

## The Number of Occurrences of Letters Versus Their Distribution in Some EOL Languages

A. EHRENFUCHT

*Department of Computer Science, University of Colorado,  
Boulder, Colorado 30302*

AND

G. ROZENBERG

*Department of Mathematics, Utrecht University, Utrecht-Uithof, The Netherlands*

A characterization theorem is given for a class of developmental languages. The theorem binds together the number of occurrences of letters in the words of the given language with the distribution of these letters.

### INTRODUCTION

This paper deals with a class of developmental languages. The theory of developmental systems and languages originated in the works of Lindenmayer (1968). This theory provided a useful theoretical framework within which the nature of cellular behavior in development can be discussed, computed and compared (see, e.g., Herman and Rozenberg, to appear, Lindenmayer, 1968, and Lindenmayer and Rozenberg, 1972). It turned out that developmental systems and languages are interesting and novel objects from the formal language theory point of view. Especially in comparison with Chomsky grammars and languages (see, e.g., Ginsburg, 1966) they provided a lot of insight into the basic problems of formal language theory.

An important subclass of developmental systems are the so-called EOL systems (see, e.g., Herman, 1974, or Herman, Lindenmayer, and Rozenberg), which were devised to allow descriptions of development which take into account the inaccuracy of our observations.

One of the basic open problems within the theory of EOL systems (and in fact within the whole theory of developmental systems) is the charac-

terization theorems that allow one, for example, to prove that some languages are not EOL languages (i.e., languages generated by the EOL systems).

This paper provides such a characterization for a subclass of EOL languages. The characterization theorem binds together the number of occurrences (in the words of the given EOL language) of letters from a given set of letters with the distribution of these letters.

The paper also discusses some applications of the main result.

## 1. PRELIMINARIES

We assume the reader to be familiar with the basics of formal language theory (see, e.g., Ginsburg, 1966, whose notation and terminology we shall mostly follow). In addition to this, we shall use the following notation:

(i)  $N$  denotes the set of nonnegative integers and  $N^+ = N - \{0\}$ . If  $n$  is an integer, then  $\text{abs}(n)$  denotes its absolute value.

(ii) If  $x$  is a word over an alphabet  $\Sigma$ , then  $|x|$  denotes the length of  $x$  and  $\text{Min}(x)$  denotes the set of letters which occur in  $x$ . For  $a$  in  $\Sigma$ ,  $\#_a(x)$  denotes the number of occurrences of the letter  $a$  in  $x$  and if  $B$  is a subset of  $\Sigma$  then  $\#_B(x) = \sum_{a \in B} \#_a(x)$ . If  $k$  is a positive integer, then  $x^k$  denotes  $x$  catenated  $k$  times with itself.

(iii) If  $A$  is a finite set, then  $\#A$  denotes its cardinality. If  $B \subseteq A$  and  $\#B = 1$  then  $B$  is called a *singleton in A*.

(iv) A *coding* is a letter to letter homomorphism. If  $h$  is a homomorphism from  $\Sigma^*$  into  $V^*$  and  $L \subseteq V^*$  then  $h^{-1}(L) = \{x \in \Sigma^* : h(x) = y \text{ for some } y \text{ in } L\}$ .

(v) If  $\tau = s_1, s_2, s_3, \dots$  is a sequence of objects and  $i_1, i_2, i_3, \dots$  are such that  $i_1 < i_2 < i_3 < \dots$ , then  $s_{i_1}, s_{i_2}, s_{i_3}, \dots$  is called a *subsequence of  $\tau$* .

(vi)  $\emptyset$  denotes the empty set and  $\Lambda$  denotes the empty word.

(vii) If  $\mathcal{A}$  is a (nondeterministic) finite automaton, then  $L(\mathcal{A})$  denotes its language.

(viii) If  $A$  is an ultimately periodic sequence (set) of nonnegative integers then  $\text{thres}(A)$  denotes the smallest integer  $j$  for which there exists a positive integer  $q$  such that, for every  $i \geq j$ , if  $i$  is in  $A$  then  $(i + q)$  is in  $A$ . The smallest positive integer  $p$  such that, for every  $i \geq \text{thres}(A)$ , whenever  $i$  is in  $A$  then also  $(i + p)$  is in  $A$ , is denoted by  $\text{per}(A)$ .

## 2. DEFINITIONS AND EXAMPLES OF EOL SYSTEMS

In this section we give basic definitions concerning developmental languages, which are relevant for this paper.

**DEFINITION 1.** An EOL system is a construct  $G = \langle V_N, V_T, P, \omega \rangle$  such that

$V_N$  is a finite alphabet (of *nonterminal letters* and *symbols*),

$V_T$  is a finite nonempty alphabet (of *terminal letters* or *symbols*), such that  $V_N \cap V_T = \emptyset$ ,

$\omega$  is an element of  $(V_N \cup V_T)^+$  (called the *axiom* of  $G$ ),

$P$  is a finite nonempty set (called the *set of productions* of  $G$ ) each element of which is of the form  $a \rightarrow \alpha$ , where the symbol " $\rightarrow$ " is not in  $V_N \cup V_T$ ,  $a$  is in  $V_N \cup V_T$ , and  $\alpha$  is in  $(V_N \cup V_T)^*$ . Moreover, for every  $a$  in  $V_N \cup V_T$  there exists a word  $\alpha$  in  $(V_N \cup V_T)^*$  such that  $a \rightarrow \alpha$  is in  $P$ .

In the sequel we shall often write " $a \rightarrow_P \alpha$ " rather than " $a \rightarrow \alpha$  is in  $P$ ." Also a production of the form  $a \rightarrow \alpha$  is called a *production for a in P*.

**DEFINITION 2.** An EOL system  $G = \langle V_N, V_T, P, \omega \rangle$  is called a 0L system if, and only if,  $V_N = \emptyset$ . (In this case we write  $G$  as  $\langle V_T, P, \omega \rangle$ ).

0L systems are investigated, for example, in Rozenberg and Doucet (1971).

**DEFINITION 3.** Let  $G = \langle V_N, V_T, P, \omega \rangle$  be an EOL system.

(i) Let  $x \in (V_N \cup V_T)^+$ , say  $x = b_1 \cdots b_t$  for some  $b_1, \dots, b_t$  in  $V_N \cup V_T$ , and let  $y \in (V_N \cup V_T)^*$ . We say that  $x$  *directly derives*  $y$  (in  $G$ ), denoted as  $x \Rightarrow_G y$ , if there exists a sequence  $\pi_1, \dots, \pi_t$  of productions from  $P$ , such that, for every  $i$  in  $\{1, \dots, t\}$ ,  $\pi_i = b_i \rightarrow \alpha_i$  and  $y = \alpha_1 \cdots \alpha_t$ .

(ii) As usual,  $\Rightarrow_G^+$  denotes the transitive closure of the relation  $\Rightarrow_G$  and  $\stackrel{*}{\Rightarrow}_G$  denotes the reflexive and transitive closure of the relation  $\Rightarrow_G$ . If  $x \stackrel{*}{\Rightarrow}_G y$  then we say that  $x$  *derives*  $y$  in  $G$ .

(iii) A finite sequence  $D = (x_0, x_1, \dots, x_r)$  of words from  $(V_N \cup V_T)^*$  such that,  $r \geq 1$  and, for each  $i$  in  $\{1, \dots, r\}$ ,  $x_{i-1} \Rightarrow_G x_i$ , is called a *derivation* (of  $x_r$  from  $x_0$ ) in  $G$ . If  $x_0 = \omega$ , then  $D$  is called a *derivation of  $x_r$  in  $G$* .

(iv) An infinite sequence  $D = (x_0, x_1, \dots)$  of words from  $(V_N \cup V_T)^*$  such that, for each  $i \geq 1$ ,  $x_{i-1} \Rightarrow_G x_i$ , is called an *infinite derivation* in  $G$ .

(v) If  $D = (x_0, x_1, \dots, x_r)$  is a derivation in  $G$ , then its *control sequence*

is any sequence  $\tau = (T_1, \dots, T_r)$  of subsets of  $P$ , such that, for each  $i$  in  $\{1, \dots, r\}$ ,  $x_{i-1} \Rightarrow_G x_i$  "using" all and only productions from  $T_i$ .

(vi) For  $x$  in  $(V_N \cup V_T)^+$ ,  $y$  in  $(V_N \cup V_T)^*$  and a positive integer  $r$  we write  $x \Rightarrow_G^r y$  if there exists a derivation  $D = (x_0 = x, x_1, \dots, x_r = y)$  in  $G$ . We also write  $x \Rightarrow_G^0 x$ , for every  $x$  in  $(V_N \cup V_T)^+$ .

(vii)  $\max(G)$  is defined as  $\max\{|\alpha| : a \Rightarrow_P \alpha \text{ for some } a \text{ in } V_N \cup V_T \text{ and } \alpha \text{ in } (V_N \cup V_T)^*\}$ .

DEFINITION 4. Let  $G = \langle V_N, V_T, P, \omega \rangle$  be an EOL system. The *language of  $G$* , denoted as  $L(G)$ , is defined by  $L(G) = \{x \in V_T^* : \omega \xrightarrow{*}_G x\}$ .

DEFINITION 5. A language  $K$  is called an EOL *language* (0L *language*) if, and only if, there exists an EOL *system* (0L *system*)  $G$  such that  $L(G) = K$ .

*Remark 1.* Given an EOL system  $G = \langle V_N, V_T, P, \omega \rangle$ , a derivation  $D = (x_0, \dots, x_r)$  and its control sequence  $\tau = (T_1, \dots, T_r)$ , the pair  $(D, \tau)$ , in general, does not tell us which productions are used to rewrite the particular occurrences of letters in the words  $x_0, \dots, x_{r-1}$ . However (to avoid cumbersome notation and to keep the size of this paper decent), we shall often assume that the pair  $(D, \tau)$  provides such information. This should not lead to confusion.

*Remark 2.* The properties we are interested in (in this paper) are trivial for finite languages, *hence we consider only infinite languages*. Thus in the sequel if we write "a language" (or "an EOL language") we mean an infinite one, unless explicitly stated otherwise. Also whenever we write "an EOL system" we mean one generating an infinite language.

*Remark 3.* Given an EOL system  $G = \langle V_N, V_T, P, \omega \rangle$  we shall sometimes consider  $P$  to be the "set of names for productions" rather than the set of productions itself. In this sense we can talk about the words over  $P$ , etc., and this should not lead to confusion.

We end this section with two examples of EOL systems.

EXAMPLE 1.  $G = \langle V_N, V_T, P, \omega \rangle$ , where  $V_N = \{S\}$ ,  $V_T = \{a, b\}$ ,  $P = \{S \rightarrow a, S \rightarrow b, a \rightarrow a^2, b \rightarrow b^2\}$  and  $\omega = S$ , is an EOL system such that  $L(G) = \{a^{2^n} : n \geq 0\} \cup \{b^{2^n} : n \geq 0\}$ .

EXAMPLE 2.  $G = \langle \Sigma, P, \omega \rangle$ , where  $\Sigma = \{a, b\}$ ,  $P = \{a \rightarrow (ab)^2, b \rightarrow \Lambda\}$  and  $\omega = ab$ , is a 0L system such that  $L(G) = \{(ab)^{2^n} : n \geq 0\}$ .

## 3. BASIC NOTIONS AND THEIR PROPERTIES

In this section we introduce basic notions describing the structure of 0L languages we are interested in and we prove some properties of these notions.

**DEFINITION 6.** Let  $L$  be a language over an alphabet  $\Sigma$  and let  $B$  be a nonempty subset of  $\Sigma$ . Let  $I_{L,B} = \{n \in \mathbb{N} : \text{there exists a word } w \text{ in } L \text{ such that } \#_B(w) = n\}$ .

(i)  $B$  is *numerically dispersed (in  $L$ )* if, and only if,  $I_{L,B}$  is infinite and for every positive integer  $k$  there exists a positive integer  $n_k$  such that, for every  $u_1, u_2$  in  $I_{L,B}$ , if  $u_1 \neq u_2$ ,  $u_1 > n_k$  and  $u_2 > n_k$  then  $\text{abs}(u_1 - u_2) > k$ .

(ii)  $B$  is *clustered (in  $L$ )* if, and only if,  $I_{L,B}$  is infinite and there exist positive integers  $k_1, k_2$  such that  $k_1 > 1, k_2 > 1$  and, for every word  $w$  in  $L$ , if  $\#_B(w) \geq k_1$ , then  $w$  contains at least two occurrences of symbols from  $B$  which are distant less than  $k_2$ .

**DEFINITION 7.** Let  $L$  be a language over an alphabet  $\Sigma$  and let  $a$  be in  $\Sigma$ . The symbol  $a$  is said to be *frequent (in  $L$ )* if, and only if, for every positive integer  $n$  there exists a word  $w$  in  $L$  such that  $\#_a(w) > n$ ; otherwise  $a$  is called *nonfrequent (in  $L$ )*.

**DEFINITION 8.** Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  and let  $a$  be in  $\Sigma$ .

(i) We define a *B-characteristic sequence of  $a$  (in  $G$ )*, denoted as  $\text{Seq}(G, B, a)$ , as an infinite sequence  $Z_1, Z_2, \dots$  of finite subsets of  $\mathbb{N}$  such that, for each  $i \geq 1$  and every nonnegative integer  $n$ ,  $n$  is in  $Z_i$  if, and only if,  $a \Rightarrow_G^i w$  for some  $w$  in  $\Sigma^*$  such that  $\#_B(w) = n$ .

(ii)  $\text{Seq}(G, B, a) = Z_1, Z_2, \dots$  is called *unique* if, and only if, for every  $i \geq 1, \#Z_i = 1$ .

(iii)  $\text{Seq}(G, B, a) = Z_1, Z_2, \dots$  is called *bounded* if, and only if, there exists a constant  $C$  such that, for every  $i \geq 1, n < C$  for every  $n$  in  $Z_i$ . In this case we also say that  $a$  is *B-bounded (in  $G$ )* and that  $C$  *bounds*  $\text{Seq}(G, B, a)$ . We say that  $a$  is *B-unbounded (in  $G$ )* otherwise.

(iv)  $\text{Seq}(G, B, a) = Z_1, Z_2, \dots$  is called *constant* if, and only if, for each  $i, j \geq 1, Z_i = Z_j$ . In this case we also say that  $a$  is *B-constant (in  $G$ )*.

**LEMMA 1.** Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  and let  $a$  be a symbol in  $\Sigma$ . Let  $\text{Seq}(G, B, a) = Z_1, Z_2, \dots$  and let

$U(G, B, a) = \{i_1, i_2, \dots\}$  be the set of positive integers such that, for every  $j \geq 1$ ,  $j$  is in  $U(G, B, a)$  if, and only if,  $Z_j \neq 0$ . Then  $U(G, B, a)$  is an ultimately periodic set.

*Proof.* Let  $G, B, a, \text{Seq}(G, B, a)$ , and  $U(G, B, a)$  be as in the statement of the lemma. Let  $\mathcal{A} = \langle Q, V, \delta, q_0, F \rangle$  be a finite automaton such that  $Q = \Sigma, V = P, q_0 = a, F = B$ , for every  $q, \bar{q}$  in  $Q$  and every  $v$  in  $V, \bar{q} \in \delta(q, v)$  if, and only if,  $v$  is a production of the form  $q \rightarrow \gamma_1 \bar{q} \gamma_2$  for some  $\gamma_1, \gamma_2$  in  $\Sigma^*$ .

We leave to the reader the easy proof of the fact that, for every  $j \geq 1, Z_j \neq \{0\}$  if, and only if, there exist a word  $y$  over  $V$  such that  $|y| = j$  and  $\delta(q_0, y) \cap F \neq \emptyset$ . Hence,  $U(G, B, a) = \{n \in \mathbb{N} : \text{there exists a word } y \text{ in } L(\mathcal{A}) \text{ such that } |y| = n\}$ .

But it is well known (see, e.g., Ginsburg, 1966, Theorems 2.1.2 and 2.1.3) that the set of lengths of a regular language is an ultimately periodic set and consequently Lemma 1 holds.

**DEFINITION 9.** Let  $G = \langle \Sigma, P, \omega \rangle$  and let  $B$  be a nonempty subset of  $\Sigma$ . A  $B$ -uniform period of  $G$ , denoted as  $m(G, B)$  is defined to be the smallest positive integer such that

- (i) for every  $b$  in  $\Sigma, m(G, B) > \text{thres}(U(G, B, b))$ , and
- (ii)  $m(G, B)$  is divisible by  $\text{per}(G, B, b)$ , for every  $b$  in  $\Sigma$ , such that  $U(G, B, b)$  is infinite.

**LEMMA 2.** Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $B$  be a nonempty subset of  $\Sigma$ . If  $B$  is numerically dispersed in  $L(G)$ , then, for every symbol  $a$  which is frequent in  $L(G)$ ,  $\text{Seq}(G, B, a)$  is unique.

*Proof.* Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $B$  be a numerically dispersed (in  $L(G)$ ) subset of  $\Sigma$ . Let  $a$  be a symbol from  $\Sigma$  which is frequent in  $L(G)$ .

Let us assume, to the contrary, that  $\text{Seq}(G, B, a) = Z_1, Z_2, \dots$  is not unique, meaning that, for some  $i_0 \geq 1, Z_{i_0}$  contains at least two nonnegative integers  $n_1$  and  $n_2$  (say  $n_2 > n_1$ , so that  $n_2 > 0$ ).

Now, let  $m$  be an arbitrary positive integer.

Let  $x$  be a word in  $L(G)$  which contains occurrences of the letter  $a$  for some  $t \geq m$  (recall that  $a$  is frequent in  $L(G)$  and so such a word  $x$  exists).

Let  $D = (x_0 = x, x_1, \dots, x_{i_0})$  be a derivation of some word  $x_{i_0}$  from the word  $x$  with a control sequence  $\tau = T_1, \dots, T_{i_0}$  which is such that each occurrence of  $a$  in  $x$  contributes  $n_2$  occurrences of symbols from  $B$  in  $x_{i_0}$  (recall that  $n_2 \in Z_{i_0}$ ).

Let  $\bar{D} = (x_0 = x, \bar{x}_1, \dots, \bar{x}_{i_0})$  be a derivation of some word  $\bar{x}_{i_0}$  from the word  $x$  with a control sequence  $\bar{\tau} = \bar{T}_1, \dots, \bar{T}_{i_0}$  which is such that exactly one occurrence of  $a$  in  $x$  contributes  $n_1$  occurrences of symbols from  $B$  in  $\bar{x}_{i_0}$  (recall that  $n_1 \in Z_{i_0}$ ) and all other occurrences in  $x$  of the letter  $a$  as well as all occurrences in  $x$  of all other letters “behave” in exactly the same way as in the derivation  $D$  with the control sequence  $\tau$ .

Thus  $\#_B(x_{i_0}) - \#_B(\bar{x}_{i_0}) = n_2 - n_1$ , where  $\#_B(x_{i_0}) \geq t \cdot n_2$ .

Consequently, there exists a positive integer  $k_0$  (put  $k_0 = n_2 - n_1 + 1$ ) such that for every integer  $m$  there exist integers  $u_1$  and  $u_2$  larger than  $m$  (put  $u_1 = tn_2 - (n_2 - n_1)$ ,  $u_2 = t \cdot n_2$  and notice that  $u_2$  and  $u_1$  are large enough if  $t$  is large enough) such that  $u_2 > u_1$  (note that  $u_2 - u_1 = n_2 - n_1$ ),  $u_1$  and  $u_2$  are in  $I_{L(G), B}$  and  $u_2 - u_1 < k_0$ . Thus  $B$  is not numerically dispersed; a contradiction.

Hence  $\text{Seq}(G, B, a)$  must be unique and Lemma 2 holds.

LEMMA 3. *Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $B$  be a subset of  $\Sigma$  which is numerically dispersed in  $L(G)$ . Let  $a$  be a symbol from  $\Sigma$  such that  $a$  is frequent in  $L(G)$  and  $a$  is  $B$ -unbounded in  $L(G)$ . Then, for every  $\alpha$  in  $\Sigma^*$ , if  $a \rightarrow_P \alpha$ , then  $\alpha$  must contain at least one occurrence of a  $B$ -unbounded letter.*

*Proof.* Let  $G, B$ , and  $a$  satisfy the statement of the lemma. Let  $\text{Seq}(G, B, a) = Z_1, Z_2, \dots$ .

Let  $a \rightarrow_P \alpha$  and let us assume, to the contrary, that either  $\alpha = A$  or  $\alpha \neq A$  and each letter which occurs in  $\alpha$  is  $B$ -bounded.

First, let us note that, if for every  $\gamma$  such that  $a \rightarrow_P \gamma$  we would have that either  $\gamma = A$  or  $\gamma \neq A$  and every letter which occurs in  $\gamma$  is  $B$ -bounded, then  $a$  itself would be  $B$ -bounded; a contradiction.

Thus for some  $\bar{\alpha}$  in  $\Sigma^+$  we have that  $a \rightarrow_P \bar{\alpha}$  and  $\bar{\alpha}$  contains an occurrence of a  $B$ -unbounded letter.

Let  $\bar{G} = \langle \Sigma, \bar{P}, \omega \rangle$  be a 0L system such that  $\bar{P}$  differs from  $P$  only in this that all productions for  $a$  that are different from  $a \rightarrow \alpha$  are deleted. Then, obviously,  $\text{Seq}(\bar{G}, B, a) = \bar{Z}_1, \bar{Z}_2, \dots$  is bounded (say for every  $i \geq 1$ , if  $n \in \bar{Z}_i$  then  $n < \bar{C}$  for some positive integer constant  $\bar{C}$  dependent on  $\bar{G}$  only).

As  $a$  is  $B$ -unbounded, for every positive integer  $C$  (in particular for  $\bar{C}$ ), there exists an integer  $i_C$  such that  $Z_{i_C}$  contains an integer larger than  $C$ . But, obviously, for each  $i \geq 1$ ,  $\bar{Z}_i \subseteq Z_i$ , and consequently, for some  $i_0 \geq 1$  (set  $i_0 = i_C$ ),  $Z_{i_0}$  must contain at least two different integers (one smaller than  $\bar{C}$  and another equal to or larger than  $\bar{C}$ ).

Consequently,  $\text{Seq}(G, B, a)$  is not unique which contradicts Lemma 2.

Thus, for every  $\alpha$  in  $\Sigma^*$ , if  $a \rightarrow_P \alpha$ , then  $\alpha$  must contain an occurrence of a  $B$ -unbounded letter. Hence Lemma 3 holds.

DEFINITION 10. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and  $n$  a positive integer such that  $n \geq 2$ . Let  $A(G, n) = \{x \in L(G) : \text{for some } i \text{ in } \{0, \dots, n - 1\}, \omega \Rightarrow_G^i x\}$ .

(i) The  $n$ -decomposition of  $G$ , denoted as  $\text{Dec}(G, n)$ , is the set of all 0L systems of the form  $\langle \Sigma, P^{(n)}, z \rangle$ , where  $z \in A(G, n)$  and  $P^{(n)} = \{a \rightarrow \alpha : a \Rightarrow_G^n \alpha\}$ . Each 0L system from  $\text{Dec}(G, n)$  is called a  $n$ -component of  $G$ .

(ii) A set  $\{G_1, \dots, G_p\}$  of 0L systems is called a decomposition of  $G$  if, and only if, for some  $n \geq 2$ ,  $\{G_1, \dots, G_p\} = \text{Dec}(G, n)$ .

We leave to the reader the obvious proof of the following result.

LEMMA 4. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be its decomposition. Let  $n$  be a positive integer such that  $n \geq 2$  and let, for  $i$  in  $\{1, \dots, p\}$ ,  $\mathcal{G}^{(i)} = \{G_1^{(i)}, \dots, G_p^{(i)}\}$  be the  $n$ -decomposition of  $G_i$ . Then  $\bigcup_{i=1}^p \mathcal{G}^{(i)}$  is a decomposition of  $G$ .

The following lemma states a number of properties that follow directly from Definitions 6, 7, 8 and 10. As the proof of these properties is straightforward, we leave it to the reader. Some of these properties are so obvious and useful that they will be used in the sequel without directly quoting them.

LEMMA 5. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and  $\{G_1, \dots, G_p\}$  be a decomposition of  $G$ .

(i)  $L(G) = \bigcup_{i=1}^p L(G_i)$ .

(ii) If  $B$  is a subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $L(G)$ , then, for every  $i$  in  $\{1, \dots, p\}$ , if  $I_{L(G), B}$  is infinite, then  $B$  is numerically dispersed in  $L(G_i)$ , and for at least one  $j$  in  $\{1, \dots, p\}$ ,  $B$  is numerically dispersed in  $L(G_j)$ .

(iii) If  $B$  is a subset of  $\Sigma$  such that, for every  $i$  in  $\{1, \dots, p\}$ ,  $B$  is either clustered in  $L(G_i)$  or  $I_{L(G), B}$  is finite, then  $B$  is clustered in  $L(G)$ .

(iv) If  $\{G_1, \dots, G_p\}$  is a  $n$ -decomposition of  $G$ ,  $B$  is a nonempty subset of  $\Sigma$  and  $a$  is in  $\Sigma$ , then, for each  $i$  in  $\{1, \dots, p\}$ ,  $\text{Seq}(G_i, B, a) = \bar{Z}_1, \bar{Z}_2, \dots$  with  $\bar{Z}_j = Z_{j \cdot n}$  for every  $j \geq 1$  (where  $\text{Seq}(G, B, a) = Z_1, Z_2, \dots$ ).

(v) If  $B$  is a nonempty subset of  $\Sigma$  and the letter  $a$  is  $B$ -bounded in  $G$ , then, for every  $i$  in  $\{1, \dots, p\}$ ,  $a$  is also  $B$ -bounded in  $G_i$ .



(vi) If the letter  $a$  is nonfrequent in  $L(G)$ , then, for each  $i$  in  $\{1, \dots, p\}$ ,  $a$  is also nonfrequent in  $L(G_i)$ .

(vii) If the letter  $a$  is frequent in  $L(G_i)$  for some  $i$  in  $\{1, \dots, p\}$ , then  $a$  is also frequent in  $L(G)$ .

(viii) If  $B$  is a nonempty subset of  $\Sigma$  and  $a$  is a symbol from  $\Sigma$  such that  $a$  is  $B$ -unique ( $B$ -constant) in  $G$ , then, for each  $i$  in  $\{1, \dots, p\}$ ,  $a$  is  $B$ -unique ( $B$ -constant) in  $G_i$ .

DEFINITION 11. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and  $\mathcal{G} = \{G_1, \dots, G_p\}$  be a decomposition of  $G$ . Let  $B$  be a nonempty subset of  $\Sigma$ .  $\mathcal{G}$  is called a  $B$ -fitted decomposition of  $G$  if, and only if, for every  $i$  in  $\{1, \dots, p\}$  and for every  $a$  in  $\Sigma$ , if  $a$  is frequent in  $L(G_i)$ , then

(i) If  $a$  is  $B$ -bounded in  $G_i$ , then  $a$  is  $B$ -constant in  $G_i$ , and

(ii) If  $a$  is  $B$ -unbounded in  $G_i$ , then, for every  $j \geq 1$ ,  $Z_j = \{z_j\}$  for some  $z_j \geq 2$  (where  $\text{Seq}(G_i, B, a) = Z_1, Z_2, \dots$ ).

The reader may easily notice that, Lemma 2 implies that if  $G$  is a 0L system,  $B$  is numerically dispersed in  $L(G)$ ,  $\mathcal{G}$  is a  $B$ -fitted decomposition of  $G$  and  $H$  is in  $\mathcal{G}$ , then if  $a$  is a symbol which is both, frequent in  $L(H)$  and  $B$ -bounded in  $H$ , then  $\text{Seq}(H, B, a) = \{z\}, \{z\}, \dots$  where  $z$  is a non-negative integer.

#### 4. THE EXISTENCE OF $B$ -FITTED DECOMPOSITIONS

In this section we prove that, for every 0L system  $G$  and for every  $B$  which is numerically dispersed in  $L(G)$ , there exists a  $B$ -fitted decomposition of  $G$ .

LEMMA 6. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $L(G)$ . Let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be a  $m(G, B)$ -decomposition of  $G$ . Let  $i$  be in  $\{1, \dots, p\}$  and let  $a$  be in  $\Sigma$ . If  $a$  is frequent in  $L(G_i)$  and  $\text{Seq}(G_i, B, a) = Z_1, Z_2, \dots$ , then either  $Z_j = \{0\}$  for every  $j \geq 1$ , or  $Z_j = \{z_j\}$  where  $z_j \neq 0$  for every  $j \geq 1$ .

*Proof.* Let  $G$ ,  $B$ , and  $\mathcal{G} = \{G_1, \dots, G_p\}$  be as in the statement of the lemma. Let  $i$  be in  $\{1, \dots, p\}$  and let  $a$  be a symbol from  $\Sigma$  such that  $a$  is frequent in  $L(G_i)$ .

By Lemma 2 and Lemma 5(vii),  $\text{Seq}(G, B, a)$  is unique and hence, by Lemma 5(viii),  $\text{Seq}(G_i, B, a)$  is unique.

By Lemma 1,  $U(G, B, a)$  is ultimately periodic. Hence, by Lemma 5(iv) and by the definition of  $m(G, B)$ , for every  $i$  in  $\{1, \dots, g\}$ , if  $\text{Seq}(G_i, B, a) = Z_1, Z_2, \dots$ , then, for every  $j \geq 1$ ,  $Z_j \neq \{0\}$  if, and only if,  $Z_1 \neq \{0\}$ .

Thus Lemma 6 holds.

LEMMA 7. *Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $G$ . Let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be the  $m(G, B)$ -decomposition of  $G$ . Let  $i$  be in  $\{1, \dots, p\}$  and let  $a$  be in  $\Sigma$ . If  $a$  is frequent in  $L(G_i)$  and  $a$  is  $B$ -bounded in  $G_i$ , then  $\text{Seq}(G_i, B, a)$  is an ultimately periodic sequence.*

*Proof.* Let  $G, B$  and  $\mathcal{G} = \{G_1, \dots, G_p\}$  satisfy the statement of the lemma. Let  $i$  be in  $\{1, \dots, p\}$  and let  $a$  be a symbol from  $\Sigma$  such that  $a$  is frequent in  $L(G)$  and  $a$  is  $B$ -bounded in  $G_i$ . Let  $\text{Seq}(G_i, B, a) = Z_1, Z_2, \dots$ .

By Lemma 6, either  $\text{Seq}(G_i, B, a) = \{0\}, \{0\}, \dots$  or  $\text{Seq}(G_i, B, a)$  consists of singletons different from  $\{0\}$  only.

If  $\text{Seq}(G_i, B, a) = \{0\}, \{0\}, \{0\}, \dots$  then obviously Lemma 7 holds.

Thus let us assume that  $\text{Seq}(G_i, B, a)$  consists of singletons different from  $\{0\}$  only.

Let  $D = (a = x_0, x_1, x_2, \dots)$  be an infinite derivation in  $G_i$  such that, for each  $i \geq 1$ ,  $x_i$  contains at least one occurrence of a letter frequent in  $L(G_i)$  such that its  $B$ -characteristic sequence consists of singletons different from  $\{0\}$  only. (Obviously such a derivation exists.) Let, for  $j \geq 1$ ,  $Y_j$  be a subset of all such letters from  $\text{Min}(x_j)$  that their  $B$ -characteristic sequences consist of singletons different from  $\{0\}$  only.

First, we shall prove that there exists a constant  $F$  such that, for every  $j \geq 1$ ,  $\#_{Y_j}(x_j) < F$ . Let us put  $F$  to be a positive integer constant such that, for every  $j \geq 1$  and every integer  $n$ , if  $n$  is in  $Z_j$  then  $n < F$  (recall that  $a$  is  $B$ -bounded). Let us assume to the contrary, that, for some  $j_0 \geq 1$ ,  $\#_{Y_{j_0}}(x_{j_0}) \geq F$ . Thus  $x_{j_0}$  has more than  $F$  occurrences of letters which are frequent in  $L(G_i)$  and  $B$ -characteristic sequences of which consist of singletons different from  $\{0\}$  only. (Recall Lemma 6 and the choice of  $Y_j$  for each  $j \geq 1$ .) Consequently, each such letter contributes at least one occurrence of an element from  $B$  to  $x_{j_0+1}$  (and, in fact, to each next word in  $D$ ), and so  $\#_B(x_{j_0+1}) > F$  which contradicts the fact that  $\text{Seq}(G_i, B, a)$  is bounded by  $F$ .

Thus our claim holds.

Now, for  $j \geq 1$ , let  $\bar{x}_j$  denote the word resulting by erasing from  $x_j$  of all occurrences of all letters in  $\text{Min}(x_j) - Y_j$ . Note that among all words of the form  $\bar{x}_j$ , for  $j \geq 1$ , there is only a finite number of different words (because none of these words is longer than  $F$ ). Hence for some  $j_1, j_2$  such

that  $j_2 > j_1$  we have  $\bar{x}_{j_2} = \bar{x}_{j_1}$ . But note that, for each  $j \geq 2$ ,  $\bar{x}_j$  "is contained" in the contribution to  $x_j$  from  $x_{j-1}$ . This is so, because (obviously) an occurrence in  $x_j$  of a letter the  $B$ -characteristic sequence of which consists only of singletons different from  $\{0\}$  may be contributed from an occurrence in  $x_{j-1}$  of a letter the  $B$ -characteristic sequence of which consists of singletons different from  $\{0\}$  only.

Thus for some  $j_1, j_2$  such that  $j_2 > j_1$  we have  $\bar{x}_{j_2} = \bar{x}_{j_1}$  and, for every  $g$  in  $\{j_1, j_1 + 1, \dots, j_2 - 1\}$ , we have  $\bar{x}_g = \bar{x}_{g+s(j_2-j_1)}$  for every  $s \geq 0$ .

But the "contributions to  $\text{Seq}(G_i, B, a)$ " from the words in the derivation  $D$ , depend (obviously) on the words  $\bar{x}_j$  (for  $j \geq 1$ ) only, and by the above, these contributions form an ultimately periodic sequence of numbers different from 0. However, (recall Lemma 6) the sequence  $\text{Seq}(G_i, B, a)$  consists of singletons only and so it is itself an ultimately periodic sequence.

Thus the lemma is proved.

LEMMA 8. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $L(G)$  and let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be a  $m(G, B)$ -decomposition of  $G$ . Let  $i \in \{1, \dots, p\}$  and let  $a$  be a symbol from  $\Sigma$  such that  $a$  is frequent in  $L(G_i)$  and  $a$  is  $B$ -unbounded in  $G_i$ . Let  $\text{Seq}(G_i, B, a) = Z_1, Z_2, \dots$ . Then for every positive integer  $C$  there exists a positive integer  $i_C$  such that, for every  $j \geq i_C$ ,  $Z_j = \{z_j\}$  where  $z_j > C$ .

*Proof.* Let  $G, B, \mathcal{G} = \{G_1, \dots, G_p\}, G_i, a$  and  $\text{Seq}(G_i, B, a)$  satisfy the statement of the lemma.

As  $a$  is frequent in  $L(G_i)$  and  $B$ -unbounded in  $G_i$ ,  $I_{L(G_i), B}$  is infinite. Then by Lemmata 2, 5(ii), 5(vii), and 5(viii),  $\text{Seq}(G_i, B, a)$  is unique, (say, for each  $l \geq 1$ ,  $Z_l = \{z_l\}$ ).

Let  $C$  be an arbitrary positive integer and let  $i_C$  be the smallest positive integer  $g$  such that  $z_g > C \cdot \max(G)$ .

Let  $D = (a, x_1, \dots, x_{i_C})$  be a derivation in  $G_i$  of a word  $x_{i_C}$  such that  $\#_B(x_{i_C}) = z_{i_C}$ . Let  $\tau$  be a control sequence of  $D$ . Now, each occurrence of a letter from  $B$  in  $x_{i_C}$  must be derived (in  $(D, \tau)$ ) from some occurrence in  $x_{i_C-1}$  of a letter whose  $B$ -characteristic sequence in  $G_i$  consists of singletons different from  $\{0\}$  only. Consequently (recall the choice of  $i_C$ )  $x_{i_C-1}$  must have more than  $C$  occurrences of letters whose  $B$ -characteristic sequences in  $G_i$  consist of singletons different from  $\{0\}$  only. Hence, one can "prolongate" the derivation  $D$  to an infinite derivation  $\bar{D} = (a, x_1, \dots, x_{i_C}, x_{i_C+1}, \dots)$  such that, for each  $g \geq i_C$ ,  $\#_B(x_g) > C$ .

But  $\text{Seq}(G_i, B, a)$  consists of singletons only, and so Lemma 8 holds.

The property stated in Lemma 8 carries over through decompositions

of OL systems in the following way. (We leave to the reader the obvious proof of the next result.)

LEMMA 9. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $L(G)$  and let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be a  $m(G, B)$ -decomposition of  $G$ . Let  $i \in \{1, \dots, p\}$ , let  $a$  be a symbol from  $\Sigma$  such that  $a$  is frequent in  $L(G_i)$  and  $a$  is  $B$ -unbounded in  $G_i$ . Let  $\mathcal{G}^{(i)} = \{G_1^{(i)}, \dots, G_{p_i}^{(i)}\}$  be a decomposition of  $G_i$  and let  $j \in \{1, \dots, p_i\}$ . Let  $\text{Seq}(G_j^{(i)}, B, a) = Z_1, Z_2, \dots$ . Then for every positive integer  $C$  there exists a positive integer  $i_C$  such that, for every  $j \geq i_C$ ,  $Z_j = \{z_j\}$ , where  $z_j > C$ .

DEFINITION 12. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $G$  and let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be the  $m(G, B)$ -decomposition of  $G$ . We define

(i)  $m_1(G, B)$  to be the smallest positive integer such that, for every  $i$  in  $\{1, \dots, p\}$  and for every  $a$  in  $\Sigma$  such that  $a$  is frequent in  $L(G_i)$  and  $B$ -bounded in  $G_i$ ,  $m_1(G, B)$  is divisible by  $\text{per}(\text{Seq}(G_i, B, a))$  and  $m_1(G, B) > \text{thres}(\text{Seq}(G_i, B, a))$ .

(ii)  $m_2(G, B)$  to be the smallest positive integer such that, for every  $i$  in  $\{1, \dots, p\}$  and for every  $a$  in  $\Sigma$  such that  $a$  is frequent in  $L(G_i)$  and  $a$  is  $B$ -unbounded in  $G_i$ ,  $m_2(G, B) > i_2$  where  $i_2$  is defined as in the statement of Lemma 8.

(iii)  $n(G, B) = m_1(G, B) \cdot m_2(G, B)$ .

LEMMA 10. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $G$  and let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be the  $(m(G, B) \cdot n(G, B))$ -decomposition of  $G$ . Let  $i \in \{1, \dots, p\}$  and let  $a$  be a letter in  $\Sigma$  such that  $a$  is frequent in  $L(G_i)$  and  $a$  is  $B$ -bounded in  $G_i$ . Then  $a$  is  $B$ -constant in  $L(G_i)$ .

*Proof.* This result follows directly from Lemmata 4, 5(ii), 5(vii), 5(viii), 7, and the definition of  $n(G, B)$ .

LEMMA 11. Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $G$ . Then there exists a  $B$ -fitted decomposition of  $G$ .

*Proof.* Let  $G$  and  $B$  satisfy the statement of the lemma. From Lemmata 5, 7, 8, 9, 10, and the definition of  $n(G, B)$  it follows that the  $(m(G, B) \cdot n(G, B))$ -decomposition of  $G$  is a  $B$ -fitted decomposition of  $G$ .

Thus Lemma 11 holds.

## 5. THE MAIN RESULT AND ITS APPLICATIONS

In this section we prove the main result of this paper which states that if  $L$  is an EOL language and  $B$  is numerically dispersed in  $L$ , then  $B$  is clustered in  $L$ . We also show some applications of this result.

**DEFINITION 13.** Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system and let  $B$  be a nonempty subset of  $\Sigma$ . A word  $x$  in  $L(G)$  is called *B-dispersed in  $G$*  if, and only if,  $\#_B(x) \geq 2$  and every two occurrences in  $x$  of symbols from  $B$  are of distance larger than  $\max(G)$ .

**LEMMA 12.** *Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  such that  $B$  is numerically dispersed in  $G$  and let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be a  $B$ -fitted decomposition of  $G$ . Let  $i \in \{1, \dots, p\}$  and let  $x$  be a word in  $L(G)$  such that  $x$  is  $B$ -dispersed in  $G_i$ . Then, if  $D = (x_0 = \omega, x_1, \dots, x_r = x)$  is a derivation of  $x$  in  $G_i$ , then, for every  $j$  in  $\{0, \dots, r-1\}$ ,  $x_j$  does not contain an occurrence of a letter which is frequent in  $L(G_i)$  and  $B$ -unbounded in  $G_i$ .*

*Proof.* Let  $G, B, G_i$  and  $x$  satisfy the statement of the lemma. Let us assume, to the contrary, that  $D = (x_0 = \omega, x_1, \dots, x_r = x)$  is a derivation of  $x$  in  $G_i$ , such that for some  $f$  in  $\{0, \dots, r-1\}$ ,  $x_f$  contains an occurrence of a letter, say  $c$ , which is frequent in  $L(G_i)$  and  $B$ -unbounded in  $G_i$ . But then (recall a definition of a  $B$ -fitted decomposition), for every  $j \geq 1$ ,  $Z_j = \{z_j\}$  for some  $z_j \geq 2$ , where  $\text{Seq}(G_i, B, c) = Z_1, Z_2, \dots$ .

Thus by Lemma 3,  $x_{r-1}$  contains an occurrence of a letter which is frequent in  $L(G_i)$  and  $B$ -unbounded in  $G_i$ , and this occurrence will contribute at least two occurrences in  $x_r$  of letters from  $B$ . These two occurrences, obviously, must be of distance smaller than  $\max(G)$  which contradicts the definition of  $x$ .

Thus Lemma 12 holds.

**LEMMA 13.** *Let  $G = \langle \Sigma, P, \omega \rangle$  be a 0L system, let  $B$  be a nonempty subset of  $\Sigma$  and let  $\mathcal{G} = \{G_1, \dots, G_p\}$ , be a  $B$ -fitted decomposition of  $G$ . Let  $i \in \{1, \dots, p\}$ . If for every positive integer  $n$  there exists a word  $y_n$  in  $L(G_i)$  such that  $\#_B(y_n) \geq n$  and  $y_n$  is  $B$ -dispersed in  $G_i$ , then  $B$  is not numerically dispersed in  $L(G_i)$ .*

*Proof.* Let  $G, B$  and  $\mathcal{G} = \{G_1, \dots, G_p\}$  satisfy the statement of the lemma. Let  $i \in \{1, \dots, p\}$  and let us assume that for every positive integer  $n$  there exists a word  $y_n$  in  $L(G_i)$  such that  $\#_B(y_n) \geq n$  and  $y_n$  is  $B$ -dispersed in  $G_i$ .

For each  $a$  in  $\Sigma$  such that  $a$  is nonfrequent in  $G_i$ , let  $\text{bound}(G_i, a)$  be the smallest positive integer larger than the maximal number of occurrences of  $a$  in any word in  $L(C_i)$ , and let  $t(a) = \max\{z: z \in Z_2^{(a)}\}$  where  $\text{Seq}(C_i, B, a) = Z_1^{(a)}, Z_2^{(a)}, \dots$ . Let  $F = \sum_{a \in N(G_i)} t(a) \cdot \text{bound}(G_i, a)$ , where  $N(G_i)$  denotes the set of all letters from  $\Sigma$  which are nonfrequent in  $G_i$ .

Let  $n$  be larger than  $(\max(G))^2 \cdot (\#_B(\omega))$  and let  $x$  be a word in  $L(G_i)$  such that  $\#_B(x) \geq n$  and  $x$  is  $B$ -dispersed in  $G_i$ .

Let  $D = (x_0 = \omega, x_1, \dots, x_r = x)$  be a derivation of  $x$  in  $G_i$  (note that by the choice of  $n$ ,  $r \geq 3$ ). Let  $\beta_0 = \#_B(x_0)$ ,  $\beta_1 = \#_B(x_1), \dots, \beta_r = \#_B(x_r)$ . (We assume also that we are given a control sequence  $\tau$  of  $D$ ). Let  $s$  be an integer in  $\{2, \dots, r\}$  such that  $\beta_r \leq \beta_{r-1} \leq \dots \leq \beta_s > \beta_{s-1}$  (such an  $s$  exists because of our choice of  $n$ ).

By Lemma 12,  $x_{s-2}$  does not contain occurrences of letters which are frequent in  $L(G_i)$  and  $B$ -unbounded in  $G_i$ . Consequently

$$\beta_{s-1} = U_1^{(s-1)} + U_2^{(s-1)} \quad \text{and} \quad \beta_s = U_1^{(s)} + U_2^{(s)}$$

where

$U_1^{(s-1)}$  is the number of such occurrences in  $x_{s-1}$  of letters from  $B$  which are contributed from occurrences in  $x_{s-2}$  of letters which are nonfrequent in  $L(G_i)$ ,

$U_1^{(s)}$  is the number of such occurrences in  $x_s$  of letters from  $B$  which are contributed from occurrences in  $x_{s-2}$  of letters which are nonfrequent in  $L(G_i)$ ,

$U_2^{(s-1)}$  is the number of such occurrences in  $x_{s-1}$  of letters from  $B$  which are contributed from occurrences in  $x_{s-2}$  of letters which are frequent in  $L(G_i)$  and  $B$ -bounded in  $G_i$ , and

$U_2^{(s)}$  is the number of such occurrences in  $x_s$  of letters from  $B$  which are contributed from occurrences in  $x_{s-2}$  of such letters which are frequent in  $L(G_i)$  and  $B$ -bounded in  $G_i$ .

Because  $\mathcal{G}$  is  $B$ -fitted,  $U_2^{(s-1)} = U_2^{(s)}$ , and by the definition of  $F$ ,  $U_1^{(s-1)} < F$ . Hence  $\beta_s - \beta_{s-1} = U_1^{(s)} - U_1^{(s-1)} < F$ .

Thus we have proved (recall that  $\beta_s \geq \beta_r$ ) that for  $n$  "large enough" if  $x$  is a word in  $L(G_i)$  such that  $\#_B(x) \geq n$  and  $x$  is  $B$ -dispersed in  $G_i$ , then there exists a word  $\bar{x}$  in  $L(G_i)$  (set  $\bar{x} = x_{s-1}$ ) such that  $\#_B(x) - \#_B(\bar{x}) < F$  where  $F$  is a constant dependent on  $G_i$  only.

Consequently  $B$  is not numerically dispersed in  $G_i$  and Lemma 13 holds.

LEMMA 14. *Let  $K$  be a 6L language over an alphabet  $\Sigma$  and let  $B$  be a nonempty subset of  $\Sigma$ . If  $B$  is numerically dispersed in  $K$ , then  $B$  is clustered in  $K$ .*

*Proof.* Let  $K$  and  $B$  satisfy the statement of the lemma. Let  $K = L(G)$  where  $G = \langle \Sigma, P, \omega \rangle$  is a 0L system. Let  $\mathcal{G} = \{G_1, \dots, G_p\}$  be a  $B$ -fitted decomposition of  $G$  (its existence is guaranteed by Lemma 11). By Lemma 5(ii), for each  $i$  in  $\{1, \dots, p\}$ , either  $B$  is numerically dispersed in  $L(G_i)$  or  $I_{L(G_i), B}$  is finite, and for every  $j$  in  $\{1, \dots, p\}$ , if  $I_{L(G_j), B}$  is infinite, then  $B$  is numerically dispersed in  $L(G_j)$ .

Let  $s$  be in  $\{1, \dots, p\}$  and let  $I_{L(G_s), B}$  be infinite. So  $B$  is numerically dispersed in  $L(G_s)$ . If we assume that  $B$  is not clustered in  $G_s$ , then for every positive integer larger than 1 there exists a word  $y_n$  in  $L(G_s)$  such that  $\#_B(y_n) \geq n$  and  $y_n$  is  $B$ -dispersed in  $G_s$ . But then, by Lemma 13,  $B$  is not numerically dispersed in  $L(G_s)$ ; a contradiction.

Thus, for every  $i$  in  $\{1, \dots, p\}$ , either  $I_{L(G_i), B}$  is finite or  $B$  is clustered in  $L(G_i)$ .

But then, by Lemma 5(iii),  $B$  must also be clustered in  $L(G)$  and so if  $B$  is numerically dispersed in  $K$  then  $B$  is clustered in  $G$ , which proves the lemma.

The following result, which was proved in Ehrenfeucht and Rozenberg (to appear) turns out to be a very useful one for this paper.

LEMMA 15. *For every EOL language  $L$  there exist a 0L language  $K$  and a coding  $\psi$  such that  $\psi(K) = L$ .*

We leave to the reader the easy proof of our next result.

LEMMA 16. *Let  $K$  and  $L$  be languages (over alphabets  $\Sigma$  and  $V$ , respectively) and let  $\psi$  be a coding such that  $\psi(K) = L$ . Then*

- (i) *If  $B$  is a subset of  $V$  such that  $B$  is numerically dispersed in  $L$ , then  $\psi^{-1}(B)$  is numerically dispersed in  $K$ , and*
- (ii) *If  $U$  is a subset of  $\Sigma$  such that  $U$  is clustered in  $K$ , then  $\psi(U)$  is clustered in  $L$ .*

THEOREM. *Let  $K$  be an EOL language over an alphabet  $\Sigma$  and let  $B$  be a nonempty subset of  $\Sigma$ . If  $B$  is numerically dispersed in  $K$ , then  $B$  is clustered in  $K$ .*

*Proof.* This result follows directly from Lemmata 14–16. It should be clear to the reader that the Theorem may be used to prove that a considerable number of languages are not EOL languages. This, by itself, fills in a gap in the developmental systems theory (for a discussion, see Herman, 1974, or Herman and Rozenberg, to appear).

We shall present now two examples of such application.

EXAMPLE 3 (See Herman, 1974). The language  $L = \{x \in \{a, b\}^*: \#_a(x) = 2^n \text{ for some } n \geq 0\}$  is not an EOL language. This is so because  $\{a\}$  is numerically dispersed in  $L$ , but, at the same time,  $\{a\}$  is not clustered in  $L$ .

EXAMPLE 4. Let  $\psi$  be a function from  $N$  to  $N$  such that, for every  $n$  in  $N$ ,  $\psi(n) \geq n$ . Let  $L_\psi = \{y_0 a y_1 a \cdots y_{r-1} a y_r : r = 2^n \text{ for some } n \geq 0 \text{ and, for each } i \text{ in } \{0, \dots, r\}, y_i \in \{b, c\}^* \text{ and } |y_i| = \psi(r)\}$ . Then  $L_\psi$  is not an EOL language.

Finally, let us discuss the "structural" character of the theorem. The theorem states the structural rather than numerical characterization of the subclass of EOL languages. This statement is not precise but it may be illustrated by the following. Whereas it was proved in Example 3 that the language  $L = \{x \in \{a, b\}^*: \#_a(x) = 2^n \text{ for some } n \geq 0\}$  is not an EOL language, the language  $\bar{L} = \{a^{2^n} b^m : n, m \geq 0\}$  is generated by the EOL grammar  $\langle V_N, V_T, P, \omega \rangle$  such that  $V_N = \{S\}$ ,  $V_T = \{a, b\}$ ,  $P = \{S \rightarrow a, S \rightarrow A, S \rightarrow Sb, a \rightarrow a^2, b \rightarrow b\}$ ,  $\omega = S$ . But  $\bar{L}$  results from  $L$  by an appropriate permutation of occurrences of letters in the words of  $L$ . Consequently, all "numerical characteristics" (such as the set of lengths, the number of occurrences of particular symbols, etc.) are the same for both languages and one of them is an EOL language, while the other is not an EOL language.

RECEIVED: October 12, 1973; REVISED: April 12, 1974

#### REFERENCES

- EHRENFEUCHT, A. AND ROZENBERG, G. (to appear), The equality of EOL languages and codings of OL languages, *Internat. J. Comput. Math.*
- GINSBURG, S. (1966), "The Mathematical Theory of Context-free Languages," McGraw-Hill, New York.
- HERMAN, G. (1974), Closure properties of some families of languages associated with biological systems, *Inform. Contr.* **24**, 101-121.
- HERMAN, G. AND ROZENBERG, G. (to appear), "Developmental Systems and Languages," North-Holland Publ. Company, Amsterdam.
- LINDENMAYER, A. (1968), Mathematical models for cellular interactions in development, *J. Theoret. Biol.* **18**, 300-315.
- LINDENMAYER, A. AND ROZENBERG, G. (1972), Developmental systems and languages, *Proc. of the 4th ACM Symp. on Theory of Comp.*
- ROZENBERG, G. AND DOUCET, P. (1971), On OL languages, *Inform. Contr.* **19**, 302-318.