Theoretical Computer Science 132 (1994) 427–434
Elsevier

Note

On a conjecture about slender context-free languages

Lucian Ilie
Faculty of Mathematics, University of Bucharest, Str. Academiei No. 14, 70109 București, Romania

Communicated by A. Salomaa
Received October 1993

Abstract

Ilie, L., On a conjecture about slender context-free languages, Theoretical Computer Science 132 (1994) 427–434.

We prove that every slender context-free language is a union of paired loops, thus confirming a conjecture of Păun and Salomaa to appear. A series of consequences of this result are inferred, most of them also left as open problems in recent papers about slender languages.

1. Slender languages

In a formal language variant of the classic Richelieu cryptosystem (hiding the message by shuffling it with some garbage text — see details in [7]), in [1], one considers *the slender* languages, namely languages for which the number of strings of every given length is bounded from above. Formally, let us denote by $|x|$ the length of a string $x \in V^*$ (V^* is the free monoid generated by the alphabet V under the operation of concatenation: the empty string is denoted by λ). A language L is said to be k-slender if $card\{w \in L \mid |w| = n\} \leqslant k$, for every $n \geqslant 0$. A language is slender if it is k-slender for some natural number k. A 1-slender language is also called *thin* language.

Such languages are useful in the cryptographic frame described in [1] in the key management: in order to rediscover the message from the cryptotext, a key of the same

Correspondence to: L. Ilie, Institute of Mathematics, University of Bucharest, Str. Academiei No. 14, 70109 Bucharest, Romania.

SSDI 0304-3975(94)00042-H

length with the cryptotext must be used; if the set of keys is a slender language, then only its grammar must be known by the legal receiver; by checking all the at most k strings of a given length (only one of them is the key), the receiver can decrypt. (Further details can be found in [1].)

The slender languages have not only good motivations, but they also raise interesting formal language theory questions. The papers [1–5] contain a series of results in this area. One of the main problems about slender context-free languages concerns their characterization. In [1] it is proved that every slender unambiguous context-free language is linear, and that slenderness is decidable for unambiguous context-free languages, and it is conjectured that this is true for all slender context-free languages. Then, in [4] the following characterization of slender regular languages is given: a regular language L is slender if and only if it is *union of single loops*, i.e., it is of the form $L=\bigcup_{i=1}^{k} u_i v_i^* w_i$, for some given strings u_i, v_i, w_i, $1 \leqslant i \leqslant k$, k, a natural number.

It is conjectured in [4] that a similar characterization holds for context-free slender languages, considering *paired loops*: a language L is said to be *union of paired loops* (UPL, in short) iff, for some $k \geqslant 1$ and strings u_i, v_i, w_i, x_i, y_i, $1 \leqslant i \leqslant k$, we have

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

A UPL language is called *disjoint union of paired loops* (DUPL) if the sets $\{u_i v_i^n w_i x_i^n y_i \mid n \geqslant 0\}$ in the previous equality are disjoint.

Păun and Salomaa [4, Theorem 4.1] show that every UPL language is a DUPL language. As every UPL language is linear and slender, and every DUPL language is unambiguous, it follows that every UPL language is a slender unambiguous linear language. The *conjecture* in [4] is that every slender context-free language is a UPL language, that is a context-free language is slender if and only if it is a UPL language (hence linear unambiguous). This conjecture is then related to several decidability and closure properties of slender languages ([4, 5]).

We shall confirm here the conjecture in [4], and then we shall point out some of its consequences.

2. The main result

Theorem 2.1. *Every slender context-free language is a UPL language.*

Proof. Let $L \subseteq V^*$ be a k-slender context-free language. According to Bar–Hillel pumping lemma, there are $p, q \in \mathbb{N}$ such that every $z \in L$ with $|z| > p$ can be written in the form $z = uvwxy$ and

$$|vwx| \leqslant q, \tag{1}$$

$$vx \neq \lambda, \tag{2}$$

$$uv^n wx^n y \in L \quad \text{for all } n \geqslant 0. \tag{3}$$

Consequently, there is a (possibly infinite) set of indices I such that if we denote

$$L_1=\{w\in L \mid |w|\leqslant p\},$$

and, for every $i\in I$,

$$A_i=\{u_i v_i^n w_i x_i^n y_i \mid n\geqslant 0\}$$

for $u_i, v_i, w_i, x_i, y_i \in V^*$, $v_i x_i \neq \lambda$, $|v_i w_i x_i|\leqslant q$, then we have

$$L=L_1\cup L_2,$$

where

$$L_2=\bigcup_{i\in I} A_i.$$

Because L_1 is finite, it is a UPL language. Therefore, it is enough to prove that L_2 is a UPL language (a finite union of UPL languages is a UPL language).

Clearly, we can assume without loss of the generality that for all $i,j\in I$, $i\neq j$, we have

$$A_i\neq A_j, \quad \text{and} \quad A_i\nsubseteq A_j. \tag{4}$$

We begin by proving the following *statement*: if $I_0\subseteq I$ such that for every $i,j\in I_0$, $i\neq j$, the set $A_i\cap A_j$ is finite, then the set I_0 is finite.

In this aim, we shall prove the relation below (which implies that I_0 is finite):

$$card(I_0)\leqslant k(q+1). \tag{5}$$

We denote

$$|u_i w_i y_i|=n_i, \quad |v_i x_i|=m_i, \qquad i\in I.$$

For every $i\in I_0$, the lengths of words in A_i form an arithmetical progression,

$$n_i,\ n_i+m_i,\ n_i+2m_i, \ldots \tag{6}$$

We suppose that $card(I_0)>k(q+1)$ and take a subset I_0' of I_0 such that $card(I_0')=k(q+1)+1$.

Obviously, there are positive integers s such that

$$s>n_i \quad \text{for all } i\in I_0', \tag{7}$$

$$u_i v_i^n w_i x_i^n y_i \neq u_j v_j^m w_j x_j^m y_j \tag{8}$$

for all $n,m\geqslant 0$, with $|u_i v_i^n w_i x_i^n y_i|>s$, and $|u_j v_j^m w_j x_j^m y_j|>s$.

Denote

$$D=\{s, s+1, \ldots, s+q\}.$$

From (1) we have $m_i\leqslant q$ for all $i\in I$, hence, in view of (7) it follows that every arithmetical progression of the form (6) has, for every $i\in I_0'$, at least one element in D. By (8) we obtain that for every $t\in D$ and for every $i,j\in I_0'$, $i\neq j$, if

$$|u_i v_i^n w_i x_i^n y_i|=|u_j v_j^m w_j x_j^m y_j|=t,$$

then

$$u_i v_i^n w_i x_i^n y_i \neq u_j v_j^m w_j x_j^m y_j.$$

Consequently, there exist $card(I_0') = k(q+1)+1$ different strings from L with the lengths in D. This implies that we can find an integer $t \in D$ with $card\{w \in L \mid |w| = t\} \geq k+1$, in contradiction with the k-slenderness of the language L.

In conclusion, the assumption that $card(I_0) > k(q+1)$ is false and (5) is true.

Consider now the set of triples

$$C = \{(v_i, w_i, x_i) \mid i \in I\}.$$

From (1) it follows that C is finite. Let d be its cardinality and write

$$C = \{(v_1, w_1, x_1), (v_2, w_2, x_2), \ldots, (v_d, w_d, x_d)\}.$$

For every r, $1 \leq r \leq d$, we denote

$$B_r = \{i \in I \mid (v_i, w_i, x_i) = (v_r, w_r, x_r)\}.$$

It follows that $I = \bigcup_{r=1}^{d} B_r$ (in fact, B_r, $1 \leq r \leq d$, constitute a partition of the set I).

We shall prove that for every r, $1 \leq r \leq d$, the set B_r is finite, and this implies that I is finite. In this aim it is sufficient to prove that there are no $i, j \in B_r$, $i \neq j$ with $A_i \cap A_j$ infinite, because in this case for every $i, j \in B_r$, $i \neq j$, the set $A_i \cap A_j$ is finite (possibly empty) and, by the *statement* proved above (relation (5)), it follows that B_r is finite. (Remark that when $B_r = \emptyset$ for all $r \in \{1, 2, \ldots, d\}$, then L is finite, hence it is a UPL language.)

Let us suppose that there is $r \in \{1, 2, \ldots, d\}$ with $i, j \in B_r$, $i \neq j$, such that $A_i \cap A_j$ is infinite. We have

$$A_i = \{u_i v_r^n w_r x_r^n y_i \mid n \geq 0\},$$

$$A_j = \{u_j v_r^m w_r x_r^m y_j \mid m \geq 0\},$$

and with the notation we have introduced we obtain $m_i = m_j = m_r$.

Two cases are possible and they can be treated in the same way: $n_i \geq n_j$ or $n_i \leq n_j$. We suppose that $n_i \leq n_j$. Three cases arise:

(a) $|u_i| \leq |u_j|$ and $|y_i| \leq |y_j|$,

(b) $|u_i| \leq |u_j|$ and $|y_i| \geq |y_j|$,

(c) $|u_i| \geq |u_j|$ and $|y_i| \leq |y_j|$.

Because case (c) is analogous to (b), we shall discuss only cases (a) and (b). The cases $x_r = \lambda$ or $v_r = \lambda$ can be treated in the same way as the case when x_r and v_r are nonempty, so we enter into details only for $x_r \neq \lambda \neq v_r$.

(a) The equality

$$u_i v_r^n w_r x_r^n y_i = u_j v_r^m w_r x_r^m y_j, \tag{9}$$

holds for infinitely many values of n and m. This implies that there are $n', m' \in \mathbb{N}$ and $\alpha, \beta, \delta, \gamma \in V^*$ such that

$$u_j = u_i v_r^{n'} \alpha, \qquad y_j = \gamma x_r^{m'} y_i, \qquad v_r = \alpha\beta = \beta\alpha, \qquad x_r = \delta\gamma = \gamma\delta. \tag{10}$$

Without loss of the generality, we can suppose that $n' \leqslant m'$. We can find $n_0, m_0 \in \mathbb{N}$ satisfying (9) and with $n_0 > \max(n', m')$ (in this aim we can separate the left side and the right side, respectively, from the following relation). We can write

$$u_i v_r^{n_0} w_r x_r^{n_0} y_i = u_j v_r^{m_0} w_r x_r^{m_0} y_j,$$

and, by (10), we get

$$u_i v_r^{n_0} w_r x_r^{n_0} y_i = u_i v_r^{n'} \alpha v_r^{m_0} w_r x_r^{m_0} \gamma x_r^{m'} y_i.$$

Because $n' \leqslant m'$, it follows that

$$|u_i v_r^{n_0}| \geqslant |u_j v_r^{m_0}|,$$

hence

$$u_i v_r^{n'+m_0} v_r^{n_0 - m_0 - n'} w_r x_r^{n_0} y_i = u_i v_r^{n'+m_0} \alpha w_r \gamma x_r^{m_0 + m' - n_0} x_r^{n_0} y_i.$$

Consequently,

$$v_r^{n_0 - m_0 - n'} w_r = \alpha w_r \gamma x_r^{m_0 + m' - n_0}. \tag{11}$$

As $n_i \leqslant n_j$ implies $n_0 \geqslant m_0$, we have $m' + m_0 - n_0 \leqslant m'$ and, in view of (11), we can write the set A_j in the following way:

$$\begin{aligned} A_j &= \{u_j v_r^{\ell} w_r x_r^{\ell} y_j \mid \ell \geqslant 0\} \\ &= \{u_i v_r^{n'} \alpha v_r^{\ell} w_r x_r^{\ell} \gamma x_r^{m'} y_i \mid \ell \geqslant 0\} \\ &= \{u_i v_r^{n'+\ell} \alpha w_r \gamma x_r^{m'+m_0-n_0} x_r^{n_0-m_0} x_r^{\ell} y_i \mid \ell \geqslant 0\} \\ &= \{u_i v_r^{n'+\ell} v_r^{n_0-m_0-n'} w_r x_r^{\ell+n_0-m_0} y_i \mid \ell \geqslant 0\} \\ &= \{u_i v_r^{\ell+n_0-m_0} w_r x_r^{\ell+n_0-m_0} y_i \mid \ell \geqslant 0\} \subseteq A_i. \end{aligned}$$

(b) Similarly, there are $n', m' \in \mathbb{N}$ and $\alpha, \beta, \delta, \gamma \in V^*$ such that

$$u_j = u_i v_r^{n'} \alpha, \qquad y_i = \gamma x_r^{m'} y_j, \qquad v_r = \alpha\beta = \beta\alpha, \qquad x_r = \delta\gamma = \gamma\delta. \tag{12}$$

As previously, we take $n_0, m_0 \in \mathbb{N}$ with $m_0 > \max(n', m')$ and

$$u_i v_r^{n_0} w_r x_r^{n_0} y_i = u_j v_r^{m_0} w_r x_r^{m_0} y_j.$$

From (12) we get

$$u_i v_r^{n_0} w_r x_r^{n_0} \gamma x_r^{m'} y_j = u_i v_r^{n'} \alpha v_r^{m_0} w_r x_r^{m_0} y_j.$$

Because

$$|x_r^{n_0} \gamma x_r^{m'} y_j| \geqslant |x_r^{m_0} y_j|,$$

we can write

$$u_i v_r^{n_0} w_r \gamma x_r^{n_0 + m' - m_0} x_r^{m_0} y_j = u_i v_r^{n_0} v_r^{m_0 + n' - n_0} \alpha w_r x_r^{m_0} y_j.$$

Consequently,

$$w_r \gamma x_r^{n_0 + m' - m_0} = v_r^{m_0 + n' - n_0} \alpha w_r. \tag{13}$$

Because $n' \geqslant n' + m_0 - n_0$, the set A_j can be written, using (13), in the following way:

$$\begin{aligned} A_j &= \{u_j v_r^\ell w_r x_r^\ell y_j \mid \ell \geqslant 0\} \\ &= \{u_i v_r^{n'} \alpha v_r^\ell w_r x_r^\ell y_j \mid \ell \geqslant 0\} \\ &= \{u_i v_r^\ell v_r^{n_0 - m_0} v^{m_0 + n' - n_0} \alpha w_r x_r^\ell y_i \mid \ell \geqslant 0\} \\ &= \{u_i v_r^\ell v_r^{n_0 - m_0} w_r \gamma x_r^{n_0 + m' - m_0} x_r^\ell y_j \mid \ell \geqslant 0\} \\ &= \{u_i v_r^{\ell + n_0 - m_0} w_r x_r^{\ell + n_0 - m_0} y_i \mid \ell \geqslant 0\} \subseteq A_i. \end{aligned}$$

Thus, we have proved that $n_i \leqslant n_j$ implies $A_j \subseteq A_i$. Similarly, $n_i \geqslant n_j$ implies $A_i \subseteq A_j$.

However, as both these possibilities are excluded by the assumption (4), we obtain a contradiction which appears from the hypothesis that there are $i, j \in B_r$, $i \neq j$, such that $A_i \cap A_j$ is infinite. Therefore, for all $i, j \in B_r$, $i \neq j$, the set $A_i \cap A_j$ is finite (possibly empty).

In conclusion, B_r is finite, which implies that I is finite, and this concludes the proof. □

3. Some consequences

Let us denote, as in [4], by SL_X the family of slender languages in a given family X; let LIN, CF be the families of linear and of context-free languages, respectively.

The following consequences of Theorem 2.1 have been already pointed out.

Corollary 3.1. *$SL_{LIN} = SL_{CF}$ and SL_{CF} contains only nonambiguous languages.*

Various closure properties of families SL_X, with X in Chomsky hierarchy, are established in [5], but the closure of SL_{CF} under morphisms, intersection and $init_t$ are left open. (Denoting by $[\alpha]$ the integral part of a rational number α, $init_t(w)$ is the prefix of $w \in V^*$ of length $[|w|/t]$, t being a positive integer. Then, for a language $L \subseteq V^*$, we define $init_t(L) = \{w_1 \mid w = w_1 w_2 \ldots w_t y \in L, |w_i| = [|w|/t], 0 \leqslant |y| < t\}$.) However, it is noticed in [5] that the positive answer to the conjecture in [4] implies the closure of SL_{CF} under all these three operations. For morphisms and $init_t$ the result is an obvious consequence of Theorem 2.1, because the morphic image of a UPL language is a UPL language, too, and the same is true for the operation $init_t$. Therefore,

Corollary 3.2. *The family SL_{CF} is closed under morphisms and $init_t$, $t \geqslant 1$.*

The argument for intersection is ommited in [5]. Because it is not at all obvious, and because we have here an interesting situation when a family X of languages is not closed under a given operation (CF is not closed under intersection), but SL_X is closed, we prove this result in some detail.

Theorem 3.3. *The family SL_{CF} is closed under intersection.*

Proof. Let $L_1, L_2 \subseteq V^*$ be two languages in SL_{CF}. According to Theorem 2.1, L_1, L_2 are unions of paired loops, hence we can write

$$L_1 = \bigcup_{i=1}^{k} A_i, \quad \text{for } A_i = \{u_i v_i^n w_i x_i^n y_i \mid n \geqslant 0\}, u_i, v_i, w_i, x_i, y_i \in V^*,$$

$$L_2 = \bigcup_{j=1}^{\ell} B_j, \quad \text{for } B_j = \{u'_j v'^m_j w'_j x'^m_j y'_j \mid m \geqslant 0\}, u'_j, v'_j, w'_j, x'_j, y'_j \in V^*,$$

Therefore,

$$L_1 \cap L_2 = \bigcup_{i=1}^{k} \bigcup_{j=1}^{\ell} (A_i \cap B_j).$$

Thus, it is sufficient to prove that for every i, j as above, $A_i \cap B_j$ is a UPL.

If $A_i \cap B_j$ is finite, it is trivially UPL. Suppose that for some i, j the set $A_i \cap B_j$ is infinite. We distinguish more cases, depending on the fact whether or not one of the strings x_i, v_i or x'_j, v'_j is empty or not. Denote

$$r = card\{z \in \{v_i, x_i\} \mid z \neq \lambda\},$$

$$s = card\{z \in \{v'_j, x'_j\} \mid z \neq \lambda\}.$$

Because we have $1 \leqslant r \leqslant 2$, $1 \leqslant s \leqslant 2$, we obtain four cases:

(a) $r = s = 2$,

(b) $r = 2$, $s = 1$,

(c) $r = 1$, $s = 2$,

(d) $r = s = 1$.

The cases (b) and (c) are analogous and (d) will be covered by the argument for (a), hence we shall consider in detail only cases (a) and (b).

(a) (All v_i, x_i, v'_j, x'_j are nonempty.) The equality

$$u_i v_i^n w_i x_i^n y_i = u'_j v'^m_j w'_j x'^m_j y'_j$$

holds for infinitely many n, $m \geqslant 1$,

$$(n_0, m_0),\ (n_1, m_1),\ (n_2, m_2), \ldots$$

Because $|u_i|, |u'_j|$ and $|y_i|, |y'_j|$ are finite, there are some constants p, q such that v_i^p is the conjugate of v'^q_j and x_i^p is the conjugate of x'^q_j.

Take k_0 such that

$$|u_i| < |u'_j| + (m_{k_0} - 1)|v'_j|,$$

$$|u'_j| < |u_i| + (n_{k_0} - 1)|v_i|,$$

$$|y_i| < |y'_j| + (m_{k_0} - 1)|x'_j|,$$

$$|y'_j| < |y_i| + (n_{k_0} - 1)|x_i|.$$

For every $k > k_0$, we have

$$u_i v_i^{n_k} v_i^p w_i x_i^{n_k} x_i^p y_i = u_j' v_j'^{m_k} v_j'^q w_j' x_j'^{m_k} x_j'^q y_j'.$$

If we take p, q the smallest integers with these properties, then for $k > k_0$, we have

$$n_{k+1} - n_k = p, \qquad m_{k+1} - m_k = q.$$

This implies that we can rewrite the set $A_i \cap B_j$ in the following way:

$$A_i \cap B_j = C \cup \{u_i v_i^{n_{k_0}} (v_i^p)^n w_i (x_i^p)^n x_i^{n_{k_0}} y_i \mid n \geqslant 0\},$$

where C is the finite set $\{z \in A_i \cap B_j \mid |z| < |u_i v_i^{n_{k_0}} w_i x_i^{n_{k_0}} y_i|\}$. In conclusion, $A_i \cap B_j$ is a UPL language.

(b) (v_i, x_i, v_j' are nonempty, x_j' is empty.) The equality

$$u_i v_i^n w_i x_i^n y_i = u_j' v_j'^m w_j' y_j'$$

holds true for infinitely many pairs n, m,

$$(n_0, m_0),\ (n_1, m_1),\ (n_2, m_2), \ldots$$

Following a similar argument as above, we take k_0 such that

$$|u_i| < |u_j'| + (m_{k_0} - 1)|v_j'|,$$

$$|u_j'| < |u_i| + (n_{k_0} - 1)|v_i|,$$

$$|y_i| < |w_j' y_j'| + (m_{k_0} - 1)|v_j'|,$$

$$|w_j' y_j'| < |y_i| + (n_{k_0} - 1)|x_i|.$$

Similarly, there are $p, q, r \in \mathbb{N}$ such that v_i^p is the conjugate of $v_j'^q$ and x_i^p is the conjugate of $v_j'^r$. Taking p, q, r the smallest with these properties, for all $k > k_0$, we obtain

$$n_{k+1} - n_k = p, \qquad m_{k+1} - m_k = q + r,$$

hence we obtain again a writing of $A_i \cap B_j$ as above, and this completes the proof. □

References

[1] M. Andraşiu, J. Dassow, Gh. Păun and A. Salomaa, Language-theoretic problems arising from Richelieu cryptosystems, *Theoret. Comput. Sci.* **116** (1993) 339–357.

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

[3] Gh. Păun and A Salomaa, Decision problems concerning the thinness of DOL languages *Bull. EATCS* **46** (1992) 171–181.

[4] Gh. Păun and A. Salomaa, Thin and slender languages, *Discrete Appl. Math.*, to appear.

[5] Gh. Păun and A. Salomaa, Closure properties of slender languages, *Theoret. Computer Sci.* **120** (1993) 293–301.

[6] A. Salomaa, *Formal Languages* (Academic Press, New York, 1973).

[7] A. Salomaa, *Public-Key Cryptography* (Springer, Berlin, 1990).