All correspondence to second author

FPOL SYSTEMS GENERATING COUNTING LANGUAGES

by

A. Ehrenfeucht*

and

G. Rozenberg**

#CU-CS-113-77 September, 1977

* A. Ehrenfeucht, Department of Computer Science, University of Colorado at Boulder, Boulder, Colorado 80309 USA

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

WORKING PAPER

ABSTRACT

Counting languages are the languages of the form $\{a_1^n a_2^n \ldots a_t^n : t \geq 2, n \geq 1\}$ where $a_1, \ldots, a_t$ are letters no two consecutive of which are identical. They possess a "clean structure" in the sense that if an arbitrary word from such a language is cut in t subwords of equal length then no two consecutive subwords contain an occurrence of the same letter. It is shown that whenever an FP0L system G is such that its language contains a "dense enough" subset of a counting language then the whole language of G cannot have such a clean structure.

I. INTRODUCTION

One of the important research areas in formal language theory is the search for results describing the structure of a single language within a given language family. The classical example of such a result is the "pumping lemma" for context free languages. It says that if certain words are in a context free language then (infinitely many) other words must be also in this language. Such results clearly shed some light on the generating abilities (restrictions) of grammars or machines) defining the given class of languages.

In this paper we establish a result in this direction for the class of languages generated by 0L systems without erasing productions and with finite axiom sets (called FP0L systems). One of the most popular type of languages (serving as examples of strict inclusions of some classes of languages in others) in formal language theory are t-counting languages. These are languages of the form $\{a_1^n a_2^n \ldots a_t^n : t \geq 2, n \geq 1\}$ where $a_1, \ldots, a_t$ are letters no two consecutive of which are identical. They possess a "clean structure" in the sense that if an arbitrary word from such a language is cut into t subwords of equal length then no two consecutive subwords share an occurrence of a common letter. We demonstrate that if an FP0L system G is such that its language contains a "dense enough" subset of a counting language, then the whole language cannot have such a clean structure (or even a structure "approximating" it!). Thus again a result in this line: if certain words are in the language from the given class, then other words must also be in the same language.

Certainly there are very few results like this for the class of FP0L languages and we believe that this result together with its proof shed some new light on the structure of derivations in FP0L systems.

Perhaps it is also worthwhile to mention that results like this are especially valuable in the theory of L forms where one is really interested in the structure of "all sentential forms" that a given system can generate. In particular our result is used in [4].

II. PRELIMINARIES

We assume the reader to be familiar with rudiments of formal language theory and in particular with the rudiments of the theory of L systems (see, e.g. [1]). We use a rather standard terminology and perhaps only the following notation requires an explanation.

1) N, N^+ and $N(t)$ denote the set of nonnegative integers, positive integers and positive integers larger than t, respectively.

2) For a finite set Z, #Z denotes its cardinality and for an integer x, $\|x\|$ denotes its absolute value.

3) If α is a word over Σ then $\underline{alph}(\alpha)$ denotes the set of all letters from Σ that occur in α, $\underline{pref}_k(\alpha)$ denotes the prefix of α of the length k and $\underline{suf}_k(\alpha)$ denotes the suffix of α of the length k. $|\alpha|$ denotes the length of α and α/x, $\alpha\backslash x$ denote the word resulting by deleting the prefix x from α and the word resulting by deleting the suffix x from α, respectively. $\#_a(\alpha)$ denotes the number of occurrences of the letter a in α.

4) If k is a language then $\underline{alph}(K) = \bigcup_{\alpha \in M} \underline{alph}(\alpha)$ and $\underline{less}_q(K) = \#\{\,|\alpha| : \alpha \in K \text{ and } |\alpha| \leq q\}$.

5) In our notation we often identify a singleton set with its element.

To establish the basic notation for this paper concerning L systems we recall now the definition of an FP0L system.

Definition 1.

1) An FP0L system is a construct $G = (\Sigma,P,A)$ where Σ is a finite nonempty alphabet, P is a finite set of productions, each of the form $a \to \alpha$ with $a \in \Sigma, \alpha \in \Sigma^+$ satisfying the condition $(\forall a)_{\Sigma}(\exists \alpha)_{\Sigma^+}[a \to \alpha \text{ is in } P]$, A is a finite nonempty set (of axioms), $A \subseteq \Sigma^+$.

2) Given words $x,y \in \Sigma^+$ we say that x directly derives y in G if $x = a_1 \ldots a_t$, $y = \alpha_1 \ldots \alpha_t$ where $\langle a_1,\alpha_1\rangle, \ldots \langle a_t,\alpha_t\rangle \in P$. We write then $x \underset{G}{\Rightarrow} y$.

3) For a positive integer m we say that x derives y in m steps if there exist $x_1,\ldots,x_m$ such that $x_0 \underset{G}{\Rightarrow} x_1$, $x_1 \underset{G}{\Rightarrow} x_2, \ldots, x_{m-1} \underset{G}{\Rightarrow} x_m$ and $x_m = y$. We denote it by $x \overset{m}{\underset{G}{\Rightarrow}} y$. If $x = y$ or there exists an m such that $x \overset{m}{\underset{G}{\Rightarrow}} y$ then we say that x derives y in G and denote it by $x \overset{*}{\underset{G}{\Rightarrow}} y$.

4) The language of G, denoted as $L(G)$, is defined by $L(G) = \{\alpha \in \Sigma^+ : (\exists w)_A[w \overset{*}{\Rightarrow} \alpha]\}$.

Definition 2. Let $G = (\Sigma,P,A)$ be an FP0L system.

1) Let $\alpha \in \Sigma^+$. Then $G_\alpha = (\Sigma,P,\alpha)$.

2) Let $n \in N^+$. Then $L(n,G) = \{\alpha \in L(G) : (\exists w)_A[w \overset{n}{\underset{G}{\Rightarrow}} \alpha]\}$ and $L(n,\alpha,G) = L(n,G_\alpha)$.

3) $\mathit{Inf}(G) \subseteq \Sigma$ where $a \in \mathit{Inf}(G)$ if and only if $\{\alpha \in L(G) : a \in \mathit{alph}(\alpha)\}$ is infinite; elements of $\mathit{Inf}(G)$ are called infinite letters (in G).

4) $\mathit{Fin}(G) = \Sigma \setminus \mathit{Inf}(G)$; elements of $\mathit{Fin}(G)$ are called finite letters (in G).

5) Mult(G) $\subseteq$ Inf(G) where $a \in$ Mult(G) if and only if

$(\forall n)_{N^+}(\exists\alpha)_{L(G)}[\#_a(\alpha) > n]$;

elements of Mult(G) are called multiple letters (in G).

6) Copy(G) $= \{m \in N^+ : (\exists\alpha)_{\Sigma^+}[\alpha^m \in L(G)]\}$.

7) The growth function of G, denoted as f_G, is a function from N^+ into finite subsets of N^+ defined by

$f_G(n) = \{|\alpha| : \alpha \in L(n,G)\}$.

7.1) If there exists a polynomial ϕ such that

$(\forall n)_{N^+}(\forall m)_{f_G(n)}[m < \phi(n)]$

then we say that f_G is of polynomial type;

otherwise f_G is exponential.

7.2) If there exists a constant C such that

$(\forall n)_{N^+}(\forall m)_{f_G(n)}[m < C]$

then we say that f_G is constant.

7.3) If $(\forall n)_{N^+}[\#f_G(n) = 1]$

then we say that f_G is deterministic.

III. AUXILIARY RESULTS

In this section we investigate certain aspects of derivations in FP0L systems in general and in the so called t-balanced FP0L systems in particular.

Lemma 1. Let G be an FP0L system such that

1. f_G is deterministic, and
2. Copy (G) is an infinite set.

Then f_G is exponential.

Proof.

Let $G = (\Sigma, P, A)$.

Let $H = (\Sigma, R, A)$ where R is a deterministic subset of G, i.e. $R \subseteq P$ and R contains precisely one production for each letter of Σ.

Obviously $f_G = f_H$.

If f_H is exponential then the result holds.

But now we will show that f_H cannot be of polynomial type.

Assume otherwise i.e. that f_H is of polynomial type.

(i) From [2] it follows that there exist positive integers n_0, n_1 and polynomials $\psi_1, \ldots, \psi_{n_1}$ such that, for $n_0+1 \leq r \leq n_0+n_1$ and for every $\ell \geq 0$,

$f_G(r+\ell \cdot n_1) = \phi_r(\ell)$.

(ii) Note that

$(\forall m)_{\text{Copy}(G)} (\exists n_m)_{N^+} (\forall n)_{N^+} [\underline{\text{if}}\ n > n_m\ \underline{\text{then}}\ m \text{ divides } f_H(n)]$.

To prove it, given $m \in$ Copy(G) choose n_m to be such that

$\alpha^m \in L(n_m,G)$ for a word α in Σ^+.

(iii) Let $r \in \{n_0+1,\ldots,n_0+n_1\}$ and let

$\phi_r(\ell) = a_k\ell^k+a_{k-1}\ell^{k-1}+\ldots+a_1\ell+a_0$.

Let $m > \max\{\ \|a_i\|\cdot(k+1)^{(k+1)} : 0 \le i \le k\}$.

$$
\begin{aligned}
\text{Let } Z_1 &= \{\phi_r(m\cdot d+1) &: d \ge 1\},\\
Z_2 &= \{\phi_r(m\cdot d+2) &: d \ge 1\},\\
&\vdots\\
Z_{k+1} &= \{\phi_r(m\cdot d+k+1) &: d \ge 1\}.
\end{aligned}
$$

We will demonstrate now that at least one of the above Z_i is such that none of its elements is divisible by m.

Since $(md+i)^t = a_i m+i^t$ for some a_i,

$\phi_r(md+i) = s_i m+\phi_r(i)$ for some s_i.

But for $i \in \{1,\ldots,k+1\}$, $\|\phi_r(i)\| < m$ and so m divides elements of Z_i only if $\phi_r(i) = 0$.

But if all $\phi_r(i) = 0$ for all $1 \le i \le (k+1)$ then the following set of equations would have to have a non-zero solution:

$$
\begin{aligned}
a_k 1^k+a_{k-1}1^{k-1}+\ldots\ldots\ldots\ldots+a_1 1+a_0 &= 0,\\
a_k 2^k+a_{k-1}2^{k-1}+\ldots\ldots\ldots\ldots+a_1 1+a_0 &= 0,\\
\vdots \qquad\qquad &\vdots\\
a_k(k+1)^k+a_{k-1}(k+1)^{k-1}+\ldots+a_1(k+1)+a_0 &= 0.
\end{aligned}
$$

Since the determinant of this system is a Vandermonde determinant it never equals zero and so the above system does not have a solution. Consequently $\phi_r(i) \neq 0$ for at least one $i \in \{1,\ldots,k+1\}$ which implies that m does not divide any element of Z_i.

(iv) This however contradicts (ii) and consequently f_H cannot be of polynomial type.

Definition 3.

(i) Let $\alpha \in \Sigma^+$ and let t be a positive integer, $t \geq 2$. A t-*disjoint decomposition of* α is a vector $\langle \alpha_1,\ldots,\alpha_t \rangle$ such that

1. $\alpha_1,\ldots,\alpha_t \in \Sigma^+$ and $\alpha_1 \ldots \alpha_t = \alpha$, and
2. for every $1 \leq i \leq t-1$, $\underline{alph}(\alpha_i) \cap \underline{alph}(\alpha_{i+1}) = \emptyset$.

(ii) Let $K \subseteq \Sigma^+$ and let t be a positive integer, $t \geq 2$. We say that K is t-*balanced* if there exist positive rational numbers $c_1,\ldots,c_t$ with $\bigcup_{i=1}^{t} c_i = 1$ and a positive integer d such that for every α in K there exists a t-disjoint decomposition $\langle \alpha_1,\ldots,\alpha_t \rangle$ such that for every $1 \leq i \leq t$,

$c_i|\alpha|-d \leq |\alpha_i| \leq c_i|\alpha|+d$.

In such a case we also say that K is (v,d)-*balanced* and that $\langle \alpha_1,\ldots,\alpha_t \rangle$ is a (v,d)-*balanced decomposition of* α, where $v = \langle c_1,\ldots,c_t \rangle$.

(iii) An FPOL G is t-*balanced* if L(G) is t-balanced.

Lemma 2. Let $G = (\Sigma,P,A)$ be a t-balanced FPOL system with $t \geq 3$. Then

$(\exists k_0)(\forall a)_\Sigma(\forall n)_N[\#f_{G_a}(n) < k_0]$.

Proof.

Clearly it suffices to show that

$(\forall a)_\Sigma(\exists k_a)(\forall n)_N[\#f_{G_a}(n) < k_a]$.

Let $v = \langle c_1,\ldots,c_t\rangle$ and d be such that $L(G)$ is (v,d)-balanced. Let $c_{min} = \min\{c_1,\ldots,c_t\}$.

(i) Let $a \in$ Inf(G).

We will prove the result by contradiction. So let us assume to the contrary that

it is not true that $(\exists k_a)(\forall n)_N[\#f_{G_a}(n) < k_a]\ldots(*)$

(i.1) $(\exists n_0)_{N^+}(\exists r)_{N(\#\Sigma)}(\exists w_1,\ldots,w_r)_{L(n_0,a,G)}(\forall i)_{\{1,\ldots,t\}}$ $[c_i|w_{i+1}| > c_i|w_i|+3d]$

Proof of (i.1).

Clearly it suffices to show (i.1) with c_i replaced by c_{min}.

Let us take an arbitrary n and let $f_{G_a}(n) = \{x_1,\ldots,x_s\}$ where elements $x_1,\ldots,x_s$ are arranged in the increasing order. Let $x_{i_1},\ldots,x_{i_r}$ be the longest subsequence of $x_1,\ldots,x_s$ defined as follows:

$x_{i_1} = x_1$, and

for $1 \le j \le r-1$, i_{j+1} is the smallest index with the property that $x_{i_{j+1}} - x_{i_j} > \frac{3d}{c_{min}}$.

If $r \le \#\Sigma$ then $s \le \#\Sigma\cdot\frac{3d}{c_{min}}$. Since n was arbitrary, if we set $k_a = (\#\Sigma \cdot \frac{3d}{c_{min}}) + 1$ then we get that

$(\forall n)_{N+}[\# f_{G_a}(n) < k_a]$

which contradicts (*).

Thus r must be larger than #Σ and so if we set

$w_1 = x_{i_1}, \ldots, w_r = x_{i_r}$ the statement (i.1) holds.

(i.2) Let α be a word containing a, so $\alpha = \alpha_1 a \alpha_2$, which is long enough, i.e.,

$(\forall i)_{\{1,..,t\}}[|\alpha| c_i > 2(|w_r|+3d)]$

where $w_1, \ldots, w_r$ is a sequence (in the order of increasing length) from (i.1).

Let

$\beta_1 = \bar{\alpha}_1 w_1 \bar{\alpha}_2 \in L(n_0, \alpha, G)$,

$\vdots$

$\beta_r = \bar{\alpha}_1 w_r \bar{\alpha}_2 \in L(n_0, \alpha, G)$,

where

$\bar{\alpha}_1 \in L(n_0, \alpha_1, G)$ and $\bar{\alpha}_2 \in L(n_0, \alpha_2, G)$.

Let for each $1 \leq i \leq r$, $u_i = \langle \beta_i(1), \ldots, \beta_i(t) \rangle$ be a $\langle v,d \rangle$-balanced decomposition of β_i.

Since $|\beta_i| \geq |\alpha_i|$ and $t \geq 3$ either w_i is contained in $\beta_i/(\beta_i(1))(\underline{pref}_{|w_r|+2d}(\beta_i(2))))$ or w_i is contained in

$\beta_i \backslash ((\underline{suf}_{|w_r|+2d}(\beta_i(t-1)))(\beta_i(t)))$; because these two cases are symmetric we assume the first one.

Note that $\beta_{i+1}(1)$ results from $\beta_i(1)$ by catenating to $\beta_i(1)$ a prefix of $\beta_i(2)$. Moreover

$|\beta_{i+1}(1) - \beta_i(1)| \geq (c_i(|\beta_i| + \frac{3d}{c_{min}}) - d) - (c_i|\beta_i| + d) =$

$= c_i \frac{3d}{c_{min}} - 2d \geq 3d-2d = d$

and so each $\beta_{i+1}(1)$ results from $\beta_i(1)$ by catenating to $\beta_i(1)$ a <u>nonempty</u> prefix of $\beta_i(2)$.

Also

$|\beta_r(1)|-|\beta_1(1)| \leq (c_1(|\bar{\alpha}_1\bar{\alpha}_2|+|w_r|)+d)-(c_1(|\bar{\alpha}_1\bar{\alpha}_2|+|w_1|)-d =$

$= c_1(|w_r|-|w_1|)+2d \leq |w_r|+2d$

and so in constructing consecutively $\beta_2(1), \beta_3(1), \ldots, \beta_r(1)$ we use nonempty subwords of a prefix of $\beta_1(2)$ (and we never reach w_1).

But $r > \#\Sigma$ and so at least two of the nonempty prefixes used to construct $\beta_{i+1}(1)$ from $\beta_i(1)$ contain an occurrence of the same letter, which implies that for a j in $\{2,\ldots,r-1\}$, $\underline{alph}(\beta_j(1)) \cap \underline{alph}(\beta_j(2)) \neq \emptyset$; a contradiction with the fact that β_j has a (v,d)-balanced decomposition.

(i,3) Thus (i.1) and (i.2) imply that (*) is not true, or in other words that, indeed,

$(\forall a)_{\underline{Inf}(G)} (\exists k_a)(\forall n)_N [\# f_{G_a}(n) < k_a]$.

ii) Let $a \in \underline{Fin}(G)$.

Let Z_a be the set of all words α such that $\underline{alph}(\alpha) \subseteq \underline{Inf}(G)$ and there exists a word β such that $\beta \overset{+}{\underset{G}{\Rightarrow}} \alpha$ and $\underline{alph}(\beta) \cap \underline{Fin}(G) \neq \emptyset$.

Note that Z_a is a finite set and so if we set

$s = \max\{|\alpha| : \alpha \in Z_a\}$,

$r = \#\{\alpha \in \underline{Fin}(G) : \underline{alph}(\alpha) \cap \{a\} \neq \emptyset$ and

$k = \max\{k_a : a \in \underline{Inf}(G)\}$

then

$(\forall n)_N[\#f_{G_a}(n) < 1+r+k^s]$.

Lemma 3. Let G be an FP0L system and let $a \in$ MULT(G). Then f_{G_a} is deterministic.

Proof

Let $G = \langle\Sigma, P, A\rangle$

Clearly there exists a letter b in Σ which for any m can derive a word β_m such that $\#_a(\beta_m) > m$. So let k_0 be a constant from the statement of Lemma 2 and let β be a word such that b derives β (in some e steps) and $\#_a(\beta) > k_0$.

We will show now that if the lemma is not true, then we get a contradiction.

If the lemma is not true then

$(\exists n_0)_{N^+}(\exists \alpha_1, \alpha_2)_{\Sigma^+}[\alpha_1, \alpha_2 \in L(n_0, a, G)$ and $|\alpha_1| \neq |\alpha_2|]$.

But then the number of words of different length that β can derive in n_0 steps is larger than k_0 and consequently $\#f_{G_b}(e+n_0) > k_0$, which contradicts Lemma 2.

IV. THE MAIN RESULT

In this section we prove the main result of this paper. First we recall a definition and a technical result needed later on.

Definition 4.

1) Let $G = (\Sigma, P, A)$ be an FP0L system and let m be a positive integer. An m-decomposition of G is a set of systems $\mathcal{G} = \{G_1, \ldots, G_m\}$ such that for every $1 \leq i \leq m$, $G_i = (\Sigma, R, A_i)$ where $A_i = \{\alpha : \alpha \in L(i-1, G)\}$ for $i \geq 2$ and $A_1 = A$, $R = \{a \to \alpha : a \overset{m}{\underset{G}{\Longrightarrow}} \alpha\}$.

2) We say that G is well-sliced if

(i) $(\forall a)_{\Sigma} (\forall \alpha)_{\Sigma^+} [(\exists m)_{N^+} [a \overset{m}{\underset{G}{\Longrightarrow}} \alpha]$ if and only if

$$(\exists \bar{\alpha})_{\Sigma^+} [(a \underset{G}{\Longrightarrow} \bar{\alpha}) \text{ and } (\underline{alph}(\alpha) = \underline{alph}(\bar{\alpha}))]],$$

(ii) $(\forall a)_{\Sigma} [$if $\bigcup_{n \geq 1} L(n, a, G)$ is finite then

$$\{\alpha : a \underset{G}{\Longrightarrow} \alpha\} = \bigcup_{n \geq 1} L(n, a, G)]$$

3) An m-decomposition $\mathcal{G}$ of G is called well-sliced if each element of it is well-sliced.

Note that $L(G) = \bigcup_{i=1}^{m} L(G_i)$ where $\mathcal{G} = \{G_1, \ldots, G_m\}$ is an m-decomposition of G.

The following result was proved in [1].

Lemma 4. For every FP0L system there exists a well-sliced decomposition.

Our main result will be concerned with the generating of counting languages by FP0L systems. These counting languages are defined formally now.

Definition 5. Let t be a positive integer, $t \geq 2$. A language M over Σ is called a t-counting language if $M = \{a_1^n a_2^n \ldots a_t^n : n \geq 1\}$ where for $1 \leq i \leq t$, $a_i \in \Sigma$ and $a_j \neq a_{j+1}$ for $1 \leq j \leq t-1$. We also say that a_j is a neighbour in M with a_{j+1}.

Theorem 1. Let $t \geq 3$, M be a t-counting language, G be a t-balanced FP0L system and $K = M \cap L(G)$. Then there exists a constant C such that

$(\forall q)_{N^+}[\mathrm{less}_q K \leq C \cdot \log q]$.

Proof

Let $G = (\Sigma, P, A)$ and $\mathrm{alph}(M) = \Delta$.

By Lemma 4, there exists a well-sliced decomposition of G and since it suffices to prove the theorem for a single component of such a decomposition let us assume that G is well-sliced.

(i) If K is finite then the result trivially holds.

(ii) Assume that K is infinite.

(ii.1) $(\exists a)_{\mathrm{Mult}(G)}(\exists b)_{\Delta}(\exists \alpha)_{\{b\}^+}[a \stackrel{+}{\Rightarrow} \alpha]$.

Proof of (ii.1).

Obvious (otherwise K could not be infinite).

(ii.2) If $a \in \mathrm{Mult}(G)$, $b \in \Delta, \alpha \in \{b\}^+$ and $a \stackrel{+}{\Rightarrow} \alpha$ then

1) f_{G_a} is either constant or exponential,

2) f_{G_b} is either constant or exponential, and

3) f_{G_a} is constant if and only if f_{G_b} is constant.

Proof of (ii.2)

1) By Lemma 3, f_{G_a} is deterministic and since G is well-sliced

$(\forall n)_{N^+}[\ell \in f_{G_a}(n)$ if and only if $b^\ell \in L(n,a,G)]$.

Let $\tau = \sigma^{i_1}, \sigma^{i_2}, \ldots$ be such that $i_j = f_{G_a}(j)$.

If τ is infinite then G_a satisfies the assumptions of Lemma 1 and so f_{G_a} is exponential.

If τ is finite then, because G is well sliced, f_{G_a} is a constant function.

2) and 3) follow from 1), a derives strings through b and so both must have the same type of growth.

(ii.3) Either $(\forall b)_\Delta[f_{G_b}$ is a constant function]

or $(\forall b)_\Delta[f_{G_b}$ is exponential].

Proof of (ii.3).

From (ii.1) and (ii.2) it follows that

$(\exists c)_\Delta[f_{G_c}$ is either constant or exponential]

Now let a be a neighbour of c in M. Then if we take a word α from K of the form $\ldots a^n c^n \ldots$ (or symmetrically $\ldots c^n a^n \ldots$) and will derive in G words from it in such a way that each occurrence of c in α will produce the same subtrees then if c is not of the same type as a then we will obtain a word β that is not t-balanced, a contradiction.

Hence any two neighbours in M must have the same type of growth.

(ii.4) It is not true that $(\forall a)_{\Delta}[f_{G_a}$ is constant]

Proof of (ii.4).

We prove it by showing that if indeed $(\forall a)_{\Delta}[f_{G_a}$ is constant] then K must be finite; a contradiction.

So assume that K is infinite. Then we can choose α in K which is arbitrarily long, e.g., so long that each derivation tree (or rather forest) in G for α is such that on each path of it there exists a label that appears at least twice.

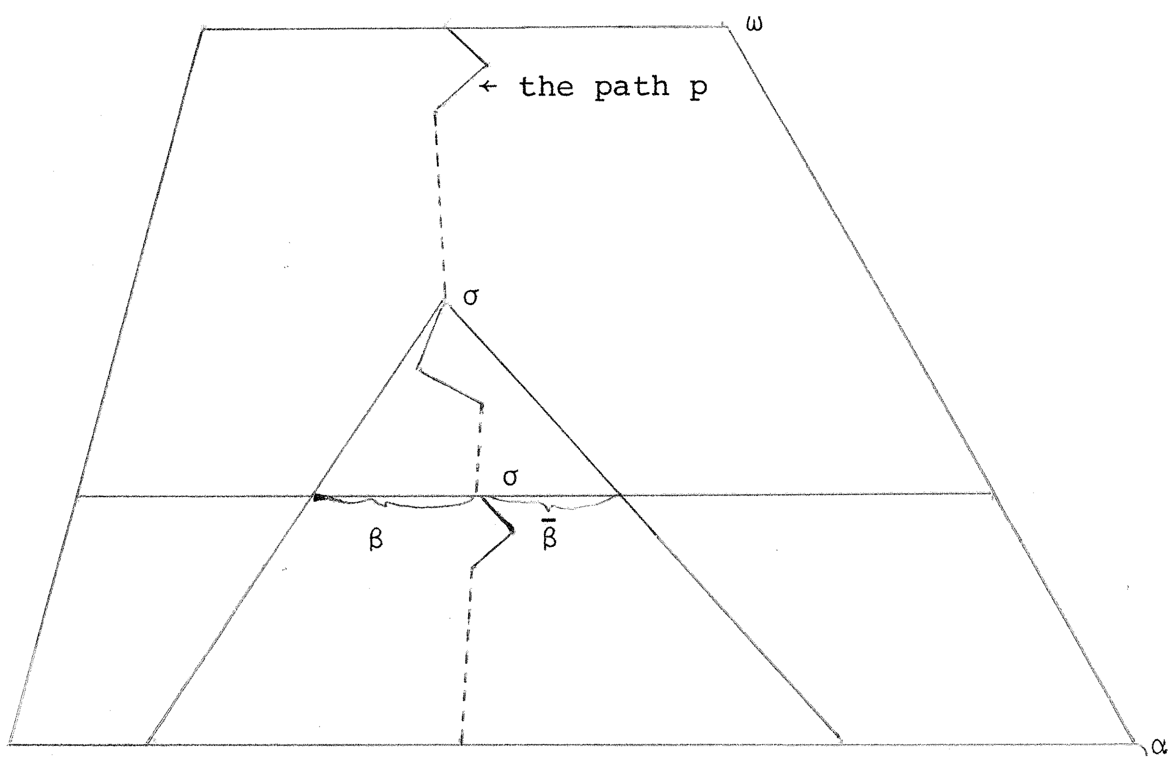
In a derivation forest from ω in A to α we choose a path $p = e_0, e_1, \ldots$ as follows:

e_0 is an occurrence in ω such that no other occurrence in ω contributes a longer subword to α,

e_{i+1} is a direct descendant of e_i such that no other direct descendant of e_i contributes a longer subword to α.

Now on p choose the first (from e_0) label σ that repeats itself at least twice on p. Then take the first repetition of σ on p.

The situation is the best explained by the following picture:


ω
← the path p
σ
σ
β
$\bar{\beta}$
α

1) $\beta\beta \neq \Lambda$.

We prove it by contradiction. Suppose $\beta\bar{\beta} = \Lambda$.

1.1) Then every label ρ on $\underline{p}$ which repeats itself must be such that $\rho \underset{G}{\overset{+}{\Longrightarrow}} \delta\rho\bar{\delta}$ implies $\delta\bar{\delta} = \Lambda$.

This is seen as follows.

Let n_0, n_1, n_2 be such that

$$\sigma \overset{n_0}{\Longrightarrow} \sigma,\ \rho \overset{n_1}{\Longrightarrow} \delta\rho\bar{\delta},\ \sigma \overset{n_2}{\Longrightarrow} \zeta\rho\bar{\zeta}.$$

Let $n = n_0 \cdot n_1 \cdot n_2$.

Then

$$\begin{array}{llllll}
\sigma \overset{n}{\Longrightarrow} \gamma_1\rho\bar{\gamma}, & \overset{n}{\Longrightarrow} \gamma_1^{(n)}\delta\rho\bar{\delta}\bar{\gamma}_1^{(n)} & & \overset{n}{\Longrightarrow} \gamma_1^{(2n)}\delta^{(n)}\delta\rho\bar{\delta}\bar{\delta}^{(n)}\bar{\gamma}^{(2n)} \\
 & & & \overset{n}{\Longrightarrow} \ldots \\
\sigma \overset{n}{\Longrightarrow} & \sigma & \overset{n}{\Longrightarrow} \gamma_1\rho\bar{\gamma}_1 & \overset{n}{\Longrightarrow} \gamma_1^{(n)}\delta\rho\bar{\delta}\gamma_1^{(n)} \\
 & & & \overset{n}{\Longrightarrow} \ldots \\
\sigma \overset{n}{\Longrightarrow} & \sigma & \overset{n}{\Longrightarrow} \sigma & \overset{n}{\Longrightarrow} \gamma_1\rho\bar{\gamma}_1 \\
 & & & \overset{n}{\Longrightarrow} \ldots \\
\vdots & & & \vdots
\end{array}$$

for some words $\gamma_1, \bar{\gamma}_1, \ldots, \gamma_1^{(i)}, \bar{\gamma}_1^{(i)}, \ldots\ \delta^{(i)}, \bar{\delta}^{(i)}, \ldots$.

Hence if $\delta\bar{\delta} \neq \Lambda$, then there exists a positive integer ℓ such that $\# f_{G_\sigma}(\ell \cdot n) > k_0$; which contradicts Lemma 2.

But if 1.1) holds then α would have to be not longer than a fixed a priori constant; a contradiction to the fact K is infinite and so we could have choosen α arbitrarily long.

Thus indeed $\beta\bar{\beta} \neq \Lambda$.

2) Let m_0, m_1 be such that $\sigma \overset{m_0}{\Longrightarrow} \beta\sigma\bar{\beta}$ and $\sigma \overset{m_1}{\Longrightarrow} \pi$ where $\pi \in \Delta^+$ (so f_{G_π} is constant). Let $m = m_0 \cdot m_1$.

Then

$$\sigma \overset{m}{\Longrightarrow} \pi \overset{m}{\Longrightarrow} \pi^{(1)} \overset{m}{\Longrightarrow} \pi^{(2)} \overset{m}{\Longrightarrow} \pi^{(3)} \Longrightarrow \dots$$

$$\sigma \overset{m}{\Longrightarrow} \gamma_1\sigma\bar{\gamma}_1 \overset{m}{\Longrightarrow} \gamma_1^{(1)}\pi\bar{\gamma}_1^{(1)} \overset{m}{\Longrightarrow} \gamma_1^{(2)}\pi^{(1)}\bar{\gamma}_1^{(2)} \overset{m}{\Longrightarrow} \dots$$

$$\sigma \overset{m}{\Longrightarrow} \gamma_1\sigma\bar{\gamma}_1 \overset{m}{\Longrightarrow} \gamma_1^{(1)}\gamma_1\sigma\bar{\gamma}_1\bar{\gamma}_1^{(1)} \overset{m}{\Longrightarrow} \gamma_1^{(2)}\gamma_1^{(1)}\pi\bar{\gamma}_1^{(1)}\bar{\gamma}_1^{(2)} \overset{m}{\Longrightarrow} \dots$$

$\vdots$ $\qquad\qquad\qquad\qquad$ $\vdots$

where all the $\gamma_1, \bar{\gamma}_1, \dots, \gamma_1^{(i)}\ \bar{\gamma}_1^{(i)}, \dots, \pi, \dots, \pi^{(i)}, \dots$ are nonempty words.

Since f_{G_π} is constant, this implies that there exists a positive integer ℓ such that $\# f_{G_0}(\ell \cdot n) > k_0$; which contradicts Lemma 2. Consequently if

$(\forall a)\ [f_{G_a}$ is constant]

then K must be finite; a contradiction.

Thus (ii.4) holds.

(ii.5) $\quad (\forall b)_\Delta [f_{G_b}$ is exponential]

<u>Proof of (ii.5)</u>

It follows directly from (ii.3) and (ii.4).

(ii.6) There exists a positive integer constant s_0 such that in every derivation without repetitions of a word from K in G, already after s_0 steps an intermediate word contains an occurrence of a multiple letter a for which there exists b in Δ and α in $\{b\}^+$ such that $a \overset{+}{\Longrightarrow} \alpha$.

Proof of (ii.6).

Obvious (otherwise all intermediate words would have to be of the length á priori bounded which contradicts the fact that K is infinite).

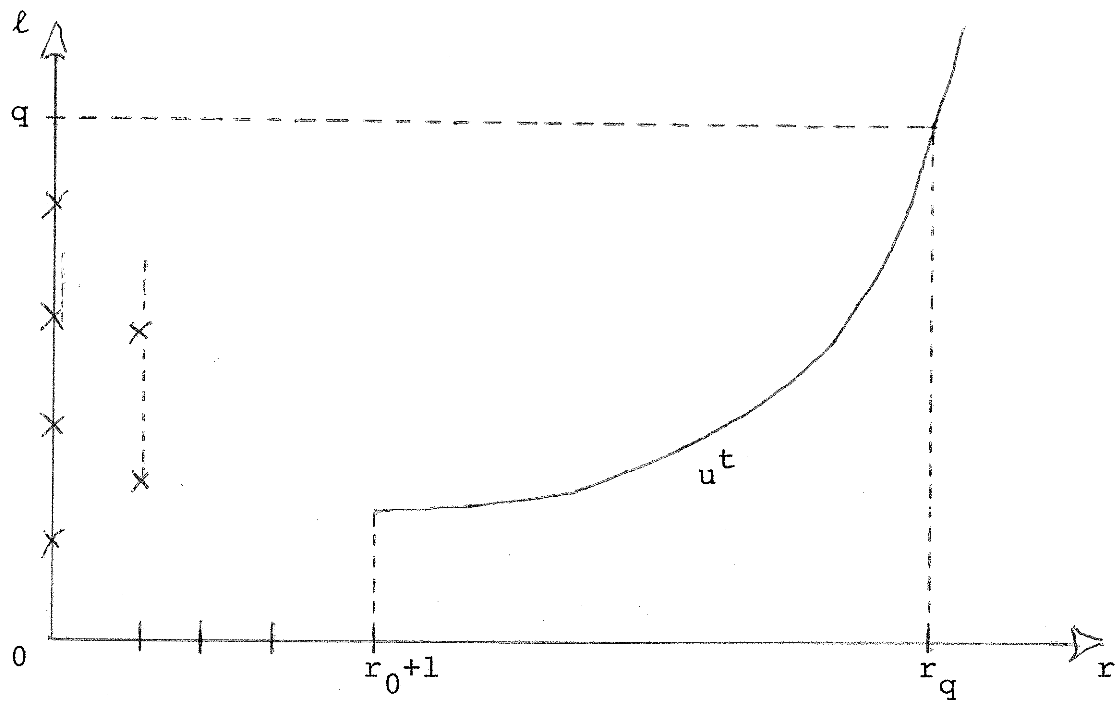
(ii.7) Now we complete the proof of the theorem as follows.

$\underline{\text{less}}_q K = U_1 + U_2$ where

U_1 are all the words from K of length not larger than q that are obtained by a derivation without a repetition which does not take more that r_0 steps, and

U_2 are all the words from K of length not larger that q that are obtained by a derivation without a repetition which takes more than r_0 steps.

The following graphic represents the situation:



where r is the number of steps (in derivations without repetitions) required to derive a word in K and ℓ is the length of a word in K (so the point (i,j) is on the graphic if in i steps one can derive a word from K of length j).

From (ii.6), (ii.2) and (ii.5) it follows that all the points (i,j) for $i > s_0$ are above the exponential line u^t for some constant u.

But then Lemma 2 implies that there exists a constant h_0 such that (note that $r_q = \log_\alpha q$)

$$\underline{less}_q K = U_1+U_2 \le h_0 s_0 + h_0 r_q = h_0 s_0 + h_0 \log_\alpha q.$$

However $\log_\alpha q = \frac{\log_2 q}{\log_2 \alpha}$ and so

$$\underline{less}_q K \le h_0 s_0 + h_0 \frac{\log_2 q}{\log_2 \alpha} \le C \log_2 q, \text{ for a suitable constant } C.$$

As a direct application of our main theorem we get for example the following result (which is used in [4]).

Corollary 1. Let G be an FP0L system such that L(G) contains $\{a^n b^n c^n : n > 1\}$. Then for no finite language F, $L(G) \setminus F$ is 3-balanced.

REFERENCES

[1]. Ehrenfeucht, A. and Rozenberg, G., 1975. On Θ-determined E0L languages, In: A. Lindenmayer and G. Rozenberg (eds) _Automata_, _Languages_, _Development_, North Holland, Amsterdam, 191-202.

[2]. Ehrenfeucht, A. and Rozenberg, G., 1976. On the structure of polynomially bounded D0L systems, _Fundamenta Informatica_, to appear.

[3]. Herman, G.T. and Rozenberg, G., 1975, _Developmental Systems and Languages_, North Holland Publishing Company, Amsterdam.

[4]. Maurer, H., Rozenberg,G., Salomaa, A. and Wood, D., On pure E0L forms, submitted for publication.

ACKNOWLEDGMENTS

The authors greatfully acknowledge the financial support by the NSF Grant Number MCS75-09972 and by IBM Belgium. They are also very much indebted to L. Fosdick for making the work on this and other papers possible.