INFORMATION AND CONTROL **48**, 109–118 (1981)

The Characterization of Nonexpansive Grammars by Rational Power Series

GERD BARON

Institut für Angewandte Mathematik, Technische Universität Wien, Wien, Austria

AND

WERNER KUICH

Institut für Mathematische Logik und Formale Sprachen, Technische Universität Wien, Wien, Austria

Given a reduced, cycle-free context-free grammar $G = (\Phi, \Sigma, P, y_1)$, the following statements are equivalent.

(i) G is nonexpansive;

(ii) the structure generating functions of the grammars $G_i = (\Phi, \Sigma, P, y_i)$, $y_i \in \Phi$, are rational.

Furthermore a helpful theorem for proving certain context-free languages to be inherently ambiguous is given.

1. INTRODUCTION

In Kuich (1970), one of the authors introduced the concept of the structure generating function of an unambiguous ε-free context-free grammar or language and showed that the structure generating function of an unambiguous nonexpansive ε-free context-free grammar is a rational function. He conjectured that the structure generating function of an unambiguous ε-free context-free language, which cannot be generated by an unambiguous nonexpansive context-free grammar, is always nonrational (see also Salomaa and Soittola, 1978, Exercises IV.3.6 and IV.3.7).

A simple example similar to that of Jones (1970) shows that this conjecture is false.

EXAMPLE 1. Let $D(a, \bar{a})$ be the Dyck language over the alphabet $\{a, \bar{a}\}$. Then it is well known that $D(a, \bar{a})$ is a deterministic context-free language and hence $\{a, \bar{a}\}^* - D(a, \bar{a})$ is again a deterministic context-free language. Hence both languages are generated by unambiguous context-free grammars

0019-9958/81/020109–10$02.00/0
Copyright © 1981 by Academic Press, Inc.
All rights of reproduction in any form reserved.

643/48/2-2

and so is $L = D(a, \bar{a}) \cup (\{b, \bar{b}\}^* - D(b, \bar{b})) - \{\varepsilon\}$. The structure generating function of L is given by the rational function $z/(1-z)$.

Using almost looping grammars (Maurer, 1969; Gruska, 1969) and the result that $D(a, \bar{a})$ cannot be generated by a nonexpansive context-free grammar (Salomaa, 1969; Salomaa, 1973, Chap. VI.10), it is easily proved that L cannot be generated by a nonexpansive context-free grammar.

In the sequel, $G = (\Phi, \Sigma, P, y_1)$ with $\Phi = \{y_1, ..., y_n\}$, $\Sigma = \{z_1, ..., z_m\}$ denotes a reduced, context-free grammar and $G_i = (\Phi, \Sigma, P, y_i)$.

Let $l_i(w)$, $w \in \Sigma^*$, be the number of distinct leftmost derivations for w according to $G_i (l_i(w) = 0$ iff $w \notin L(G_i))$, i.e., the ambiguity of w according to G_i and assume $l_i(w) < \infty$.

Then the power series $g_i \in \mathbb{N}\langle\langle \Sigma^* \rangle\rangle$, $1 \leqslant i \leqslant n$, are defined by

$$g_i = \sum_{w \in \Sigma^*} l_i(w) w.$$

Denote by $c(\Sigma^*)$ the free commutative monoid generated by Σ and by h_c the natural homomorphism mapping Σ^* into $c(\Sigma^*)$.

Then the power series $h_i \in \mathbb{N}\langle\langle c(\Sigma^*) \rangle\rangle$, $1 \leqslant i \leqslant n$, are defined by

$$h_i = h_c(g_i),$$

i.e., the coefficient of $z_1^{i_1} z_2^{i_2} \cdots z_m^{i_m}$ in h_i equals the number of distinct leftmost derivation for all w according to G_i, such that the Parikh vector of w is $(i_1, i_2, ..., i_m)$.

Denote by z a complex variable and by $h: c(\Sigma^*) \to z^*$ the homomorphism defined by $h(z_i) = z$, $1 \leqslant i \leqslant m$.

Then the power series $f_i \in \mathbb{N}\langle\langle z^* \rangle\rangle$, $1 \leqslant i \leqslant n$, are defined by

$$f_i(z) = \sum_{i=0}^{\infty} u_i(n) z^n,$$

where $u_i(n) = \sum_{|w|=n} l_i(w)$, i.e., $u_i(n)$ is the number of distinct leftmost derivations for words $w \in L(G_i)$ of length n according to G_i.

The homomorphisms h_c and h are nonerasing and

$$f_i = h(h_i) = h \circ h_c(g_i), \qquad 1 \leqslant i \leqslant n.$$

We denote $f_1(z)$ by $f_G(z)$ and call it *structure generating function of* G (Takaoka, 1974).

We call G *cycle-free* if, for each nonterminal y_i, $1 \leqslant i \leqslant n$, $y_i \Rightarrow^* y_i$ is impossible.

Then we show, that the power series g_i, $1 \leqslant i \leqslant n$, and hence h_i and f_i are well defined if G is cycle-free. This leads to the following characterization result:

Let G be cycle-free. Then G is nonexpansive iff $h_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$ for all i, $1 \leqslant i \leqslant n$.

The rest of the paper deals with unambiguity of context-free grammars and inherently ambiguous context-free languages.

2. The Characterization of Nonexpansive Grammars

The *algebraic system induced by G* is defined by

$$y_i = p_i, \quad p_i \in \mathbb{N}\langle(\Phi \cup \Sigma)^*\rangle, \qquad 1 \leqslant i \leqslant n,$$

where p_i is the polynomial formed by the right sides of the productions for y_i.

Since G is cycle-free, the induced algebraic system has a strong solution by Lemma 3 of Kuich (1981) and this strong solution equals $(g_1,..., g_n)$. Hence the power series $g_1,..., g_n$; $h_1,..., h_n$; $f_1,...,f_n$ are in $\mathbb{N}^{\text{semi-alg}}\langle\langle\Sigma^*\rangle\rangle$; $\mathbb{N}^{\text{semi-alg}}\langle\langle c(\Sigma^*)\rangle\rangle$; $\mathbb{N}^{\text{semi-alg}}\langle\langle z^*\rangle\rangle$, respectively, by Theorems IV.6.4 and IV.3.3 of Salomaa and Soittola (1978) and the fact that the homomorphisms h and h_c are nonerasing.

The *dependence graph* $D(G)$ of the context-free grammar G is defined to be the directed graph with vertex set Φ, such that there is a line from y_i to y_j iff $y_j \to \alpha y_i \beta$ is a production of G.

If y_i and y_j are points in a strong component of $D(G)$, then there exist $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (\Phi \cup \Sigma)^*$ such that $y_j \Rightarrow^* \alpha_1 y_i \alpha_2$ and $y_i \Rightarrow^* \beta_1 y_j \beta_2$.

A strong component of $D(G)$ is called *expansive*, if it contains vertices y_i, y_j, y_k and there exist $\alpha_1, \alpha_2, \alpha_3 \in (\Phi \cup \Sigma)^*$ such that $y_i \Rightarrow^* \alpha_1 y_j \alpha_2 y_k \alpha_3$. Otherwise it is called *nonexpansive*.

The context-free grammar G is called *expansive*, if there exists an $y_i \in \Phi$ and $\alpha_1, \alpha_2, \alpha_3 \in (\Phi \cup \Sigma)^*$ such that $y_i \Rightarrow^* \alpha_1 y_i \alpha_2 y_i \alpha_3$. Otherwise it is called *nonexpansive*.

Lemma 1. *G is expansive iff at least one strong component of D(G) is expansive.*

Proof. If $D(G)$ has an expansive strong component, then there exist y_i, y_j, y_k and $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \gamma_1, \gamma_2 \in (\Phi \cup \Sigma)^*$ such that $y_i \Rightarrow^* \alpha_1 y_j \alpha_2 y_k \alpha_3$, $y_j \Rightarrow^* \beta_1 y_i \beta_2$, $y_k \Rightarrow^* \gamma_1 y_i \gamma_2$ and hence G is expansive.

If G is expansive, then there exist y_i and $\alpha_1, \alpha_2, \alpha_3 \in (\Phi \cup \Sigma)^*$ such that $y_i \Rightarrow^* \alpha_1 y_i \alpha_2 y_i \alpha_3$ and hence the strong component containing y_i is expansive.

In the sequel let $C_1, C_2, \ldots, C_r$ with vertex sets $\Phi_1, \Phi_2, \ldots, \Phi_r$ be the strong components of $D(G)$. Then we define the following partial order over the set of strong components of $D(G)$: $C_i \geqslant C_j$ iff there exist $y_{i_1} \in \Phi_i$, $y_{i_2} \in \Phi_j$ and $\alpha_1, \alpha_2 \in (\Phi \cup \Sigma)^*$ such that $y_{i_1} \Rightarrow^* \alpha_1 y_{i_2} \alpha_2$. If $C_i \geqslant C_j$ and $C_i \neq C_j$ then $C_i > C_j$.

THEOREM 1. *Let G be cycle-free. Let C be a strong component of $D(G)$ such that all strong components D of $D(G)$ with $D \leqslant C$ are nonexpansive. Then $h_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$ for y_i a point of D, $D \leqslant C$.*

Proof. Without loss of generality, let $C_1, C_2, \ldots, C_r$ be the strong components of $D(G)$ such that $i < j$ implies $C_i < C_j$ or C_i and C_j are incomparable. Let $C = C_l$. By Lemma 1 the y_i-productions of G, $y_i \in \Phi_k$, are linear in the variables of Φ_k, $1 \leqslant k \leqslant l$.

Hence the commutative variant of the algebraic system induced by G has the form

$$\begin{pmatrix} Y_r \\ \vdots \\ Y_1 \end{pmatrix} = \begin{pmatrix} P_r \\ \vdots \\ P_1 \end{pmatrix} + \begin{pmatrix} Q_r & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & Q_1 \end{pmatrix} \begin{pmatrix} Y_r \\ \vdots \\ Y_1 \end{pmatrix},$$

where Y_j and P_j are of dimension $|\Phi_j| \times 1$ and Q_j are of dimension $|\Phi_j| \times |\Phi_j|$, $1 \leqslant j \leqslant r$.

The components of P_k and Q_k, $1 \leqslant k \leqslant l$, are in $\mathbb{N}\langle c((\Sigma \cup \Phi_1 \cup \cdots \cup \Phi_{k-1})^*)\rangle$, the components of P_k, $l < k \leqslant r$, are in $\mathbb{N}\langle c((\Sigma \cup \Phi_1 \cup \cdots \cup \Phi_{k-1})^*)\rangle$ and the components of Q_k, $l < k \leqslant r$, are in $\mathbb{N}\langle c((\Sigma \cup \Phi_1 \cup \cdots \cup \Phi_k)^*)\rangle$.

By Kuich (1981), the strong solution of this system is $H = (h_1, \ldots, h_n)$.

We now proceed by induction on the index of the strong components of $D(G)$.

(i) Let $k = 1$. Consider $Y_1 = P_1 + Q_1 Y_1$, $P_1, Q_1 \in \mathbb{N}\langle c(\Sigma^*)\rangle$. Since G is cycle-free, Q_1 has the form $Q_1 = (Q_1, \varepsilon) + S_1$, S_1 quasiregular matrix and (Q_1, ε) nilpotent matrix. Hence $h_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$, $y_i \in \Phi_1$.

(ii) Let $1 < k \leqslant l$. Consider the subsystem

$$Y_k = P_k + Q_k Y_k.$$

By induction hypothesis $h_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$ for $y_i \in \Phi_1 \cup \cdots \cup \Phi_{k-1}$. Since H is solution of the whole system, $H_k = (h_i)_{y_i \in \Phi_k}$ is solution of

$$Y_k = H \cdot P_k + H \cdot Q_k Y_k.$$

Since $P_k, Q_k \in \mathbb{N}\langle c((\Sigma \cup \Phi_1 \cup \cdots \cup \Phi_{k-1})^*)\rangle$, $H \cdot P_k$ and $H \cdot Q_k$ are in $\mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$. $H \cdot Q_k$ can be written in the form $H \cdot Q_k = (H \cdot Q_k, \varepsilon) + (H \cdot S_k)$, $H \cdot S_k$ quasiregular matrix and $(H \cdot Q_k, \varepsilon)$ nilpotent matrix. This implies $h_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$, $y_i \in \Phi_k$.

COROLLARY 1. *Let G be cycle-free and nonexpansive. Then the structure generating function $f_G(z)$ is in $\mathbb{N}^{\mathrm{rat}}\langle\langle z^*\rangle\rangle$.*

Next we need a few technical lemmas.

Let G be cycle-free and let R_i be the radius of convergence of $f_i(z)$, $1 \leqslant i \leqslant n$.

LEMMA 2. *Let G be cycle-free. If $y_i \Rightarrow^* \alpha_1 y_j \alpha_2$, then $R_i \leqslant R_j$.*

Proof. Analogous to the proof of Lemma 3 of Kuich (1970).

LEMMA 3. *Let G be cycle-free. If y_i and y_j are vertices in a strong component of $D(G)$, then $R_i = R_j$.*

Proof. Since y_i and y_j are vertices in a strong component, there exist $\alpha_1, \alpha_2, \beta_1, \beta_2$ such that $y_i \Rightarrow^* \alpha_1 y_j \alpha_2$ and $y_j \Rightarrow^* \beta_1 y_i \beta_2$. Hence by Lemma 2 $R_i \leqslant R_j$ and $R_j \leqslant R_i$.

Let $l \geqslant 2$ and

$$y_i = \sum_{0 \leqslant k_1 + \cdots + k_m \leqslant l} p_{i;k_1,\ldots,k_m}(u_1,\ldots,u_n;z)\, y_1^{k_1} \cdots y_m^{k_m},$$

$$p_{i;k_1,\ldots,k_m}(u_1,\ldots,u_n;z) \in \mathbb{N}\langle c(\{u_1,\ldots,u_n,z\}^*)\rangle,$$

$1 \leqslant i \leqslant m$, $0 \leqslant k_1 + \cdots + k_m \leqslant l$ be an algebraic system of equations. The *dependence graph* of this system has vertex set $\{y_1,\ldots,y_m\}$. There is a line from y_j to y_i iff there exists a $p_{i;k_1,\ldots,k_m}(u_1,\ldots,u_n;z) \not\equiv 0$ with $k_j > 0$.

LEMMA 4. *Let $l \geqslant 2$ and*

$$y_i = \sum_{0 \leqslant k_1 + \cdots + k_m \leqslant l} p_{i;k_1,\ldots,k_m}(u_1,\ldots,u_n;z)\, y_1^{k_1} \cdots y_m^{k_m},$$

$1 \leqslant i \leqslant m$, be an algebraic system of equations with the following properties:

(1) *$p_{i;k_1,\ldots,k_m}(u_1,\ldots,u_n;z) \in \mathbb{N}\langle c(\{u_1,\ldots,u_n,z\}^*)\rangle$ for all $1 \leqslant i \leqslant m$, $0 \leqslant k_1 + \cdots + k_m \leqslant l$.*

(2) *There exists an index i and $l_1,\ldots,l_m$ such that $l_1 + \cdots + l_m \geqslant 2$ and $p_{i;l_1,\ldots,l_m}(u_1,\ldots,u_n;z) \not\equiv 0$.*

(3) *The dependence graph of the system of equations is strongly connected.*

(4) $u_j(z) \in \mathbb{N}^{\mathrm{rat}}\langle\langle z^* \rangle\rangle$, $u_j(z) \not\equiv 0$, *with radius of convergence* $\rho_j > 0$, $1 \leqslant j \leqslant n$.

(5) *The system of equations has a strong solution* $(f_1(z),\ldots,f_m(z))$ *with* $f_i(z) \in \mathbb{N}^{\text{semi-alg}}\langle\langle z^* \rangle\rangle, f_i(z) \not\equiv 0$, $1 \leqslant i \leqslant m$, *such that all* $f_i(z)$, $1 \leqslant i \leqslant m$, *have a common radius of convergence* ρ *with* $0 < \rho \leqslant \min\{\rho_j \mid 1 \leqslant j \leqslant n\}$.

Then there exists an index i, $1 \leqslant i \leqslant m$, *such that* $f_i(z) \notin \mathbb{N}^{\mathrm{rat}}\langle\langle z^* \rangle\rangle$ *or* $f_i(z) \equiv 1$ *for all* $1 \leqslant i \leqslant m$.

Proof. For proof by contradiction assume $f_i(z) \in \mathbb{N}^{\mathrm{rat}}\langle\langle z^* \rangle\rangle$, $1 \leqslant i \leqslant m$.

(a) $0 < \rho < +\infty$.

According to Pringsheim's theorem, each power series with center $z = 0$ and coefficients in $\mathbb{N}$ and radius of convergence $0 < \rho < +\infty$ represents a function which has a singular point at $z = \rho$ (see also Salomaa and Soittola, 1978, Theorem II.10.1). Together with (5) this implies

$$f_i(z) = \frac{f_{i1}(z)}{f_{i2}(z)(\rho - z)^{\lambda_i}}, \qquad \lambda_i > 0, \qquad \frac{f_{i1}(z)}{f_{i2}(z)} > 0$$

for $0 < z \leqslant \rho, f_{i1}(z), f_{i2}(z) \in \mathbb{R}\langle z^* \rangle$, $1 \leqslant i \leqslant m$.

Pringsheim's theorem and (4) imply

$$u_j(z) = \frac{u_{j1}(z)}{u_{j2}(z)(\rho - z)^{\mu_j}}, \qquad \mu_j \geqslant 0, \qquad \frac{u_{j1}(z)}{u_{j2}(z)} > 0$$

at least for $0 < z \leqslant \rho$, $u_{j1}(z)$, $u_{j2}(z) \in \mathbb{R}\langle z^* \rangle$, $1 \leqslant j \leqslant n$.

Hence

$$\frac{f_{i1}(z)}{f_{i2}(z)(\rho - z)^{\lambda_i}} = \frac{h_{i1}(z)}{h_{i2}(z)(\rho - z)^{\sigma_i}}, \qquad \frac{h_{i1}(z)}{h_{i2}(z)} > 0$$

for $0 < z \leqslant \rho$, $h_{i1}(z)$, $h_{i2}(z) \in \mathbb{R}\langle z^* \rangle$, $1 \leqslant i \leqslant m$, with

$$\sigma_i \geqslant \max\{k_1 \cdot \lambda_1 + \cdots + k_m \cdot \lambda_m \mid 0 \leqslant k_1 + \cdots + k_m \leqslant l, p_{i;k_1,\ldots,k_m} \not\equiv 0\}.$$

Since $\lambda_i = \sigma_i$ for all $0 \leqslant k_1 + \cdots + k_m \leqslant l$ with $p_{i;k_1,\ldots,k_m} \not\equiv 0$, $\lambda_i \geqslant k_1 \cdot \lambda_1 + \cdots + k_m \cdot \lambda_m$, $1 \leqslant i \leqslant m$.

This implies $\lambda_1 = \lambda_2 = \cdots = \lambda_m = \lambda$. For proof by contradiction assume without loss of generality

$$\lambda = \lambda_1 = \cdots = \lambda_t < \lambda_{t+1} \leqslant \lambda_{t+2} \leqslant \cdots \leqslant \lambda_m, \qquad t < m.$$

Then $\lambda \geqslant (k_1 + \cdots + k_t) \cdot \lambda + k_{t+1} \cdot \lambda_{t+1} + \cdots + k_m \cdot \lambda_m$, $1 \leqslant i \leqslant t$, implies $k_{t+1} = \cdots = k_m = 0$ for all $0 \leqslant k_1 + \cdots + k_m \leqslant l$ with $p_{i;k_1,\ldots,k_m}(u_1,\ldots, u_n; z) \not\equiv 0$, contradicting (3).

Hence $\lambda \geqslant (k_1 + \cdots + k_m)\lambda$.

By (2) $\lambda \geqslant 2\lambda$, which implies $\lambda = 0$, contradicting $\rho < +\infty$.

(b) $\rho = +\infty$.

Then $f_i(z) \in \mathbb{N}\langle z^*\rangle$, $1 \leqslant i \leqslant m$, and by (4) and (5) $u_j \in \mathbb{N}\langle z^*\rangle$, $1 \leqslant j \leqslant n$.

For $q(z) \in \mathbb{N}\langle z^*\rangle$, let $[q]$ denote the degree of q. Then by (4) and (5) $[f_i] \geqslant 0$, $1 \leqslant i \leqslant m$, $[u_j] \geqslant 0$, $1 \leqslant j \leqslant n$, and $[f_i] \geqslant k_1 \cdot [f_1] + \cdots + k_m \cdot [f_m]$ for all $0 \leqslant k_1 + \cdots + k_m \leqslant l$ with $p_{i;k_1,\ldots,k_m} \not\equiv 0$, $1 \leqslant i \leqslant m$.

Similar to (a) this implies $[f_1] = \cdots = [f_m] = 0$. Hence $f_i(z) \in \mathbb{N}$, $1 \leqslant i \leqslant m$.

Let $f_i(z) \equiv a_i > 0$, $1 \leqslant i \leqslant m$. Then $u_j(z) \in \mathbb{N}$, $1 \leqslant j \leqslant n$, and

$$a_i \geqslant a_1^{k_1} \cdots a_m^{k_m}$$

for all $0 \leqslant k_1 + \cdots + k_m \leqslant l$ with $p_{i;k_1,\ldots,k_m} \not\equiv 0$, $1 \leqslant i \leqslant m$.

Similar to (a) this implies $a_1 = \cdots = a_m = 1$.

THEOREM 2. *Let G be expansive and cycle-free. Then there exists an index i, $1 \leqslant i \leqslant n$, such that $f_i(z) \notin \mathbb{N}^{\mathrm{rat}}\langle\langle z^*\rangle\rangle$.*

Proof. Since G is expansive, there exists a smallest strong component C that is expansive. Hence if $D < C$, then D is nonexpansive. By Theorem 1 $h_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$ and hence $f_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle z^*\rangle\rangle$ for y_i vertex of D, $D < C$.

We are now in the position to apply Lemma 4: $y_1,\ldots, y_m$ are the vertices of C, $u_1,\ldots, u_n$ are the structure generating functions corresponding to the vertices of the strong components D, $D < C$. The system of equations is induced by the y_i-rules of G, y_i vertex of C. Condition (1) is trivially satisfied, (2) is implied by the fact that C is expansive, (3) is satisfied since C is a strong component, (4) is implied by Lemma 2 and Theorem 1 and (5) is implied by Lemma 3 and the fact that G is reduced. Since G is cycle-free, (2) implies that $f_i(z) \equiv 1$, $1 \leqslant i \leqslant m$, is no strong solution. Hence there exists an y_i, y_i a vertex of C, such that $f_i \notin \mathbb{N}^{\mathrm{rat}}\langle\langle z^*\rangle\rangle$.

COROLLARY 2. *Let G be expansive and cycle-free. Then there exists an index i, $1 \leqslant i \leqslant n$, such that $h_i \notin \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$.*

Proof. Since $f_i = h(h_i)$, h nonerasing, the Corollary is implied by Theorem 2 and Theorem IV.3.3 of Salomaa and Soittola (1978).

We are now in the position to characterize nonexpansive grammars.

THEOREM 3. *Let G be reduced and cycle-free. Then G is nonexpansive iff for all i, $1 \leqslant i \leqslant n$, $h_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$.*

COROLLARY 3. *Let G be reduced and cycle-free. Then G is nonexpansive iff for all i, $1 \leqslant i \leqslant n$, $f_i \in \mathbb{N}^{\mathrm{rat}}\langle\langle z^*\rangle\rangle$.*

Let G be cycle-free and denote h_1 by h_G. Let $L \subseteq \Sigma^*$ be a formal language and denote the commutative variant of char L by h_L. Then we can formulate the following conjecture:

(A) Let G be cycle-free and expansive. Then $h_G \notin \mathbb{N}^{\text{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$.

This would imply at once:

(B) Let G be unambiguous and expansive with $L = L(G)$. Then $h_L \notin \mathbb{N}^{\text{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$.

Hence together with Theorem 1 this would imply:

(C) Let G be unambiguous with $L = L(G)$. Then G is nonexpansive iff $h_L \in \mathbb{N}^{\text{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$.

3. Inherently Ambiguous Languages

Let L be a formal language and $v(n)$, $n \geqslant 0$, be the number of distinct words of length n in L. Then the function $f_L(z)$ of the complex variable z

$$f_L(z) = \sum_{n=0}^{\infty} v(n)\, z^n$$

is called *structure generating function of L*.

The next theorem was noted by several authors (Kuich and Maurer, 1971; Semenov, 1973; quoted in Salomaa and Soittola, 1978, Chap. IV.5; Takaoka, 1974).

THEOREM 4. *Assume that L is a context-free language and G is a context-free grammar of which it is known that $L(G) \supseteq L$.*

Then $f_L(z) = f_G(z)$ iff G is unambiguous and $L(G) = L$.

The next theorem is useful in proving context-free languages of a certain form to be inherently ambiguous.

THEOREM 5. *Let G_1 and G_2 be unambiguous context-free grammars with $L_1 = L(G_1)$ and $L_2 = L(G_2)$.*

Then $L_1 \cup L_2$ is an inherently ambiguous context-free language if $f_{L_1 \cap L_2}(z) \notin \mathbb{Z}^{\text{semi-alg}}\langle\langle z^*\rangle\rangle$.

Proof. Since $f_{L_1 \cup L_2}(z) = f_{L_1}(z) + f_{L_2}(z) - f_{L_1 \cap L_2}(z)$, $f_{L_1}(z)$, $f_{L_2}(z) \in \mathbb{N}^{\text{semi-alg}}\langle\langle z^*\rangle\rangle$ and $f_{L_1 \cap L_2}(z) \notin \mathbb{Z}^{\text{semi-alg}}\langle\langle z^*\rangle\rangle$, Theorem IV.3.1 of Salomaa and Soittola (1978) implies $f_{L_1 \cup L_2} \notin \mathbb{Z}^{\text{semi-alg}}\langle\langle z^*\rangle\rangle$. Hence Theorem IV.1.6 of Salomaa and Soittola (1978) implies that $L_1 \cup L_2$ is inherently ambiguous.

COROLLARY 4. *Let G_1 and G_2 be unambiguous context-free grammars with $L_1 = L(G_1)$ and $L_2 = L(G_2)$.*

Let $f_{L_1 \cap L_2}(z) = \sum_{n=0}^{\infty} a_n z^{k_n}$ *with* $\lim_{n\to\infty} k_n/n = +\infty$.

Then $L_1 \cup L_2$ is inherently ambiguous.

Proof. Exercise IV.5.8 of Salomaa and Soittola (1978).

EXAMPLE 2 (due to Ginsburg and Spanier (1971)). Let $L_1 = \{ba^iba^{i+2} \mid i \geqslant 1\}^* ba^*b$ and $L_2 = ba^2\{ba^iba^{i+2} \mid i \geqslant 1\}^* b$. Then L_1 and L_2 are unambiguous context-free languages and $L_1 \cap L_2 = \{ba^2ba^4b \cdots ba^{4k+2}b \mid k \geqslant 0\}$.

Hence $f_{L_1 \cap L_2}(z) = \sum_{k=0}^{\infty} z^{4(k+1)^2}$ and by Corollary 4 $L_1 \cup L_2$ is inherently ambiguous.

EXAMPLE 3 (due to Kemp (1980)). Let $L_1 = a\{b^ia^i \mid i \geqslant 1\}^*$ and $L_2 = \{a^ib^{2i} \mid i \geqslant 1\}^* a^+$. Then L_1 and L_2 are unambiguous context-free languages and $L_1 \cap L_2 = \{ab^2a^2b^4a^4 \cdots b^{2^k}a^{2^k} \mid k \geqslant 1\} \cup \{a\}$.

Hence $f_{L_1 \cap L_2}(z) = \sum_{k=0}^{\infty} z^{2^{k+2}-3}$ and by Corollary 4 $L_1 \cup L_2$ is inherently ambiguous.

Since the languages L_1 and L_2 of Examples 2 and 3 are generated by nonexpansive context-free grammars the following theorem holds.

THEOREM 6. *There are reduced cycle-free nonexpansive context-free grammars G with $L = L(G) \subseteq \Sigma^*$ such that $h_L \notin \mathbb{N}^{\mathrm{rat}}\langle\langle c(\Sigma^*)\rangle\rangle$.*

Hence in Conjecture (C) "*unambiguous*" *cannot be replaced by* "*cycle-free.*"

RECEIVED: December 20, 1980

REFERENCES

GINSBURG, S. AND SPANIER, E. H. (1971), AFL with the semilinear property, *J. Comput. Systems Sci.* **5**, 365–396.

GRUSKA, J. (1969), Some classifications of context-free languages, *Inform. Contr.* **14**, 152–179.

HARRISON, M. A. (1978), "Introduction to Formal Language Theory," Addison-Wesley, Reading, Mass.

JONES, N. D. (1970), A note on the index of a context-free language, *Inform. Contr.* **16**, 201–202.

KEMP, R. (1980), A note on the density of inherently ambiguous context-free languages, *Acta Inform.* **14**, 295–298.

KUICH, W. (1970), On the entropy of context-free languages, *Inform. Contr.* **16**, 173–200.

KUICH, W. (1981), "Cycle-Free $\mathbb{N}$-Algebraic Systems," *in* "Theoretical Computer Science" (P. Deussen, Ed.), pp. 5–12, Springer-Verlag, Berlin/New York/Heidelberg.

KUICH, W. AND MAURER, H. (1971), On the inherent ambiguity of simple tuple languages, *Computing* **7**, 194–203.

MAURER, H. (1969), A direct proof of the inherent ambiguity of a simple context-free language, *J. Assoc. Comput. Math.* **16**, 256–260.

SALOMAA, A. (1969), On the index of a context-free grammar and language, *Inform. Contr.* **14**, 474–477.

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

SALOMAA, A. AND SOITTOLA, M. (1978), "Automata-Theoretic Aspects of Formal Power Series," Springer-Verlag, New York/Heidelberg/Berlin.

SEMENOV, A. L. (1973), Algoritmitseskie problemy dlja stepennykh rjadov i kontekstnosvobodnykh grammatik, *Dokl. Akad. Nauk SSSR* **212**, 50–52.

TAKAOKA, T. (1974), A note on the ambiguity of context-free grammars, *Inform. Process. Lett.* **3**, 35–36.