JOURNAL OF COMPUTER AND SYSTEM SCIENCES **52**, 185–190 (1996)
ARTICLE NO. 0014

On Parikh Slender Languages and Power Series

JUHA HONKALA*

Department of Mathematics, University of Turku, SF-20500 Turku, Finland

the power series attains finitely many values

Received June 27, 1995; revised September 11, 1995

We define and study Parikh slender languages and power series. A language is Parikh slender if the number of words in the language with the same Parikh vector is bounded from above. As an application we get a new method for ambiguity proofs of context-free languages and a new proof of an earlier result of Autebert, Flajolet, and Gabarro concerning prefixes of infinite words.

1. INTRODUCTION

Length considerations are often useful in language theory. For example, Flajolet [8] has shown that the inherent ambiguity of very many context-free languages can be deduced from the transcendentality of their generating functions. Other deep results based on length considerations are well known, e.g., in the theory of Lindenmayer systems (see Rozenberg and Salomaa [18]).

Andraşiu, Dassow, Păun, and Salomaa [1] have defined and studied languages with the property that for each n the number of words in the language of length n is bounded from above by a constant. They have termed such languages *slender*. By now the theory of slender languages has been developed in many directions in Păun and Salomaa [14–16], Dassow, Păun, and Salomaa [5], Ilie [10], Raz [17], and Nishida and Salomaa [13]. We mention only that slender languages are also of cryptographic interest.

In this paper we consider the following modification of the approach of Andraşiu, Dassow, Păun, and Salomaa [1]. Instead of words of length n we count the number of words with the same Parikh vector. By definition, if $\Sigma = \{a_1, ..., a_m\}$ is an alphabet and $w \in \Sigma^*$ is a word, the Parikh vector $\psi(w)$ of w is defined by

$$\psi(w) = (\#_{a_1}(w), ..., \#_{a_m}(w)).$$

Here $\#_a(w)$ is the number of the occurrences of the letter a in w. Now, a language $L \subseteq \Sigma^*$ is termed *Parikh slender* if there is a positive integer k such that for each $(i_1, ..., i_m) \in \mathbf{N}^m$ there are at most k words in L with the Parikh vector $(i_1, ..., i_m)$. Below we consider also the natural generalization of Parikh slenderness to formal power series.

Standard terminology and notation concerning formal languages and power series will be used in this paper. Whenever necessary, the reader may consult Salomaa [19], Berstel and Reutenauer [3], Kuich and Salomaa [12], and Salomaa and Soittola [20]. We now outline the contents of the paper.

Section 2 contains the basic definitions. In Section 3 we characterize the Parikh slender regular languages and $\mathbf{N}$-rational series. The related decision problems are shown to be solvable. In Section 4 we discuss Parikh slender context-free languages and algebraic series. In particular, we establish a deep result concerning bounded $\mathbf{Z}$-algebraic series with commuting variables. This result is used in Section 5 to obtain a new method for ambiguity proofs of context-free languages. We also give a new simple proof of the well known result of Autebert, Flajolet, and Gabarro [2] concerning the prefixes of infinite words.

2. DEFINITIONS AND PREVIOUS RESULTS

Consider a language L over the alphabet Σ. L is said to be *thin* if for some n_0,

$$\text{card}(\{w \in L \mid |w| = n\}) \leqslant 1 \qquad \text{whenever} \quad n \geqslant n_0.$$

L is said to be *slender* if there exists a positive integer k such that

$$\text{card}(\{w \in L \mid |w| = n\}) \leqslant k \qquad \text{for all} \quad n \geqslant 0.$$

The definitions of thin and slender languages are due to Andraşiu, Dassow, Păun, and Salomaa [1]. Furthermore, L is *sparse* or *poly-slender* if there is a polynomial $p(n)$ such that

$$\text{card}(\{w \in L \mid |w| = n\}) \leqslant p(n) \qquad \text{for all} \quad n \geqslant 0.$$

Clearly, a thin language is slender and a slender language is sparse.

The following notions are used in the characterization of slender languages. A language $L \subseteq \Sigma^*$ is said to be a *union*

[*] E-mail: jhonkala@sara.cc.utu.fi.

0022-0000/96 $12.00

of single loops (briefly, USL) if for some k and words $u_i, v_i, w_i \in \Sigma^*$,

$$L = \bigcup_{i=1}^{k} u_i v_i^* w_i.$$

$L \subseteq \Sigma^*$ is said to be a *union of paired loops* (UPL) if for some k and words $u_i, v_i, w_i, x_i, y_i \in \Sigma^*$,

$$L = \bigcup_{i=1}^{k} \{u_i v_i^n w_i x_i^n y_i \mid n \geqslant 0\}.$$

The following result was established in Păun and Salomaa [16] (see also Shallit [22]).

THEOREM 2.1. *A regular language L is slender if and only if L is USL.*

The next result was conjectured in Păun and Salomaa [16] and proved by Ilie [10] and Raz [17].

THEOREM 2.2. *A context-free language L is slender if and only if L is UPL.*

Raz [17] also shows that it is decidable whether or not a given context-free language is slender.

Let now Σ be an alphabet and denote by $\Sigma^{\oplus}$ the free commutative monoid generated by Σ. Furthermore, denote by c the canonical morphism $c\colon \mathbf{R}\langle\langle \Sigma^* \rangle\rangle \to \mathbf{R}\langle\langle \Sigma^{\oplus} \rangle\rangle$. (Here $\mathbf{R}$ is the semiring of the real numbers.) A power series $r \in \mathbf{R}_+\langle\langle \Sigma^* \rangle\rangle$ is said to be *Parikh slender* if there exists a positive integer k such that

$$(c(r), w) \leqslant k \qquad \text{for all} \quad w \in \Sigma^{\oplus}.$$

Hence, a series $r \in \mathbf{R}_+\langle\langle \Sigma^* \rangle\rangle$ is Parikh slender if and only if the coefficients of $c(r)$ are bounded from above by a constant. A series $r \in \mathbf{R}_+\langle\langle \Sigma^* \rangle\rangle$ is said to be *Parikh thin* if

$$(c(r), w) \leqslant 1 \qquad \text{for almost all} \quad w \in \Sigma^{\oplus}.$$

The definitions of Parikh thin and Parikh slender languages are now obtained as a special case. If $L \subseteq \Sigma^*$ is a language, denote by $\operatorname{char}(L) \in \mathbf{N}\langle\langle \Sigma^* \rangle\rangle$ the characteristic series of L. A language L is said to be *Parikh thin* or *Parikh slender* if the series $\operatorname{char}(L)$ is Parikh thin or Parikh slender, respectively. Intuitively, a language L is Parikh slender if the number of words in L with the same Parikh vector is bounded from above.

The following are typical examples of Parikh slender languages:

$$L_1 = a^* b^* c^*,$$

$$L_2 = a(a^2)^* \, b(ab)^* \, a^2,$$

$$L_3 = \{a^m b^m c^n \mid m, n \geqslant 1\} \cup \{a^m b^n c^n \mid m, n \geqslant 1\}.$$

Clearly, a slender language is Parikh slender. The language L_1 is Parikh thin, but not slender. Intuitively, the reason is that when considering Parikh thinness the words of length n are divided into $\binom{n+\operatorname{card}(\Sigma)-1}{n}$ classes each consisting of words with equal Parikh vectors. Note that a Parikh slender language is sparse.

Parikh slender regular or context-free languages have weaker closure properties than slender regular or context-free languages, respectively. For example, the language $L_4 = a^* b^* a^*$ is a morphic image of L_1 but is not Parikh slender, whereas it follows from Theorem 2.1 that the morphic image of a slender regular language is slender. L_3 above is an example of a Parikh slender context-free language which is inherently ambiguous. It is a consequence of Theorem 2.2 that every slender context-free language is unambiguous (see Păun and Salomaa [16]).

3. PARIKH SLENDER REGULAR LANGUAGES

We begin with a characterization of Parikh slender regular languages.

A language $L \subseteq \Sigma^*$ is said to be a *multiple loop language* if there exist $k \geqslant 1$ and $u_1, v_1, u_2, v_2, ..., u_k, v_k, u_{k+1} \in \Sigma^*$ such that

$$L = u_1 v_1^* u_2 v_2^* u_3 \cdots u_k v_k^* u_{k+1} \tag{1}$$

and

$$\psi(v_1), ..., \psi(v_k) \quad \text{are linearly independent over } \mathbf{Q}. \tag{2}$$

A language $L \subseteq \Sigma^*$ is said to be a *union of multiple loops* (UML) if L is a finite disjoint union of multiple loop languages. Note that if (1) and (2) hold and $w \in L$ there exist unique integers $i_1, ..., i_k$ such that

$$w = u_1 v_1^{i_1} u_2 v_2^{i_2} u_3 \cdots u_k v_k^{i_k} u_{k+1}.$$

Furthermore, if (1) and (2) hold then k is less than or equal to the cardinality of Σ.

THEOREM 3.1. *A regular language L is Parikh slender if and only if L is UML.*

Proof. Clearly, a multiple loop language is Parikh thin. Hence, every UML language is Parikh slender. Conversely, let $L \subseteq \Sigma^*$ be a Parikh slender regular language. The regularity of L implies that L is an unambiguous rational subset of Σ^* (see Eilenberg [6]). Hence,

$$L = \bigcup_{i=1}^{n} w_{i0} L_{i1}^* w_{i1} L_{i2}^* w_{i2} \cdots L_{ik_i}^* w_{ik_i}, \tag{3}$$

where n and k_i, $1 \leqslant i \leqslant n$, are nonnegative integers, $w_{ij} \in \Sigma^*$ are words, and $L_{ij} \subseteq \Sigma^+$ are codes and the union is disjoint.

Furthermore, if $w \in L$ there exist unique i, $1 \leqslant i \leqslant n$, and $u_{ij} \in L_{ij}^*$, $1 \leqslant j \leqslant k_i$, such that

$$w = w_{i0} u_{i1} w_{i1} u_{i2} w_{i2} \cdots u_{ik_i} w_{ik_i}.$$

Next, suppose $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant k_i$ and let v_{ij} be a word of minimal length in L_{ij}. Then $L_{ij} = v_{ij}$. Indeed, if $v_1 \in L_{ij}$ and $v_1 \neq v_{ij}$ then $v_{ij} v_1 \neq v_1 v_{ij}$ and L contains the language

$$w_{i0} \cdots w_{i,j-1} \{v_{ij} v_1, v_1 v_{ij}\}^* w_{ij} \cdots w_{ik_i}.$$

Hence, for each $m \geqslant 1$, L contains at least 2^m distinct words with the same Parikh vector.

To conclude the proof it suffices to prove that for $1 \leqslant i \leqslant n$, the vectors $\psi(v_{i1}), ..., \psi(v_{ik_i})$ are linearly independent over **Q**. Suppose this is not true. Then there exist positive integers $a_1, ..., a_{k_i}, b_1, ..., b_{k_i}$ such that

$$a_1 \psi(v_{i1}) + \cdots + a_{k_i} \psi(v_{ik_i}) = b_1 \psi(v_{i1}) + \cdots + b_{k_i} \psi(v_{ik_i})$$

and $a_t \neq b_t$ for some t, $1 \leqslant t \leqslant k_i$. For $m \geqslant 1$, consider the words

$$w_{i0} v_{i1}^{xa_1 + (m-x) b_1} w_{i1} \cdots w_{i,k_i - 1} v_{ik_i}^{xa_{k_i} + (m-x) b_{k_i}} w_{ik_i}$$

for $0 \leqslant x \leqslant m$. These words are distinct elements of L and have the same Parikh vector. Therefore L is not Parikh slender. This contradiction shows that $\psi(v_{i1}), ..., \psi(v_{ik_i})$ are linearly independent and concludes the proof. ▌

Note that the characterization of slender regular languages given by Păun and Salomaa is a consequence of Theorem 3.1.

THEOREM 3.2. *It is decidable whether or not a given regular language L is Parikh slender.*

Proof. It is possible to obtain effectively the words w_{ij}, $1 \leqslant i \leqslant n$, $0 \leqslant j \leqslant k_i$, and codes L_{ij}, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant k_i$, such that (3) holds and the union and the products in (3) are unambiguous (see Eilenberg [6]). By the proof of Theorem 3.1, the language L is Parikh slender if and only if there exist words v_{ij} such that $L_{ij} = v_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant k_i$, and the vectors $\psi(v_{ij})$, $1 \leqslant j \leqslant k_i$, are linearly independent over **Q** for each $1 \leqslant i \leqslant n$. Both conditions are clearly decidable. ▌

Theorems 3.1 and 3.2 generalize easily to **N**-rational series.

THEOREM 3.3. *A series $r \in \mathbf{N}^{\text{rat}}\langle\langle \Sigma^* \rangle\rangle$ is Parikh slender if and only if r is a finite* **N**-*linear combination of series of the form*

$$u_1 v_1^* u_2 v_2^* u_3 \cdots u_n v_n^* u_{n+1},$$

where $\psi(v_1), ..., \psi(v_n)$ are linearly independent, $n \geqslant 0$, $u_i, v_j \in \Sigma^*$.

Proof. If the series $r \in \mathbf{N}^{\text{rat}}\langle\langle \Sigma^* \rangle\rangle$ is Parikh slender, its coefficients are bounded by a constant. Hence r is a finite **N**-linear combination of characteristic series of Parikh slender regular languages. The claim follows by Theorem 3.1. ▌

THEOREM 3.4. *It is decidable whether or not a given series $r \in \mathbf{N}^{\text{rat}}\langle\langle \Sigma^* \rangle\rangle$ is Parikh slender.*

Proof. Boundedness of the coefficients of r is a decidable necessary condition for the Parikh slenderness of r (see Jacob [11]). If the coefficients are bounded, r is effectively an **N**-linear combination of characteristic series of regular languages and the claim follows by Theorem 3.2. ▌

4. PARIKH SLENDER CONTEXT-FREE LANGUAGES AND ALGEBRAIC SERIES

In this section we establish necessary conditions for Parikh slender context-free languages and algebraic series and discuss the related decision problems.

A language $L \subseteq \Sigma^*$ is said to be *bounded* if there exist words $w_1, ..., w_n \in \Sigma^*$ such that $L \subseteq w_1^* \cdots w_n^*$.

THEOREM 4.1. *Each Parikh slender context-free language is bounded.*

Proof. If $L \subseteq \Sigma^*$ is context-free and Parikh slender, L is sparse. Hence the claim follows by Theorem 3 of Raz [17]. ▌

The converse of Theorem 4.1 is not true. Indeed, $a^*b^*a^*$ is a bounded context-free language which is not Parikh slender. The characterization of Parikh slender context-free languages among the bounded languages is an open problem.

Note that Theorem 4.1 implies the decidability of many problems concerning Parikh slender context-free languages. For example, the equivalence problem for Parikh slender context-free languages is decidable (see Ginsburg [9]).

To proceed we need a result concerning bounded **Z**-algebraic series with commuting variables. The proof relies heavily on earlier deep results due to Kuich and Salomaa [12] and Semenov [21].

THEOREM 4.2. *Suppose $r \in \mathbf{Z}^{\text{alg}}\langle\langle \Sigma^* \rangle\rangle$ satisfies $|(c(r), w)| \leqslant k$ for every $w \in \Sigma^{\oplus}$, where k is a constant. Then the series $c(r) \in \mathbf{Z}^{\text{alg}}\langle\langle \Sigma^{\oplus} \rangle\rangle$ is a finite* **Z**-*linear combination of series in $\mathbf{N}^{\text{rat}}\langle\langle \Sigma^{\oplus} \rangle\rangle$ of the form $uv_1^* \cdots v_m^*$, where $u, v_1, ..., v_m \in \Sigma^{\oplus}$.*

Proof. Without loss of generality we assume that r is quasiregular. By Corollary 16.11 of Kuich and Salomaa

[12] there exists a nonzero polynomial $P(x_1, ..., x_n, y) \in \mathbf{Z}\langle (\Sigma \cup y)^{\oplus} \rangle$ such that

$$P(x_1, ..., x_n, c(r)) = 0. \tag{4}$$

(Here $\Sigma = \{x_1, ..., x_n\}$.) Next, fix an integer j, $-k \leqslant j \leqslant k$, and denote

$$D_j = \{(i_1, ..., i_n) \in \mathbf{N}^n \mid (c(r), x_1^{i_1} \cdots x_n^{i_n}) = j\}.$$

To study the properties of the set D_j, choose a large prime p and denote by v the canonical morphism

$$v\colon \mathbf{Z}\langle\langle \Sigma^{\oplus} \rangle\rangle \to \mathbf{Z}_p\langle\langle \Sigma^{\oplus} \rangle\rangle.$$

Define the sequence s: $\mathbf{N}^n \to \mathbf{Z}_p$ by

$$s(i_1, ..., i_n) = (v(c(r)), x_1^{i_1} \cdots x_n^{i_n}).$$

It follows by (4) that

$$v(P)(x_1, ..., x_n, v(c(r))) = 0$$

or

$$v(P)\left(x_1, ..., x_n, \sum_{i_1, ..., i_n \geqslant 0} s(i_1, ..., i_n)\, x_1^{i_1} \cdots x_n^{i_n}\right) = 0.$$

Hence the sequence s is p-algebraic. By Theorem 5.1 in Bruyère *et al.* [4] the sequence s is p-recognizable. Consequently the set D'_j defined by

$$D'_j = \{(i_1, ..., i_n) \in \mathbf{N}^n \mid (c(r), x_1^{i_1} \cdots x_n^{i_n}) \equiv j \pmod{p}\}$$

is a p-recognizable subset of $\mathbf{N}^n$. Because clearly $D_j = D'_j$, the set D_j is a p-recognizable subset of $\mathbf{N}^n$.

Now, by replacing in the argument above the prime p by another large prime q it follows that D_j is also q-recognizable. Therefore, by a deep result of Semenov [21], the set D_j is a rational subset of $\mathbf{N}^n$. Denote

$$E_j = \{x_1^{i_1} \cdots x_n^{i_n} \mid (i_1, ..., i_n) \in D_j\}.$$

Clearly, E_j is a rational subset of $\Sigma^{\oplus}$. Because $\Sigma^{\oplus}$ is a commutative monoid, E_j is an unambiguous rational subset of $\Sigma^{\oplus}$ (see Eilenberg and Schützenberger [7]). It follows that

$$\operatorname{char}(E_j) \in \mathbf{N}^{\mathrm{rat}}\langle\langle \Sigma^{\oplus} \rangle\rangle.$$

Hence $\operatorname{char}(E_j)$ is an $\mathbf{N}$-linear combination of series of the form $uv_1^* \cdots v_m^*$, where $u, v_1, ..., v_m \in \Sigma^{\oplus}$. Because $c(r)$ is a finite $\mathbf{Z}$-linear combination of $\operatorname{char}(E_j)$, $-k \leqslant j \leqslant k$, the claim follows. ▮

Theorem 4.2 implies a necessary condition for Parikh slender series in $\mathbf{Z}^{\mathrm{alg}}\langle\langle \Sigma^* \rangle\rangle \cap \mathbf{N}\langle\langle \Sigma^* \rangle\rangle$.

Theorem 4.3. *Suppose $r \in \mathbf{Z}^{\mathrm{alg}}\langle\langle \Sigma^* \rangle\rangle \cap \mathbf{N}\langle\langle \Sigma^* \rangle\rangle$ is Parikh slender. Then $c(r)$ is a finite $\mathbf{N}$-linear combination of series of the form $uv_1^* \cdots v_m^*$, where $u, v_1, ..., v_m \in \Sigma^{\oplus}$ and $\psi(v_1), \ldots, \psi(v_m)$ are linearly independent over $\mathbf{Q}$.*

If L is a language, the generating function $f_L(z)$ of L is defined by

$$f_L(z) = \sum_{n \geqslant 0} l_n z^n,$$

where

$$l_n = \operatorname{card}(\{w \in L \mid |w| = n\})$$

for $n \geqslant 0$.

Theorem 4.4. *Suppose $L \subseteq \Sigma^*$ is a Parikh slender unambiguous context-free language. Then $f_L(z)$ is a rational function.*

Proof. The claim follows by Theorem 4.3. ▮

Next we discuss decision problems concerning Parikh slenderness of context-free languages and algebraic series. The decidability status of the following questions is open:

(1) Is a given context-free language Parikh slender?

(2) Is a given unambiguous context-free language Parikh slender?

(3) Is a given series in $\mathbf{Z}^{\mathrm{alg}}\langle\langle \Sigma^* \rangle\rangle \cap \mathbf{N}\langle\langle \Sigma^* \rangle\rangle$ Parikh slender?

Clearly, the decidability of (1) or (3) implies the decidability of (2). To prove the decidability of (1) it is enough to give a method to decide whether or not a given bounded context-free language is Parikh slender. The decidability of (3) would follow if there is a method to decide whether or not a series $r \in \mathbf{Z}^{\mathrm{rat}}\langle\langle \Sigma^{\oplus} \rangle\rangle$ is bounded. Indeed, if $r \in \mathbf{Z}^{\mathrm{alg}}\langle\langle \Sigma^* \rangle\rangle \cap \mathbf{N}\langle\langle \Sigma^* \rangle\rangle$ is Parikh slender, then by Theorem 4.3, the series $c(r)$ is in $\mathbf{Z}^{\mathrm{rat}}\langle\langle \Sigma^{\oplus} \rangle\rangle \cap \mathbf{N}\langle\langle \Sigma^{\oplus} \rangle\rangle$. Because the condition $c(r) \in \mathbf{Z}^{\mathrm{rat}}\langle\langle \Sigma^{\oplus} \rangle\rangle$ is decidable (see Kuich and Salomaa [12]), the decidability of (3) follows if there is a method to decide whether $c(r)$ is bounded.

A similar argument shows that for the decidability of (3) it would also suffice to have a method to decide whether or not a given $r \in \mathbf{Z}^{\mathrm{rat}}\langle\langle \Sigma^{\oplus} \rangle\rangle$ belongs to $\mathbf{N}^{\mathrm{rat}}\langle\langle \Sigma^{\oplus} \rangle\rangle$.

We conclude this section with a positive decidability result.

Theorem 4.5. *It is decidable whether or not a given series $r \in \mathbf{Z}^{\mathrm{alg}}\langle\langle x^* \rangle\rangle$ is bounded. Consequently, it is decidable whether or not a series $r \in \mathbf{Z}^{\mathrm{alg}}\langle\langle x^* \rangle\rangle \cap \mathbf{N}\langle\langle x^* \rangle\rangle$ is Parikh slender. (Here $x \in \Sigma$ is a letter.)*

Proof. If r is bounded, Theorem 4.2 implies that r is in $\mathbf{Z}^{\text{rat}}\langle\langle x^* \rangle\rangle$. This property is decidable. The claim follows because it is decidable whether or not a series $r \in \mathbf{Z}^{\text{rat}}\langle\langle x^* \rangle\rangle$ is bounded (see Jacob [11]). ▮

5. APPLICATIONS TO AMBIGUITY PROOFS OF CONTEXT-FREE LANGUAGES

The purpose of this section is to show that the results of Section 4 give a new method to prove the inherent ambiguity of many context-free languages. In particular, we obtain a new simple proof of a well-known theorem of Autebert, Flajolet and Gabarro concerning the prefixes of infinite words.

If $L \subseteq \Sigma^*$ is a language, the *length set* of L consists of the nonnegative integers n such that there is a word of length n in L.

In what follows we denote $L^c = \Sigma^* - L$.

Theorem 5.1. *Suppose* $L \subseteq \Sigma^*$ *is a context-free language. If there is a regular language* $R \subseteq \Sigma^*$ *such that the language* $L \cap R$ (*or* $L^c \cap R$) *is Parikh slender and has a non-rational generating function, then* L *is inherently ambiguous. In particular, if* $L^c \cap R$ *is Parikh slender and has a length set which is not ultimately periodic then* L *is inherently ambiguous.*

Proof. Suppose on the contrary that $L \subseteq \Sigma^*$ is an unambiguous context-free language such that $L^c \cap R$ is Parikh slender, where $R \subseteq \Sigma^*$ is a regular language. (If $L \cap R$ is Parikh slender the argument is similar.) Then $\operatorname{char}(L) \in \mathbf{N}^{\text{alg}}\langle\langle \Sigma^* \rangle\rangle$ and, hence, the series

$$r = \operatorname{char}(L^c \cap R) = \operatorname{char}(L^c) \odot \operatorname{char}(R)$$

belongs to $\mathbf{Z}^{\text{alg}}\langle\langle \Sigma^* \rangle\rangle$. Because r is Parikh slender it follows by Theorem 4.3 that the generating function of $L^c \cap R$ is rational. This contradiction proves the theorem. ▮

As an example consider the *Goldstine language* L over the alphabet $\Sigma = \{a, b\}$ consisting of the words

$$a^{n_1} b a^{n_2} b \cdots a^{n_p} b,$$

where $p \geqslant 1$, $n_i \geqslant 0$, and $n_j \neq j$ for some j, $1 \leqslant j \leqslant p$. L is a context-free language which is inherently ambiguous (see Flajolet [8]). The ambiguity of L follows also by Theorem 5.1. Indeed, let $R = \Sigma^* b$. Then

$$L^c \cap R = \{ab, aba^2b, aba^2ba^3b, \ldots\}$$

which clearly does not have an ultimately periodic length set.

If $w \in \Sigma^\omega$ is an infinite word and

$$w = w_1 w_2 w_3 \cdots \qquad w_j \in \Sigma,$$

its *prefix language* and *coprefix language* are defined by

$$\operatorname{Pref}(w) = \{w_1, w_2 \cdots w_m \mid m \geqslant 0\}$$

and

$$\operatorname{Copref}(w) = \Sigma^* - \operatorname{Pref}(w),$$

respectively.

Theorem 5.2. (Autebert, Flajolet, and Gabarro [2]). *Suppose* w *is an infinite word which is not ultimately periodic. If the language* $L = \operatorname{Copref}(w)$ *is context-free, it is inherently ambiguous.*

Proof. Because w is not ultimately periodic, there is a letter $a \in \Sigma$ such that $L^c \cap \Sigma^* a$ has a length set which is not ultimately periodic. Clearly $L^c \cap \Sigma^* a$ is Parikh slender. Therefore the claim follows by Theorem 5.1. ▮

Note that the proof of Theorem 5.2 given above is purely algebraic. The original proof in Autebert, Flajolet, and Gabarro [2] uses deep analytic tools.

The proof of Theorem 5.2 can be further simplified by replacing the use of Semenov's theorem by an application of Cobham's theorem stating that if a set A of natural numbers is both p-recognizable and q-recognizable, then A is ultimately periodic provided that p and q are multiplicatively independent (see Eilenberg [6]). Indeed, suppose $L = \operatorname{Copref}(w)$ is an unambiguous context-free language and denote $R = \Sigma^* a$, where $a \in \Sigma$. Then $r = \operatorname{char}(L^c \cap R)$ and the generating function of r are $\mathbf{Z}$-algebraic. It follows that the length set of $\operatorname{supp}(r)$ is p-recognizable for any large prime p. Hence, the length set of $\operatorname{supp}(r)$ is ultimately periodic, which implies that w is ultimately periodic.

ACKNOWLEDGMENT

The author is deeply grateful to Professor Arto Salomaa for useful discussions concerning this paper.

REFERENCES

1. M. Andraşiu, J. Dassow, G. Păun, and A. Salomaa, Language-theoretic problems arising from Richelieu cryptosystems, *Theoret. Comput. Sci.* **116** (1993), 339–357.
2. J.-M. Autebert, P. Flajolet, and J. Gabarro, Prefixes of infinite words and ambiguous context-free languages, *Inform. Process. Lett.* **25** (1987), 211–216.

3. J. Berstel and C. Reutenauer, "Rational Series and Their Languages," Springer-Verlag, Berlin, 1988.
4. V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belgian Math. Soc.* **1** (1994), 191–237.
5. J. Dassow, G. Păun, and A. Salomaa, On thinness and slenderness of L languages, *EATCS Bull.* **49** (1993), 152–158.
6. S. Eilenberg, "Automata, Languages and Machines," Vol. A, Academic Press, New York, 1974.
7. S. Eilenberg and M. P. Schützenberger, Rational sets in commutative monoids, *J. Algebra* **13** (1969), 173–191.
8. P. Flajolet, Analytic models and ambiguity of context-free languages, *Theoret. Comput. Sci.* **49** (1987), 283–309.
9. S. Ginsburg, "The Mathematical Theory of Context-Free Languages," McGraw–Hill, New York, 1966.
10. L. Ilie, On a conjecture about slender context-free languages, *Theoret. Comput. Sci.* **132** (1994), 427–434.
11. G. Jacob, La finitude des représentations linéaires des semi-groupes est décidable, *J. Algebra* **52** (1978), 437–459.
12. W. Kuich and A. Salomaa, "Semirings, Automata, Languages," Springer-Verlag, Berlin, 1986.
13. T. Nishida and A. Salomaa, Slender 0L languages, *Theoret. Comput. Sci.*, to appear.
14. G. Păun and A. Salomaa, Decision problems concerning the thinness of DOL languages, *EATCS Bull.* **46** (1992), 171–181.
15. G. Păun and A. Salomaa, Closure properties of slender languages, *Theoret. Comput. Sci.* **120** (1993), 293–301.
16. G. Păun and A. Salomaa, Thin and slender languages, *Discrete Appl. Math.* **61** (1995), 257–270.
17. D. Raz, Length considerations in context-free languages, *Theoret. Comput. Sci.*, to appear.
18. G. Rozenberg and A. Salomaa, "The Mathematical Theory of L Systems," Academic Press, New York, 1980.
19. A. Salomaa, "Formal Languages," Academic Press, New York, 1973.
20. A. Salomaa and M. Soittola, "Automata-Theoretic Aspects of Formal Power Series," Springer, Berlin, 1978.
21. A. L. Semenov, Presburgerness of predicates regular in two number systems, *Sibirsk. Mat. Zh.* **18** (1977), 403–418 [Russian]. Engl. transl. *Siberian Math. J.* **18** (1977), 289–299.
22. J. Shallit, Numeration systems, linear recurrences, and regular sets, *Inform. Comput.* **113** (1994), 331–347.