UNIVERSITAS COMENIANA
ACTA MATHEMATICA UNIVERSITATIS COMENIANAE
XLVIII—XLIX — 1986

COMPLEXITY CLASSES OF g-SYSTEMS ARE AFL

BRANISLAV ROVAN, Bratislava

The notion of a generative system (g-system) was introduced in [3] in an attempt to facilitate a unified approach to various types of rewriting systems studied in the literature. The complexity of language description by sequential and parallel grammars was studied there. An important question arising in the complexity theory is the study of complexity classes. In the present paper we shall consider the complexity classes of g-systems. We shall show that each complexity class for the measure STATE$_{,1}$ (the number of states of the a-transducer of the g-system in normal form) is an abstract family of languages.

The paper consists of two sections. The first section briefly reviews the main notions studied. The results are presented in the second section. The reader is referred to [3] for further background and motivation concerning g-systems. Language theory notions unexplained here can be found in [1] and [2].

1. Preliminaries

We shall define the main notions here. The reader is assumed to be familiar with the basic notions and notations of the theory of formal languages (see e.g. [1] and [2]).

Definition 1.1: A *one-input finite state transducer with accepting states* (1-a-transducer) is a 6-tuple $M = (K, X, Y, H, q_0, F)$, where K is a finite set of states, X and Y are finite alphabets (input and output resp.), q_0 in K is the initial state, $F \subseteq K$ is the set of accepting (final) states, and H is a finite subset of $K \times X \times Y^* \times K$. In case H is a subset of $K \times X \times Y^* \times K$, M is said to be *ε-free*.

By a *computation* of such a 1-a-transducer a word $h_1 \ldots h_n$ in H^* is understood such that (i) $\mathrm{pr}_1(h_1) = q_0$, (ii) $\mathrm{pr}_4(h_n)$ is in F, and (iii) $\mathrm{pr}_1(h_{i+1}) = \mathrm{pr}_4(h_i)$ for $1 \leqslant i \leqslant n - 1$, where pr_i are homomorphisms on H^* defined by $\mathrm{pr}_i((x_1, x_2,$

$x_3, x_4)) = x_i$ for $i = 1, 2, 3$, and 4. The set of all computations of M is denoted by Π_M.

A *1-a-transducer mapping* is then defined for each language $L \subseteq X^*$ by $M(L) = \mathrm{pr}_3(\mathrm{pr}_2^{-1}(L) \cap \Pi_M)$. For a word w let $M(w) = M(\{w\})$.

Definition 1.2: An *[ε-free] generative system* (g-system) is a 4-tuple $G = (N, T, M, S)$, where N and T are finite alphabets of nonterminal and terminal symbols resp. (not necessarily disjoint), S in N is the initial nonterminal symbol, and M is a 1-a-transducer mapping [ε-free], with $M(w) = \emptyset$ for each w in $(T - N)^*$.

Definition 1.3: The *language generated by a g-system* $G = (N, T, M, S)$ is the language $L(G) = \{w \text{ in } T^*;\ S \overset{*}{\underset{G}{\Rightarrow}} w\}$, where $\overset{*}{\underset{G}{\Rightarrow}}$ is the transitive and reflexive closure of the *rewrite relation* $\underset{G}{\Rightarrow}$ defined by $u \underset{G}{\Rightarrow} v$ iff v is in $M(u)$.

When discussing g-systems we shall frequently employ the following expressions. We shall say that x is a *computation enabling a derivation step* $u \underset{G}{\Rightarrow} v$ in a g-system $G = (N, T, M, S)$, if x is a computation in Π_M such that $\mathrm{pr}_2(x) = u$ and $\mathrm{pr}_3(x) = v$. If h is a 4-tuple in H such that $\mathrm{pr}_2(h) = a$, we shall say that the 4-tuple h *is rewriting* the symbol a. The 4-tuple (p, a, a, p) is said to be a *copying cycle* (for the symbol a in the state p).

We shall now define a special type of g-systems to be studied in this paper.

Definition 1.4: A g-system $G = (N, T, M, S)$ with $M = (K, V, V, H, q_0, \{q_F\})$, $V = N \cup T$, is said to be in *normal form* if the following holds:

(i) H contains 4-tuples (q_0, x, x, q_0) and (q_F, x, x, q_F) for each x in V (i.e., M contains a copying cycle for each symbol in its initial state and in its single final state).

(ii) If (p, x, y, q_0) is in H, then $p = q_0$ and $x = y$ (i.e., the only 4-tuples h in H such that $\mathrm{pr}_4(h) = q_0$ are the copying cycles in the initial state).

(iii) If (q_F, x, y, p) is in H, then $p = q_F$ and $x = y$ (i.e., the only 4-tuples h in H such that $\mathrm{pr}_1(h) = q_F$ are the copying cycles in the final state).

(iv) If (p, x, y, q) is in H for some x in T, then $y = x$ and either $p = q = q_0$ or $p = q = q_F$ (i.e., the only 4-tuples rewriting terminal symbols are the copying cycles in the initial and final states).

(v) M is ε-free.

It can be shown [3] that for each ε-free g-system G there exists a g-system G' in normal form such that $L(G) = L(G')$.

Several complexity measures were introduced in [3] to study the complexity of g-systems. In what follows only one is considered.

Definition 1.5: Let $G = (N, T, M, S)$ be a g-system with $M = (K, V, V, H, q_0, \{q_F\})$. Define STATE($G$) to be the number of states in K. For a language L (definable by some g-system in normal form) let

$$\mathrm{STATE}_{\mathscr{N}}(L) = \min_{G \text{ in } \mathscr{N}} \{\mathrm{STATE}(G);\ L(G) = L\},$$

where $\mathscr{N}$ is the family of all g-systems in normal form.

We shall use the following notation for the complexity classes of the measure STATE.

Notation 1.6: For each $k \geqslant 1$ let $\boldsymbol{L}(k) = \{L;\ \mathrm{STATE}_{\mathscr{N}}(L) \leqslant k\}$.

Note that $\boldsymbol{L}(1)$ contains only $\emptyset$ and all languages of the form $\{x\}$, where x is a letter. The families $\boldsymbol{L}(k)$ for $k \geqslant 2$ are more interesting. We shall show that each of them is an AFL.

Definition 1.7 [1]: A family of languages containing a nonempty language is said to be an *abstract family of languages* (AFL) if it is closed under ε-free homomorphism, inverse homomorphism, intersecion with regular sets, union, concatenation and Kleene $+$.

2. Results

We shall consider the families $\boldsymbol{L}(k)$ for $k \geqslant 2$ in this section. Through a sequence of theorems we shall show that each of them is an AFL.

Theorem 2.1: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under union.

Proof: We shall prove the above theorem by constructing a g-system in normal form generating the union of languages generated by two given g-systems in normal form. The fact the given g-systems are in normal form enables us to "lay one over the other" and let them use the same set of states. Due to the normal form assumption we can rename nonterminals so that no cross-talk occurs.

Formally, let L_1 and L_2 be in $\boldsymbol{L}(k)$, i.e., there are two g-systems G_1 and G_2 in normal form such that $L(G_1) = L_1$, $L(G_2) = L_2$, $\mathrm{STATE}(G_1) \leqslant k$, and $\mathrm{STATE}(G_2) \leqslant k$. Let for i in $\{1, 2\}$ $G_i = (N_i, T_i, M_i, S_i)$ and $M_i = (K_i, V_i, V_i, H_i, q_{0i}, \{q_{Fi}\})$. Let $\mathrm{STATE}(G_1) = m$ and $\mathrm{STATE}(G_2 = n$. Let us assume that $m \geqslant n$. Let f be an injective mapping from K_2 to K_1 such that $f(q_{02}) = q_{01}$ and $f(q_{F2}) = q_{F1}$. Clearly we can assume that $(N_1 - T_1) \cap V_2 = (N_2 - T_2) \cap V_1 = \emptyset$. Let us construct g-system $G_3 = (N_3, T_3, M_3, S_3)$, $M_3 = (K_3, V_3, V_3, H_3, q_{03}, \{q_{F3}\})$ as follows. Let $K_3 = K_1$, $q_{03} = q_{01}$, $q_{F3} = q_{F1}$, $N_3 = N_1 \cup N_2 \cup \{S_3\}$, $T_3 = T_1 \cup T_2$, where S_3 is a new symbol, and the set $H_3 = H_1 \cup H'_2 \cup Q$, where $H'_2 = \{(f(q), x, y, f(p));\ (q, x, y, p) \text{ in } H_2\}$ and $Q = \{(q_{03}, S_3, S_1, q_{F3}), (q_{03}, S_3, S_2, q_{F3}), (q_{03}, S_3, S_3, q_{03}), (q_{F3}, S_3, S_3, q_{F3})\}$. Since G_1 and G_2 are in normal form, G_3 is also in normal form. We shall prove that $L(G_3) = L_1 \cup L_2$. Let w be in $L_1 \cup L_2$. Suppose w is in L_1, i.e., $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} w$. Since $H_1 \subseteq H_3$, it holds $S_1 \overset{*}{\underset{G_3}{\Rightarrow}} w$. Thus $S_3 \underset{G_3}{\Rightarrow} S_1 \overset{*}{\underset{G_3}{\Rightarrow}} w$ and w is in

$L(G_3)$. Suppose w is in L_2, i.e., $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} w$. Then $S_2 \overset{*}{\underset{G_3}{\Rightarrow}} w$. (For, if $(q_{02}, a_1, v_1, q_1)(q_1, a_2, v_2, q_2) \ldots (q_{r-1}, a_r, v_r, q_r)$ is a computation in M_2, then $(f(q_{02}), a_1, v_1, f(q_1))(f(q_1), a_2, v_2, f(q_2)) \ldots (f(q_{r-1}), a_r, v_r, f(q_r))$ is a computation in M_3.) Thus $S_3 \underset{G_3}{\Rightarrow} S_2 \overset{*}{\underset{G_3}{\Rightarrow}} w$ and w is in $L(G_3)$. To prove the converse inclusion, let w be in $L(G_3)$ and let $S_3 \underset{G_3}{\Rightarrow} v_1 \underset{G_3}{\Rightarrow} v_2 \underset{G_3}{\Rightarrow} \ldots \underset{G_3}{\Rightarrow} v_r = w$ be a derivation of w in G_3. It follows from the definition of H_3 that $v_1 = S_1$ or $v_1 = S_2$. Suppose $v_1 = S_1$. Since $\mathrm{pr}_3(x)$ is in V_1^+ for each x in H_3 with $\mathrm{pr}_2(x)$ in V_1^*, it follows that v_i is in V_1^+ for each i, $1 \leqslant i \leqslant r$. Since G_3 is in normal form, terminal symbols can be rewritten by copying cycles in the initial and final states only. From this and the assumption $(N_2 - T_2) \cap \cap V_1 = \emptyset$ it follows that $v_i \underset{G_1}{\Rightarrow} v_{i+1}$ for each i, $1 \leqslant i \leqslant r-1$. We can proceed similarly in case $v_1 = S_2$. (In each computation of M_3 on a word in V_2^+ only 4-tuples from H_2' can be used, thus a computation with the same input and output exists in M_2.) We obtain $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} w$, hence w is in $L(G_2) = L_2$. In each case w is in $L_1 \cup L_2$ and the proof is complete.

Theorem 2.2: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under intersection with regular sets.

Proof: The proof is based on an idea similar to that of the proof of the same property for the family of context-free languages. Given a g-system and a finite state automaton we shall construct a g-system which (in a sense) generates all computations of the given automaton on the words generated by the given g-system as sentential forms. The computations terminating in an accepting state are then rewritten to terminal words.

Let L be in $\boldsymbol{L}(k)$ and R be a regular set. Thus $L = L(G)$ for some g-system $G = (N, T, M, S)$, $M = (K, V, V, H, q_0, \{q_F\})$ in normal form such that $\mathrm{STATE}(G) = k$ and $R = L(A)$ for some finite state atomaton $A = (\bar{K}, \bar{T}, \delta, \bar{q}_0, F)$. Without the loss of generality we can assume that $\bar{T} = T$. Let us construct g-system $G' = (N', T, M', S')$, $M' = (K, V', V', H', q_0, \{q_F\})$ as follows. Let $N' = \bar{K} \times V \times \bar{K} \cup \{S'\}$, with S' new, and let $H' = H_1 \cup H_2 \cup H_3 \cup H_4$, where $H_1 = \{(q_0, S', (\bar{q}_0, S, q), q_F); q \text{ in } F\}$, $H_2 = \{(q, (r, x, t), (r, a_1, s_1)(s_1, a_2, s_2) \ldots (s_{j-1}, a_j, t), p); j \geqslant 1, a_i \text{ is in } V, 1 \leqslant i \leqslant j, s_i \text{ is in } \bar{K} \text{ for } 1 \leqslant i \leqslant j-1, (q, x, a_1 \ldots a_j, p) \text{ is in } H, r \text{ and } t \text{ are in } \bar{K}\}$, $H_3 = \{q_0, (r, x, t), x, q_F); \delta(r, x) = t \text{ and } x \text{ is in } T\}$, and $H_4 = \{(q, x, x, q); q \text{ is in } \{q_0, q_F\}, x \text{ in } V'\}$.

We shall prove that $L(G') = L \cap R$. First we shall show the following.

$(+)$ Let $S \overset{*}{\underset{G}{\Rightarrow}} a_1 \ldots a_n$, a_i in V for each i, $1 \leqslant i \leqslant n$. Then

$$S \overset{*}{\underset{G'}{\Rightarrow}} (s_0, a_1, s_1)(s_1, a_2, s_2) \ldots (s_{n-1}, a_n, s_n)$$

for each sequence of states $s_0, \ldots, s_n$ in $\bar{K}$ such that $s_0 = q_0$ and $s_n = q_F$.

The proof is by induction on the length of the derivation. Let the length be $d = 0$. Then $n = 1$, $a_1 = S$ and $(q_0, S', (s_0, S, s_1), q_F)$ is in $H_1 \subseteq H'$. Thus $S' \underset{G'}{\Rightarrow} (s_0, S, s_1)$. Suppose now that (+) holds for all sentential forms in G derivable by derivations of length at most d. Let $a_1 ... a_n$ be a sentential form in G with the shortest derivation of length $d + 1$. Let it be the derivation $S \overset{*}{\underset{G}{\Rightarrow}} b_1 ... b_m \underset{G}{\Rightarrow}$ $\underset{G}{\Rightarrow} a_1 ... a_n$, with each b_i in V. Let the computation enabling the last step of this derivation be the computation (q_0, b_1, v_1, q_1) (q_1, b_2, v_2, q_2) ... (q_{m-1}, b_m, v_m, q_m), with $q_m = q_F$. Let (i_1, j_1), (i_2, j_2), ..., (i_m, j_m) be the sequence of pairs of integers such that j_r is the length of v_r and $i_r = \sum_{t=0}^{r-1} j_t$ (i.e., we can write $v_r = a_{i_r+1} ... a_{i_r+j_r}$) for each r. By the inductive hypothesis we have

$$S' \overset{*}{\underset{G'}{\Rightarrow}} (s_0, b_1, s_{i_2})(s_{i_2}, b_2, s_{i_3}) ... (s_{i_m}, b_m, s_n).$$

Since (q_r, b_r, v_r, q_{r+1}) is in H for each r, $1 \leqslant r \leqslant m$, the set H_2 (thus H') contains for each such r the fourtuple $(q_r, (s_{i_r}, b_r, s_{i_{r+1}}), (s_{i_r}, a_{i_r+1}, s_{i_r+1}) (s_{i_r+1}, a_{i_r+2}, s_{i_r+2}) ... (s_{i_r+j_r-1}, a_{i_r+j_r-1}, s_{i_{r+1}}), q_{r+1})$. Thus

$$(s_0, b_1, s_{i_2})(s_{i_2}, b_2, s_{i_3}) ... (s_{i_m}, b_m, s_n) \underset{G'}{\Rightarrow} (s_0, a_1, s_1)(s_1, a_2, s_2) ... (s_{n-1}, a_n, s_n),$$

proving the inductive step.

We can now prove $L \cap R \subseteq L(G')$. Let w be in $L \cap R$, $w = a_1 ... a_n$ for some a_1, ..., a_n in T, $n \geqslant 1$. Since w is in $R = L(A)$, there is an accepting computation of A on w $(\bar{q}_0, a_1 ... a_n) \vdash (s_1, a_2 ... a_n) \vdash ... \vdash (s_n, a_n) \vdash (s_n, \varepsilon)$, s_n in F. Since w is in $L = L(G)$, there is a derivation $S \overset{*}{\underset{G}{\Rightarrow}} a_1 ... a_n$ of w in G. It follows from (+) that

$$S' \overset{*}{\underset{G'}{\Rightarrow}} (\bar{q}_0, a_1, s_1)(s_1, a_2, s_2) ... (s_{n-1}, a_n, s_n).$$

Using the fact the g-system G' is in normal form (hence it has a copying cycle for each symbol in q_0 and q_F), we have by the definition of H_3

$$(\bar{q}_0, a_1, s_1)(s_1, a_2, s_2) ... (s_{n-1}, a_n, s_n) \underset{G'}{\Rightarrow}$$

$$\underset{G'}{\Rightarrow} a_1(s_1, a_2, s_2) ... (s_{n-1}, a_n, s_n) \underset{G'}{\Rightarrow}$$

$$\underset{G'}{\Rightarrow} a_1 a_2 (s_2, a_3, s_3) ... (s_{n-1}, a_n, s_n) \underset{G'}{\Rightarrow} ... \underset{G'}{\Rightarrow} a_1 a_2 ... a_n$$

Thus $S' \overset{*}{\underset{G'}{\Rightarrow}} a_1 ... a_n$ and w is in $L(G')$.

To prove $L(G') \subseteq L \cap R$ we have to prove several properties of derivations in G'.

(++) If $S' \underset{G'}{\Rightarrow} (s_1, x_1, s_1') (s_2, x_2, s_2') ... (s_n, x_n, s_n')$, then $s_i' = s_{i+1}$ for each

i, $1 \leqslant i \leqslant n-1$, $s_1 = q_0$, and s'_n is in F.

We prove $(++)$ by induction on the length of the derivation. For all sentential forms having derivations of length one $(++)$ holds due to the definition of H' (see definition of H_1). Now suppose $(++)$ holds for all sentential forms in $(\bar{K} \times V \times \bar{K})^+$ having a derivation of length at most d. Let a minimal derivation $S' \overset{*}{\underset{G'}{\Rightarrow}} u \underset{G'}{\Rightarrow} x$ for a sentential form $x = (s_1, x_1, s'_1)\,(s_2, x_2, s'_2) \ldots (s_n, x_n, s'_n)$ be of length $d+1$. By the definition of H' u is in $(\bar{K} \times V \times \bar{K})^+$ (for otherwise x could not be in $(\bar{K} \times V \times \bar{K})^+$) and by the inductive hypothesis $u = (r_1, u_1, r_2)\,(r_2, u_2, r_3) \ldots (r_m, u_m, r_{m+1})$ with $r_1 = \bar{q}_0$ and r_{m+1} in F. By the definition of H' the only 4-tuples (besides the copying cycles in H_4) that can be used in the computation enabling the derivation step $u \underset{G'}{\Rightarrow} x$ are those in H_2. In case a symbol (r_i, u_i, r_{i+1}) is rewritten by a 4-tuple in H_2, it is replaced by some nonempty string $(t_1, v_1, t'_1) \ldots (t_s, v_s, t'_s)$ such that $t_1 = r_i$, $t'_s = r_{i+1}$, and $t'_j = t_{j+1}$ for all j, $1 \leqslant j \leqslant s-1$. Thus x clearly satisfies $(++)$, proving the inductive step. Hence $(++)$ holds.

We now prove

$(+++)$ If $S' \overset{*}{\underset{G'}{\Rightarrow}} (s_1, x_1, s_2)\,(s_2, x_2, s_3) \ldots (s_n, x_n, s_{n+1})$, then $S \overset{*}{\underset{G}{\Rightarrow}} x_1 \ldots x_n$.

The proof is again by induction on the length of the derivation. Derivations of length one are in this case enabled by computations consisting of a single 4-tuple in H_1. Since $S \overset{*}{\underset{G}{\Rightarrow}} S$, $(+++)$ holds in this case. Let it hold for all sentential forms $(s_1, x_1, s_2)\,(s_2, x_2, s_3) \ldots (s_n, x_n, s_{n+1})$ with a derivation of length at most d. Let x be a sential form with a minimal derivation $S' \overset{*}{\underset{G'}{\Rightarrow}} u \underset{G'}{\Rightarrow} x$ of length $d+1$. Similarly as in the proof of $(++)$ we have that u is in $(\bar{K} \times V \times \bar{K})^+$. Thus, by $(++)$, we can write u in a form $u = (r_1, u_1, r_2)\,(r_2, u_2, r_3) \ldots (r_m, u_m, r_{m+1})$. By the inductive hypothesis $S \overset{*}{\underset{G}{\Rightarrow}} u_1 \ldots u_m$. Let the computation of M' enabling $u \underset{G'}{\Rightarrow} x$ be $(q_0, (r_1, u_1, r_2), y_1, q_1)\,(q_1, (r_2, u_2, r_3), y_2, q_2) \ldots (q_{m-1}, (r_m, u_m, r_{m+1}), y_m, q_m)$, $q_m = q_F$. Each 4-tuple is in $H_2 \cup H_4$ due to the form of x. Let h be a homomorphism, h: $(\bar{K} \times V \times \bar{K})^+ \to V^+$, defined by $h((p, a, q)) = a$ for each p, q, a. By the definition of H_2 and the fact G is in normal form it then follows that for each i, $0 \leqslant i \leqslant m-1$, $(q_i, h((r_{i+1}, u_{i+1}, r_{i+2})), h(y_{i+1}), q_{i+1})$ is in H. Thus $(q_0, u_1, h(y_1), q_1)\,(q_1, u_2, h(y_2), q_2) \ldots (q_{m-1}, u_m, h(y_m), q_m)$ is a computation of M. Clearly $x = y_1 \ldots y_m$, thus $h(y_1) \ldots h(y_m) = x_1 \ldots x_n$ and $S \overset{*}{\underset{G}{\Rightarrow}} x_1 \ldots x_n$. This completes the proof of $(+++)$.

Let us now consider symbols in $\bar{K} \times T \times \bar{K}$. Since G is in normal form, all symbols in $\bar{K} \times T \times \bar{K}$ are in a given derivation step rewritten by copying cycles in q_0 or q_F, or there is only one symbol changed in this step and that is by a 4-tuple in H_3. The terminal symbol obtained using a 4-tuple in H_3 can be later rewritten only by copying cacles in q_0 or q_F. That means, for each computation of M' enabling a derivation step $uav \underset{G'}{\Rightarrow} u'av'$ with a in T, there is a computation of M' (differing by one copying cycle only) enabling $u(p, a, q)v \underset{G'}{\Rightarrow} u'(p, a, q)v'$ for any p and q. Furthermore, regardless of u and v, if $q = \delta(p, a)$ then $u(p, a, q)v \underset{G'}{\Rightarrow}$ $\underset{G'}{\Rightarrow} uav$. Thus the following holds.

(+ + + +) For each x in $L(G')$ there is a derivation in G' $S' \underset{G'}{\overset{*}{\Rightarrow}} u \underset{G'}{\overset{*}{\Rightarrow}} x$ such that u is in $(\bar{K} \times T \times \bar{K})^+$ and all computations enabling the derivation steps in $u \underset{G'}{\overset{*}{\Rightarrow}} x$ are in $H_4^* H_3 H_4^*$.

Now we are ready to prove $L(G') \subseteq L \cap R$. Let w be in $L(G')$. By (+ + + +) there is a derivation $S' \underset{G'}{\overset{*}{\Rightarrow}} u \underset{G'}{\overset{*}{\Rightarrow}} w$ such that u is in $(\bar{K} \times T \times \bar{K})^+$. By (+ +) u is of the form $u = (\bar{q}_0, u_1, q_1)\ (q_1, u_2, q_2)\ \dots\ (q_{n-1}, u_n, q_n)$, with q_n in F. By the definition of H_3 for each i, $1 \leqslant i \leqslant n-1$, $\delta(q_i, u_{i+1}) = q_{i+1}$ and $\delta(\bar{q}_0, u_1) = q_1$. Thus $(\bar{q}_0, u_1 \dots u_n) \vdash^* (q_n, \varepsilon)$ is an accepting computation of A on $u_1 \dots u_n$. Furthermore it follows from the definition of H_3 that $u_1 \dots u_n = w$, thus w is in R. By (+ + +) $S \underset{G}{\overset{*}{\Rightarrow}} u_1 \dots u_n$, hence w is also in L. This completes the proof.

Theorem 2.3: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under substitution by context-free languages not containing the empty word.

Proof: Let L be in $\boldsymbol{L}(k)$ and let $G = (N, T, M, S)$ be a g-system $M = (K, V, V, H, q_0, q_F)$, such that $L(G) = L$ and STATE$(G) \leqslant k$. Let τ be a substitution such that $\tau(a)$ is an ε-free context-free language for each symbol a, with $\tau(a) =$ $= L(G_a)$ for an ε-free context-free grammar $G_a = (N_a, T_a, P_a, S_a)$. Denoting $N_a \cup T_a$ by V_a as usual, we can assume without loss of generality that $N \cap$ $\cap \bigcup_a V_a = \emptyset$ and for each a $N_a \cap \left(\bigcup_{b \neq a} V_b \cup V\right) = \emptyset$. Let $h\colon V^* \to \left(N \cup \bigcup_a \{S_a\}\right)$ be a homomorphism defined by $h(x) = x$ for x in N and $h(x) = S_x$ for x in T. We shall construct a g-system G' that will first generate a word $h(w)$ for w in L and then derive from each S_a in $h(w)$ a word in $\tau(a)$. Let $G' = (N', T', M', S)$, where $M' = (K, V', V', H', q_0, q_F)$, $N' = N \cup \bigcup_a N_a$, $T' = \bigcup_a T_a$ and $H' = H_1 \cup$ $\cup H_2 \cup H_3$, where

$$H_1 = \{(p, h(x), h(y), q);\ (p, x, y, q) \text{ in } H\},$$

$$H_2 = \left\{(q_0, x, y, q_F);\ x \to y \text{ in } \bigcup_a P_a\right\}, \text{ and}$$

$$H_3 = \{(q, x, x, q);\ x \text{ in } V',\ q \text{ in } \{q_0, q_F\}\}.$$

It is easy to prove that G' is in normal form and STATE(G') = STATE(G). We shall show that $L(G') = L(G)$.

From the definition of H_2 and the fact G' is in normal form (thus it contains a copying cycle for each symbol in q_0 and q_F) we have

$(+)$ If $u \underset{G_a}{\Rightarrow} v$ for some a and words u, v in V_a^+, then $xuy \underset{G'}{\Rightarrow} xvy$ for all words for all words x, y in $(V')^*$.

It is easy to show by induction on the length of the derivation (based on the definition of H_1) that the following holds.

$(++)$ $S \overset{*}{\underset{G}{\Rightarrow}} x$ if and only if $S \overset{*}{\underset{G'}{\Rightarrow}} h(x)$.

Consider now w in $\tau(L)$. Then there is a word $x = a_1 \ldots a_n$ in L, $n \geqslant 1$ and each a_i in T, and words w_1 in $\tau(a_1)$, $\ldots$, w_n in $\tau(a_n)$ such that $w = w_1 \ldots w_n$. Since x is in L, $S \overset{*}{\underset{G}{\Rightarrow}} x$ and by $(++)$ we have

$(+++)$ $S \overset{*}{\underset{G'}{\Rightarrow}} S_{a_1} \ldots S_{a_n}$.

Now, for each i, $1 \leqslant i \leqslant n$, $S_{a_i} \overset{*}{\underset{G_{a_i}}{\Rightarrow}} w_i$, since w_i is in $\tau(a_i)$. Thus by $(+)$ we have

$(++++)$ $S_{a_1} \ldots S_{a_n} \overset{*}{\underset{G'}{\Rightarrow}} w_1 S_2 \ldots S_n \overset{*}{\underset{G'}{\Rightarrow}} w_1 w_2 S_{a_3} \ldots S_{a_n} \overset{*}{\underset{G'}{\Rightarrow}} w_1 w_2 \ldots w_n = w$.

By $(+++)$ and $(++++)$ $S \overset{*}{\underset{G'}{\Rightarrow}} w$, thus w is in $L(G')$.

Consider now the converse inclusion. By the definition of H' there are only two types of computations in M', namely those in $H_3^* H_1^* H_3^*$ and those in $H_3^* H_2 H_3^*$. Similarly to the case of the symbols in $\bar{K} \times T \times \bar{K}$ in the proof of Theorem 2.2 the symbols in $V' - N$ have a special role in the derivations in G'. Each symbol from $V' - N$ in a sentential form is during one derivation step either just copied (by a copying cycle in q_0 or q_F) or rewritten by a fourtuple in H_2. In the latter case it is at most one symbol in the sentential form that is changed and it is replaced by a string in $(V' - N)^+$. It therefore follows that

(§) If $uxv \underset{G'}{\Rightarrow} uyv$, where x is in $V' - N$ and y is in $(V' - N)^+$, is enabled by a computation in $H_3^* H_2 H_3^*$ and some computation in $H_3^* H_1^* H_3^*$ enables $uyv \underset{G'}{\Rightarrow} u'y'v'$, then $y' = y$, and $uxv \underset{G'}{\Rightarrow} u'xv' \underset{G'}{\Rightarrow} u'yv'$.

The assertion (§) thus says that the derivation steps enabled by computations in $H_3^* H_2 H_3^*$ may be deferred. Thus the following holds.

(§§) For each word x in $L(G')$ there is a derivation of x in G' $S \overset{*}{\underset{G'}{\Rightarrow}} u \overset{*}{\underset{G'}{\Rightarrow}} x$ with u in $\{S_a;\ a \text{ in } T\}^*$ such that the derivation steps enabled by computations in $H_3^* H_2 H_3^*$ are exactly those in $u \overset{*}{\underset{G'}{\Rightarrow}} x$.

The assertion (§§) follows (by induction) directly from (§). It suffices to show that u is in $\{S_a;\ a \text{ in } T\}^*$. This follows from the following facts. (i) All computations enabling derivation steps in $S \overset{*}{\underset{G'}{\Rightarrow}} u$ are in $H_3^* H_1^* H_3^*$. (ii) S is in N. (iii) There is no fourtuple in $H_1 \cup H_3$ rewriting a symbol in N to a string containing a terminal symbol. (iv) Only fourtuples in H_2 can rewrite a nonterminal symbol to a string containing a terminal symbol. (v) The fourtuples in H_2 can rewrite symbols from $V' - N$ only. And finally. (vi) The symbols in $\{S_a;\ a \text{ in } T\}$ are the only symbols in $V' - N$ to which symbols in N can be rewritten.

Now, let x, y be in $N' - N$ and r, s, and t be in $(V')^*$. Clearly, if

$$rxsyt \underset{G'}{\Rightarrow} rxsy't \underset{G'}{\Rightarrow} rx'sy't,$$

then it also holds that

$$rxsyt \underset{G'}{\Rightarrow} rx'syt \underset{G'}{\Rightarrow} rx'sy't.$$

It thus follows

(§§§) In (§§) we can moreover assume that in each step of the derivation $u \overset{*}{\underset{G'}{\Rightarrow}} x$ the fourtuple in H_2 rewrites the leftmost occurrence of a nonterminal symbol.

Now it is easy to show $L(G') \subseteq \tau(L)$. Let w be in $L(G')$. Let $S \overset{*}{\underset{G'}{\Rightarrow}} u \overset{*}{\underset{G'}{\Rightarrow}} w$ be a derivation of w satisfying (§§) and (§§§). Let $u = S_{a_1} \ldots S_{a_n}$ for $n \geqslant 1$ and some $a_1, \ldots, a_n$ in T. By $(++)$ $S \overset{*}{\underset{G}{\Rightarrow}} a_1 \ldots a_n$, thus $a_1 \ldots a_n$ is in L. Consider now the derivation $u \overset{*}{\underset{G'}{\Rightarrow}} w$. Let this derivation be

$$u \underset{G'}{\Rightarrow} u_1 \underset{G'}{\Rightarrow} u_2 \underset{G'}{\Rightarrow} \ldots \underset{G'}{\Rightarrow} u_r = w$$

and let $i_1, \ldots, i_n$ be those indices for which $u_{i_j} = w_1 w_2 \ldots w_{j-1} S_{a_j} \ldots S_{a_n}$ for each j, $1 \leqslant j \leqslant n$ and w_j in $(T')^+$. Since all computations enabling the derivation steps in this derivation are in $H_3^* H_2 H_3^*$, we have $S_{a_j} \overset{*}{\underset{G'}{\Rightarrow}} w_j$ for each j, $1 \leqslant j \leqslant n$. Since we assumed $N_a \cap \left(\bigcup_{b \neq a} V_b \cup V \right) = \emptyset$, it follows from the definition of H' (H_2 in particular) that $S_{a_j} \overset{*}{\underset{G_{a_j}}{\Rightarrow}} w_j$, $1 \leqslant j \leqslant n$. Thus each w_j is in $\tau(a_j)$ and w in $\tau(a_1 \ldots a_n)$ is in $\tau(L)$. This completes the proof.

The following two corollaries are immediate consequences of the above theorem.

Corollary 2.4: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under ε-free regular substitution.

Corollary 2.5: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under ε-free homomorphism.

In order to show that each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under inverse homomorphism we need three lemmas.

Lemma 2.6: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under union with regular sets.

Proof: It is easy to see that for $k \geqslant 2$ each $\boldsymbol{L}(k)$ contains the family of regular sets. The result then follows by Theorem 2.1.

We shall need the following special type of homomorphism [1].

Definition 2.7: Let A be an alphabet and c a symbol not in A. Let h be a homomorphism defined by $h(a) = a$ for each a in A and $h(c) = \varepsilon$. Let $L \subseteq$

$\subseteq (A^+\{c^i;\ 1 \leqslant i \leqslant n\})^+$ for some fixed n. The homomorphism h is said to be *limited symbol erasing* (on L).

Lemma 2.8: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under limited symbol erasing.

Proof: Let L be in $\boldsymbol{L}(k)$ and let $L \subseteq (\bar{T}\{c^i;\ 0 \leqslant i \leqslant n\})^+$, where c is a symbol not in $\bar{T}$. Let h be a homomorphism such that $h(a) = a$ for a in $\bar{T}$ and $h(c) = \varepsilon$. Let $G = (N, T, M, S)$, with $M = (K, V, V, H, q_0, \{q_F\})$ be a g-system such that $L(G) = L$ and $\mathrm{STATE}(G) \leqslant k$. We shall define a g-system $G' = (N', T, M', S)$, with $M' = (K, V', V', H', q_0, \{q_F\})$ such that $L(G') = h(L)$ (and clearly $\mathrm{STATE}(G') = \mathrm{STATE}(G)$). Let the nonterminals be $N' = \{S\} \cup \{[w];\ w \text{ in } N^+,\ 1 \leqslant |w| \leqslant 2n+1\}$ and let $H' = H_1 \cup H_2 \cup H_3 \cup H_4 \cup H_5$, where $H_1 = \{(q_0, S, f(y), q_F);\ (q_0, S, y, q_F) \text{ in } H\}$, $H_2 = \{(q_0, [x], f(y_1 \ldots y_m), q_F);\ 1 \leqslant |x| \leqslant n,\ x = x_1 \ldots x_m$, each x_i in V, and $(q_0, x_1, y_1, q_1)(q_1, x_2, y_2, q_2) \ldots (q_{m-1}, x_m, y_m, q_F)$ is a computation of M for some $q_1, \ldots, q_{m-1}\}$, $H_3 = \{(p, [x], f(y_1 \ldots y_m), q);\ p, q$ in K, $n+1 \leqslant |x| \leqslant 2n+2$, $x = x_1 \ldots x_m$, with at least one x_i in N, and there exists a sequence of fourtuples $(p, x_1, y_1, q_1)(q_1, x_2, y_2, q_2) \ldots (q_{m-1}, x_m, y_m, q)$ in H^+ for some $q_1, \ldots, q_{m-1}\}$, $H_4 = \{(q_0, [x], h(x), q_F);\ x \text{ in } T^+ - c^+\}$, and $H_5 = \{(q, a, a, q);\ q \text{ in } \{q_0, q_F\},\ a \text{ in } V'\}$, with the function $f\colon V^* \to (N')^*$ defined by $f(a_1 \ldots a_m) = [a_1 \ldots a_m]$ if $1 \leqslant m \leqslant n$ and

$$f(a_1 \ldots a_m) = [a_1 \ldots a_{n+1}][a_{n+2} \ldots a_{2n+2}] \ldots [a_{(t-1)n+t} \ldots a_{tn+t}][a_{tn+t+1} \ldots a_m]$$

if $m = (t+1)(n+1) + r$ for some $t \geqslant 0$ and $0 \leqslant r \leqslant n$.

The function f thus assigns to each nonempty string of length at most n one symbol (the symbol $[x]$ to a string x) and to each string x of length exceeding n it assigns a sequence of symbols $[x_0][x_1] \ldots [x_{t-1}][x_t]$ such that $|x_i| = n+1$, $1 \leqslant i \leqslant t-1$, $n+1 \leqslant |x_t| \leqslant 2n+1$, and $x = x_0 \ldots x_t$.

It is easy to verify that G' is in normal form. It thus remains to show that $L(G') = h(L)$. We shall use the following assertions.

(+) If $S \overset{*}{\underset{G'}{\Rightarrow}} [x_1] \ldots [x_m]$, then $S \overset{*}{\underset{G}{\Rightarrow}} x_1 \ldots x_m$.

(++) If $S \overset{*}{\underset{G}{\Rightarrow}} u$ and $u \neq S$, then there is $m \geqslant 1$ and $u_1, \ldots, u_m$ such that $u = u_1 \ldots u_m$ and $S \underset{G'}{\Rightarrow} [u_1] \ldots [u_m]$.

We shall prove (+) by induction on the length of the derivation of the sential form $[x_1] \ldots [x_m]$. For the derivation of length one the assertion follows from the definition of H_1. Suppose (+) holds for the sentential forms having a derivation of length at most P. Let $[x_1] \ldots [x_m]$ be a sentential form with a minimal derivation of length $P+1$. Let $S \overset{*}{\underset{G'}{\Rightarrow}} u \underset{G'}{\Rightarrow} x$ be such a derivation and $u = [u_1] \ldots [u_p]$. If $p = 1$, then the compuation of M' enabling the derivation step $u \underset{G'}{\Rightarrow} x$ consists of a single fourtuple (in H_2 or H_3, depending on the length of u_1) and thus by the definition of H_2 or H_3 resp. it follows $u_1 \underset{G}{\Rightarrow} x_1 \ldots x_m$. Now suppose $p \geqslant 2$ and let for each i, $1 \leqslant i \leqslant p$, $u_i = u_{i_1} \ldots u_{ir_i}$ with all u_{ij} in V. Let the computation of M' enabling $u \underset{G'}{\Rightarrow} x$ be $(q_0, [u_1], z_1, q_1)\,(q_1, [u_2], z_2, q_2) \ldots (q_{p-1}, [u_p], z_p, q_p)$, $q_p = q_F$. By the definition of H_3 for each fourtuple $(q_{i-1}, [u_i], z_i, q_i)$ in this computation there is a sequence of fourtuples of M $(q_{i-1}, u_{i_1}, y_{i_1}, q_{i_1})\,(q_{i_1}, u_{i_2}, y_{i_2}, q_{i_2}) \ldots (q_{i(r_i-1)}, u_{ir_i}, y_{ir_i}, q_i)$ such that $z_i - f(y_{i_1} \ldots y_{ir_i})$. Concatenating these sequences we clearly obtain a computation of the 1-a-transducer M on the word

$$u_{11}u_{12}\ldots u_{1r_1}u_{21}\ldots \quad \ldots u_{pr_p} = u_1 \ldots u_p$$

with the output

$$y_{11}y_{12}\ldots y_{1r_1}y_{21}\ldots \quad \ldots y_{pr_p} = x_1 \ldots x_m.$$

(The last equality follows from the fact that $x = z_1 \ldots z_p = [x_1] \ldots [x_m]$.) It thus holds $u_1 \ldots u_p \underset{G}{\Rightarrow} x_1 \ldots x_m$. By the inductive hypothesis $S \overset{*}{\underset{G}{\Rightarrow}} u_1 \ldots u_p$, hence $S \overset{*}{\underset{G}{\Rightarrow}} x_1 \ldots$ $\ldots x_m$. This proves $(+)$.

To prove $(++)$ we shall again proceed by induction on the length of the derivation of the sentential form u. For the derivation of length one the assertion follows by the definition of H_1. Suppose $(++)$ holds for sentential forms having a derivation of length at most P. Let u be a sentential form with a minimal derivation of length $P+1$. Let $S \overset{*}{\underset{G}{\Rightarrow}} v \underset{G}{\Rightarrow} u$ be such a derivation. By the inductive hypothesis there exist $m \geqslant 1$ and $v_1, \ldots, v_m$ such that $v = v_1 \ldots v_m$ and $S \overset{*}{\underset{G'}{\Rightarrow}} [v_1] \ldots$ $\ldots [v_m]$. Let the computation of M enabling $v \underset{G}{\Rightarrow} u$ be $(q_0, a_1, x_1, q_1)\,(q_1, a_2, x_2, q_2)$ $\ldots (q_{r-1}, a_r, x_r, q_r)$, with $q_r = q_F$. Clearly $v = a_1 \ldots a_r$. Let $i_1, \ldots, i_{m+1}$ be indices such that for each j, $1 \leqslant j \leqslant m$, $v_j = a_{i_j}a_{i_j+1} \ldots a_{i_{j+1}-1}$. (Thus $i_1 = 1$ and $i_{m+1} =$ $= r+1$). By the definition of H_3 (or H_2 if $r = n$) H' contains $(q_{i_1-1}, [v_j],$ $f(x_{i_j} \ldots x_{i_{j+1}-1}), q_{i_{j+1}-1})$ for each j, $1 \leqslant j \leqslant m$. Thus, noting that $i_1 - 1 = 0$, $i_{m+1} - 1 = r$, and $q_r = q_F$, we can see that

$$(q_{i_1-1}, [v_1], f(x_{i_1} \ldots x_{i_2-1}), q_{i_2-1})\,(q_{i_2-1}, [v_2],$$
$$f(x_{i_2} \ldots x_{i_3-1}), q_{i_3-1}) \ldots (q_{im-1}, [v_m], f(x_{i_m} \ldots x_{i_{m+1}-1}), q_{i_{m+1}-1})$$

is a computation of M' and therefore

$$S \overset{*}{\underset{G'}{\Rightarrow}} f(x_{i_1} \ldots x_{i_2-1}) f(x_{i_2} \ldots x_{i_3-1}) \ldots f(x_{i_m} \ldots x_{i_{m+1}-1}).$$

By the definition of f there exist $u_1, \ldots, u_t$ such that

$$f(x_{i_1} \ldots x_{i_2-1}) f(x_{i_2} \ldots x_{i_3-1}) \ldots f(x_{i_m} \ldots x_{i_{m+1}-1}) = [u_1] \ldots [u_t]$$

and clearly $u_1 \ldots u_t = x_{i_1} \ldots x_{i_{m+1}-1} = x_1 \ldots x_r = u$. Thus $(++)$ holds.

We shall now show that $h(L) \subseteq L(G')$. Let w be in $h(L)$. Thus there exists a (nonempty) word u in L such that $w = h(u)$. By $(++)$ there exist words $u_1, \ldots,$

u_m. It is easy to see that $\{u_1, \ldots, u_m\} \cap c^+ = \emptyset$, i.e., none of the words u_i consists entirely of letters c. [For, with the exception of the case $m = 1$ the length of all words $u_1, \ldots, u_m$ is at least $n + 1$. Since $L \subseteq (\bar{T}\{c^i;\ 0 \leqslant i \leqslant n\})^+$, no word in L contains more than n consecutive letters c. Thus each of the words $u_1, \ldots, u_m$ contains at least one letter distinct from c. In the case $m = 1$ it may hold $|u_1| \leqslant n$. However, in this case $u = u_1$ and due to the fact that $L \cap c^+ = \emptyset$ u_1 contains at least one letter distinct from c as well.] Thus $(q_0, [u_i], h(u_i), q_F)$ is in H_4 for each symbol $[u_i]$. It thus follows

$$[u_1] \ldots [u_m] \underset{G'}{\Rightarrow} h(u_1)[u_2] \ldots [u_m] \underset{G'}{\Rightarrow}$$
$$\underset{G'}{\Rightarrow} h(u_1)h(u_2)[u_3] \ldots [u_m] \underset{G'}{\Rightarrow} \ldots \underset{G'}{\Rightarrow} h(u_1)h(u_2) \ldots h(u_m).$$

Clearly $h(u_1) \ldots h(u_m) = h(u) = w$. Thus $S \overset{*}{\underset{G'}{\Rightarrow}} w$ and w is in $L(G')$.

We shall now prove the converse inclusion, $L(G') \subseteq h(L)$. Since G is in normal form, only the fourtuples in H_4 or H_5 may have in their second component symbols $[x]$ for x in T^+. Besides, only the fourtuples in H_5 may have a terminal symbol in their second component. It is hence easy to see that

(+ + +) If $u[x]v \underset{G'}{\Rightarrow} uh(x)v \underset{G'}{\Rightarrow} u'h(x)v'$ for some x in T^+ and words u, v, u', v' in $(V')^*$, then it also holds that $u[x]v \underset{G'}{\Rightarrow} u'[x]v' \underset{G'}{\Rightarrow} u'h(x)v'$.

Based on (+ + +) we can show by an easy induction proof that for each derivation in G' we can find a derivation yielding the same word in which the derivation steps enabled by the computations in $H_5^* H_4 H_5^*$ occur at the end, i.e.,

(+ + + +) Each word u in $L(G')$ has a derivation $S \overset{*}{\underset{G'}{\Rightarrow}} v \overset{*}{\underset{G'}{\Rightarrow}} u$ such that none of the computations enabling the derivation steps in $S \overset{*}{\underset{G'}{\Rightarrow}} v$ contains a fourtuple in H_4 and all computations enabling the derivation steps in $v \overset{*}{\underset{G'}{\Rightarrow}} u$ are in $H_5^* H_4 H_5^*$.

Let w be in $L(G')$. Let $S \overset{*}{\underset{G'}{\Rightarrow}} v \overset{*}{\underset{G'}{\Rightarrow}} w$ be a derivation satisfying (+ + + +). Clearly v is of the form $v = [v_1] \ldots [v_m]$ for some $v_1, \ldots, v_m$ in T^+. By the definition of H_4 we have $w = h(v_1) \ldots h(v_m)$, i.e., $w = h(v_1 \ldots v_m)$. Since $S \overset{*}{\underset{G'}{\Rightarrow}} [v_1] \ldots [v_m]$, we have by (+) $S \overset{*}{\underset{G}{\Rightarrow}} v_1 \ldots v_m$. Hence $v_1 \ldots v_m$ is in L and $w = h(v_1 \ldots v_m)$ in $h(L)$. This proves the second inclusion and the proof is complete.

Lemma 2.9 [1]: Let $\mathscr{L}$ be a family of languages closed under ε-free regular substitution, limited symbol erasing, union with ε-free regular sets, and intersection with regular sets. Then $\mathscr{L}$ is closed under inverse homomorphism.

Based on the above lemma, Theorem 2.2, Lemmas 2.6 and 2.8, and Corollary 2.4 we have the following theorem.

Theorem 2.10: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under inverse homomorphism.

To prove the closure of $\boldsymbol{L}(k)$ under Kleene $+$ we need the following auxiliary result.

Lemma 2.11: Let L be in $\boldsymbol{L}(k)$, $k \geqslant 2$, and let c be a new symbol. Then $(Lc)^+$ is in $\boldsymbol{L}(k)$.

Proof: Let $L = L(G)$ for some g-system $G = (N, T, M, S)$ in normal form, with $M = (K, V, V, H, q_0, q_F)$ and $\mathrm{STATE}(G) \leqslant k$. Let S' be a new symbol. Let us construct a g-system $G' = (N \cup \{S'\}, T \cup \{c\}, M', S')$, where $M' = (K, V', V', H', q_0, q_F)$ and $H' = H \cup H_1 \cup H_2$, with H_1 and H_2 defined by $H_1 = \{(q, x, x, q);$ q in $\{q_0, q_F\}$ and x in $\{S', c\}\}$ and $H_2 = \{(q_0, S', Sc, q_F), (q_0, S', S'Sc, q_F)\}$. It is easy to see that G' is in normal form and $\mathrm{STATE}(G') = \mathrm{STATE}(G)$. We shall prove that $L(G') = (Lc)^+$. Since $H \subseteq H'$, we have

(+) If $x \underset{G}{\Rightarrow} y$, then $x \underset{G'}{\Rightarrow} y$.

Besides, since G' is in normal form, we have

(+ +) If $x \underset{G'}{\Rightarrow} y$, then $uxv \underset{G'}{\Rightarrow} uyv$ for each u and v in $(V')^*$.

Suppose w is in $(Lc)^+$. Then there exist $n \geqslant 1$ and words $w_1, \ldots, w_n$ in L such that $w = w_1cw_2c\ldots w_nc$. By the definition of H' we have $S' \overset{*}{\underset{G'}{\Rightarrow}} S_1cS_2c\ldots S_nc$, where $S_1 = S_2 = \ldots = S_n = S$. Since each w_i is in L, $S_i \overset{*}{\underset{G}{\Rightarrow}} w_i$. By (+) and (+ +) thus

$$S' \overset{*}{\underset{G'}{\Rightarrow}} S_1cS_2c\ldots S_nc \overset{*}{\underset{G'}{\Rightarrow}} w_1cS_2c\ldots S_nc \overset{*}{\underset{G'}{\Rightarrow}} \ldots \overset{*}{\underset{G'}{\Rightarrow}} w_1cw_2c\ldots w_nc.$$

Hence w is in $L(G')$, proving $(Lc)^+ \subseteq L(G')$.

To prove the converse inclusion we shall need several assertions concerning derivations in G'. Since c is a new symbol, it follows from the definition of H' that c can be rewritten only by copying cycles in q_0 and q_F. Since G' satisfies the conditions (ii) and (iii) of Definition 1.4, we have

(§) If $u_1cu_2c\ldots u_nc \underset{G'}{\Rightarrow} v_1cv_2c\ldots v_nc$ for some $n \geqslant 1$ and some $u_1, \ldots, u_n, v_1, \ldots, v_n$ in V^+, then there is at most one index r, $1 \leqslant r \leqslant n$, such that $u_r \neq v_r$.

and furthermore

(§§) If $u_1cu_2c\ldots u_nc \underset{G'}{\Rightarrow} u_1c\ldots u_{j-1}cv_jcu_{j+1}c\ldots u_nc \underset{G'}{\Rightarrow} u_1c\ldots u_{i-1}cv_icu_{i+1}c\ldots u_{j-1}\cdot$ $\cdot cv_jcu_{j+1}c\ldots u_nc$ for some $u_1, \ldots, u_n, v_i, v_j$ in V^+ and $1 \leqslant i < j \leqslant n$, then $u_1cu_2c\ldots u_nc \underset{G'}{\Rightarrow} u_1c\ldots u_{i-1}cv_iu_{i+1}c\ldots u_nc \underset{G'}{\Rightarrow} u_1c\ldots u_{i-1}cv_iu_{i+1}c\ldots u_{j-1}cv_j\cdot$ $\cdot u_{j+1}c\ldots u_nc$.

By the definition of H' the initial nonterminal S' can be rewritten either by the copying cycles in q_0 or q_F, or its rewriting causes the transition from q_0 to q_F. It thus follows that

(§§§) If $S'u \underset{G'}{\Rightarrow} S'u' \underset{G'}{\Rightarrow} S'Scu'$ or $S'u \underset{G'}{\Rightarrow} S'u' \underset{G'}{\Rightarrow} Scu'$ and the computations enabling the first step of these derivations do not contain any fourtuple from H_2, then it also holds that $S'u \underset{G'}{\Rightarrow} S'Scu \underset{G'}{\Rightarrow} S'Scu'$ or $S'u \underset{G'}{\Rightarrow}$ $\underset{G'}{\Rightarrow} Scu \underset{G'}{\Rightarrow} Scu'$ resp.

By (§), (§§), and (§§§) we thus have

(§§§§) Each word w in $L(G')$ has a derivation of the form $S' \overset{*}{\underset{G'}{\Rightarrow}} u \overset{*}{\underset{G'}{\Rightarrow}} w$, where all computations enabling the derivation steps in $S' \overset{*}{\underset{G'}{\Rightarrow}} u$ are in $H_2\{(q_F, a, a, q_F); a \text{ in } V'\}^*$, and all computations enabling the derivation steps in $u \underset{G'}{\Rightarrow} w$ are in $\{(q_0, a, a, q_0); a \text{ in } T \cup \{c\}\}^* \; H^*\{(q_F, a, a, q_F); a \text{ in } V'\}^*$.

For, by (§§§) the steps in which S' is not copied can be "moved" towards the beginning of the derivation, by (§) one derivation step can change only one group of symbols separated by c's, and by (§§) we can change the order of derivation steps to make it the leftmost group containing a nonterminal symbol.

We shall now prove $L(G') \subseteq (Lc)^+$. Let w be in $L(G')$ and let $S' \overset{*}{\underset{G'}{\Rightarrow}} u \overset{*}{\underset{G'}{\Rightarrow}} w$ be its derivation satisfying (§§§§). Clearly u does not contain the symbol S' and it is the first such sentential form in this derivation of w. It then follows from the form of computations enabling the derivation steps in $S' \overset{*}{\underset{G'}{\Rightarrow}} u$ that $u = S_1cS_2c\ldots \ldots S_nc$, where $S_1 = S_2 = \ldots = S_n = S$, for some $n \geqslant 1$. Due to the form of the computations enabling the derivation steps in $u \overset{*}{\underset{G'}{\Rightarrow}} w$ given in (§§§§), there exist $w_1, \ldots, w_n$ in T^+ such that

$$u = S_1c\ldots S_nc \overset{*}{\underset{G'}{\Rightarrow}} w_1cS_2c\ldots S_nc \overset{*}{\underset{G'}{\Rightarrow}} \ldots \overset{*}{\underset{G'}{\Rightarrow}} w_1cw_2c\ldots w_nc.$$

Since except for the copying cycles in q_0 and q_F all the fourtuples used in these computations are in H, we have for each i, $1 \leqslant i \leqslant n$, $S_i \overset{*}{\underset{G}{\Rightarrow}} w_i$. Thus w_i is in L. w is in $(Lc)^n \subseteq (Lc)^+$ and the inclusion $L(G') \subseteq (Lc)^+$ holds. Both inclusions imply the equality and the proof is complete.

Theorem 2.12: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is closed under Kleene $+$.

Proof: Let L be in $\boldsymbol{L}(k)$ and let c be a new symbol. By Lemma 2.11, $(Lc)^+$ is in $\boldsymbol{L}(k)$. Let h be a homomorphism defined by $h(c) = \varepsilon$ and $h(a) = a$ for each a in the alphabet of L. Clearly h is a symbol limited erasing on $(Lc)^+$. By Lemma 2.8, $h((Lc)^+)$ is in $\boldsymbol{L}(k)$. Since $h((Lc)^+) = L^+$, the proof is complete.

We are now ready to state the main result of this paper.

Theorem 2.13: Each family $\boldsymbol{L}(k)$, $k \geqslant 2$, is an AFL, i.e., it is closed under ε-free homomorphism, inverse homomorphism, intersection with regular sets, union, Kleene $+$, and concatenation.

Proof: The proof follows from Theorems 2.1, 2.2, 2.10, 2.12, Corollary 2.5, and the fact [1] that each family closed under the first five operations is closed under concatenation as well.

REFERENCES

[1] Ginsburg, S.: Algebraic and automata-theoretic properties of formal languages, North-Holland, Amsterdam, 1975.

[2] Hopcroft, J.—Ullman, J.: Formal languages and their relation to automata, Addison-Wesley, Reading, 1969.

[3] Rovan, B.: A framework for studying grammars, in: Lecture Notes in Computer Science 118, Springer Verlag, 1981, pp. 473—482.

Acknowledgement: This research was supported in part by the grant SPZV I-5-7/7.

Author's address:
Doc. RNDr. Branislav Rovan, CSc.,
Katedra teoretickej kybernetiky
Matematicko-fyzikálnej fakulty Univerzity Komenského,
Mlynská dolina,
842 15 Bratislava,
Czechoslovakia

Received: 8. 2. 1984

SÚHRN

TRIEDY ZLOŽITOSTI g-SYSTÉMOV SÚ AFL

V práci sa skúmajú triedy zložitosti pre mieru $\mathrm{STATE}_{\mathcal{N}}$ (počet stavov a-prekladača v g-systéme). Dokazuje sa, že sú to abstraktné triedy jazykov, t. j., triedy jazykov uzavreté na nevymazávajúci homomorfizmus, inverzný homomorfizmus, prienik s regulárnymi množinami, zjednotenie, zreťazenie a iteráciu.

РЕЗЮМЕ

КЛАССЫ СЛОЖНОСТИ g-СИСТЕМ ЯВЛЯЮТСЯ АФЛ

Бранислав Рован, Братислава

В работе доказывается, что классы сложности меры $\mathrm{STATE}_{\mathcal{N}}$ (число состояний конечного преобразователя с допускающими состояниями) являются абстрактными семействами языков, т. е. они замкнуты относительно операций несокращающего гомоморфизма, обратного гомоморфизма, пересечения с регулярными множествами, соединения, сцепления и итерации.