INFORMATION AND CONTROL 17, 359–394 (1970)

Simple Matrix Languages

OSCAR H. IBARRA

Department of Computer, Information, and Control Sciences, University of Minnesota, Minneapolis, Minnesota 55455

Simple matrix languages and right-linear simple matrix languages are defined as subfamilies of matrix languages by putting restrictions on the form and length (degree) of the rewriting rules associated with matrix grammars. For each $n \geqslant 1$, let $\mathscr{S}(n)[\mathscr{R}(n)]$ be the class of simple matrix languages [right-linear simple matrix languages] of degree n, and let

$$\mathscr{S} = \bigcup_{n\geqslant 1} \mathscr{S}(n) \left[\mathscr{R} = \bigcup_{n\geqslant 1} \mathscr{R}(n)\right].$$

It is shown that $\mathscr{S}(1)[\mathscr{R}(1)]$ coincides with the class of context-free languages [regular sets] and that $\mathscr{S}$ is a proper subset of the family of languages accepted by deterministic linear bounded automata. It is proved that $\mathscr{S}(n)[\mathscr{R}(n)]$ forms a hierarchy of classes of languages in $\mathscr{S}[\mathscr{R}]$. The closure properties and decision problems associated with $\mathscr{S}(n)$, $\mathscr{S}$, $\mathscr{R}(n)$, and $\mathscr{R}$ are thoroughly investigated.

Let $\mathscr{S}_B[\mathscr{R}_B]$ be the bounded languages in $\mathscr{S}[\mathscr{R}]$. It is shown that $\mathscr{S}_B = \mathscr{R}_B$ and that most of the positive closure and decision results which are true for bounded context-free languages are carried over in $\mathscr{S}_B$. A characterization of $\mathscr{S}_B$ as the smallest family of languages which contains the bounded context-free languages and which is closed under the operations of union and intersection is proved.

INTRODUCTION

In recent years, a number of new families of languages which are richer than context-free languages has been introduced and studied, such as context-sensitive languages [11] and indexed languages [2]. In [1], Abraham introduced the concept of a matrix language in his study of some questions related to language theory. The purpose of the present investigation is to study a restricted family of matrix languages called simple matrix languages. This family properly contains the context-free languages and is richer than the family of equal matrix languages reported in [15].

The paper is divided into six sections. Section 1 contains basic notations and definitions and introduces simple matrix languages (right-linear simple matrix languages) and relates them with context-free languages (regular sets) over the direct product of free monoids and with the family of deterministic linear bounded automata.

In Section 2, a hierarchy of classes in the family of simple matrix languages (right-linear simple matrix languages) is established. In Section 3, the positive closure properties of simple matrix languages (right-linear simple matrix languages) are given. In particular, it is shown that simple matrix languages (right-linear simple matrix languages) are closed under the operations of substitution by context-free languages (regular sets), union, transposition, intersection with regular sets, sequential transducer mapping, and mapping and pseudoinverse mapping by a nondeterministic generalized sequential machine.

In Section 4, negative results are presented. It is proved that simple matrix languages and right-linear simple matrix languages are not closed under the operations of intersection, complementation, concatenation, and closure. The existence of a context-free language which is not a right-linear simple matrix language is also proved.

In Section 5, a result which connects simple matrix languages with semi-linear sets is derived. Bounded simple matrix languages are considered and are shown to be the smallest family of languages containing the bounded context-free languages which is closed with respect to finite union and intersection. In section 6, decision questions associated with the languages under study are briefly investigated.

1. Definitions and Basic Results

An alphabet is a finite nonempty set of symbols. Let Σ be an alphabet. A word over Σ is a finite (possibly empty) sequence of symbols in Σ. The *empty word* will be denoted by Λ. The set of all words including the empty word will be denoted by Σ^*. An n-tuple of words over Σ is an n-tuple $(x_1, ..., x_n)$ where each x_i is a word in Σ. We make no distinction between a 1-tuple of word and a word. The set of all n-tuples of words over Σ is $\Sigma^* \times \cdots \times \Sigma^*$ (n times) and is written $[\Sigma^*]^n$.

Let $x = a_1 a_2 \cdots a_m$ and $y = b_1 b_2 \cdots b_n$ be words over Σ, each a_i in Σ, b_j in Σ, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$.

(1) The *concatenation* of x and y is the word $xy = a_1 \cdots a_m b_1 \cdots b_n$. We note that $x\Lambda = \Lambda x = x$ for every x in Σ^*.

(2) x^l is defined inductively as follows: $x^0 = x$ and $x^{l+1} = x^l x$.

(3) The *transpose* of x is the word $x^T = a_m \cdots a_1$. We have $\Lambda^T = \Lambda$ and $a^T = a$ for every a in Σ.

(4) The *length* of x, denoted by $lg(x)$, is the number of occurrences of symbols of Σ in x. Thus, $lg(a_1 \cdots a_m) = m$ and $lg(\Lambda) = 0$.

Let X and Y be sets of words.

(5) The *concatenation* of X and Y is the set $XY = \{xy \mid x \text{ in } X, y \text{ in } Y\}$.

(6) The *transpose* of X is the set $X^T = \{x^T \mid x \text{ in } X\}$.

(7) Let $X^0 = \{\Lambda\}$. For $i \geqslant 0$, let $X^{i+1} = X^i X$.

The *closure* of X is the set $X^* = \bigcup_{i \geqslant 0} X^i$ ($*$ is sometimes called the *star operator*).

Concatenation, transposition, and closure for n-tuples of words and sets of n-tuples of words are defined in the same manner except that the operations are done componentwise.

We now define the notion of a matrix grammar as developed in [1].

DEFINITION. Let $n \geqslant 1$. A *matrix grammar of degree n* (*n – MG*, for short) is a 4-tuple $G_n = \langle V, P, S, \Sigma \rangle$, where

(1) V is a finite nonempty set of *nonterminal symbols*;

(2) Σ is a finite nonempty set of *terminal symbols*, $V \cap \Sigma = \varnothing$[1];

(3) S in V is the *start symbol*;

(4) P is a finite set of *matrix rewriting rules* of the form $[A_1 \to w_1, \ldots, A_k \to w_k]$, where $1 \leqslant k \leqslant n$ and for $1 \leqslant i \leqslant k$, A_i is in V and w_i is in $(V \cup \Sigma)^*$.

We describe how G_n generates words over Σ^*.

DEFINITION. Let $G_n = \langle V, P, S, \Sigma \rangle$ be an n-MG. For α, β in $(V \cup \Sigma)^*$, let $\alpha \Rightarrow \beta$ if there exist $k \geqslant 1$, $\alpha_1, \ldots, \alpha_{k+1}$, $x_1, \ldots, x_k$, $w_1, \ldots, w_k$, $y_1, \ldots, y_k$ in $(V \cup \Sigma)^*$, $A_1, \ldots, A_k$ in V such that (1) $\alpha_1 = \alpha$, $\alpha_{k+1} = \beta$; (2) $\alpha_i = x_i A_i y_i$, $\alpha_{i+1} = x_i w_i y_i$ for $1 \leqslant i \leqslant k$; (3) $[A_1 \to w_1, \ldots, A_k \to w_k]$ is in P. Let $\alpha \overset{*}{\Rightarrow} \beta$ if there exist $r > 0$, $\alpha_0, \alpha_1, \ldots, \alpha_r$ such that $\alpha_0 = \alpha$, $\alpha_r = \beta$, and $\alpha_i \Rightarrow \alpha_{i+1}$ for $0 \leqslant i < r$. The language generated by G_n is $L(G_n) = \{x \text{ in } \Sigma^* \mid S \overset{*}{\Rightarrow} x\}$ and is called the *matrix language of degree n* (*or n-ML*) generated by G_n. $L \subseteq \Sigma^*$ is an n-ML if and only if there exists an n-MG G_n such that $L = L(G_n)$.

[1] $\varnothing$ denotes the empty set.

EXAMPLE. Let $G_4 = \langle\{A_1, A_2, A_3, A_4, A_5\}, P, A_1, \{0, 1\}\rangle$, where $P = \{[A_1 \to A_2A_3A_4A_5], [A_2 \to 0A_2, A_3 \to A_30, A_4 \to A_40, A_5 \to 0A_5], [A_2 \to 1A_2, A_3 \to A_31, A_4 \to A_41, A_5 \to 1A_5], [A_2 \to \Lambda, A_3 \to \Lambda, A_4 \to \Lambda, A_5 \to \Lambda]\}$. Then G_4 is a 4-MG, and $L(G_4) = \{xx^Tx^Tx \mid x \text{ in } \{0, 1\}^*\}$.

Remark. A *context-free grammar* or *CFG* is a matrix grammar of degree 1, i.e., all the matrix rewriting rules have length 1. The language associated with a CFG is called a *context-free language* or *CFL*. In the case of a CFG, we need not carry the brackets around each rule. The language $L(G_4)$ in the above example is not a CFL, therefore, matrix grammars have more generating power than context-free grammars.

In this paper, we shall be concerned with a subfamily of matrix grammars. These are grammars which have certain restrictions on the lengths of the matrix rewriting rules and the mode of derivations of words.

DEFINITION. Let $n \geqslant 1$. A *simple matrix grammar of degree n* (or *n-SMG*) is an $(n + 3)$-tuple $G_n = \langle V_1, \ldots, V_n, P, S, \Sigma\rangle$, where

(1) $V_1, \ldots, V_n$ are finite nonempty pairwise disjoint sets of nonterminal symbols;

(2) Σ is a finite nonempty set of terminal symbols, $\Sigma \cap V_j = \varnothing$ for $1 \leqslant j \leqslant n$;

(3) S is not in $V_1 \cup \cdots \cup V_n \cup \Sigma$ and is called the *start symbol*;

(4) P is a finite set of matrix rewriting rules of the form:

(a) $[S \to w]$, where w is in Σ^*.

(b) $[S \to x_{11}A_{11}x_{12}A_{12} \cdots x_{1k}A_{1k} \cdots x_{n1}A_{n1} \cdots x_{nk}A_{nk}y]$, where $k \geqslant 1$, y is in Σ^*, and for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant k$, A_{ij} is in V_i and x_{ij} is in Σ^*.

(c) $[A_1 \to w_1, \ldots, A_n \to w_n]$, where for $1 \leqslant i \leqslant n$, A_i is in V_i and w_i is in Σ^*.

(d) $[A_1 \to x_{11}A_{11}x_{12}A_{12} \cdots x_{1k}A_{1k}y_1, \ldots, A_n \to x_{n1}A_{n1} \cdots x_{nk}A_{nk}y_n]$, where $k \geqslant 1$ and for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant k$, y_i, x_{ij} are in Σ^* and A_i, A_{ij} are in V_i.

DEFINITION. Let $G_n = \langle V_1, \ldots, V_n, P, S, \Sigma\rangle$ be an n-SMG. For α, β in $(V_1 \cup \cdots \cup V_n \cup \Sigma \cup \{S\})^*$, let $\alpha \Rightarrow \beta$ if either (1) or (2) holds:

(1) $\alpha = S$ and $[S \to \beta]$ is in P.

(2) There exist $y_1, \ldots, y_n$ in Σ^*, $w_1, \ldots, w_n$, $z_1, \ldots, z_n$ each w_i, z_i in $(V_i \cup \Sigma)^*$, $A_1, \ldots, A_n$ each A_i in V_i such that $\alpha = y_1A_1z_1 \cdots y_nA_nz_n$,

$\beta = y_1w_1z_1 \cdots y_nw_nz_n$, and $[A_1 \to w_1, ..., A_n \to w_n]$ is in P. Let $\alpha \overset{*}{\Rightarrow} \beta$ if there exist $r > 0$, $\alpha_0, ..., \alpha_r$ such that $\alpha_0 = \alpha$, $\alpha_r = \beta$, and $\alpha_i \Rightarrow \alpha_{i+1}$ for $0 \leqslant i < r$. The language generated by G_n is $L(G_n) = \{x \text{ in } \Sigma^* \mid S \overset{*}{\Rightarrow} x\}$ and is called the *simple matrix language of degree n* (or *n-SML*) generated by G_n. $L \subseteq \Sigma^*$ is an n-SML if and only if there exists an n-SMG G_n such that $L = L(G_n)$.

An important subfamily of simple matrix grammars which we shall also consider is the following.

DEFINITION. Let $n \geqslant 1$. A *right-linear simple matrix grammar of degree n* (abbreviated, *n-RLSMG*) is an $(n+3)$-tuple $G_n = \langle V_1, ..., V_n, P, S, \Sigma\rangle$, where $V_1, ..., V_n$, Σ, S have the same significance as is an n-SMG and P is a finite set of matrix rewriting rules of the form:

(1) $[S \to w]$, where w is in Σ^*.

(2) $[S \to x_{11}A_{11}x_{12}A_{12} \cdots x_{1k}A_{1k} \cdots x_{n1}A_{n1} \cdots x_{nk}A_{nk}y]$, where $k \geqslant 1$, y is in Σ^*, and for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant k$, A_{ij} is in V_i and x_{ij} is in Σ^*.

(3) $[A_1 \to w_1, ..., A_n \to w_n]$, where for $1 \leqslant i \leqslant n$, w_i is in Σ^* and A_i is in V_i.

(4) $[A_1 \to x_1B_1, ..., A_n \to x_nB_n]$, where for $1 \leqslant i \leqslant n$, x_i is in Σ^* and A_i, B_i are in V_i.

The language generated by an n-RLSMG will be called a *right-linear simple matrix language of degree n* (or *n-RLSML*).

Remark. Every n-SML is an n-ML, and every CFG can be reduced to a 1-SMG. Thus, the class of 1-SML's is just the class of CFL's. Also, the class of 1-RLSML's is precisely the class of *right-linear context-free languages* (*RLCFL*'s) which is just the class of *regular sets*.[2] RLCFL's (= regular sets) are those languages generated by CFG's with right linear rules. Recently, a subfamily of simple matrix languages called equal matrix languages appeared in the literature [15]. This subfamily is essentially the family of right linear simple matrix languages. We shall show in Section 4 that there are CFL's which are not right linear simple matrix languages and therefore not equal

[2] A *finite automaton* is a 5-tuple $A = \langle K, \Sigma, M, q_0, F\rangle$, where K and Σ are finite nonempty sets of *states* and *inputs*, respectively, q_0 in K is the *start state*, $F \subseteq K$ is the set of *final states*, and M is a function from $K \times \Sigma$ into K. M is extended into a function from $K \times \Sigma^*$ into K as follows: $M(q, \Lambda) = \Lambda$, $M(q, xa) = M(M(q, x), a)$ for all x in Σ^*, a in Σ, and q in K. $L \subseteq \Sigma^*$ is a *regular set* if and only if there exists a finite automaton $A = \langle K, \Sigma, M, q_0, F\rangle$ such that $L = T(A) = \{x \text{ in } \Sigma^* \mid M(q_0, x) \text{ in } F\}$. Regular sets are discussed in [8, 13].

matrix languages. This result was not shown in [15]. The language in the previous example is a 4-SML generated by $G_4 = \langle\{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}, P, A_1, \Sigma\rangle$ with P defined in the example. Thus, again, simple matrix grammars have more generating power than context-free grammars.

We have defined a simple matrix grammar and thus a simple matrix language by putting essentially two types of restrictions on the definition of a matrix grammar. One type of restriction is on the form of the rewriting rules and the other type of restriction is on the way the words of the language are generated. In an n-SMG $G_n = \langle V_1, ..., V_n, P, S, \Sigma\rangle$, each rule in P of the form $[A_1 \to \omega_1, ..., A_n \to \omega_n]$ satisfies the conditions that (1) for each $1 \leqslant i \leqslant n$, A_i and the nonterminals in ω_i are in V_i, and (2) for $1 \leqslant i, j \leqslant n$, the number of nonterminals in ω_i is equal to the number of nonterminals in ω_j. Derivation of words in G_n is defined by the relation $\Rightarrow$. (3) Intuitively, $\alpha \Rightarrow \beta$ if β can be obtained from α using a rule which replaces the leftmost nonterminal in each of the n "disjoint subwords" of α (see the formal definition of $\Rightarrow$).

Restriction (2) is essential in the sense that there is a 2-MG satisfying only (1) and (3) and which generates a nonsimple matrix language. Consider $G_2 = \langle\{A_1, A_2\}, \{B_1, B_2\}, P, S, \{a, b, c\}\rangle$, where $P = \{[S \to A_1A_2B_1B_2], [S \to \Lambda], [A_1 \to aA_1b, B_1 \to B_1B_1], [A_1 \to ab, B_1 \to B_1], [A_2 \to cA_2, B_1 \to \Lambda], [A_2 \to A_1A_2, B_2 \to B_1B_2], [A_2 \to \Lambda, B_2 \to \Lambda]\}$. Clearly, $L(G_2) = \{a^nb^nc^n \mid n \geqslant 1\}^*$ which is not a simple matrix language as we shall see in Section 4 (Theorem 4.6.).

Now suppose in G_n, we impose only restrictions (1) and (2) and that the application of a rule in a derivation of β from α need not be a replacement of the leftmost nonterminal in each of the n disjoint subwords of α. With these conditions, we shall see that a nonsimple matrix language can be generated. Let $G_2 = \langle\{A_1, A_2, A_3\}, \{B_1, B_2, B_3, B_4, B_5\}, P, S, \{a, b, c\}\rangle$, where $P = \{[S \to \Lambda], [S \to A_1A_2A_3B_1B_4B_5], [A_1 \to aA_1, B_1 \to B_2], [A_2 \to bA_2, B_2 \to B_3], [A_3 \to cA_3, B_3 \to B_1], [A_1 \to a, B_1 \to \Lambda], [A_2 \to b, B_4 \to \Lambda], [A_3 \to c, B_5 \to \Lambda], [A_3 \to cA_1A_2A_3, B_5 \to B_1B_4B_4]\}$. Then $L(G_2) = \{a^nb^nc^n \mid n \geqslant 1\}^*$ which is again not a simple matrix language. Thus, restriction (3) is also essential in the definition of a simple matrix language.

It can be shown that restrictions (2) and (3) can be removed without altering the family of right-linear simple matrix languages.

In [9, 10], the notions of an n–context-free language and an n–right-linear context-free language were introduced in the study of multitape pushdown automata and multitape finite automata. We shall show that n–context-free languages (n–right-linear context-free languages) are related in a natural way to n-SML's (n-RLSML's).

Notation. Let Σ be an alphabet and n be a positive integer. For each a in Σ and $1 \leqslant i \leqslant n$, let $[a, i]$ be an abstract symbol. Let Σ_n be the set of all such abstract symbols. The mapping τ_n from $\Sigma_n{}^*$ into $[\Sigma^*]^n$ is defined as follows:

(1) $\tau_n(\Lambda) = (\Lambda,..., \Lambda)$ (n-occurences of Λ),

(2) For each $[a, i]$ in Σ_n, let $\tau_n([a, i]) = (\Lambda,..., a,..., \Lambda)$ (with $n - 1$ occurrences of Λ and a occurring in the i-th coordinate),

(3) For each $\alpha_1,..., \alpha_m$ in Σ_n ($m \geqslant 1$) let $\tau_n(\alpha_1 \cdots \alpha_m) = \tau_n(\alpha_1) \cdots \tau_n(\alpha_m)$ (Thus, $\tau_n(\alpha_1 \cdots \alpha_m) = (x_1,..., x_n)$ for some $(x_1,..., x_n)$ in $[\Sigma^*]^n$).

EXAMPLE. If $\Sigma = \{a, b\}$ and $n = 2$, then $\Sigma_2 = \{[a, 1], [a, 2], [b, 1], [b, 2]\}$ and $\tau_n([a, 1][b, 2][b, 1]) = (a, \Lambda)(\Lambda, b)(b, \Lambda) = (ab, b)$, $\tau_n^{-1}((ab, b)) = \{[a, 1][b, 2][b, 1], [a, 1][b, 1][b, 2], [b, 2][a, 1][b, 1]\}$.

DEFINITION. A subset $L \subseteq [\Sigma^*]^n$ is called an *n–context-free language* or *n-CFL* (*n–right-linear context-free language* or *n-RLCFL*) if and only if there exists a CFL (RLCFL, or alternatively, a regular set) $L' \subseteq \Sigma_n{}^*$ such that $L = \tau_n(L')$.

Remark. In [14], n-RLCFL's were called n-regular sets and were studied in connection with multitape finite automata.

We now relate n-CFL's (n-RLCFL's) with n-SML's (n-RLSML's).

LEMMA 1.1. *If* $L \subseteq [\Sigma^*]^n$ *is an n-CFL, then the set* $L' = \{x_1 \cdots x_n \mid (x_1,..., x_n)$ *in* $L\}$ *is an n-SML. Furthermore, an n-SMG* G_n *generating* L' *can be effectively constructed.*

Proof. Let $G = \langle V, P, S, \Sigma_n\rangle$ be a CFG such that $\tau_n(L(G)) = L$. Without loss of generality, we may assume that the rules in P are of the form $A \to x$ or $A \to BC$, where A, B, C are in V and x is in $\Sigma_n{}^*$ [5]. For each A in V, let $(A, 1),..., (A, n)$ be abstract symbols. Let S' be a new symbol. Define an n-SMG $G_n = \langle V_1,..., V_n, P', S', \Sigma\rangle$, where $V_i = \{(A, i) \mid A \text{ in } V\}$ for $1 \leqslant i \leqslant n$, and P' is defined as follows:

(1) $[S' \to (S, 1)(S, 2) \cdots (S, n)]$ is in P'.

(2) For all A, B, C in V, if $A \to BC$ is in P, then let $[(A, 1) \to (B, 1)(C, 1),..., (A, n) \to (B, n)(C, n)]$ be in P'.

(3) For each A in V and x in $\Sigma_n{}^*$ if $A \to x$ is in P, then let $[(A, 1) \to p_1(\tau_n(x)),..., (A, n) \to p_n(\tau_n(x))]$ be in P', where for $1 \leqslant i \leqslant n$, p_i is a mapping from $[\Sigma^*]^n$ into Σ^* defined by: $p_i(x_1,..., x_n) = x_i$ for each $(x_1,..., x_n)$ in $[\Sigma^*]^n$. Clearly, $L(G_n) = \{x_1 \cdots x_n \mid (x_1,..., x_n) \text{ in } \tau_n(L(G)) = L\} = L'$.

LEMMA 1.2. *If $L \subseteq \Sigma^*$ is an n-SML, then there exists an n-CFL $L' \subseteq [\Sigma^*]^n$ such that $L = \{x_1 \cdots x_n \mid (x_1,..., x_n)$ in $L'\}$. Moreover, a CFG $G = \langle V, P', S, \Sigma_n\rangle$ can be effectively constructed such that $L' = \tau_n(L(G))$.*

Proof. Let $G_n = \langle V_1,..., V_n, P, S, \Sigma\rangle$ be an n-SMG such that $L = L(G_n)$. Define a CFG $G = \langle V, P', S, \Sigma_n\rangle$, where $V = V_1 \times \cdots \times V_n \cup (S\}$ and P' is defined as follows:

(1) For each w in Σ^*, if $[S \to w]$ is in P, then let $S \to \alpha$ be in P' for each α in $\tau_n^{-1}((\Lambda,..., \Lambda, w))$.

(2) For $k \geqslant 1$, y in Σ^*, x_{ij} in Σ^*, A_{ij} in V_i $(1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant k)$, if $[S \to x_{11}A_{11}x_{12}A_{12} \cdots x_{1k}A_{1k} \cdots x_{n1}A_{n1} \cdots x_{nk}A_{nk}y]$ is in P, then let $S \to \alpha_1[A_{11}, A_{21},..., A_{n1}]\, \alpha_2[A_{12}, A_{22},..., A_{n2}] \cdots \alpha_k[A_{1k}, A_{2k},..., A_{nk}]\, \alpha_{k+1}$ be in P' for each α_j in $\tau_n^{-1}((x_{1j},..., x_{nj}))$ and α_{k+1} in $\tau_n^{-1}((\Lambda,..., \Lambda, y))$.

(3) For each A_i in V_i and w_i in Σ^* $(1 \leqslant i \leqslant n)$, if $[A_1 \to w_1,..., A_n \to w_n]$ is in P, then let $[A_1,..., A_n] \to \alpha$ be in P' for each α in $\tau_n^{-1}((w_1,..., w_n))$.

(4) For $k \geqslant 1$, y_i in Σ^*, x_{ij} in Σ^*, A_{ij} in V_i $(1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant k)$, if $[A_1 \to x_{11}A_{11}x_{12}A_{12} \cdots x_{1k}A_{1k}y_1,..., A_n \to x_{n1}A_{n1}x_{n2}A_{n2} \cdots x_{nk}A_{nk}y_n]$ is in P, then let $[A_1,..., A_n] \to \alpha_1[A_{11}, A_{21},..., A_{n1}]\, \alpha_2[A_{12}, A_{22},..., A_{n2}] \cdots \alpha_k[A_{1k}, A_{2k},..., A_{nk}]\alpha_{k+1}$ be in P' for each α_j in $\tau_n^{-1}((x_{1j}\ ..., x_{nj}))$ and α_{k+1} in $\tau_n^{-1}((y_1,..., y_n))$. It is easily verified that $L = L(G_n) = \{x_1 \cdots x_n \mid (x_1,..., x_n)$ in $\tau_n(L(G)) = L'\}$.

Combining Lemmas 1.1 and 1.2, we have:

THEOREM 1.1. *$L \subseteq \Sigma^*$ is an n-SML if and only if there exists an n-CFL $L' \subseteq [\Sigma^*]^n$ such that $L = \{x_1 \cdots x_n \mid (x_1,..., x_n)$ in $L'\}$.*

The proofs of Lemmas 1.1 and 1.2 can be easily modified to give the following result.

THEOREM 1.2. *$L \subseteq \Sigma^*$ is an n-RLSML if and only if there exists an n-RLCFL $L' \subseteq [\Sigma^*]^n$ such that $L = \{x_1 \cdots x_n \mid (x_1,..., x_n)$ in $L'\}$.*

COROLLARY 1.1. *Let $\Sigma = \{a\}$. If $L \subseteq \Sigma^*$ is an n-SML, then L is a regular set.*

Proof. Since L is an n-SML, there exists a CFG $G = \langle V, P, S, \Sigma_n\rangle$, $\Sigma_n = \{[a, i] \mid 1 \leqslant i \leqslant n\}$ such that $L = \{x_1 \cdots x_n \mid (x_1,..., x_n)$ in $\tau_n(L(G))\}$ (by Theorem 1.1). We may assume that the rules in P are of the form $A \to x$ or $A \to BC$, where x is in $\Sigma_n{}^*$ and A, B, C are in V. Define a CFG $G' = \langle V, P', S, \{a\}\rangle$, where P' is defined by cases.

(1) For each A in V, if $A \to A$ is in P, then let $A \to A$ be in P'.

(2) For all A, B, C in V, if $A \to BC$ is in P, then let $A \to BC$ be in P'.

(3) For $k \geqslant 1$, $1 \leqslant i_j \leqslant n$ $(1 \leqslant j \leqslant k)$, and each A in V, if $A \to [a, i_1][a, i_2] \cdots [a, i_k]$ is in P, then let $A \to a^k$ be in P'. Clearly, $L = L(G')$. The corollary now follows from the well-known fact that a context-free language over a single symbol is a regular set [5].

COROLLARY 1.2. *For each n-SMG $G_n = \langle V_1, \dots, V_n, P, S, \Sigma\rangle$ and x in Σ^*, it is recursively solvable*[3] *to determine whether x is in $L(G_n)$. Furthermore, it is recursively solvable to determine whether $L(G_n)$ is empty, finite, or infinite.*

Proof. Let $G = \langle V, P', S', \Sigma_n\rangle$ be a CFG such that $L(G_n) = \{x_1 \cdots x_n \mid (x_1, \dots, x_n) \text{ in } \tau_n(L(G))\}$. Now let x be in Σ^* and $P(x)$ be the set of all n-tuples $(x_1, \dots, x_n)$ such that $x = x_1 \cdots x_n$. Since $lg(x)$ is finite, $P(x)$ is finite. Then $\tau_n^{-1}(P(x))$ is finite. Hence, x is in $L(G_n)$ if and only if there exists an α in $\tau_n^{-1}(P(x))$ such that α is in $L(G)$. Since it is recursively solvable to determine whether an arbitrary word is in the language generated by a context-free grammar [5], the first part of the corollary follows. Now $L(G_n)$ is empty, finite, or infinite if and only if $L(G)$ is empty, finite, or infinite. This is again solvable since this problem is known to be solvable for context-free grammars [5].

We now use Theorem 1.1 to establish a connection between simple matrix languages and a well-known family of devices, the deterministic linear bounded automata. Intuitively, these devices are Turing machines [11] which are constrained to work on a finite input tape. The formal definition follows.

DEFINITION. A *deterministic linear bounded automaton* (abbreviated, *dlba*) is a 7-tuple $A = \langle K, \Gamma, ¢, \$, M, s_0, F\rangle$, where

(1) K and Γ are finite nonempty sets (of *states* and *inputs*, respectively), with $K \cap \Gamma = \varnothing$,

(2) s_0 in K (*start state*) and $F \subseteq K$ (*final states*),

(3) $¢$ and $\$$ are symbols not in $K \cup \Gamma$,

(4) M is a mapping from $K \times (\Gamma \cup \{¢, \$\})$ into $K \times (\Gamma \cup (¢, \$\}) \times \{-1, 0, 1\}$ satisfying the following requirements: For s and s' in K, a in Γ, b in $\Gamma \cup \{¢, \$)$, and d in $\{-1, 0, 1\}$:

[3] See [4] for the definition of recursively solvable problems.

(a) $M(s, ¢) = (s', b, d)$ implies $b = ¢$ and $d \geqslant 0$,

(b) $M(s, \$) = (s', b, d)$ implies b $= \$$,

(c) $M(s, a) = (s', b, d)$ implies b in Γ.

DEFINITION. *A configuration* of a dlba $A = \langle K, \Gamma, ¢, \$, M, s_0, F\rangle$ is any element of $(\Gamma \cup \{¢, \$\})^* K(\Gamma \cup \{¢, \$\})^*$. A configuration $a_1 \cdots a_{i-1} s a_i \cdots a_m$, each a_j in $\Gamma \cup \{¢, \$\}$, s in K is to be interpreted as the dlba A reading the i-th symbol of $a_1 \cdots a_m$ in state s.

We now describe a relation $\overset{*}{\vdash}$ on configurations.

DEFINITION. Let $A = \langle K, \Gamma, ¢, \$, M, s_0, F\rangle$ be a dlba. Define the relation $\vdash$ on configurations as follows. Let u, v in $(\Gamma \cup \{¢, \$\})^*$; a, b, c in $\Gamma \cup \{¢, \$\}$; s, s' in K. Then

(1) $ucsav \vdash us'cbv$ if $M(s, a) = (s', b, -1)$,

(2) $usav \vdash us'bv$ if $M(s, a) = (s', b, 0)$,

(3) $usav \vdash ubs'v$ if $M(s, a) = (s', b, 1)$.

For configurations α and β, we write $\alpha \overset{*}{\vdash} \beta$ if and only if there exist $r > 0$ and configurations $\alpha_0, \ldots, \alpha_r$ such that $\alpha_0 = \alpha$, $\alpha_r = \beta$, and $\alpha_i \vdash \alpha_{i+1}$ for $0 \leqslant i < r$.

DEFINITION. A word w in Γ^* is accepted by a dlba $A = \langle K, \Gamma, ¢, \$, M, s_0, F\rangle$ if there exist s in F and v in $(\Gamma \cup \{¢, \$\})^*$ such that $s_0 ¢ x \$ \overset{*}{\vdash} vs$ (note that by the constraints on M, v is always of the form $¢y\$$ for some y in Γ^*). The set of all words accepted by A is denoted by $T(A)$.

Notation. Let $\Sigma = \{a_1, \ldots, a_k\}$. Then $\Sigma_n = \{[a_i, j] \mid 1 \leqslant i \leqslant k, 1 \leqslant j \leqslant n\}$. Order the elements of Σ_n as follows. For $[a_i, j]$ and $[a_t, l]$ in Σ_n, let $[a_i, j] < [a_t, l]$ if either $j < l$ or $j = l$ and $i < t$. Let $[a_i, j] \leqslant [a_t, l]$ if either $[a_i, j] = [a_t, l]$ or $[a_i, j] < [a_t, l]$. For $m \geqslant 1$, $u_1, \ldots, u_m$, $v_1, \ldots, v_m$ in Σ_n, let $u_1 \cdots u_m < v_1 v_2 \cdots v_m$ if there exists $s (1 \leqslant s \leqslant m)$ such that $u_s < v_s$ and $u_{s+r} \leqslant v_{s+r}$ for each $r \geqslant 0$ such that $s + r \leqslant m$. Clearly, there are $\lambda = (kn)^m$ distinct words of length m in Σ_n^*. For each $m \geqslant 1$, let $\Sigma_n^m = \{\delta_1, \delta_2, \ldots, \delta_\lambda\}$ be the set of all words of length m in Σ_n^* such that $\delta_1 < \delta_2 < \cdots < \delta_\lambda$.

THEOREM 1.3. *Let $\Sigma = \{a_1, \ldots, a_k\}$. If $L \subseteq \Sigma^*$ is an n-SML, then a dlba $B = \langle K_B, \Gamma_B, ¢, \$, M_B, q_0, F_B\rangle$ can be effectively constructed such that $T(B) = L$. Moreover, there are sets accepted by dlba's which are not simple matrix languages.*

Proof. We need only prove the first statement of the theorem since the second statement follows from Corollary 1.1 and the fact that there are nonregular sets over single-letter alphabets which are accepted by dlba's.

By Theorem 1.1, there exists a context-free language $L' \subseteq \Sigma_n{}^*$ such that $L = \{x_1 \cdots x_n \mid (x_1, ..., x_n) \text{ in } \tau_n(L')\}$, and a context-free grammar generating L' can be effectively constructed. Since for every context-free grammar G, we can effectively construct a dlba accepting $L(G)$ [11], we can effectively construct a dlba $A = \langle K, \Gamma, ¢, \$, M, s_0, F\rangle$ such that $T(A) = L'$. Without loss of generality, we may assume that Λ is not in L. Then Λ is not in L'. For each x in $\Sigma^*\Sigma$, let $P(x) = \{(x_1, ..., x_n) \mid x = x_1 \cdots x_n\}$. Intuitively, the dlba B operates as follows. Given x in $\Sigma^*\Sigma$, $lg(x) = m \geqslant 1$, B successively generates each δ_p in $\Sigma_n{}^m$ $(1 \leqslant p \leqslant \lambda)$ (we are using the notation above) and checks whether δ_p is in $\tau_n^{-1}(P(x)) \cap T(A)$. If there exists δ_p in $\Sigma_n{}^m$ $(1 \leqslant p \leqslant \lambda)$ such that δ_p is in $\tau_n^{-1}(P(x)) \cap T(A)$, then B accepts the input; otherwise, B rejects the input. It is clear that B operating in this way would accept L. We now define B formally.

Let $*$, β, $q_0, q_1, q_2, q_3, q_4, q_5, q_6, t_1, t_2, t_3, p_1, p_2$ be new symbols not in $(K \cup \Gamma \cup \{¢, \$\})$. For each $1 \leqslant i \leqslant k$ and each $1 \leqslant j \leqslant n$, let $[t_1, j]$, $[t_2, j]$, $[t_3, j]$, $[p_1, i, j]$, $[p_2, i, j]$ be abstract symbols. Let $B = \langle K_B, \Gamma_B, ¢, \$, M_B, q_0, F\rangle$, where $K_B = K \cup \{q_i \mid 0 \leqslant i \leqslant 6\} \cup \{[t_1, j], [t_2, j], [t_3, j] \mid 1 \leqslant j \leqslant n\} \cup \{[p_1, i, j], [p_2, i, j] \mid 1 \leqslant i \leqslant k, 1 \leqslant j \leqslant n\}$, $\Gamma_B = \Sigma \cup \{\beta, *\} \times \Sigma \times \Sigma_n \times \Gamma \times \{\beta, *\}$, and M_B is defined as follows (for ease of exposition, we describe M_B in 4 phases).

Phase 1: *Initialization*

Given x in $\Sigma^*\Sigma$, $lg(x) = m \geqslant 1$, B begins by dividing the input tape into five tracks. Tracks 1 and 5 will initially contains all β's (β stands for a blank). Track 2 will always contain x, and tracks 3 and 4 will initially contain the word $\delta_1 = [a_1, 1] \cdots [a_1, 1]$ in $\Sigma_n{}^m$. Thus, the track symbols of B are five-tuples, each component representing a symbol of one track. Upon completion of Phase 1, B then goes to Phase 2 in state $[t_1, 1]$.

For $1 \leqslant i \leqslant k$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ in $\Gamma_B - \Sigma$:

(1) $M_B(q_0, ¢) = (q_0, ¢, 1)$;

(2) $M_B(q_0, a_i) = (q_0, (\beta, a_i, [a_1, 1], [a_1, 1], \beta), 1)$;

(3) $M_B(q_0, \$) = (q_1, \$, -1)$;

(4) $M_B(q_1, (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)) = (q_1, (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5), -1)$;

(5) $M_B(q_1, ¢) = ([t_1, 1], ¢, 1)$.

Phase 2: *Comparison*

B checks whether the word δ_p $(1 \leqslant p \leqslant \lambda)$ which appears in tracks 3 and 4 is in $\tau_n^{-1}(P(x))$. If successful, B goes to Phase 4 in state s_0; otherwise, B goes to Phase 3 in state q_3.

For $1 \leqslant i \leqslant k$, $1 \leqslant j, l \leqslant n$, $(\alpha_1, \alpha_2, [a_i, l], [a_i, l], \alpha_5)$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ in $\Gamma_B - \Sigma$:

$$(6) \quad M_B([t_1, j], (\alpha_1, \alpha_2, [a_i, l], [a_i, l], \alpha_5) = \begin{cases} ([t_1, j], (\alpha_1, \alpha_2, [a_i, l], [a_i, l], \alpha_5), 1) \text{ if } j \neq l \text{ or } \alpha_5 = *, \\ ([p_1, i, j], (\alpha_1, \alpha_2, [a_i, j], [a_i, j], *), -1) \text{ if } j = l \text{ and } \alpha_5 = \beta; \end{cases}$$

$$(7) \quad M_B([p_1, i, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)) = ([p_1, i, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5), -1);$$

$$(8) \quad M_B([p_1, i, j], ¢) = ([p_2, i, j], ¢, 1);$$

$$(9) \quad M_B([p_2, i, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)) = \begin{cases} ([t_3, j], (*, \alpha_2, \alpha_3, \alpha_4, \alpha_5), -1) \text{ if } \alpha_1 = \beta \text{ and } \alpha_2 = a_i, \\ ([p_2, i, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5), 1) \text{ if } \alpha_1 = *, \\ (q_2, (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5), -1) \text{ if } \alpha_1 = \beta \text{ and } \alpha_2 \neq a_i; \end{cases}$$

$$(10) \quad M_B([p_2, i, j], \$) = (q_2, \$, -1);$$

$$(11) \quad M_B([t_3, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)) = ([t_3, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5), -1);$$

$$(12) \quad M_B([t_3, j], ¢) = ([t_1, j], ¢, 1);$$

$$(13) \quad M_B([t_1, j], \$) = ([t_2, j], \$, -1);$$

$$(14) \quad M_B([t_2, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)) = ([t_2, j], (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5), -1);$$

$$(15) \quad M_B([t_2, j], ¢) = \begin{cases} ([t_1, j+1], ¢, 1) & \text{if } j < n, \\ (s_0, ¢, 0) & \text{if } j = n; \end{cases}$$

$$(16) \quad M_B(q_2, (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)) = (q_2, (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5), -1);$$

$$(17) \quad M_B(q_2, ¢) = (q_3, ¢, 1).$$

Phase 3: *Modification of the Contents of Tracks* 1, 3, 4, *and* 5

Let δ_p $(1 \leqslant p \leqslant \lambda)$ be the word which appears in track 3 when B enters Phase 3. If $p < \lambda$, then (a) δ_p in track 3 is replaced by δ_{p+1}; (b) δ_{p+1} is copied in track 4; (c) tracks 1 and 5 are filled with β's, and (d) B goes to Phase 2 via state q_1. If $p = \lambda$, then B rejects the input in a nonfinal state q_5.

For $1 \leqslant i \leqslant k$, $1 \leqslant j \leqslant n$, $(\alpha_1, \alpha_2, [a_i, j], \alpha_4, \alpha_5)$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ in $\Gamma_B - \Sigma$:

(18) $M_B(q_3, (\alpha_1, \alpha_2, [a_i, j], \alpha_4, \alpha_5))$

$$= \begin{cases} (q_4, (\beta, \alpha_2, [a_{i+1}, j], [a_{i+1}, j], \beta), 1) \text{ if } i < k, \\ (q_4, (\beta, \alpha_2, [a_1, j+1], [a_1, j+1], \beta), 1) \text{ if } i = k \text{ and } j < n, \\ (q_3, (\beta, \alpha_2, [a_1, 1], [a_1, 1], \beta), 1) \text{ if } i = k \text{ and } j = n; \end{cases}$$

(19) $M_B(q_4, (\alpha_1, \alpha_2, [a_i, j], \alpha_4, \alpha_5) = (q_4, (\beta, \alpha_2, [a_i, j], [a_i, j], \beta), 1)$;

(20) $M_B(q_4, \$) = (q_1, \$, -1)$;

(21) $M_B(q_3, \$) = (q_5, \$, 1)$.

Phase 4: *Simulation of A*

In this phase, B simulates the action of A on track 4 (leaving the contents of tracks 1, 2, 3, and 5 unaltered). If A would go off the right end of the input in a final state, so would B; otherwise B would go to the left end of the input and enter Phase 3 via state q_2.

For s in K, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ in $\Gamma_B - \Sigma$, d in $\{-1, 0, 1\}$:

(22) $M_B(s, (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)) = (s', (\alpha_1, \alpha_2, \alpha_3, \alpha_4', \alpha_5), d)$ if $M(s, \alpha_4) = (s', \alpha_4', d)$;

(23) $M_B(s, \$) = (s', \$, d)$ if $M(s, \$) = (s', \$, d)$ and either $d \neq 1$ or s' in F;

(24) $M_B(s, \$) = (q_2, \$, -1)$ if $M(s, \$) = (s', \$, 1)$ and s' not in F.

In addition to the rules given above, let $M_B(r, \gamma) = (q_6, \gamma, 1)$ for all (r, γ) in $K_B \times \Gamma_B$ not previously defined (including $r = q_6$).

It is clear that M_B is single-valued and B is a dlba. It is easily argued formally that $T(B) = L$.

2. Hierarchy of Classes of Simple Matrix Languages (Right-Linear Simple Matrix Languages)

In this section, we shall prove that there is a hierarchy of classes of languages generated by simple matrix grammars (right-linear simple matrix grammars). In order to do this, we must establish a necessary condition for a subset of Σ^* to be an n-SML (n-RLSML). First, we quote a lemma in [3].

LEMMA 2.1. *For each infinite context-free language L, there exist integers p and q with the property that each z in L, $lg(z) > p$, is of the form $z = xuwvy$, where $uv \neq \Lambda$, $lg(uwv) \leqslant q$, and xu^kwv^ky is in L for each $k \geqslant 1$.*

We now prove an analog of this result for n-SML's.

THEOREM 2.1. *Let $L \subseteq \Sigma^*$ be an infinite n-SLM. There exist integers p and q with the property that each z_1 in L, $lg(z_1) > p$, is of the form $z_1 = x_1'u_1w_1v_1x_2' \cdots x_n'u_nw_nv_nx'_{n+1}$, where $u_1v_1 \cdots u_nv_n \neq \Lambda$, $lg(u_1w_1v_1 \cdots u_nw_nv_n) \leqslant q$ and $x_1'u_1{}^kw_1v_1{}^kx_2' \cdots x_n'u_n{}^kw_nv_n{}^kx'_{n+1}$ is in L for each $k \geqslant 1$.*

Proof. By Theorem 1.1, there exists a CFL $L' \subseteq \Sigma_n{}^*$ such that $L = \{\alpha_1 \cdots \alpha_n \mid (\alpha_1, ..., \alpha_n) \text{ in } \tau_n(L')\}$. Clearly, L' is infinite since L is. For the CFL L', let p and q have the same significance as in Lemma 2.1. Now let z_1 in L be such that $lg(z_1) > p$. Then there exists $(\alpha_1, ..., \alpha_n)$ in $\tau_n(L')$ such that $z_1 = \alpha_1 \cdots \alpha_n$. Hence, there exist $a_1, ..., a_m$ in Σ_n ($m \geqslant 1$) such that $a_1 \cdots a_m$ is in L', $\tau_n(a_1 \cdots a_m) = (\alpha_1, ..., \alpha_n)$, and $m = lg(\alpha_1) + \cdots + lg(\alpha_m) = lg(z_1) > p$. Thus, $a_1 \cdots a_m$ will decompose as in Lemma 2.1, that is, $a_1 \cdots a_m = xuwvy$, $uv \neq \Lambda$, $lg(uwv) \leqslant q$, and xu^kwv^ky is in L' for each $k \geqslant 1$. Then $\tau_n(xu^kwv^ky)$ is in $\tau_n(L')$ for every $k \geqslant 1$. Let $\tau_n(x) = (x_1, ..., x_n)$, $\tau_n(u) = (u_1, ..., u_n)$, $\tau_n(w) = (w_1, ..., w_n)$, $\tau_n(v) = (v_1, ..., v_n)$, and $\tau_n(y) = (y_1, ..., y_n)$. Then $u_1v_1u_2v_2 \cdots u_nv_n \neq \Lambda$ since $lg(u_1v_1 \cdots u_nv_n) = lg(uv)$, and

$$lg(u_1w_1v_1 \cdots u_nw_nv_n) = lg(uwv) \leqslant q.$$

Also,

$$\begin{aligned}\tau_n(xu^kwv^ky) &= \tau_n(x)(\tau_n(u))^k\tau_n(w)(\tau_n(v))^k\tau_n(y)\\ &= (x_1, ..., x_n)(u_1, ..., u_n)^k(w_1, ..., w_n)(v_1, ..., v_n)^k(y_1, ..., y_n)\\ &= (x_1u_1{}^kw_1v_1{}^ky_1, ..., x_nu_n{}^kw_nv_n{}^ky_n)\end{aligned}$$

is in $\tau_n(L')$ for each $k \geqslant 1$. We have, therefore, $z_1 = x_1u_1w_1v_1y_1 \cdots x_nu_nw_nv_ny_n$ in L, $u_1v_1 \cdots u_nv_n \neq \Lambda$, $lg(u_1w_1v_1 \cdots u_nw_nv_n) \leqslant q$, and $x_1u_1{}^kw_1v_1{}^ky_1 \cdots x_nu_n{}^kw_nv_n{}^ky_n$ is in L for each $k \geqslant 1$. The theorem follows by letting $x_1' = x_1$, $x_2' = y_1x_2, ..., x_n' = y_{n-1}x_n$, and $x'_{n+1} = y_n$.

Notation. $\mathscr{S}(n)$ will denote the class of n-SML's and $\mathscr{S} = \bigcup_{n\geqslant 1} \mathscr{S}(n)$. Similarly, $\mathscr{R}(n)$ will denote the class of n-RLSML's and $\mathscr{R} = \bigcup_{n\geqslant 1} \mathscr{R}(n)$.

We shall show that for each $n \geqslant 1$, $\mathscr{S}(n+1)$ properly contains $\mathscr{S}(n)$, thus establishing a hierarchy of classes of simple matrix languages.

THEOREM 2.2. *For each $n \geqslant 1$, $\mathscr{S}(n+1)$ properly contains $\mathscr{S}(n)$.*

Proof. Let $L \subseteq \Sigma^*$ be in $\mathscr{S}(n)$. Then by Theorem 1.1, there exists a CFL $L' \subseteq \Sigma_n{}^*$ such that $L = \{x_1 \cdots x_n \mid (x_1, ..., x_n) \text{ in } \tau_n(L')\}$. But L' is also a subset of Σ_{n+1}^* and $\tau_{n+1}(L') = \{(x_1, ..., x_n, \Lambda) \mid (x_1, ..., x_n) \text{ in } \tau_n(L')\}$. Thus, by Theorem 1.1, L is in $\mathscr{S}(n+1)$.

We now show that there is a language in $\mathscr{S}(n+1)$ which is not in $\mathscr{S}(n)$.

Let $a_1, a_2, \ldots, a_n, b, c_n, \ldots, c_2, c_1$ be distinct symbols and let Σ denote this set of symbols. Let $L_n = \{a_1^k a_2^k \cdots a_n^k b^k c_n^k \cdots c_2^k c_1^k \mid k \geqslant 1\}$. Clearly, L_n is in $\mathscr{S}(n+1)$. We shall show that L_n is not in $\mathscr{S}(n)$. Suppose it is. Then by Theorem 2.1 (with a change in notation), there exists an integer p, $x_1, y_1, x_2, y_2, \ldots, x_n, y_n, v, w_n, z_n, \ldots, w_2, z_2, w_1, z_1$ in $\{a_1, \ldots, a_n, b, c_n, \ldots, c_1\}^*$ such that $a_1^p a_2^p \cdots a_n^p b^p c_n^p \cdots c_2^p c_1^p = x_1 y_1 x_2 y_2 \cdots x_n y_n v w_n z_n \cdots w_2 z_2 w_1 z_1$, $y_1 y_2 \cdots y_n w_n \cdots w_2 w_1 \neq \Lambda$, and $x_1 y_1^k x_2 y_2^k \cdots x_n y_n^k v w_n^k z_n \cdots w_2^k z_2 w_1^k z_1$ is in L_n for each $k \geqslant 1$. Note that $y_1 y_2 \cdots y_n w_n \cdots w_2 w_1$ contains at least one symbol, say α, from Σ. Since $a_1, a_2, \ldots, a_n, b, c_n, \ldots, c_2, c_1$ are distinct, it is clear that none of the y_i, w_i $(1 \leqslant i \leqslant n)$ can contain 2 distinct symbols from Σ. Thus, $y_1 y_2 \cdots y_n w_n \cdots w_2 w_1$ contains at most $2n$ distinct symbols from Σ. Let β in Σ be a symbol not in $y_1 y_2 \cdots y_n w_n \cdots w_2 w_1$. Let m be the maximum of the lengths of $x_1, \ldots, x_n, v, z_n, \ldots, z_1$. Consider the word $x_1 y_1^{m+1} x_2 y_2^{m+1} \cdots x_n y_n^{m+1} v w_n^{m+1} z_n \cdots w_2^{m+1} z_2 w_1^{m+1} z_1$ in L_n. This word contains at least $m+1$ occurrences of α and at most m occurrences of β. This is a contradiction. It follows that L_n is not in $\mathscr{S}(n)$.

Remark. It is well-known that for each infinite regular set R, there exists an integer p with the property that each z in R, $lg(z) > p$, is of the form $z = xuy$, where $u \neq \Lambda$, and $xu^k y$ is in R for each $k \geqslant 1$ [13]. Using the same technique as in the proof of Theorem 2.1, we can prove that for each infinite n-RLSML L, there exists an integer p with the property that each z_1 in L, $lg(z_1) > p$, is of the form $z_1 = x_1 u_1 x_2 \cdots x_n u_n x_{n+1}$, $u_1 \cdots u_n \neq \Lambda$, and $x_1 u_1^k x_2 \cdots x_n u_n^k x_{n+1}$ is in L for each $k \geqslant 1$. Then we can prove the following theorem.

THEOREM 2.3. *For each $n \geqslant 1$, let $a_1, \ldots, a_{n+1}$ be distinct symbols. Let $L_n = \{a_1^k \cdots a_{n+1}^k \mid k \geqslant 1\}$. Then $\mathscr{R}(n+1)$ properly contains $\mathscr{R}(n)$ with L_n in $\mathscr{R}(n+1) - \mathscr{R}(n)$.*

COROLLARY 2.1. *For each $n \geqslant 1$, $\mathscr{S}(n)$ properly contains $\mathscr{R}(n)$.*

Proof. $L_n = \{a_1^k \cdots a_{n+1}^k \mid k \geqslant 1\}$ is certainly in $\mathscr{S}(n)$.

3. POSITIVE CLOSURE PROPERTIES

In this section, we shall demonstrate the positive closure properties of $\mathscr{S}(n)[\mathscr{R}(n)]$ under various operations. In particular, we shall show that $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under the operations of substitution by context-free languages [regular sets], union, transposition, intersection with regular sets, sequential transducer mapping, and mapping and pseudoinverse mapping by a nondeterministic generalized sequential machine.

DEFINITION. Let Σ be an alphabet. For each a in Σ let Σ^a be a finite nonempty set and $\phi(a)$ a subset of Σ^{a*}. Let $\phi(\Lambda) = \{\Lambda\}$ and $\phi(a_1 \cdots a_r) = \phi(a_1) \cdots \phi(a_r)$, each a_i in Σ. Then the mapping ϕ defined on Σ^* is called a *substitution*. If $L \subseteq \Sigma^*$, then $\phi(L) = \bigcup_{x \text{ in } L} \phi(x)$ is called the substitution of L by $\phi(a)$.

THEOREM 3.1. *For each $n \geqslant 1$, $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under substitution by context-free languages [regular sets].*

Proof. Let $L \subseteq \Sigma^*$ be in $\mathscr{S}(n)[\mathscr{R}(n)]$. For each a in Σ, let $\phi(a) \subseteq \Sigma^{a*}$ be a context-free language [regular set]. We shall show that $\phi(L)$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$. Since L is in $\mathscr{S}(n)[\mathscr{R}(n)]$, by Theorem 1.1 [Theorem 1.2], there exists a context-free language [regular set] $L' \subseteq \Sigma_n^*$ such that $L = \{x_1 \cdots x_n \mid (x_1, \ldots, x_n) \text{ in } \tau_n(L')\}$. Define a substitution $\hat{\phi}$[4] on Σ_n^* as follows. For each $[a, i]$ in Σ_n, let $\hat{\phi}([a, i]) = \{[b_1, i] \cdots [b_k, i] \mid k \geqslant 1, b_1, \ldots, b_k \text{ in } \Sigma^a, b_1 \cdots b_k \text{ in } \phi(a)\} \cup Q$ where $Q = \{\Lambda\}$ if Λ is in $\phi(a)$ and $Q = \varnothing$ otherwise. Obviously, $\hat{\phi}([a, i])$ is a context-free language [regular set] over $\Sigma_n{}^a$ if and only if $\phi(a)$ is a context-free language [regular set] over Σ^a. Let $\Sigma' = \bigcup_{a \text{ in } \Sigma} \Sigma^a$. Since context-free languages [regular sets] are closed under substitution by context-free languages [regular sets] (see [5]), $\hat{\phi}(L')$ is a context-free language [regular set] over $\Sigma_n{}'$. The theorem follows from the fact that $\phi(L) = \{y_1 \cdots y_n \mid (y_1, \ldots, y_n) \text{ in } \tau_n(\hat{\phi}(L'))\}$.

COROLLARY 3.1. *For each $n \geqslant 1$, $\mathscr{S}(n)$ is closed under substitution by regular sets.*

COROLLARY 3.2. *For each $n \geqslant 1$, $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under homomorphism.*[5]

In Section 4, we shall show that for each $n \geqslant 1$, $\mathscr{S}(n)[\mathscr{R}(n)]$ is not closed under substitution by languages in $\mathscr{S}(2)[\mathscr{R}(2)]$.

The proof of the following theorem is immediate.

THEOREM 3.2. *For each $n \geqslant 1$, $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under union and under transposition.*

[4] Note that we need only specify $\hat{\phi}$ for each element in Σ_n for $\hat{\phi}$ to be well defined.

[5] Let Σ and Δ be nonempty sets. A *homomorphism* ϕ from Σ^* into Δ^* is any mapping from Σ^* into Δ^* such that $\phi(\Lambda) = \Lambda$ and $\phi(a_1 \cdots a_k) = \phi(a_1) \cdots \phi(a_k)$ for each $k \geqslant 1$, a_i in Σ for $1 \leqslant i \leqslant k$.

THEOREM 3.3. *For each* $n \geqslant 1$, $\mathscr{S}(n)[\mathscr{R}(n)]$ *is closed under concatenation with context-free languages* [regular sets].

Proof. Let L be in $\mathscr{S}(n)[\mathscr{R}(n)]$ and L' be a context free language [regular set]. Let $G_n = \langle V_1, ..., V_n, P, S, \Sigma\rangle$ be an n-SMG [n-RLSMG] generating L. Let c be a new symbol not in $\Sigma \cup V_1 \cup \cdots \cup V_n \cup \{S\}$. Define an n-SMG [n-RLSMG] $G_n' = \langle V_1, ..., V_n, P', S, \Sigma \cup \{c\}\rangle$, where $P' = (P - \{$rules in P of the form $[S \to w]\}) \cup \{[S \to wc] \mid [S \to w]$ in $P\}$. Clearly, $L(G_n') = L(G_n)\{c\} = L\{c\}$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$. Define a substitution ϕ on $(\Sigma \cup \{c\})^*$ by: $\phi(a) = \{a\}$ for each a in Σ, and $\phi(c) = L'$. Then $\phi(L\{c\}) = LL'$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$ by Theorem 3.1. By a similar argument, we can show that $L'L$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$.

In the next section, we shall show that for each $n \geqslant 1$, $\mathscr{S}(n)[\mathscr{R}(n)]$ is not closed under concatenation with languages in $\mathscr{S}(2)[\mathscr{R}(2)]$ and therefore not closed under arbitrary concatenation.

We now examine the result of intersecting a language in $\mathscr{S}(n)[\mathscr{R}(n)]$ with a regular set. The following lemma will prove useful. The proof is simple and is omitted.

LEMMA 3.1. *Let* $G' = \langle V_1, ..., V_n, P', S, \Sigma\rangle$ *be an* n-SMG [n-RLSMG]. *Let* c *be a new symbol not in* $V_1 \cup \cdots \cup V_n \cup \Sigma \cup \{S\}$. *Define an* n-SMG $G_n = \langle V_1, ..., V_n, P', S, \Sigma \cup \{c\}\rangle$, *where* $P' = \{[S \to wc] \mid [S \to w]$ *in* $P\} \cup \{[A_1 \to w_1c, ..., A_n \to w_nc] \mid [A_1 \to w_1, ..., A_n \to w_n]$ *in* $P\}$. *Now let* $A' = \langle K, \Sigma, M', q_0, F\rangle$ *be a finite automaton. Modify* A' *to* $A = \langle K, \Sigma \cup \{c\}, M, q_0, F\rangle$, *where* M *is defined by*: $M(q, a) = M'(q, a)$ *for all* (q, a) *in* $K \times \Sigma$, *and* $M(q, c) = q$ *for all* q *in* K. *Let* ϕ *be a homomorphism from* $(\Sigma \cup \{c\})^*$ *into* Σ^* *which maps* a *into* a *and* c *into* Λ. *Then* $\phi(L(G_n) \cap T(A)) = L(G_n') \cap T(A')$.

THEOREM 3.4. *If* L *is in* $\mathscr{S}(n)[\mathscr{R}(n)]$ *and* R *is a regular set, then* $L \cap R$ *is in* $\mathscr{S}(n)[\mathscr{R}(n)]$.

Proof. Let $G_n' = \langle V_1, ..., V_n, P', S, \Sigma\rangle$ be an n-SMG [n-RLSMG] generating L and let $A' = \langle K, \Sigma, M', q_0, F\rangle$ be a finite automaton such that $R = T(A')$. Construct the corresponding $G_n = \langle V_1, ..., V_n, P, S, \Sigma \cup \{c\}\rangle$ and $A = \langle K, \Sigma \cup \{c\}, M, q_0, F\rangle$ of the preceding lemma. If we can show that $L(G_n) \cap T(A) \subseteq (\Sigma \cup \{c\})^*$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$, then by the preceding lemma and the fact that $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under homomorphism, $\phi(L(G_n) \cap T(A)) = L(G_n') \cap T(A')$ would be in $\mathscr{S}(n)[\mathscr{R}(n)]$, ϕ being the homomorphism defined in Lemma 3.1. Now $T(A) = \bigcup_{q \text{ in } F} T(A_q)$, where $A_q = \langle K, \Sigma \cup \{c\}, M, q_0, \{q\}\rangle$, and $L(G_n) \cap T(A) = \bigcup_{q \text{ in } F} (L(G_n) \cap T(A_q))$. Since $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under union, it suffices to prove that $L(G_n) \cap T(A)$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$ when A has only one final state, that is, $A = \langle K, \Sigma \cup \{c\}, M, q_0, \{f\}\rangle$.

Define an n-SMG [n-RLSMG] $\bar{G}_n = \langle \bar{V}_1, ..., \bar{V}_n, \bar{P}, \bar{S}, \Delta \rangle$, where $\bar{V}_i = \{(p, A, q) \mid p, q \text{ in } K, A \text{ in } V_i\} (1 \leqslant i \leqslant n)$, $\Delta = \{(p, a, q) \mid p, q \text{ in } K, a \text{ in } (\Sigma \cup \{c\})\}$, $\bar{S} = (q_0, S, f)$, and $\bar{P}$ is defined by cases.

(1) If $[S \to \alpha_1\alpha_2 \cdots \alpha_k]$ is in P, where $\alpha_1, ..., \alpha_k$ are in $(\Sigma \cup V_1 \cup \cdots \cup V_n \cup \{c\})$ (note that $k \geqslant 1$ by Lemma 3.1), then let $[(q_0, S, f) \to (p_0, \alpha_1, p_1)(p_1, \alpha_2, p_2) \cdots (p_{k-1}, \alpha_k, p_k)]$ be in $\bar{P}$, where $p_0, p_1, ..., p_k$ are in K, $p_0 = q_0$, $p_k = f$, and for $1 \leqslant i \leqslant k$, $M(p_{i-1}, \alpha_i) = p_i$ if α_i is in $(\Sigma \cup \{c\})$ (note that if $\alpha_i = c$, $p_i = p_{i-1}$ by Lemma 3.1).

(2) If $[A_1 \to \alpha_{11}\alpha_{12} \cdots \alpha_{1k(1)}, ..., A_n \to \alpha_{n1}\alpha_{n2} \cdots \alpha_{nk(n)}]$ is in P, each A_i in V_i $(1 \leqslant i \leqslant n)$, α_{ij} in $(V_i \cup \Sigma \cup \{c\})$ for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant k(i)$ (note that $k(i) \geqslant 1$ for $1 \leqslant i \leqslant n$ by Lemma 3.1), then let $[(p_{10}, A_1, p_{1k(1)}) \to (p_{10}, \alpha_{11}, p_{11})(p_{11}, \alpha_{12}, p_{12}) \cdots (p_{1(k(1)-1)}, \alpha_{1k(1)}, p_{1k(1)}), ..., (p_{n0}, A_n, p_{nk(n)}) \to (p_{n0}, \alpha_{n1}, p_{n1})(p_{n1}, \alpha_{n2}, p_{n2}) \cdots (p_{n(k(n)-1)}, \alpha_{nk(n)}, p_{nk(n)})]$ be in $\bar{P}$, where p_{ij} is in K and $M(p_{i(j-1)}, \alpha_{ij}) = p_{ij}$ for α_{ij} in $\Sigma \cup \{c\}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant k(i)$ (note again that $p_{ij} = p_{i(j-1)}$ if $\alpha_{ij} = c$ by Lemma 3.1).

It is easily verified by induction on t that if $S \overset{*}{\Rightarrow} a_1a_2 \cdots a_t$ in G_n, $a_1, a_2, ..., a_t$ in $\Sigma \cup \{c\}$, and $M(q_0, a_1) = p_1$, $M(p_1, a_2) = p_2, ..., M(p_{t-1}, a_t) = f$ in A for some $p_1, ..., p_{t-1}$ in K, then $(q_0, S, f) \overset{*}{\Rightarrow} (q_0, a_1, p_1)(p_1, a_2, p_2) \cdots (p_{t-1}, a_t, f)$ in $\bar{G}_n$ and conversely. Let $\hat{\phi}$ be a homomorphism from Δ^* into $(\Sigma \cup \{c\})^*$ which maps each (p, a, q) in Δ into a. Then $\hat{\phi}(L(\bar{G})) = L(G_n) \cap T(A)$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$ since $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under homomorphism, completing the proof.

COROLLARY 3.3. *If L is in $\mathscr{S}(n)[\mathscr{R}(n)]$ and R is a regular set, then $L - R$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$.*

Proof. $L - R = L \cap (\Sigma^* - R)$. The corollary now follow from Theorem 3.4, and the fact that regular sets are closed under complementation.

Let us now consider the effect of mapping devices to the languages in $\mathscr{S}(n)[\mathscr{R}(n)]$.

DEFINITION. A *sequential transducer* is a 5-tuple $M = \langle K, \Sigma, \Delta, H, q_0 \rangle$, where

(1) K, Σ, Δ are finite nonempty sets (of *states*, *inputs*, and *outputs*, respectively);

(2) q_0 is in K (the *start state*);

(3) H is a finite subset of $K \times \Sigma^* \times \Delta^* \times K$.

The sequential transducer operates as follows.

DEFINITION. For each sequential transducer $M = \langle K, \Sigma, \Delta, H, q_0 \rangle$ and x in Σ^*, let $M(x)$ be the set of words y in Δ^* with the property that there exist $x_1, \ldots, x_k$ in Σ^*, $y_1, \ldots, y_k$ in Δ^*, and $q_1, \ldots, q_k$ in K such that $x = x_1 \cdots x_k$, $y = y_1 \cdots y_k$, and $(q_i, x_{i+1}, y_{i+1}, q_{i+1})$ is in H for each $0 \leqslant i \leqslant k-1$. Let $M(L) = \bigcup_{x \text{ in } L} M(x)$ for each $L \subseteq \Sigma^*$. The function M so defined is called a *sequential transducer mapping*.

The theorem that follows was proved for context-free languages and regular sets in [5]. However, in the proof of the theorem, the only requirements for the theorem to hold for an arbitrary class of languages is for the class to contain all regular sets and for it to be closed under the operations of union, substitution by regular sets, and intersection with regular sets. Thus, the theorem also holds for $\mathscr{S}(n)$ and $\mathscr{R}(n)$ since we have already shown that these classes have the required properties.

THEOREM 3.5. *For each $n \geqslant 1$, $\mathscr{S}(n)$ and $\mathscr{R}(n)$ are closed under sequential transducer mappings.*

COROLLARY 3.4. *If ϕ is a homomorphism from Σ^* into Δ^* and $L \subseteq \Delta^*$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$, then $\phi^{-1}(L) = \{x \text{ in } \Sigma^* \mid \phi(x) \text{ in } L\}$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$. Thus, $\mathscr{S}(n)[\mathscr{R}(n)]$ is closed under inverse homomorphism.*

Proof. Consider the sequential transducer $M = \langle \{q_0\}, \Delta, \Sigma, H, q_0 \rangle$, where $H = \{(q_0, \Lambda, \Lambda, q_0)\} \cup \{(q_0, \phi(a), a, q_0) \mid a \text{ in } \Sigma\}$. Clearly, $M(L) = \phi^{-1}(L)$, and by Theorem 3.5, $M(L)$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$.

A special case of a sequential transducer is given in the next definition.

DEFINITION. A *nondeterministic gsm* (or *nondeterministic generalized sequential machine*) is a 5-tuple $M = \langle K, \Sigma, \Delta, \lambda, q_0 \rangle$, where

(1) K, Σ, Δ, and q_0 have the same significance as in a sequential transducer, and

(2) λ is a mapping from $K \times (\Sigma \cup \{\Lambda\}) \times K$ into the finite subsets of Δ^* satisfying the requirements that $\lambda(q, \Lambda, q) = \{\Lambda\}$ for each q in K, and $\lambda(q, \Lambda, q') = \varnothing$ if $q \neq q'$, q, q' in K.

The nondeterministic *gsm* effects an operation as follows.

DEFINITION. For each nondeterministic *gsm* $M = \langle K, \Sigma, \Delta, \lambda, q_0 \rangle$ and each x in Σ^*, let $M(x)$ be the union of all sets of the form $\lambda(q_0, x_{i_1}, q_{i_1})$

$\lambda(q_{i_1}, x_{i_2}, q_{i_2}) \cdots \lambda(q_{i_{n-1}}, x_{i_n}, q_{i_n})$ where $n \geqslant 1$, $x = x_{i_1} \cdots x_{i_n}$, each x_{i_j} in $\Sigma \cup \{\Lambda\}$, each q_{i_j} in K. If $L \subseteq \Sigma^*$, then $M(L) = \bigcup_{x \text{ in } L} M(x)$ is called the *nondeterministic gsm mapping of L*. If $L \subseteq \Sigma^*$, then $M^{-1}(L) = \{x \text{ in } \Sigma^* \mid M(x) \cap L \neq \varnothing\}$ is called the *pseudoinverse nondeterministic gsm mapping* of L.

In [6], it is shown that any class of languages closed under the operations of union, homomorphism, inverse homomorphism, and intersection with regular sets is also closed under mappings and pseudo-inverse mappings by nondeterministic *gsm*'s. Thus, we have the following theorem for $\mathscr{S}(n)$ and $\mathscr{R}(n)$.

THEOREM 3.6. *For each $n \geqslant 1$, $\mathscr{S}(n)$ and $\mathscr{R}(n)$ are closed under mappings and pseudoinverse mappings by nondeterministic gsms.*

The following result is of interest.

THEOREM 3.7. *Let $M = \langle K, \Sigma, \Delta, \lambda, q_0 \rangle$ be a nondeterministic gsm. Let d be a symbol not in $\Sigma \cup \Delta$. Let $L = \{xdy \mid x \text{ in } \Sigma^*, y \text{ in } M(x)\}$. Then L is in $\mathscr{R}(2)$.*

Proof. Given $M = \langle K, \Sigma, \Delta, \lambda, q_0 \rangle$, let S and B be two new symbols not in K. Define a 2-RLSMG $G_2 = \langle K, \{B\}, P, S, \Sigma \cup \{d\} \rangle$, where $P = \{[S \to q_0 dB]\} \cup \{[q \to aq', B \to wB] \mid a \text{ in } \Sigma \cup \{\Lambda\}, \lambda(q, a, q') \neq \varnothing, w \text{ in } \lambda(q, a, q')\}$. Clearly, $L(G_2) = \{xdy \mid x \text{ in } \Sigma^*, y \text{ in } M(x)\}$.

We close this section with the following theorem which was shown to be true for any class of languages closed under the operations of union, homomorphism, inverse homomorphism, and intersection with regular sets [6].

THEOREM 3.8. *For each $n \geqslant 1$, L in $\mathscr{S}(n)[\mathscr{R}(n)]$ and R a regular set:*

(1) $L/R = \{x \mid xy \text{ in } L \text{ for some } y \text{ in } R\}$,

(2) $R\backslash L = \{x \mid yx \text{ in } L \text{ for some } y \text{ in } R\}$,

(3) *Init* $(L) = \{x \neq \Lambda \mid xy \text{ in } L \text{ for some } y\}$,

(4) *Fin* $(L) = \{x \neq \Lambda \mid yx \text{ in } L \text{ for some } y\}$,

(5) *Sub* $(L) = \{x \neq \Lambda \mid yxz \text{ in } L \text{ for some } y, z\}$,

are all in $\mathscr{S}(n)[\mathscr{R}(n)]$.

4. NEGATIVE CLOSURE PROPERTIES

In this section, we shall prove a number of negative results about $\mathscr{S}(n)$, $n \geqslant 1$ [$\mathscr{R}(m)$, $m \geqslant 2$]. In particular, we shall show that $\mathscr{S}(n)[\mathscr{R}(m)]$ is not closed under intersection, under complementation, under concatenation, and under the star operator. In fact, we shall prove that $\mathscr{S}[\mathscr{R}]$ is not closed under the operations of intersection, complementation, and star. In the case of $\mathscr{R}$, we shall demonstrate that there are context free languages not in $\mathscr{R}$.

THEOREM 4.1. *There exist languages L_1 and L_2 in $\mathscr{R}(2)$ such that $L_1 \cap L_2$ is not in $\mathscr{R}$.*

Proof. Let a, b, c, $\hat{a}$, $\hat{b}$ be distinct symbols. Let $\hat{\Lambda} = \Lambda$ and $\hat{x} = \hat{\alpha}_1 \cdots \hat{\alpha}_k$ for each $x = \alpha_1 \cdots \alpha_k$, all α_i in $\{a, b\}$. Let $L_1 = \{xc\hat{x} \mid x \text{ in } \{a, b\}^*\}$ and $L_2 = \{a^{i_1}b^{j_1} \cdots a^{i_k}b^{j_k}c\hat{a}^{l_1}\hat{b}^{i_1} \cdots \hat{a}^{l_k}\hat{b}^{i_k} \mid k \geqslant 1, i_r, j_r, l_r \geqslant 1 \text{ for } 1 \leqslant r \leqslant k\} \cup \{c\}$. Clearly, L_1 is in $\mathscr{R}(2)$. Let $G_2 = \langle\{A_1, A_2, A_3, A_4\}, \{B_1, B_2, B_3, B_4\}, P, S, \{a, b, c, \hat{a}, \hat{b}\}\rangle$ be a 2-RLSMG, where $P = \{[S \to c], [S \to A_1cB_1], [A_1 \to A_2, B_1 \to \hat{a}B_2], [A_2 \to A_2, B_2 \to \hat{a}B_2], [A_2 \to aA_3, B_2 \to \hat{b}B_3], [A_3 \to aA_3, B_3 \to \hat{b}B_3], [A_3 \to bA_4, B_3 \to B_4], [A_4 \to bA_4, B_4 \to B_4], [A_4 \to \Lambda, B_4 \to \Lambda], [A_4 \to A_2, B_4 \to \hat{a}B_2]\}$. It is easily verified that $L(G_2) = L_2$. Hence, L_2 is in $\mathscr{R}(2)$. Suppose $L_1 \cap L_2$ is in $\mathscr{R}$. Let ϕ be a homomorphism which maps a into a, b into b, c into Λ, $\hat{a}$ into Λ, and $\hat{b}$ into Λ. Then $\phi(L_1 \cap L_2)$ is in $\mathscr{R}$ since $\mathscr{R}$ is closed under homomorphism. However, $\phi(L_1 \cap L_2) = \{a^kb^k \mid k \geqslant 1\}^*$ is not in $\mathscr{R}$ as we shall see in Theorem 4.7.

THEOREM 4.2. *There exist languages L_1 and L_2 in $\mathscr{S}(1)$ such that $L_1 \cap L_2$ is not in $\mathscr{S}$.*

Proof. Let $L_1 = \{a^nb^nc^m \mid n, m \geqslant 1\}^*$ and $L_2 = \{a^mb^nc^n \mid n, m \geqslant 1\}^*$. Clearly, L_1 and L_2 are in $\mathscr{S}(1)$. In Theorem 4.6, we shall show that $L_1 \cap L_2 = \{a^nb^nc^n \mid n \geqslant 1\}^*$ is not in $\mathscr{S}$.

COROLLARY 4.1. *$\mathscr{S}(n)$, $n \geqslant 1$ $(\mathscr{S})[\mathscr{R}(m), m \geqslant 2]\{\mathscr{R}\}$ is not closed under the operations of intersection and complementation.*

Proof. $\mathscr{S}(n)(\mathscr{S})[\mathscr{R}(m)]\{\mathscr{R}\}$ is closed under union by Theorem 3.2. The corollary now follows from Theorem 4.1 and 4.2 and DeMorgan's law.

We shall show through a sequence of lemmas that for each $n \geqslant 1$, $\mathscr{S}(n)[\mathscr{R}(n)]$ is not closed under concatenation with languages in $\mathscr{S}(2)[\mathscr{R}(2)]$.

Notation. Let $n \geqslant 1$ and $a_1, \ldots, a_{9n}$ be distinct symbols. Let $x = a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_{9n-2}^{k_{3n}}a_{9n-1}^{k_{3n}}a_{9n}^{k_{3n}}$ for some $k_i \geqslant 1$, $1 \leqslant i \leqslant 3n$. Let $x_1, \ldots, x_n$ in $\{a_1, \ldots, a_{9n}\}^*$ be such that $x = x_1 \cdots x_n$. Define $f(x_1, \ldots, x_n) = (i_0, \ldots, i_n)$ where $i_0 = 0$ and for $1 \leqslant j \leqslant n$, $i_j = i_{j-1}$ if $x_j = \Lambda$ and $i_j = r$ if $x_j \neq \Lambda$ and the final symbol of x_j is a_r $(1 \leqslant r \leqslant 9n)$.

LEMMA 4.1. *Let* $x = a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_{9n-2}^{k_{3n}}a_{9n-1}^{k_{3n}}a_{9n}^{k_{3n}}$. *Let* $x_1, ..., x_n$ *be such that* $x = x_1 \cdots x_n$, *and* $f(x_1, ..., x_n) = (i_0, ..., i_n)$. *Then the following properties hold:*

(1) $0 = i_0 \leqslant \cdots \leqslant i_n = 9n$.

(2) *There exists* j, $0 \leqslant j \leqslant n-1$ *such that* $i_{j+1} - i_j \geqslant 9$.

(3) *There exists an* r *in* $R = \{3m + 1 \mid 1 \leqslant m \leqslant 3n-2\}$, α_{k_p} *in* $\{a_1, ..., a_{r-1}\}^*$, β_{k_p} *in* $\{a_{r+3}, ..., a_{9n}\}^*$, $t_{k_p} \geqslant 1$, $q_{k_p} \geqslant 1$ *such that*

$$x_{j+1} = \alpha_{k_p} a_{r-1}^{t_{k_p}} a_r^{k_p} a_{r+1}^{k_p} a_{r+2}^{k_p} a_{r+3}^{q_{k_p}} \beta_{k_p},$$

where $p = [r/3] + 1$.[6]

Proof. That (1) is true is obvious by the definition of $f(x_1, ..., x_n)$. Now assume (2) is false. Then for all j, $i_{j+1} - i_j \leqslant 8$. Then $(i_n - i_{n-1}) + \cdots + (i_1 - i_0) = i_n - i_0 \leqslant 8n$, contradicting (1).

Now let j, $0 \leqslant j \leqslant n-1$ be such that $i_{j+1} - i_j \geqslant 9$. Let $i_j = l$. Then $i_{j+1} \geqslant l + 9$. So, the initial symbol of x_{j+1} is a_l and the final symbol of x_{j+1} is a_z for some $z \geqslant l + 9$. Thus,

$$x_{j+1} = a_l^{s_l} a_{l+1}^{s_{l+1}} a_{l+2}^{s_{l+2}} \cdots a_{l+9}^{s_{l+9}} \gamma$$

for some $s_l, ..., s_{l+9}$ and some γ in $\{a_{l+9}, ..., a_{9n}\}^*$. We consider 3 cases.

Case 1. $l + 2$ is in R. Let $r = l + 2$. Then clearly, $s_{l+2} = s_{l+3} = s_{l+4} = k_p$ and $p = [r/3] + 1$. Hence

$$x_{j+1} = \alpha_{k_p} a_{r-1}^{t_{k_p}} a_r^{k_p} a_{r+1}^{k_p} a_{r+2}^{k_p} a_{r+3}^{q_{k_p}} \beta_{k_p}$$

where $\alpha_{k_p} = a_l^{s_l}$, $t_{k_p} = s_{l+1} \geqslant 1$, $q_{k_p} = s_{l+5} \geqslant 1$, and $\beta_{k_p} = a_{l+6}^{s_{l+6}} \cdots a_{l+9}^{s_{l+9}} \gamma$.

Case 2. $l + 2$ is not in R but $l + 3$ is in R. Let $r = l + 3$. Then $s_{l+3} = s_{l+4} = s_{l+5} = k_p$ and $p = [r/3] + 1$. Hence $x_{j+1} = \alpha_{k_p} a_{r-1}^{t_{k_p}} a_r^{k_p} a_{r+1}^{k_p} a_{r+2}^{k_p} a_{r+3}^{q_{k_p}} \beta_{k_p}$, where $\alpha_{k_p} = a_l^{s_l} a_{l+1}^{s_{l+1}}$, $t_{k_p} = s_{l+2} \geqslant 1$, $q_{k_p} = s_{l+6} \geqslant 1$, and $\beta_{k_p} = a_{l+7}^{s_{l+7}} \cdots a_{l+9}^{s_{l+9}} \gamma$.

Case 3. If $l + 2$ and $l + 3$ are not in R, then by the definition of the set R, we must have $l + 4$ in R. Let $r = l + 4$. Then $s_{l+4} = s_{l+5} = s_{l+6} = k_p$ and $p = [r/3] + 1$. Hence $x_{j+1} = \alpha_{k_p} a_{r-1}^{t_{k_p}} a_r^{k_p} a_{r+1}^{k_p} a_{r+2}^{k_p} a_{r+3}^{q_{k_p}} \beta_{k_p}$, where $\alpha_{k_p} = a_l^{s_l} a_{l+1}^{s_{l+1}} a_{l+2}^{s_{l+2}}$, $t_{k_p} = s_{l+3} \geqslant 1$, $q_{k_p} = s_{l+7} \geqslant 1$, and $\beta_{k_p} = a_{l+8}^{s_{l+8}} a_{l+9}^{s_{l+9}} \gamma$.

[6] For any real number d, $[d]$ is the largest integer $\leqslant d$.

Notation. If $X \subseteq [\Sigma^*]^n$, then for each i, $1 \leqslant i \leqslant n$, $P_i(X)$ will denote the set $\{x_i \mid \exists x_1, ..., x_{i-1}, x_{i+1}, ..., x_n$ such that $(x_1, ..., x_n)$ is in $X\}$.

LEMMA 4.2. *Let $n \geqslant 1$ and $a_1, ..., a_{9n}$ be distinct symbols. Let $L_n = \{a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_{9n-2}^{k_{3n}}a_{9n-1}^{k_{3n}}a_{9n}^{k_{3n}} \mid k_i \geqslant 1,\ 1 \leqslant i \leqslant 3n\}$. Let Q_n be any subset of $[\{a_1, ..., a_{9n}\}^*]^n$ such that $L_n = \{x_1 \cdots x_n \mid (x_1, ..., x_n)$ in $Q_n\}$. Then there exist g, $1 \leqslant g \leqslant n$, r in $R = \{3m + 1 \mid 1 \leqslant m \leqslant 3n - 2\}$, $p = [r/3 + 1]$ such that for infinitely many k_p's there exist corresponding α_{k_p} in $\{a_1, ..., a_{r-1}\}^*$, β_{k_p} in $\{a_{r+3}, ..., a_{9n}\}^*$, $t_{k_p} \geqslant 1$, $q_{k_p} \geqslant 1$* such that

$$\alpha_{k_p} a_{r-1}^{t_{k_p}} a_r^{k_p} a_{r+1}^{k_p} a_{r+2}^{k_p} a_{r+3}^{q_{k_p}} \beta_{k_p} \quad \text{is in} \quad P_g(Q_n).$$

Proof. For each $(i_0, ..., i_n)$ such that $0 = i_0 \leqslant \cdots \leqslant i_n = 9n$, let $F_{(i_0,...,i_n)} = \{(x_1, ..., x_n)$ in $Q_n \mid f(x_1, ..., x_n) = (i_0, ..., i_n)\}$. Since there are only a finite number of distinct n-tuples $(i_0, ..., i_n)$ satisfying $0 = i_0 \leqslant \cdots \leqslant i_n = 9n$, there exists $(i_0', ..., i_n')$ such that $F_{(i_0',...,i_n')}$ is infinite. Then by Lemma 4.1, there exists j, $0 \leqslant j \leqslant n - 1$, such that $i_{j+1}' - i_j' \geqslant 9$ and each x_{j+1} in $(x_1, ..., x_n)$ such that $(x_1, ..., x_n)$ is in $F_{(i_0',...,i_n')}$ has the form:

$$x_{j+1} = \alpha_{k_p} a_{r-1}^{t_{k_p}} a_r^{k_p} a_{r+1}^{k_p} a_{r+2}^{k_p} a_{r+3}^{q_{k_p}} \beta_{k_p}, \tag{$*$}$$

where r is in $R = \{3m + 1 \mid 1 \leqslant m \leqslant 3n - 2\}$, $p = [r/3] + 1$, $t_{k_p} \geqslant 1$, $q_{k_p} \geqslant 1$, α_{k_p} in $\{a_1, ..., a_{r-1}\}^*$, and β_{k_p} in $\{a_{r+3}, ..., a_{9n}\}^*$. Since there are only three possible values of r corresponding to the three cases in the proof of Lemma 4.1, and since $F_{(i_0',...,i_n')}$ is infinite, a fixed r can be chosen for which $(*)$ is true for infinitely many k_p's. Now let $g = j + 1$. Then $\{x_{j+1} \mid \exists x_1, ..., x_j, x_{j+2}, ..., x_n$ such that $(x_1, ..., x_n)$ is in $F_{(i_0',...,i_n')}\} \subseteq P_g(Q_n)$ satisfies the requirements of the lemma.

LEMMA 4.3. *Let $L \subseteq [\Sigma^*]^n$ be an n-CFL [n-RLCFL]. Then for each i $1 \leqslant i \leqslant n$, $P_i(X)$ is a CFL [regular set].*

Proof. Let $G = \langle V, P, S, \Sigma_n \rangle$ be a CFG [RLCFG] such that $L = \tau_n(L(G))$. We may assume that the rules in P are of the form $A \to x$ or $A \to BC$, where x is in Σ_n^* and A, B, C are in V. For each i, $1 \leqslant i \leqslant n$, let ϕ_i be a homomorphism from Σ_n^* into Σ^* which maps $[a, i]$ into a and $[a, j]$ into Λ for $j \neq i$. For each i, $1 \leqslant i \leqslant n$, let $G^i = \langle V, P^i, S, \Sigma \rangle$ be a CFG, where $P^i = \{A \to BC \mid A, B, C$ in V, $A \to BC$ in $P\} \cup \{A \to \phi_i(x) \mid A$ in V, $A \to x$ in $P\}$. Clearly, $P_i(L) = L(G^i)$ for $1 \leqslant i \leqslant n$.

We will need the following lemma which was proved in [5].

LEMMA 4.4. *Let b_1, b_2 be distinct symbols. If M is an infinite subset of $\{b_1{}^n b_2{}^n b_1{}^n \mid n \geqslant 1\}$, then M is not a context-free language.*

LEMMA 4.5. *Let $n \geqslant 1$ and $a_1, ..., a_{9n}$ be distinct symbols. Then the set $L_n = \{a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_{9n-2}^{k_{3n}}a_{9n-1}^{k_{3n}}a_{9n}^{k_{3n}} \mid k_i \geqslant 1, 1 \leqslant i \leqslant 3n\}$ is not in $\mathscr{S}(n)$.*

Proof. Assume that L_n is in $\mathscr{S}(n)$. Then by Theorem 1.1, there is an n-CFL Q_n such that $L_n = \{x_1 \cdots x_n \mid (x_1, ..., x_n) \text{ in } Q_n\}$. Then by Lemma 4.3, for each i, $1 \leqslant i \leqslant n$, $P_i(Q_n)$ is a context-free language. Now by Lemma 4.2, there exist g, $1 \leqslant g \leqslant n$, r in $R = \{3m + 1 \mid 1 \leqslant m \leqslant 3n - 2\}$, $p = [r/3] + 1$ such that for infinitely many k_p's there exist corresponding α_{k_p} in $\{a_1, ..., a_{r-1}\}^*$, β_{k_p} in $\{a_{r+3}, ..., a_{9n}\}^*$, $t_{k_p} \geqslant 1$, $q_{k_p} \geqslant 1$ such that $\alpha_{k_p}a_{r-1}^{t_{k_p}}a_r^{k_p}a_{r+1}^{k_p}a_{r+2}^{k_p}a_{r+3}^{q_{k_p}}\beta_{k_p}$ is in $P_g(Q_n)$. Let $T = \{a_1^{i_1} \cdots$

$$a_{r-2}^{i_{r-2}}a_{r-1}^{i_{r-1}}a_r^{i_r}a_{r+1}^{i_{r+1}}a_{r+2}^{i_{r+2}}a_{r+3}^{i_{r+3}}a_{r+4}^{i_{r+4}} \cdots a_{9n}^{i_{9n}} \mid i_j \geqslant 0 \qquad \text{for} \quad 1 \leqslant j \leqslant r-2$$

and $r + 4 \leqslant j \leqslant 9n$, and $i_k \geqslant 1$ for $r - 1 \leqslant k \leqslant r + 3\}$. Clearly, T is a regular set. Since $P_g(Q_n)$ is a context-free language and context-free languages are closed under intersection with regular sets, $P_g(Q_n) \cap T$ is a context free language. Furthermore, $P_g(Q_n) \cap T$ is infinite and all of its elements are of the form

$$\alpha_{k_p}a_{r-1}^{t_{k_p}}a_r^{k_p}a_{r+1}^{k_p}a_{r+2}^{k_p}a_{r+3}^{q_{k_p}}\beta_{k_p}.$$

Let ϕ be a homomorphism which maps a_r into b_1, a_{r+1} into b_2, a_{r+2} into b_1, and all other a_j ($j \neq r, j \neq r + 1, j \neq r + 2$) into Λ. Then $\phi(P_g(Q_n) \cap T)$ is a context-free language since context-free languages are closed under homomorphism. Moreover, $\phi(P_g(Q_n) \cap T)$ is an infinite subset of $\{b_1{}^n b_2{}^n b_1{}^n \mid n \geqslant 1\}$. This is a contradiction to Lemma 4.4.

COROLLARY 4.2. *Let $n \geqslant 1$ and a_1, a_2, a_3 be distinct symbols. Then the set $L_n = \{a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_1^{k_{3n}}a_2^{k_{3n}}a_3^{k_{3n}} \mid k_i \geqslant 1, 1 \leqslant i \leqslant 3n\}$ is not in $\mathscr{S}(n)$.*

Proof. This follows from the preceding lemma and the fact that $\mathscr{S}(n)$ is closed under nondeterministic gsm mappings (Theorem 3.6).

Remark. In the case of $\mathscr{R}(n)$, a result similar to Corollary 4.2 can be derived using an approach very similar to the one we have for $\mathscr{S}(n)$, utilizing the fact that if M is an infinite subset of $\{b_1{}^n b_2{}^n \mid n \geqslant 1\}$, then M is a nonregular set. We leave it up to the reader to provide the necessary modifications to Lemmas 4.1–4.5 to prove the following result.

COROLLARY 4.3. *Let* $n \geqslant 1$ *and* a_1, a_2 *be distinct symbols. Then the set* $L_n = \{a_1^{k_1}a_2^{k_1} \cdots a_1^{k_{3n}}a_2^{k_{3n}} \mid k_i \geqslant 1, 1 \leqslant i \leqslant 3n\}$ *is not in* $\mathscr{R}(n)$.

COROLLARY 4.4. *For each* $n \geqslant 1$, $\mathscr{S}(n)$ *is not closed under substitution by languages in* $\mathscr{S}(2)$.

Proof. Let $n \geqslant 1$ and a_1, a_2, a_3 be distinct symbols. Let $L_1 = \{a_1^{3n}\}$ and $L_2 = \{a_1{}^k a_2{}^k a_3{}^k \mid k \geqslant 1\}$. Clearly, L_1 is in $\mathscr{S}(n)$ and L_2 is in $\mathscr{S}(2)$. Define a substitution ϕ which maps a_1 into L_2. Then $\phi(L_1) = \{a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_1^{k_{3n}}a_2^{k_{3n}}a_3^{k_{3n}} \mid k_i \geqslant 1, 1 \leqslant i \leqslant 3n\}$ is not in $\mathscr{S}(n)$ by Corollary 4.2.

THEOREM 4.3. *For each* $n \geqslant 1$, $\mathscr{S}(n)$ *is not closed under concatenation with languages in* $\mathscr{S}(2)$.

Proof. Let $L_1 = \{\Lambda\}$ and L_2 be the language defined in the proof of Corollary 4.4. Then L_1 is in $\mathscr{S}(n)$ and L_2 is in $\mathscr{S}(2)$. If $\mathscr{S}(n)$ were closed under concatenation with languages in $\mathscr{S}(2)$, then $L_1L_2^{3n} = \{a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_1^{k_{3n}}a_2^{k_{3n}}a_3^{k_{3n}} \mid k_i \geqslant 1, 1 \leqslant i \leqslant 3n\}$ would be in $\mathscr{S}(n)$, contradicting Corollary 4.2.

In the case of $\mathscr{R}(n)$, we have the following result which is easily verified using Corollary 4.3.

THEOREM 4.4. *For each* $n \geqslant 1$, $\mathscr{R}(n)$ *is not closed under the following operations*:

(1) substitution by languages in $\mathscr{R}(2)$,

(2) concatenation with languages in $\mathscr{R}(2)$.

COROLLARY 4.5. *For each* $n \geqslant 2$, $\mathscr{S}(n)[\mathscr{R}(n)]$ *is not closed under concatenation.*

It is interesting to find out whether $\mathscr{S}[\mathscr{R}]$ is closed under concatenation. The next theorem gives a positive answer.

THEOREM 4.5. *For* $n, m \geqslant 2$, *if* L_1 *is in* $\mathscr{S}(n)[\mathscr{R}(n)]$ *and* L_2 *is in* $\mathscr{S}(m)[\mathscr{R}(m)]$, *then* L_1L_2 *is in* $\mathscr{S}(n+m)[\mathscr{R}(n+m)]$.

Proof. Since $\mathscr{S}(n+m)[\mathscr{R}(n+m)]$ is closed under homomorphism, we may assume that L_1 and L_2 are over disjoint alphabets, that is, $L_1 \subseteq \Sigma^*$ and $L_2 \subseteq \Delta^*$, $\Sigma \cap \Delta = \varnothing$. By Theorem 1.1 [Theorem 1.2], there are context-free languages [regular sets] L_3 and L_4 over Σ_n and Δ_m, respectively, such that $L_1 = \{x_1 \cdots x_n \mid (x_1,..., x_n) \text{ in } \tau_n(L_3)\}$ and $L_2 = \{y_1 \cdots y_m \mid (y_1,..., y_m) \text{ in } \tau_m(L_4)\}$. Let $\Gamma = \Sigma \cup \Delta$ and ϕ be a homomorphism from $(\Sigma_n \cup \Delta_m)^*$ into

Γ^*_{n+m} which maps each $[a,i]$ in Σ_n into $[a,i]$ and each $[b,j]$ in Δ_m into $[b, n+j]$. Then $\phi(L_3L_4)$ is a context-free language [regular set] over Γ_{n+m} and clearly, $L_1L_2 = \{x_1 \cdots x_n x_{n+1} \cdots x_{n+m} \mid (x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}) \text{ in } \tau_{n+m}(\phi(L_3L_4))\}$. The theorem now follows from Theorem 1.1 [Theorem 1.2].

Let us now turn our attention to the operation of closure.

THEOREM 4.6. *There exists a language L in $\mathscr{S}(2)$ such that L^* is not in $\mathscr{S}$.*

Proof. Let a_1, a_2, a_3 be distinct symbols and $L = \{a_1{}^k a_2{}^k a_3{}^k \mid k \geqslant 1\}$. Then L is in $\mathscr{S}(2)$. Suppose that L^* is in $\mathscr{S}$. Then there exists an $n \geqslant 1$ such that L is in $\mathscr{S}(n)$. Let $T = \{a_1^{i_1}a_2^{i_2}a_3^{i_3} \cdots a_1^{i_{9n-2}}a_2^{i_{9n-1}}a_3^{i_{9n}} \mid i_j \geqslant 1,\ 1 \leqslant j \leqslant 9n\}$. Clearly, T is a regular set. By Theorem 3.4, $L^* \cap T$ is in $\mathscr{S}(n)$. However, $L^* \cap T = \{a_1^{k_1}a_2^{k_1}a_3^{k_1} \cdots a_1^{k_{3n}}a_2^{k_{3n}}a_3^{k_{3n}} \mid k_i \geqslant 1, 1 \leqslant i \leqslant 3n\}$ is not in $\mathscr{S}(n)$ by Corollary 4.2, a contradiction.

COROLLARY 4.6. *For each $n \geqslant 2$, $\mathscr{S}(n)$ is not closed under the star operator.*

We now show that there are context-free languages which are not in $\mathscr{R}$.

THEOREM 4.7. *There are context-free languages which are not in $\mathscr{R}$.*

Proof. Let a_1, a_2 be distinct symbols and $L = \{a_1{}^k a_2{}^k \mid k \geqslant 1\}$. Clearly, L and L^* are context-free languages. We claim that L^* is not in $\mathscr{R}$. Assume for contradiction that it is. Then L^* is in $\mathscr{R}(n)$ for some $n \geqslant 1$. Let $T = \{a_1^{i_1}a_2^{i_2}a_1^{i_3}a_2^{i_4} \cdots a_1^{i_{6n-1}}a_2^{i_{6n}} \mid i_j \geqslant 1,\ i \leqslant j \leqslant 6n\}$. Since T is a regular set, $L^* \cap T = \{a_1^{k_1}a_2^{k_1} \cdots a_1^{k_{3n}}a_2^{k_{3n}} \mid k_i \geqslant 1,\ 1 \leqslant i \leqslant 3n\}$ would be in $\mathscr{R}(n)$ (by Theorem 3.4), contradicting Corollary 4.3.

COROLLARY 4.7. *There exists a language L in $\mathscr{R}(2)$ such that L^* is not in $\mathscr{R}$. Thus, for each $n \geqslant 2$, $\mathscr{R}(n)$ is not closed under the star operator.*

Proof. The language L defined in the proof of Theorem 4.7 is certainly in $\mathscr{R}(2)$.

5. BOUNDED SIMPLE MATRIX LANGUAGES AND SEMILINEAR SETS

In this section, we shall present a result which connects simple matrix languages with semilinear sets. This generalizes the result for context-free languages proved by Parikh [12]. The concept of bounded languages first introduced in [7] is extended to include bounded simple matrix languages. We then show that the family of bounded simple matrix languages coincides with the family of bounded right-linear simple matrix languages. A characterization of the family of bounded simple matrix languages as the smallest family of languages containing the bounded context free languages closed under the operations of union and intersection is proved.

Notation. Let N denote the nonnegative integers. For each $k \geqslant 1$, let $N^k = N \times \cdots \times N$ (k times). We shall regard N^k as a subset of the vector space R^k of all n-tuples of rational numbers over the rational numbers. Thus for elements $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$ in N^k and c in N, $x + y = (x_1 + y_1, \ldots, x_k + y_k)$, $x - y = (x_1 - y_1, \ldots, x_k - y_k)$ and $cx = (cx_1, \ldots, cx_k)$.

DEFINITION. A subset Q of N^k is said to be a *linear set* if there exist c, $p_1, \ldots, p_m$ in N^k such that $Q = \{x \mid x = c + n_1 p_1 + \cdots + n_m p_m$, each n_i in $N\}$. c is called the *constant* and $p_1, \ldots, p_m$ are called the *periods* of Q. We write $Q = Q(c; p_1, \ldots, p_m)$ if Q is a linear set with constant c and periods $p_1, \ldots, p_m$. Q is said to be a *semilinear set* if Q is a finite union of linear sets.

The following lemma was proved in [7].

LEMMA 5.1. *Let r and m be positive integers, and f be a linear function of N^r into N^m. If Q is a linear (semilinear) subset of N^r, then $f(N^r)$ is a linear (semilinear) subset of N^m.*

COROLLARY 5.1. *Let $k \geqslant 1$, $n \geqslant 1$. If $Q \subseteq N^{kn}$ is a semilinear set, then the set $\bar{Q} = \{(\alpha_1, \ldots, \alpha_k) \mid \exists x_1, \ldots, x_{kn}$ such that $\alpha_i = x_{(i-1)n+1} + x_{(i-1)n+2} + \cdots + x_{in}$ for $1 \leqslant i \leqslant k$ and $(x_1, \ldots, x_{kn})$ in $Q\}$ is a semilinear subset of N^k.*

Notation. Let $\Sigma = \{a_1, \ldots, a_k\}$. The mapping $\psi_{\langle a_1, \ldots, a_k \rangle}$ or $\psi_\alpha(\alpha = \langle a_1, \ldots, a_k \rangle)$ is the function from Σ^* into N^k defined by $\psi_\alpha(x) = (\#_{a_1}(x), \ldots, \#_{a_k}(x))$, where $\#_{a_i}(x)$ is the number of occurrences of a_i in x. Thus $\psi_\alpha(\Lambda) = (0, \ldots, 0)$ and $\psi_\alpha(x_1 \cdots x_m) = \sum_{i=1}^m \psi_\alpha(x_i)$ for each x_i in Σ^*. If $L \subseteq \Sigma^*$, $\psi_\alpha(L) = \bigcup_{x \text{ in } L} \psi_\alpha(x)$ is sometimes called the *Parikh map* of L.

The following lemma is due to Parikh [12].

LEMMA 5.2. *If $L \subseteq \{a_1, \ldots, a_k\}^*$ is a context-free language, then $\psi_\alpha(L)$ ($\alpha = \langle a_1, \ldots, a_k \rangle$) is a semilinear set.*

We now extend this result to the family of simple matrix languages.

THEOREM 5.1. *If $L \subseteq \{a_1, \ldots, a_k\}^*$ is in $\mathscr{S}(n)$, then $\psi_\alpha(L)$ ($\alpha = \langle a_1, \ldots, a_k \rangle$) is a semilinear set.*

Proof. By Theorem 1.1, there exists a context-free language

$$L' \subseteq \{[a_1, 1], [a_1, 2], ..., [a_1, n], ..., [a_k, 1], [a_k, 2], ..., [a_k, n]\}^*$$

such that $L = \{x_1 \cdots x_n \mid (x_1, ..., x_n) \text{ in } \tau_n(L')\}$. Then by Lemma 5.2, $\psi_{\alpha'}(L')(\alpha' = \langle [a_1, 1], [a_1, 2], ..., [a_1, n], ..., [a_k, 1], [a_k, 2], ..., [a_k, n] \rangle)$ is a semilinear set. Thus, there exists a semilinear set $Q \subseteq N^{kn}$ such that $\psi_{\alpha'}(L') = Q$. We claim that $\psi_\alpha(L) = \bar{Q} = \{(\alpha_1, ..., \alpha_k) \mid \exists x_1, ..., x_{kn}$ such that $\alpha_i = x_{(i-1)n+1} + \cdots + x_{in}$ for $1 \leqslant i \leqslant k$ and $(x_1, ..., x_{kn})$ in $Q\}$. Suppose w is in L. Then there exists z in L' such that $\tau_n(z) = (w_1, ..., w_n)$ and $w = w_1 \cdots w_n$. Now $\psi_\alpha(w) = (\#_{a_1}(w), ..., \#_{a_k}(w))$ and $\psi_{\alpha'}(z) = (\#_{[a_1,1]}(z), \#_{[a_1,2]}(z), ..., \#_{[a_1,n]}(z), ..., \#_{[a_k,1]}(z), \#_{[a_k,2]}(z), ..., \#_{[a_k,n]}(z))$. Since $\tau_n(z) = (w_1, ..., w_n)$ and $w = w_1 \cdots w_n$, obviously, $\#_{a_i}(w) = \sum_{j=1}^{n} \#_{[a_i,j]}(z)$ for each i, $1 \leqslant i \leqslant k$. Therefore, $\psi_\alpha(w)$ is in $\bar{Q}$, and $\psi_\alpha(L) \subseteq \bar{Q}$. Reversing the argument would show that $\bar{Q} \subseteq \psi_\alpha(L)$. Thus, $\bar{Q} = \psi_\alpha(L)$. By Corollary 5.1, $\bar{Q} = \psi_\alpha(L)$ is a semilinear set, completing the proof.

There is a converse to Theorem 5.1 which can be stated as follows.

THEOREM 5.2. *Let $\Sigma = \{a_1, ..., a_k\}$. If $L \subseteq a_1^* a_2^* \cdots a_k^*$*[7] *and $\psi_\alpha(L) = Q(\alpha = \langle a_1, ..., a_k \rangle)$ is a semilinear set, then L is in $\mathscr{R}(k)$.*

Proof. Since $\mathscr{R}(k)$ is closed under union (Theorem 3.2), we may assume that Q is a linear set. So let $Q = Q(c; p_1, ..., p_m)$, where $c = (c_1, ..., c_k)$ and $p_i = (p_{i1}, ..., p_{ik})$ for $1 \leqslant i \leqslant m$. Let $A_1, ..., A_k$, S be distinct symbols not in Σ. Construct a k-RLSMG $G_k = \langle \{A_1\}, \{A_2\}, ..., \{A_k\}, P, S, \Sigma \rangle$, where $P = \{[S \to a_1^{c_1} A_1 a_2^{c_2} A_2 \cdots a_k^{c_k} A_k]\} \cup \{[A_1 \to a_1^{p_{i1}} A_1, ..., A_k \to a_k^{p_{ik}} A_k] \mid 1 \leqslant i \leqslant m\} \cup \{[A_1 \to \Lambda, ..., A_k \to \Lambda]\}$. Clearly, $\psi_\alpha(L(G_k)) = Q$ and $L = L(G_k)$.

In the remainder of this section, we shall be concerned with bounded languages. The following definition is taken from [7].

DEFINITION. A subset L of Σ^* is said to be *bounded* if there exist words $w_1, ..., w_k$ in Σ^* such that $L \subseteq w_1^* \cdots w_k^*$.

Notation. $\mathscr{S}_B(n)$ will denote the bounded languages in $\mathscr{S}(n)$ and $\mathscr{S}_B = \bigcup_{n \geqslant 1} \mathscr{S}_B(n)$. Similarly, $\mathscr{R}_B(n)$ will denote the bounded languages in $\mathscr{R}(n)$ and $\mathscr{R}_B = \bigcup_{n \geqslant 1} \mathscr{R}_B(n)$. Note that $\mathscr{S}_B(1)$ coincides with the class of bounded context-free languages studied in [7].

[7] For words $w_1, ..., w_k$ in Σ^*, we write $w_1^* \cdots w_k^*$ to denote the set

$$\{w_1^{i_1} \cdots w_k^{i_k} \mid i_j \geqslant 0, 1 \leqslant j \leqslant k\}.$$

LEMMA 5.3. *Let* $w_1, \ldots, w_k$ *be words in* Σ^* *and* $a_1, \ldots, a_k$ *be distinct symbols. Let* $n \geqslant 1$. *If* $L \subseteq w_1^* \cdots w_k^*$ *is in* $\mathscr{S}_B(n)[\mathscr{R}_B(n)]$, *then the set* $\{a_1^{i_1} \cdots a_k^{i_k} \mid w_1^{i_1} \cdots w_k^{i_k} \text{ in } L\}$ *is in* $\mathscr{S}_B(n)[\mathscr{R}_B(n)]$.

Proof. Since L is in $\mathscr{S}_B(n)[\mathscr{R}_B(n)]$, L must be in $\mathscr{S}(n)[\mathscr{R}(n)]$. Let ϕ be a homomorphism which maps each a_i into w_i $(1 \leqslant i \leqslant n)$. Then $\phi^{-1}(L)$ is in $\mathscr{S}(n)[\mathscr{R}(n)]$ by Corollary 3.4. Let $L' = \phi^{-1}(L) \cap a_1^* \cdots a_k^*$. Then $L' = \{a_1^{i_1} \cdots a_k^{i_k} \mid w_1^{i_1} \cdots w_k^{i_k} \text{ in } L\}$. Since $a_1^* \cdots a_k^*$ is regular, L' is in $\mathscr{S}(n)[\mathscr{R}(n)]$ by Theorem 3.4. Since L' is bounded, L' is in $\mathscr{S}_B(n)[\mathscr{R}_B(n)]$.

COROLLARY 5.2. *Let* $w_1, \ldots, w_k$ *be words in* Σ^* *and* $a_1, \ldots, a_k$ *be distinct symbols. If* $L \subseteq w_1^* \cdots w_k^*$ *is in* $\mathscr{S}_B[\mathscr{R}_B]$, *then the set* $\{a_1^{i_1} \cdots a_k^{i_k} \mid w_1^{i_1} \cdots w_k^{i_k} \text{ in } L\}$ *is in* $\mathscr{S}_B[\mathscr{R}_B]$.

THEOREM 5.3. *Let* $w_1, \ldots, w_k$ *be words in* Σ^*. *A subset* $L \subseteq w_1^* \cdots w_k^*$ *is in* $\mathscr{R}_B$ *if and only if the set* $\{(i_1, \ldots, i_k) \mid w_1^{i_1} \cdots w_k^{i_k} \text{ in } L\} = Q$ *is semilinear.*

Proof. Let $L \subseteq w_1^* \cdots w_k^*$ be in $\mathscr{R}_B$. Let $a_1, \ldots, a_k$ be distinct symbols. Then the set $L' = \{a_1^{i_1} \cdots a_k^{i_k} \mid w_1^{i_1} \cdots w_k^{i_k} \text{ in } L\}$ is in $\mathscr{R}_B$ by Corollary 5.2. Obviously, L' is in $\mathscr{S}$ and by Theorem 5.1, the set Q is semilinear.

Now let Q be a semilinear set. Then by Theorem 5.2, $L'' = \{a_1^{i_1} \cdots a_k^{i_k} \mid (i_1, \ldots, i_k) \text{ in } Q\}$ is in $\mathscr{R}$. Let ϕ be a homomorphism which maps each a_i into w_i $(1 \leqslant i \leqslant k)$. Then $\phi(L'')$ is in $\mathscr{R}$ since $\mathscr{R}$ is closed under homomorphism. Since $\phi(L'') = L$, L is in $\mathscr{R}_B$.

COROLLARY 5.3. $\mathscr{S}_B = \mathscr{R}_B$.

Proof. It suffices to show that $\mathscr{S}_B \subseteq \mathscr{R}_B$. So let $L \subseteq w_1^* \cdots w_k^*$ be in $\mathscr{S}_B$. Let $a_1, \ldots, a_k$ be distinct symbols. Then the set $L' = \{a_1^{i_1} \cdots a_k^{i_k} \mid w_1^{i_1} \cdots w_k^{i_k} \text{ in } L\}$ is in $\mathscr{S}_B$ (hence in $\mathscr{S}$) by Corollary 5.2. By Theorem 5.1, the set $Q = \{(i_1, \ldots, i_k) \mid a_1^{i_1} \cdots a_k^{i_k} \text{ in } L'\}$ is a semilinear set. Then by Theorem 5.3, L is in $\mathscr{R}_B$.

We now discuss briefly the closure properties of $\mathscr{R}_B$. Since bounded sets are closed under union, under concatenation, and under transposition [7] and since $\mathscr{R}$ contains $\mathscr{R}_B$ and $\mathscr{R}$ is closed under these operations, we have the following result.

THEOREM 5.4. $\mathscr{R}_B$ *is closed under the operations of union, concatenation, and transposition.*

THEOREM 5.5. *If* L_1 *is in* $\mathscr{R}_B$ *and* L_2 *is in* $\mathscr{S}$, *then* $L_1 \cap L_2$ *is in* $\mathscr{R}_B$.

Proof. Let $w_1, \ldots, w_k$ be words in Σ^* and $L_1 \subseteq w_1^* \cdots w_k^*$. Let $L_3 = L_2 \cap w_1^* \cdots w_k^*$. Then L_3 is in $\mathscr{S}$ since $w_1^* \cdots w_k^*$ is regular. Since L_3 is bounded, L_3 is in $\mathscr{R}_B$ by Corollary 5.3. We have $L_1 \cap L_2 = L_1 \cap L_2 \cap w_1^* \cdots w_k^* = L_1 \cap L_3$. By Theorem 5.3, the sets $Q_1 = \{(i_1, \ldots, i_k) \mid w_1^{i_1} \cdots w_k^{i_k} \text{ in } L_1\}$ and $Q_3 = \{(j_1, \ldots, j_k) \mid w_1^{j_1} \cdots w_k^{j_k} \text{ in } L_3\}$ are semilinear. Since the intersection of semilinear subsets of N^k is semilinear [7], $Q_1 \cap Q_3$ is semilinear. The theorem follows from Theorem 5.3.

COROLLARY 5.4. *$\mathscr{R}_B$ is closed under intersection.*

The following theorem is easily verified using the fact that if Q_1 and Q_2 are semilinear subsets of N^k, then $Q_1 - Q_2$ is semilinear [7].

THEOREM 5.6. *If L_1 is in $\mathscr{R}_B$ and L_2 is in $\mathscr{S}$, then $L_2 - L_1$ is in $\mathscr{S}$ and $L_1 - L_2$ is in $\mathscr{R}_B$.*

COROLLARY 5.5. *If L_1 and L_2 are in $\mathscr{R}_B$, then $L_1 - L_2$ is in $\mathscr{R}_B$.*

We shall show that $\mathscr{S}_B$ is just the closure of $\mathscr{S}_B(1)$ (= class of bounded context free languages) under the operations of union and intersection.

LEMMA 5.4. *Let $L \subseteq \Sigma^*$ be in $\mathscr{R}_B$. Then L is a finite union of sets of the form*

(1) $\{x_{10}x_{11}^{k_1} \cdots x_{1m}^{k_m}x_{20}x_{21}^{k_1} \cdots x_{2m}^{k_m} \cdots x_{n0}x_{n1}^{k_1} \cdots x_{nm}^{k_m} \mid k_i \geqslant 0,\ 1 \leqslant i \leqslant m\}$,

where $n, m \geqslant 1$ and the x_{ij}'s are words in Σ^* $(1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m)$. Conversely, each finite union of sets of the form (1) is in $\mathscr{R}_B$.

Proof. Let L be in $\mathscr{R}_B$. Then there are words $w_1, \ldots, w_n$ in Σ^* such that $L \subseteq w_1^* \cdots w_n^*$. Then by Theorem 5.3, L is a finite union of sets of the form $A = \{w_1^{i_1} \cdots w_n^{i_n} \mid (i_1, \ldots, i_n) \text{ in } Q\}$, where Q is a linear set. Let $Q = Q(c; p_1, \ldots, p_m)$ where $c = (c_1, \ldots, c_n)$ and $p_i = (p_{i1}, \ldots, p_{in})$ for $1 \leqslant i \leqslant m$. Let $a_1, \ldots, a_n$ be distinct symbols. From the construction in the proof of Theorem 5.2, it is easily seen that the set $B = \{a_1^{i_1} \cdots a_n^{i_n} \mid (i_1, \ldots, i_n) \text{ in } Q\}$ is equal to the set $E = \{a_1^{c_1}(a_1^{p_{11}})^{k_1} \cdots (a_1^{p_{m1}})^{k_m} a_2^{c_2}(a_2^{p_{12}})^{k_1} \cdots (a_2^{p_{m2}})^{k_m} \cdots a_n^{c_n}(a_n^{p_{1n}})^{k_1} \cdots (a_n^{p_{mn}})^{k_m} \mid k_i \geqslant 0, 1 \leqslant i \leqslant m\}$. Let ϕ be a homomorphism which maps each a_i into w_i $(1 \leqslant i \leqslant n)$. Then $\phi(E) = A$ and $\phi(E)$ is of the form (1).

To show the converse, it suffices to show that any set of the form (1) is in $\mathscr{R}_B$ (because $\mathscr{R}_B$ is closed under union). Let S, A_{ij} $(1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m)$ be distinct symbols. Let $G_n = \langle V_1, \ldots, V_n, P, S, \Sigma\rangle$, where $V_i = \{A_{ij} \mid 1 \leqslant j \leqslant m\}$ $(1 \leqslant i \leqslant n)$ and $P = \{[S \to x_{10}A_{11} \cdots A_{1m}x_{20}A_{21} \cdots A_{2m} \cdots x_{n0}A_{n1} \cdots A_{nm}]\} \cup \{[A_{1j} \to x_{1j}A_{1j}, \ldots, A_{nj} \to x_{nj}A_{nj}], [A_{1j} \to \Lambda, \ldots, A_{nj} \to \Lambda] \mid 1 \leqslant j \leqslant m\}$. Clearly, G_n is an n-RLSMG, and $L(G_n)$ is exactly the set (1).

LEMMA 5.5. *Let* $n, m \geqslant 1$ *and* x_{ij} in Σ^* *for* $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$. *Then the set*

$$L = \{x_{10}x_{11}^{k_1} \cdots x_{1m}^{k_m} x_{20} x_{21}^{k_1} \cdots x_{2m}^{k_m} \cdots x_{n0} x_{n1}^{k_1} \cdots x_{nm}^{k_m} \mid k_i \geqslant 0, 1 \leqslant i \leqslant m\}$$

is a finite intersection of languages in $\mathscr{S}_B(1)$.

Proof. For positive integers a, b, c, d such that $1 \leqslant a, c \leqslant n$, $1 \leqslant b, d \leqslant m$, let $L(a, b, c, d) = \{x_{10}x_{11}^{k_{11}} \cdots x_{1m}^{k_{1m}} x_{20} x_{21}^{k_{21}} \cdots x_{2m}^{k_{2m}} \cdots x_{n0} x_{n1}^{k_{n1}} \cdots x_{nm}^{k_{nm}} \mid k_{ij} \geqslant 0$ for $1 \leqslant i \leqslant n$, $i \leqslant j \leqslant d$ and $k_{ab} = k_{cd}\}$. Clearly, L is a finite intersection of suitable $L(a, b, c, d)$'s. Thus, it suffices to show that $L(a, b, c, d)$ is in $\mathscr{S}_B(1)$. Let e_1, e_2, e_3, e_4, e_5 be distinct symbols. The set $L' = \{e_1 e_2{}^k e_3 e_4{}^k e_5 \mid k \geqslant 0\}$ is clearly in $\mathscr{S}(1)$. Let ϕ be a substitution defined by $\phi(e_1) = x_{10}x_{11}^* \cdots x_{a(b-1)}^*$, $\phi(e_2) = x_{ab}$, $\phi(e_3) = x_{a(b+1)}^* \cdots x_{c(d-1)}^*$, $\phi(e_4) = x_{cd}$, and $\phi(e_5) = x_{c(d+1)}^* \cdots x_{nm}^*$. By Theorem 3.1, $\phi(L')$ is in $\mathscr{S}(1)$. Moreover, since $\phi(L') = L(a, b, c, d)$, $L(a, b, c, d)$ is in $\mathscr{S}(1)$. Now since $L(a, b, c, d)$ is bounded, $L(a, b, c, d)$ is in $\mathscr{S}_B(1)$.

From Lemmas 5.4 and 5.5, we have:

THEOREM 5.7. *Let* $L \subseteq \Sigma^*$ *be in* $\mathscr{R}_B$. *Then* L *is a finite union and intersection of languages in* $\mathscr{S}_B(1)$.

THEOREM 5.8. $\mathscr{R}_B$ $(= \mathscr{S}_B)$ *is the smallest family of languages containing* $\mathscr{S}_B(1)$ *and closed with respect to finite union and finite intersection.*

Proof. Follows from Corollary 5.3, Theorem 5.4, Corollary 5.4, and Theorem 5.7.

6. DECISION QUESTIONS

In this section, we briefly investigate some decision questions associated with simple matrix languages, right-linear simple matrix languages, and bounded languages.

THEOREM 6.1. *For each* $n \geqslant 2$, *it is recursively unsolvable to determine whether an arbitrary simple matrix language in* $\mathscr{S}(n)$ *is in* $\mathscr{S}(k)$ *for some* $1 \leqslant k < n$. (*Thus, it is recursively unsolvable to determine whether an arbitrary language in* $\mathscr{S}(n)$ *is a context-free language.*)

Proof. Let Σ_1 and Σ_2 be two alphabets such that $\Sigma_1 \cap \Sigma_2 = \varnothing$. Let $L_2 \subseteq \Sigma_2{}^*$ be in $\mathscr{S}(n)$ and not in $\mathscr{S}(k)$, $k < n$ (L_2 exists by Theorem 2.2).

For each context-free language $L_1 \subseteq \Sigma_1^*$, let $L(L_1)$ be the set $L_1\Sigma_2^* \cup \Sigma_1^*L_2$. Clearly, $L(L_1)$ is in $\mathscr{S}(n)$ since $\mathscr{S}(n)$ is closed under union, and under concatenation with regular sets. We now prove that $L(L_1)$ is in $\mathscr{S}(k)$ if and only if $L_1 = \Sigma_1^*$. Suppose $L_1 = \Sigma_1^*$, the $L(L_1) = \Sigma_1^*\Sigma_2^*$ is regular and therefore in $\mathscr{S}(k)$. Suppose $L_1 \neq \Sigma_1^*$ and $L(L_1)$ is in $\mathscr{S}(k)$. Let w be in $(\Sigma_1^* - L_1)$. Then $L(L_1) \cap w\Sigma_2^* = wL_2$ is in $\mathscr{S}(k)$ since $\mathscr{S}(k)$ is closed under intersection with regular sets. Moreover, by Theorem 3.8, L_2 is in $\mathscr{S}(k)$, a contradiction. Thus $L(L_1)$ is in $\mathscr{S}(k)$ if and only if $L_1 = \Sigma_1^*$. Since it is unsolvable to determine if an arbitrary context-free language L_1 is Σ_1^* [3], the theorem follows.

The following theorem follows from corresponding results for context-free languages [3].

THEOREM 6.2. *For arbitrary simple matrix languages L_1 and L_2 is $\mathscr{S}$ and a regular set R, it is recursively unsolvable to determine whether*

(1) $L_1 = R$,

(2) $L_1 = \Sigma^*$,

(3) $L_1 \cap L_2$ *is empty, finite, infinite, regular, or in* $\mathscr{S}$,

(4) $L_1 \subseteq L_2$,

(5) $L_1 = L_2$.

We now mention a few unsolvable problems pertaining to $\mathscr{R}$.

THEOREM 6.3. *Let $x = (x_1, \ldots, x_n)$ be an arbitrary n-tuple of nonempty words in $\{a, b\}^*$ and $L(x) = \{ba^{i_1} \cdots ba^{i_k}cx_{i_1} \cdots x_{i_k} \mid k \geqslant 1,\ 1 \leqslant i_j \leqslant n\}$. Then $\{a, b, c\}^* - L(x)$ is in $\mathscr{R}(2)$.*

Proof. Let $R = \{ba^{i_1} \cdots ba^{i_k}cz \mid k \geqslant 1,\ 1 \leqslant i_j \leqslant n,\ z \text{ in } \{a, b\}^*\}$. Clearly, R is regular. Since regular sets are closed under complementation, $\{a, b, c\}^* - R$ is regular. Thus, $\{a, b, c\}^* - R$ is in $\mathscr{R}(2)$. Let $Q = \{ba^{i_1} \cdots ba^{i_k}cz \mid k \geqslant 1,\ 1 \leqslant i_j \leqslant n,\ z \text{ in } \{a, b\}^*,\ z \neq x_{i_1} \cdots x_{i_k}\}$. Then $\{a, b, c\}^* - L(x) = Q \cup (\{a, b, c\}^* - R)$. Since $\mathscr{R}(2)$ is closed under union (Theorem 3.2), we need only show that Q is in $\mathscr{R}(2)$.

Let $G_2 = \langle\{A_1, B_1, C_1, D_1, E_1\}, \{A_2, B_2, C_2, D_2, E_2\}, P, S, \{a, b, c\}\rangle$ be a 2-RLSMG, where P consists of the following rules ($1 \leqslant i \leqslant n$):

(1) $[S \to A_1cA_2]$;

(2) $[S \to E_1cE_2]$;

(3) $[A_1 \to ba^iA_1,\ A_2 \to x_iA_2]$;

(4) $[A_1 \to ba^iB_1, A_2 \to zB_2]$ for each z in $\{a, b\}^*$ such that $lg(z) < lg(x_i)$;

(5) $[B_1 \to ba^i B_1, B_2 \to B_2]$;

(6) $[B_1 \to \Lambda, B_2 \to \Lambda]$;

(7) $[A_1 \to ba^i C_1, A_2 \to zC_2]$ for each z in $\{a, b\}^*$ such that $lg(z) = lg(x_i)$ and $z \neq x_i$;

(8) $[C_1 \to ba^i C_1, C_2 \to C_2]$;

(9) $[C_1 \to D_1, C_2 \to D_2]$;

(10) $[D_1 \to D_1, D_2 \to aD_2]$;

(11) $[D_1 \to D_1, D_2 \to bD_2]$;

(12) $[D_1 \to \Lambda, D_2 \to \Lambda]$;

(13) $[E_1 \to ba^i E_1, E_2 \to x_i E_2]$;

(14) $[E_1 \to D_1, E_2 \to aD_2]$;

(15) $[E_1 \to D_1, E_2 \to bD_2]$.

It is easily verified that $S \overset{*}{\Rightarrow} w$ if and only if w is one of the following forms:

(a) $w = ba^{i_1} \cdots ba^{i_j} ba^{i_{j+1}} \cdots ba^{i_{j+m}} c x_{i_1} \cdots x_{i_j} z$, where $j \geqslant 0$, $m \geqslant 1$, $1 \leqslant i_r \leqslant n$, z in $\{a, b\}^*$ with $lg(z) < lg(x_{i_{j+1}})$.

(b) $w = ba^{i_1} \cdots ba^{i_j}\, ba^{i_{j+1}} \cdots ba^{i_{j+m}} c x_{i_1} \cdots x_{i_j} z\alpha$, where $j \geqslant 0$, $m \geqslant 1$, $1 \leqslant i_r \leqslant n$, z and α in $\{a, b\}^*$ with $lg(z) = lg(x_{i_{j+1}})$ and $z \neq x_{i_{j+1}}$.

(c) $w = ba^{i_1} \cdots ba^{i_k} c x_{i_1} \cdots x_{i_k} \alpha$, where $k \geqslant 1$, $1 \leqslant i_r \leqslant n$, α in $\{a, b\}^*$ with $lg(\alpha) \geqslant 1$.

Clearly, $L(G_2) \subseteq Q$. We now show that $Q \subseteq L(G_2)$. So suppose $ba^{i_1} \cdots ba^{i_k} cz$ is in Q for some $k \geqslant 1$, $1 \leqslant i_r \leqslant n$, z in $\{a, b\}^*$, $z \neq x_{i_1} \cdots x_{i_k}$. If $z = \Lambda$, then by (a) (letting $j = 0$ and $m = k$), we have that $ba^{i_1} \cdots ba^{i_k} c$ is in $L(G_2)$. Suppose $z \neq \Lambda$. Clearly, there exists j $(0 \leqslant j \leqslant k)$ such that $x_{i_1} \cdots x_{i_j}$ is a maximal initial subword[8] of z (i.e., $x_{i_1} \cdots x_{i_j}$ is an initial subword of z, but $x_{i_1} \cdots x_{i_{j+1}}$ is not an initial subword of z). Note that if $j = 0$, we interpret $x_{i_1} \cdots x_{i_j}$ as Λ. We consider two cases.

Case 1. Suppose $j < k$. Let $z = x_{i_1} \cdots x_{i_j} z'$ for some z' in $\{a, b\}^*$. If $lg(z') < lg(x_{i_{j+1}})$, then by (a), $ba^{i_1} \cdots ba^{i_k} cz$ is in $L(G_2)$. If $lg(z') \geqslant lg(x_{i_{j+1}})$, then by the choice of j, $x_{i_{j+1}}$ is not an initial subword of z'. Thus (b) applies and again $ba^{i_1} \cdots ba^{i_k} cz$ is in $L(G_2)$.

[8] If x is in Σ^*, then w in Σ^* is a *subword* of x if there exist u and v in Σ^* such that $x = uwv$. $u(v)$ is an *initial* (*final*) subword of x.

Case 2. Suppose $j = k$. Then $z = x_{i_1} \cdots x_{i_k}\alpha$ for some α in $\{a, b\}^*$. Since $x_{i_1} \cdots x_{i_k} \neq x_{i_1} \cdots x_{i_k}\alpha$, we must have $lg(\alpha) \geqslant 1$. Thus by (c), $ba^{i_1} \cdots ba^{i_k}cz$ is in $L(G_2)$.

We have shown that $L(G_2) = Q$, completing the proof.

THEOREM 6.4. *Let $m \geqslant 2$. It is recursively unsolvable to determine for an arbitrary language $L \subseteq \Sigma^*$ in $\mathscr{R}(m)$ whether $L = \Sigma^*$.*

Proof. It suffices to prove the theorem for $m = 2$. Let $\Sigma = \{a, b, c\}$ and $x = (x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ be two arbitrary n-tuples of nonempty words in $\{a, b\}^*$. Let $L(x) = \{ba^{i_1} \cdots ba^{i_k}cx_{i_1} \cdots x_{i_k} \mid k \geqslant 1,\ 1 \leqslant i_j \leqslant n\}$ and $L(y) = \{ba^{i_1} \cdots ba^{i_k}cy_{i_1} \cdots y_{i_k} \mid k \geqslant 1,\ 1 \leqslant i_j \leqslant n\}$. Clearly, $L(x)$ and $L(y)$ are in $\mathscr{R}(2)$. By Theorem 6.3, $\Sigma^* - L(x)$ and $\Sigma^* - L(y)$ are in $\mathscr{R}(2)$. Let $L = (\Sigma^* - L(x)) \cup (\Sigma^* - L(y))$. By Theorem 3.2, L is in $\mathscr{R}(2)$. If we can decide whether $L = \Sigma^*$, then by DeMorgan's Law, we can decide whether $L(x) \cap L(y) = \varnothing$. The theorem now follows from the unsolvability of the Post Correspondence problem [5].

THEOREM 6.5. *The following problems, with L_1, L_2 varying over $\mathscr{R}$ and Y varying over the class of regular sets, are unsolvable:*

(1) $L = Y$?

(2) $L_1 = L_2$?

(3) $L_1 \subseteq L_2$?

Proof. (1) and (2) are easily seen by letting $L_1 = (\Sigma^* - L(x)) \cup (\Sigma^* - L(y))$ and $Y = L_2 = \Sigma^*$. (3) follows from the unsolvability of the Post Correspondence problem by letting $L_1 = L(x)$ and $L_2 = \Sigma^* - L(y)$.

The following theorem is easily verified using Theorem 6.4 and a technique similar to that used in the proof of Theorem 6.1.

THEOREM 6.6. *For each $n \geqslant 2$, it is recursively unsolvable to determine whether an arbitrary right-linear simple matrix language in $\mathscr{R}(n)$ is a context-free language or in $\mathscr{R}(k)$ for some $1 \leqslant k < n$. Moreover, it is recursively unsolvable to determine whether an arbitrary context-free language is in $\mathscr{R}$.*

The proof of the following theorem is again straightforward and is omitted.

THEOREM 6.7. *For arbitrary languages L_1 and L_2 in $\mathscr{R}$, it is recursively unsolvable to determine whether $L_1 \cap L_2$ is empty, finite, infinite, or in $\mathscr{R}$.*

Finally, we consider decision questions related to bounded simple matrix languages.

LEMMA 6.1. *Let* $L \subseteq \Sigma^*$ *be in* $\mathscr{S}(n)$ *and* $L' \subseteq [\Sigma^*]^n$ *be any n-CFL such that* $L = \{x_1 \cdots x_n \mid (x_1, ..., x_n) \text{ in } L'\}$. *Then L is bounded if and only if for each i,* $1 \leqslant i \leqslant n$, $P_i(L')$[9] *is bounded. If L is bounded, then* $L \subseteq w_{11}^* \cdots w_{1k(1)}^* \cdots w_{n1}^* \cdots w_{nk(n)}^*$, *where for each i,* $1 \leqslant i \leqslant n$, $w_{i1}, ..., w_{1k(i)}$ *are words in* Σ^* *such that* $P_i(L') \subseteq w_{i1}^* \cdots w_{ik(i)}^*$.

Proof. For each i, $1 \leqslant i \leqslant n$, let $P_i(L')$ be bounded. Then there are words $w_{i1}, ..., w_{ik(i)}$ in Σ^* such that $P_i(L') \subseteq w_{i1}^* \cdots w_{ik(i)}^*$. Now let x be in L. Then by Theorem 1.1, there exist $x_1, ..., x_n$ in Σ^*, $x = x_1 \cdots x_n$, and $(x_1, ..., x_n)$ in L'. Then x_i is in $P_i(L')$ for each i, $1 \leqslant i \leqslant n$ (by the definition of $P_i(L')$). Hence, x is in $w_{11}^* \cdots w_{1k(1)}^* \cdots w_{n1}^* \cdots w_{nk(n)}^*$. We conclude that L is bounded and $L \subseteq w_{11}^* \cdots w_{1k(1)}^* \cdots w_{n1}^* \cdots w_{nk(n)}^*$.

Now suppose L is bounded. Clearly, for each i, $1 \leqslant i \leqslant n$, $P_i(L')$ is a set of subwords of words in L. Since a set of subwords of words in a bounded set is bounded [7], $P_i(L')$ is bounded for each i, $1 \leqslant i \leqslant n$.

We will need the following theorem whose proof can be found in [7].

THEOREM 6.8. *It is recursively solvable to determine whether an arbitrary context-free grammar* $G = \langle V, P, S, \Sigma \rangle$ *generates a bounded context-free language. If* $L(G)$ *is bounded, then words* $w_1, ..., w_k$ *in* Σ^* *can be effectively found such that* $L(G) \subseteq w_1^* \cdots w_k^*$.

THEOREM 6.9. *It is recursively solvable to determine whether an arbitrary n-SMG* $G = \langle V_1, ..., V_n, P, S, \Sigma \rangle$ *generates a bounded simple matrix language. If* $L(G_n)$ *is bounded, then words* $w_1, ..., w_k$ *in* Σ^* *can be effectively found such that* $L(G) \subseteq w_1^* \cdots w_k^*$.

Proof. Let $G_n = \langle V_1, ..., V_n, P, S, \Sigma \rangle$ be an n-SMG. Then by Theorem 1.1, we can effectively construct a context-free grammar $G' = \langle V', P', S', \Sigma_n \rangle$ such that $L(G_n) = \{x_1 \cdots x_n \mid (x_1, ..., x_n) \text{ in } \tau_n(L(G'))\}$. Now by Lemma 4.3, we can effectively find for each i, $1 \leqslant i \leqslant n$, a context-free grammar $G^i = \langle V^i, P^i, S^i, \Sigma \rangle$ such that $P_i(\tau_n(L(G'))) = L(G^i)$. The theorem now follows from Lemma 6.1 and Theorem 6.8.

The next theorem has been shown for the case of context-free languages (see, e.g., [7]). Since the same technique of proof applies to the general case of simple matrix languages, we choose to omit the proof.

[9] Recall that $P_i(L') = \{x_i \mid \exists x_1, ..., x_{i-1}, x_{i+1}, ..., x_n \text{ such that } (x_1, ..., x_n) \text{ in } L'\}$.

THEOREM 6.10. *It is recursively solvable to determine for arbitrary simple matrix languages L_1 and L_2, one of them bounded, whether* (a) $L_1 \subseteq L_2$, (b) $L_2 \subseteq L_1$, (c) $L_1 = L_2$.

ACKNOWLEDGMENT

The author thanks the referee for a critical reading of the manuscript and for his many useful comments and suggestions.

RECEIVED: December 1, 1969; REVISED: May 6, 1970.

REFERENCES

1. S. ABRAHAM, Some questions of phrase structure grammars I, *Computational Linguistics* 4 (1965), 61–70.
2. A. V. AHO, Indexed grammars—an extension of context-free grammars, *IEEE Conf. Rec. 8th Ann. Symp. Switching Automata Theory*, No. 16-C-56 (1967).
3. Y. BAR-HILLEL, M. PERLIS, AND E. SHAMIR, On formal properties of simple phrase structure grammars, *Z. Phonetik Sprachwissen. Kommunikat.* **14** (1961), 143–172.
4. M. DAVIS, "Computability and Unsolvability," McGraw-Hill Book Co., New York, 1958.
5. S. GINSBURG, "The Mathematical Theory of Context-Free Languages," McGraw-Hill Book Co., New York, 1966.
6. S. GINSBURG AND S. GREIBACH, Abstract families of languages, *IEEE Conf. Rec. 8th Ann. Symp. Switching Automata Theory*, Austin, Texas (1967).
7. S. GINSBURG AND E. H. SPANIER, Bounded Algol-like languages, *Trans. Amer. Math. Soc.* **113** (1964), 333–368.
8. M. A. HARRISON, "Introduction to Switching and Automata Theory," McGraw-Hill Book Co., New York, 1965.
9. M. A. HARRISON AND O. H. IBARRA, Multi-tape and multi-head pushdown automata, *Information and Control* **13** (1968), 433–470.
10. O. H. IBARRA, "Generalizations of Pushdown Automata," Ph.D. Thesis, University of California, Berkeley, 1967.
11. S. Y. KURODA, Classes of languages and linear bounded automata, *Information and Control* **7** (1964), 207–223.
12. R. J. PARIKH, On context-free languages, *J. Assoc. Comput. Mach.* **13** (1966), 570–581.
13. M. O. RABIN AND D. SCOTT, Finite automata and their decision problems, *IBM J. Res. Develop.* **3** (1959), 114–125.
14. A. L. ROSENBERG, "Nonwriting Extensions of Finite Automata," Ph.D. Thesis, Harvard University, 1965.
15. R. SIRMONEY, On equal matrix languages, *Information and Control* **14** (1969), 135–151.