, G_k is Θ-determined and so the lemma holds.

Thus now we can carry over into Θ-determined EOL languages the result (Lemma 3) on the bound on the length of words in Θ-determined EPOL systems.

Theorem 2. Let K be a Θ-determined EOL language. There exist positive integer constants C and D such that, for every x in K, if $\#_\Theta(x) > C$, then $|x| < D^{\#_\Theta(x)}$.

Proof.

This result follows directly from Lemma 3, Lemma 4 and the quoted before fact that EPOL systems generate all EOL languages (modulo Λ).

We will show now that in the case that an EOL language K over Δ is Δ-determined it can be decomposed into a finite union of Λ-free homomorphic images of PDOL languages.

Theorem 3. Let K be an EPOL language over an alphabet Δ. If K is Δ-determined then there exists a finite set of PDOL systems $H_0, H_1, \ldots, H_f$ and a Λ-free homomorphism ϕ such that $K = \bigcup_{i=0}^{f} \phi(L(H_i))$

Proof.

This follows from:

1) Lemma 3 and Lemma 4,

2) the proof of Lemma 2, see Claim 2.1, where one easily sees that if $\Theta = \Delta$, then $L(\overline{G}) = L(G) = U \cup \bigcup_{i=1}^{f} \phi(L(H_i))$, and

3) the simple observation that a finite set is an image under a Λ-free homomorphism of a PDOL language.

Based on Theorem 3 we can get a very useful result on the number of subwords in Δ-determined EOL languages. (In what follows for a language L and a nonnegative integer $n, \pi_n(L)$ denotes the number of subwords of length n which occur in words of L).

Corollary 1. Let K be an EQL language over an alphabet Δ. If K is Δ-determined, then there exists a constant C such that, for every n in N, $\pi_n(K) \leq C \cdot n^3$.

Proof.

First, we recall the following result from [Ehrenfeucht, Lee and Rozenberg, 1975]:

- If L is a PDOL language then there exists a constant D such that, for every n in N, $\pi_n(L) \leq Dn^2$.

By Theorem 1, $K = \bigcup_{i=0}^{f} \phi(L(H_i))$ where ϕ is a Λ-free homomorphism and H_i's are PDOL systems. Hence it suffices to show that if L_1, L_2 are languages and g is a Λ-free homomorphism such that $L_2 = g(L_1)$ then if, for every n in N, $\pi_n(L_1) \leq C_1 \cdot n^2$ for a constant C_1, then, for every n in N, $\pi_n(L_2) \leq C_2 \cdot n^3$ for some constant C_2. This is done as follows.

If $w \epsilon \pi_n(L_2)$ then there exists z in $\pi_\ell(L_1)$ for some $\ell \leq n$ such that w is a subword of $h(z)$. Let $z = a_1 \ldots a_\ell$ with $a_1, \ldots, a_\ell$ being single letters. It suffices to consider the following situation:
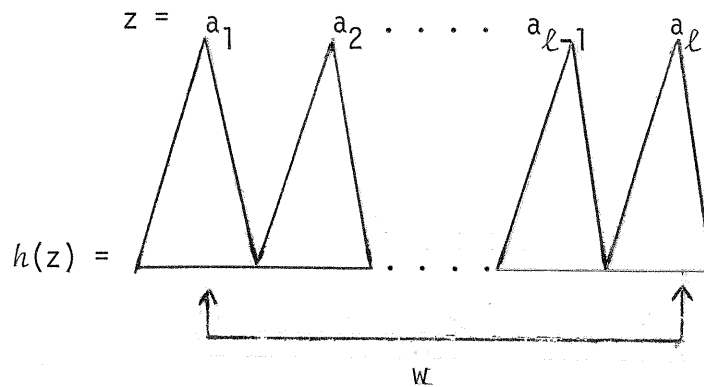
Hence, for every n in N,

$$\pi_n(L_2) \leq (\#\Sigma)^{C_0} \cdot \sum_{m=1}^{n} \pi_n(L_1) \ , \quad \text{where } L_1 \text{ is over the alphabet } \Sigma \text{ and}$$

and C_0 equal twice the maximum length of $h(a)$ with a in Σ.
Thus

$$\pi_n(L_2) \leq C_1 \cdot (\#\Sigma)^{C_0} \cdot \sum_{m=1}^{n} m^2 \leq C_2 \cdot m^3, \text{ where } C_2 = C_1 \cdot (\#\Sigma)^{C_0}$$

Thus the Corollary holds.

We end this section by demonstrating how our results can be used to get elegant proofs that some languages are not EOL languages.

Application 1. Let $K = \{a^n b^m a^n : m \geq n \geq 1\}$. Clearly K is $\{a\}$-determined but it is not true that the number of a's in a word from K bounds the length of the string. Thus by Theorem 2, K is not an EOL language.

Application 2. Let $\ell \geq 3$ and let $K_\ell = \{w^\ell : w \in \{a,b\}^+\}$. Clearly K is $\{a,b\}$-determined but at the same time, for every n in N, $\pi_n(K) = 2^n$. Thus by Corollary 1, K_ℓ is not an EOL language.

REFERENCES

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

[Lee and Rozenberg, 1974], K. P. Lee and G. Rozenberg, The length sets of DOL languages are uniformly bounded, Information Processing Letters, 2, 185-188, 1974.

[Nielsen, 1974], M. Nielsen, On the decidability of some equivalue problems for DOL systems, Information and Control, 25, 166-193, 1974.

[Rozenberg, 1975], G. Rozenberg, On slicing of K-iteration grammars, Technical Report of the Department of Mathematics, Utrecht University, LOCOS 29, 1975.

[Rozenberg and Salomaa, 1974], G. Rozenberg and A. Salomaa, L Systems, Lecture Notes in Computer Science No. 15, Springer-Verlag, 1974.

[Rozenberg and Salomaa, 1975], G. Rozenberg and A. Salomaa, The mathematical theory of L systems, to appear in Progress in Information Processing, ed. I. Tou.