View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

JOURNAL OF COMPUTER AND SYSTEM SCIENCES 17, 376–387 (1978)

A Definition of Measures over Language Space

TADAO TAKAOKA

Department of Information Science, Faculty of Engineering, Ibaraki University, Hitachi, Ibaraki 316, Japan

Received June 20, 1977

As an attempt to associate a real number with a language, entropies of languages are computed by Banerji, Kuich, and others. As mappings from languages to real numbers, in this paper, measures over languages are presented. These measures satisfy additivity while entropies do not. Two kinds of measures, p-measure and ω-measure, are defined, and the computing method of these measures is shown for regular languages and context-free languages. Some properties of these measures are applied to show the nonregularity of several languages.

1. Introduction

As an attempt to associate a real number with a language, entropies of languages are computed by Banerji [1] and Kuich [2]. In this paper, another kind of quantity, measures, is associated with languages. These measures satisfy additivity while entropies do not. Using several properties of these measures, we can prove the nonregularity of several languages, such as the set of binary expansions of perfect squares and the set of binary expansions of primes. The presented proofs are simpler than that for perfect squares by Ritchie [3] and that for primes by Minsky and Papert [4].

2. Definitions and Basic Properties

Let Σ be a finite alphabet and $\#(\Sigma) = r$. The free monoid generated from Σ is denoted by Σ^*. The null string is expressed by λ. Let $\wp(\Sigma^*)$ denote the set of all subsets of Σ^*. Thus the densities of Σ^* and $\wp(\Sigma^*)$ are countable and continuous. We call an element L of $\wp(\Sigma^*)$ a language and a subset $\mathscr{A}$ of $\wp(\Sigma^*)$ a language class.

DEFINITION 1. A language class $\mathscr{A}$ is said to be additive if for any $A, B \in \mathscr{A}$ it holds that

$$A \cap B = \varnothing \text{ (empty set)} \Rightarrow A \cup B \in \mathscr{A}.$$

For an additive language class $\mathscr{A}$, a mapping μ: $\mathscr{A} \to R$ (R: set of real numbers) is said to be additive if for any $A, B \in \mathscr{A}$ it holds that

$$A \cap B = \varnothing \Rightarrow \mu(A \cup B) = \mu(A) + \mu(B).$$

0022-0000/78/0173-0376$02.00/0
Copyright © 1978 by Academic Press, Inc.
All rights of reproduction in any form reserved.

For an additive language class $\mathscr{A}$, if a mapping $\mu: \mathscr{A} \to R$ is additive, then the mapping μ is said to be a measure over the language class $\mathscr{A}$, and the system $\langle \mathscr{A}, \mu \rangle$ is called a language measure space.

DEFINITION 2. For a language $A \in \wp(\Sigma^*)$, the sequence $\{p_n(A)\}$ is defined by

$$p_n(A) = \frac{1}{n} \sum_{t=0}^{n-1} N(A, t)\, r^{-t},$$

where the quantity $N(A, t)$ is defined as the number of words x in A such that the length of x, $lg(x)$, is equal to t. If the sequence $\{p_n(A)\}$ converges, the quantity $p(A)$ is defined by

$$p(A) = \lim_{n\to\infty} p_n(A).$$

The class of languages A such that the sequence $\{p_n(A)\}$ converges is denoted by $\mathscr{A}_p$.

THEOREM 1. *The system $\langle \mathscr{A}_p, p \rangle$ forms a language measure space, and the class $\mathscr{A}_p$ is closed under complementation.*

Proof. For any $A, B \in \mathscr{A}_p$ it holds that

$$A \cap B = \varnothing \Rightarrow N(A \cup B, t) = N(A, t) + N(B, t).$$

From this we see that the class $\mathscr{A}_p$ and the mapping p are additive. The latter part of the theorem is verified from

$$p_n(\bar{A}) = 1 - \frac{1}{n} \sum_{t=0}^{n-1} N(A, t)\, r^{-t},$$

and hence for A

$$p(\bar{A}) = 1 - p(A),$$

where $\bar{A}$ denotes the complement of A. Q.E.D.

For the measure p we have the properties that

$$p(\varnothing) = 0, \qquad p(\Sigma^*) = 1,$$
$$A \subseteq B \Rightarrow p(A) \leqslant p(B)$$

Thus the measure p is analogous to the probability measure.

LEMMA 1. *There exists a context-sensitive language such that the sequence $\{p_n(A)\}$ does not converge.*

Proof. Over the one symbol alphabet $\Sigma = \{\sigma\}$ the context-sensitive language A is defined by

$$A = \{x \in \Sigma^* \mid 2^{2k-2} \leqslant lg(x) < 2^{2k-1};\ k = 1, 2,...\}.$$

Now for $n = 2^{2m-1}$ ($m = 1, 2,\ldots$), we have that

$$p_n(A) = \frac{1}{2^{2m-1}} \sum_{k=1}^{m} (2^{2k-1} - 2^{2k-2}).$$

For this n the sequence $\{p_n(A)\}$ converges to $\frac{2}{3}$. Similarly calculating, the sequence $\{p_n(A)\}$ for $n = 2^{2m-2}$ ($m = 1, 2,\ldots$) converges to $\frac{1}{3}$. Q.E.D.

THEOREM 2. *The language family $\mathscr{A}_p$ is not closed under $\cup$ (set union) nor under $\cap$ (set intersection).*

Proof. Let the language L_1 be defined by

$$\begin{aligned} L_1 = &\{x \mid lg(x) \text{ is even and } 2^{2k-2} \leqslant lg(x) < 2^{2k-1};\ k = 1, 2,\ldots\} \\ &\cup \{x \mid lg(x) \text{ is odd and } 2^{2k-1} \leqslant lg(x) < 2^{2k};\ k = 1, 2,\ldots\}. \end{aligned}$$

Next the language L_2 is defined by

$$L_2 = \{x \mid lg(x) \text{ is even}\}.$$

Obviously it holds that $p(L_1) = p(L_2) = \frac{1}{2}$. By a calculation similar to that of Lemma 1 we have that the sequence $\{p_n(L_1 \cap L_2)\}$ converges to $\frac{1}{3}$ for $n = 2^{2m-1}$ ($m = 1, 2,\ldots$) and converges to $\frac{1}{6}$ for $n = 2^{2m-2}$ ($m = 1, 2,\ldots$). Thus the class $\mathscr{A}_p$ is proved not to be closed under the operation $\cap$. On the other hand we have that

$$p(\bar{L}_1) = p(\bar{L}_2) = \tfrac{1}{2}$$

and that

$$p_n(\bar{L}_1 \cup \bar{L}_2) = p_n(\overline{L_1 \cap L_2}) = 1 - p_n(L_1 \cap L_2).$$

From these facts it is concluded that the class $\mathscr{A}_p$ is not closed under set union. Q.E.D.

It is an interesting open problem whether or not class $\mathscr{A}_p$ is closed under concatenation.

DEFINITION 3. For a language $A \in \wp(\Sigma^*)$ the sequence $\{\omega_n^{(k)}(A)\}$ is defined by

$$\omega_n^{(k)}(A) = (k-1)! \sum_{t=0}^{n-1} \binom{t+k-1}{k-1} N(A, t)\, r^{-t}.$$

If the sequence $\{\omega_n^{(k)}(A)\}$ converges, the quantity $\omega^{(k)}(A)$ is defined by

$$\omega^{(k)}(A) = \lim_{n\to\infty} \omega_n^{(k)}(A).$$

The notations $\omega_n^{(1)}(A)$ and $\omega^{(1)}(A)$ can be simplified as $\omega_n(A)$ and $\omega(A)$. Thus the definition of $\omega_n(A)$ is given by

$$\omega_n(A) = \sum_{t=0}^{n-1} N(A, t)\, r^{-t}.$$

The class of languages A such that the sequence $\{\omega_n^{(k)}(A)\}$ converges is denoted by $\mathscr{A}_{\omega^{(k)}}$.

LEMMA 2. *The language class $\mathscr{A}_{\omega^{(k)}}$ is closed under the operations of $\cup$ and $\cap$, but not closed under complementation.*

Proof. Obvious.

From this lemma we have the following.

THEOREM 3. *The system $\langle \mathscr{A}_{\omega^{(k)}}, \omega^{(k)} \rangle$ forms a language measure space.*

Note that for an arbitrary positive integer k and an arbitrary language A it holds that

$$\omega^{(k)}(A) \leqslant \omega^{(k+1)}(A),$$

including the case in which the right-hand or both terms are infinite.

LEMMA 3.

$$\forall A \in \mathscr{A}_p \qquad (p(A) > 0 \Rightarrow \omega(A) = \infty),$$
$$\forall A \in \mathscr{A}_\omega \qquad (\omega(A) < \infty \Rightarrow p(A) = 0).$$

That is, for any $A \in \not{p}(\Sigma^*)$, $p(A)$ and $\omega(A)$ can not take nonzero finite value at the same time.

Proof. Obvious.

3. MEASURES OVER THE CLASS $\mathscr{R}$ OF REGULAR LANGUAGES

DEFINITION 4. A finite automaton of Moore-type $S = \langle S, M, s_1, F \rangle$ over an input alphabet Σ is defined as follows.

- S: a finite set of internal states.
- M: a mapping $S \times \Sigma \to S$ (state transition function).
- s_1: the initial state in S.
- F: the set of final states (subset of S).

The mapping M is uniquely extended to $S \times \Sigma^* \to S$ by the following condition.

$$M(s, \lambda) = s,$$
$$M(s, \sigma x) = M(M(s, \sigma), x),$$

where $s \in S$, $\sigma \in \Sigma$, and $x \in \Sigma^*$. The language $\beta(S)$ accepted by the finite automaton S is defined by

$$\beta(S) = \{x \in \Sigma^* \mid M(s_1, x) \in F\}.$$

DEFINITION 5. A language $A \in \not{p}(\Sigma^*)$ is said to be regular if there exists a finite automaton S such that $A = \beta(S)$. The class of regular languages is denoted by $\mathscr{R}$.

DEFINITION 6. For a finite automaton S, where the set S is expressed by $\{s_1, \ldots, s_n\}$, the (n, n)-matrix $Q_\sigma(S)$ is defined for each $\sigma \in \Sigma$ by

$$\begin{aligned}[Q_\sigma(S)]_{ij} &= 1, \qquad \text{if } M(S_i, \sigma) = S_j \\ &= 0, \qquad \text{otherwise.}\end{aligned}$$

Then the matrix $Q(S)$ is defined by

$$Q(S) = \sum_{\sigma \in \Sigma} Q_\sigma(S).$$

Finally, the matrix $P(S)$ is defined by

$$P(S) = Q(S) \cdot r^{-1}.$$

The matrix $P(S)$ is obviously a probability matrix.

DEFINITION 7. The vector η_F is defined to be a row vector whose ith component is 1 for $s_i \in F$ and 0 otherwise. The vector $(1, 0, \ldots, 0)$ is expressed by $\mathbf{a}$.

LEMMA 4.

$$N(\beta(S), t) = \mathbf{a}\{Q(S)\}^t \eta_F{}^T$$

$$p_n(\beta(S)) = \frac{1}{n} \sum_{t=0}^{n-1} \mathbf{a}\{Q(S)\}^t \, \eta_F{}^T \cdot r^{-t}$$

$$= \mathbf{a} \left\{ \frac{1}{n} \sum_{t=0}^{n-1} \{P(S)\}^t \right\} \eta_F{}^T,$$

where the symbol "T" means transposition.

Proof. Obvious.

THEOREM 4. *If a language $A \in \not p(\Sigma^*)$ is regular, then there exists a finite value for $p(A)$, and the value is a rational number.*

Proof. By the theory of Markov chains (Kemeny and Snell [5]), for a probability matrix P,

$$(1/n)(P + \cdots + P^n)$$

converges to P^*, where the matrix P^* is obtained by solving the equation

$$P^*P = P^*,$$

$$\sum_{j=1}^{n} [P^*]_{ij} = 1 \qquad (i = 1, \ldots, n).$$

In our case, $p(\beta(S))$ is given by

$$p(\beta(S)) = \mathbf{a}[P(S)]^* \eta_F^T.$$

The right-hand term of the above equation is therefore rational. Q.E.D.

As is well known, the class $\mathscr{R}$ forms a Boolean algebra with regard to set operations. The system $\langle \mathscr{R}, p \rangle$ is therefore a language measure space. It is open whether or not the p-measure can be defined over the class of context-free languages.

Hereafter a finite automaton S is supposed, without loss of generality, to be connected and minimal.

DEFINITION 8. For a finite automaton, a set $E \subseteq S$ is said to be ergodic if for any $s, t \in E$,

$$(\exists x \in \Sigma^+ M(s, x) = t) \wedge (\exists y \in \Sigma^+ M(t, y) = s),$$

and

$$M(E, \Sigma) = E.$$

A state not belonging to any ergodic set is said to be transient, and the set of transient states is said to be a transient set.

THEOREM 5. *If a language $A \in \not{p}(\Sigma^*)$ is regular and nonempty, one of the values $p(A)$ and $\omega(A)$ is nonzero finite.*

Proof. First suppose that some $s_i \in F$ belongs to some ergodic set. Then we have that $p(\beta(S)) > 0$. Second, every state in F is transient. Let M be a permutation matrix which permutes the state sequence $(s_1, ..., s_n)$ to $(s_{i_1}, ..., s_{i_n})$ where $s_{i_1}, ..., s_{i_k}$ for some k belong to ergodic sets, and the remainder are transient. Then the probability matrix $P(S)$ is transformed as follows.

$$P'(S) = M^{-1}P(S)M = \begin{bmatrix} P & 0 \\ R & Q \end{bmatrix},$$

where 0 is a zero matrix. This notation is due to Kemeny and Snell [5]. Let $\mathbf{a}M = \mathbf{a}'$ and $\eta_F M = \eta_F'$. Then

$$\omega_n(\beta(S)) = \mathbf{a}' \sum_{t=0}^{n-1} [P'(S)]^t \eta_F'^T$$

$$= \mathbf{a}' \sum_{t=0}^{n-1} \tilde{Q}^t \eta_F'^T,$$

where

$$\tilde{Q} = \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix}.$$

Since the absolue value of any eigenvalue of the matrix Q is smaller than 1, $\omega_n(\beta(S))$ converges to a nonzero finite value. These facts and Lemma 3 prove the theorem. Q.E.D.

THEOREM 6. *For a regular language A, if $\omega(A)$ has a finite value then $\omega^{(k)}(A)$ also has a finite value for any $k \geqslant 1$.*

Proof. Suppose that $\omega(A)$ has a finite value. Then from the above theorem the absolute values of the eigenvalues of the matrix Q are smaller than 1. In this case we have the following identity.

$$(k-1)! \sum_{t=0}^{\infty} \binom{t+k-1}{k-1} Q^t = (I-Q)^{-k}.$$

Since $\omega_n^{(k)}(A)$ is given by the formula

$$\omega_n^{(k)}(A) = (k-1)!\, \mathbf{a}' \sum_{t=0}^{n-1} \binom{t+k-1}{k-1} Q^t \eta_F'^T$$

$$= (k-1)!\, \mathbf{a}'(I-\tilde{Q})^{-k} \eta_F'^T,$$

$\omega_n^{(k)}(A)$ has a finite value. Q.E.D.

THEOREM 7. *For a regular language A and a positive integer k, if $\omega^{(k)}(A)$ has a finite value, then it is a rational number.*

Proof. Obvious from the above theorem. Q.E.D.

Remark. For concatenation we have that

$$\omega(AB) \leqslant \omega(A)\, \omega(B).$$

Hence if $\omega(A) < 1$ then $\omega(A^*)$ is finite, and it holds that

$$\omega(A^*) \leqslant 1/(1-\omega(A)).$$

EXAMPLE 1. A Knuth language (Knuth [6]) L is given by

$$L = \{x \in \{a, b\}^* \mid \text{equal number of } a\text{'s and } b\text{'s occur in } x\}.$$

For the language L we have

$$N(L, 2n) = {}_{2n}C_n = \frac{(2n)!}{(n!)^2} \sim \frac{1}{(\pi n)^{1/2}}\, 2^{2n},$$

$$N(L, 2n) \cdot 2^{-2n} \sim 1/(\pi n)^{1/2}$$

From this the following relations hold.

$$N(L, 2n) \cdot 2^{-2n} \to 0 \qquad (n \to \infty),$$

$$p(L) = 0,$$

$$\sum_{t=0}^{n-1} N(L, 2t) \cdot 2^{-2t} \to \infty \qquad (n \to \infty),$$

$$\omega(L) = \infty.$$

Hence the language L is not regular. By the way, there is an unambiguous context-free grammar generating L (Hopcroft and Ullman [7]).

By another method we can show the nonregularity of L. Let z be a complex variable. Then we have the following expansion with the radius 1 of convergence.

$$\frac{1}{(1-z)^{1/2}} = 1 - \frac{1}{2}z + \cdots + \frac{(2n)!}{(n!)^2} 2^{-2n}(-z)^n + \cdots.$$

If the language L is assumed to be defined over the alphabet $\Sigma' = \{a, b, c, d\}$, where c and d are dummy symbols, we have $\omega(L)$ as follows.

$$\omega(L) = \sum_{n=0}^{\infty} \frac{(2n)!}{(n!)^2} 4^{-2n}.$$

On the other hand it holds that

$$\frac{1}{(1+\frac{1}{4})^{1/2}} = \frac{2}{5^{1/2}} = \sum_{n=0}^{\infty} \frac{(2n)!}{(n!)^2} \cdot 4^{-2n}.$$

Hence $\omega(L) = 2/5^{1/2}$, which is irrational, and L is proved to be nonregular over the alphabet Σ'. Since the symbols c and d are dummy, L is not regular over Σ either.

EXAMPLE 2. The kth Knuth language L_k over the alphabet $\Sigma_k = \{a_1, a_2, \ldots, a_k\}$ is defined by

$$L_k = \{x \in \Sigma_k^* \mid \text{equal number of } a_1\text{'s}, \ldots, a_k\text{'s occur in } x\}.$$

The following relations hold.

$$N(L_k, kn) = (kn)!/(n!)^k$$
$$\sim n^{(1-k)/2} k^{kn}$$
$$N(L_k, kn) \cdot k^{-kn} \sim n^{(1-k)/2}.$$

For $k = 2$ and 3, L_k is nonregular because $p(L_k) = 0$ and $\omega(L_k) = \infty$. For $k \geqslant 4$, from Theorem 6 L_k is nonregular because $\omega(L_k) < \infty$ and $\omega^{(m)}(L_k) = \infty$, where $m = \lceil (k-1)/2 \rceil$.

EXAMPLE 3. Let L be the set of binary expansions of prime numbers, that is,

$$L = \{10, 11, 101, \ldots\}.$$

From the prime number theorem the number of primes between 0 and m is nearly equal to $m/\log m$, where m is a large positive integer. The number $N(L, n)$ is estimated as follows.

$$N(L, n) \sim \frac{2^n}{n} - \frac{2^{n-1}}{n-1}$$
$$= \frac{(n-2)}{n(n-1)} \cdot 2^{n-1}$$
$$\sim \frac{1}{n} \cdot 2^n,$$

$$N(L, n) \cdot 2^{-n} \to 0 \qquad (n \to \infty),$$
$$p(L) = 0$$
$$\sum_{t=0}^{n-1} N(L, t) \cdot 2^{-t} \to \infty \qquad (n \to \infty),$$
$$\omega(L) = \infty$$

Hence the language is not regular.

The language L is, by the way, proved to be non-context-free by Hartmanis and Shank [8].

EXAMPLE 4. Let L be the set of binary expansions of perfect squares, that is,

$$L = \{1, 100, 1001, \ldots\}.$$

In Fig. 1 the number of perfect squares contained in part $P(2m-1)$ is larger by one than that in part $Q(2m-1)$, and that in part $P(2m)$ is smaller by one than that in part

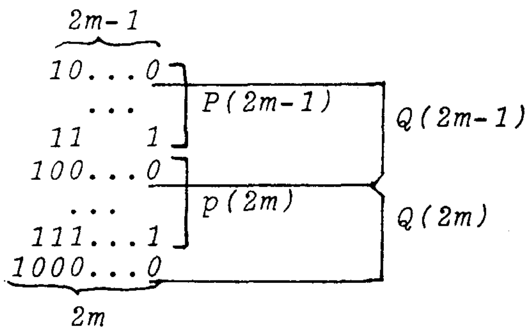

FIG. 1. Binary expansions of perfect squares.

$Q(2m)$. Using the fact that the number of perfect squares not larger than n is equal to $\lfloor n^{1/2} \rfloor$, $\omega(L)$ is computed as follows.

$$\omega_{2n}(L) = \sum_{m=1}^{n} \{(\lfloor 2^{m-1/2} \rfloor - 2^{m-1} + 1) \cdot 2^{-(2m-1)} + (2^m - \lfloor 2^{m-1/2} \rfloor - 1) \cdot 2^{-2m}\}$$
$$= \sum_{m=1}^{n} \{\lfloor 2^{m-1/2} \rfloor \cdot 2^{-2m} + 2^{-2m}\},$$
$$\omega(L) = \sum_{m=1}^{\infty} \lfloor {}^{m-1/2} \rfloor \cdot 2^{-2m} + \tfrac{1}{3}.$$

Now let $\cdot\alpha_1\alpha_2\alpha_3 \cdots$ be the binary expansion of $2^{1/2}/2$, which is not ultimately periodic. Then, as demonstrated in Fig. 2, the binary expansion of the first term of $\omega(L)$ becomes $\cdot\alpha_1 0 \alpha_2 0 \alpha_3 0 \cdots$, which is not ultimately periodic either. Hence $\omega(L)$ is irrational and L is nonregular. Q.E.D.

$$
\begin{array}{ll}
m & \\
1 & .0\alpha_1 \\
2 & .0\ 0\alpha_1\alpha_2 \\
3 & .0\ 0\ 0\alpha_1\alpha_2\alpha_3 \\
4 & .0\ 0\ 0\ 0\alpha_1\alpha_2\alpha_3\alpha_4 \\
5 & .0\ 0\ 0\ 0\ 0\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5 \\
6 & .0\ 0\ 0\ 0\ 0\ 0\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6 \\
+) & \cdots \\
\hline
 & .\alpha_1 0 \alpha_2 0 \alpha_3 0 \ldots
\end{array}
$$

FIG. 2. First term of $\omega(L)$.

4. MEASURES OVER CONTEXT-FREE LANGUAGES

DEFINITION 9. Let $G = (V, \Sigma, P, S)$ be a context-free grammar (CFG), where V is the set of nonterminals; Σ, terminals; P, rewriting rules; and S is the starting symbol. The grammar G is assumed to be reduced. A homomorphism φ is defined as follows. For $A_i \in V$, $\varphi(A_i) = y_i$, where $i = 1,..., n$, n is the number of nonterminals, $A_1 = S$, and y_i is a complex variable. For $a \in \Sigma$, $\varphi(a) = z$, where z is a complex variable. For $x \in (V \cup \Sigma)^+$, where $x = B_1 B_2 \cdots B_m$,

$$\varphi(x) = \varphi(B_1)\, \varphi(B_2) \cdots \varphi(B_m).$$

The generating equation of G is defined as follows.

$$\varphi(A_i) = \sum_{j=1}^{r_i} \varphi(\alpha_j), \qquad i = 1,..., n,$$

where $A_i \to \alpha_j$ $(j = 1,..., r_i,\ i = 1,..., n)$ is in P.

We can easily show that for any CFG its generating equation has a unique analytical solution,

$$y_i = f_i(z), \qquad f_i(0) = 0 \ (i = 1,..., n),$$

in a disk with center $z = 0$. We expand $f_1(z)$ into a power series in the following manner:

$$f_1(z) = \sum_{n=1}^{\infty} g(n)\, z^n.$$

The coefficient $g(n)$ expresses the number of derivations of words of length n generated by G. We call $f_1(z)$ the generating function of G, and write $f(z)$ for $f_1(z)$ for simplicity.

EXAMPLE 5. Consider a CFG $G = (\{S\}, \{a, b\}, \{S \to aSS, S \to b\}, S)$. The context-free language (CFL) $L(G)$ is the set of polish notations with operators a's and operands b's. For G we have the following generating equation:

$$y_1 = zy_1^2 + z.$$

The generating function becomes as follows.

$$f(z) = \frac{1 - (1 - 4z^2)^{1/2}}{2z},$$

$$f(z) = \sum_{n=0}^{\infty} \frac{(2n)!}{n!\,(n+1)!} z^{2n+1}.$$

DEFINITION 10. The structure-generating function of a language L, $h(z)$, is defined as follows

$$h(z) = \sum_{n=1}^{\infty} N(L, n)\, z^n.$$

This function was defined by Kuich [2] in the case in which L is an unambignous CFL. It is obvious that for a CFG G the CFL $L(G)$ is unambiguous if and only if its generating function is equal to its structure generating function. The relation between ambiguity and generating functions of CFGs is discussed by Takaoka [9].

Whenever the power series for $f(z)$ converges for a given z, the power series for $h(z)$ converges. If we denote the radii of convergence of power series by R_f and R_h, respectively, then they are given by

$$R_f = \overline{\lim_{n\to\infty}}\, (g(n))^{1/n}, \qquad R_h = \overline{\lim_{n\to\infty}}\, (N(L, n))^{1/n}.$$

Note that $R_f \geqslant R_h$.

THEOREM 8. *For a* CFL L *if* $r > 1/R_h$ *then* $\omega^{(k)}(L)$ *has a finite value. For a* CFG G *if* $r > 1/R_f$ *then* $\omega^{(k)}(L(G))$ *has a finite value. For an unambiguous* CFG G *if* $r > 1/R_f$ *then* $\omega^{(k)}(L(G))$ *has a finite value which is an algebraic number and computed by*

$$\omega^{(k)}(L(G)) = f^{(k-1)}(r^{-1}),$$

where $f^{(m)}(z)$ *is the derivative of* $f(z)$ *of mth order.*

Proof. Obvious.

EXAMPLE 6. The CFG of Example 5 is unambiguous. Then

$$\omega(L) = 1 \qquad \text{over} \quad \Sigma = \{a, b\},$$

$$\omega(L) = \frac{3 - 5^{1/2}}{2} \qquad \text{over} \quad \Sigma' = \{a, b, c\},$$

where c is a dummy symbol. The second value is an irrational number and hence $L(G)$ is nonregular.

On the other hand, $\omega^{(2)}(L(G))$ over $\Sigma = \{a, b\}$ is computed as follows.

$$f(z) = h(z) = \sum_{n=1}^{\infty} \frac{(2n)!}{n!\,(n+1)!} z^{2n+1},$$

$$f'(z) = \sum_{n=1}^{\infty} \frac{(2n+1)(2n)!}{n!\,(n+1)!} z^{2n},$$

$$\omega^{(2)}(L(G)) = f'(2^{-1}) = \infty.$$

This shows the nonregularity of $L(G)$, too.

It is open whether or not for any CFG G (ambiguous or unambiguous) $\omega(L(G))$ is an algebraic number.

Acknowledgment

The author wishes to thank Dr. N. Ikeno, who discussed the subject and showed the proof of Theorem 2.

References

1. R. Banerji, Phrase structure languages, finite machines, and channel capacity, *Inform. Contr.* **6** (1963), 153–162.
2. W. Kuich, On the entropy of context-free languages, *Inform. Contr.* **16** (1970), 173–200.
3. R. W. Ritchie, Finite automata and the set of squares, *J. Assoc. Comput. Mach.* **10** (1963), 528–531.
4. M. Minsky and S. Papert, Unrecognizable set of numbers, *J. Assoc. Comput. Mach.* **13** (1966), 281–286.
5. J. G. Kemeny and J. Snell, "Finite Markov Chains," Van Nostrand, Princeton, N.J., 1960.
6. D. E. Knuth, On the translation of languages from left to right, *Inform. Contr.* **8** (1965), 607–639.
7. J. E. Hopcroft and J. D. Ullman, "Formal Languages and Their Relation to Automata," Addison–Wesley, Reading, Mass., 1969.
8. J. Hartmanis and H. Shank, On the recognition of primes by automata, *J. Assoc. Comput. Mach.* **15** (1968), 382–389.
9. T. Takaoka, A note on the ambiguity of context-free grammars, *Inform. Processing Lett.* **3** (1974), 35–36.