ON INHERENTLY AMBIGUOUS EOL LANGUAGES

by

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

CU-CS-257-83 July, 1983

*Department of Computer Science, University of Colorado, Boulder, CO 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., Antwerp, Belgium

ABSTRACT

An EOL system G is called *ambiguous* if its language contains a word with (at least) two different derivations in G. An EOL language is called inherently EOL-*ambiguous* if every EOL system generating it is ambiguous. It is demonstrated that there exist inherently ambiguous EOL languages and in particular that the language $\{a^m b^{2^n} : 1 \leq m \leq n\} \cup \{a^{m^2} b^{2^n} : 1 \leq m \leq n\}$ is inherently EOL-ambiguous.

INTRODUCTION

The class of EOL languages forms a very natural extension of the class of context free languages and it is a very central class in the theory of parallel rewriting systems (see, e.g., [RS] and the references there). Quite a number of results are available concerning the combinatorial structure of EOL languages (see e.g., [RS]).

A particularly interesting topic concerned with the combinatorial structure of EOL languages is that of ambiguity. An EOL system G is called *ambiguous* if its language contains a word with (at least) two different derivation trees in G. An EOL language is called *inherently* EOL-*ambiguous* if every EOL system that generates it is ambiguous. The topic of ambiguity of EOL systems and languages is investigated in [MSW], [ReS] and [ER2]. In particular in [ER2] it is demonstrated that the degree of ambiguity of a context free language K in the class of EOL systems is not larger than the degree of ambiguity of K in the class of context free grammars. Perhaps the most natural question concerning ambiguity of EOL systems and languages that was open until now is whether or not there exist EOL languages that are inherently EOL-ambiguous. The analogous question concerning inherently "(context free)-ambiguous" context free languages was settled at the beginning stage of the development of the theory of context free languages (see e.g., [H] and [S]).

In this paper we demonstrate that there exist inherently ambiguous EOL languages, settling in this way the open problem from [MSW].

0. PRELIMINARIES

We assume the reader to be familiar with the basic theory of EOL systems, e.g., in the scope of [RS]; with the exception of some minor changes we follow the notation and terminology from [RS]. To facilitate the reading of this paper we recall now some basic notation and terminology.

We use N, N^+ and R^+ to denote the set of natural numbers, the set of positive integers and the set of positive reals respectively. $\emptyset$ denotes the empty set and for sets A and B, $A \backslash B$ denotes their difference. We often identify a singleton set with its element. A set $Z \subseteq N$ is called *numerically dispersed* if for every $r \in N$ there exists $n_r \in N$ such that for every $m_1, m_2 \in Z$ if $n_r < m_1 < m_2$ then $m_2 - m_1 > r$. For a real r, $\lfloor r \rfloor$ denotes the biggest integer n such that $n \leq r$.

For a word x, $|x|$ denotes its length and *alph*(x) denotes the set of all symbols that occur in x. If Σ is an alphabet and $\Delta \subseteq \Sigma$ then $pres_{\Sigma,\Delta}$ is the homomorphism of Σ^* defined by: for $b \in \Sigma$, $pres_{\Sigma,\Delta}(b) = b$ if $b \in \Delta$ and $pres_{\Sigma,\Delta}(b) = \Lambda$ if $b \in \Sigma \backslash \Delta$. We will write $pres_{\Delta}$ rather than $pres_{\Sigma,\Delta}$ whenever the alphabet Σ is understood.

An EOL *system* is specified in the form $G = (\Sigma, h, \omega, \Delta)$ where Σ is its alphabet, h its finite substitution, ω its axiom and Δ its terminal alphabet. If $z \in \Sigma^+$ and $z \stackrel{*}{\Rightarrow} \gamma_1 b \gamma_2$ where $\gamma_1, \gamma_2 \in \Sigma^*$ and $b \in \Sigma$ then we say that b is *reachable* from z and we write $z \leq b$. If $b \leq b$ then we say that b is *recursive*. We assume that G is *reduced*, i.e., $\omega \leq b$ for every $b \in \Sigma$. If G is synchronized then F is the synchronization symbol of G; if additionally $\omega \in \Sigma \backslash \Delta$ then we use $W(G)$ to denote the set $\Sigma \backslash (\Delta \cup \{F, \omega\})$ and $S(G)$ to denote the set of all sentential forms z such that $alph(z) \subseteq W(G) \cup \{S\}$. If G is a DOL system then E(G) denotes its sequence.

Since problems considered in this paper become trivial otherwise, we consider only infinite EOL *systems and languages. Also, we deal with propagating* EOL *systems only.*

We recall now from [ER1] the notion of a DOL system with rank. This notion forms a very essential tool in the proof of our main result. We assume the reader to be familiar with the topic of DOL systems with rank.

Definition 1. Let $G = (\Sigma, h, \omega)$ be a DOL system where $\omega \in \Sigma$.

(1). For a letter $b \in \Sigma$ the *rank of* b *in* G, denoted $rank_G(b)$, is defined inductively as follows.

(i). If $L(G_b)$ is finite, then $rank_G(b) = 0$, where $G_b = (\Sigma, h, b)$.

(ii). Let, for $i \geq 1$, $\Sigma_{(i)} = \Sigma \setminus \{a \in \Sigma: rank_G(a) < i\}$ and let $f_{(i)}$ be the homomorphism of Σ^* defined by:

$f_{(i)}(a) = a$ for $a \in \Sigma_{(i)}$ and $f_{(i)}(a) = \Lambda$ for $a \in \Sigma \setminus \Sigma_{(i)}$. Then let $h_{(i)}$ be the homomorphism of $\Sigma^*_{(i)}$ defined by $h_{(i)}(a) = f_{(i)}(h(a))$. If b is such that the language of the DOL system $(\Sigma_{(i)}, h_{(i)}, b)$ is finite then $rank_G(b) = i$.

(2). We say that G is a DOL *system with rank* if every $b \in \Sigma$ reachable from ω has a rank. The *rank of* G is the highest of the ranks of letters reachable from ω. □

We define now the basic notion of this paper.

Definition 2.

(1). Let $G = (\Sigma, h, \omega, \Delta)$ be an EOL system such that $\omega \in \Sigma \setminus \Delta$. We say that G is *unambiguous* is every word in L(G) possesses precisely one derivation tree in G. Otherwise G is *ambiguous*.

(2). Let K be an EOL language. We say that K is *inherently* EOL-*ambiguous* if every EOL system generating K is ambiguous. Otherwise we say that K is EOL-*unambiguous*. □

In the sequel we say simply "inherently ambiguous" and "unambiguous" rather than "inherently EOL-ambiguous" and "EOL-unambiguous" respectively.

Lemma 1. Let G be an unambiguous EOL system. There exists a constant $\alpha \in R^+$ such that, for every derivation D in G of a word w, $|w| \geq \alpha |D|$, where $|D|$ is the length of D.

Proof.

This lemma follows directly from the fact that G is propagating and unambiguous. □

We define now a subclass of the class of EOL systems.

Definition 3. An EOL system $G = (\Sigma, h, S, \Delta)$ is *clean* if it satisfies the following properties.

(1). $S \in \Sigma \backslash \Delta$ and, for each $b \in \Sigma$, $\beta \in h(b)$, $S \notin alph(\beta)$.

(2). If $\alpha \in h(S)$ and $\alpha \in (W(G))^+$ then $|\alpha| \geq 2$.

(3). G is propagating.

(4). G is reduced.

(5). G is synchronized.

(6). If $b \in W(G)$ then for every $n \in N^+$ there exists a word $\beta \in (W(G))^+$ such that $\beta \in h^n(b)$.

(7). If $b \in W(G)$ then the set $\{\beta \in S(G): b \in alph(\beta)\}$ is infinite.

(8). If $b \in W(G)$ then $h(b) \cap \Delta^+ \neq \emptyset$. □

The usefulness of clean EOL systems for our considerations stems from the following result. Its proof is standard (using the speed-up technique) and so we leave it to the reader.

Lemma 2. For every unambiguous EOL language K there exists an unambiguous EOL system G such that L(G) = K and G is clean. □

I. THE MAIN RESULT

In this section we prove that there exist inherently ambiguous EOL languages.

Theorem. There exist inherently ambiguous EOL languages.

Proof.

Let $K_0 = \{a^m b^{2^n} : 1 \le m \le n\}$, $K_1 = \{a^{m^2} b^{2^n} : 1 \le m \le n\}$ and $K_2 = K_0 \cup K_1$.

We will prove that K_2 is an inherently ambiguous EOL language.

To see that K_2 is an EOL language consider the following EOL system $G_2 = (\Sigma_2, h_2, S, \{a,b\})$ where

$\Sigma_2 = \{ S, X, \overline{X}, U, A, B, \$, F, a, b\}$

and h_2 is defined as follows:

$h_2(S) = \{X B^2, A B^2\}$, $h_2(X) = \{X, U \overline{X}, a\}$, $h_2(B) = \{B^2, b\}$,

$h_2(\overline{X}) = \{U \overline{X}, a\}$, $h_2(U) = \{U, a\}$, $h_2(A) = \{A, U^2C \$, a\}$,

$h_2(\$) = \{U^2C \$, a\}$, $h_2(C) = \{U^2C, a\}$, $h_2(a) = \{F\}$, $h_2(b) = \{F\}$

and $h_2(F) = \{F\}$.

It is rather easy to see that $L(G_2) = K_2$.

To prove that K_2 is an inherently ambiguous EOL language we proceed as follows.

Assume that $G = (\tilde{\Sigma}, \tilde{h}, \tilde{S}, \tilde{\Delta})$ is an unambiguous EOL system generating K_2; by Lemma 2 we can assume that G is clean. We need now some additional terminology and notation.

In the sequel we identify derivations with their traces; that is a derivation in G is a sequence $(x_0, x_1, \ldots, x_k)$, $k \ge 1$, such that $x_i \Rightarrow x_{i+1}$ for $0 \le i \le k-1$. (This will not lead to a confusion, because we have defined ambiguity through derivation trees!). Given a derivation

$D = (x_0,\dots,x_k)$, its length, denoted $|D|$, is equal to k and its result, denoted *res*(D), is equal to x_k. We also say that D is a derivation *of* x_k, D is a derivation *from* x_0 and for $0 \le i \le k$, x_i is the i'*th level* of D. D is called *complete* if $x_0 = S$ and $x_k \in \Delta^+$. We use $\hat{\mathcal{D}}_G$ to denote the set of all derivations in G and $\mathcal{D}_G$ to denote the set of all complete derivations in G. Correspondingly, we use $\hat{\mathcal{T}}_G$ to denote the set of all derivation trees (forests) in G and $\mathcal{T}_G$ to denote the set of all complete derivation trees in G (that is trees corresponding to complete derivations in G).

Given a tree (forest) in $\hat{\mathcal{T}}_G$ we call it *deterministic* if the nodes with the same labels are rewritten in the same way; otherwise the tree is called *nondeterministic*.

Given a letter $X \in W(G)$ we say that

- X is *directly* t-*nondeterministic* if h(X) contains two different words $\beta_1, \beta_2 \in \Delta^+$,
- X is *directly* nt-*nondeterministic* if h(X) contains two different words $\beta_1, \beta_2 \in (W(G))^+$.

We say that X is t-*nondeterministic* (nt-*nondeterministic*) if there exist a directly t-nondeterministic (directly nt-nondeterministic) letter Y such that $X \le Y$. We say that X is *nondeterministic* if it is either t-nondeterministic or nt-nondeterministic; otherwise X is *deterministic*.

Now through a (long) sequence of lemmas we will demonstrate that the assumption that G is unambiguous leads to a contradiction.

Lemma 3. Let θ be a finite alphabet and let b be a symbol not in θ. Let $K \subseteq \theta^+ b^+$ be an unambiguous EOL language such that

(i). there exists a growing function f: $N^+ \to N^+$ such that, for every $z \in K$, $|pres_\theta(z)| \leq f(|pres_b z|)$ and

(ii). $\{|pres_b(z)| : z \in K\}$ is numerically dispersed.

There exists an unambiguous EOL system $M = (\Sigma, h, S, \Delta)$ such that $\Delta = \theta \cup \{b\}$, $L(M) = K$, M is clean and if $\alpha \in h(S)$ then either $\alpha \in \Delta^+$ or $\alpha = AB$ where A, B $\in$ W(M) are such that $L(M_A) \subseteq \theta^+$ and $L(M_B) \subseteq b^+$.

Proof of Lemma 3.

Let H be an unambiguous clean EOL system such that L(H) = K; let $H = (\bar{\Sigma}, \bar{h}, S, \Delta)$ where Δ is as in the statement of the lemma.

Consider a symbol $C \in W(H)$; C is called b-*determined* if for every $m \in N^+$ there exists an $r \in N^+$ such that if $z \in h^m(C) \cap \theta^* b^+$, then $pres_b\, z = b^r$.

First we prove the following claim.

Claim 1. If $C \in W(H)$ such that $L(H_C) \cap \theta^* b^+ \neq \emptyset$, then C is b-determined.

Proof. Let C be as in the statement of the claim. The fact that C is b-determined is proved by contradiction as follows.

Assume that C is not b-determined. Consequently there exists a positive integer m and $z_1, z_2 \in h^m(C) \cap \theta^* b^+$ such that $pres_b\,(z_1) \neq pres_b\,(z_2)$. Without loss of generality assume that $pres_b\,(z_1) = b^{r_1}$ and $pres_b\,(z_2) = b^{r_2}$ where $r_1 > r_2 > 0$.

Since $\{|pres_b\,(z)| : z \in K\}$ is numerically dispersed, there exists a nonnegative integer n_{r_1} such that for every $m_1, m_2 \in \{|pres_b\,(z)| : z \in K\}$ if $n_{r_1} < m_1 < m_2$, then $m_2 - m_1 > r_1$. Since H is clean and $C \in W(H)$, $\{\beta \in S(H) : C \in alph\,(\beta)\}$ is infinite. Let $w \in S(H)$, $C \in alph\,(w)$ and $|w| > f(n_{r_1}) + n_{r_1}$. Then $w = x_1 C x_2$ where $x_1, x_2 \in W(H)^*$.

For i = 1, 2 let $y_i \in h^m(x_i) \cap \Delta^*$. Then $y_1 z_1 y_2 \in K$ and $y_1 z_2 y_2 \in K$. Moreover $|y_1 z_i y_2| > f(n_{r_1}) + n_{r_1}$. Hence $|pres_b (y_1 z_i y_2)| > n_{r_1}$ since otherwise $|y_1 z_i y_2| = |pres_\theta (y_1 z_i y_2)| + |pres_b (y_1 z_i y_2)| \le f(n_{r_1}) + n_{r_1}$. So we have $|pres_b (y_1 z_1 y_2)| > |pres_b (y_1 z_2 y_2)| > n_{r_1}$ and $|pres_b (y_1 z_1 y_2)| - |pres_b (y_1 z_2 y_2)| = r_1 - r_2 < r_1$, which contradicts (ii).

This ends the proof of Claim 1. □

Consider now all derivation trees in T_H. Given a derivation tree $T \in T_H$ of a word $z \in K$ we relabel all the nodes of it (*except* for its root and its leafs) in such a way that if the label of a node e is E and e contributes to z a subword in θ^+ then we change E to E_θ; if e contributes to z a subword in b^+ then we change E to E_b; if e constitutes to z a subword containing an occurrence of a letter from θ and an occurrence of b then we change E to $E_{\theta;b}$.

After we relabel all derivation trees in T_H, we get (in the obvious way) the set of (indexed) productions P_{in} corresponding to the way that indexed nodes of relabelled trees are rewritten. The productions in P_{in} are over the alphabet $V = \{S\} \cup \Delta \cup V_{\theta;b} \cup V_\theta \cup V_b \cup \{F\}$ where $V_{\theta;b} = \{E_{\theta;b}: E \in W(H)\}$, $V_\theta = \{E_\theta : E \in W(H)\}$, $V_b = \{E_b : E \in W(H)\}$.

Now let $\overline{V} = V \cup \overline{V}_\ell \cup \overline{V}_r$ where $\overline{V}_\ell = \{[Y, \ell]: Y \in V_{\theta;b}\}$ and $\overline{V}_r = \{[Y,r] : Y \in V_{\theta;b}\}$.

Based on P_{in} we construct two sets of productions R and $\overline{R}$ as follows.
If $S \to \alpha \in P_{in}$ is such that either $\alpha \in \Delta^+$ or $\alpha \in V_\theta^+ V_b^+$ then $S \to \alpha \in R$.
If $Y \to \alpha \in P_{in}$ where $Y \in V_\theta \cup V_b$ then $Y \to \alpha \in R$.
If $S \to \alpha \in P_{in}$ where $\alpha = \alpha_1 U \alpha_2$ with $\alpha_1 \in V_\theta^*$, $U \in V_{\theta;b}$ and $\alpha_2 \in V_b^*$ then $S \to \alpha_1[U,\ell]\, [U,r]\alpha_2 \in R$.
If $Y \to \alpha \in P_{in}$ where $Y \in V_{\theta;b}$ and $\alpha = \alpha_1 \alpha_2$ with $\alpha_1 \in V_\theta^+$, $\alpha_2 \in V_b^+$ or $\alpha_1 \in \theta^+$, $\alpha_2 \in b^+$ then $[Y,\ell] \to \alpha_1 \in R$ and $[Y,r] \to \alpha_2 \in \overline{R}$.
If $Y \to \alpha \in P_{in}$ where $Y \in V_{\theta;b}$ and $\alpha = \alpha_1 U \alpha_2$ with $\alpha_1 \in V_\theta^*$, $U \in V_{\theta;b}$, $\alpha_2 \in V_b^*$ then $[Y,\ell] \to \alpha_1\, [U,\ell] \in R$ and $[Y,r] \to [U,r]\alpha_2 \in \overline{R}$.

Now for each letter $X \in \overline{V}_r$ we choose one fixed (but arbitrary) production π_X from $\overline{R}$ of the form $X \to \alpha$ with $\alpha \in b^+$ and one fixed (but arbitrary) production $\overline{\pi}_X$ from $\overline{R}$ of the form $X \to \alpha$ with $\alpha \notin b^+$. Both productions (π_X and $\overline{\pi}_X$) are added to R; we also add to R productions of the form $x \to F$ where $x \in \Delta$. Moreover, the only productions in R are the productions specified as above.

Now we change R to the set of productions R_1 as follows.

(1). If $S \to \alpha_1 \alpha_2 \in R$ where $\alpha_1 \in V_\theta^+$ and $\alpha_2 \in V_b^+$, then we replace this production by three productions $S \to AB$, $A \to \alpha_1$, $B \to \alpha_2$, where A, B are two new symbols. We take care that the sets of symbols {A,B} used for different productions are pairwise disjoint.

(2). If $S \to \alpha_1[U,\ell][U,r]\alpha_2 \in R$ where $\alpha_1 \in V_\theta^*$, $\alpha_2 \in V_b^*$, $[U,\ell] \in \overline{V}_\ell$ and $[U,r] \in \overline{V}_r$, then we replace this production by three new productions $S \to AB$, $A \to \alpha_1[U,\ell]$ and $B \to [U,r]\alpha_2$ where A, B are two new symbols. We take care that the sets of symbols {A,B} used for different productions (accounting also for productions from (1) above) are pairwise disjoint.

(3). All other productions from R go unchanged to R_1.

Finally we set $M = (\Sigma, h, S, \Delta)$ where $\Sigma = \overline{V}$ and h is the finite substitution corresponding to productions in R_1. Clearly M satisfies the conclusion of the lemma. The equality $L(H) = L(M)$ is guaranteed by the construction of R_1 and by Claim 1 above. It is easily seen that M is unambiguous.

Thus Lemma 3 holds. □

Recall that $G = (\tilde{\Sigma}, \tilde{h}, \tilde{S}, \tilde{\Delta})$ is an unambiguous EOL system such that $L(G) = K_2$. By Lemma 3 we can assume that G satisfies the conclusion of Lemma 3 where $\theta = \{a\}$. Let $S \to AB$ be a production in G where A, $B \in W(G)$ are such that $L(G_A) \subseteq a^+$ and $L(G_B) \subseteq b^+$.

Lemma 4. Let $z \in S(G_B)\backslash L(G_B)$ and let $X \in alph(z)$. Then X is deterministic.

Proof of Lemma 4.

Assume to the contrary that X is nondeterministic. Hence for some directly nondeterministic letter Y we have $X \leq Y$.

(i). Assume that Y is t-nondeterministic.

Hence $b^r \in h(Y)$ and $b^s \in h(Y)$ for some $1 \leq r < s$. Since G is clean, for every $m \in N^+$ there exists a $n \in N^+$, $n > m$, such that $\alpha b^{2^n} \in K_2$ and $\beta b^{2^n+(s-r)} \in K_2$ where $\alpha, \beta \in \{a\}^+$. This clearly yields a contradiction.

(ii). Assume that Y is nt-nondeterministic.

Hence $\alpha, \beta \in h(Y)$, where $\alpha \neq \beta$, $\alpha, \beta \in (W(G))^+$. Since G is clean, for every $n \in N^+$ there exists a $m \in N^+$, $m > n$, and a derivation $D \in \mathcal{D}_G$ of a word w of length (m+2) such that on the level m of this derivation (at least one) Y occurs and it is rewritten by α. Since G is unambiguous, if we now change D in such a way that this one fixed occurrence of Y is rewritten by β (and on the level (m+1) letters from β are rewritten into a word from b^+, with all other letters on the $(m+1)^{th}$ level being rewritten as in D) then we get a derivation of a different word ω' such that $pres_b(\omega) \neq pres_b(\omega')$.

Thus for each $r \in N^+$ there exists a $s \in N^+$, $s > r$, such that $\alpha b^{2^s} \in K$ and $\alpha b^{2^s+q} \in K$ where $\alpha \in a^+$ and q is a constant dependent on G only. This clearly yields a contradiction.

From (i) and (ii) it follows that X must be a deterministic letter. □

If $S \to AB$ is a production in G where $A, B \in W(G)$ then we call B a *right letter* and A a *left letter*; R_G denotes the set of right letters in G and L_G denotes the set of left letters in G.

Thus Lemma 4 tells us that with each right letter B we can associate the DOL system G(B) and the HDOL systems $\hat{G}(B)$ as follows. G(B) includes all letters in W(G) that are reachable from B; if C is such a letter then the production for it in G(B) is $C \to \gamma$ where $\gamma \in h(C)$ and $\gamma \in (W(G))^+$. The HDOL system $\hat{G}(B)$ has G(B) as its underlying DOL system and the homomorphism g mapping L(G(B)) is defined by $g(C) = \gamma'$ if and only if $\gamma' \in b^+$ and $\gamma' \in h(C)$. By Lemma 4 both G(B) and $\hat{G}(B)$ are well defined.

The following result is very crucial in our further considerations. It establishes a lower and an upper bound on the length of a complete derivation D in terms of the length of $pres_b(res(D))$.

Lemma 5. There exist constants $n_0 \in N^+$ and $\varepsilon_1, \varepsilon_2 \in R^+$, $\varepsilon_1 \le 1$, $\varepsilon_1 < \varepsilon_2$ such that if $(w_0 = S, w_1 = A\,B, \ldots, w_{k+1})$ is a derivation in G where $k \ge 1$, A is a left letter, B is a right letter such that L(G(B)) is infinite, and $w_{k+1} \Rightarrow \gamma\, b^{2^n}$ for some $\gamma \in a^+$ and $n \ge n_0$, then $\varepsilon_1 n \le k \le \varepsilon_2 n$.

Proof.

Let $G(B) = (\Sigma_B, h_B, B)$. Since L(G(B)) is infinite and $L(\hat{G}(B)) \subseteq \{b^{2^n} : n \ge 1\}$, B is a letter without rank.

Hence (see, e.g., [ER1]) there exist integers $m_{B,1} \ge 0$ and $m_{B,2} > 0$ such that $B \overset{m_{B,1}}{\underset{G(B)}{\Rightarrow}} x_1\, \alpha\, x_2$ and $\alpha \overset{m_{B,2}}{\underset{G(B)}{\Rightarrow}} x_3\, \alpha\, x_4\, \alpha\, x_5$ where $\alpha \in \Sigma_B$ and $x_1\, x_2\, x_3\, x_4\, x_5 \in \Sigma_B^*$.

Let $E(G(B)) = \omega_0, \omega_1, \ldots$. Then clearly there exist $Q_{1B}, Q_{2B} \in R^+$, $Q_{1B} > Q_{2B} \ge 2$ such that for every nonnegative integer i and $0 \le j < m_{B,2}$,

$$Q_{2B}^{i} < |\omega_{m_{B,1}+im_{B,2}+j}| < Q_{1B}^{m_{B,1}+im_{B,2}+j} \qquad \text{or}$$

$$Q_{2B}^{\frac{-m_{B,1}-j}{m_{B,2}}} \left(Q_{2B}^{\frac{1}{m_{B,2}}}\right)^{m_{B,1}+im_{B,2}+j} < |\omega_{m_{B,1}+im_{B,2}+j}| < Q_{1B}^{m_{B,1}+im_{B,2}+j} .$$

Let $q_1 = \min\{Q_{2B}^{\frac{-m_{B,1}-j}{m_{B,2}}}$: B right letter, $L(G(B))$ infinite$\}$,

$q_2 = \min\{Q_{2B}^{\frac{1}{m_{B,2}}}$: B right letter, $L(G(B))$ infinite$\}$

and $q_3 = \max\{Q_{1B}$: B right letter, $L(G(B))$ infinite$\}$.

Then if we choose j such that $m_{B,1}+im_{B,2}+j = k$,
$q_1\, q_2^k \le |\omega_k| \le q_3^k$. Now let $r' = \min\{g(\alpha) : \alpha \in \Sigma_B\}$ and $s = \max\{g(\alpha) : \alpha \in \Sigma_B\}$.
Then $r'\, q_1\, q_2^k \le 2^n \le s\, q_3^k$ and consequently if we denote $r'\, q_1 = r$,
$\log_2 r + k \log_2 q_2 \le n \le \log_2 s + k \log_2 q_3$.

Since $\log_2 q_2 > 0$ and $\log_2 q_3 > 0$ we get

$$\frac{n - \log_2 s}{\log_2 q_3} \le \quad \le \frac{n - \log_2 r}{\log_2 q_2} \quad .$$

Consequently there exist constants $n_0 \in N^+$ and $\varepsilon_1, \varepsilon_2 \in R^+$, $\varepsilon_1 \le 1$, $\varepsilon_1 < \varepsilon_2$ such that for every $n \ge n_0$, $\varepsilon_1 n \le k \le \varepsilon_2 n$. □

Remark 1. We will often apply Lemma 5 in the sequel. To avoid unnessary technicalities we will assume that $n_0 = 1$ and that $L(G(B))$ is infinite for each $B \in R(G)$. Since the number of words $z \in L(G)$ such that
$|pres_b\,(z)| < 2^{n_0}$ or
$pres_b\,(z) \in \{w : w \in L(\hat{G}(B)),\ B \in R_G$ and $L(G(B))$ is finite$\}$, is finite, it is easily seen that such a simplifying assumption does not affect the validity of our proof of the theorem. □

We will analyze now EOL systems G_A where $A \in L_G$. So let $A \in L_G$ and let $Y \in W(G)$ be such that $A \le Y$.

If Y is a deterministic letter then for every Y' such that $Y \le Y'$ there is precisely one production of the form $Y' \to \alpha$ where $\alpha \in (W(G))^+$ and one production of the form $Y' \to \beta$ where $\beta \in a^+$. Thus, once again, we can associate with Y the (unique) DOL system $G(Y)$ and the (unique) HDOL system $\hat{G}(Y)$.

Lemma 6. G(Y) is a DOL system with rank and moreover *rank*(G(Y)) ≤ 2.

Proof of Lemma 6.

From the form of words in K_2 it follows that G(Y) is a DOL system with rank.

Assume that *rank*(G(Y)) ≥ 3.

Thus there exist constants $p, q \in R$, $p > 0$ such that if Y derives in G(Y) a word α in k steps, then $|\alpha| \geq pk^3 + q$.

Let ℓ be an integer such that we have $S \Rightarrow AB \overset{\ell}{\Rightarrow} \gamma_1 Y \gamma_2 \beta \overset{k}{\Rightarrow} \gamma_1' \alpha \gamma_2' \beta' \Rightarrow \gamma b^{2^n}$ for some $A \in L_G$, $B \in R_G$, $\gamma_1, \gamma_2, \gamma_1', \gamma_2', \beta, \beta' \in \Sigma^*$, $n \geq 1$ and $\gamma \in a^+$.

Thus by Lemma 5, $\varepsilon_1 n \leq \ell + k \leq \varepsilon_2 n$, where $\varepsilon_1, \varepsilon_2 \in R^+$ are constants dependent on G only. Hence $k \geq \varepsilon_1 n - \ell$ and consequently

$|\gamma| \geq p(\varepsilon_1 n - \ell)^3 + q$ ....(1)

But from the definition of K_2 it follows that

$|\gamma| \leq n^2$..(2)

Since (1) and (2) must hold for arbitrary long γ's, we get a contradiction. Consequently it must be that *rank*(G(Y)) ≤ 2 and the lemma holds. □

Hence all deterministic nonterminals reachable from nonterminals in L_G have associated DOL systems either of rank 0 or of rank 1 or of rank 2.

Remark 2. Notice that the above conclusion also holds if we consider a nondeterministic nonterminal (reachable from a letter in L_G), where we choose for it, and for each nonterminal reachable from it, one arbitrary but fixed production with its right-hand side consisting of nonterminals. In this way we have "selected" a DOL system for the nonterminal considered. □

We will analyze now nondeterministic nonterminals reachable from letters in L_G.

Lemma 7. Let T be a nondeterministic letter reachable from a letter in L_G. If $z \in S(G)$ is such that $T \in alph(z)$ then z contains exactly one occurrence of T.

Proof of Lemma 7.

Assume to the contrary that there exists a word $z \in S(G)$ such that z contains two occurrences of T. Let X be a directly nondeterministic letter reachable from T. Hence there exists a word $z' \in S(G)$ such that z' contains two occurrences of X.

We consider separately two cases.

(i). X is directly t-nondeterministic. Hence there exist $\alpha_1, \alpha_2 \in a^+$, $\alpha_1 \neq \alpha_2$, such that $X \Rightarrow \alpha_1$ and $X \Rightarrow \alpha_2$. We consider then two ways of rewriting z' in a terminal word. In one way the leftmost occurrence of X in z' is rewritten by α_1 and the rightmost occurrence of X in z' is rewritten by α_2; in the other way the leftmost occurrence of X in z' is rewritten by α_2 and the rightmost occurrence of X in z' is rewritten by α_1; all other occurrences of all letters are rewritten in the same way in both cases.

Consequently we get two different derivation trees of the same word in L(G); this contradicts the fact that G is unambiguous.

(ii). X is directly nt-nondeterministic. The reasoning is analogous to the one above except that there is a step in-between z' and a terminal word.

Thus the lemma holds. □

We will demonstrate now that one can assume that G satisfies also the following condition:

each element of $\mathcal{S}(G)$ contains at most one occurrence of one letter that is nondeterministic ....(3)

From Lemma 7 we know that if $z \in \mathcal{S}(G)$ contains an occurrence of a nondeterministic letter then z contains precisely one occurrence of this letter; consequently z contains no more than a bounded number of occurrences of nondeterministic letters. Each of these occurrences is reachable from (an occurrence of) a letter in L_G and (consequently) each of them will (eventually) contribute a (sub) word from a^+ if z is considered to be a word in a specific derivation. The key observation now is that z can be written in the form $z = z_1 z_2$ where z_1 consists of (occurrences of) letters reachable from a letter in L_G and z_2 consists of (occurrences of) letters reachable from a letter in R_G. But if we consider z to be a word used in a specific derivation of a word in L(G) then permuting (occurrences of) letters in z_1 with the fixed application of productions attached to (occurrences of) letters being permuted (and to their descendants) does not affect the final result of a derivation (which is a word in L(G))!!!

Formally this observation leads us to the following transformation of G.

For $z \in \mathcal{S}(G)$ we define its type, denoted *type*(z), to be the subset of *alph*(z) consisting of all nondeterministic letters in *alph*(z). Clearly the number of different types of words in $\mathcal{S}(G)$ is finite. Let

TYPE = $\{W \subseteq \Sigma : W = type(z) \text{ for a } z \in \mathcal{S}(G)\}$

and let

$\Pi = \{Z_t : t \in \text{TYPE}\}$

be a new set of letters.

Let $A \in L_G$.

Each production from G of the form

$A \to \alpha_0 N_1 \alpha_1 N_2 \dots N_m \alpha_m$

where $m \geq 1$, $N_1, \dots, N_m$ are nondeterministic letters and $\alpha_0, \alpha_1, \dots, \alpha_m \in \Sigma^*$ do not include nondeterministic letters is replaced by the production

$A \to \alpha_0 \alpha_1 \dots \alpha_m Z_t$ where $t = \{N_1, \dots, N_m\}$.

All other productions from G are not changed.

For symbols in Π we add the following productions.

If $t = \{N_1, \dots, N_q\}$, where $N_1, \dots, N_q$ are nondeterministic letters and, for $1 \leq i \leq q$,

$N_i \to \alpha_{i0} N_{i1} \alpha_{i1} N_{i2} \dots N_{im_i} \alpha_{im_i}$

is a production in G, where $m_i \geq 0$, $\alpha_{i0}, \dots, \alpha_{im} \in \Sigma^*$ do not contain nondeterministic letters and $N_{i1}, \dots, N_{im_i}$ are nondeterministic letters, then we add the production

$Z_t \to \alpha_{10} \dots \alpha_{1m_1} \alpha_{20} \dots \alpha_{2m_2} \dots \alpha_{q0} \dots \alpha_{qm_q} Z_{\overline{t}}$

where $\overline{t} = \{N_{11}, \dots, N_{1m_1}, \dots, N_{q1}, \dots, N_{qm_q}\}$

(and we set formally $Z_{\overline{t}} = \Lambda$ if $\overline{t} = \emptyset$).

Lemma 7 and the "permutational property" discussed above guarantee that the so obtained EOL system is equivalent to G. Obviously the so constructed EOL system has all the properties that we have required so far from G and additionally it has the property (3).

To avoid a cumbersome notation, rather than to consider the new system constructed above, we simply go on analyzing G *but we assume from now on that* G *satisfies* (3).

Let M be a nondeterministic recursive letter. A M-derivation $(\gamma_0 = M, \gamma_1, \dots, \gamma_m)$ such that $m \geq 1$, $M \notin alph(\gamma_i)$ for $1 \leq i \leq m-1$ and $M \in alph(\gamma_m)$ is called *elementary*.

Lemma 8. Let M be a nondeterministic recursive letter. There exists precisely one elementary M-derivation

Proof of Lemma 8.

We prove the lemma by contradiction.

Assume that there exist two different elementary M-derivations: $(\gamma_0 = M, \gamma_1, \ldots, \gamma_{m_1})$ and $(\overline{\gamma}_0 = M, \overline{\gamma}_1, \ldots, \overline{\gamma}_{m_2})$.
Let us consider the derivation
$(\delta_0 = S, \delta_1, \delta_2, \ldots, \delta_{k_0+1}, \delta_{k_0+2}, \ldots, \delta_{k_0+k+1})$
where $k_0 \geq 0$, $M \notin alph(\delta_i)$ for $1 \leq i \leq k_0$, $M \in alph(\delta_{k_0+1})$, $k > m = \max\{m_1, m_2\}$ and $\delta_{k_0+k+1} \Rightarrow \gamma b^{2^n}$ for some $n \geq 1$, $\gamma \in a^+$.

Thus there exist at least $2^{\frac{k}{m}}$ different derivations in G of words of the form γb^{2^n}, where $\gamma \in a^+$. But Lemma 5 implies that $k_0 + k \geq \varepsilon_1 n$ where $\varepsilon_1 \in R^+$ is a constant dependent on G only. Consequently there are at least $2^{\frac{1}{m}(\varepsilon_1 n - k_0)}$ different derivations in G of words of the form γb^{2^n}, where $\gamma \in a^+$. Since K_2 contains no more than 2n different words x with the property $pres_b(x) = 2^n$, for n big enough we get several different derivation trees of the same word, which contradicts the fact that G is unambiguous.

Consequently, there exists precisely one elementary M-derivation. □

Let M be a nondeterministic recursive letter and let us consider the unique elementary M-derivation, $(M, \gamma_1, \ldots, \gamma_m)$, denoted *elem*(M). We know that $\gamma_m = \gamma_{m1} M \gamma_{m2}$ where γ_{m1}, γ_{m2} do not contain nondeterministic letters. Then we can write $\gamma_{m-1} = \gamma_{(m-1)1} M_{m-1} \gamma_{(m-1)2}$, $\gamma_{m-2} = \gamma_{(m-2)1} M_{m-2} \gamma_{(m-2)2}$, $\ldots$, $\gamma_1 = \gamma_{11} M_1 \gamma_{12}$, where M_{m-1} is the ancestor of M in

γ_m, M_{m-2} is the ancestor of M_{m-1} in γ_{m-1}, ..., M_1 is the ancestor of M_2 in γ_2. The sequence $M_0 = M, M_1, \ldots, M_{m-1}$ is referred to as the *elementary cycle* (of M).

The following result is obvious.

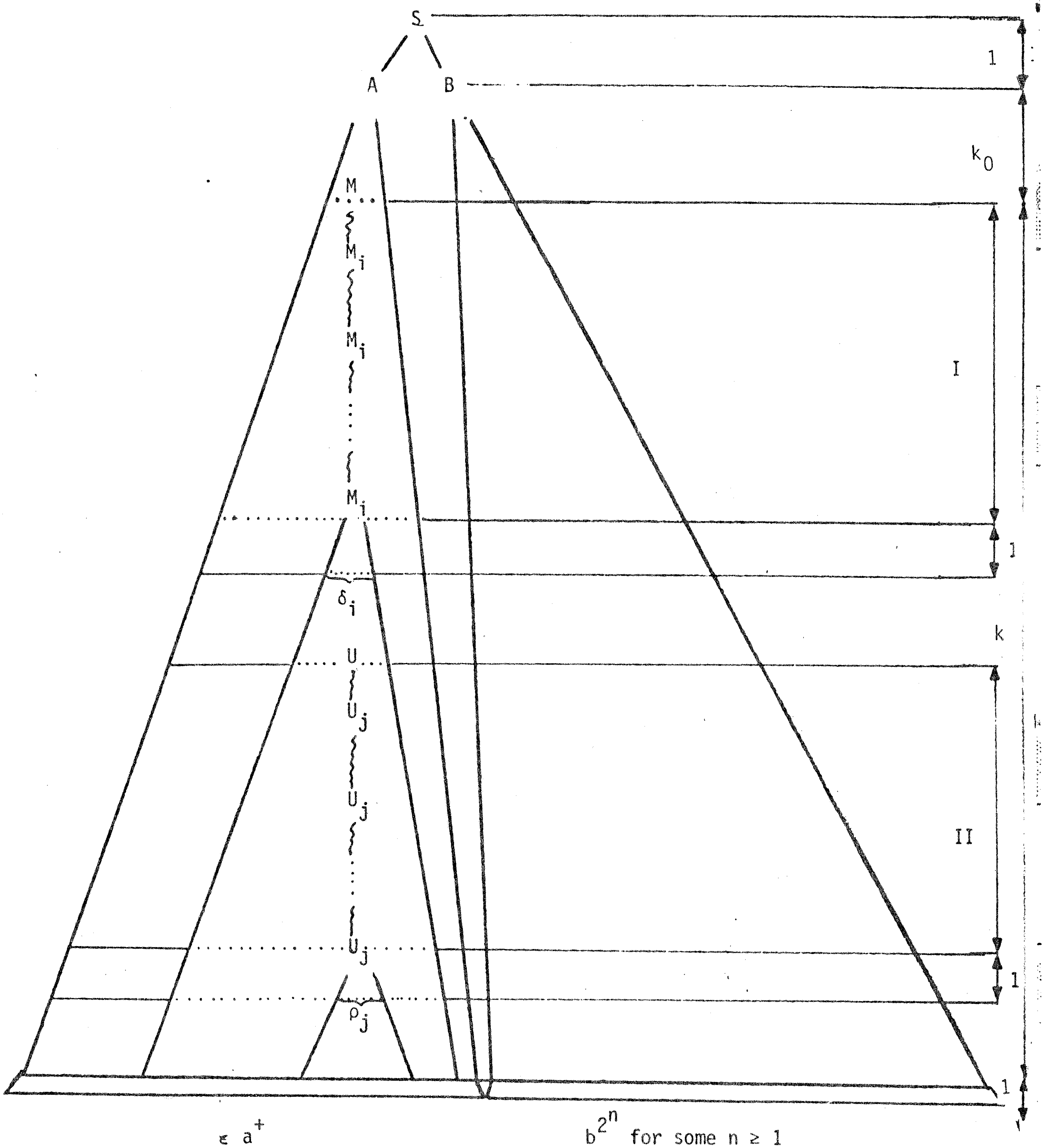
Lemma 9. If M is a nt-nondeterministic recursive letter then the elementary cycle of M contains a directly nt-nondeterministic recursive letter. □

Let M be a nt-nondeterministic recursive letter and let $M_0 = M, M_1, \ldots, M_{m-1}$ be the elementary cycle of M. Let $0 \le i \le m-1$ be such that M_i is a directly nt-nondeterministic recursive letter (by Lemma 9 we know that such an i exists). Thus M_i has at least two productions in G with the right-hand sides consisting of nonterminals from W(G): one of these is the production used in *elem*(M), it is of the form $M_i \to \alpha_i M_{i+1} \beta_i$ (where for $i = m-1$, $i+1$ is set to 0) and the other one is of the form $M_i \to \delta_i$ for some $\delta_i \in (W(G))^+$. Let $NT(\delta_i)$ be the set of nonterminal letters reachable from (the letters in) δ_i.

Lemma 10. $NT(\delta_i)$ does not contain recursive nt-nondeterministic letters.

Proof of Lemma 10.

Assume to the contrary that $NT(\delta_i)$ contains a recursive nt-nondeterministic letter, say $U = U_0$ Let the elementary cycle of U be $U_0, U_1, \ldots, U_{r-1}$ and let $0 \le j \le r-1$ be such that U_j is directly nt-nondeterministic; let $U_j \to \rho_j$ be a production that is not used by U_j in *elem*(U). Consider then a derivation depicted by the following derivation tree.


S
A
B
1
k_0
M
M_i
M_i
M_i
I
1
δ_i
k
U
U_j
U_j
U_j
II
1
ρ_j
1
$\varepsilon\ a^+$
b^{2^n} for some $n \geq 1$

Part I of the derivation starts at the sentential form where M is for the first time derived (in k_0 steps from A); it ends at the sentential form containing M_i which will be rewritten (to get the next sentential form) using production $M_i \to \delta_i$. Part II of the derivation starts at the sentential form where U is for the first time derived (from δ_i); it ends at the sentential form containing U_j which will be rewritten (to get the next sentential form) using production $U_j \to \rho_j$.

The whole derivation is of length $k_0 + k + 2$, where part I is of length at least $\frac{k}{3}$ and part II is of length at least $\frac{k}{3}$.

Now we can modify this derivation (within its subtree rooted at A) as follows.

When the production $M_i \to \delta_i$ is used, we say that we *exit* the elementary cycle of M and when the production $U_j \to \rho_j$ is used we say that we *exit* the elementary cycle of U. Thus within the part I we can exit the elementary cycle of M on the 1st or 2nd or ... or($\left\lfloor \frac{k}{3m} \right\rfloor$)th occurrence of M_i; similarly within part II we can exit the elementary cycle of U on the 1st or 2nd or ... or($\left\lfloor \frac{k}{3r} \right\rfloor$)th occurrence of U_j. (Within these changes "the rest of the derivation" remains intact.)

In this way we get at least $\left\lfloor \frac{k}{3m} \right\rfloor \cdot \left\lfloor \frac{k}{3r} \right\rfloor$ different derivations (of the same length) of words x such that $pres_b(x) = b^{2^n}$. By Lemma 5 we know that $k^2 \geq \varepsilon_1^2 n^2$ for some constant $\varepsilon_1 \in R^+$ and consequently we get at least $s\,n^2$ different derivation trees of words x such that $pres_b(x) = b^{2^n}$ (where s is a constant dependent on G only). Since we may take n arbitrary large and since K_2 contains at most $2n$ words x such that $pres_b(x) = b^{2^n}$, G must be ambiguous; a contradiction.

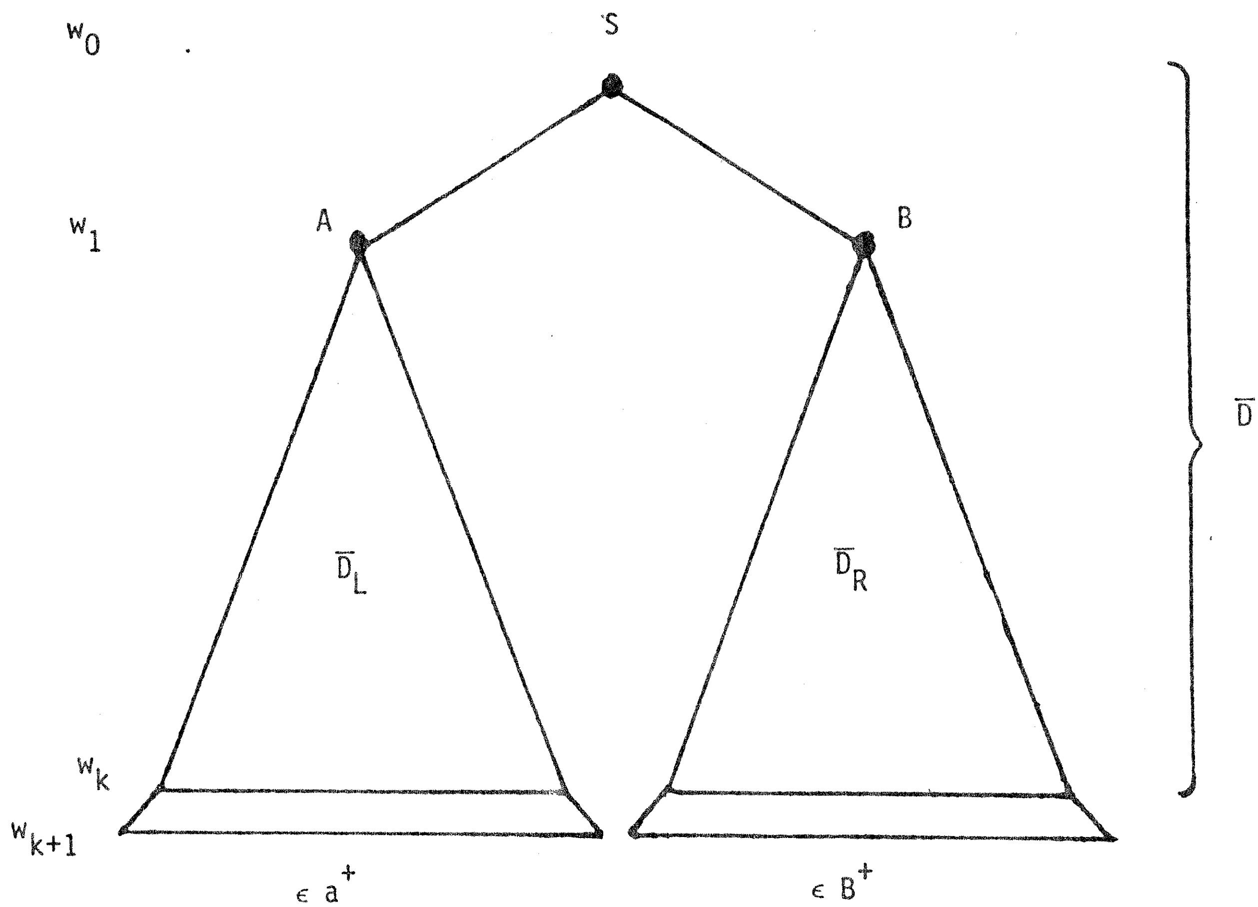
Consequently $NT(\delta_i)$ does not contain recursive nt-nondeterministic letters.

Hence the lemma holds. □

Let us summarize now what we know already about the structure of derivations in G.

Consider a derivation D of a terminal word z from S where D is of length at least two, $D = (w_0, w_1, \dots, w_{k+1})$, $k \geq 1$, $w_0 = S$, $w_{k+1} = z$. Thus $w_1 = AB$ where $A \in L_G$ and $B \in R_G$. The last step of D ($w_k \Rightarrow w_{k+1}$) is a finite substitution into (subsets of) $\{a,b\}^+$. In our classification of derivations in G we will ignore this final step and so we consider the derivation $\bar{D} = (w_0, \dots, w_k)$.

Thus, except for the first step ($S \Rightarrow AB$), $\bar{D}$ consists of two derivations $\bar{D}_L$ and $\bar{D}_R$ "running in parallel"; $\bar{D}_L$ is the derivation originating in A and $\bar{D}_R$ is the derivation originating in B. The situation may be illustrated as follows:



w_0
S
w_1
A
B
$\overline{D}_L$
$\overline{D}_R$
$\overline{D}$
w_k
w_{k+1}
$\in a^+$
$\in B^+$

$\overline{D}_R$ is a DOL derivation, it is a derivation in the DOL system G(B) (without rank).

In considering $\overline{D}_L$ we distinguish several cases.

(I). Only deterministic letters occur in (words of) $\overline{D}_L$. Then $\overline{D}_L$ is a derivation in a DOL system with rank and the rank of the system is not bigger than 2. If the DOL system corresponding to $\overline{D}_L$ is of rank i, $0 \le i \le 2$, then we say that D is of *type* I i.

(II). Nondeterministic letters occur in $\overline{D}_L$. We consider here separately two cases.

(IIa). The derivation tree $T(\overline{D}_L)$ corresponding to $\overline{D}_L$ is deterministic. Then the situation is as in (I): $\overline{D}_L$ is a derivation in a DOL system with rank and the rank of this system is not bigger than 2. If the DOL system corresponding to $\overline{D}_L$ is of rank i, $0 \le i \le 2$, then we say that D is of *type* IIi.

(IIb). The derivation tree $T(\overline{D}_L)$ is nondeterministic. Hence on a path of $T(\overline{D}_L)$ we have (possibly repeating) the elementary cycle of a recursive nt-nondeterministic letter, say M, from which the exit is taken at some point (that is a production leading out of the cycle is applied to a directly nondeterministic letter from the cycle). From this moment on the tree $T(\overline{D}_L)$ is deterministic. As a matter of fact we have the following situation. $\overline{D}_L$ is the "superposition" of two derivations $\overline{D}_L^{(1)}$ and $\overline{D}_L^{(2)}$. $\overline{D}_L^{(1)}$ is a derivation in a DOL system with rank where the rank of the system is not bigger than 2. Also $\overline{D}_L^{(2)}$ is a derivation in a (different) DOL system with rank where the rank of the system is not bigger than 2. If the DOL system corresponding to $\overline{D}_L^{(1)}$ is of rank i and the DOL system corresponding to $\overline{D}_L^{(2)}$ is of rank j, $0 \le i, j \le 2$, then we say that $\overline{D}_L$ (and also D) is of *type* (i,j).

Clearly there is only a finite number of DOL systems that (either directly or by their superposition) generate all (types of) derivations discussed above.

We have classified now all complete derivations in G(for the sake of completeness, let one step complete derivations be derivations of *type* 0).

Given a type X (of a derivation) we use C(X) to denote the set of all words in L(G) that have a complete derivation of type X. Hence

$$L(G) = C(0) \cup \bigcup_{i=0}^{2} C(Ii) \cup \bigcup_{i=0}^{2} C(IIi) \cup \bigcup_{i=0}^{2}\left(\bigcup_{j=0}^{2} C((i,j))\right).$$

We will also use the following notation:

$$\bigcup_{i=0}^{2} C(Ii) = C(I) \text{ and } \bigcup_{i=0}^{2} C(II\,i) = C(II);$$

also we write C(i,j) rather than C((i,j)).

Since G is assumed to be unambiguous, $C(X) \cap C(Y) = \emptyset$ if $X \neq Y$. Also, it is clear that a derivation of type (2,0) cannot exist and so $C(2,0) = \emptyset$.

Lemma 11. There exist constants $p, r \in N^+$ and $q \in R^+$ such that $Z \subseteq C(0,1) \cup C(1,0)$, where $Z = \{a^m b^{2^n} : p \leq m \leq q\,n \text{ and } n \geq r\}$.

Proof of Lemma 11.

In considering how Z (which will be constructed "on line") is generated by G we will eliminate systematically all C(X) except for $X \in \{(0,1),(1,0)\}$.

Clearly there exists $p_1 \in N^+$ such that if $a^m b^{2^n} \in C(0) \cup C(I0)$ then $m \leq p_1$.

If $w = a^m b^{2^n} \in C(I1) \cup C(I2)$ and w is derived in $(k+2)$ steps in G then $m > \alpha_1 k$ where $\alpha_1 \in R^+$ is a constant dependent on G only. Thus, by Lemma 5, $m > \alpha_2 n$ for some constant $\alpha_2 \in R^+$ dependent on G only.

Hence indeed, there exist $p_1, r_1 \in N^+$ and $q_1 \in R^+$ such that $\{a^m b^{2^n} : p_1 \le m \le q_1 n \text{ and } n \ge r\} \cap (C(0) \cup C(I)) = \emptyset$.

By similar arguments we eliminate $C(II)$, $C(0,0)$, $C(1,1)$, $C(1,2)$, $C(2,1)$ and $C(2,2)$; that is we demonstrate that there exist $\overline{p}, \overline{r} \in N^+$ and $\overline{q} \in R^+$ such that $\overline{Z} \cap E = \emptyset$, where

$\overline{Z} = \{a^m b^{2^n} : \overline{p} \le m \le \overline{q} n \text{ and } n \ge \overline{r}\}$ and

$E = C(0) \cup C(I) \cup C(II) \cup C(0,0) \cup C(1,1) \cup C(1,2) \cup C(2,1) \cup C(2,2)$.

To eliminate $C(0,2)$ we will demonstrate that if it is not true that $\overline{Z} \subseteq C(0,1) \cup C(1,0)$ then it is also not true that $\overline{Z} \subseteq C(0,1) \cup C(1,0) \cup C(0,2)$. To this aim we proceed as follows.

First of all we can assume that if it is not true that $\overline{Z} \subseteq C(0,1) \cup C(1,0)$ then it is not the case that $\overline{Z} \setminus (C(0,1) \cup C(1,0))$ is finite. (Otherwise we adjust parameters $\overline{p}$, $\overline{q}$ and $\overline{r}$ and obtain $\overline{\overline{Z}}$ such that $\overline{\overline{Z}} \subseteq C(0,1) \cup C(1,0)$.)

Let $\overline{M} = \{m : a^m b^{2^n} \in \overline{Z}$ for some $n\}$ and

$M' = \{m : a^m b^{2^n} \in C(0,1) \cup C(1,0)$ for some $n\}$.

Observe that for sufficient big n, if $a^m b^{2^n} \in C(0,1) \cup C(1,0)$, $m \ge \overline{p}$, then by "pumping in the DOL system of rank 0" also $a^m b^{2^{n_1}} \in C(0,1) \cup C(1,0)$ where $\overline{q}\, n_1 \ge m$.

Also observe that using "pumping in the DOL system of rank 1" one can prove the existence of a positive integer n_2 such that if $m \in M'$, $m \ge n_2$, then also $m + s \in M'$.

The above two observations imply that $\overline{M} \setminus M'$ contains (at least one) infinite arithmetic progression.

Clearly, for every $s \in R^+$ there exists a $t \in R^+$ such that, for every $n \geq \overline{r}$,

$$\#\{m : a^m b^{2^n} \in C(0,2) \text{ and } m < s\} \leq t\sqrt{s}$$

and consequently, for every $n \geq \overline{r}$,

$$\#\{m : a^m b^{2^n} \in C(0,2) \quad \text{and } m < \overline{q}n\} \leq \overline{t}\sqrt{\overline{q}n}$$

for some $\overline{t} \in R^+$.

Thus on the one hand we know that, for each $n \geq \overline{r}$, the number of elements in the set

$$\hat{Z} = \{m : a^m b^{2^n} \in \overline{Z}\} \setminus \{m : a^m b^{2^n} \in C(0,1) \cup C(1,0)\}$$

is at least $t'\overline{q}n$ for some positive real constant t' dependent on G only.

On the other hand we know that, for each $n \geq \overline{r}$, the number of elements in $C(0,2) \cap \hat{Z}$ is not larger than $\overline{t}\sqrt{\overline{q}n}$. Since for n large enough $t'\overline{q}n > \overline{t}\sqrt{\overline{q}n}$, we have proved that if it is not true that $\overline{Z} \subseteq C(0,1) \cup C(1,0)$ then it is also not true that $\overline{Z} \subseteq C(0,1) \cup C(1,0) \cup C(0,2)$.

Consequently we have "eliminated" $C(0,2)$ and the lemma holds. □

Lemma 12. $(C(0,2) \cup C(1,2) \cup C(2,1)$ is infinite.

Proof of Lemma 12.

(1). It is easily seen (using Lemma 5) that if $z = a^m b^{2^n}$ and $z \in C(I0) \cup C(I1) \cup C(II0) \cup C(II1) \cup C(0,0) \cup C(0,1) \cup C(1,0) \cup C(1,1)$,

then $m \leq \rho n$ for some constant $\rho \in N^+$ dependent on G only.

(2). If $z = a^m b^{2^n} \in C(I2) \cup C(II2) \cup C(2,2)$ then, again using Lemma 5 it is easily seen that there exist $\overline{n} \in N^+$ and $\overline{\rho} \in R^+$ such that for $n \geq \overline{n}$, $m \geq \overline{\rho}\, n^2$.

(3). $C(0)$ contains only a finite number of words.

Thus if $n \geq \bar{n}$ and $\rho n < \pi < \bar{\rho} n^2$, then, with perhaps a finite number of exceptions,

$a^{\pi^2} b^{2^n} \notin (C(I) \cup C(II) \cup C(0,0) \cup C(0,1) \cup C(1,0) \cup C(1,1) \cup C(2,2))$.

Thus the lemma holds. □

Lemma 13. Each of the languages: $C(0,2)$, $C(1,2)$ and $C(2,1)$ is finite.

Proof.

We will separately consider each of the three cases.

(1). $C(0,2)$.

A derivation of type (0,2) looks as follows

- the first step is $S \Rightarrow AB$ where $A \in L_G$ and $B \in R_G$,
- then in the part of the derivation originating in A we have
- $k_1 \geq 1$ steps of rewriting in a DOL system G_1 of rank 0,
- $k_2 \geq 1$ steps of rewriting in a DOL system G_2 of rank 2, and
- the final derivation step yielding a word $a^{m_0} b^{2^{n_0}}$ in $L(G)$.

By Lemma 11 we know that there exist constants $p, r \in N^+$ and $q \in R^+$ such that

$Z = \{a^m b^{2^n} : p \leq m \leq qn \text{ and } n \geq r\} \subseteq C(0,1) \cup C(1,0)$.

But for n big enough, $qn - p$ is also big enough so that by taking k_1 big enough we can generate a word $a^{m_1} b^{2^{n_1}}$ where $p \leq m_1 \leq qn_1$. Then however, $a^{m_1} b^{2^{n_1}} \in C(0,2)$ and $a^{m_1} b^{2^{n_1}} \in C(0,1) \cup C(1,0)$; a contradiction to our assumption that G is unambiguous.

Hence $C(0,2)$ must be finite.

(2). $C(1,2)$.

A derivation of type (1,2) looks as follows

- the first step is $S \Rightarrow A B$ where $A \in L_G$ and $B \in R_G$,
- then in the part of the derivation originating in A we have
- $k_1 \geq 1$ steps of rewriting in a DOL system G_1 of rank 1,
- $k_2 \geq 1$ steps of rewriting in a DOL system G_2 of rank 2, and
- the final derivation step yielding the word $a^m b^{2^n}$ for some $m, n \geq 1$.

Let $G_1 = (V_1, f_1, \alpha_1)$ and $G_2 = (V_2, f_2, \alpha_2)$. Clearly the following properties hold.

(1) There exists a positive integer r depending on G_2 only such that $h^k(\alpha) \geq r k^2$ for each $k > 0$ and α a symbol of rank 2 of G_2.

(2) There exists a positive integer p depending on G_1 only such that if $\alpha \in V_1$ where α is of rank i, $i \in \{0, 1\}$ and $\alpha \overset{j_1 p}{\Rightarrow} x_1$, $\alpha \overset{j_2 p}{\Rightarrow} x_2$, then $pres_{V_{1,i}}(x_1) = pres_{V_{1,i}}(x_2)$ where $V_{1,i}$ denotes the set of all symbols of rank i of G_1.

Choose k_1 and k_2 to be multiples of p such that

$$k_1 > p\,(\# V_1 + 4) \text{ and } r\,k_2^2 > \frac{k_1 + k_2 + 3\,\#\,V_1\,p}{\varepsilon_1} \quad (\varepsilon_1 \text{ as in Lemma 5}).$$

Now k_1 is big enough such that on a path of the corresponding derivation tree there is a node on levels $l_1\,p$ and $l_2\,p$, $1 \leq l_1 < l_2 \leq p\,(\# V_1 + 1)$ labelled by M such that M is a (recursive) letter of rank 1. Repeating such a "cycle" (with the rest of the derivation remaining "the same") once again, then twice and then three times, we get three new derivations in G deriving words $a^{m_1} b^{2^{n_1}}$, $a^{m_2} b^{2^{n_2}}$ and $a^{m_3} b^{2^{n_3}}$.

Observe that after level p and hence after level $l_1\,p$ the number of symbols of rank 1 of G_1 cannot grow any more. The above observation, (1) and the fact that $k_1 > l_2\,p + 3\,p$ yield that m, m_1, m_2 and m_3 form an arithmetic progression.

The choice of k_2 together with (2) yield

$$m, m_1, m_2, m_3 > \frac{k_1 + k_2 + 3 (\# V_1) p}{\varepsilon_1} \quad \text{and}$$

$$n, n_1, n_2, n_3 \leq \frac{k_1 + k_2 + 3 (\# V_1) p}{\varepsilon_1} .$$

Consequently m, m_1, m_2 and m_3 are all squares. This, however, contradicts the well known fact from number theory (see, e.g., [D] p. 404) that in the set of squares there is no arithmetic progression of length larger than 3.

Consequently $C(1,2)$ must be finite. □

(3). $C(2,1)$.

This case is proved analogously to case (2).

Hence the lemma holds. □

However, Lemma 13 contradicts Lemma 12 and consequently our assumption that there exists an unambiguous EOL system G generating K_2 is false.

Thus K_2 is inherently ambiguous and the theorem holds. □

ACKNOWLEDGEMENTS

The first and the second author gratefully acknowledge the financial support of NSF grant number MCS79-03838. The second and the third author are indebted to NFWO Belgium for supporting their research.

REFERENCES

[D] Dickson, L. E., 1944, *History of the theory of numbers*, vol. II, Chelsea Publishing Company.

[ER1] Ehrenfeucht, A. and Rozenberg, G., 1979, On the structure of polynomially bounded DOL systems, *Fundamenta Informaticae*, vol. 2, 187-197.

[ER2] Ehrenfeucht, A. and Rozenberg, G., 1980, On ambiguity in EOL systems, *Theoretical Computer Science*, vol. 12, no. 2, 127-134.

[H] Harrison, M., 1978, *Introduction to formal language theory*, Addison-Wesley, Reading, Massachusetts.

[MSW] Maurer, H., Salomaa, A. and Wood, D., 1980, Synchronized EOL forms, *Theoretical Computer Science*, vol. 12, 135-159.

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

[ReS] Reedy, A. and Savitch, W., 1975, Ambiguity in the developmental systems of Lindenmayer, *Journal of Computer and Systems Science*, v. 11, 262-283.

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