A COMBINATORIAL PROPERTY OF EOL LANGUAGES

by

A. Ehrenfeucht*, G. Rozenberg**, and
R. Verraedt***

CU-CS-228-82 August, 1982

*Department of Computer Science, University of Colorado, Boulder, Colorado 80309, USA

**Institute of Applied Mathematics and Computer Science, University of Leiden, Leiden, The Netherlands

***Research Assistant of the National Fund for Scientific Research (Belgium), Department of Mathematics, University of Antwerp, U.I.A., B-2610 Wilrijk, Belgium

All correspondence to second author.

ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE NATIONAL SCIENCE FOUNDATION.

THIS MATERIAL IS BASED UPON WORK SUPPORTED IN PART BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT NO. MCS 79-03838.

ABSTRACT

Let Δ be an alphabet and Π its nontrivial binary partition. Each word over Δ can uniquely be decomposed in subwords (called blocks) consisting of letters of Π_i only, $i \in \{1,2\}$. Let $K \subseteq \Delta^*$. K has a LB-property (with respect to Π) if there exists a function $f : \mathbb{N}^+ \to \mathbb{N}^+$ such that for every $w \in K$ and every positive integer m the number of blocks of lenght at most m in w is bounded by $f(m)$. K has a CB-property (with respect to Π) if there exists a positive integer n_0 and a growing function $g : \mathbb{N}^+ \to \mathbb{N}^+$ such that for every $w \in K$ and every positive integer m the blocks of length at most m can be covered by at most n_0 segments of length at most $g(m)$.

It is proved that a CB-property always implies a LB-property but not necessarily otherway around. It is proved that an EOL language has a LB-property if and only if it has a CB-property.

INTRODUCTION

A study of combinatorial properties of languages in various language classes constitutes an active and important research area within formal language theory. A typical result here is of the form : if K is a language of type X, then $\mathcal{P}(K)$ where $\mathcal{P}$ is a combinatorial property of K. Such a property can be expressed directly , as e.g. in all kinds of pumping theorems, or indirectly (conditionally) getting then the following form : if K is a language of type X and $\mathcal{P}_1(K)$, then $\mathcal{P}_2(K)$ where $\mathcal{P}_1$, $\mathcal{P}_2$ are combinatorial properties.

This paper is concerned with a combinatorial property of the letter-type concerning EOL languages. Let Δ be an alphabet and Π its nontrivial binary partition. Each word over Δ can uniquely be decomposed in subwords (called blocks) consisting of letters of Π_i only, $i \in \{1, 2\}$. Let $K \subseteq \Delta^*$. K has a LB-property (with respect to Π) if there exists a function $f : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ such that for every $w \in K$ and every positive integer m the number of blocks of lenght at most m in w is bounded by $f(m)$. K has a CB-property (with respect to Π) if there exists a positive integer n_0 and a growing function $g : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ such that for every $w \in K$ and every positive integer m the blocks of length at most m can be covered by at most n_0 segments of length at most $g(m)$.

A CB-property always implies a *LB*-property but not necessarily otherway around. It is proved that an EOL language has a LB-property if and only if it has a CB-property. This result is proved by the "in depth" analysis of derivations in EOL systems and in this way we believe that this paper contributes to our understanding of the nature of derivations in EOL systems. Also we provide some applications of our main result. The first of this yields an example of a language which is an ETOL but not an EOL language. The second example is given gramatically (using the grammatical mechanism of the so-called regular pattern grammars, see e.g. [KR1]) . We prove this language to be not an EOL language, which allows one to prove an important strict inclusion in [KR2] ; we can do this without knowing precisely the form of strings belonging to this language.

1. PRELIMINARIES

We assume the reader to be familiar with the basic theory of EOL systems and languages, e.g. in the scope of [RS]. In this section we recall some basic terminology concerning EOL systems, fixing in this way the notation for our paper. Also, some new notions are introduced.

For a finite set X, $\#X$ denotes the number of elements of X. $\mathbb{N}$ denotes the set of nonnegative integers and $\mathbb{N}^+$ denotes the set of positive integers. For a finite subset X of $\mathbb{N}$, $\min X$ and $\max X$ denote the minimum and maximum of X respectively. An alphabet is a finite nonempty set of symbols. $f : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ denotes a total function with domain $\mathbb{N}^+$ and range $\mathbb{N}^+$; f is called *growing* if for every $n \in \mathbb{N}^+$, $f(n) \geq n$.

Let Δ be an alphabet. Λ denotes the empty word. For a word $w \in \Delta^*$, $|w|$ denotes its length and $\operatorname{alph} x$ denotes the set of letters occurring in x. For an alphabet Θ, $\#_\Theta\, x$ denotes the number of occurrences of letters from Θ in x. For a nonnegative integer i, $w(i)$ denotes the i-th letter of w if $1 \leq i \leq |w|$ and $w(i) = \Lambda$ otherwise. If x is nonempty , then $\operatorname{last} x$ denotes $x(|x|)$. Let $w = uv$ where $u, v \in \Delta^*$. Then u is a *prefix of* w and v is a *suffix of* w; we write $u \operatorname{pref} w$ and $v \operatorname{suf} w$ respectively.

For a language K, $K^0 = \{\Lambda\}$ and for a nonnegative integer n, $K^{n+1} = K^n \cdot K$. For a nonnegative integer n, $K^{\leq n} = \cup_{i=0}^{n} K^i$.

An E0L *system* will be denoted as $G = (\Sigma, P, \omega, \Delta)$ where Σ is its *total alphabet*, Δ is its *terminal alphabet*, $\omega \in \Sigma^+$ is its *axiom* ,and P is the *set of productions*. In the notation of an EOL system we use a production set instead of a finite substitution since this seems to be more plausible for the purpose of this paper.

If $\alpha \in \Sigma$ and $\alpha \rightarrow x$ belongs to P then $\alpha \rightarrow x$ is called an *α-production of* G. The fact that $\alpha \rightarrow x$ belongs to P is often abbreviated as $\alpha \underset{P}{\rightarrow} x$.

If $\alpha \in \Sigma$ and G is as above then $G_\alpha = (\Sigma, P, \alpha, \Delta)$.

Since the problems concerned in this paper become trivial otherwise we consider infinite EOL systems only (i.e. EOL systems which generate infinite languages), unless explicitly clear otherwise.

Let $G = (\Sigma, P, \omega, \Delta)$ be an EOL system.

(1) A letter $\alpha \in \Sigma$ is called *recursive* if $\alpha \overset{+}{\underset{G}{\Rightarrow}} u\alpha v$, $uv \in \Sigma^*$ The set of recursive letters of G is denoted by $\operatorname{rec} G$.

(2) G is called *propagating* if $\alpha \underset{P}{\rightarrow} x$ implies $x \neq \Lambda$. In this case we say that G is an EPOL *system*.

(3) G is called *synchronized* if for every $\alpha \in \Delta$, $\alpha \overset{+}{\underset{G}{\Rightarrow}} x$ implies $x \notin \Delta^*$. Without loss of generality we can assume that if G is a synchronized EOL system then there exists a symbol $F \in \Sigma - \Delta$, the *synchronization symbol* of G ,such that $F \rightarrow F$ is

the only F-production of G and for each $\alpha \in \Delta$, $\alpha \to F$ is the only α-production of G. In the rest of this paper whenever we consider a synchronized EOL system it will be assumed that its synchronization symbol equals F.

If G is synchronized , then we use $\text{us}\, G$ to denote $\Sigma - (\Delta \cup \{S, F\})$.

(4) G is called *standard* if the following conditions hold:

(i) $\omega = S \in \Sigma - \Delta$.

(ii) G is propagating and synchronized;

(iii) for each $\alpha \in \Sigma$, $\alpha \underset{P}{\to} x$ implies $S \notin \text{alph}\, x$;

(iv) for each $\alpha \in \text{us}\, G$, $S \overset{*}{\underset{G}{\Rightarrow}} u\alpha v$ for some $uv \in (\text{us}\, G)^*$, $\alpha \underset{G}{\Rightarrow} x$ for some $x \in (\text{us}\, G)^+$ and $\alpha \underset{G}{\Rightarrow} y$ for some $y \in \Delta^+$.

Let $G = (\Sigma, P, \omega, \Delta)$ be an EOL system and let ℓ be a positive integer. Let us recall that a *derivation in G (of length ℓ leading from $x \in V^*$ to $y \in V^*$)* is a sequence $(x = x_0, x_1, \ldots, x_\ell = y)$, such that $x_0 \underset{G}{\Rightarrow} x_1$, $x_1 \underset{G}{\Rightarrow} x_2$, ..., $x_{\ell-1} \underset{G}{\Rightarrow} x_\ell$ together with a precise description of how all the occurrences in x_i are rewritten to obtain x_{i+1} for $0 \leq i \leq \ell - 1$. Such a description can be formalized (see, e.g. [RS]). For the purpose of this paper it is suffices to depict a derivation D by

$$D : x_0 \underset{G}{\Rightarrow} x_1 \underset{G}{\Rightarrow} x_2 \underset{G}{\Rightarrow} \ldots \underset{G}{\Rightarrow} x_\ell .$$

A derivation in G leading from ω to $x \in \Sigma^*$ is called a *successful derivation in G*.

To each derivation there corresponds a derivation tree ; if a derivation tree of G corresponds to a successful derivation in G, then it is called a *successful derivation tree (in G)*.

If $\Sigma_1 \subseteq \Sigma$ and T is a derivation tree of G whose nodes are labelled by elements of Σ_1, then T is called a *Σ_1-labelled derivation tree of G*.

In addition to the rather standard notation and terminology concerning derivation trees we will also use the following.

For a tree T, $\text{height}\, T$ denotes its *height*.

For a node v of a derivation tree we will use $\ell(v)$ to denote the *label of v*.

Let G be an EPOL system and let T be a derivation tree in G of height ℓ. Then for $0 \leq i \leq \ell$, $\text{set}_i\, T$ denotes the set of nodes whose distance to the root equals i, and $\text{result}_i\, T$ denotes the word which results from the sequence of all nodes (ordered from left to right) from $\text{set}_i\, T$ by replacing each node by its label. Whenever we omit the index i in the above notation, it is assumed that i equals $\text{height}\, T$.

2.LONG BLOCK AND CLUSTERED BLOCK PROPERTY

In this section we define two combinatorial properties of languages forming the subject of investigation of this paper : a long block property and a clustered block property. We need a number of auxilliary notions first.

Definition 2.1 Let Δ be an alphabet and let $w \in \Delta^*$. A *segment of* w is a construct (u, k, ℓ) where $u \in \Delta^+$, $k, \ell \in \mathbb{N}$, $1 \leq k \leq \ell \leq |w|$ and $u = w(k)w(k+1)\ldots w(\ell)$. The set of segments of w is denoted by $\mathrm{SEG}(w)$. ∎

In the sequel the usual terminology concerning words will also be used for segments (e.g. ,the length of a segment (u, k, ℓ) is defined as $|u|$); however, this should not lead to confusion.

Definition 2.2 Let $X, Y \subseteq \mathrm{SEG}(w)$. We say that X *covers* Y if for every $(u, k, \ell) \in Y$ there exists a segment $(u', k', \ell') \in X$ such that $k' \leq k$ and $\ell' \geq \ell$. ∎

Definition 2.3 Let Δ be an alphabet and let $\Pi = (\Delta_1, \Delta_2)$ be a binary partition of Δ (i.e. Δ_1, $\Delta_2 \neq \emptyset$, $\Delta_1 \cup \Delta_2 = \Delta$ and $\Delta_1 \cap \Delta_2 = \emptyset$).

Then a *block of* w (*with respect to* Π) is a construct $(u, k, \ell) \in \mathrm{SEG}(w)$ such that either $u \in \Delta_1^+$ and $w(k-1)$, $w(\ell+1) \notin \Delta_1$ or $u \in \Delta_2^+$and $w(k-1)$, $w(\ell+1) \notin \Delta_2$.

The set of all blocks of w (with respect to Π) is denoted $\mathrm{BL}_\Pi(w)$ ($\mathrm{BL}(w)$ if Π is understood).

For a positive integer m we also denote

$$\mathrm{BL}_\Pi^m(w) = \{(u, k, \ell) \in \mathrm{BL}_\Pi(w) \mid |u| \leq m\}.$$

∎

We are now ready to state the definitions of a long block property and a clustered block property.

Definition 2.4 Let $K \subseteq \Delta^*$ and let Π be a binary partition of Δ. Then K has a *long block property* (*with respect to* Π), written $K \in \mathrm{LB}(\Pi)$ or $K \in \mathrm{LB}$ if Π is understood if there exists a function $f : \mathbb{N}^+ \to \mathbb{N}^+$ such that for every $w \in K$ and every positive integer m, $\#\mathrm{BL}_\Pi^m(w) \leq f(m)$. We also say that $K \in \mathrm{LB}(\Pi)$ *with parameter* f. ∎

Definition 2.5 Let $K \subseteq \Delta^*$ and let Π be a binary partition of Δ. Then K has a *clustered block property* (*with respect to* Π), written $K \in \mathrm{CB}(\Pi)$ or $K \in \mathrm{CB}$ if Π is understood, if there exists a positive integer n_0 and a growing function $g : \mathbb{N}^+ \to \mathbb{N}^+$ such that for every $w \in K$ and every positive integer m there exists a $X \subseteq \mathrm{SEG}(w)$ such that the following conditions hold:

(i) $\#X \leq n_0$;

(ii) for every $z \in X$, $|z| \leq g(m)$;
(iii) X covers $\mathrm{BL}_{\Pi}^{m}(w)$.

We also say that $K \in \mathrm{CB}(\Pi)$ *with parameters* n_0 *and* g. ∎

The following example illustrates the above definitions

Example 2.1 Let $\Delta = \{a, b\}$, $\Pi = (\{a\}, \{b\})$ and let

$$K = \{aba^2b^2a^3b^3 \ldots a^nb^n \mid n \geq 1\}.$$

Then $K \in \mathrm{LB}(\Pi) \cap \mathrm{CB}(\Pi)$.

Proof Clearly for every $w \in K$ and every positive integer m, $\# \mathrm{BL}^m(w) \leq 2m$. Thus, if we define $f : \mathbb{N}^+ \to \mathbb{N}^+$ by $f(m) = 2m$, then $K \in \mathrm{LB}$ with parameter f.

Also if we define $g : \mathbb{N}^+ \to \mathbb{N}^+$ by $g(m) = 2(1 + 2 + \cdots + m) = m(m+1)$, then $K \in \mathrm{CB}$ with parameters 1 and g. This can be seen as follows . For $w = ab \ldots a^kb^k \in K$ take

$$X = \{(ab \ldots a^\ell b^\ell, 1, \ell(\ell + 1))\}$$

where $\ell = \min\{k, m\}$. ∎

We investigate now the relationship between LB and CB properties.

First of all we demonstrate that a language which has a clustered block property with respect to Π, always has a long block property with respect to Π.

Theorem 2.1 Let $K \subseteq \Delta^*$ and let Π be a binary partition of Δ. Then $K \in \mathrm{CB}(\Pi)$ implies $K \in \mathrm{LB}(\Pi)$.

Proof Assume that $K \in \mathrm{CB}(\Pi)$ with parameters n_0 and g. Define $f : \mathbb{N}^+ \to \mathbb{N}^+$ by $f(m) = n_0 \cdot g(m)$. We claim that $K \in \mathrm{LB}(\Pi)$ with parameter f. This is proved as follows. Let $w \in K$, let m be a positive integer and let $X \subseteq \mathrm{SEG}(w)$ be such that it satisfies (i) through (iii) of Definition 2.5. Then the total length of all segments of X is not longer than $\# X$ times the maximal length of a segment in X, i.e. $n_0 \cdot g(m)$. Thus — because the length of a block is always positive — $\# \mathrm{BL}_{\Pi}^{m}(w) \leq n_0 \cdot g(m) = f(m)$ and consequently $K \in \mathrm{LB}(\Pi)$ with parameter f. ∎

On the other hand, the following example demonstrates that the converse of Theorem 2.1 does not hold.

Example 2.2 Let $\Delta = \{a, b\}$, $\Pi = (\{a\}, \{b\})$ and let

$$K = \{a^{i_1}b^{j_1}a^{i_2}b^{j_2} \ldots a^{i_n}b^{j_n} : n \geq 1 \text{ and } i_\ell, j_\ell \geq \ell \text{ for } 1 \leq \ell \leq n\}.$$

Then $K \in \mathrm{LB}(\Pi) \setminus \mathrm{CB}(\Pi)$.

Proof Obviously $K \in \mathrm{LB}(\Pi)$ with parameter $f : \mathbb{N}^+ \to \mathbb{N}^+$ where $f(m) = 2m$ for every positive integer m. The fact that $K \notin \mathrm{CB}(\Pi)$ is proved by contradiction. Assume that $K \in \mathrm{CB}(\Pi)$ with parameters n_0 and g. Consider $m = n_0 + 1$ and

$$w = ab^{g(n_0+1)+1}a^2b^{g(n_0+1)+2}\dots a^{n_0+1}b^{g(n_0+1)+n_0+1}.$$

Obviously every element of $\mathrm{BL}_{\Pi}^{n_0+1}$ consists of a's only and $\#\,\mathrm{BL}_{\Pi}^{n_0+1}(w) = n_0+1$. Now let $X \subseteq \mathrm{SEG}(w)$ satisfy conditions (i) through (iii) of Definition 2.5 for $m = n_0 + 1$ and w. Since for every $z \in X$, $|z| \leq g(n_0 + 1)$, every $z \in X$ can cover at most one element of $\mathrm{BL}_{\Pi}^{m}(w)$. Consequently $\#\,\mathrm{BL}_{\Pi}^{n_0+1}(w) \leq n_0$; a contradiction. ∎

In the following theorem we state the obvious fact that if for a language there exists a positive integer which limites the number of blocks in every word then this language has a long block property and a clustered block property.

The following result is obvious and hence given without a proof.

Theorem 2.2 Let $K \subseteq \Delta^*$ and let Π be a binary partition of Δ. If there exists a positive integer n such that for each $w \in K$, $\#\,\mathrm{BL}_{\Pi}(w) \leq n$, then $K \in \mathrm{CB}(\Pi)$ (and hence $K \in \mathrm{LB}(\Pi)$). ∎

We proceed to investigate operations on languages which preserve the CB-property.

Lemma 2.1 Let K_1 , $K_2 \subseteq \Delta^*$ and let Π be a binary partition of Δ. If $K_1, K_2 \in \mathrm{CB}(\Pi)$ and $K \subseteq K_1 \cup K_2$, then $K \in \mathrm{CB}(\Pi)$.

Proof Let K_1, K_2, Π be as in the statement of the lemma. Let $K_1 \in \mathrm{CB}(\Pi)$ with parameters n_1 and g_1, and let $K_2 \in \mathrm{CB}(\Pi)$ with parameters n_2 and g_2. Then clearly $K \in \mathrm{CB}(\Pi)$ with parameters $\max\{n_1, n_2\}$ and g where $g(n) = \max\{g_1(n), g_2(n)\}$ for a positive integer n. ∎

Definition 2.6 Let $X \subseteq \mathrm{SEG}(w)$. X is *disjoint* if for every (u_1, k_1, ℓ_1), $(u_2, k_2, \ell_2) \in X$ either $k_2 > \ell_1 + 1$ or $k_1 > \ell_2 + 1$.

The *join of X*, denoted $\mathrm{JOIN}(X)$, is defined as follows.

(i) $\mathrm{JOIN}(X) \subseteq \mathrm{SEG}(w)$, $\mathrm{JOIN}(X)$ is disjoint and covers X.

(ii) For every disjoint $Y \subseteq \mathrm{SEG}(w)$ which covers X, Y also covers $\mathrm{JOIN}(X)$. ∎

Thus given $X \subseteq \mathrm{SEG}(w)$ its join is obtained by combining into one segment those segments that either overlap or touch each other.

Lemma 2.2 Let $K_1 \subseteq \Delta^*$ and let Π be a binary partition of Δ. Let k be a positive integer. If $K_1 \in \mathrm{CB}(\Pi)$ and $K \subseteq K_1^k$, then $K \in \mathrm{CB}(\Pi)$.

Proof Let K_1, k, Π be as in the statement of the lemma. Let $K_1 \in \mathrm{CB}(\Pi)$ with parameters n_1 and g_1. We will demonstrate now that $K \in \mathrm{CB}(\Pi)$ with parameters $k \cdot n_1$ and g where, for all positive integers n, $g(n) = k \cdot n_1 \cdot g_1(n)$ for a positive integer n.

Let $w \in K$ and let m be a positive integer. Then $w = w_1 w_2 \ldots w_k$ where $w_i \in K_1$ for $1 \leq i \leq k$. Since $K_1 \in \mathrm{CB}(\Pi)$ for each w_i, $1 \leq i \leq k$,there exists an $X_i \subseteq \mathrm{SEG}(w_i)$ such that :

(i) $\# X_i \leq n_1$;
(ii) for every $z \in X_i$, $|z| \leq g_1(m)$; and
(iii) X_i covers $\mathrm{BL}_{\Pi}^{m}(w_i)$.

For $1 \leq i \leq k$ let

$$X'_i = \{(u, k + \sum_{j=1}^{i-1} |w_j|, \ell + \sum_{j=1}^{i-1} |w_j|) \mid (u, k, \ell) \in X_i\}.$$

Let $X = \mathrm{JOIN}(\cup_{i=1}^{k} X'_i)$.

Then the following conditions hold:

(i) $\# X \leq \#(\cup_{i=1}^{k} X_i) = \sum_{i=1}^{k} \# X_i \leq k \cdot n_1$,
(ii) for every $z \in X$,

$$|z| \leq \#(\cup_{i=1}^{k} X_i) \cdot \max\{|w| \mid w \in \cup_{i=1}^{m} X_i\} \leq k \cdot n_1 \cdot g_1(m)$$

and

(iii) X covers $\mathrm{BL}_{\Pi}^{m}(w)$.

To see that (iii) holds consider a block $u \in \mathrm{BL}_{\Pi}^{m}(w)$. Thus either $u = u_j u_{j+1} \ldots u_{j+s}$, $1 \leq j \leq j + s \leq k$ where u_j suf w_j , u_{j+1} pref w_{j+s} and $u_{j+\ell} = w_{j+\ell}$ for $j < \ell < j + s$, or u is a subword of u_j, $1 \leq j \leq k$.

Each of those u_i 's, $j \leq i \leq j + s$ is covered by a segment from X_i and so u is covered by a segment of $\mathrm{JOIN}(\cup_{i=1}^{k} X_i)$. Thus $K \in \mathrm{CB}(\Pi)$ with parameters $k \cdot n_1$ and g. ∎

Lemma 2.1 and Lemma 2.2 yield the following theorem.

Theorem 2.3 Let k, ℓ be positive integers. Let $K_1, \ldots, K_\ell \subseteq \Delta^*$ and let Π be a binary partition of Δ such that $K_1, \ldots, K_\ell \in \mathrm{CB}(\Pi)$. If $K \subseteq (K_1 \cup \ldots \cup K_\ell)^{\leq k}$, then $K \in \mathrm{CB}(\Pi)$. ∎

We end this section by a result which states that in the study of "block properties" we can restrict ourselves to the study of languages over a two letter alphabet.

Theorem 2.4 Let $K \subseteq \Delta^*$ and let $\Pi = (\Delta_1, \Delta_2)$ be a binary partition of Δ. Let h be the homomorphism on Δ^* defined by $h(\alpha) = a$ for $\alpha \in \Delta_1$, and $h(\alpha) = b$ for $\alpha \in \Delta_2$, where a, b are two fixed different letters. Then

(1) $K \in \mathrm{LB}(\Pi)$ if and only if $h(K) \in \mathrm{LB}(\{a\}, \{b\})$, and
(2) $K \in \mathrm{CB}(\Pi)$ if and only if $h(K) \in \mathrm{CB}(\{a\}, \{b\})$.

Proof (1) Obvious.

(2) If $K \in \mathrm{CB}(\Pi)$, then $h(K) \in \mathrm{CB}(\{a\}, \{b\})$. This can be seen as follows. If n_0 , g are parameters proving that $K \in \mathrm{CB}(\Pi)$, then the same n_0, g will prove that $h(K) \in \mathrm{CB}(\{a\}, \{b\})$; for a word $h(w)$ we consider covering by $h(X)$.

If $h(K) \in \mathrm{CB}(\{a\}, \{b\})$, then $K \in \mathrm{CB}(\Pi)$. This can be seen as follows. If n_0, g are parameters proving that $h(K) \in \mathrm{CB}(\{a\}, \{b\})$ then the same n_0,g will prove that $K \in \mathrm{CB}(\Pi)$. Let $w \in K$ and let m be a positive integer. Consider $h(w)$ and $X \subseteq \mathrm{SEG}(h(w))$ such that conditions (i) through (iii) from Definition 2.5 are satisfied for $h(w)$. Let

$$X' = \{(u, k, \ell) \in \mathrm{SEG}(w) \mid (h(u), k, \ell) \in X\}.$$

Clearly X' satisfies conditions (i) through (iii) from Definition 2.5 for w. Since w and m were arbitrary, $K \in \mathrm{CB}(\Pi)$. ∎

3. BASIC LETTER TYPES

The aim of this paper is to prove that if $K \subseteq \Delta^*$ is an EOL language and Π is a binary partition of Δ such that $K \in \mathrm{LB}(\Pi)$, then $K \in \mathrm{CB}(\Pi)$.

In view of Theorem 2.4 it suffices to consider $\Delta = \{a, b\}$ and $\Pi = (\{a\}, \{b\})$. Therefore in the rest of this paper, unless explicitly stated otherwise, we assume that $\Delta = \{a, b\}$, $\Pi = (\{a\}, \{b\})$, $K \subseteq \{a, b\}^*$ is an EOL language and $K \in \mathrm{LB}(\Pi)$. Let $G = (\Sigma, P, S, \Delta)$ be an EOL system generating K. Without loss of generality we can assume that G is standard (see the proof of Theorem II.2.2 in [RS]).

To prove that $K \in \mathrm{CB}$ we proceed as follows. Observe that

$$K = L(G) \subseteq \left(\left(\bigcup_{\alpha \in \mathrm{us}\, G} L(G_\alpha) \right) \cup M_0 \right)^{\mathrm{maxr}\, G}$$

where M_0 is a finite language (equal to all words of K which can be derived in one step from S). Thus in view of Theorem 2.2 and Theorem 2.3 it suffices to prove that for each $\alpha \in \mathrm{us}\, G$, $L(G_\alpha) \in \mathrm{CB}$. To this aim letters of $\mathrm{us}\, G$ are divided into various categories. First of all we need a subdivision of the words of K.

Definition 3.1 A nonempty word $w \in K$ is

- *of type* 1 if $\# \mathrm{BL}(w) = 1$,
- *of type* 2 if $\# \mathrm{BL}(w) = 2$, and
- *of type* 3 if $\# \mathrm{BL}(w) \geq 3$.

Note that type 1 words are words of the form either a^n or b^n, $n > 0$ (the former are referred to as type $1a$ and the latter as type $1b$). Type 2 words are words of the form either $a^n b^m$ or $b^n a^m$, $n, m > 0$ (the former are referred to as type $2a$ and the latter as type $2b$. ∎

Now letters of $\mathrm{us}\, G$ can be divided into various types.

Definition 3.2 For a letter $\alpha \in \mathrm{us}\, G$,

$$\mathrm{type}\, \alpha = \{x \in \{1a, 1b, 2a, 2b, 3\} \mid \alpha \overset{+}{\underset{G}{\Rightarrow}} w \text{ and } w \text{ is of type } x\}.$$

G is called *promissing* if for every $\alpha \in \mathrm{us}\, G$, $\# \mathrm{type}\, \alpha = 1$.

Then if $\mathrm{type}\, \alpha = \{x\}$, α is said to be a *letter of type* x. Furthermore , for a derivation tree of G, each node labelled by a letter of type x is called a *node of type* x.

If G is promissing then for $x \in \{1a, 1b, 2a, 2b, 3\}$,

$$\mathrm{type}\, x = \{\alpha \mid \mathrm{type}\, \alpha = \{x\}\}.$$

∎

Lemma 3.1 There exists an EOL system $G = (\Sigma, P, S, \{a, b\})$ such that $L(G) = K$ and G satisfies the following conditions

(1) G is standard.

(2) G is promissing.

(3) For every alphabet $\Theta \subseteq \mathrm{us}\, G$ and every $\alpha \in \mathrm{us}\, G$ either for all $n > 0$, there exists an $x \in (\mathrm{us}\, G)^*$ such that $\alpha \overset{n}{\underset{G}{\Rightarrow}} x$ and $\#_\Theta\, x \geq n$ or there exists a $n_0 > 0$ such that for every positive integer n, $\alpha \overset{n}{\underset{G}{\Rightarrow}} x$ for some $x \in (\mathrm{us}\, G)^*$ where $\#_\Theta\, x \leq n_0$.

(4) If $\alpha \in \mathrm{us}\, G \cap \mathrm{rec}\, G$, then $\alpha \underset{G}{\Rightarrow} u\alpha v$ for some $uv \in \Sigma^*$.

(5) Let $\alpha \in \mathrm{us}\, G \cap \mathrm{type}\, 3$ be such that $\mathrm{alph}\, x \cap \mathrm{rec}\, G \cap \mathrm{type}\, 3 = \emptyset$ whenever $\alpha \overset{+}{\underset{G}{\Rightarrow}} x$. Then $\alpha \underset{P}{\Rightarrow} y$ and $\mathrm{alph}\, y \subseteq \mathrm{us}\, G$ imply $\mathrm{alph}\, y \subseteq \mathrm{type}\, 1 \cup \mathrm{type}\, 2$. ∎

For the rest of this paper we fix an EOL system $G = (\Sigma, P, S, \{a, b\})$ which generates K and satisfies conditions (1) through (5) of Lemma 3.1.

We consider now several properties of derivation trees in G.

Lemma 3.2 Let T be an arbitrary derivation tree of G.

For every node of type 3 its direct ancestor is either the root (labelled by S) or a node of type 3.

For every node of type 2a its direct ancestor is either the root or a node of type x, $x \in \{3, 2a\}$.

For every node of type 2b its direct ancestor is either the root or a node of type x, $x \in \{3, 2b\}$.

For every node of type 1a its direct ancestor is either the root or a node of type x, $x \in \{3, 2a, 2b, 1a\}$.

For every node of type 1b its direct ancestor is either the root or a node of type x, $x \in \{3, 2a, 2b, 1b\}$. ∎

Lemma 3.3 There exists a positive integer k such that, for every $(\mathrm{us}\, G)$-labelled derivation tree T of G and every $0 \leq i \leq \mathrm{height}\, T$, $\#_{\mathrm{type}\, 3}\, \mathrm{result}_i\, T \leq k$.

Proof Since $K \in \mathrm{LB}$, there exists a function $f : \mathbb{N}^+ \to \mathbb{N}^+$ such that $\#\, \mathrm{BL}^{\mathrm{maxr}\, G}(w) \leq f(\mathrm{maxr}\, G)$. Let $k = f(\mathrm{maxr}\, G)$. We prove that for this choice of k the lemma holds. This is proved by contradiction.

Assume that T is a $(\mathrm{us}\, G)$-labelled tree of G and $0 \leq i \leq \mathrm{height}\, T$ is such that $\#_{\mathrm{type}\, 3}\, \mathrm{result}_i\, T > k$. Clearly based on T a successful derivation tree T' of G can be constructed such that there exists a $0 \leq j \leq \mathrm{height}\, T'$ such that $\#_{\mathrm{type}\, 3}\, \mathrm{result}_j\, T' > k$.

Then, since each letter of type 3 does create at least one block and since G is standard, we have

$$S \overset{j}{\underset{G}{\Rightarrow}} \mathrm{result}_j\, T' \underset{G}{\Rightarrow} w \in K \text{ and } \#\, \mathrm{BL}^{\mathrm{maxr}\, G}(w) > k = f(\mathrm{maxr}\, G);$$

a contradiction.

Hence the lemma holds. ∎

For the rest of the paper we will assume that k_1 is an arbitrary but fixed integer satisfying the statement of Lemma 3.3.

The following theorem shows that if α is a letter of type 1 or a letter of type 2 , then $L(G_\alpha) \in \mathrm{CB}$.

Theorem 3.1 If $\alpha \in \mathrm{type}\,1 \cup \mathrm{type}\,2$, then $L(G_\alpha) \in \mathrm{CB}$.

Proof If α is either of type 1 or of type 2, then each word in $L(G_\alpha)$ has at most two blocks. Thus by Theorem 2.2 $L(G_\alpha) \in \mathrm{CB}$. ∎

We will also show that for every letter α of type 3, $L(G_\alpha) \in \mathrm{CB}$. To this aim we divide letters of type 3 into two categories.

Definition 3.3 The set type $3I$ is the set of all letters of type 3 such that $\alpha \overset{+}{\underset{G}{\Rightarrow}} x$ implies $\mathrm{alph}\, x \cap \mathrm{rec}\, G \cap \mathrm{type}\,3 = \emptyset$ and type $3II$ is the set type $3 -$ type $3I$.

Nodes *of type* $3I$ ($3II$ respectively) are nodes labelled by letters of type $3I$ (type $3II$ respectively). ∎

We can now prove the analogon of Theorem 3.1 for letters of type $3I$.

Theorem 3.2 If $\alpha \in \mathrm{type}\,3I$, then $L(G_\alpha) \in \mathrm{CB}$.

Proof Follows immediately from Lemma 3.1 (5), Theorem 2.3 and Theorem 3.1. ∎

4. THE MAIN RESULT

In the last section we have divided letters of G into basic types and for each letter type *except for type 3II* we have proved that a letter of this type gives rise to a language in CB. Thus to complete the proof of our main result it suffices to demonstrate that also letters of type $3II$ yield languages in CB. This will be done in this section. We need a number of definitions first.

Definition 4.1 Let $\alpha \in$ type $3II$. Consider a successful derivation tree in G_α. The subtree spanned on all nodes of type $3II$ is called the *skeleton of* T and denoted skel T.

A node in skel T which has at most one successor (in skel T) is called *simple*; otherwise it is called *complex*.

A maximal (looked at top-down) branch in skel T which contains only simple nodes and ends either on a complex node or a leaf is called a *limb* (of T). Such a limb will be denoted as $\langle v_1, \ldots, v_k \rangle$ where $v_1, \ldots, v_k$ are subsequent nodes of the limb.

A *spike* (*of* a limb) is a subbranch of a limb consisting of nodes $v_\ell, \ldots, v_m$, $m \geq \ell$ such that v_{m+1} still belongs to the same limb and $\ell(v_\ell) = \ell(v_{m+1})$. A spike as above is said to be *of type* $\ell(v_\ell)$ and will be denoted as $\ll v_\ell, \ldots, v_m \gg$. ∎

Definition 4.2 Let $\alpha \in$ type $3II$, let T be a successful derivation tree of G_α of height ℓ, let $\rho = \langle v_{r_1}, \ldots, v_{r_2} \rangle$ be a limb of T such that for $r_1 \leq i \leq r_2$, v_i has distance i to the root and let $\sigma = \ll v_{i_1}, \ldots, v_{i_2} \gg$ be a spike of ρ.

Then for $i_1 < i \leq \ell$, $\mathrm{lset}_i(T, \sigma)$ denotes the following subset of $\mathrm{set}_i T$.

If $i_1 < i \leq i_2 + 1$ then $\mathrm{lset}_i(T, \sigma)$ consists of all nodes of $\mathrm{set}_i T$ which are descendants of $v_{i_1}, \ldots, v_{i-1}$ and which are to the left of v_i.

If $i_2 + 1 < i \leq \ell$ then $\mathrm{lset}_i(T, \sigma)$ consists of all nodes of $\mathrm{set}_i T$ which are descendants of $v_{i_1}, \ldots, v_{i_2}$ and which are to the left of all elements of $\mathrm{set}_i T$ which are descendants of v_{i_2+1}.

If in the above definition we replace left by right then we get the definition of $\mathrm{rset}_i(T, \sigma)$.

Finally $\mathrm{lresult}_i(T, \sigma)$ ($\mathrm{rresult}_i(T, \sigma)$ respectively) is the word which results from the sequence of all nodes (ordered from left to right) from $\mathrm{lset}_i(T, \sigma)$ ($\mathrm{rset}_i(T, \sigma)$ respectively) by replacing each node by its label.

If $i =$ height T then in the above the subscript i is omitted. ∎

Definition 4.3 Let $\alpha \in$ type $3II$. We associate with α two languages $\mathrm{LC}(\alpha)$ – the *left contributions of* α ,and $\mathrm{RC}(\alpha)$ – the *right contributions of* α as follows.

(1) $x \in \mathrm{LC}(\alpha)$

if and only if

there exists a letter β of type 3II, a successful derivation tree T of G_β , a limb $\rho = \langle v_1, \ldots, v_k \rangle$ of T and a spike $\sigma = \ll v_\ell, \ldots, v_m \gg$ of ρ of type α such that $x = \mathrm{lresult}(T, \sigma)$.

(2) $x \in \mathrm{RC}(\alpha)$

if and only if

there exists a letter β of type 3II, a successful derivation tree T of G_β , a limb $\rho = \langle v_1, \ldots, v_k \rangle$ of T and a spike $\sigma = \ll v_\ell, \ldots, v_m \gg$ of ρ of type α such that $x = \mathrm{rresult}(T, \sigma)$. ∎

The following lemma gives an upper bound for the number of complex nodes occurring in a successful derivation tree of G_α where $\alpha \in$ type $3II$.

Lemma 4.1 There exists a positive integer k such that for every $\alpha \in$ type $3II$ and every successful derivation tree T of G_α , skel T contains no more than k complex nodes.

Proof The lemma is proved by contradiction. Assume that no bound k exists. Let $\alpha \in$ type $3II$ and let T be a successful derivation tree of G_α such that skel T has more than k_1 complex nodes. This T can now easily be transformed (using Lemma 3.1 (4)) into a (us G)-labelled tree T' of G with at least k_1+1 complex nodes and such that each node of type 3II has a son of type 3II. Consequently there is a $i \geq 0$ such that $\#_{\text{type } 3} \, \mathrm{result}_i \, T' > k_1$ which contradicts Lemma 3.3. Thus the lemma holds. ∎

For the rest of the paper let k_2 be an arbitrary but fixed positive integer satisfying the statement of Lemma 4.1.

Now, given $\alpha \in$ type $3II$ we will prove that $L(G_\alpha) \subseteq M^k$ for a positive integer k. Then we will prove that $M \in$ CB and thus by Theorem 2.3 $L(G_\alpha) \in$ CB. We need the following definitions first.

Definition 4.4 Let $\alpha \in$ type $3II$. Let T be a successful derivation tree of G_α and let $\rho = \langle v_1, \ldots, v_k \rangle$ be a limb of T.

Let SPIKE(ρ) be the (possibly empty) sequence of spikes defined inductively as follows.

If ρ does not have spikes then SPIKE(ρ) $= \emptyset$.

Otherwise SPIKE(ρ) $= (\rho_1, \ldots, \rho_t)$, $t \geq 1$ where ρ_1, ... , ρ_t are spikes constructed one-by-one as follows.

Choose the first (top-down) node v_{i_1} on ρ such that ρ contains another node labelled by the same letter ; let v_{j_1+1} be the last node on ρ labelled by the same letter as v_{i_1}. Then $\rho_1 \ll v_{i_1}, v_{i_1+1}, \ldots, v_{j_1} \gg$. Then start with v_{j_1+1} and proceed as above. One stops when there are no spikes anymore in the remaining part of the limb.

By Rest ρ we denote the set of all nodes from ρ not involved in any of ρ_1, ..., ρ_t. Elements ρ_1 , ..., ρ_t of SPIKE(ρ) are called *true spikes of* ρ. For every limb ρ of T, a true spike of ρ is called a *a true spike of* T. ∎

Definition 4.5 Let $\alpha \in \text{type}\,3II$ and let T be a successful derivation tree of G_α. Let C_1 be the set of all nodes which occur in a true spike of T and let C_2 be the set of all nodes which contribute in one step to the terminal word. Observe that $C_1 \cap C_2 = \emptyset$ (the last node of a limb is never included in a spike). C_3 is the set of all nodes of skel T not contained in $C_1 \cup C_2$.

Then the *extended skeleton of* T, denoted eskel T is obtained from skel T by including also all nodes (and their edges to nodes from skel T) that can be obtained in one step from nodes in C_3.

The nodes of eskel T which do not belong to T are referred to as *added nodes of* T. Observe that none of the added nodes is a leaf of T and all added nodes belong to $\text{type}\,1 \cup \text{type}\,2 \cup \text{type}\,3$. ∎

In the following lemma we calculate bounds on the number of true spikes of an arbitrary successful derivation tree of G_α and the number of its added nodes ($\alpha \in \text{type}\,3II$).

Lemma 4.2 Let $\alpha \in \text{type}\,3II$ and let T be a successful derivation tree of G_α.

(1) The number of true spikes of T is bounded by $\#\Sigma \cdot (k_2 \cdot \text{maxr}\,G + 1)$.

(2) The number of added nodes of T is bounded by $\#\Sigma \cdot (k_2 \cdot \text{maxr}\,G + 1) \cdot \text{maxr}\,G$.

Proof (1) Since each limb of T either starts from a son of a complex node or from the root, the number of limbs of T is bounded by $k_2 \cdot \text{maxr}\,G + 1$. Since every limb can contain at most $\#\Sigma$ true spikes, the number of true spikes of T is bounded by $\#\Sigma \cdot (k_2 \cdot \text{maxr}\,G + 1)$.

(2) Clearly the number of added nodes of T is bounded by the number of limbs of T times the number of nodes on a limb not included in a true spike of T times $\text{maxr}\,G$, thus by $\#\Sigma \cdot (k_2 \cdot \text{maxr}\,G + 1) \cdot \text{maxr}\,G$. ∎

Next we present a language M such that $L(G_\alpha) \subseteq M^k$ for a positive integer K ($\alpha \in \text{type}\,3II$).

Lemma 4.3 Let $\alpha \in \text{type}\,3II$ and let M be the following language: $M = M_1 \cup M_2 \cup M_3 \cup M_4$ where

$$M_1 = \bigcup_{\beta \in \text{type}\,1 \cup \text{type}\,2 \cup \text{type}\,3I} L(G_\beta),$$

$$M_2 = \{x \in \{a,b\}^+ \mid \beta \underset{G}{\Rightarrow} x \text{ and } \beta \in \text{type}\,3II\},$$

$$M_3 = \bigcup_{\beta \in \text{type}\,3II} \text{LC}(\beta), \quad \text{and}$$

$$M_4 = \bigcup_{\beta \in \text{type } 3II} \mathrm{RC}(\beta).$$

Then there exists a positive integer k such that $L(G_\alpha) \subseteq M^k$.

Proof Let $k = (k_2 \cdot \mathrm{maxr}\, G+1)(\#\,\Sigma \cdot (\mathrm{maxr}\, G+2)+1)$. Consider a successful derivation tree T in G_α of w. The word w is divided into subwords as follows :

(i) contributions of added nodes,

(ii) contributions of nodes of the C_2 category,

(iii) left contributions of true spikes of T,

(iv) right contributions of true spikes of T.

Obviously in this way $w = w_1 \ldots w_p$ where for $1 \leq i \leq p$, $w_i \in M$. To prove the lemma it suffices to prove that $p \leq k$.

(1) $\#\{i \mid 1 \leq i \leq p,\ w_i \in M_1\}$ is bounded by the number of added nodes of T thus by $\#\,\Sigma \cdot (k_2 \cdot \mathrm{maxr}\, G + 1) \cdot \mathrm{maxr}\, G$ (see Lemma 4.2 (2)).

(2) $\#\{i \mid 1 \leq i \leq p,\ w_i \in M_2\}$ is bounded by the number of nodes of the C_2 category, thus by the number of limbs, i.e. by $k_2 \cdot \mathrm{maxr}\, G + 1$

(3) $\#\{i \mid 1 \leq i \leq p,\ w_i \in M_3\}$ is bounded by the number of true spikes of T, thus by $\#\,\Sigma \cdot (k_2 \cdot \mathrm{maxr}\, G + 1)$ (see Lemma 4.2 (1)).

(4) $\#\{i \mid 1 \leq i \leq p,\ w_i \in M_4\}$ is bounded by the number of true spikes of T, thus by $\#\,\Sigma \cdot (k_2 \cdot \mathrm{maxr}\, G + 1)$ (see Lemma 4.2 (1)).

Combining (1) through (4) we get that p is bounded by

$$(k_2 \cdot \mathrm{maxr}\, G + 1)(\#\,\Sigma \cdot (\mathrm{maxr}\, G + 2) + 1) = k.$$

Thus the lemma holds. ∎

To prove that for $\alpha \in \text{type } 3II$, $L(G_\alpha) \in \mathrm{CB}$ it remains to prove that $\mathrm{LC}(\beta) \in \mathrm{CB}$ and $\mathrm{RC}(\beta) \in \mathrm{CB}$ for each $\beta \in \text{type } 3II$. We need the following definition and lemma first.

Definition 4.6 Let $x = a_1 \ldots a_n$, $a_i \in \mathrm{us}\, G$ for $1 \leq i \leq n$, $(y, k, \ell) \in \mathrm{SEG}(x)$.

Then y is called a *promissed block of type* A if $x = a_1 \ldots a_{k-1} y a_{\ell+1} \ldots a_n$ and one of the following conditions holds.

(1) $y \in (\text{type } 2b)(\text{type } 1a)^*(\text{type } 2a)$.

(2) $y \in (\text{type } 2b)(\text{type } 1a)^*$ and $a_{\ell+1} \in (\text{type } 1b \cup \text{type } 2b)$.

(3) $y \in (\text{type } 1a)^*(\text{type } 2a)$ and $a_{k-1} \in (\text{type } 1b \cup \text{type } 2a)$.

(4) $y \in (\text{type } 1a)^+$, $a_{k-1} \in (\text{type } 1b \cup \text{type } 2a)$ and $a_{\ell+1} \in (\text{type } 1b \cup \text{type } 2b)$.

We call y a *promissed block of type* B if $x = a_1 \ldots a_{k-1} y a_{\ell+1} \ldots a_n$ and one of the following conditions holds.

(1) $y \in (\text{type } 2a)(\text{type } 1b)^*(\text{type } 2b)$.

(2) $y \in (\text{type } 2a)(\text{type } 1b)^*$ and $a_{\ell+1} \in (\text{type } 1a \cup \text{type } 2a)$.

(3) $y \in (\text{type } 1b)^*(\text{type } 2b)$ and $a_{k-1} \in (\text{type } 1a \cup \text{type } 2b)$.

(4) $y \in (\text{type } 1b)^+$, $a_{k-1} \in (\text{type } 1a \cup \text{type } 2b)$ and $a_{\ell+1} \in (\text{type } 1a \cup \text{type } 2a)$.

If y is a promissed block of type A or a promissed block of type B then y is called a *promissed block.* ■

Lemma 4.4 Let $\alpha \in \text{type}\, 3II$ and let T be a successful derivation tree of G_α which contains a spike $\sigma = \ll v_1, \ldots, v_p \gg$ and there exists a positive integer t such that either $\text{lresult}_t(T, \sigma)$ or $\text{rresult}_t(T, \sigma)$ contains as subword a promissed block x. Then

(a) If x is of type A then there exists a $\beta \in \text{alph}\, x$ such that for every $n > 0$, $\beta \overset{n}{\underset{G}{\Rightarrow}} w$, $w \in (\text{us}\, G)^+$ implies $\#_{\text{type 1}a}\, w \geq n$.

(b) If x is of type B then there exists a $\beta \in \text{alph}\, x$ such that for every $n > 0$, $\beta \overset{n}{\underset{G}{\Rightarrow}} w$, $w \in (\text{us}\, G)^+$ implies $\#_{\text{type 1}b}\, w \geq n$.

Proof We will prove (a); the proof of (b) is analogous. That (a) holds is proved by contradiction. Assume that (a) does not hold. Then Lemma 3.1 (3) implies that for every $\beta \in \text{alph}\, x$ there exists a positive integer C_β such that for every $n > 0$ there exists a $w_n \in (\text{us}\, G)^+$ where $\beta \overset{n}{\underset{G}{\Rightarrow}} w_n$ and $\#_{type1a}\, w_n \leq C_\beta$. Let

$$D = \max\{C_\beta \mid \beta \in \text{alph}\, x\} \cdot |x| \cdot \text{maxr}\, G.$$

Let f be such that $K \in \text{LB}$ with parameter f. Without loss of generality we assume that X is a subword of $\text{lresult}_t(T, \sigma)$.

Let T' be the subtree with root v_1 where we delete

(i) all descendants of v_{p+1},

(ii) all nodes of set T, and

(iii) all nodes of $\text{lset}_i(T, \sigma)$, $i > t$.

Let $T_1, \ldots, T_{f(D)+1}$ be $f(D)+1$ disjoint copies of T'. Let $v_1^1, \ldots, v_1^{f(D)+1}$ be the nodes corresponding to v_1 and let $v_{p+1}^1, \ldots, v_{p+1}^{f(D)+1}$ be the nodes corresponding to v_{p+1}. Let T'' be the tree which results from $T_1, \ldots, T_{f(D)+1}$ by identifying v_1^{j+1} and v_{p+1}^j for $1 \leq j < f(D)+1$. Let T''' be the tree which results from T by removing all nodes of set T and by replacing the subtree rooted at v_1 by T''.

Finally T''' is completed to a successful derivation tree of G_α as follows. Let $m = \text{height}\, T''' + 1$. In each leaf node v of T''' with $\ell(v) = \beta \in \text{alph}\, x$ which occurs in $\text{set}_j\, T'''$, $j < m$, append a tree representing a derivation

$$D_{\beta,j} : \beta \overset{m-j}{\underset{G}{\Rightarrow}} w_{\beta,j} \in (\text{us}\, G)^+ \text{ where } \#_{type1a}\, w_{\beta,j} \leq C_\beta$$

In each leaf node v of T''', with $\ell(v) = \beta \not\in \text{alph}\, x$ which occurs in $\text{set}_j\, T'''$, $j < m$ append an arbitrary tree representing a derivation $D_{\beta,j} : \beta \overset{m-j}{\underset{G}{\Rightarrow}} w_{\beta,j} \in (\text{us}\, G)^+$. Let $\overline{T}$ denote the resulting tree.

Since G is standard the above construction implies the existence of a word $w \in K$ such that $\#\, \text{BL}^D(w) \geq f(D)+1$; a contradiction. Hence (a) holds. ■

We are now ready to prove that for $\alpha \in \text{type}\,3II$, $\mathrm{RC}(\alpha)$ and $\mathrm{LC}(\alpha)$ belong to CB.

Lemma 4.5 Let $\alpha \in \text{type}\,3II$. Then $\mathrm{RC}(\alpha) \in \mathrm{CB}$ and $\mathrm{LC}(\alpha) \in \mathrm{CB}$.

Proof Let $g : \mathbb{N}^+ \to \mathbb{N}^+$ be defined by $g(n) = (\mathrm{maxr}\, G)^{n+4}$. We will show that $\mathrm{RC}(\alpha) \in \mathrm{CB}$ with parameters 4 and g. The proof for $\mathrm{LC}(\alpha)$ is analogous.

Let $x \in \mathrm{RC}(\alpha)$. Thus we have a situation as expressed by Definition 4.3. We will use the notation of Definition 4.3. We number the levels from T bottom-up and we consider the covering of $\mathrm{BL}^t(x)$ where t is a positive integer. The situation is depicted in figure 1.

Let $x = x_1 \ldots x_q$, $q \geq 1$ where $x_i \in \mathrm{BL}(x)$ for $1 \leq i \leq q$. Blocks x_1 and x_q are called *outside blocks* and blocks $x_2, \ldots, x_{q-1}$ are called *inside blocks.*

Let $1 \leq i \leq q$ be such a block. Let $p(x_i)$ be the number of the level on which a node v_j of σ lies such that v_j is an ancestor of last x_i but v_{j+1} is not an ancestor of last x_i. We consider inside blocks only.

Let x_i be such a block. Now x_i is called

t-young if $p(x_i) \leq t+4$,

t-old if $p(x_{i-1} \geq t+4$,

t-middle if $p(x_{i-1}) \leq t+4$ and $p(x_i) \geq t+4$.

Claim 4.1 If x_i is t-old, then $|x_i| \geq t+1$.

Proof Let u_1 (u_2 respectively) be the word formed by all ancestors of nodes from x_i on the level $p(x_{i-1})-1$ ($p(x_{i-1})-2$ respectively). The situation is depicted in figure 2.

All letters of u_1 are of one of the follwing types: $3I$, $2a$, $2b$, $1a$, $1b$, and consequently all letters of u_2 are of one of the following types : $2a$, $2b$, $1a$, $1b$. Consequently u_2 is a promissed block (see Definition 4.6) and thus by Lemma 4.4 u_2 contributes at least $p(x_{i-1})-3$ letters a (or letters b) to the block x_i. Since x_i is t-old, $p(x_{i-1}) \geq t+4$ and consequently $p(x_{i-1})-3 \geq t+1$; thus $|x_i| \geq t+1$. ∎

Claim 4.2 The joint length of all t-young blocks is bounded by $(\mathrm{maxr}\, G)^{t+4}$; moreover, all t-young blocks are adjacent.

Proof Both facts follow from the fact that to the left of a t-young block (different from the first inside block) there is always a t-young block and all of them are included in the contribution to w of the node in σ which is an ancestor of the last letter of the rightmost t-young block (and this node is on a level not higher than $t+4$. ∎

Claim 4.3 Among $x_2, \ldots, x_{q-1}$ there is at most one t-middle block.

Proof This follows immediately from Claim 4.2 and the observation that each block to the right of a t-middle block is t-old and each block to the left of a t-middle block is t-young. ∎

We have now the following situation concerning blocks in x.

– all t-old blocks are outside $\mathrm{BL}^t(w)$,

– there are at most three ("special") blocks to handle: x_1, x_q and the t-middle block, each of them has length at most t and can be covered by a segment of length $(\operatorname{maxr} G)^{t+4}$ if $\operatorname{maxr} G \geq 2$ (clearly, this can be assumed without loss of generality).

– all t-young blocks form a subword of length bounded by $(\operatorname{maxr} G)^{t+4}$.

Then clearly we can choose $X \subseteq \mathrm{SEG}(w)$ such that

(i) $\# X \leq 4$;

(ii) for every $z \in X$, $|z| \leq (\operatorname{maxr} G)^{t+4} = g(t)$; and

(iii) X covers $\mathrm{BL}^t(w)$.

Thus indeed $\mathrm{RC}(\alpha) \in \mathrm{CB}$ with parameters 4 and g. ∎

We are now ready to state the analogon of Theorem 3.1 and Theorem 3.2 for letters of type $3II$.

Theorem 4.1 If $\alpha \in \text{type}\, 3II$, then $L(G_\alpha) \in \mathrm{CB}$.

Proof If $\alpha \in \text{type}\, 3II$, then $L(G_\alpha) \subseteq M^k$ for a positive integer k where M is as in the statement of Lemma 4.3 ,

$$M = M_1 \cup M_2 \cup M_3 \cup M_4.$$

We know that $M_1 \in \mathrm{CB}$ by Theorem 3.1 and Theorem 3.2 , $M_2 \in \mathrm{CB}$ by Theorem 2.2 , M_3, $M_4 \in \mathrm{CB}$ by Lemma 4.5. Thus $M \in \mathrm{CB}$ by Lemma 2.1 and consequently Lemma 2.2 implies that $L(G_\alpha) \in \mathrm{CB}$. ∎

Theorem 3.1 , Theorem 3.2 and Theorem 4.1 can be combined into the following result.

Theorem 4.2 $K \in \mathrm{CB}$.

Proof For every $\alpha \in \mathrm{us}\, G$, $\alpha \in \text{type}\, 1 \cup \text{type}\, 2 \cup \text{type}\, 3I \cup \text{type}\, 3II$. Then by Theorem 3.1, Theorem 3.2 and Theorem 4.1 , $L(G_\alpha) \in \mathrm{CB}$.

Since

$$K = L(G) \subseteq \left(\bigcup_{\alpha \in \mathrm{us}\, G} L(G_\alpha) \cup M_0 \right)^k$$

for a finite language M_0 and a positive integer k, Theorem 2.2 and Theorem 2.3 imply $K \in \mathrm{CB}$. ∎

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

Theorem 4.3 Let Δ be an alphabet, let $\Pi = (\Delta_1, \Delta_2)$ be a binary partition of Δ and let $L \subseteq \Delta^*$ be an EOL language such that $L \in \mathrm{LB}(\Pi)$. Then $L \in \mathrm{CB}(\Pi)$.

Proof Let L be as in the statement of the theorem. Let h be the homomorphism defined by $h(\alpha) = a$ if $\alpha \in \Delta_1$ and $h(\alpha) = b$ if $\alpha \in \Delta_2$ where a and b are two different symbols.

Then by Theorem 2.4, $h(L) \in \mathrm{LB}(\{a\}, \{b\})$. Then Theorem 4.2 implies $h(L) \in \mathrm{CB}(\{a\}, \{b\})$ and consequently, again applying Theorem 2.4 yields $L \in \mathrm{CB}(\Pi)$. ∎

5. APPLICATIONS

In this section we present examples to show the usefulness of Theorem 4.3 for proving that a language is not an EOL language.

Example 5.1

$$K = \{a^{i_1}b^{j_1}a^{i_2}b^{j_2}\ldots a^{i_n}b^{j_n} \mid n \geq 1 \text{ and } i_\ell, j_\ell \geq \ell \text{ for } 1 \leq \ell \leq n\}.$$

It was proved in Section 2 that for $\Pi = (\{a\},\{b\})$, $K \in \mathrm{LB}(\Pi) - \mathrm{CB}(\Pi)$. Consequently Theorem 4.3 implies $K \notin \mathcal{L}(EOL)$. Observe that $K \in \mathcal{L}(ETOL)$. ∎

Example 5.2 Let G be a rewriting system with letters a and b, productions $a \to b^2$ and $b \to a^2$, axiom a and with the following rewriting rule.

Let $x = a_1 \ldots a_n$, $n \geq 1$, $a_i \in \{a, b\}$ for $1 \leq i \leq n$, and let $y = y_1 \ldots y_n$. Then $x \underset{G}{\Rightarrow} y$ if either $a_i \underset{P}{\to} y_i$ for $1 \leq i \leq n$,or there exists a $1 \leq j \leq n$ such that $y_i = a_i$ and for $1 \leq i \leq n$, $i \neq j$, $a_i \underset{P}{\to} y_i$. Thus in a string either all letters are rewritten or all but one letter are rewritten.

The system G is a regular pattern grammar (see, e.g., [KR1]). We will now show that $K = L(G)$ (which consists of all strings derivable from the axiom) is not an EOL language. Observe that we are going to prove this without knowing any explicit expression for the language K.

Let $\Pi = (\{a\},\{b\})$.

Claim 5.1 $K \in \mathrm{LB}(\Pi)$ with parameter f where $f(1) = 1$ and for $m > 1$, $f(m) = f(\lfloor m/2 \rfloor) + 3$. ($\lfloor m/2 \rfloor$ denotes the integer k such that $2k \leq m$ and $2(k+1) > m$).

Proof The proof goes by induction on the length of the blocks m.

(1) $m = 1$

We must prove that for every $w \in K$, $\#\mathrm{BL}^1_\Pi(w) \leq 1$. This is proved by induction on the number of derivation steps needed to derive w.

If w is the axiom then $w = a$ and thus $\#\mathrm{BL}^1_\Pi(a) = 1$. Assume that for any $w \in K$ which can be derived in less than or equal to n steps, $\#\mathrm{BL}^1_\Pi(w) \leq 1$. Then let $a \overset{n+1}{\Rightarrow} w_{n+1}$, i.e. $a \overset{n}{\Rightarrow} w_n \Rightarrow w_{n+1}$ and by induction $\#\mathrm{BL}^1_\Pi(w_n) \leq 1$.

To obtain w_{n+1} either all occurrences of letters from w_n are rewritten or all but one occurrences of letters from w_n are rewritten. If all occurrences of letters from w_n are rewritten then the form of the productions implies that every block of w_{n+1} has length at least 2, hence $\#\mathrm{BL}^1_\Pi(w_{n+1}) \leq 1$. If one occurrence of a letter is not rewritten, then this occurrence is the only possible candidate to belong to a block of length 1, hence $\#\mathrm{BL}^1_\Pi(w_{n+1}) \leq 1$.

This concludes the proof for the case $m = 1$.

(2) Assume that for every $w \in K$, $\#\mathrm{BL}^k_\Pi(w) \leq f(k)$ for $1 \leq k \leq m$. Then we prove that for every $w \in K$, $\#\mathrm{BL}^{m+1}_\Pi(w) \leq f(m+1)$. This is proved by induction on the length of a derivation of w.

If w is the axiom , $\#\mathrm{BL}^{m+1}_\Pi(w) \leq f(m+1)$ clearly holds. Assume that for any $w \in K$ such that $a \overset{\leq n}{\Rightarrow} w$, $\#\mathrm{BL}^{m+1}_\Pi(w) \leq f(m+1)$. Then let $a \overset{n}{\Rightarrow} w_n \Rightarrow w_{n+1}$.

As in (1) there are two possible ways to derive w_{n+1} from w_n. If productions are applied to all occurrences of letters of w_n then every element of $\mathrm{BL}_{\Pi}^{m+1}(w_{n+1})$ must come from an element of $\mathrm{BL}_{\Pi}^{\lfloor (m+1)/2 \rfloor}(w_n)$. Thus

$$\#\,\mathrm{BL}_{\Pi}^{m+1}(w_{n+1}) \leq \#\,\mathrm{BL}_{\Pi}^{\lfloor (m+1)/2 \rfloor}(w_n) \leq f(\lfloor (m+1)/2 \rfloor) \leq f(m).$$

If one occurrence of a letter of w_n is not rewritten then

$$\#\,\mathrm{BL}_{\Pi}^{m+1}(w_{n+1}) \leq \#\,\mathrm{BL}_{\Pi}^{\lfloor (m+1)/2 \rfloor}(w_n) + 3 \leq f(\lfloor (m+1)/2 \rfloor) + 3 \leq f(m).$$

This concludes the proof of (2).

From (1) and (2) the claim follows. ∎

Claim 5.2 $K \notin \mathrm{CB}(\Pi)$.

Proof The proof goes by contradiction. Assume that $K \in \mathrm{CB}(\Pi)$ with parameters n_0 and g. We will derive a word w which violates the property.

Let $n > g(2^{n_0+1}) \cdot (n_0+1)$ and let $t = g(2^{n_0+1})$. To get w we first derive a^{2^n} and then proceed $n_0 + 1$ steps (using the second rewriting rule) according to the following scheme (for each step we have underlined the occurrence which is not rewritten in that step).

$$\begin{aligned} \underline{a}a^{2^n-1} &\Rightarrow ab^t\underline{b}b^{x_1} \\ &\Rightarrow b^2a^{2t}ba^t\underline{a}a^{x_2} \\ &\Rightarrow a^4b^{4t}a^2b^{2t}ab^t\underline{b}b^{x_3} \\ &\Rightarrow b^8a^{8t}b^4a^{4t}b^2a^{2t}ba^t\underline{a}a^{x_4} \\ &\Rightarrow \ldots \end{aligned}$$

If (n_0+1) is odd we get

$$w = a^{2^{n_0+1}}b^{(2^{n_0+1})\cdot t}a^{2^{n_0}}b^{2^{n_0}\cdot t}\ldots a^{2^2}b^{2^2\cdot t}a^{2^1}b^{2^1\cdot t}ab^tbb^{x_{n_0+1}}$$

and if (n_0+1) is even we get the same word with the roles of a and b interchanged.

Without loss of generality assume that n_0 is odd (n_0 is even is symmetric). Then all blocks of w consisting of b's only are longer than t. Let $m_0 = 2^{n_0+1}$. All blocks of length at most m_0 consist only of a's. However no two different elements of $\mathrm{BL}_{\Pi}^{m_0}(w)$ can be covered by a segment of length at most $g(m_0) = t$. Thus if X covers $\mathrm{BL}_{\Pi}^{m_0}(w)$ then $\#X > n_0$; acontradiction. This ends the proof of Claim 5.2. ∎

Since $K \in \mathrm{LB}(\Pi) - \mathrm{CB}(\Pi)$, by Theorem 4.3 , $K \notin \mathcal{L}(EOL)$. ∎

ACKNOWLEDGEMENT.

The first and the second author gratefully acknowledge the financial support of NSF grant MCF 79-038038 .The second and the third author are indebted to NFWO Belgium for supporting their research ;all authors are indebted to H.C.M. Kleijn for comments on the first version of the paper.

REFERENCES

[KR1] H.C.M. Kleijn and G. Rozenberg, Context-free like restrictions on selective rewriting, Theoretical Computer Science, v.16, 1981, 237-269.

[KR2] H.C.M. Kleijn and G. Rozenberg, On the generative power of regular pattern grammars, manuscript.

[RS] G. Rozenberg and A. Salomaa, The mathematical theory of L systems, 1980, Academic Press, New York.

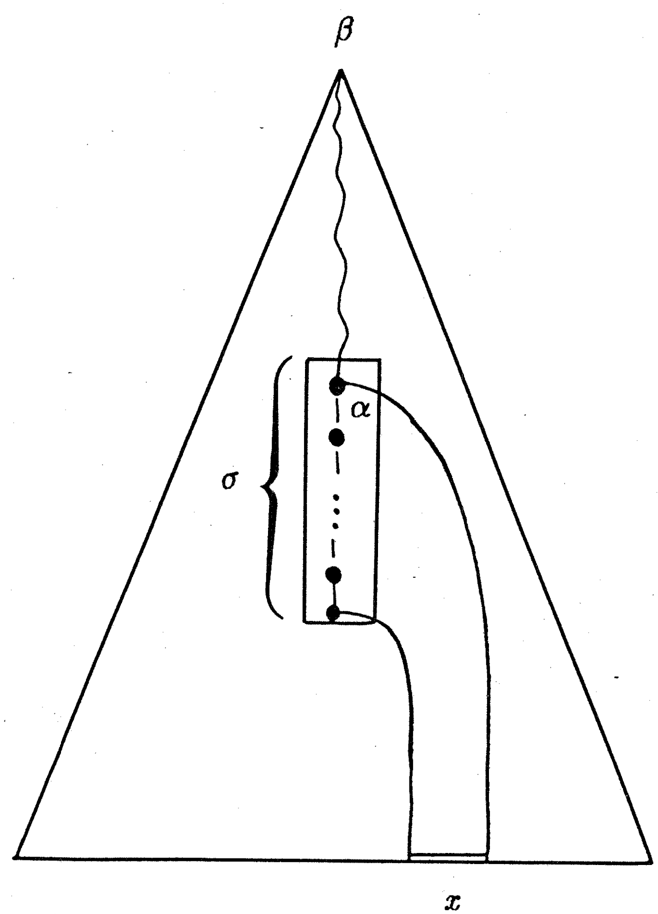

fig. 1.

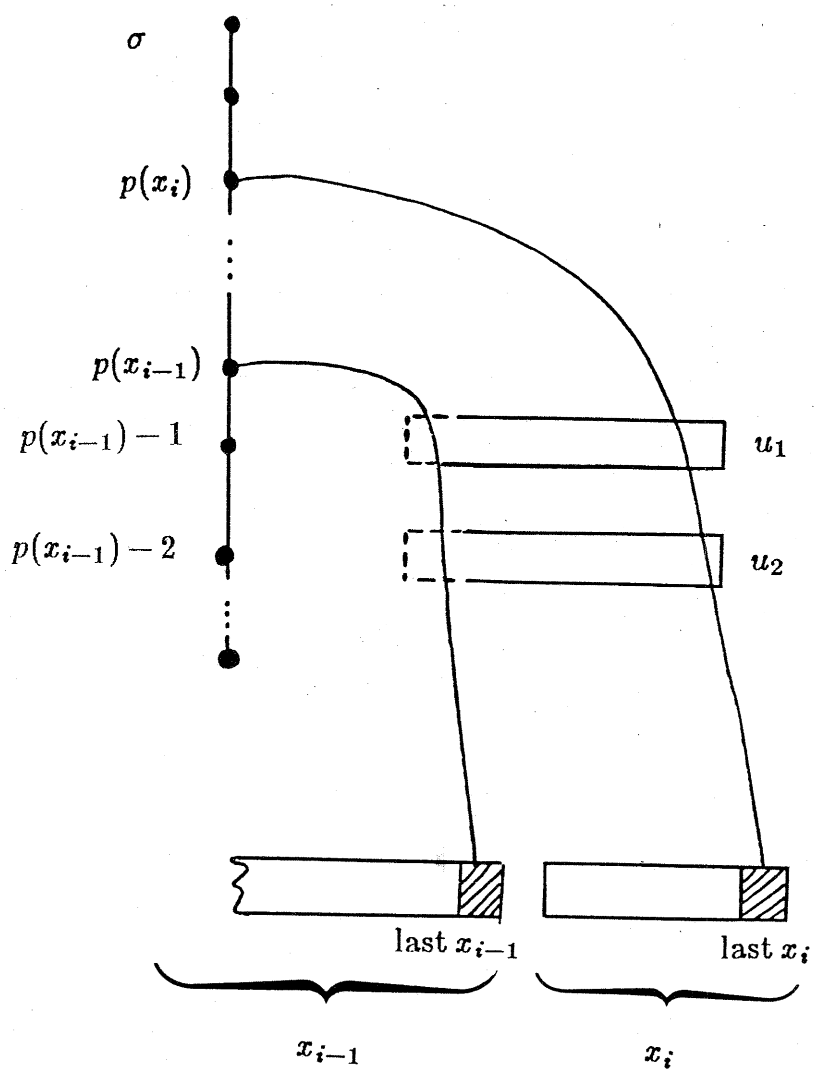

fig. 2.

A COMBINATORIAL PROPERTY OF EOL LANGUAGES

A. Ehrenfeucht
Department of Computer Science
University of Colorado–Boulder
Boulder, CO 80309, USA

G. Rozenberg
Institute of Applied Mathematics and Computer Science
University of Leiden
The Netherlands

R. Verraedt
Research Assistant of the National Fund for Scientific Research (Belgium)
Department of Mathematics
University of Antwerp, U.I.A.
B-2610 Wilrijk
Belgium

All correspondence to:

G. Rozenberg
Institute of Applied Mathematics and Computer Science
University of Leiden
Leiden, The Netherlands