# FPOL SYSTEMS GENERATING COUNTING LANGUAGES

by

A. Ehrenfeucht*

and

G. Rozenberg**

CU-CS-156-79 August, 1979

Note:

Report #156 is a revised version of report # CU-CS-113-77.

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

**G. Rozenberg, Institute of Applied Mathematics and Computer Science, University of Leiden, 2300 RA Leiden, Holland

All correspondence to G. Rozenberg

ABSTRACT

Counting languages are the languages of the form $\{a_1^n a_2^n \ldots a_t^n \mid t \geq 2, n \geq 1\}$ 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 FPOL 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 OL systems without erasing productions and with finite axiom sets (called FPOL 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. Those are languages of the form $\{a_1^n a_2^n \ldots a_t^n \mid t \geq 2, n \geq 1\}$ 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 FPOL 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 FPOL languages and we believe that this result together with its proof sheds some new light on the structure of derivations in FPOL 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 [3].

## 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., [2]). 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.

(3). If $\alpha$ is a word over $\Sigma$ then $\underline{\text{alph}}\ \alpha$ denotes the set of all letters from $\Sigma$ that occur in $\alpha$, $\underline{\text{pref}}_k\ (\alpha)$ denotes the prefix of $\alpha$ of the length k and $\underline{\text{suf}}_k\ (\alpha)$ denotes the suffix of $\alpha$ of the length k. $|\alpha|$ denotes the length of $\alpha$ and $\#_a\ \alpha$ denotes the number of occurrences of the letter a in $\alpha$.

(4). If K is a language then

$\underline{\text{alph}}\ K = \bigcup_{\alpha\ \varepsilon\ K} \underline{\text{alph}}\ \alpha$, $\text{ALPH}(K) = \{\underline{\text{alph}}\ \alpha \mid \alpha \varepsilon K\}$ and

$\underline{\text{less}}_q\ K = \#\{|\alpha| \mid \alpha \varepsilon K \text{ and } |\alpha| \leq q\}$.

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

To establish the basic notation for this paper we recall now the definition of an FPOL system.

Definition

(1). An FPOL <u>system</u> is a construct $G = (\Sigma, P, A)$ where $\Sigma$ is a finite nonempty alphabet, P is a finite set of <u>productions</u>, each of the form $a \to \alpha$ with $a \varepsilon \Sigma, \alpha \varepsilon \Sigma^+$ 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 $x = a_1 \ldots a_t$ and $y = \alpha_1 \ldots \alpha_t$ where $< a_1,\alpha_1 >, \ldots <a_t,\alpha_t > \in P$. 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 $L(G) = \{\alpha \in \Sigma^+ \mid (\exists w)_A [w \overset{*}{\Rightarrow} \alpha]\}$. □

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

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

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

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

(4). fin $G = \Sigma \backslash$ inf G; elements of 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 $G = \{m \in N^+ \mid (\exists \alpha)_{\Sigma+}[\alpha^m \in L(G)]\}$.

(7). The growth relation of G, denoted as $f_G$, is a function from $N^+$ into finite subsets of $N^+$ defined by $f_G(n) = \{|\alpha| \mid \alpha \in L(n,G)\}$.

(7.1). If there exists a polynomial $\Phi$ 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

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

then we can say that $f_G$ is limited.

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

then we can say that fG is deterministic. □

## III. AUXILIARY RESULTS

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

Definition. Let $\Sigma$ be a finite alphabet.

(1). Let $\alpha \in \Sigma^+$ and let t be a positive integer $t \geq 2$. A t-disjoint decomposition of $\alpha$ is a vector $(\alpha_1,\ldots,\alpha_t)$ such that $\alpha_1,\ldots,\alpha_t \in \Sigma^+$, $\alpha_1\ldots\alpha_t = \alpha$ and, for every i in $\{1,\ldots,t-1\}$, alph $\alpha_i \cap$ alph $\alpha_{i+1} = \emptyset$.

(2). Let $K \subseteq \Sigma^+$ and let t be a positive integer, $t > 2$. We say that K is t-balanced if there exist positive rational numbers $c_1,\ldots,c_t$ with $\sum_{i=1}^{t} c_i = 1$ and a positive integer d such that for every $\alpha$ in K there exists a t-disjoint decomposition $(\alpha_1,\ldots,\alpha_t)$ of $\alpha$ such that, for every $i \in \{1,\ldots,t\}$, $c_i \cdot |\alpha| - d \leq |\alpha_i| \leq c_i \cdot |\alpha| + d$. In such a case we also say that K is (v,d)-balanced and that $(\alpha_1,\ldots,\alpha_t)$ is a (v,d)-balanced decomposition of $\alpha$, where $v = (c_1,\ldots,c_t)$.

(3). An FPOL system G is t-balanced if L(G) is t-balanced. □

The following three lemmas describe the basic property of growth relations of t-balanced FPOL systems.

Lemma 1. If $G = (\Sigma,P,A)$ is a t-balanced FPOL system with $t \geq 3$, then there exists a positive integer $k_0$ such that, for every a in $\Sigma$ and for every positive integer n, $\#f_{G_a}(n) < k_0$.

Proof. Clearly it suffices to show that for every a in $\Sigma$ there exists a positive integer $k_a$ such that, for every positive integer n, $\#f_{G_a}(n) < k_a$.

Let $v = (c_1,\ldots,c_t)$ and $d$ be such that $L(G)$ is $(v,d)$-balanced. Let $c_{min} = \min\{c_1,\ldots,c_t\}$. If $a \in \Sigma$ then either $a \in \underline{inf}\ G$ or $a \in \underline{fin}\ G$. We will consider these cases separately.

(i). Let $a \in \underline{inf}\ G$.

In this case we will prove the result by contradiction. Thus let us assume that:

there does not exist a positive integer $k_a$ such that, for every positive integer $n$, $\#f_{G_a}(n) < k_a$. .....(*).

Then we proceed as follows.

(i.1). There exist a positive integer $n_0$, a positive integer $r$ larger than $\#\Sigma$ and words $w_1,\ldots,w_r$ in $L^{n_0}(G_a)$ such that, for every $i$ in $\{1,\ldots,t\}$ and for every $j$ in $\{1,\ldots,r-1\}$, $c_i|w_{j+1}| > c_i|w_j| + 2d$.

This is proved as follows.

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 $x_{i_{j+1}} - x_{i_j} > \frac{2d}{c_{min}}$.

If $r \le \#\Sigma$ then $s \le \#\Sigma \frac{2d}{c_{min}}$. Since $n$ was arbitrary, if we set $k_a$ equal to the smallest positive integer larger than $(\#\Sigma \frac{2d}{c_{min}}) + 1$ then we get that, for every positive integer $n$, $\#f_{G_a}(n) < k_a$, which contradicts (*).

(i.2). Let $\alpha = \alpha_1\ a\ \alpha_2$ be a word in $L(G)$ that is long enough, meaning that, for every $i \in \{1,\ldots,t\}$, $|\alpha|c_i > 3|w_r| + 5d$ where $w_1,\ldots,w_r$ is a

sequence (in the order of increasing length) from (i.1) for some fixed $n_0$ and r. Let $\beta_1 = \bar{\alpha}_1 w_1 \bar{\alpha}_2 \in L^{n_0}(G,\alpha)$,

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

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

where $\bar{\alpha}_1, \alpha_2$ are some fixed words such that $\bar{\alpha}_1 \in L^{n_0}(G,\alpha_1)$ and $\bar{\alpha}_2 \in L^{n_0}(G,\alpha_2)$.

Let, for each $i \in \{1,\ldots,r\}$, $(\beta_i[1],\ldots,\beta_i[t])$ be a (v,d)-balanced decomposition of $\beta_i$.

Since $|\beta_i| \geq |\alpha|$ and $t \geq 3$ the condition on the length of $\alpha$ assures us that either $w_i$ is contained in the word resulting from $\beta_i$ by cutting off its prefix $(\beta_i[1])(\underline{\text{pref}}_{|w_r|+2d}(\beta_i[2]))$ or $w_i$ is contained in the word resulting from $\beta_i$ by cutting off its suffix $(\underline{\text{suf}}_{|w_r|+2d}(\beta_i[t-1]))(\beta_i[t])$. Because these two cases are symmetric we assume the first one.

Since, for each $i \in \{1,\ldots,r-1\}$, $|w_{i+1}|-|w_i| > \frac{2d}{c_{min}}$, $|\beta_{i+1}|-|\beta_i| > \frac{2d}{c_{min}}$. Consequently $|\beta_{i+1}[1]|-|\beta_i[1]| > 0$ and so $\beta_{i+1}[1]$ results from $\beta_i[1]$ by catenating to $\beta_i[1]$ a nonempty prefix of $\beta_i[2]$. Also

$$|\beta_r[1]|-|\beta_1[1]| \leq (c_1.(|\bar{\alpha}_1\bar{\alpha}_2|+|w_r|)+d) - (c_1(|\bar{\alpha}_1\bar{\alpha}_2|+|w_1|) - d) = c_1(|w_r|-|w_1|)+2d \leq$$
$$\leq |w_r|+2d.$$

Thus 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 the occurrence of $w_1$ indicated by the equality $\beta_1 = \bar{\alpha}_1 w_1 \bar{\alpha}_2$. However $r > \#\Sigma$ and so at least two nonempty subwords used in the process of constructing $\beta_2[1], \beta_3[1],\ldots,\beta_r[1]$ contain an occurrence of the same letter. This implies that there exists a j in $\{2,\ldots,r-1\}$ such that $\underline{\text{alph}}(\beta_j[1]) \cap \underline{\text{alph}}(\beta_j[2] \neq \emptyset$ which contradicts the fact that $(\beta_j[1],\ldots,\beta_j[t])$ is a (v,d)-balanced decomposition of $\beta_j$.

integers m dividing all numbers $f_{\overline{G}}(n)$ provided that $n \geq n_m$ for suitably chosen m.

The lemma follows now by the following easy to prove property of DOL growth functions. Assume that a DOL growth function f not identically zero has the following property. For every positive integer m, there are integers $m_0 \geq m$ and $n_0$ such that $m_0$ divides f(n) wherever $n \geq n_0$. Then f is not of polynomial type. □

After we have established the basic properties of growth relations of t-balanced FPOL systems we move to investigate the structure of t-balanced FPOL systems the languages of which contain counting languages. Those counting languages are defined now.

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

To prove our main theorem we need the following transformation of an FPOL system.

Definition. Let $G = (\Sigma,P,A)$ be an FPOL system and k a positive integer. The k-decomposition of G is a set $\mathcal{G} = \{G_1,\ldots,G_k\}$ of FPOL systems (called components) such that, for every $i \in \{1,\ldots,k\}$, $G_i = (\Sigma,P^k,A_i)$ where $A_1 = A$ and $A_i = \{\alpha | \alpha \in L^{i-1}(G)\}$ for $i \in \{2,\ldots,k\}$, and $(a \to \alpha) \in P^k$ if and only if either $a \overset{k}{\underset{G}{\Rightarrow}} \alpha$ or $a \overset{m}{\Rightarrow} \Lambda$ for $m \leq k$ and $\alpha = \Lambda$. □

It follows directly from the above definition that $L(G) = \bigcup_{i=1}^{k} L(G_i)$ where $\mathcal{G} = \{G_1,\ldots,G_k\}$ is a k-decomposition of G.

A particular kind of decomposition will be useful for our purposes. It is defined as follows. Let $G = (\Sigma,h,A)$ be an FPOL system. We say that G is well-sliced if:

(1). for every a in $\Sigma$ and every $k,\ell \geq 1$, $ALPH(L^k(G_a)) = ALPH(L^\ell(G_a))$ and moreover if x is a word such that $|x| \geq 2$ and #alph x = 1 then $x \in L^k(G_a)$ if and only if there exists a word y such that $|y| \geq 2$, alph x = alph y and $y \in L^\ell(G_a)$,

(2). for every a in $\Sigma$ if $\bigcup_{n \geq 1} L^n(G_a)$ is finite then $\bigcup_{n \geq 1} L^n(G_a) = \{\alpha \mid a \Rightarrow \alpha\}$.

The proof of the following result is rather standard (see, e.g., [1]) and so it is omitted. (By a well-sliced decomposition of an FPOL system we understand a decomposition each component of which is well-sliced).

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

We are ready now to prove the main result of this paper.

Theorem 1. Let $t \geq 3$, M be a t-counting language, G be a t-balanced FPOL system and $K = M \cap L(G)$. There exists a constant C such that $\underline{less}_q K \leq C.\log_2 q$ for every positive integer q.

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

Since the result holds trivially when K is finite, let us assume that K is infinite.

(1). For every letter b in $\Delta$ there exists a multiple letter a and a

word $\alpha$ in $\{b\}^*$ such that $a \overset{+}{\Longrightarrow} \alpha$. This is obvious.

(2). If $a \in \underline{\text{mult}}\ G$, $b \in \Delta$, $\alpha \in \{b\}^+$ and $a \overset{+}{\Longrightarrow} \alpha$ then

(i). $f_{G_a}$ is either constant or exponential,

(ii). $f_{G_b}$ is either constant or exponential, and

(iii). $f_{G_a}$ is constant if and only if $f_{G_b}$ is constant.

We prove (2) as follows.

By Lemma 2, $f_{G_a}$ is deterministic and because G is well-sliced, for every positive integer n, $\ell \in f_{G_a}(n)$ if and only if $b^{\ell} \in L^n(G_a)$.

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

If $\tau$ contains infinitely many different words then $G_a$ satisfies the assumptions of Lemma 3 and so $f_{G_a}$ is exponential.

Otherwise, because G is well-sliced, $f_{G_a}$ is a constant function.

Thus (i) is proved. But a derives strings "through" b and so a and b must have the same type of growth. Consequently (i) implies (ii) and (iii).

(3). Either, for every b in $\Delta$, $f_{G_b}$ is a constant function, or, for every b in $\Delta$, $f_{G_b}$ is exponential.

This is proved as follows.

Let $b \in \Delta$. From (1) and (2) it follows that $f_{G_b}$ is either constant or exponential. Now let a be a neighbor of b (in M). Then if we take a word $\alpha$ from K of the form $\ldots a^n b^n \ldots$ (or symmetrically $\ldots b^n a^n \ldots$) and will derive in G words from it in such a way that each occurrence of b in $\alpha$ will produce the same subtree, then if b is not of the same type as a, we obtain a word $\beta$ in L(G) that is not t-balanced; a contradiction. Consequently any two neighbors in M must have the same type of growth and (3) holds.

(4). It is not true that $f_{G_a}$ is constant for every a in $\Delta$.

We prove it by showing that if $f_{G_a}$ if constant for every a in $\Delta$ then the fact that K is infinite leads to a contradiction.

Since K is infinite we can choose $\alpha$ in K which is arbitrarily long, e.g., so long that each derivation graph for $\alpha$ in G is such that on each path in it there exists a label that appears at least twice.

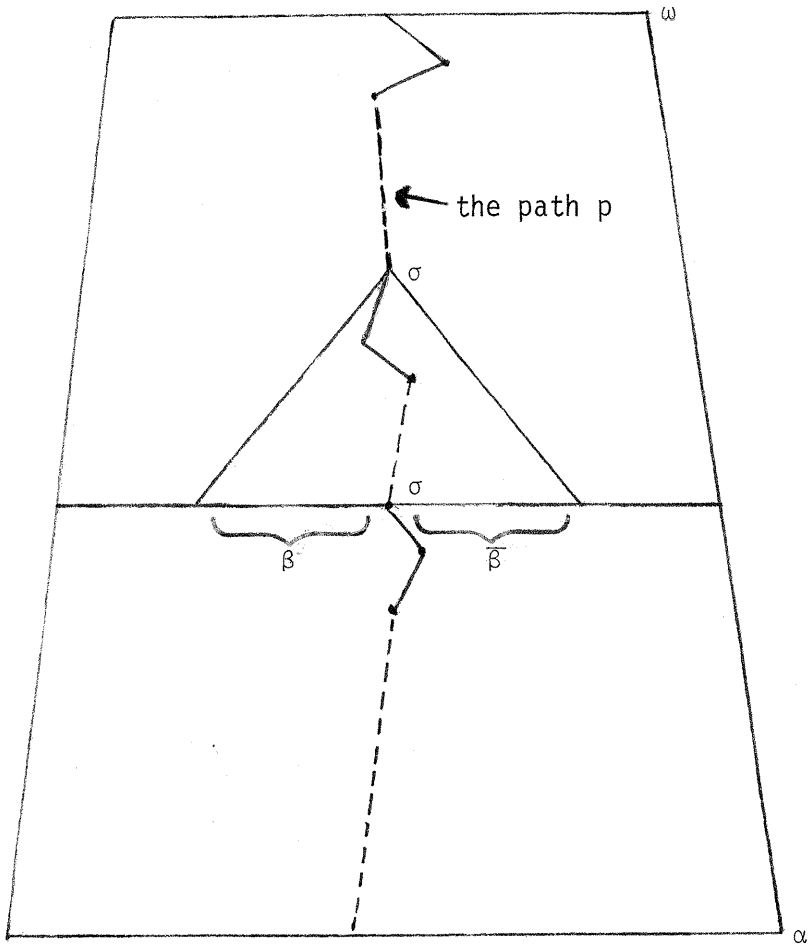
In a derivation graph corresponding to a derivation of $\alpha$ from $\omega$ in A we choose a path $p = e_0, e_1, \ldots$ as follows:

$e_0$ is an occurrence in $\omega$ such that no other occurrence in $\omega$ contributes a longer subward to $\alpha$,

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

Now on p we choose the first (from $e_0$) label $\sigma$ that repeats itself on p. Then we take the first repetition of $\sigma$ on p (and we let $\beta, \overline{\beta}$ to be the words such that the contribution of the first $\sigma$ on p to the level on which the first repetition of $\sigma$ p occurs is $\beta\ \sigma\ \overline{\beta}$ where the indicated occurrence of $\sigma$ is the occurrence of $\sigma$ on p).

The situation is illustrated by the following figure:




Now we proceed as follows.

(i). $\beta\bar{\beta}\neq\Lambda$.

We prove it by contradiction. To this aim assume that $\beta\bar{\beta}=\Lambda$.

(i.1). Then every label $\rho$ on p that repeats itself must be such that $\rho \overset{+}{\Longrightarrow} \delta\rho\bar{\delta}$ implies $\delta\bar{\delta} = \Lambda$.

This is seen as follows.

Since G is well-sliced, $\sigma \Rightarrow \sigma, \sigma \Longrightarrow \zeta\rho\bar{\zeta}$ and $\rho \Longrightarrow \mu\rho\mu$ for some words $\zeta, \bar{\zeta}, \mu, \bar{\mu}$ such that alph $\mu\bar{\mu}$ = alph $\delta\bar{\delta}$.

Then

$$\sigma \Rightarrow \zeta\rho\bar{\zeta} \Rightarrow \zeta^{(1)}{}_{\mu\rho\bar{\mu}\bar{\zeta}}{}^{(1)} \Rightarrow \zeta^{(2)}{}_{\mu}{}^{(1)}{}_{\mu\rho\bar{\mu}\bar{\mu}}{}^{(1)}{}_{\bar{\zeta}}{}^{(2)} \Rightarrow \ldots$$

$$\sigma \Rightarrow \sigma \Rightarrow \zeta\rho\bar{\zeta} \Rightarrow \zeta^{(1)}{}_{\zeta\rho\bar{\zeta}\bar{\zeta}}{}^{(1)} \Rightarrow \ldots$$

$$\sigma \Rightarrow \sigma \Rightarrow \sigma \Rightarrow \zeta\rho\bar{\zeta} \Rightarrow \ldots$$

$$\sigma \Rightarrow \sigma \Rightarrow \sigma \Rightarrow \sigma \Rightarrow \ldots$$

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

for some words $\zeta^{(1)}, \bar{\zeta}^{(1)}, \zeta^{(2)}, \bar{\zeta}^{(2)}, \ldots, \mu^{(1)}, \bar{\mu}^{(1)}, \mu^{(2)}, \bar{\mu}^{(2)}, \ldots$ where all the words $\mu^{(1)}\bar{\mu}^{(1)}, \mu^{(2)}\bar{\mu}^{(2)}, \ldots$ are nonempty if $\delta\bar{\delta}$ is nonempty. Consequently if $\delta\bar{\delta} \neq \Lambda$ then there exists a positive integer $\ell$, such that $\#f_{G_\sigma}(\ell) > k_0$, which contradicts Lemma 1 (where $k_0$ is the constant from the statement of Lemma 1).

Thus (i.1) holds.

But (i.1) implies that $\alpha$ cannot be longer than a fixed a priori constant; since $\alpha$ was an arbitrary word in K this contradicts the fact that K is infinite.

Thus indeed $\beta\bar{\beta} \neq \Lambda$ and (i) holds.

(ii). Since G is well-sliced, $\sigma \Rightarrow \gamma\sigma\bar{\gamma}$ for some words $\gamma, \bar{\gamma}$ such that $\underline{\text{alph}}\ \gamma\bar{\gamma} = \underline{\text{alph}}\ \beta\bar{\beta}$ and $\sigma \Rightarrow \pi$ for some $\pi \in \Delta^+$. Since we have assumed that $f_{G_a}$ is constant for every a in $\Delta$, $f_{G_\pi}$ is constant.

Then

$$\sigma \Rightarrow \pi \Rightarrow \pi^{(1)} \Rightarrow \pi^{(2)} \Rightarrow \pi^{(3)} \Rightarrow \ldots$$

$$\sigma \Rightarrow \gamma\sigma\bar{\gamma} \Rightarrow \gamma^{(1)}{}_{\pi\bar{\gamma}}{}^{(1)} \Rightarrow \gamma^{(2)}{}_{\pi}{}^{(1)}{}_{\bar{\gamma}}{}^{(2)} \Rightarrow \ldots$$

$$\sigma \Rightarrow \gamma\sigma\bar{\gamma} \Rightarrow \gamma(1)\ \gamma\sigma\bar{\gamma}\bar{\gamma}^{(1)} \Rightarrow \gamma^{(2)}{}_{\gamma}{}^{(1)}{}_{\pi\bar{\gamma}}{}^{(1)}{}_{\bar{\gamma}}{}^{(2)} \Rightarrow \ldots$$

$$\sigma \Rightarrow \gamma\sigma\bar{\gamma} \Rightarrow \gamma^{(1)}\ \gamma\sigma\bar{\gamma}\bar{\gamma}^{(1)} \Rightarrow \gamma^{(2)}{}_{\gamma}{}^{(1)}{}_{\gamma\sigma\bar{\gamma}\bar{\gamma}}{}^{(1)}{}_{\bar{\gamma}}{}^{(2)} \Rightarrow \gamma^{(3)}{}_{\gamma}{}^{(2)}{}_{\gamma}{}^{(1)}{}_{\pi\bar{\gamma}}{}^{(1)}{}_{\bar{\gamma}}{}^{(2)}{}_{\bar{\gamma}}{}^{(3)} \Rightarrow \ldots$$

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

where all $\gamma\bar{\gamma}, \gamma^{(1)}\bar{\gamma}^{(1)}, \ldots, \pi, \pi^{(1)}, \ldots$ are nonempty words.

Since $f_{G_\pi}$ is constant, the above implies that there exists a positive integer $\ell$ such that $\#f_{G_\sigma}(\ell) > k_0$ which contradicts Lemma 1 (where $k_0$ is the constant from the statement of Lemma 1).

Consequently it cannot be true that $f_{G_a}$ is constant for every a in $\Delta$, and so (4) holds.

(5). $f_{G_b}$ is exponential for every b in $\Delta$. This follows directly from (3) and (4).

(6). There exists a positive integer constant $s_0$ such that in every derivation without repetitions (in its trace) of a word from K, already after $s_0$ steps an intermediate word contains an occurrence of a multiple letter a for which there exist b in $\Delta$ and $\alpha$ in $\{b\}^+$ such that $a \overset{+}{\Rightarrow} \alpha$. This is obvious.

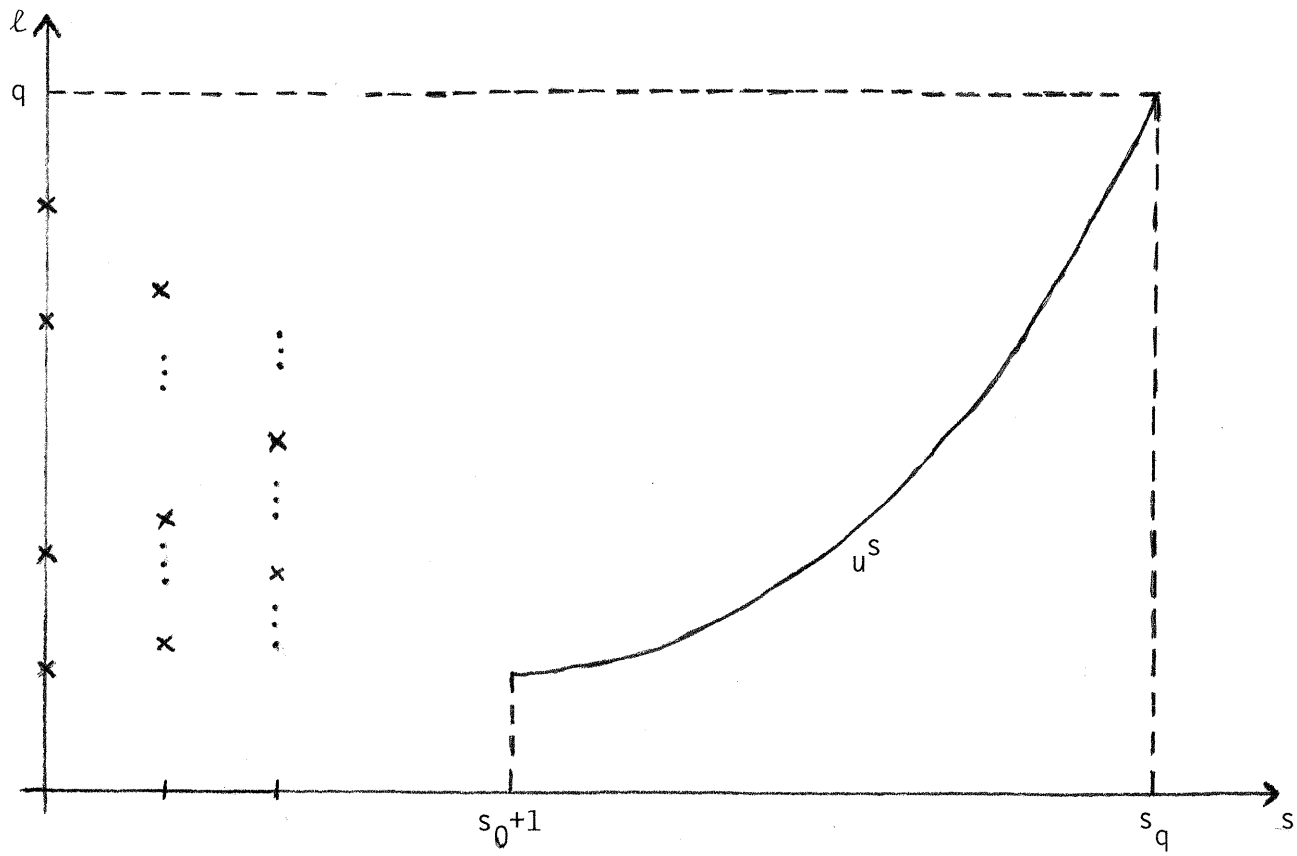
(7). Now we complete the proof of the theorem as follows.

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

$U_1$ is the number of 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 than $s_0$ steps, and

$U_2$ is the number of all the words from K of length not larger than q that are obtained by a derivation without a repetition which takes more than $s_0$ steps.

The following graphic represents the situation:




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

From (2), (5) and (6) it follows that for $i > s_0$ all the points (i,j) are above the exponential line $u^s$ for some constant $u > 1$. But then Lemma 1 implies that there exists a constant $h_0$ such that (note that $s_q = \log_u q$) $\underline{less}_q K \leq U_1+U_2 \leq h_0 s_0 + h_0 \log_u q$. Since $\log_u q = \frac{\log_2 q}{\log_2 u}$ $\underline{less}_q K \leq h_0 \cdot s_0 + h_0 \frac{\log_2 q}{\log_2 u} \leq C.\log_2 q$ for a suitable constant C.

Thus the theorem holds. □

As a corollary of the above theorem we get the following result which turns out to be useful in the theory of EOL forms (see [3]).

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

Proof. Directly from Theorem 1. □

REFERENCES

[1]. Ehrenfeucht, A. and Rozenberg, G., 1975. On $\Theta$-determined EOL languages, In: A. Lindenmayer and G. Rozenberg (eds.) Automata, Languages, Development, North Holland, Amsterdam, 191-202.

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

[3]. Maurer, H., Rozenberg, G., Salomaa, A. and Wood, D., On pure EOL forms, Revue Francaise d'Automatique, d'Informatique et de Recherche Opérationnelle, to appear.