数理解析研究所講究録
910 巻 1995 年 67-74

A few notes on slender 0L languages

(Extended abstract)

Taishin Y. Nishida

西田 泰伸

Faculty of Engineering
Toyama Prefectural University
5180 Kurokawa, Kosugi-machi, Toyama 939-03, Japan
e-mail nishida@pu-toyama.ac.jp

富山県立大学工学部
富山県射水郡小杉町黒河 5180

1. Introduction

The purpose of this paper is to proceed with the classification of slender 0L languages. Slender languages are introduced in [1] and many decision problems for them are investigated in [10]. Slender 0L languages are considered in [2], [9], and [12]. Since all finite 0L languages and D0L languages are known to be slender [9], we consider nondeterministic and infinite 0L systems.

The first result is that all quasi-deterministic 0L systems (or D$'$0L systems for short, cf. [6]) are slender (Theorem 2.1). Next we establish a necessary condition for slender 0L systems. To describe the condition, we first establish the notion of unbounded letters. Let $G = \langle \Sigma, \tau, w \rangle$ be a 0L system. A letter $a \in \Sigma$ is said to be unbounded if, for any integer C, $L(G)$ has a word u in which a occurs more than C. Then, if $L(G)$ is slender, G satisfies the next condition (Theorem 3.1):

For every unbounded letter a and for every positive integer n, there exist $zz' \in \Sigma^*$ and a finite set I of nonnegative integers such that $\tau^n(a) = \{(zz')^i z \mid i \in I\}$.

This condition is not sufficient (see Examples 2.3 and 3.1 below). Putting z the empty word and $I = \{1\}$, every deterministic letter trivially satisfies this condition. Of course, every D$'$0L system satisfies this condition because all unbounded letters for a D$'$0L system are deterministic (cf. [6]). If a word has n occurrences of a nondeterministic letter, then there are at least n possibilities in making words of the same length by substituting them. But these possibilities must yield the same word in a slender 0L language. Therefore the condition forces for nondeterministic unbounded letters to have a restricted form on the right hand side of the substitution, which makes it to be decidable whether they satisfy this condition or not.

We establish a subclass of slender 0L languages, other than D$'$0L systems, which are effectively decidable. This class, called ultimately extended free-generated, is based on nondeterministic unbounded letters (Theorem 4.3). The characterization of slender 0L languages is, however, not completed. For example, it is an open problem to decide for a given 0L system which is not D′0L and does not have any nondeterministic unbounded letters whether or not it generates a slender language.

2. Slender 0L systems and D′0L systems

Slender languages, first introduced by Andraşiu *et al.* [1], are defined as follows:

Definition ([1]). Let L be a language over Σ and k be a positive integer.

(i) L is said to be *thin* if, for some n_0,

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

(ii) L is said to be *k-thin* if, for some n_0,

$$\mathrm{card}(\{w \in L \mid |w| = n\}) \leq k \quad \text{whenever} \quad n_0 \leq n.$$

(iii) L is called *slender* if it is k-thin for some k. □

Since we are interested in slender 0L languages, we mention 0L systems briefly. We assume the reader to be familiar with 0L systems and their variations (see [3] and [11]).

Definition. A *0L system* G is a triple $\langle \Sigma, \tau, w \rangle$, where Σ is a finite alphabet, τ is a substitution on Σ^*, and w is a word over Σ called the *axiom* of G. If τ is an endomorphism on Σ^* rather than a substitution, then G is called a *deterministic 0L system* or a *D0L system* for short. The language generated by G is denoted by $L(G)$ and defined by $L(G) = \tau^*(w)$. A language L is called a 0L language if it is generated by some 0L system. □

We call a 0L system slender if it generates a slender language.

Let $\langle \Sigma, \tau, w \rangle$ be a 0L system. We will use the following classification of letters in Σ in the sequel.

(1) A letter a is said to be *deterministic* if the restriction of τ on $\mathrm{alph}(\tau^*(a))$† is an endomorphism, in other words, a and every descendant of a have only one descendant. A letter which is not deterministic is called *nondeterministic*.

(2) A letter a is called *infinite* if $\mathrm{card}(\tau^*(a)) = \infty$. A letter a is called *finite* if $\mathrm{card}(\tau^*(a)) < \infty$.

(3) A letter a is said to be *persistent* if $uav \in \tau^n(a)$ for some $uv \in \Sigma^*$ and $n > 0$. The smallest positive integer k satisfying $uav \in \tau^k(a)$ is called the *period* of a.

(4) A letter which always generates 1 (the empty word is denoted by 1) is called *mortal*. We note that $\tau^m(a) = 1$ for every mortal letter a where $m = \mathrm{card}(\Sigma)$.

(5) A persistent letter a is said to be *self-productive* if it has a descendant which has more than one occurrences of a, i.e., $u \in \tau^+(a)$ such that $|u|_a \geq 2$‡.

(6) A persistent letter a is said to be *stem* if a satisfies the following conditions:

† $\mathrm{alph}(L)$ denotes the set of all and only letters appearing in the words of L.

‡ $|u|_a$ denotes the number of occurrences of a in u.

(i) a is not self-productive.

(ii) $uav \in \tau^n(a)$ for some $n > 0$ such that uv has occurrences of some persistent letters.

We note that a letter is effectively decided whether it belongs each class defined above or not. Finally we define unbounded letters.

Definition. Let $G = \langle \Sigma, \tau, w \rangle$ be a 0L system. A letter $a \in \Sigma$ is said to be *unbounded* in $L(G)$ if, for every integer N, there exists a words w in $L(G)$ such that $|w|_a > N$. □

A letter is effectively decided whether it is unbounded or not by the following considerations. A letter b is an *unbounded product* of a if

(i) a is self-productive and b is a descendant of a,

(ii) a is a stem letter and b is persistent which occurs in uv where $uav \in \tau^k(a)$ for some $k > 0$, or

(iii) b is a descendant of a letter which satisfies (ii).

Now a letter a is unbounded in G if and only if a is an unbounded product of $c \in \mathrm{alph}(\cup_{n=0}^{m} \tau^n(w))$ where w is the axiom and $m = \mathrm{card}(\Sigma)$.

The notion of quasi-deterministic 0L systems somewhat resembles that of slender 0L systems (therefore in [12] it is called *derivation slender*). Let $\mathbb{N}$ and $\mathbb{N}_+$ denote the sets of nonnegative and positive integers, respectively.

Definition ([6]). Let $G = \langle \Sigma, \tau, w \rangle$ be a 0L system. Then G is said to be *quasi-deterministic* 0L system, or D′0L system for short, if there is a positive integer k and G satisfies

$$\mathrm{card}(\tau^n(w)) \leq k$$

for every $n \in \mathbb{N}$. □

There is a close relation between slender 0L systems and D′0L systems.

Theorem 2.1. *Every D′0L language is slender.*

Since every D′0L language is an HD0L language [7], this theorem follows from the next lemma.

Lemma 2.2. *Every HD0L language is slender.*

Proof. Since $\mathcal{L}(\mathrm{HD0L}) = \mathcal{L}(\mathrm{CPFD0L})$ (see [5]), we will show every CPFD0L language is slender. Let L be a CPFD0L language and let $G = \langle \Sigma, \theta, \Gamma, h, F \rangle$ be a CPFD0L system which generates L. Then

$$L = \theta(h^*(F)) = \bigcup_{w_i \in F} \theta(h^*(w_i))$$

provided $F = \{w_1, w_2, \ldots, w_n\}$ where $w_i \in \Gamma^*$ with $i = 1, 2, \ldots, n$. Now $\theta(h^*(w_i))$ is a CPD0L language generated by the CPD0L system $G_i = \langle \Sigma, \theta, \Gamma, h, w_i \rangle$. Therefore L is slender because every PD0L language is slender [9] and the family of slender languages is closed under coding and finite union [10]. □

There are 0L systems which are not D'0L and have no nondeterministic unbounded letters. Some of these systems are slender and the others are not, that is, the converse of Theorem 2.1 is not true.

Example 2.1. Let $G_1 = \langle \{a,b,c\}, \tau_1, a\rangle$ where $\tau_1(a) = \{ab, a, cb\}$, $\tau_1(b) = b$, and $\tau_1(c) = cb$. Then G_1 is not D'0L while $L_1 = L(G_1) = \{ab^i \mid i \in \mathbb{N}\} \cup \{cb^i \mid i \in \mathbb{N}_+\}$ is slender. But L_1 is a D'0L language generated by $G_1' = \langle \{a,b,c\}, \tau_1', a\rangle$ where $\tau_1'(a) = \{ab, cb\}$, $\tau_1'(b) = b$, and $\tau_1'(c) = cb$. □

Example 2.2. Let $G_2 = \langle \{a,b\}, \tau_2, a\rangle$ where $\tau_2(a) = \{bab, b\}$ and $\tau_2(b) = b$. Then G_2 is not D'0L and $L_2 = L(G_2) = \{b^i ab^i \mid i \in \mathbb{N}\} \cup \{b^{2i+1} \mid i \in \mathbb{N}\}$ is slender. But L_2 is not a D'0L language.

Assume that a D'0L system $G_2' = \langle \{a,b\}, \tau_2', w\rangle$ generates L_2. Then $\tau_2'(b) = b^n$ for some $n > 0$ and a is not erased for otherwise b^{2i} is contained in $\tau_2'^*(w)$. Since no word derives shorter words, a is the axiom. If $\tau(b) = b$, then all words in b^+ are repeatable which is impossible with D'0L system (cf. [6]). If $\tau_2'(a)$ has two words $b^i ab^i$ and $b^j ab^j$ with $j < i$, then $\tau_2'^k(a)$ contains at least k words, namely $b^x ab^x$ where $x = \frac{in^{k-l}(n^l-1)+j(n^{k-l}-1)}{n-1}$ for $l = 0, 1, \ldots, k-1$. Again G_2' is not D'0L. Thus $\tau_2'(a) \cap b^*ab^* = bab$. But now $bbabb$ cannot be generated because $\tau_2'(bab) \cap b^*ab^* = b^{n+1}ab^{n+1}$ and $n+1 > 2$. □

Example 2.3. Let $G_3 = \langle \{a,b,c\}, \tau_3, a\rangle$ where $\tau_3(a) = \{bab, cac\}$, $\tau_3(b) = b$, and $\tau_3(c) = c$. Then $L(G_3) = \{waw^R \mid w \in \{b,c\}^*\}$ is not slender. □

As the above examples suggest, it would be difficult for a 0L system which is not D'0L and has no nondeterministic unbounded letters to decide whether or not it generates a slender language. Or the problem might be undecidable. Anyway, the problem is left for future investigation.

3. A necessary condition for slender 0L systems

Theorem 3.1. *If a 0L system $G = \langle \Sigma, \tau, w\rangle$ is slender, then G satisfies the following condition:*

For every unbounded letter a in Σ and for every $n \in \mathbb{N}_+$ there exist $zz' \in \Sigma^+$ and a finite set $I \subset \mathbb{N}$ such that $\tau^n(a) = \{(zz')^i z \mid i \in I\}$. (3.1)

Proof. Let x, y be words in $\tau^n(a)$. Let w be a word in $L(G)$ with k occurrences of a, i.e., $w = u_0 a u_1 a \cdots a u_k$ for some $u_0 u_1 \cdots u_k \in \Sigma^*$. Let u_i' be a word in $\tau^n(u_i)$ for every $i = 0, 1, \ldots, k$. Then $\tau^n(w)$ contains the next subset:

$$L(w,n) = \{u_0' y u_1' x \cdots x u_k',\ u_0' x u_1' y \cdots x u_k',\ \ldots,\ u_0' x u_1' x \cdots y u_k'\},$$

that is, $L(w,n)$ consists of the words which are obtained by substituting one occurrence of a in w with y and the others with x. Obviously all words in $L(w,n)$ have the same length. Because G is slender, the cardinality of $L(w,n)$ is bounded. Then we have $xuy = yux$ for some $u \in \Sigma^*$ and hence $uxuy = uyux$. By Lyndon and Schützenberger's theorem on free monoids, we have that if $xy = yx$ for some $x, y \in \Sigma^+$, then $x = z^m$ and $y = z^n$ for some $z \in \Sigma^+$ and $m, n \in \mathbb{N}_+$ (see [4] and [13]). Now the theorem immediately follows. □

This theorem, unfortunately, does not give sufficient condition for a 0L system to generate slender language. In addition to Example 2.3, the following example show that the converse of Theorem 3.1 does not hold.

Example 3.1. Let us consider $\tau(a) = abc$, $\tau(b) = \{b, 1\}$, $\tau(c) = c$. The letters b and c are unbounded and satisfy the condition of Theorem 3.1. But $\tau^*(a)$ is not slender. □

Condition (3.1) is a statement on all words in $\tau^*(a)$. But we will show that (3.1) is decidable. If a is a deterministic letter, then (3.1) is trivially fulfilled. As for nondeterministic letters, the next proposition guarantees the decidability of condition (3.1). The proof of Proposition 3.2 will be found elsewhere [8].

Proposition 3.2. *(i) Let a be a nondeterministic unbounded letter and let $\tau^n(a) = \{(u_n u'_n)^i u_n \mid i \in I_n\}$ for each $n > 0$. For every $n \in \mathbb{N}_+$, $|u_n u'_n| \leq \alpha^k$ where $\alpha = \max_{w \in \tau(\Sigma)} |w|$ and k is the period of a.*
(ii) A nondeterministic unbounded letter satisfies Condition (3.1) if and only if $\tau^n(a) = \{(u_n u'_n)^i u_n \mid i \in I_n\}$ for every $n \leq (\mathrm{card}(\Sigma))^{\alpha^k}$ and $u_n u'_n = u_m u'_m$ for some $n < m \leq (\mathrm{card}(\Sigma))^{\alpha^k}$.

Proposition 3.2 suggests a characterization of slender 0L systems which have occurrences of nondeterministic unbounded letters. This will be done in the next section

4. Classification of slender 0L systems

In this section, we show that slender 0L systems are classified into four types: D′0L, free-generated, not D′0L and generating no nondeterministic unbounded letters, and mixture of these types. The first type is already known. The second type is defined here and we show that every 0L system is decidable whether it generates free-generated slender language or not.

Definition. Let τ be a substitution over Σ. Let $\Lambda = (L_0, L_1, \ldots)$ be a sequence of languages over Σ such that every L_i is finite and the 'flat language' $\mathcal{U}(\Lambda)$ of Λ given by

$$\mathcal{U}(\Lambda) = \bigcup_{i=0}^{\infty} L_i$$

is infinite.

(i) Λ is said to be *free-generated with respect to* τ if there exist $2k$ words $u_0, u'_0, \ldots, u_{k-1}, u'_{k-1}$ such that

$$L_{i+nk} \subset \{(u_i u'_i)^j u_i \mid j \in \mathbb{N}\}$$

and that

$$\begin{aligned}
\tau(u_i u'_i) &= \{(u_{i+1} u'_{i+1})^j u_{i+1} \mid j \in I_i\} \quad (i < k-1) \\
\tau(u_i) &= \{(u_{i+1} u'_{i+1})^j u_{i+1} \mid j \in I'_i\} \quad (i < k-1) \\
\tau(u_{k-1} u'_{k-1}) &= \{(u_0 u'_0)^j u_0 \mid j \in I_{k-1}\} \\
\tau(u_{k-1}) &= \{(u_0 u'_0)^j u_0 \mid j \in I'_{k-1}\}
\end{aligned}$$

where I_i is a finite subset of $\mathbb{N}$ and $\text{card}(I_i) \geq 2$ at least one i ($i = 0, 1, \ldots, k-1$). The words $u_0, u_0', \ldots, u_{k-1}, u_{k-1}'$ are called the *generating words* of Λ and k is called the *period* of $u_i u_i'$. We call simplely Λ free-generated if τ is understood.

(ii) Λ is said to be *extended free-generated* if there exist a finite language F such that

$$L_i = F_i \bar{L}_i F_i', \quad F_i \subseteq F, \quad \text{and} \quad F_i' \subseteq F$$

for every $0 \leq i$ and that the sequence $(\bar{L}_i)$ is free-generated.

(iii) Λ is called *ultimately (extended) free-generated* if there exists a positive integer N and p (extended) free-generated sequences $(L_i^{(1)}), \ldots, (L_i^{(p)})$ such that

$$L_i = \bigcup_{j=1}^{p} L_{i-N}^{(j)}$$

for every $i \geq N$. □

As shown in the next property, the flat language of a ultimately extended free-generated sequence is slender.

Property 4.1. *Let $(L_i)_{i \geq 0}$ be a ultimately extended free-generated sequence. Then*

$$L = \mathcal{U}((L_i)_{i \geq 0}) = (\bigcup_{i=1}^{n} F_i \{(u_i u_i')^j u_i \mid j \in \mathcal{I}_i\} F_i') \cup F$$

for some finite languages F, F_i, and F_i' ($i = 1, \ldots, n$) and L is slender.

Proof. Let $(K_i)_{i \geq 0}$ be an extended free-generated sequence and let $(\bar{K}_i)_{i \geq 0}$ be the free-generated sequence of (K_i), i.e., $K_i = E_i \bar{K}_i E_i'$ for every $i \in \mathbb{N}$, where E_i and E_i' are subsets of a finite set. Let $u_0, u_0', \ldots, u_{k-1}, u_{k-1}'$ be the generating words of $(\bar{K}_i)$. Then $L_{kn+i} = E_{kn+i}\{(u_i u_i')^j u_i \mid j \in I_{kn+i}\} E_{kn+i}'$ for $0 \leq n$ and $0 \leq i < k$ and hence

$$\bigcup_{n=0}^{\infty} L_{kn+i} = F_i \{(u_i u_i')^j u_i \mid j \in \mathcal{I}_i\} F_i'$$

where $F_i = \cup_{n=0}^{\infty} E_{kn+i}$ and $F_i' = \cup_{n=0}^{\infty} E_{kn+i}'$ are finite and $\mathcal{I}_i = \bigcup_{n=0}^{\infty} I_{kn+i}$. Then we have

$$\mathcal{U}((K_i)_{i \geq 0}) = \bigcup_{i=1}^{k} F_i \{(u_i u_i')^j u_i \mid j \in \mathcal{I}_i\} F_i'$$

and it is obviously slender. Since a ultimately extended free-generated sequence is a finite union of extended free-generated sequences after an initial mess, the conclusion follows immediately. □

By the definition of extended free-generated sequences, the next property, which give a necessary and sufficient condition for a catenation of two extended free-generated sequences to be extended free-generated, is obvious.

Property 4.2. *Let $(L_{1,j})_{j\geq 0}$ and $(L_{2,j})_{j\geq 0}$ be extended free-generated sequences with the following descriptions ($i = 1, 2$)*

$$L_{i,k_in+j} = F_{i,j}\bar{L}_{i,k_in+j}F'_{i,j}$$

and let $u_{i,0}, u'_{i,0}, \ldots, u_{i,l_i}, u'_{i,l_i}$ be the generating words of $(\bar{L}_{i,j})_{j\geq 0}$ for $i = 1, 2$, respectively. Then $(L_{1,j}L_{2,j})$ is extended free-generated if and only if $l_1 = l_2$, $F'_{1,j} = F_{2,j} = 1$ for $0 \leq j \leq \max(k_1, k_2)$, and $u'_{1,j}u_{1,j} = u_{2,j}u'_{2,j}$ for $j = 0, \ldots, l_1$. □

Example 4.1. *Let $\tau(a) = \{aba, a\}$, $\tau(b) = b$, and $\tau(c) = a$ be a substitution over $\{a, b, c\}$. Then $(\tau^i((ab)^j))_{i\geq 0}$, $(\tau^i((ba)^j))_{i\geq 0}$, and $(\tau^i((ab)^j a))_{i\geq 0}$ are free-generated for every $j \in \mathbb{N}$ and c is ultimately free-generated.* □

Theorem 4.3. *Let $G = \langle \Sigma, \tau, w\rangle$ be a 0L system.*

(i) If $L(G)$ is slender, then one of the following conditions holds:

(1) G is D′0L.

(2) $(\tau^i(w))_{i\geq 0}$ is a ultimately extended free-generated sequence.

(3) $L(G)$ has no nondeterministic unbounded letters but G is not D′0L.

(4) w has a factorization $w = w_1w_2\ldots w_l$ such that, for every $i \in \{1, 2, \ldots, l\}$, the 0L system $G_i = \langle \Sigma, \tau, w_i\rangle$ generates a slender language of type (1), (2), or (3).

(ii) G is effectively decidable whether or not it is type (1) or (2).

Type (1) is already known. The proof for the other assertions is lengthy and will appear in [8].

At the end of this section, we give some examples of type (2) and type (4) slender languages.

Example 4.2. Let $G = \langle \{a, b, c\}, \tau, w\rangle$ be a 0L system where τ is given in Example 4.1. Then $L(G)$ is slender if the axiom w is of the form: $w = (ab)^j$, $w = (ba)^j$, or $w = (ab)^j a$ for some $j \in \mathbb{N}$ or $w = c$. But the other cases, for example, if $w = cc$, then $L(G)$ is not slender. □

Example 4.3. Let $G = \langle \{a, b, c\}, \tau, abc\rangle$ be a 0L system where $\tau(a) = \{aba, a\}$, $\tau(b) = b$, and $\tau(c) = c^5$. Then

$$\tau^i(abc) = \{(ab)^j c^{5^i} \mid j \in \{1, 2, \ldots, 2^i\}\}$$

for every $i \in \mathbb{N}$. Because the length sets of $\tau^k(a)$ and $\tau^l(a)$ do not overlap for every $k \neq l$, $L(G)$ is slender, indeed it is thin. □

References

[1] Mircea Andraşiu, G. Păun, J. Dassow, and A. Salomaa, Language-theoretic problems arising from Richelieu cryptosystems, *Theoretical Computer Science*, (1993) 339–357.

[2] J. Dassow, G. Păun, and A. Salomaa, On thinness and slenderness of L languages, *EATCS Bulletin* 49 (1993) 152–158.

[3] G. T. Herman and G. Rozenberg, *Developmental systems and languages*, (North-Holland, Amsterdam, 1975).

[4] R. C. Lyndon and M. P. Schützenberger, The equation $a^M = b^N c^P$ in a free group, *Michigan Math. J.*, (1962) 289–298.

[5] M. Nielsen, G. Rozenberg, A. Salomaa, and S. Skyum, Nonterminals, homomorphisms and codings in different variations of 0L-systems, I and II, *Acta Informatica* **3** (1974) 357–364; **4** (1974) 87–106.

[6] T. Nishida, Quasi-deterministic 0L systems, In: *Proc. ICALP '92*, Lecture Notes in Computer Science, Vol. 623, (Springer-Verlag, Berlin 1992) 65–76.

[7] T. Nishida, Quasi-deterministic 0L systems and their representation, *Theoretical Computer Science*, to appear.

[8] T. Nishida and A. Salomaa, Slender 0L languages, *Theoretical Computer Science*, to appear.

[9] G. Păun and A. Salomaa, Decision problems concerning the thinness of D0L languages *EATCS Bulletin* 46 (1992) 171–181.

[10] G. Păun and A. Salomaa, Closure properties of slender languages, *Theoretical Computer Science*, (1993) 293–301.

[11] G. Rozenberg and A. Salomaa, *The mathematical theory of L systems*, (Academic Press, New York 1980).

[12] A. Salomaa, Slenderness and immigration: New aspects of L systems, In: *Proceedings of the 9th International Conference of Automata and Formal Languages, Salgotarjan, Hungary*, to appear.

[13] H. J. Shyr, *Free Monoids And Languages*, Lecture Notes, National Chung-Hsing University, Taiwan, 1991.