INFORMATION AND CONTROL 36, 309–319 (1978)

Propagating Chain-Free Normal Forms for EOL Systems

K. CULIK II

Department of Computer Science, University of Waterloo, Canada

AND

H. A. MAURER

Institut f. Informationsverarbeitung, Technische Universität Graz, A-8010 Graz, Austria

We establish two types of normal forms for EOL systems. We first show that each ϵ-free EOL language can be generated by a propagating EOL system in which each derivation tree is chain-free. By this we mean that it contains at least one path from the root to the grandfather of a leaf in which each node has more than one son. We use this result to prove that each ϵ-free EOL language can be generated by a propagating EOL system in which each production has a right side of length at most two and which does not contain nonterminal chain-productions, i.e., productions $A \to B$ for nonterminals A and B. As applications of our results we give a simple proof for the decidability of the finiteness problem for EOL systems and solve an open problem concerning completeness of EOL forms.

1. INTRODUCTION AND PRELIMINARIES

In this paper we are concerned with new normal forms for EOL systems and ramifications thereof.

Obtaining simple normal forms is an important aspect of the study of grammars. While a number of important normal form results for EOL systems belong now to standard knowledge (Herman and Rozenberg, 1975) other results valid for context-free grammars do not hold at all for EOL systems or have not yet been established. A typical example is the well-known Chomsky normal form (Salomaa, 1973) for context-free grammars. Every ϵ-free context-free language can be generated by a context-free grammar whose only productions are of the form $A \to a$, $A \to BC$ (A, B, C nonterminals, a is a terminal). Thus, in particular, "nonterminal chain-productions" $A \to B$ (A, B nonterminals) are not required. A similar result for synchronized EOL systems does clearly not hold. With productions just of the form, say, $A \to a$, $A \to BC$ (A, B, C and a as above, every terminal "blocking") many EOL languages cannot be generated (Maurer, Salomaa and Wood, 1977). Thus, nonterminal chain productions are

0019-9958/78/0363–0309$02.00/0

required in synchronized EOL systems to be able to generate all EOL languages. Further, while a number of "expansive" normal-forms are known for context-free grammars (i.e. normal forms assuring that the lengths of words in each derivation form a strictly growing sequence except when terminals are introduced)—a typical example is the Chomsky normal form—no such results have been available for EOL systems, sofar.

In this paper we show that a normal form even stronger than expansive can be obtained for EOL systems. In particular, we establish that each ϵ-free EOL language can be generated by a propagating EOL system in which each derivation tree is what we would like to call "chain-free", i.e., contains a path from the root to the grandfather of a leaf along which each node has at least two sons. (Note that any such chain-free normal form is certainly expansive).

Using the result on chain-free normal forms we also establish two normal forms showing that nonterminal chain-productions (i.e. productions $A \rightarrow B$ for nonterminals A, B) are not required to generate all EOL languages, if one does not insist on synchronization.

We finally give two applications of our normal form results. We present short direct proofs for the (known) decidability of emptiness and finiteness of the languages generated by EOL systems (emptiness also for ETOL systems) and we solve an open problem concerning complete EOL forms by establishing that the form with productions $S \rightarrow a$, $S \rightarrow aS$, $S \rightarrow Sa$, $a \rightarrow a$, $a \rightarrow S$, $a \rightarrow SS$ (S nonterminal, a terminal) is complete.

In the remainder of this section we present some standard definitions from language theory. For any notions not explicitly defined we refer to (Salomaa, 1973) and, as far as EOL forms are concerned, to (Maurer, Salomaa, Wood, 1977).

An EOL *system* G is a quadruple $G = (N, T, P, S)$. N and T are disjoint alphabets of *nonterminals* and *terminals*, respectively, S in N is called the *startsymbol* and $P \subseteq (N \cup T) \times (N \cup T)^*$ is a finite set of *productions* such that for each α in $N \cup T$, P contains at least one production (α, x). Productions (α, x) are usually written as $\alpha \rightarrow x$. For words $x = \alpha_1\alpha_2 \cdots \alpha_n$ with $\alpha_i \in (N \cup T)$ and $y = y_1y_2 \cdots y_n$ we write $x \Rightarrow_G y$ (or just $x \Rightarrow y$) if $\alpha_i \rightarrow y_i$ is a production of p for $i = 1, 2,..., n$. We define $\overset{+}{\Rightarrow}$ and $\overset{*}{\Rightarrow}$ as the transitive, transitive and reflexive respectively, closure of $\Rightarrow$ and call the set $L(G) = \{x \mid S \overset{*}{\Rightarrow} x, x \in T^*\}$ the *language generated* by G. With each *derivation* $S = x_0 \Rightarrow x_1 \Rightarrow x_2 \Rightarrow \cdots \Rightarrow x_n = x \in L(G)$ we associate a *derivation tree* t as usual. S is considered to be the label of the *root* of the tree, the symbols of x_{i+1} are label of nodes which are the *sons* or *successors* of the nodes corresponding to symbols in x_i, and the nodes corresponding to x_{i+1} are considered to lie below those corresponding to x_i. The word x is obtained by reading the *leaves* of t, i.e. the *frontier* of t, from left to right. A node corresponding to a symbol in x_i is considered to be of *depth* i, and the maximal depth of any node is called the *height* of the tree. A path in the derivation tree leading from the root of the tree to a leaf is called a *leaf-path* throughout this paper.

Let $G = (N, T, P, S)$ be an EOL system. A production $A \to B$ in P with $A, B \in N$ is called a *nonterminal chain-production.* A production $\alpha \to x$ in P is called *short* if $|x| \leqslant 2$, and is called *propagating* if $x \neq \epsilon$ (the empty word). For a word $x \in (N \cup T)^*$ we denote by alph(x) the set of all symbols occuring in x and extend alph to sets of words M by alph$(M) = \bigcup_{x \in M}$ alph(x) as usual.

A terminal a is called *pseudo*-terminal if $a \notin$ alph$(L(G))$.

Let $G = (N, T, P, S)$ be an EOL system. G is called *propagating* if each production is propagating, *short* if each production is short, *synchronized* if for each $a \in T$, $a \overset{+}{\Rightarrow} x$ implies $x \notin T^*$, and is called *separated* if $P \subseteq N \times T \cup (T \cup N) \times N^+$.

2. Results

We first show that each EOL language can be generated by what we call a chain-free EOL system. We thereby establish the strongest EOL analogue to context-free Chomsky normal form known to date. Based on this result we prove that each EOL language can be generated by a propagating EOL system containing no production $A \to B$ for nonterminals A and B, and containing no production with more than two symbols on the right-hand side. As a byproduct, we then solve a problem on complete EOL forms mentioned in (Maurer, Salomaa and Wood 1977) and give a simple proof for the decidability of emptiness and finiteness for EOL systems.

Definition. An EOL system $G = (N, T, P, S)$ is called *stretching*, if for each nonterminal chain-production $A \to B \in P$ and each $k \geqslant 1$, $A \Rightarrow_G^k B$ holds.

We now establish an auxiliary normal form result which we believe is also of interest in itself.

Lemma 2.1 (Stretching lemma). *Each ϵ-free* EOL *language L can be generated by a stretching* EOL *system.*

Proof. By e.g. (Salomaa, 1973) we may assume $L = L(G)$, where $G = (N, T, P, S)$ is a propagating, synchronized and separated EOL system. For $A \in N$, let m_A be the minimal $k \geqslant 1$ such that $A \Rightarrow_G^k A$ or 1, if such k does not exist. Let m be the least common multiple of the numbers in the set $\{m_A \mid A \in N\} \cup \{|N|\}$.

We now define a new EOL system G_1 as "m-times speed-up" of G as follows:

$$G_1 = (N_1, T, P_1, Z), \qquad \text{where} \quad N_1 = N \cup \{Z\}, \quad Z \notin N,$$

$$P_1 = \{\alpha \to x \mid \alpha \in N \cup T, \alpha \overset{m}{\underset{G}{\Rightarrow}} x\} \cup \{Z \to x \mid S \overset{i}{\underset{G}{\Rightarrow}} x, m \leqslant i \leqslant 2m - 1\}$$

$$\cup \{Z \to x \mid S \overset{i}{\underset{G}{\Rightarrow}} x, x \in T^+, 1 \leqslant i \leqslant m - 1\}.$$

Evidently, $L(G) = L(G_1)$. Now suppose that $A \to B \in P_1$. Then $A \Rightarrow_G^m B$. Thus for some $C \in N$ we have $A \overset{*}{\Rightarrow}_G C \Rightarrow_G^j C \overset{*}{\Rightarrow}_G B$ for some $j \geqslant 1$. Hence also $C \Rightarrow_G^{m_c} C$ and $C \Rightarrow_G^m C$. Thus we have $A \overset{*}{\Rightarrow} C \Rightarrow_G^{j+(k-1)m} C \overset{*}{\Rightarrow}_G B$ for each $k \geqslant 1$, i.e., $A \Rightarrow_{G_1}^k B$ for each $k \geqslant 1$ as desired. ∎

DEFINITION. Let G be a propagating EOL system. A derivation tree in G is called *chain-free* if it contains at least one leaf-path $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}, \alpha_n$ in which each node α_i $(0 \leqslant i \leqslant n-2)$ has at least two successors. The EOL system G is called *chain-free* if each derivation tree in G is chain-free.

Our first major aim is to show that each EOL language can be generated by a chain-free EOL system. We start with an important lemma.

LEMMA 2.2. For each stretiking, synchronized, propagating, separated EOL system $G = (N, T, P, S)$ an equivalent propagating EOL system $G_2 = (N_2, T, P_2, S)$ can be constructed such that for each $x \in L(G_2)$ a chain-free derivation tree in G_2 exists.

Proof. We first define an auxiliary EOL system $G' = (N, N', P', S)$ with $N' = \{A' \mid A \in N\}$ as follows:

$$P' = P \cap N \times N^+ \cup \{A \to B' \mid A \to B \in P, A, B \in N\} \cup \{B' \to B' \mid B \in N\}.$$

Clearly, $L(G')$ is an ϵ-free context-free language. By (Salomaa, 1973) there exists a context-free Chomsky Normal Form Grammar $G_1 = (N_1 \cup \{S\}, N', P_1, S)$ with $L(G_1) = L(G')$.

We may further assume that S occurs only on the left side of productions. Assuming without loss of generality that N, N' and N_1 are pairwise disjoint we define an EOL system $G_2 = (N_2, T, P_2, S)$ with $N_2 = N \cup N' \cup N_1$ as follows:

$$P_2 = P \cup P_1 \cup \{B' \to B' \mid B' \in N'\} \cup \{B' \to w \mid B \to w \in P\}$$
$$\cup \{A \to w \mid A \to B' \in P_1, B \to w \in P\}.$$

We establish that G_2 has the desired properties.

(a) Since $P \subseteq P_2$, evidently $L(G) \subseteq L(G_2)$.

(b) To prove $L(G_2) \subseteq L(G)$, consider an arbitrary $x \in L(G_2)$ and a derivation tree t of x in G_2. This derivation tree either starts with a production of P, and then $x \in L(G)$, or else starts with a production of P_1.

In this case the nodes in each leaf-path of t occur in the following order: first a node labelled S, then some nodes with labels in N_1, then some with labels in N', then some with labels in N, finally one with a label in T.

We construct now a new derivation tree t' of x in G_2 by modifying tree t slightly, increasing its height by one. (This is just to assure that along each leaf-path at least one node with a label in N' occurs).

For each leaf-path, let α be the last node with a label in $N_1 \cup N'$. If the label of α is $B' \in N'$, replace the production $B' \to w$ at node α by $B' \to B' \to w$. If the label of α is $A \in N_1$ then by construction of G_2, $A \to w \in P_2$ with $w \in (T \cup N)^+$ implies $A \to B' \in P_1$, $B' \in N'$, $B \to w \in P$. Hence $B' \to w \in P_2$. Thus we can replace $A \to w$ at node α by $A \to B' \to w$.

In the tree t', let n_i be the last node in the ith leaf path with a label B_i' in N'. Then $y = B_1'B_2' \cdots B_n' \in L(G_1) = L(G')$. If x_i is the frontier of the subtree of t' induced by n_i then clearly $B_i' \overset{*}{\Rightarrow}_{G_2} x_i$, $B_i \overset{*}{\Rightarrow}_G x_i$, $B_i \in N$ and $x = x_1 x_2 \cdots x_i \cdots x_n$.

Let $\hat{t}$ be a derivation tree of $y = B_1'B_2' \cdots B_n'$ in G' and let A_i be the last label in N in the i-th leaf-path of $\hat{t}$. (Thus, only productions of G are used up to this point in tree $\hat{t}$). Since $A_i \to B_i' \in P'$ we have $A_i \to B_i \in P$ and, since P is stretiking, $A_i \Rightarrow_G^k B_i$ for each $k \geqslant 1$. Together with the fact that $B_i \overset{*}{\Rightarrow}_G x_i$ we have $S \overset{*}{\Rightarrow}_G x$ and $x \in L(G)$, as desired.

(c) It remains to show that for every $x \in L(G_2)$ there is a chain-free derivation tree. For arbitrary $x \in L(G_2) = L(G)$ let t be a derivation tree in G. If t is chain-free, nothing has to be shown. Suppose therefore that t is not chain-free. We will modify t to obtain a chain-free tree t' for x in G_2.

We observe that t contains nodes $n_1, n_2, \ldots, n_s$ as follows:

(i) on each leaf-path of t there is exactly one of the nodes $n_1, n_2, \ldots, n_s$;

(ii) the subtrees $t_1, t_2, \ldots, t_s$ of t with roots $n_1, n_2, \ldots, n_s$ are chain-free and the concatenation of their frontiers $x_1, x_2, \ldots, x_s$ from left to right is x;

(iii) there is a subtree t_0 of t with root labeled S and leaves $n_1, n_2, \ldots, n_s$ from left to right.

The situation described is shown in Fig. 2.1.

Let $B_1, B_2, \ldots, B_s$ be the labels of $n_1, n_2, \ldots, n_s$, respectively. Clearly, $B_1', \ldots, B_s' \in L(G')$ and

$$B_k \overset{i_k}{\underset{G}{\Rightarrow}} x_k \qquad \text{for} \quad i_k \geqslant 1 \quad \text{and} \quad k = 1, 2, \ldots, s.$$

Evidently, $S \overset{*}{\Rightarrow}_{G'} A_1 \cdots A_s \Rightarrow_{G'} B_1' \cdots B'_{r-1} B_r' B'_{r+1} \cdots B_s'$ with $A_i \in N_1$, $B_i' \in N'$.

Consider the derivation tree t_{CF} in G' of the sentential form $A_1 \cdots A_s$ and let $m_1, \ldots, m_s$ be its leaves from left to right, with labels $A_1, \ldots, A_s$, respectively. Let j_k be the depth of node m_k for $k = 1, 2, \ldots, s$. Finally, choose r such that $i_r + j_r \geqslant i_k + j_k$, for $k = 1, \ldots, s$.

We are now in a position to exhibit a derivation tree t' in G_2 of x based on t_{CF} and the subtrees $t_1, t_2, \ldots, t_s$ as follows:

(i) for each l with $i_r + j_r = i_l + j_l$ (in particular for $l = r$) attach to A_l the tree t_l from which the root has been removed,

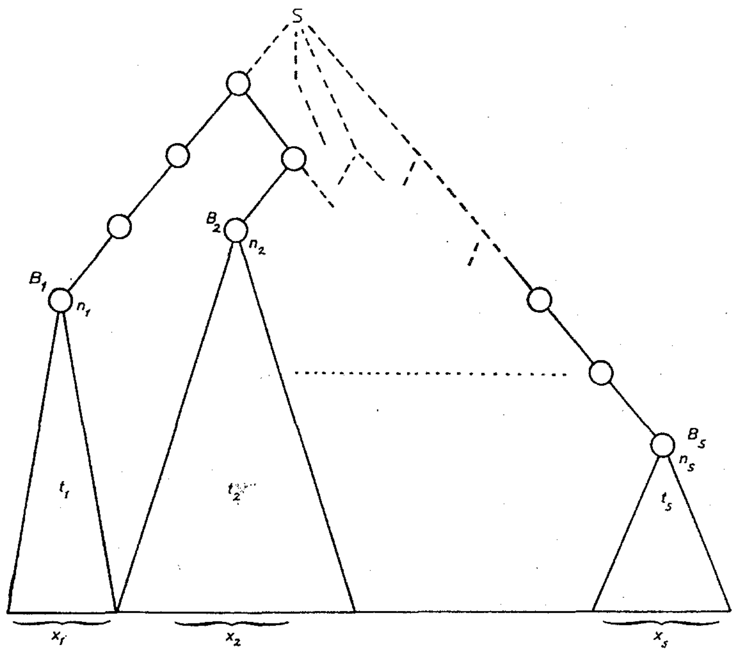

FIGURE 2.1

(ii) for each l with $p = i_r + j_r - i_l - j_l > 0$ attach to A_l a chain of $p\, B_l'$ (i.e. $B_l' \to B_l' \to \cdots \to B_l'$) and attach to the last of these B_l' the tree t_l from which the root has been removed.

The tree obtained is shown in Fig. 2.2. It is clearly a chain-free derivation tree of x in G_2, completing the proof. ∎

We are now ready to prove our first main theorem.

THEOREM 2.3. *Every ϵ-free* EOL *language L can be generated by a propagating, synchronized, separated and chain-free* EOL *system.*

Proof. By Lemma 2.2 we may assume that L is generated by a propagating, synchronized and separated EOL system $G = (N, T, P, S)$ such that for every $x \in L(G)$ a chain-free derivation tree exists. We will construct an EOL system $\bar{G} = (N \cup \bar{N}, T, \bar{P}, \bar{S})$, with $\bar{N} = \{\bar{A} \mid A \in N\}$ an alphabet disjoint from N, which will yield exactly all the chain-free derivation trees of G. To this end define

$$\bar{P} = P \cup \{\bar{A} \to x\bar{B}y \mid A \to xBy \in P,\, A, B \in N,\, xy \neq \epsilon\}$$
$$\cup \{\bar{A} \to a \mid A \to a \in P,\, a \in T\}.$$

Note the critical condition $xy \neq \epsilon$. It insures that the "bar" can only be "inherited" if a node has at least two sons. Since we start with the "barred

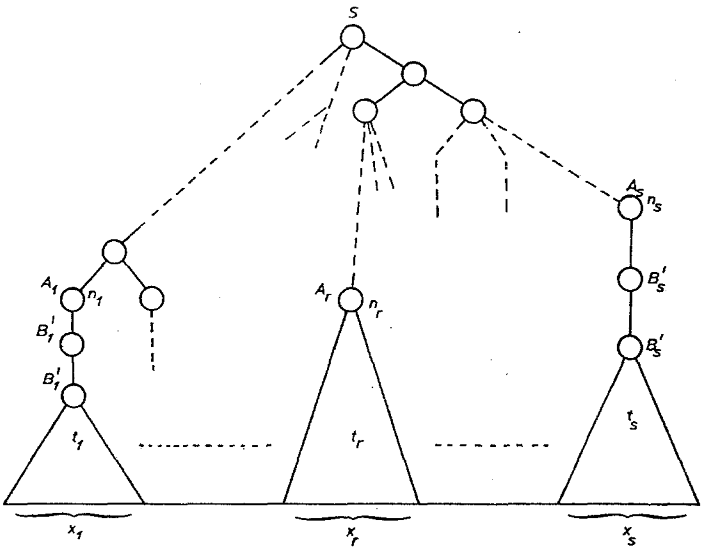

FIGURE 2.2

symbol" $\bar S$, $\bar G$ is synchronized and a bar can only disappear when producing a terminal. Every derivation tree in $\bar G$ is chain-free, and every chain-free derivation tree in G occurs in $\bar G$ with the nodes along one leaf-path being barred. ∎

We next use Theorem 2.3 to establish further normal forms.

THEOREM 2.4. *Every ϵ-free* EOL *language L can be generated by a propagating* EOL *system which contains no nonterminal chain production.*

Proof. Let $\bar G = (N \cup \bar N, T, \bar P, \bar S)$ be the chain-free EOL system generating L according to the proof of Theorem 2.3. Based on $\bar G$ we will construct the desired system $\hat G$ as follows.

Let $\hat G = (N \cup \bar N \cup \{Q\}, T \cup \hat T, \hat P, \bar S)$ with $\hat T = \{\hat A \mid A \in N \cup \bar N\}$ and Q a new nonterminal. We will call the symbols of $\hat T$ "pseudoterminals" since our construction will assure that $L(\hat G) \cap \hat T^* = \varnothing$.

Let $N_u = N \cup \bar N$.

Let μ be the homomorphism on $N_u{}^*$ defined by $\mu(A) = A$ for $A \in N$ and $\mu(\bar A) = \bar A$ for $\hat{\bar A} \in \bar N$.

Define:

$$\begin{aligned}\hat P = {} & \{\bar A \to \mu(w) \mid \bar A \to w \in \bar P \cap \bar N \times N_u{}^+\} \\ & \cup \{\hat A \to w \mid A \to w \in \bar P \cap N \times N_u{}^+\} \\ & \cup \{A \to \hat A \mid A \in N\} \cup \{\hat{\bar A} \to \bar A \mid \bar A \in \bar N\} \\ & \cup (\bar P \cap \bar N \times T) \cup \{a \to Q \mid a \in T\} \cup \{Q \to QQ\} \\ & \cup \{\hat A \to a \mid A \to a \in \bar P \cap N \times T\}.\end{aligned}$$

Note that $\hat{G}$ contains no nonterminal chain production We conclude the proof by establishing $L(\hat{G}) = L(\bar{G}) = L$

A possible derivation tree in $\hat{G}$ is shown in Fig 2.3.

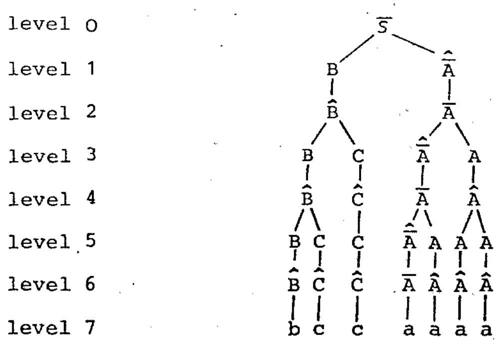

FIGURE 2.3

Clearly, $\hat{G}$ simulates $\bar{G}$ with "half speed". To each level of a tree in $\bar{G}$ correspond two levels of a derivation tree in $\hat{G}$ except for the last one. Terminal words can be obtained only on odd levels and we can easily verify that $L(\hat{G}) \cap \hat{T}^* = \varnothing$, i.e., $\hat{T}$ is really a set of pseudoterminals. This is assured by the fact that on every odd level there can be at most one pseudoterminal (obtained from a barred symbol) which is always produced together with a nonterminal symbol. While there can be many pseudoterminals on even levels there is always one barred nonterminal on such a level. Thus $L(\hat{G}) = L(\bar{G}) = L$ as desired. ▮

Note that the EOL system constructed for the proof of Theorem 2.4 may contain productions with long right-hand sides. While it is possible to eliminate such productions by well-known techniques, those techniques introduce chain productions, thus defeating the aim of the theorem. We will now show, however, that Theorem 2.4 holds even for short EOL systems.

THEOREM 2.5. *Every ϵ-free* EOL *language L can be generated by a propagating and short* EOL *system which contains no nonterminal chain production.*

Proof. Let G be an EOL system generating L and let $\bar{G} = (N \cup \bar{N}, T, \bar{P}, \bar{S})$, $N_u = N \cup \bar{N}$, be the EOL system constructed for the proof of Theorem 2.3 with $L(\bar{G}) = L$. By modifying the idea of the proof of Theorem 2.4 we will construct a propagating and short EOL system G_1 without chain productions which generates L.

In tuitively, each derivation step in $\bar{G}$ will be simulated by several steps in G_1 (rather than by two steps as in $\hat{G}$).

Let p be the maximum length of the right side of productions of $\bar{G}$, i.e. $p = \max\{|w| \mid A \to w \in \bar{P}\}$. Let q be the smallest even number such that

$2^q \geqslant p$. Clearly, for each production $p \in \bar{P}$ with $p = A \to w$, $w \in N_u^+$, we can construct a set P_p of short productions such that $A \Rightarrow_{P_p}^q u$ iff $u = w$, and such that in the intermediate steps new nonterminals, say N_p, are used. Furthermore, if $\bar{A} \Rightarrow_{P_p} u_1 \Rightarrow_{P_p} u_2 \Rightarrow_{P_p} \cdots \Rightarrow_{P_p} u_q = u$, i.e. $|u| \geqslant 2$, then each u_i contains exactly one barred symbol and in the first step a production of the form $\bar{A} \to \bar{B}C$ or $\bar{A} \to B\bar{C}$ is used, $\bar{A} \in \bar{N}$, $\bar{B}, C$ or $B, \bar{C}$ in $\bigcup_{p \in \bar{P}_N} N_p$, where $\bar{P}_N = \bar{P} \cap N_u \times N_u^+$.

Assume that $N_{p_1} \cap N_{p_2} = \varnothing$ for $p_1 \neq p_2$ and let Q be a new nonterminal. Let $G_2 = (N_2, T, P_2, \bar{S})$ be an auxiliary EOL system as follows:

$$N_2 = \{Q\} \cup N_u \cup \bigcup_{p \in \bar{P}_N} N_p, \qquad P_2 = (P_2 - \bar{P}_N) \cup \bigcup_{p \in \bar{P}_N} P_p.$$

It is easy to verify that $L(G_2) = L(\bar{G})$.

We are now ready to define the desired EOL system $G_1 = (N_2, T \cup \hat{T}, P_1, \bar{S})$, where $\hat{T} = \{\hat{A} \mid A \in N_2\}$ is a set of pseudoterminals. Let N_1 be the subset of N_2 of the nonbarred symbols and let $\bar{N}_1 = N_2 - N_1$ be the set of barred symbols. Let μ be the homomorphism on N_2^* defined by $\mu(A) = A$ for $A \in N_1$ and $\mu(\bar{A}) = \hat{\bar{A}}$ for each $\bar{A} \in \bar{N}_1$.

Let

$$\begin{aligned} P_1 = {} & \{\bar{A} \to \mu(w) \mid \bar{A} \to w \in P_2 \cap \bar{N}_1 \times N_2^+\} \\ & \cup \{\hat{A} \to w \mid A \to w \in P_2 \cap N_1 \times N_2^+\} \\ & \cup \{A \to \hat{A} \mid A \in N_1\} \cup \{\hat{\bar{A}} \to \bar{A} \mid \bar{A} \in \bar{N}\} \\ & \cup (\bar{P} \cap \bar{N} \times T) \cup \{a \to QQ \mid a \in T\} \\ & \cup \{Q \to \hat{Q}Q, \hat{Q} \to Q\} \\ & \cup \{\hat{A} \to a \mid A \to a \in \bar{P} \cap N \times T\}. \end{aligned}$$

Note that G_1 is short and contains no nonterminal chain production. We verify that $L(G_1) = L(G_2) = L(\bar{G}) = L$ using a number of observations.

Note that terminals in T can only be produced from nonterminals in N_u, not from symbols in $\bigcup_{p \in \bar{P}_N} N_p$. The proof that $L(G_1) = L(G_2)$ is analogous to that of $L(\hat{G}) = (\bar{G})$, the most important difference being that a terminal word can be produced only in $1 + kq$, $k \geqslant 0$, steps in a G_1 derivation.

Indeed on all "even levels" of any derivation tree in G_1 there is a barred nonterminal, and on the "odd levels" different from levels $1 + kq$, $k = 0, 1, \ldots$, there is at least one nonbarred nonterminal, since none of the new symbols in $\bigcup_{p \in \bar{P}_N} N_p$ produces a terminal symbol. There is clearly one path from the root to a leaf along which all symbols are barred. Finally, as an alternative to a terminal symbol on levels $1 + kq$ $k = 0, 1, \ldots$ either a nonbarred nonterminal or a terminal together with a barred nonterminal is produced. Thus no terminal word containing a symbol of $\hat{T}$ is produced and the proof is complete. ∎

We will now consider two applications of our results.

Let $G = (N, T, P, S)$ be an EOL system. We first note that $L(G) \neq \varnothing$ iff

$S \Rightarrow^i x$, $x \in T^*$ for some $i \leqslant 2^{|N+T|}$. For suppose $S \Rightarrow^j x$, $x \in T^*$ for some $j > 2^{|N+T|}$. Then $S \Rightarrow^{i_1} x_1 \Rightarrow^{i_2} x_2 \Rightarrow^{i_3} x$ with $\text{alph}(x_1) = \text{alph}(x_2)$ and $i_2 > 0$ Thus $S \Rightarrow^{i_1} x_1 \Rightarrow^{i_3} y$, i.e. $S \Rightarrow^{j'} y$ for $y \in T^*$ and $j' < j$. (Note that above proof also holds for ETOL systems).

Modifying above idea and using the chain-free normal form result of Theorem 2.3 we will show that the finiteness problem for EOL system is decidable.

While it is well-known that both emptiness and finiteness is decidable for EOL systems, these facts are usually proved based on decidability results for indexed-languages and the fact that every EOL language is an indexed-language (Culik II, 1974), (Rosenberg, Salomaa, 1976). Thus our simple direct proofs may be of some interest.

COROLLARY 1. *There exists an algorithm which decides for any* EOL *system* $G = (N, T, P, S)$ *whether* $L(G)$ *is finite.*

Proof. Let $\bar{G}$ be the chain-free EOL system generating $L(G) - \{\epsilon\}$ as constructed in the proof of Theorem 2.1 and let $m = 2^{|N \cup \bar{N}|}$.

We maintain that $L(G)$ is infinite iff $\bar{S} \Rightarrow^i x$ for some $x \in T^*$ and $m < i \leqslant 2m$.

Part 1. Suppose $\bar{S} \Rightarrow^i x$ for some $x \in T^*$ and $m < i \leqslant 2m$. Then $\bar{S} \Rightarrow^{i_1} x_1 \Rightarrow^{i_2} x_2 \Rightarrow^{i_3} x$ for some $i_1 + i_2 + i_3 = i$, $i_2 > 0$, $\text{alph}(x_1) = \text{alph}(x_2)$ and $x_1 = u_1\bar{A}v_1$, $x_2 = u_2 y\bar{A}zv_2$ with $\bar{A} \in \bar{N}$, $u_1 \Rightarrow^{i_2} u_2$, $v_1 \Rightarrow^{i_2} v_2$, $\bar{A} \Rightarrow^{i_2} y\bar{A}z$, $|yz| > 0$. Thus $\bar{S} \Rightarrow^{i_1} x_1 \Rightarrow^{ki_2} w_k \Rightarrow^{i_3} \bar{w}_k \in T^+$ for $k \geqslant 1$, with $\bar{w}_1 = x$ and $|\bar{w}_1| < |\bar{w}_2| < \cdots$. Thus $L(\bar{G}) = L(G) - \{\epsilon\}$ is infinite.

Part 2. Suppose $L(\bar{G}) = L(G) - \{\epsilon\}$ is infinite and for no i with $m < i \leqslant 2m$ and $x \in T^*$ does $\bar{S} \Rightarrow^i x$ hold. We will derive a contradiction.

Since $L(\bar{G})$ is infinite, $\bar{S} \Rightarrow^j y_0$ holds for some $j > 2m$ and $y_0 \in T^*$. We will construct a sequence $y_0, y_1, \ldots$ with $y_i \in L(\bar{G})$ and $|y_0| > |y_1| > |y_2| > \cdots$ and $|y_i| - |y_{i+1}| \leqslant m$, contradicting our assumption. Suppose we have $\bar{S} \Rightarrow^k y_i \in T^*$ for some $k > 2m$. Then $\bar{S} \Rightarrow^{i_1} x_1 \Rightarrow^{i_2} x_2 \Rightarrow^{i_3} y_i$ for some i_1, i_2, i_3 with $0 < |i_2| \leqslant m$, $i_1 + i_2 + i_3 = k$. Thus $\bar{S} \Rightarrow^{i_1} x_1 \Rightarrow^{i_3} y_{i+1} \in T^*$ for some y_{i+1}. Clearly, $|y_i| - |y_{i+1}| = i_2 \leqslant m$ and $|y_i| > |y_{i+1}|$ as desired. ∎

We conclude the paper with an application of Theorem 2.5 to EOL forms.

From the productions used in P_1 it is clear that there are propagating complete EOL forms with a single nonterminal S and a single terminal a with no production $S \to S$, answering a question raised in (Maurer, Salomaa and Wood, 1977). More specifically we have:

COROLLARY 2. *The* EOL *form* $F = (\{S\}, \{a\}, P, S)$ *with productions* $P = \{S \to a, S \to aS, S \to Sa, a \to a, a \to S, a \to SS\}$ *is complete.*

RECEIVED: March 10, 1977; REVISED: April 21, 1977

References

MAURER, H. A., SALOMAA, A., AND WOOD, D. (1977), EOL forms, *Acta Informatica* 8, 75–96.

SALOMAA, A. (1973), "Formal Languages," Academic Press, New York.

CULIK II, K. (1974, On some families of languages related to developmental systems, *Intern. J. Computer Math.* **4**, Section A, 31–42.

ROZENBERG, G., AND SALOMAA, A., (1976), The mathematical theory of *L* systems, *in* "Advances in Information Systems Science," Vol. 6, Plenum Press, New York, pp. 161–206.

HERRMANN, G. T., AND ROZENBERG, G. (1975), "Developmental Systems and Languages, North Holland, Amsterdam.