JOURNAL OF COMBINATORIAL THEORY 4, 222-239 (1968)

Complexity of Grammars by Group Theoretic Methods

JOHN RHODES* AND ELIAHU SHAMIR†

Department of Mathematics, University of California, Berkeley, California 94720

Communicated by Gian-Carlo Rota

ABSTRACT

Let $\mathscr{G}$ be a context free (phrase) structure grammar generating the context free language L. The set $P = P(\mathscr{G})$ of all "generation histories" of words in L can be coded as words in some augmented alphabet. It is proved here that $P = R \cap G$ where R is a regular (finite automaton definable) set and G is a "free group kernel" or Dyck set, a result first proved by Chomsky and Schützenberger [3].

We can construct the *lower central series* of the free group kernel

$$G_1 \supset G_2 \supset \cdots \supset G_n \supset \ldots,$$

so $\cap\, G_n = G$. Let $P_n = R \cap G_n$, so $\cap\, P_n = P$. P_n is the n-th order approximation of P. P_n need not be a context free language but it can be computed by n cascade or sequential banks of counters (integers). We give two equivalent characterizations of P_n, one "grammatical" and one "statistical," which follow from the theorems of Magnus, Witt, M. Hall, etc. for free groups. The main new theoretical tool used here for the study of grammars is the Magnus transform on the free group, $a \to 1 + a$, $a^{-1} \to 1 - a + a^2 - a^3 + a^4 \ldots$, which acts like a non-commutative Fourier transform.

1. INTRODUCTION

Let V be a finite set. $\sum V$ is the free non-commutative semigroup without identity generated by V. A context free (phrase) structure grammar $\mathscr{G}$ over V is given by a finite set of rewriting rules

$$\xi_i \to a_i, \qquad \xi_i \in V, \qquad a_i \in \textstyle\sum V, \qquad 1 \leqslant i \leqslant n. \tag{1.1}$$

For $x, y \in \sum V$, $x \Rightarrow y$ signifies that

$$x = u\xi_i v, \qquad y = ua_i v \qquad \text{and} \quad \xi_i \to a_i \quad \text{for some} \quad 1 \leqslant i \leqslant n. \tag{1.2}$$

* This research was sponsored in part by the Office of Naval Research, Information Systems Branch, Contract Nonr-4138(00), and the United States Air Force Office of Scientific Research, Contracts AF-AFOSR-848-65 and AF-49(638)-1550.

† This research was sponsored in part by the Office of Naval Research, Contract Nonr-222(60).

We write

$$x \Rightarrow \Rightarrow z \qquad \text{if} \quad x = y_0 \Rightarrow y_1 \Rightarrow \cdots \Rightarrow y_r = z, \qquad r \geqslant 0 \quad (1.3)$$

for suitable $y_1, ..., y_r$ in $\sum V$. By taking $r = 0$, we see that $x \Rightarrow x$ is always true. The set $V_1 = \{\xi_i : 1 \leqslant i \leqslant n\}$ is the set of *auxiliaries* and $V_2 = V - V_1$ is the set of *terminals*.

The set of words

$$L_i(\mathscr{G}) = \left\{x \in \sum V_2 : \xi_i \Rightarrow \Rightarrow x\right\} \qquad (1.4)$$

is called the language generated by $\mathscr{G}$ with ξ_i as initial symbol; and, abstractly, L is called a context free (CF) language iff it can be obtained as $L_i(\mathscr{G})$ for some $\mathscr{G}$. An algorithm (a decision procedure) for the relation $x \in L_i(\mathscr{G})$ exists, but the actual computation is quite tedious (in particular it involves finding all the generation histories of x). In fact the very definition (1.4) makes it hopeless to give a direct characterization of the language $L_i(\mathscr{G})$ in an "easily computable" form.

One can pose the same general questions, not for $L_i(\mathscr{G})$, but for the set of all "generation histories" of words in $L_i(\mathscr{G})$. These generations can be coded as linear words in some augmented alphabet—and these codes will be called *phrase-markers.* We can indeed characterize the collection of all phrase-markers, $P(\mathscr{G})$, as the intersection of a regular (finite automaton definable) set and a "free group kernel" (or Dyck set). This result (which under some slight restriction on the form of $\mathscr{G}$, was first proved by Chomsky and Schützenberger) is proved in Section 2. Our straightforward proof is obtained by separating the canonical role of bracketing engendered in all phrase structures (the free group kernel) from the particular form of the rewriting rule given by the specified grammar (the regular set).

Next we use the representation of the set $P = P(\mathscr{G})$ of phrase-markers above to construct approximations P_n of P which are easier to compute. Let $P = R \cap G$ where R is a regular set and G is a free group kernel (or Dyck set). Using some classical group theoretic results (which for the convenience of the reader are reviewed in Section 3 and the appendix of this paper) we can construct the *lower central series* of the free group kernel

$$G_1 \supset G_2 \supset \cdots \supset G_n \supset \cdots$$

so that $\cap \{G_k : k = 1, 2, 3, ...\} = G$. And thus

$$P_1 \supset P_2 \supset \cdots \supset P_n \supset \cdots, \qquad \cap \{P_k : k = 1, 2, 3, ...\} = P$$

where $P_n = R \cap G_n$. P_n is the nth order approximation of P.

P_n need not be a context free (phrase structure) language, but this is of little importance. However, P_n can be computed by n cascade or sequential banks of counters (the integers). Hence, it is "easy to compute P_n."

Further, we can give the following two equivalent characterizations of P_n: (a) We call the elements in the free group kernel (i.e., those elements which represent the identity) "grammatical." The standard collection process of free group elements [4, Chapter 11] can be likened to a human or a machine reading a string and trying to compute its grammatical status. Then P_n consists of those strings in R whose grammatical construction is correct up to depth n; that is, when they are unraveled up nth commutators, they agree with some string in P (or equivalently in G). By the same token, the set R (which is regular and determined by local conditions only) can be considered as the set of strings whose "spelling" is correct. Thus $x \in P_n$ if and only if the spelling is correct and the grammatical constructions are correct up to depth n. (b) P_n consists of those strings in R which agree with a string in P (or equivalently in G) up to "nth order statistical information." For two positive strings (meaning no inverses occur) t and r, we say that t and r agree up to "nth order statistical information" iff the number of times s occurs as a subsequence of t and r is the same for all sequences s of length $\leqslant n$. This concept is generalized to arbitrary strings via the Magnus transform.

The main new theoretical tool used here for the study of grammars is the Magnus transform, $a \to 1 + a$, $a^{-1} \to 1 - a + a^2 - a^3 + a^4 - \cdots$, which is essentially a non-commutative Fourier transform. In changing the strings of $\sum A$ into statistical coordinates via this transform, we find sequential computations with counters easy and natural to perform. Further, by (a) and (b) above, we see that these coordinates have a natural grammatical interpretation since (by well-known theorems) these coordinates are simply related to nth order commutators and "nth order statistical information."

It seems remarkable that, from only the context free phrase structure grammar assumption, we can show that humans and machines which employ structural approximations to those grammars actually use language on a statistical basis.

The results contained herein were formulated by the authors during a seminar entitled "Mathematical Theory of Languages" given at the University of California, Spring 1964, by Professor E. Spanier and the authors. We thank Professor Spanier and Mr. B. F. Wells for stimulating conversations and lectures. The first named authors also thanks Dr. S. Ginsburg, whose publications and lectures aroused his interest in languages.

We thank Mr. Y. Zalcstein and Mr. W. Leffler for their generous help in preparing this manuscript.

2. Recognition of Phrase-Markers

Let $\mathscr{G}$ be a context free grammar with vocabulary V and rewriting rules

$$\begin{gathered}\xi_p \to a_p = a_{p1} \cdots a_{p\pi(p)}, \qquad 1 \leqslant p \leqslant m, \quad \xi_p \in V, \\ a_p \in \sum V, \quad a_{pj} \in V.\end{gathered} \tag{2.1}$$

The set $\{\xi_p : 1 \leqslant p \leqslant m\}$ is the set of auxiliary symbols of $\mathscr{G}$.

Consider a set $V^{\#}$ which we take to be disjoint from V with

$$V^{\#} = \{A_{pq} : 1 \leqslant p \leqslant m,\ 0 \leqslant q \leqslant \pi(p)\}.$$

Define $\varphi : V^{\#} \to V$ by $\varphi(A_{pq}) = a_{pq}$ for $q > 0$ and $\varphi(A_{p0}) = \xi_p$. Thus we have made the A_{pq} formally different even they denote the same element of V (under φ). We extend φ to a length preserving homomorphism $\varphi : \sum V^{\#} \to \sum V$ by $\varphi(x_1, ..., x_n) = (\varphi(x_1), ..., \varphi(x_n))$. We define an equivalence relation $\sim$ on $V^{\#}$ by $x, y \in V^{\#}$, $x \sim y$ iff $\varphi(x) = \varphi(y)$.

We also consider a new grammar $\mathscr{G}^*$ with vocabulary

$$V^* = V^{\#} \cup \{L^j_{pq}, R^j_{pq} : 1 \leqslant p \leqslant m \quad \text{and} \quad A_{j0} \sim A_{pq}\}$$

and rewriting rules

$$A_{j0} \to L^j_{pq} A_{j1} \cdots A_{j\pi(j)} R^j_{pq} \qquad \text{for} \qquad (p, q, j) \tag{2.2a}$$

satisfying $A_{j0} \sim A_{pq}$,

$$A_{pq} \to A_{j0} \qquad \text{if} \qquad A_{pq} \sim A_{j0}. \tag{2.2b}$$

We say that (L^j_{pq}, R^j_{pq}) is a matching pair of $L - R$ symbols when (p, q, j) satisfies $A_{j0} \sim A_{pq}$.

If $A_{pq} \sim A_{j0}$, then the triple

$$D = (x, j, y) = (uA_{pq}v, j, uA_{j1} \cdots A_{j\pi(j)}v)$$

with $x, y \in \sum V^{\#}$ is called a *direct generation* of $\mathscr{G}$ with left component x and right component y. A sequence $D_1, ..., D_n$ of direct generations is *connected* iff

$$D_i = (x_i, j_i, y_i), \quad 1 \leqslant i \leqslant n, \quad \varphi(y_i) = \varphi(x_{i+1}) \quad \text{for } 1 \leqslant i \leqslant n - 1.$$

A *generation* of $\mathscr{G}$ is a connected sequence $D_1, ..., D_n$ of direct generations. A *j-derivation* of $y \in \sum V^{\#}$ in $\mathscr{G}$ is a generation $D_1, ..., D_n$ of $\mathscr{G}$ such that the left component of D_1 is $A_{j1} \cdots A_{j\pi(j)}$ and the right component of D_n is y.

Clearly, if there exists a j-derivations of y in $\mathscr{G}$, then $\varphi(A_{j0}) \Rightarrow \Rightarrow \varphi(y)$. Conversely, if $s \Rightarrow \Rightarrow w$ in $\mathscr{G}$, $s \neq w$, $s = \varphi(A_{j0})$, then for some $y \in \sum V^{\#}$, $w = \varphi(y)$ and there exists a j-derivation of y in $\mathscr{G}$.

A direct generation in $\mathscr{G}^*$ is a triple

$$E = (x, j, y) = (u^* A_{pq} v^*, j, u^* L^j_{pq} A_{j1} \cdots A_{j\pi(j)} R^j_{pq} v^*),$$

where $A_{pq} \sim A_{j0}$. The definitions of generation and j-derivation in $\mathscr{G}^*$ are exactly the same as those for $\mathscr{G}$ except that $\mathscr{G}$ is replaced by $\mathscr{G}^*$ and $V^{\#}$ by V^*.

We define a homomorphism $h : \sum V^* \to (\sum V^{\#})^1$ (where for an arbitrary semigroup S, S^1 is obtained from S by adjoining a new symbol 1 and letting it operate as the identity if S has none, and letting $S^1 = S$ otherwise)

$$h(A_{pq}) = A_{pq}, \; h(L^j_{pq}) = h(R^j_{pq}) = 1.$$

For $x = (x_1, ..., x_n) \in \sum X$, let $| x | = n$ (the length of x). Also let $| 1 | = 0$.

Lemma 2.1. *Let $D_1, ..., D_n$ be a j-generation of $\mathscr{G}$. Then there exists a unique generation $E_1, ..., E_n$ of $\mathscr{G}^*$ such that for $1 \leqslant i \leqslant n$ we have:*

$$\text{if} \quad D_i = (x_i, j_i, y_i), \quad \text{then} \quad E_i = (x_i^*, j_i, y_i^*), \tag{2.3a}$$

where $x_i = h(x_i^*)$ and $y_i = h(y_i^*)$. (Note also that $x_{i+1} = y_i$ and $x^*_{i+1} = y^*_i$ for $1 \leqslant i \leqslant n - 1$.) Further, if $\pi(p) \geqslant 2$ for $1 \leqslant p \leqslant n$, i.e., all the rewriting rules of (2.1) *are strictly length increasing, then*

$$| x_i^* | \leqslant 2 | x_i |, \quad | y_i^* | \leqslant 2 | y_i |. \tag{2.3b}$$

Proof: We construct $E_1, ..., E_n$ inductively. Given $A_{j1} \cdots A_{j\pi(j)} = uA_{jq}v$ and $D_1 = (A_{j1} \cdots A_{j\pi(j)}, k, uA_{k1} \cdots A_{k\pi(k)} v)$ let

$$E_1 = (A_{j1} \cdots A_{j\pi(j)}, k, uL^k_{jq} A_{k1} \cdots A_{k\pi(k)} R^k_{jq} v).$$

Assume that $E_1, ..., E_i$ have been defined and satisfy (2.3a), and let

$$D_i = (..., ..., uA_{st}v), \qquad D_{i+1} = (uA_{st}v, l, uA_{l1} \cdots A_{l\pi(l)} v)$$

and

$$E_i = (..., ..., u^* A_{st} v^*),$$

where by induction $u = h(u^*)$, $v = h(v^*)$. Let

$$E_{i+1} = (u^* A_{st} v^*, l, u^* L^l_{st} A_{l1} \cdots A_{l\pi(l)} R^l_{st} v^*).$$

Then clearly E_{i+1} satisfies (2.3a). This proves (2.3a).

Now (2.3b) is immediate for $i = 1$ and by the above construction the right component of D_{i+1} has length

$$(|u| + |v| + 1) + (\pi(l) - 1)$$

while the right component of E_{i+1} has length

$$(|u^*| + |v^*| + 1) + (\pi(l) + 1).$$

Now by induction hypothesis,

$$|u^*| + |v^*| + 1 \leqslant 2(|u| + |v| + 1).$$

Hence

$$(|u^*| + |v^*| + 1) + (\pi(l) + 1) \leqslant 2[(|u| + |v| + 1) + \pi(l) - 1],$$

and so (2.3b) is proved.

The sequence $E_1, ..., E_n$ is thus a code for the generation $D_1, ..., D_n$. However, the right component of E_n contains essentially all the information about the generation, except for its order. More precisely, an interchange of two consecutive generations D_i, D_{i+1} will not change the right component of E_n when D_i and D_{i+1} only rewrite two distinct symbols of the right component of D_{i-1}.

We call the right component of E_n the *phrase-marker* of $D_1, ..., D_n$ and define two generations to be *equivalent* iff they give rise to the same phrase-marker. (If one represents the generation in "tree form," then two generations are equivalent iff they have the same tree.) We also include $A_{p1} \cdots A_{p\pi(p)}$ among the phrase-markers of the j-derivations of $\mathscr{G}$.

We now define subsets K_1, K_2 of

$$\sum \{A_{pq}, L^j_{pq}, R^j_{pq} : 1 \leqslant p, j \leqslant m, 0 \leqslant q \leqslant \pi(p), A_{j0} \sim A_{pq}\}$$

as follows: K_1 is the smallest set containing

$$\sum \{A_{pq} : 1 \leqslant p \leqslant m, 0 \leqslant q \leqslant \pi(p)\}$$

such that if $x, y \in K_1$, then $xy \in K_1$ and $L^j_{pq} x R^j_{pq} \in K_1$. K_1 is called the *free half-group kernel.*

K_2 is the smallest set containing $\sum \{A_{pq} : 1 \leqslant p \leqslant m, 0 \leqslant q \leqslant \pi(p)\}$ such that if $x, y \in K_2$ then $xy \in K_2$, $L^j_{pq} x R^j_{pq} \in K_2$, and $R^j_{pq} x L^j_{pq} \in K_2$. K_2 is called the *free group kernel.*

THEOREM 1. *There is a regular set F_j such that the set of all phrase-markers of j-derivations of $\mathscr{G}$ is given by $K_1 \cap F_j$. Moreover,*

$$K_1 \cap F_j = K_2 \cap F_j.$$

PROOF: We construct a finite automaton $\mathscr{M}$ which will defined (or accept) all the sets F_j by suitable choice of initial and final states.

The alphabet of $\mathscr{M}$ is

$$X = \{A_{pq}, L^j_{pq}, R^j_{pq} : 1 \leqslant p \leqslant m, 0 \leqslant q \leqslant \pi(p), A_{j0} \sim A_{pq}\}.$$

The set of states is

$$Q = \{\alpha_{pq} : 1 \leqslant p \leqslant m, \quad 0 \leqslant q \leqslant \pi(p)\} \cup \{\beta^j : 1 \leqslant j \leqslant m\} \cup \{\gamma\}.$$

The next state function $M : Q \times Y \to Q$ is defined by the following transition rules

$$M(\alpha_{pq}, L^k_{pq}) = \alpha_{k1} \text{ (an } L\text{-transition)} \tag{2.4a}$$

$$M(\beta^k, R^k_{pq}) = \begin{cases} \alpha_{p,q+1} & \text{if } q \neq \pi(p) \\ \beta^p & \text{if } q = \pi(p) \end{cases} \text{ (an } R \text{ transition)} \tag{2.4b-b'}$$

$$M(\alpha_{pq}, A_{pq}) = \begin{cases} \alpha_{p,q+1} & \text{if } q \neq \pi(p) \\ \beta^p & \text{if } q = \pi(p) \end{cases} \text{ (an } A \text{ transition)} \tag{2.4c-c'}$$

$$M(s, x) = \gamma \text{ in all other cases (an absorbing transition)} \tag{2.4d}$$

As usual we extend M to $Q \times (\sum X)^1$ by

$$M(s, 1) = s \quad \text{and} \quad M(s, xy) = M(M(s, x), y).$$

The proof of Theorem 1 proceeds via Lemmas 2.2–2.4.

LEMMA 2.2. *Let z be a phrase-marker of a j-derivation of $\mathscr{G}$. Then $z \in K_1$ and $M(\alpha_{j1}, z) = \beta^j$.*

LEMMA 2.3. *Let $z \in K_1$, and let $M(\alpha_{j1}, z) = \beta^j$. Then z is a phrase-marker of a j-derivation of $\mathscr{G}$.*

PROOF OF LEMMA 2.2: By (2.3a) there is an associated j-derivation of $\mathscr{G}^*$, $E_1, \ldots, E_n$ such that z is the right component of E_n. Let

$$E_n = (uA_{pq}v, k, uL^k_{pq}A_{k1} \cdots A_{k\pi(k)}R^k_{pq}v = z).$$

We proceed by induction on n. Since the (inductive) proof that $z \in K_1$ is trivial, and since $M(\alpha_{j1}, A_{j1} \cdots A_{j\pi(j)}) = \beta^j$ is immediately verified, it suffices to show that

$$M(\alpha_{j1}, uA_{pq}v) = \beta^j \quad \text{implies} \quad M(\alpha_{j1}, z) = \beta^j. \tag{2.5}$$

Now $M(\alpha_{j1}, uA_{pq}) = M(M(\alpha_{j1}, u), A_{pq}) \neq \gamma$ so by (2.4) we see that this state was obtained by an A-transition.

Thus

$$M(\alpha_{j1}, u) = \alpha_{pq} \quad \text{and} \quad M(\alpha_{j1}, uA_{pq}) = \alpha_{p,q+1}$$
$$(\text{or } \beta^p \text{ if } q = \pi(p)); \tag{2.6}$$

continuing along the initial substrings of z we have

$$\begin{aligned} M(\alpha_{j1}, uL_{pq}^k A_{k1} \cdots A_{k\pi(k)} R_{pq}^k) &= M(\alpha_{pq}, L_{pq}^k A_{k1} \cdots A_{k\pi(k)} R_{pq}^k) \\ &= M(\beta^k, R_{pq}^k) \\ &= \alpha_{p,q+1} \, (\text{or } \beta^p \text{ if } q = \pi(p)) \end{aligned} \tag{2.7}$$

Now from (2.6) and (2.7) and the hypothesis of (2.5):

$$M(\alpha_{j1}, z) = M(\alpha_{p,q+1}, v) = M(\alpha_{j1}, uA_{pq}v) = \beta^j.$$

This proves Lemma 2.2.

PROOF OF LEMMA 2.3: Since $z \in K_1$, we have either

$$z = x \in \sum \{A_{pq} : 1 \leqslant p \leqslant m, \ 0 \leqslant q \leqslant \pi(p)\}$$

or

$$\begin{aligned} &z = uL_{pq}^k x R_{pq}^k v \quad \text{for some} \quad u, v \in K_1 \quad \text{and} \\ &x \in \sum \{A_{pq} : 1 \leqslant p \leqslant m, 0 \leqslant q \leqslant \pi(p)\} \end{aligned} \tag{2.8}$$

and also $uA_{pq}v \in K_1$. We claim that in this case

$$M(\alpha_{j1}, uA_{pq}v) = \beta^j = M(\alpha_{j1}, z). \tag{2.9}$$

Indeed

$$M(\alpha_{j1}, uL_{pq}^k) = M(M(\alpha_{j1}, u), L_{pq}^k) \neq \gamma. \tag{2.10}$$

So the left transitions was an L-transition and

$$M(\alpha_{j1}, u) = \alpha_{pq}, \tag{2.11}$$

$$M(\alpha_{j1}, uL_{pq}^k) = \alpha_{k1}. \tag{2.12}$$

Again, since we do not encounter any absorbing transitions,

$$M(\alpha_{j1}, uL_{pq}^k x R_{pq}^k) = \alpha_{p,q+1} \quad (\text{or} \quad \beta^p \quad \text{if} \quad q = \pi(p)), \tag{2.13}$$

$$M(\alpha_{j1}, uL_{pq}^k x) = \beta^k. \tag{2.14}$$

From (2.11) we obtain

$$M(\alpha_{j1}, uA_{pq}) = \alpha_{p,q+1} \qquad (\text{or } \beta^p \text{ if } q = \pi(p)).$$

Comparing this with (2.13), we clearly obtain (2.19). Now by comparing (2.12) with (2.14), we have

$$M(\alpha_{k1}, x) = \beta^k. \tag{2.15}$$

Since $x \in \sum \{A_{pq}\}$, the only transitions in (2.15) are of type A, and (2.15) is possible only if

$$x = A_{k1} \cdots A_{k\pi(k)} . \tag{2.16}$$

In particular if $x = z$ (then (2.15) holds with $k = j$), we find z is the phrase-marker of the trivial j-derivation. Otherwise we proceed by induction (on the length of z). By induction hypothesis and (2.9), we have $uA_{pq}v$ the phrase-marker of $E_1, \ldots, E_r$. Consequently by (2.16), z itself will be the phrase-marker of $E_1, \ldots, E_{r+1}$ where

$$E_{r+1} = (uA_{pq}v, k, uL_{pq}^k A_{k1} \cdots A_{k\pi(k)} R_{pq}^k v).$$

This proves Lemma 2.3.

Lemmas 2.2 and 2.3 clearly establish Theorem 1, except for the last assertion, which will follow from the next lemma.

LEMMA 2.4. *Let* $x \in \sum \{A_{pq} : 1 \leqslant p \leqslant m,\ 0 \leqslant q \leqslant \pi(p)\}$ *and let* $M(s, R_{pq}^k x L_{pq'}^k) = s'$ *with* $s, s' \neq \gamma$. *Then* $q < q'$.

PROOF: Indeed by (2.4) s must be β^k and $M(s, R_{pq}^k) = \alpha_{p,q+1}$, cannot be β^p since the next transition cannot be of type R; indeed if $x \neq 1$ the next transition must be of type A and any additional transition increases the second index of the state until finally (or immediately if $x = 1$) we end with an L transition with $q' > q$. This proves Lemma 2.4.

Now we can prove that $K_1 \cap F_j = K_2 \cap F_j$. If not, there is a $z \in K_2 \cap F_j$ with $z \notin K_1$. Then every substring of z (we say x is a substring of y iff $y = uxv$ for some u, v) effects a non-absorbing transition of the automaton (starting from some state) and there is a substring w of the form

$$w = R_{oq_0} x_1 L_{oq_1} y_1 R_{oq_1} x_2 \cdots x_{s-1} L_{oq_s} y_s R_{oq_s} x_s L_{oq_0}$$

where L_{oq_j} and R_{oq_j} $(0 \leqslant j \leqslant s)$ match (we have omitted the upper index and the first lower index), and where $x_j \in \sum \{A_{pq}\}$ (and $y_j \in K_1$, but we will not use this fact). But then, by Lemma 2.4, $q_0 < q_1 < \cdots < q_s < q_0$ (in case $s = 0$, we have $q_0 < q_0$ too) and this is a contradiction.

This proves Theorem 1.

COROLLARY 1. *If* $L = L(\mathscr{G})$ *is the language generated by the auxiliary* *symbol* $\xi = \varphi(A_{j0})$, *then there is a regular set F and a non-length increasing homomorphism* $\psi : \sum V^* \to (\sum V)^1$ *such that*

$$L = \psi(K_1 \cap F) = \psi(K_2 \cap F).$$

PROOF: Let $F_j' = \{z \in F_j : z$ contains no A_{p0} and no A_{pq} with $A_{pq} \sim A_{j0}$, $1 \leqslant j, p \leqslant m\}$. Then $F_j' \cap K_1$ is the set of all terminated j-derivations of $\mathcal{G}$. Let

$$F = \cup\{F_j' : \varphi(A_{j0}) = \xi\}.$$

Clearly F is regular since F_j is regular. Let $\psi(A_{pq}) = \varphi(A_{pq})$ for $q \geqslant 1$ and let $\psi(L_{pq}^j) = \psi(R_{pq}^j) = \psi(A_{p0}) = 1$. That is, $\psi = \varphi h$. Corollary 1 now follows from Theorem 1.

REMARK 2.1. (a) If $\mathcal{G}$ has only strictly length increasing rules (2.1) (i.e., $\pi(p) \geqslant 2$ in all the rules), then

$$K_1 \cap F = K_1 \cap F \cap E = K_2 \cap F \cap E$$

where E is the regular set $\{t : |t| \leqslant 2|h(t)|\}$. The proof is immediate by Lemma 2.1.

(b) We note in passing that neither of the restrictions that we have put on the grammar, namely that $1 \leqslant \pi(p)$ (i.e., no rule $\xi_i \to 1$) and that $2 \leqslant \pi(p)$ (i.e., that all rules are strictly length increasing) involves a loss of generality, as has been shown in [1].

(c) The finite semigroup of the reduced automaton defining F of Corollary 1 has no subgroups of order greater than one.

3. APPROXIMATION OF FREE GROUP AND FREE HALF-GROUP KERNELS

Let A be a set. $(\sum A)^1$ is the free semigroup with identity generated by A; i.e., $(\sum A)^1 = \{(a_1, ..., a_n) : n \geqslant 0$ and $a_i \in A$ for $i = 1, ..., n\}$ with concatenation serving as the multiplication in $(\sum A)^1$.

Let $\bar{A} = \{\bar{a} : a \in A\}$ where $A \cap \bar{A}$ is empty.

Let $W_n(A) = \{t \in (\sum A)^1 : |t| \leqslant n\}$ for $n = 1, 2, ..., \omega$, and $W_\omega = (\sum A)^1$.

Let Z be the ring of integers.

Let $R_n(A)$, $n = 1, 2, ..., \omega$ be the set of integer valued functions on $W_n(A)$.

For $f \in R_n(A)$, $t \in (\sum A)^1$ we denote $f(t)$ by $\langle f, t\rangle$. Then f can be represented as $\sum \{\langle f, t\rangle t : t \in (\sum A)^1\}$ and we will identify f with its formal expansion. R_n can be given a ring structure by defining

$$\langle f + g, t\rangle = \langle f, t\rangle + \langle g, t\rangle$$
$$\langle fg, t\rangle = \sum \{\langle f, t_1\rangle\langle g, t_2\rangle : t_1 t_2 = t\}$$

$R_\omega(A)$ is thus the ring of formal power series in non-commuting variates $a \in A$ and with coefficients in Z.

There is a canonical ring homomorphism $P_n : R_\omega(A) \to R_n(A)$ given by $P_n(f) = f$ restricted to $W_n(A)$.

Let $H_n(A) = \{f \in R_n(A) : \langle f, 1\rangle = 1\}$ $n = 1, 2,..., \omega$. It is easy to check that $H_n(A)$ is a group under multiplication. In fact,

$$(1 + f)^{-1} = 1 - f + f^2 - \cdots + (-1)^n f^n$$

when $\langle f, 1\rangle = 0$. We denote $H_\omega(A)$ by $H(A)$.

For any group G, let

$$G_1 = G,\ G_2 = \{\{(g_1, g_2) : g_1, g_2 \in G\}\},..., G_n = \{\{(g_1,..., g_n) : g_i \in G\}\}$$

where, for $S \subseteq G$, $\{S\}$ denotes the subgroup generated by S, and the higher order commutators are defined inductively by

$$(g_1, g_2) = g_1^{-1} g_2^{-1} g_1 g_2$$
$$(g_1,..., g_n) = ((g_1,..., g_{n-1}), g_n)$$

The series $G_1 \supset G_2 \supset \cdots$ is called the lower central series of G. Let $G(A)$ denote the free group with generators A. The following lemma is well known.

LEMMA 3.1. *Let $G = G(A)$ and $G_1 \supset G_2 \supset \cdots$ be its lower central series.*

(a) *The map $a \to 1 + a$*

$$a^{-1} \to 1 - a + a^2 - a^3 + \cdots = \sum \{(-1)^n a^n : n = 0, 1,...\}$$

extends to a 1-1 *homomorphism τ of $G(A)$ into $H(A)$.*

(b) *Kernel* $(P_n \circ \tau) = G_{n+1}$

(c) $\cap \{G_n : n = 1, 2,...\} = \{e\}$.

The map τ will be referred to as the Magnus transform.

(d) *Kernel* $(P_n \circ \tau) = \{g \in G(A)$: *the unique expansion of g, by the collection process, into basic commutators has no terms of weight* $\leqslant n\}$ *(see the Appendix). Notice, for $(a_1,..., a_n) \in \sum A \subset G(A)$,*

$$\tau(a_1,..., a_n) = (1 + a_2) \cdot ... \cdot (1 + a_n) = f,$$

where $\langle f, t\rangle$ is the number of times t occurs as a subsequence of $(a_1,..., a_n)$. Thus we can view the Magnus transform as the unique extension of the "statistical transform" τ on $\sum A$, defined above, to a homomorphism on $G(A)$.

PROOF: The reader can find all the proofs in Fox [7]. An alternative method is sketched below. The assertion of (c) is the Magnus theorem (see Kurosh [5, p. 38]). To prove (a) we need some properties of the collecting process (Hall [4, Chapter 11]). For the convenience of the reader these properties are summarized in the appendix.

Let $g \in G$, then g has a unique ordered expansion in terms of basic commutators (the expansion may have infinite length). Now (c) implies that, conversely, this expansion determines g uniquely. For, if g_1 and g_2 have identical expansions, then, for all n, $g_1 g_2^{-1} \in G_n$; thus

$$g_1 g_2^{-1} \in \cap \{G_n : n = 1, 2, \ldots\} = \{e\} \qquad \text{and} \qquad g_1 = g_2 .$$

Let $g_1 \neq g_2$. Then by the above, the expansions of g_1 and g_2 differ. Let w be the maximal initial subword common to both expansions. Then

$$g_1 \equiv w c_{i_1}^{k_1} \cdots c_{i_m}^{k_m} \bmod G_{n+1} ,$$

$$g_2 \equiv w c_{i_1}'^{k_1} \cdots c_{i_l}'^{k_l} \bmod G_{n+1} ,$$

where $c_{i_1} \neq c'_{i_1}$, c_i, c'_i basic commutators of weight w. Now, by lemma proved in the appendix, the leading term of $\tau(c_{i_1}) \neq$ leading term of $\tau(c'_{i_1})$. Hence $\tau(g_1) \neq \tau(g_2)$ and (a) is proved. Moreover, by the lemma quoted above, the leading term of $\tau(c_j)$ is a ring commutator of weight n and so $P_n \circ \tau(g_1) \neq P_n \circ \tau(g_2)$. Now, let $g \notin G_{n+1}$; then $g \not\equiv e \bmod G_{n+1}$. Thus $g \equiv x \bmod G_{n+1}$ where $x \neq e$ is a product of basic commutators of weight n and thus, as above, $P_n \circ \tau(g) \neq P_n \circ \tau(e) = 1$, i.e., $g \notin$ Kernel $(P_n \circ \tau)$. On the other hand, if $g \in G_{n+1}$, g is a product of basic commutators of weight $n + 1$ and their inverses. Hence the leading term of $\tau(g)$ has weight $\geqslant n + 1$, implying $P_n \circ \tau(g) = 1$ and $g \in$ Kernel $(P_n \circ \tau)$. Hence (b) is proved.

Finally, (d) follows from (b) and Theorem 11.-24 of Hall [4]. See the appendix.

Let T be a finite set. Let

$$\varphi : \left(\sum (T \cup A \cup \bar{A})\right)^1 \to G(A)$$

be the semigroup homomorphism given by $\varphi(a) = a$, $\varphi(\bar{a}) = a^{-1}$, $\varphi(t) = e$ for $a \in A$, $\bar{a} \in \bar{A}$, $t \in T$.

Let ${}^T K_G^A = \operatorname{Ker} \varphi = \{x \in (\sum (T \cup A \cup \bar{A}))^{-1} : \varphi(x) = e\}$. ${}^T K_G^A$ is called the *free group kernel*. Let $K_G^A = {}^T K_G^A$ where T is empty.

Let ${}^T K_S^A$ be the smallest set containing 1 such that $\alpha, \beta \in {}^T K_B^A$, $u \in T$, $a \in A$ implies $\alpha u \beta$, $\alpha\beta$, $a\alpha\bar{a} \in {}^T K_S^A$ (so $\alpha a \bar{a} \beta \in {}^T K_S^A$ by $1 \in {}^T K_S^A$). ${}^T K_S^A$ is called the *free half-group kernel*. Let $K_S^A = {}^T K_S^A$ where T is empty.

By virtue of Lemma 3.1 it is reasonable to consider the elements of G_n as approximate to the identity and thus to try to approximate ${}^T K_G^A$ by ${}^T K_{G_n}^A = \{t \in (\sum (T \cup A \cup \bar{A}))^1 : \varphi(t) \in G_n\}$.

4. An Approximation Theorem for Context Free Languages

To begin with, we observe that the sets K_1 and K_2 defined in Section 2 above are ${}^T K_S^A$ and ${}^T K_G^A$, respectively, where we let $T = \{A_{pq}\}$, $A = \{L_{pq}^j\}$, $\bar{A} = \{R_{pq}^j\}$.

Let $\mathscr{G}$ be a grammar with strictly length increasing rules (see Remark 2.1(b)). Let $L = L_1(\mathscr{G})$ and $P = P_1(\mathscr{G})$, the phrase-markers of 1-derivations of $\mathscr{G}$. Then by Theorem 1, Corollary 1, and Remark 2.1(a) we have

$$L = \psi({}^T K_G^A \cap F \cap E) \tag{4.1}$$
$$P = {}^T K_G^A \cap F \cap E$$

where ψ is defined in Corollary 1, F is a regular set, and

$$E = \left\{ t \in \sum (T \cup A \cup \bar{A}) : |t| \leqslant 2 \, |\psi(t)| \right\}.$$

Let $P_n = {}^T K_{G_n}^A \cap F \cap E$, and $L_n = \psi(P_n)$.

THEOREM 2. (a) $L_1 \supseteq \cdots \supseteq L_n \supseteq \cdots \supseteq \cap \{L_i : i = 1, 2, 3, \ldots\} = L$

(b) $P_1 \supseteq \cdots \supseteq P_n \supseteq \cdots \supseteq \cap \{P_i : i = 1, 2, 3, \ldots\} = P$

(c) *P_k has "grammatical depth" $\leqslant k$ in the sense that $t \in P_k$ iff $t \in F$, $t \in E$ and the unique expansion of $\varphi(t)$ in $G(A)$ by the collection process into basic commutators has no terms with weight $\leqslant k$.*

(d) *P_k is the k-th order "grammatical statistical approximation" in the sense that $t \in P_k$ iff $t \in F$, $t \in E$ and the Magnus transform of t, $\tau \circ \varphi(t)$, has no terms of n-th order for $n \leqslant k$; i.e., $\langle \tau \circ \varphi(t), r \rangle = 0$ for $1 \leqslant |r| \leqslant k$*

(e) *P_k is "easy to compute" in the sense that P_k is accepted by an automaton $\mathscr{M}_k$ which is the direct sum of a finite state machine and another automaton which is the sequential composition of a finite number of (infinite) counters (precise definitions given below).*

PROOF: The assertion of (b) follows from Lemma 3.1(c) and equation (4.1).

It is immediate that $\{L_n\}$ is decreasing and $L \subseteq \cap L_n$. Now let $y \in \cap L_n$. Then for each n there is an x_n satisfying

$$x_n \in {}^T K_{G_n}^A \cap F, \quad \psi(x_n) = y, \quad |x_n| \leqslant 2 \, |y|.$$

Only finitely many words satisfy the last conditions; so some $x = x_i$ occurs in infinitely many, and hence all, ${}^T K_{G_n}^A$. Thus $x \in \cap\, {}^T K_{G_n}^A = {}^T K_G^A$. Thus $x \in {}^T K_G^A \cap F \cap E$ and $y \in L$. This proves (a).

The assertions of (c) and (d) follow from Lemmas 3.1(d) and (b), respectively, and equation (4.1).

We now prove (e). We introduce the following notation. Let S be a semigroup. Then $S^f : \sum S \to S$ with $S^f(s_1, ..., s_n) = s_1 \cdot ... \cdot s_n$. Let $f : \sum A \to B$; such an f is called a machine, and we define $f^\sigma : \sum A \to \sum B$ by $f^\sigma(a_1, ..., a_n) = (f(a_1), f(a_1, a_2), ..., f(a_1, ..., a_n))$. In particular

$$S^{f\sigma}(s_1, ..., s_n) = (s_1, s_1 s_2, ..., s_1 ... s_n).$$

If $f : \sum A \to \{0, 1\}$ then $\mathscr{A}(f) = \{t \in \sum A : f(t) = 1\}$, the set of tapes accepted by f. If $h : A \to B$ then $h^\Gamma : \sum A \to \sum B$ with $h^\Gamma(a_1, ..., a_n) = (h(a_1), ..., h(a_n))$. Z denotes the additive group of integers. Z^f is called the (infinite) *counter*. Let $f_k : \sum A_k \to B_k$ for $k = 1, ..., n$. Then

$$f_1 \times \cdots \times f_n : \sum (A_1 \times \cdots \times A_n) \to B_1 \times \cdots \times B_n$$

(the direct sum of the f_i) with

$$f_1 \times \cdots \times f_n((a_{11}, ..., a_{1n}), ..., (a_{m1}, ..., a_{mn})) = \\ = (f_1(a_{11}, ..., a_{m1}), ..., f_n(a_{1n}, ..., a_{mn})).$$

We first notice that $E = \mathscr{A}(f)$ where $f : \sum (A \cup \bar{A} \cup T) \to \{0, 1\}$ with $f = h_2 Z^f h_1{}^\Gamma$ where $h_1 : A \cup \bar{A} \cup T \to Z$ with $h_1(a) = h_1(\bar{a}) = -1$ for $a \in A$, and $h_1(t) = 2$ for $t \in T$. Further, $h_2 : Z \to \{0, 1\}$ with $h_2(x) = 1$ when $x \geqslant 0$ and $h_2(x) = 0$ when $x < 0$.

We next show that ${}^T K_{G_n}^A$ are accepted by automata constructed from a finite number of counters. In fact, we can construct a representation of ${}^T K_G{}^A$ by infinite triangular matrices using the Magnus transform, and using Lemma 3.1(b) one could "abstract" to ${}^T K_{G_n}^A$. Then one could use the methods of Schützenberger [6]. However, we follow the methods of [8], using the fact that the right regular representation of $R_\omega(S_0)$ is in the wreath product of counters with respect to the natural coordinates. Another way to view the matter is that the construction of [8, Lemma 3.4] is effective for any group G and any subnormal series of G (e.g., $G(A)$ and its lower central series). We will now show that, for $r \in \sum (A \cup \bar{A} \cup T)$, $t \to \tau \cdot \varphi(r)$ can be computed by a "bank" of counters of depth n.

We use the following notation. We write

$$\alpha_n = (f_n, \varphi_n, ..., f_2, \varphi_2, f_1, \varphi_1),$$

where $f_k : \sum A_k \to B_k$ is a machine for $k = 1,..., n$ and $\varphi_1 : B_0 \to A_1$, $\varphi_2 : B_1 \times B_0 \to A_2 ,..., \varphi_{n-1} : B_{n-2} \times \cdots \times B_0 \to A_{n-1}$,

$$\varphi_n : B_{n-1} \times \cdots \times B_0 \to A_n .$$

Then α_n determines a machine $M(\alpha_n) : \sum B_0 \to B_1 \times \cdots \times B_n$ in a natural way. We consider φ_1 as mapping basic inputs in B_0 into basic inputs of f_1, and φ_k as mapping the last basic outputs of $f_{k-1} ,..., f_1$ and the original last basic input into a basic input of f_k. Precisely

$$\begin{aligned}&M(\alpha_0)^\sigma \text{ equals the identity map on } \textstyle\sum B_0 ,\\ &M(\alpha_1) = f_1\varphi_1{}^\Gamma,\\ &M(\alpha_k) = f_k\varphi_k{}^\Gamma(M(\alpha_{k-1}) \times \cdots \times M(\alpha_1) \times M(\alpha_0))^\sigma.\end{aligned}$$

Let $\mathcal{M}_n : \sum (A \cup \bar{A} \cup T) \to R_n(A)$ be defined by $\mathcal{M}_n(r) = (P_n \circ \tau \circ \varphi(r))$ (see Section 3 for the definitions).

Let $\Omega_n(\sum X) = \{r \in \sum X : | r | = n\}$ and

$$\Delta_n \left(\sum X\right) = \left\{r \in \sum X : 1 \leqslant | r | \leqslant n\right\}.$$

We now let

$$\alpha_n = (S_n{}^f, \varphi_n ,.... S_2{}^f, \varphi_2 , S_1{}^f, \varphi_1),$$

where $S_k = F(\Omega_k(\sum A), Z) = \{f : \Omega_k(\sum A) \to Z\}$ with pointwise multiplication $(f_1 + f_2)(x) = f_1(x) + f_2(x)(Z \times \cdots \times Z$, the free Abelian group of rank $| A |^k)$. Now $S_{k-1} \times \cdots \times S_1$ can be naturally identified with $T_{k-1} = F(\Delta_{k-1}(\sum A), Z)$ by $(f_{k-1} ,..., f_1) \leftrightarrow f$ where $f(r) = f_j(r)$ for $1 \leqslant | r | = j \leqslant k - 1$.

Let $S_0 = A \cup \bar{A} \cup T$. Next we define $\varphi_1 : S_0 \to S_1$ and for $k = 2,..., n$ $\varphi_k : ((S_{k-1} \times \cdots \times S_1) \times S_0) \cong T_{k-1} \times S_0 \to S_k$. For $2 \leqslant k \leqslant n$,

$$g_{k-1} \in T_{k-1} , a \in A, \bar{a} \in \bar{A}, t \in T, x \in \Omega_n(\textstyle\sum A)$$

we let

$$\varphi_k(g_{k-1} , a)(x) = ((1 + g_{k-1})a)(x) = g_{k-1}(y),$$

where $x = y \cdot a$ with the convention that $g_{k-1}(1) = 1$ and $g_{k-1}(y) = 0$ when y is not defined. Further,

$$\begin{aligned}\varphi_k(g_{k-1} , t)(x) &= 0,\\ \varphi_k(g_{k-1} , \bar{a})(x) &= ((1 + g_{k-1})(-a + a^2 - a^3 + \cdots\\ &\quad + (-1)^{k-1}a^{k-1} + (-1)^k a^k))(x)\\ &= -g_{k-1}(y_1) + g_{k-1}(y_2) - g_{k-1}(y_3) + \cdots\\ &\quad + (-1)^{k-1}g_{k-1}(y_{k-1}) + (-1)^k \delta(x, a^k),\end{aligned}$$

where $y_j a^j = x$ for $1 \leqslant j \leqslant k-1$ with $g_{k-1}(y^j) = 0$ when y_j is not defined and $\delta(x, a^k) = 1$ if $x = a^k$ and zero otherwise. Finally for $a, b \in A$, $t \in T$

$$\varphi_1(t)(b) = 0,$$
$$\varphi_1(a)(b) = \delta(a, b),$$
$$\varphi_1(\bar{a})(b) = -\delta(a, b).$$

Thus we have defined α_n. Now it can be verified by induction that

$$\mathscr{M}_n = M(\alpha_n)$$

when 1 is struck from the domain of $M(\alpha_n)(r)$. Now let $\beta_n(f_n) = 1$ if $f_n(x) = 0$ all x, and $\beta_n(f_n) = 0$ otherwise. Then by Lemma 3.1(b)

$$\mathscr{A}(\beta_n M(\alpha_n)) = \mathscr{A}(\beta_n(\mathscr{M})_n) = {}^T K^A_{G_{n+1}}.$$

This proves (e) and hence Theorem 2.

APPENDIX

The Collecting Process (see Hall [4, Chapter 11])

Let $G = (A)$, $A = \{a_1, ..., a_n\}$. We define *commutators* c_i and weight $w(c_i)$ as follows:

(1) The a_i's are commutators of weight 1.

(2) If c_i and c_j are commutators, then (c_i, c_j) is a commutator of weight $w(c_i) + w(c_j)$.

The commutators are ordered by weight and commutators of the same weight are ordered in an arbitrary but fixed manner. The subclass of *basic commutators* is defined as follows:

(1) The a_i's are basic commutators of weight 1.

(2) Having defined basic commutators of weight $< n$, we define the basic commutators of length n as (c_i, c_j) where c_i and c_j are basic, $w(c_i) + w(c_j) = n$ and

(a) $c_i > c_j$,

(b) if $c_i = (c_k, c_l)$, then $c_j \geqslant c_l$.

The collecting process consists of successive replacements of a subword of the form $c_i c_j$ by $c_j c_i (c_i, c_j)$. Note that this does not change the group element represented by the word.

Every element of G has a *unique* ordered expansion of the form

$$g = c_1^{k_1} \cdots c_m^{k_m} \bmod G_n ,$$

where the $c_i{}^s$ are basic commutators of weight $< n$ [4, Theorem 11.2.4].

We define the bracketing operation $[\cdot, \cdot]$ in $R_\omega(A)$, the ring of power series, by $[u, v] = uv - uv$. Using it we can define ring commutators in $R_\omega(A)$ by letting the $a_i \in A$ be commutators of weight 1 and continue as in the group case. For $f \in H_n(A)$, we define the leading term of f as the formal sum of all monomials of smallest positive degree. By Corollary 11.2.1 of Hall [4], the basic commutators of degree m are linearly independent.

Let $\tau : G \to R_\omega(A)$ be the Magnus transform.

LEMMA [4, pp. 173–174]. (a) *For $f, g \in H_n(A)$, $f, g \neq 1$, with leading terms f_k, g_l of degrees k, l, respectively; the leading term of f^{-1} is $-f_k$ and the leading term of*

$$f_k \cdot g_l = \begin{cases} f_k & \text{if} \quad k < l \\ g_l & \text{if} \quad l < k \\ f_k + g_l & \text{if} \quad l = k \quad \text{and} \quad f_k + g_l \neq 0. \end{cases}$$

If $[f_k, g_l] \neq 0$, it is the leading term of (f, g).

(b) *The leading term of $\tau(c_i)$ is c_i' where c_i is a basic (group) commutator and c_i' is the ring commutator obtained by replacing round by square brackets in the expression for c_i.*

PROOF: The assertions about f^{-1} and $f \cdot g$ follow by straightforward computations. Now if $f = 1 + f', g = 1 + g', f^{-1} = 1 + f'', g^{-1} = 1 + g''$, then

$$(f, g) = 1 + f'g' - g'f' + f'f''g' - g'g''f' + g''f'g' + f''g''f' + f''g''f'g'.$$

Thus

$$(f, g) = 1 + [f_k, g_l] + \text{higher terms}.$$

We prove (*b*) by inductions on the weight of c_i, the assertion being immediate for weight 1. Thus let $c_k = (c_i, c_j)$; $f_i = \tau(c_i)$ has leading term c_i' and $f_j = \tau(c_j)$ has leading term c_j'. By part (a), (f_i, f_j) has leading term $[c_i', c_j'] = c_k'$ provided the latter expression is not zero; but being a basic commutator it cannot be zero by the theorem quoted above, and $(f_i, f_j) = (\tau(c_i), \tau(c_j)) = \tau(c_i, c_j) = \tau(c_k)$. Hence the lemma is proved.

References

1. Y. Bar-Hillel, M. Perles, and E. Shamir, On Formal Properties of simple Phrase-Structure Grammars, *Z. für Phonetik Sprachwiss. Kommunikat.* **14**, 2 (1961), 143–172, (reproduced as Chapter 9 in Bar-Hillel, *Language and Information*, Addison-Wesley, Reading, Mass., 1964).
2. N. Chomsky and G. A. Miller, Chapters 11, 12, 13 in D. Luce. ed., *Handbook of Mathematical Psychology*, Vol. 2, Wiley, New York, 1964 (in particular Chapter 12, "Formal Properties of Grammars," and its reference list).
3. N. Chomsky and M. P. Schützenberger, "Algebraic Theory of Context-Free Languages" in F. Braffort *et al.*, eds., *Computer Programming and Formal Systems*, North-Holland, Amsterdam, 1963.
4. M. Hall, *The Theory of Groups*, Macmillan, New York, 1959.
5. A. G. Kurosh, *Theory of Groups*, Vol. II, Chelsea, New York, 1960.
6. M. P. Schützenberger, On the Definition of a Family of Automata, *Information and Control* **4** (1961), 245–270.
7. R. H. Fox, Free Differential Calculus. I, *Ann. Math.* **57**, (1953), 547–560.
8. K. Krohn and J. Rhodes, Algebraic Theory of Machines, I. Prime Decomposition Theorem for Finite Semigroups and Machines, *Trans. Amer. Math. Soc.* **116** (1965), 450–464.