all correspondence to the second author

EOL LANGUAGES ARE NOT CODINGS OF FPOL LANGUAGES

by

A. Ehrenfeucht

and

G. Rozenberg †

#CU-CS-098-76 September, 1976

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

working report

ABSTRACT

One of the useful results concerning EOL languages states that a language is an EOL language if and only if it is a coding of a OL language. In this paper we refine this result by demonstrating that there exist EOL languages that are not codings of languages that are generated by propogating OL systems with finite axiom sets. This solves Problem 10 from the "L SYSTEMS PROBLEM BOOK '75" (see [4]).

I. INTRODUCTION

One of the useful results about EOL languages says that a language is an EOL language if and only if it is a coding of a OL language (see [1]). The proof of this result from [1] (see also [6]) essentialy requires that the "underlying" OL system contains erasing productions. As it is much easier to deal with the structure of derivations in OL systems that *do not* use erasing, the open question for the last few years was: can one get every EOL language as a coding of a language generated by a *propogating* OL system with finite axiom set? (See Problem 10 in the "L SYSTEMS PROBLEM BOOK '75 from [4]).

We settle this question in the negative. Our solution is effective in the sense that we provide a result characterizing a subclass of CFPOL languages which allows one to construct examples of EOL languages that are not CFPOL languages.

II. PRELIMINARIES

We assume the reader to be familiar with the rudiments of L systems theory (see, e.g., [3] or [6]) and with the basics of formal language theory (see, e.g., [7]).

The basic type of L systems considered here are propagating OL systems with finite axiom sets (abbreviated FPOL systems). For such a system we use the notation $G = \langle \Sigma, P, Ax \rangle$ where Σ is the alphabet of G, P its set of productions and Ax its axiom set. We will write $a \xrightarrow[P]{} \alpha$ for "$a \to \alpha$ is in P". $L^{(n)}(G)$ denotes the set of all words derivable in n steps in G. A letter a in Σ is called (*strictly*) *recursive* in G if $a \overset{+}{\Longrightarrow} \alpha a \beta$ for some $\alpha, \beta \in \Sigma^*$ (such that $\alpha\beta \neq \Lambda$). Unless clear otherwise we consider only *reduced* FPOL systems, it is such FPOL systems G in which

every letter of Σ occurs in a word of L(G). For a positive integer m, the m-*slicing of* G, is an FPOL system $G^{(m)} = \langle \Sigma^{(m)}, P^{(m)}, Ax^{(m)} \rangle$, where $\Sigma^{(m)} = \Sigma$, $P^{(m)} = \{a \rightarrow \alpha : a \overset{m}{\underset{G}{\Longrightarrow}} \alpha\}$, and $Ax^{(m)} = \{x : x \in \bigcup_{i=0}^{m-1} L^{(i)}(G)\}$. It is obvious that $L(G^{(m)}) = L(G)$. For a letter a, $Acc_G(a)$ denotes the set of all letters accessible from a in G.

If x is a word than $|x|$ denotes the length of x, *min* x denotes the set of letters that occur in x, $pref_n x$ is the prefix of x of length n and $suf_n x$ denotes the suffix of x of length n. If Σ is an alphabet then $\phi_\Sigma x$ denotes the word resulting from x by erasing from it all letters that are not in Σ. *To avoid very cumbersome wordings we will often talk about "a letter in a word" when we really mean "an occurrence of a letter in a word", this however should not lead to a confusion.*

For a language K, *Length* (K) denotes the length set of K, $Pref_n K = \{pref_n x : x \in K\}$, $Suf_n K = \{suf_n x : x \in K\}$ and $\phi_\Sigma K = \{\phi_\Sigma x : x \in K\}$.

We will use N, N^+, R, R^+ to denote the sets of nonnegative integers, positive integers, nonnegative reals and positive reals respectively. For an ultimately periodic set Z we use *thres* Z to denote its smallest threshold, and for this threshold we use *per* Z to denote its smallest period.

We end this section by proving a result that will be very useful in the sequel.

Definition 1. Let G be an FPOL system and let K be a language. Then the *existential spectrum of* G *with respect to* K, denoted as *Espec* (G,K), is defined by $Espec\ (G,K) = \{n \in N : (\exists w)_{L^{(n)}(G)} [w \in K]\}$.

The following result is from [2].

Theorem 1. If G is an FPOL system and K is a regular language, than *Espec* (G,K) is ultimately periodic.

We introduce now a subclass of FPOL systems that is (mathematically) quite pleasant to deal with.

Definition 2. Let $G = \langle \Sigma, P, Ax \rangle$ be an FPOL system. We say that G is *impatient* if

$(\forall a,b)_{\Sigma}(\forall r,s)_{N^+}$[b is accessible from a in r steps if and only if b is accessible from a in s steps].

Let us recall that a coding is a letter-to-letter homomorphism, and a coding of an FPOL language is referred to as an CFPOL language.

Lemma 1. A language K is a coding of an FPOL language if and only if it is a coding of an FPOL language that can be generated by an impatient FPOL system.

Proof.

Clearly it suffices to show that if K is a coding of an FPOL language then it is a coding of a language generated by an impatient FPOL system.

Thus let ρ be a coding and $G = \langle \Sigma, P, Ax \rangle$ be an FPOL system such that $K = \rho(L(G))$. Let for every a in Σ, $G_a = \langle \Sigma, P, a \rangle$. From Theorem 1 it follows that for every b in Σ, *Espec* $(G_a, \Sigma^*\{b\}\Sigma^*)$ is ultimately periodic. Let $m_{a,b}$ be a fixed integer larger than *thres* $(Espec(G_a, \Sigma^*\{b\}\Sigma^*))$ and divisible by *per* $(Espec(G_a, \Sigma^*\{b\}\Sigma^*))$. Finally let $n = \prod_{a,b \in \Sigma} m_{a,b}$.

Now if we consider the system $\overline{G}$ resulting from the n-slicing of G then, clearly, it is impatient.

III. SOME SPECIAL CLASSES OF LANGUAGES

In this section we introduce several basic notions needed for our analysis of CFPOL languages.

Definition 3. Let K be a language, $K \subseteq \Sigma^+$.

1) We say that K is *left tight* if

(i) $(\forall \alpha,\beta)_{\Sigma^+}[$*if* $\beta\alpha \in K$ *then* $\alpha \notin K]$,

(ii) $(\forall \alpha,\beta)_{\Sigma^+}[$*if* $\alpha \in K$ *then* $\beta\alpha \notin K]$, and

(iii) $(\forall k)_{N^+}(\exists n_k)_{N^+}(\forall \alpha_1,\alpha_2,\beta)_{\Sigma^+}$

$[$*if* $\alpha_1\beta \in K, \alpha_2\beta \in K, |\alpha_1| \leq k$ *and* $|\beta| \geq n_k$ *then* $\alpha_1 = \alpha_2]$.

2) We say that K is *right tight* if

(iv) $(\forall \alpha_1\beta)_{\Sigma^+}[$*if* $\alpha\beta \in K$ *then* $\alpha \notin K]$,

(v) $(\forall \alpha_1,\beta)_{\Sigma^+}[$*if* $\alpha \in K$ *then* $\alpha\beta \notin K]$ and

(vi) $(\forall k)_{N^+}(\exists n_k)_{N^+}(\forall \alpha_1,\alpha_2,\beta)_{\Sigma^+}$

$[$*if* $\beta\alpha_1 \in K, \beta\alpha_2 \in K, |\alpha_1| \leq k$ *and* $|\beta| \geq n_k$ *then* $\alpha_1 = \alpha_2]$.

Example 1. Let Σ_1,Σ_2 be two disjoint alphabets, let ρ_{12} be a coding from Σ_1 into Σ_2 and let ρ_{21} be a coding from Σ_2 into Σ_1. Then

$K_1 = \{\rho_{21}(\beta)\cdot\beta : \beta \in \Sigma_2^+\}$ is left tight, and

$K_2 = \{\alpha\cdot\rho_{12}(\alpha) : \alpha \in \Sigma_1^+\}$ is right tight.

Definition 4. Let K be a language over Σ.

1) We say that K is *finitely prefixed* if

$(\exists k)_{N^+}(\forall n)_{N^+}(\exists z_1,\ldots,z_k)_{\Sigma^+}$

$[(|z_1|=\ldots=|z_k|=n)$ *and* $\{x \in K : |x|>n$ *and* $pref_n(x) \notin \{z_1,\ldots,z_k\}\}$ is finite$]$.

2) We say that K is *finitely suffixed* if

$(\exists k)_{N^+}(\forall n)_{N^+}(\exists z_1,\ldots,z_k)_{\Sigma^+}$

$[(|z_1|=\ldots=|z_k|=n)$ *and* $\{x \in K : |x|>n$ *and* $suf_n(x) \notin \{z_1,\ldots,z_k\}\}$ is finite$]$.

Example 2. It was proved in [5] that every DOL language is both finitely prefixed and fintely suffixed.

Definition 5. Let $G = \langle \Sigma, P, Ax \rangle$ be an FOL system. We say that G is *extreme* if

$$(\exists a)_N (\exists b)_{R^+} [\text{if } x_0 \Longrightarrow x_1 \Longrightarrow \ldots \Longrightarrow x_m \text{ is a derivation in } G$$
$$\text{then either } |x_m| < a \text{ or } |x_m| > b \cdot m]$$

Definition 6. Let K be a language. We say that K is CFPOL-*extreme* if

$$(\forall K_1 \subseteq K)\ [\text{if } K_1 = \rho(L(G_1)) \text{ where } \rho \text{ is a coding and } G_1 \text{ is an FPOL system}$$
$$\text{then } G_1 \text{ is extreme}].$$

It is rather difficult to provide examples of CFPOL-extreme languages, unless one has a result that binds together a "structural" property of a language with the "grammatical" property of being CFPOL-extreme. Such a result is provided now.

Theorem 2. Let K be a language over Σ such that

i) $(\forall x_1, x_2, y)_{\Sigma^*} (\forall a)_{\Sigma} [\text{if } x_1 a x_2 \in K \text{ and } y \neq a \text{ then } x_1 y x_2 \notin K]$,

and

ii) $(\forall k)_{N^+} (\exists n_k)_{N^+} (\forall \alpha, \beta, \overline{\beta}, \gamma)_{\Sigma^*}$
$[\text{if } \alpha\beta\gamma \in K, \alpha\overline{\beta}\gamma \in K, |\beta| \leq k, |\overline{\beta}| \leq k \text{ and } |\alpha\gamma| \geq n_k \text{ then } \beta = \beta']$.

Then K is CFPOL extreme.

Proof.

Let us assume that $K_1 \subseteq K$ and that $K_1 = \rho(L(G))$ where ρ is a coding and $G = \langle \Sigma, P, Ax \rangle$ is an FPOL system. We will show that G is extreme.

1) First of all let us notice that we can assume that G is impatient. If we construct, as in the proof of Lemma 1, an impatient FPOL system $\overline{G}$ equivalent to G and we find constants $a_{\overline{G}}$, $b_{\overline{G}}$

from Definition 5, then it suffices to take constants $a_G = a_{\overline{G}}$ and $b_G = \frac{b_{\overline{G}}}{n}$ (where $\overline{G}$ results from G by n-slicing) to show that G is extreme.

2) We shall divide letters in Σ into two categories.

$\Sigma_1 = \{a\varepsilon\Sigma : if\ a \underset{P}{\longrightarrow} \alpha\ then\ |\alpha| = 1\}$, and

$\Sigma_2 = \{a\varepsilon\Sigma : if\ a \underset{P}{\longrightarrow} \alpha\ then\ |\alpha| \geq 2\}$.

That $\Sigma_1 = \Sigma_1 \cup \Sigma_2$ is proved as follows.

Let us assume, to the contrary, that there is a letter a in Σ such that $a \underset{P}{\longrightarrow} b$ and $a \underset{P}{\longrightarrow} \alpha$, where $b\varepsilon\Sigma$ and $\alpha\varepsilon\Sigma^+$ with $|\alpha| \geq 2$. Let z be a word in L(G) of the form $z = w_1 a w_2$ and let us consider two one-step derivations from z:

$z \Longrightarrow z_1 = \overline{w}_1 b \overline{w}_2$ and $z \Longrightarrow z_2 = \overline{w}_1 \alpha \overline{w}_2$

which differ only in the way the given occurrence of a is rewritten. But then both $\rho(z_1)$ and $\rho(z_2)$ are in K which contradicts assumption (i) from the statement of the theorem.

3) Let us now make a subdivision of letters in Σ_1.

$\Sigma_{11} = \{a\varepsilon\Sigma_1 : if\ a \underset{P}{\longrightarrow} b\ then\ \rho(a) = \rho(b)\}$, and

$\Sigma_{12} = \{a\varepsilon\Sigma_1 : if\ a \underset{P}{\longrightarrow} b\ then\ \rho(a) \neq \rho(b)\ and\ b\varepsilon\Sigma_{11}\}$.

That $\Sigma_1 = \Sigma_{11} \cup \Sigma_{12}$ is proved as follows.

Let $a\varepsilon\Sigma_1$ and let us consider b_1 and b_2 such that $a \underset{P}{\longrightarrow} b_1$ and $a \underset{P}{\longrightarrow} b_2$. Let z be a word from L(G) of the form $z = w_1 a w_2$ and let us again consider two one step derivations from z:

$z \Longrightarrow z_1 = \overline{w}_1 b_1 \overline{w}_2$ and $z \Longrightarrow z_2 = \overline{w}_1 b_2 \overline{w}_2$

which differ only in the way that the given occurrence of a is rewritten. As $\rho(z_1)\varepsilon K$ and $\rho(z_2)\varepsilon K$ we conclude

$if\ a \underset{P}{\longrightarrow} b_1\ and\ a \underset{P}{\longrightarrow} b_2\ then\ \rho(b_1) = \rho(b_2)$ ... (*) .

Let us now consider a production $b_1 \longrightarrow \beta$ from P. Since G is impatient and $a \in \Sigma_1, |\beta| = 1$. But G is impatient and so $a \underset{G}{\Longrightarrow} \beta$ thus, by (*), $\rho(\beta) = \rho(b_1)$ which completes the proof.

4) G satisfies the following property:

if $w_0 \Longrightarrow w_1 \Longrightarrow \ldots \Longrightarrow w_i \Longrightarrow w_{i+1} \Longrightarrow \ldots \Longrightarrow w_j$ *is a derivation in G and* $|w_i| = |w_{i+1}|$ *then* $|w_j| = |w_i|$.

This is proved as follows.

If $|w_i| = |w_{i+1}|$ *then* w_i consists of letters from Σ_1 only. Since letters from Σ_1 derive letters from Σ_1 only, we have $|w_j| = |w_i|$.

5) G satisfies the following property:

there exists a positive integer constant n_0, such that

if $w_0 \Longrightarrow w_1 \Longrightarrow \ldots \Longrightarrow w_{n_0} \Longrightarrow w_{n_0+1} \Longrightarrow \ldots \Longrightarrow w_j$ *is a derivation in G, and* $|w_{n_0}| < |w_{n_0+1}|$ *then* $|w_j| > |w_{j-1}|$.

This is proved as follows.

Let $\mathcal{D}$ be a set of derivations in G constructed in the following way. One starts with an axiom and carry on a derivation $w_0 \Longrightarrow w_1 \Longrightarrow \ldots \Longrightarrow w_i$ as long as:

(i) either $|w_i| = |w_{i-1}|$, or

(ii) there exists a strict recursive symbol a in w_f, for some $f<i$, such that it contributes to w_i a word of the form $\alpha a \beta$ with $\alpha\beta \neq \Lambda$.

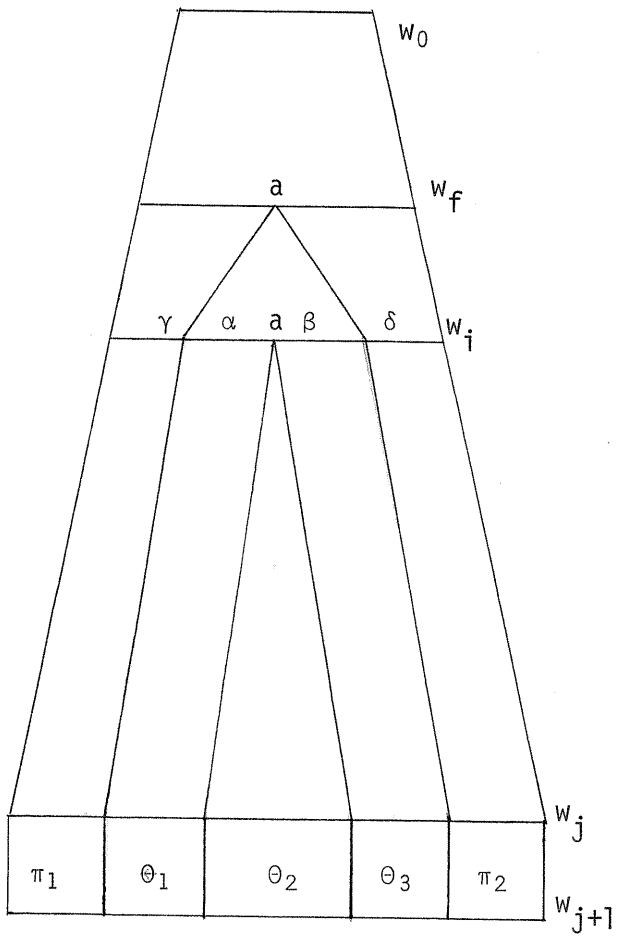
Clearly one of the above two possibilities must occur because if a derivation does not contain the second situation than the length of every word in it is bounded by a constant dependent on G only. Let $\mathcal{D}_2$ be the subset of $\mathcal{D}$ which consists of derivations which do stop by the second condition. Let n_0 be the maximal length of a derivation in $\mathcal{D}$.

We will demonstrate now that if one takes a derivation $D: w_0 \Longrightarrow w_1 \Longrightarrow \ldots \Longrightarrow w_i$ from $\mathcal{D}_2$ and then continues it further as a derivation

$\overline{D}: w_0 \Longrightarrow \ldots \Longrightarrow w_i \Longrightarrow \ldots \Longrightarrow w_j \Longrightarrow w_{j+1}$, *then* $|w_j| > |w_{j+1}|$.

To prove this let us assume, to the contrary, that $|w_j| = |w_{j+1}|$. The situation is the best visualized by the following picture.

$\overline{D}$:



Note that $w_j \in \Sigma_1^+$, $w_{j+1} \in \Sigma_{11}^+$, and $\Theta_1\Theta_2 \neq \Lambda$.

Let $|\Theta_1\Theta_2\Theta_3| = k$ and let n_k be a positive integer constant satisfying condition (ii) from the statement of the theorem. We will construct now two derivations $\overline{D}_1$ and $\overline{D}_2$ resulting from $\overline{D}$ in such a way that

(i) both $\overline{D}_1$ and $\overline{D}_2$ have the same inital piece which is simply a derivation $w \Longrightarrow \ldots \Longrightarrow w_f$ as in $\overline{D}$,

(ii) then in $\overline{D}_1$ we iterate (n_k+1)-times the piece from w_f to w_i in such a way that in each iteration a contributes the same segment $\alpha a \beta$ but elements of γ and δ are rewritten in such a way as in $\overline{D}$, that is they "aim" at π_1 and π_2 respectively and if they reach them then they obviously do not change anymore as far as their coding through ρ is concerned (remember that $w_{j+1} \subseteq \Sigma_{11}^+$); after the completion of this iteration we continue as in the piece from w_i to w_{j+1} in $\overline{D}$.

(iii) in $\overline{D}_2$ the situation is almost the same except that we iterate (n_k+2)-times the piece from w_f to w_i .

Let us now look closer on the results of $\overline{D}_1$ and $\overline{D}_2$. (In what follows for a word x, $x^{(i)}$ denotes the result of rewriting of x by a single iteration step from the above description.) Let $s = n_k$.

First of all after the iteration is completed we get

(iv) in $\overline{D}$ the string of the form
$\gamma^{(s)}{}_{\alpha}{}^{(s)} \ldots \alpha^{(1)}{}_{\alpha a \beta}{}^{(1)} \ldots \beta^{(s)}{}_{\delta}{}^{(s)}$, and

(v) in $\overline{D}_2$ the string of the form
$\gamma^{(s+1)}{}_{\alpha}{}^{(s+1)} \ldots \alpha^{(1)}{}_{\alpha a \beta}{}^{(1)} \ldots \beta^{(s+1)}{}_{\delta}{}^{(s+1)}$.

Then after completing the derivations we get the following strings in K.

(vi) from $\overline{D}_1$ in K
$z_1 = \rho(\pi_1)(\rho(\theta_1))^{s+1}\rho(\theta_2)(\rho(\theta_3))^{s+1}\rho(\pi_2)$, and

(vii) from $\overline{D}_2$
$z_2 = \rho(\pi_1)(\rho(\theta_1))^{s+1}\rho(\theta_1)\rho(\theta_2)\rho(\theta_3)(\rho(\theta_3))^{s+1}\rho(\pi_2)$.

But $|\rho(\theta_2)| \leq k, |\rho(\theta_1)\rho(\theta_2)\rho(\theta_3)| \leq k$ while, remember that $\theta_1\theta_3 \neq \Lambda$, $|\rho(\pi_1)(\rho(\theta_1))^{s+1}(\rho(\theta_3))^{s+1}\rho(\pi_2)| > n_k$ which contradicts conditon (ii) from the statement of the theorem.

Consequently $|w_{j+1}|>|w_j|$ and so (5) holds.

Now to complete the proof we choose

a_G to be (the maximal length of any word appearing in any derivation in $\mathcal{D}$)+1, and

b_G to be any number smaller than 1, for example, 0.9.

Example 3. Let K be a language over V and let V_1, V_2 be alphabets such that V, V_1, V_2 are mutually disjoint. Let ρ_1 be a one-to-one coding from V to V_1 and ρ_2 be a coding from from V to V_2. Than, by Theorem 2, $\{w\rho_1(w)\rho_2(w): w\in K\}$ is CFPOL extreme.

IV. AUXILIARY RESULTS

In this section we investigate the structure of an FPOL system that generates (through a coding) a language $K \subseteq \Sigma_1^+\Sigma_2^+$ where Σ_1, Σ_2 are disjoint alphabets. Our investigation in this section proceeds in such a way that starting with the very simple assumption about K (namely that $K \subseteq \Sigma_1^+\Sigma_2^+$) we will be adding more and more constraints on K at the same time observing their implication on the structure of an "underlying" FPOL system.

To avoid trivial considerations we will assume that K *is an infinite language and according to Lemma 1 we will restrict ourselves to impatient FPOL systems only.*

Thus let Σ_1, Σ_2 be two disjoint alphabets. Let $K \subseteq \Sigma_1^+\Sigma_2^+$ be a language such that $K = \rho(L(G))$ where ρ is a coding and $G = \langle \Sigma, P, Ax \rangle$ is an (impatient) FPOL system.

Definition 7. Let $\sim$ be an equivalence relation on Σ defined by : $a \sim b$ if and only if $\rho(a) = \rho(b)$. Let $a \in \Sigma$.

1) a is called *early* if $\rho(a) \in \Sigma_1$,
2) a is called *late* if $\rho(a) \in \Sigma_2$,
3) a is called *strong* if
   $(\forall b)_\Sigma [\text{if } b \in Acc_G(a) \text{ then } b \sim a]$,
4) a is called *weak* if
   $(\forall b)_\Sigma [\text{if } b \in Acc_G(a) \text{ then } b \not\sim a]$,
5) a is called *mixed* if
   $(\exists b,c)_\Sigma [b \in Acc_G(a), c \in Acc_G(a) \text{ and } b \not\sim c]$.

We will use the following notation to denote various subsets of Σ.

Ea G - early letters in Σ,

La G - late letters in Σ,

Str G - strong letters in Σ,

We G - weak letters in Σ,

Mi G - mixed letters in Σ,

Eas G - early and strong letters in Σ,

Eaw G - early and weak letters in Σ,

Eam G - early and mixed letters in Σ,

Las G - late and strong letters in Σ,

Law G - late and weak letters in Σ,

Lam G - late and mixed letters in Σ,

We leave to the reader the obvious but tedious proof of the following result. This result although not always explicitly mentioned underlies most of the further considerations in this section.

Lemma 2. $L(G) = Z_1 \cup Z_2 \cup Z_3 \cup Z_4 \cup Z_5$, where

$Z_1 \subseteq (EasG)^+(LasG)^+$,

$Z_2 \subseteq (EasG)^+(LawG)^{\ell_2}(LasG)^+$,

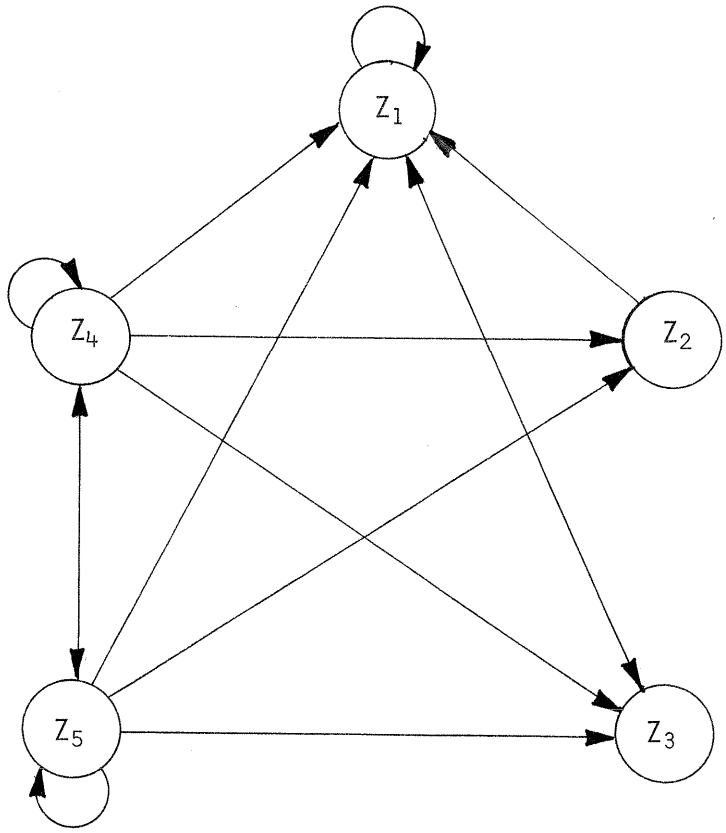
$Z_3 \subseteq (EasG)^+(EawG)^{\ell_3}(LasG)^+$,

$Z_4 \subseteq (EasG)^+(LawG)^{\ell_4}(LamG)(LasG)^*$, and

$Z_5 \subseteq (EasG)^*(EamG)(EawG)^{\ell_5}(LasG)^+$,

with $0<\ell_2,\ell_3<\ell_G, 0\leq\ell_4,\ell_5<\ell_G$ and ℓ_G is a positive integer constant dependent on G only.

Moreover the structure of derivations between words from different components of L(G) looks as follows:



Definition 8. Let $y \in L(G)$.

1) We say that (an occurrence of) a letter b in y is a *last early ancestor* (l.e.a) if there is a derivation from y to a word x such that in this derivation b is an ancestor of the rightmost occurrence of an early letter in x.

2) We say that (an occurrence of) a letter b in y is a *first late ancestor* (f.l.a) if there is a derivation from y to a word x such that in this derivation b is an ancestor of the leftmost occurrence of a late letter in x.

Lemma 3. Let us assume that $\phi_{\Sigma_1}(K)$ is right tight and $\phi_{\Sigma_2}(K)$ is left tight. Then

1) If $y\varepsilon Z_1$ then the rightmost early strong letter in y is a l.e.a. and the leftmost late strong letter in y is a f.l.a.

2) If $y\varepsilon Z_2$ then the rightmost late weak letter in y is a l.e.a. and the leftmost late strong letter in y is a f.l.a.

3) If $y\varepsilon Z_3$ then the rightmost early strong letter in y is a l.e.a. and the leftmost early weak letter in y is a f.l.a.

4) If $y\varepsilon Z_4$ then the late mixed letter in y is both l.e.a. and f.l.a.

5) If $y\varepsilon Z_5$ then the early mixed letter in y is both l.e.a. and f.l.a.

Proof.

The obvious (based on Lemma 2) proofs of 1), 2) and 3) are left to the reader.

4) Let $y\varepsilon Z_4$. Let a be the late mixed letter in y and let us assume that y contains also a letter b which is a l.e.a. or f.l.a.

4.1) Let us assume that b occurs to the right of a. Then clearly it must be a f.l.a. Let us now consider a derivation D from y into z in which b contributes to z the leftmost late letter in z. Then let us change D to $\overline{D}$ in such a way that all letters from y except for a behave precisely as in D and a contributes now a late letter to the last word (let's call it $\overline{z}$). But then $\phi_{\Sigma_2}(\rho(z))$ is a suffix of $\phi_{\Sigma_2}(\rho(\overline{z}))$ which contradicts the assumption that $\phi_{\Sigma_2}(K)$ is left tight.

4.2) Analogously we got a contradiction if we assume that b occurs to the right of a.

5) This is proved anologously to the proof of 4).

Lemma 4. If a is an early mixed recursive letter and $a \longrightarrow \alpha$ is a recursive production in P, then α must contain either an early weak letter or a late strong letter.

Proof.

Let us assume to the contrary that the lemma is not true. Let us consider a derivation $D: y=y_1 a y_2 \Longrightarrow z=z_1 a z_2$ from a word in Z_5 where a is an early mixed recursive letter. Clearly, because of our assumption, all letters in z_2 are strong late and a production used to rewrite a in y is of the form $a \longrightarrow \beta a$ (where β consists of early letters only).

Now let us change D to $\overline{D}$ in such a way that y_1 and y_2 are rewritten exactly as in D but a introduces a late letter. Let the word obtained from y in $\overline{D}$ be $\overline{z}$. Then $\phi_{\Sigma_2}\rho(z)$ is a suffix of $\phi_{\Sigma_2}\rho(\overline{z})$ which contradicts the fact that $\phi_{\Sigma_2}(K)$ is left tight.

Lemma 5. If a is a late mixed recursive letter and $a \longrightarrow \alpha$ is a recursive production in P, then α must contain either late weak letter or early strong letter.

Proof.

Analogous to the proof of Lemma 4.

As a corollary from Lemma 4 and Lemma 5 we get the following result.

Lemma 6. If a is a mixed recursive letter and $a \longrightarrow \alpha$ is a recursive production in P then $|\alpha| \geq 2$. Moreover if $\alpha=\alpha_1 a \alpha_2$ then if a is an early letter then $|\alpha_2| \geq 1$ and if a is a late letter then $|\alpha_1| \geq 1$.

Lemma 7. Let a be a mixed recursive letter and let $D: a \Longrightarrow y \Longrightarrow \ldots \Longrightarrow y_s$ be a derivation in G such that in each step of

this derivation a is rewritten by a recursive production. Let $\overline{D}: a \Longrightarrow \overline{y}_1 \Longrightarrow \overline{y}_2 \Longrightarrow \ldots \Longrightarrow \overline{y}_s$ be a derivation in G. Then $\Phi_{\Sigma_2}\rho(y_s)=\phi_{\Sigma_2}\rho(\overline{y}_s)$.

Proof.

1) Let us assume that a is an early letter. Let for each k, n_k be a constant from the third condition of the definition of a left tight language. Now let T, $\overline{T}$ be two derivations from a word z in Z_5 constructed as follows. Both of them are identical on the first $n_{|y_s|}$ steps and in each of these steps a is rewritten by a recursive production. Then T continues further for s steps with the condition that a is rewritten as in D and $\overline{T}$ continues further for s steps rewriting all letters precisely as in T with the exception that a is rewritten precisely as in $\overline{D}$.

Hence T looks as follows

$$z = z_1^{(\ell)} a z_1^{(r)} \Longrightarrow z_2^{(\ell)} a z_2^{(r)} \Longrightarrow \ldots \Longrightarrow z_{n_{|y_s|}}^{(\ell)} a z_{n_{|y_s|}}^{(r)} \Longrightarrow$$

$$\Longrightarrow z_{n_{|y_s|}}^{(\ell)} {}_{+1} y_1 z_{n_{|y_s|}}^{(r)} {}_{+1} \Longrightarrow \ldots \Longrightarrow z_{n_{|y_s|}}^{(\ell)} {}_{+s} y_s z_{n_{|y_s|}}^{(r)} {}_{+s} ,$$

and $\overline{T}$ looks as follows

$$z = z_1^{(\ell)} a z_1^{(r)} \Longrightarrow z_2^{(\ell)} a z_2^{(r)} \Longrightarrow \ldots \Longrightarrow z_{n_{|y_s|}}^{(\ell)} a z_{n_{|y_s|}}^{(r)}$$

$$\Longrightarrow z_{n_{|y_s|}}^{(\ell)} {}_{+1} \overline{y}_1 z_{n_{|y_s|}}^{(r)} {}_{+1} \Longrightarrow \ldots \Longrightarrow z_{n_{|y_s|}}^{(\ell)} {}_{+s} \overline{y}_s z_{n_{|y_s|}}^{(r)} {}_{+s} .$$

But by Lemma 6, $|z_{n_{|y_s|}}^{(r)}| \geq n_s$. On the other hand $\phi_{\Sigma_2}\rho(z_{n_{|y_s|}}^{(\ell)} {}_{+s} y_s z_{n_{|y_s|}}^{(r)} {}_{+s})$ and $\phi_{\Sigma_2}\rho(z_{n_{|y_s|}}^{(\ell)} {}_{+s} y_s z_{n_{|y_s|}}^{(r)} {}_{+s})$ differ at most on their prefix parts which are $\phi_{\Sigma_2}\rho(y_s)$ and $\phi_{\Sigma_2}\rho(\overline{y}_s)$ respectively. Since $\phi_{\Sigma_2}(K)$ is left tight, we conclude that

$$\phi_{\Sigma_2}\rho(y_s)=\phi_{\Sigma_2}\rho(\overline{y}_s).$$

2) If we assume that a is a late letter then we can prove the lemma analogously.

Lemma 8. Let a be an early mixed recursive letter and let $D:a\Longrightarrow y_1\Longrightarrow y_2\ldots\Longrightarrow y_s$ be a derivation from a in G. Then there exists a sequence of nonempty words $\alpha_1,\ldots,\alpha_s$ such that

$\phi_{\Sigma_2}\rho(y_1)=\alpha_1$,

$\phi_{\Sigma_2}\rho(y_2)=\alpha_1\alpha_2$,

.
.
.

$\phi_{\Sigma_2}\rho(y_s)=\alpha_1\alpha_2\ldots\alpha_s$.

Proof.

By Lemma 7 we can assume that D is such that in each step of D a is rewritten by the same recursive production. Hence D looks as follows

$a\Longrightarrow\gamma_1a\beta_1\Longrightarrow\gamma_2a\beta_1\beta_2\Longrightarrow\ldots\Longrightarrow\gamma_sa\beta_1\beta_2\ldots\beta_s$ where

$a\longrightarrow\gamma_1a\beta_1$ is a production in P,

$\beta_1\Longrightarrow\beta_2,\ldots,\beta_{s-1}\Longrightarrow\beta_s$, and

by Lemma 6 $\beta_1,\ldots,\beta_s$ are nonempty.

From this the lemma follows.

Analogously we prove the following.

Lemma 9. Let a be a late mixed recursive letter and let $D:a\Longrightarrow y_1\Longrightarrow\ldots\Longrightarrow y_s$ be a derivation from a in G. Then there exists a sequence of nonempty words $\alpha_1,\ldots,\alpha_s$ such that

$\phi_{\Sigma_1}\rho(y_1)=\alpha_1$,

$\phi_{\Sigma_1}\rho(y_2)=\alpha_2\alpha_1$,

$\phi_{\Sigma_1}\rho(y_s)=\alpha_s\ldots\alpha_2\alpha_1$.

Lemma 10. Let us assume that K satisfies also the following two conditions:

1) $\phi_{\Sigma_1}(K)$ is CFPOL extreme and infinite, and

2) $\phi_{\Sigma_2}(K)$ is not finitely prefixed.

Then $(\exists u_1)_{\phi_{\Sigma_1}(K)}(\exists u_2)_{\phi_{\Sigma_2}(K)}(u_1u_2 \notin K)$.

Proof.

Let us consider the set of all derivations in G starting with an axiom and continuing untill the obtained word either contains a mixed recursive letter or it consists of strong letters only. Clearly such derivations cannot be longer than m+1 steps where m is the number of mixed letters in Σ. Hence the set of last words in these derivations forms a finite set, say W. We can position W as follows: $W=W_1 \cup W_2 \cup W_3$ where $W_1 \subseteq Z_1$, $W_2 \subseteq Z_4$ and $W_3 \subseteq Z_5$. If a word in L(G) can be derived from W_i then we call it a W_i-*word*.

Now the proof of the lemma goes through two claims.

Claim 1. $(\exists k_o)_{N^+}(\forall z)_{\phi_{\Sigma_1}(K)}$ *if* $|z|>k_o$ *then*

$T_z=\{t \varepsilon \phi_{\Sigma_2}(K):(\exists x)_{\Sigma^+}[(x$ *is a* W_1-word or x *is a* W_2-word) *and* $(zt=\rho(x))]\}$

is finite.

Proof of Claim 1.

Let $U_1=\{\phi_{\Sigma_1}\rho(x):x$ *is a* W_1-word$\}$.

Clearly U_1 is a CFPOL language. To see this take $H=<Eas\ G,R,Early\ W_1>$ where

$Early\ W_1=\{y \varepsilon (Eas\ G)^+:(\exists \bar{y})_{(Las\ G)^+}(y\bar{y} \varepsilon W_1)\}$, and

$R=\{a \longrightarrow \alpha : a \underset{P}{\longrightarrow} \alpha$ *and* $a \varepsilon Eas\ G\}$.

The correctness of the definition of R is insured by the fact that an early strong letter in G derives only early strong letters. It is also clear that $\phi_{\Sigma_1}\rho(L(H))=U_1$.

Since $U_1 \subseteq \phi_{\Sigma_1}(K)$ and $\phi_{\Sigma_1}(K)$ is CFPOL-extreme, H must be extreme. Let then a and b be constants satisfying Definition 6. Let $k_0=a$. Then if a word y is longer than a and it is derived in H in m steps then $|y|>b\cdot m$ hence $m<\frac{|y|}{b}$. Consequently y occurs as the final word only in derivations shorter then $\frac{|y|}{b}$. From the construction of H it follows then that

$(\forall z)_{\phi_{\Sigma_1}(K)}$ *if* $|z|>k_0$ *then*

$T_z^{(1)}=\{t\varepsilon\phi_{\Sigma_2}(K):(\exists x)_{\Sigma^+}[x \text{ is a } W_1\text{-word and } zt=\rho(x)]\}$

is finite.

Let $U_2=\{\phi_{\Sigma_1}\rho(x):x \text{ is a } W_2\text{-word}\}$.

Let us consider an arbitrary derivation $D:u_1\Longrightarrow\ldots\Longrightarrow u_s$ in G such that $u_1\varepsilon W_2$. By Lemma 9 the sequence $|\phi_{\Sigma_1}\rho(u_1)|,|\phi_{\Sigma_1}\rho(u_2)|,\ldots,|\phi_{\Sigma_1}\rho(u_s)|$ is strictly growing and so if z is one of the $\phi_{\Sigma_1}\rho(u_i),1\leq i\leq s$, then $T_z^{(2)}=\{t\varepsilon\phi_{\Sigma_2}(K):(\exists x)_{\Sigma^+}[x \text{ is a } W_2\text{-word and } zt=\rho(x)]\}$ is finite. Hence if we set k_0 to be the maximal length of a word in W and $|z|>k_0$ then $T_z^{(2)}$ is finite.

But for every z in $\phi_{\Sigma_1}(K)$, $T_z=T_z^{(1)}\cup T_z^{(2)}$ and so Claim 1 holds.

Claim 2.

$Y=\{y\varepsilon\phi_{\Sigma_2}(K):\neg(\exists x)_{\Sigma^+}[x \text{ is a } W_3\text{-word and } \phi_{\Sigma_2}\rho(x)=y]\}$

is infinite.

Proof of Claim 2.

Let $a_1,\ldots,a_k$ be the set of all early mixed recursive letters occurring in words of W_3. Let for each $a_i,\alpha_1^{(i)}\ldots\alpha_j^{(i)}\ldots$ be the (infinite) string resulting from the catenation of words

$\alpha_1^{(i)},\ldots,\alpha_j^{(i)},\ldots$ which satisfy the statement of Lemma 8. Let the set of all these catenated words be X.

Since $\phi_{\Sigma_2}(K)$ is not finitely prefixed, there exists a positive integer n such that

$\{y\varepsilon\phi_{\Sigma_2}(K): |y|>n$ and $pref_n(y)\notin Pref_n(X)\}$ is infinite.

Consequently, by Lemma 8, Y is infinite, and so Claim 2 holds.

Now let k_o be a constant from the statement of Claim 1 and let u_1 be an arbitrary word from $\phi_{\Sigma_1}(K)$ such that $|u_1|>k_o$. (Since $\phi_{\Sigma_1}(K)$ is infinite such a word exists.) For this given u_1 let us choose u_2 to be such an element of the set Y (from the statement of Claim 2) which do not belong to the set T_{u_1} (from the statement of Claim 1).

Since $u_2\varepsilon Y$, there do not exist a W_3-word x such that $u_1u_2=\rho(x)$ and because of our choice of u_2 for the given u_1 there do not exist a word y which is either a W_1-word or a W_2-word such that $u_1u_2=\rho(x)$.

Consequently $u_1u_2\notin K$ and the theorem holds.

V. THE MAIN RESULT

Now we can easily prove the main result of this paper.

Theorem 3. $L(EOL)/L(CFPOL)\neq\emptyset$.

Proof.

Let Σ_1,Σ_2 be two disjoint alphabets. Let $K_1\subseteq\Sigma_1^+$ be an EOL language such that K_1 is infinite, right tight and a CFPOL extreme. Let $K_2\subseteq\Sigma_2^+$ be an EOL language such that K_2 is left tight and not finitely prefixed. Then

(i) since $L(EOL)$ is closed w.r.t. catenation (see, e.g., [3]), $K_1\cdot K_2\varepsilon L(EOL)$.

(ii) from Lemma 10 it follows that $K_1\cdot K_2\notin L(CFPOL)$.

Example 4.

Let M be an infinite DOL language over an alphabet Σ. Let $\overline{\Sigma}$ and $\overline{\overline{\Sigma}}$ be two new alphabets such that Σ, $\overline{\Sigma}$, $\overline{\overline{\Sigma}}$ are pairwise disjoint and let ρ_1 be a one-to-one coding from Σ into $\overline{\Sigma}$ and ρ_2 be a one-to-one coding from $\overline{\Sigma}$ into $\overline{\overline{\Sigma}}$.

Let $K_1=\{w\cdot\rho_1(w)\cdot\rho_2(\rho_1(w)):w\in M\}$.

Clearly K_1 is infinite, right tight and (by Theorem 2) also CFPOL extreme. It is also obvious that K_1 is an EOL language.

Let $G=\langle V,P,a\rangle$ be an OL system such that

$V=\{a,b,\overline{b},c,\overline{c}\}$, $V\cap(\Sigma\cup\overline{\Sigma}\cup\overline{\overline{\Sigma}})=\emptyset$, and

$P=\{a\longrightarrow ba\overline{b},a\longrightarrow ca\overline{c},b\longrightarrow b,\overline{b}\longrightarrow\overline{b},c\longrightarrow c,\overline{c}\longrightarrow\overline{c}\}$.

Then, by Lemma 10, $K_1\cdot K_2$ is not in L(CFPOL) while obviously $K_1\cdot K_2$ is an EOL language.

Let us notice that both K_1 and K_2 from Example 4 are CFPOL languages, and so as a corollary from Lemma 10 we also get the following result.

Theorem 4. L(CFPOL) is not closed with respect to catenation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge support by NSF Grant No. DCR75-09972 and by Belgian National Foundation for Scientific Research (NFWO). We are indebted to L. Fosdick for providing us with an opportunity to get together and work on this and other papers.

REFERENCES

[1] Ehrenfeucht, A. and Rozenberg, G., 1974. The equality of EOL languages and codings of OL languages, International Journal of Computer Mathematics, 4: 95-104.

[2] Ehrenfeucht, A. and Rozenberg, G., 1976. On existential and universal spectra of ETOL systems, to appear.

[3] Herman, G. T. and Rozenberg, G., 1975. Developmental systems and languages, North Holland Publishing Company, Amsterdam.

[4] Lindenmayer, A. and Rozenberg, G., (eds.), 1976. Automata, languages and development, North Holland Publishing Company, Amsterdam.

[5] Rozenberg, G., 1974. DOL sequences, Discrete Mathematics, 7: 323-347.

[6] Rozenberg, G. and Salomaa, A., 1976. The mathematical theory of L systems, In: J. Tou (ed.), Advances in Information Systems Science, Plenum Press, New York, 161-206.

[7] Salomaa, A., 1973. Formal languages, Academic Press, London-New York.