Slender context-free languages are strongly linear

Lucian Ilie
Turku Centre for Computer Science TUCS
FIN-20520 Turku, Finland

Grzegorz Rozenberg
Department of Computer Science, Leiden University
P.O.Box 9512, 2300 RA Leiden, The Netherlands
and
Department of Computer Science
University of Colorado at Boulder
Boulder, CO 80309, U.S.A.

Arto Salomaa
Turku Centre for Computer Science TUCS
FIN-20520 Turku, Finland

Turku Centre for Computer Science
TUCS Technical Report No 248
March 1999

ISBN 952-12-0400-1
ISSN 1239-1891

Abstract

Solving an open problem of [3], we prove that any slender context-free language L is strongly linear, i.e., both L and its complement are linear languages.

TUCS Research Group
Theory Group: Mathematical Structures in Computer Science

A language L is called *slender* if the number of words of the same length is bounded from above by a constant, that is, there is $k \geq 0$ such that, for any $n \geq 0$, $\mathrm{card}(\{w \in L \mid |w| = n\}) \leq k$. It is known (a direct consequence of Theorem 1, see below) that any slender context-free language is linear. In [3], one defines the *strongly* linear languages as those linear languages L such that the complement of L, $\overline{L}$, is also linear. An open problem in [3] asks whether or not any slender context-free language is strongly linear. In this note, we answer this question in the affirmative. Moreover, our proof is constructive, that is, given a slender context-free language L, a linear grammar generating $\overline{L}$ can be effectively constructed.

The following two theorems on slender context-free languages from [1] and [2] will be essential for our purpose. We need few notions. A *paired loop* is a set of the form $\{uv^nwx^ny \mid n \geq 0\}$, for some words u, v, w, x, y. A *single loop* is a set $\{uv^nw \mid n \geq 0\}$.

Theorem 1 ([1]). *A context-free language is slender iff it is a finite union of paired loops.*

Theorem 2 ([2]). *A given slender context-free language can be effectively written as a finite union of paired loops.*

Consider a (non-empty) slender context-free language L. By Theorem 1, L has a form

$$L = \bigcup_{i=1}^{m} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\} \tag{1}$$

for some $m \geq 1, u_i, v_i, w_i, x_i, y_i \in \Sigma^*$, for any $1 \leq i \leq m$. Moreover, by Theorem 2, a decomposition of L as in (1) can be effectively found. Denote, for each $1 \leq i \leq m$, $L_i = \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\}$. Then, the complement of L, denoted $\overline{L}$, is $\overline{L} = \bigcap_{i=1}^{m} \overline{L_i}$.

We now make the following very important *assumption*: if, for some $1 \leq i \leq m$, v_i and x_i are both non-empty, then L_i cannot be written as a finite union of single loops. We call such an L_i a *pure* paired loop. So, all loops of L in (1) are either single or pure paired. (Notice also that one can easily decide whether L_i is a finite union of single loops and, in the affirmative case, L_i can be effectively written as such.)

Under this assumption, we have, for L_i a pure paired loop, that

$$\overline{L_i} = \overline{u_i v_i^* w_i x_i^* y_i} \cup \{u_i v_i^n w_i x_i^m y_i \mid n \neq m\}. \tag{2}$$

Indeed, if we have an identity $u_i v_i^{n_0} w_i x_i^{n_0} y_i = u_i v_i^{m_0} w_i x_i^{p_0} y_i$, for some $m_0 > n_0 > p_0$, then $v_i^{m_0 - n_0} w_i = w_i x_i^{n_0 - p_0}$ and the primitive roots of v_i and x_i are conjugates. More precisely, $v_i = (t_1t_2)^r, x_i = (t_2t_1)^s, w_i = (t_1t_2)^k t_1$, for some integers r, s, k and words t_1, t_2 with t_1t_2 primitive. Therefore, for any $n \geq 0$, $u_i v_i^n w_i x_i^n y_i = u(t_1t_2)^k (t_1t_2)^{(r+s)n} t_1 y$, so L_i is a single loop, a contradiction. Thus (2) holds.

Denote, for any $1 \leq i \leq m$, $A_{0,i} = \overline{u_i v_i^* w_i x_i^* y_i}$, $A_{1,i} = \{u_i v_i^n w_i x_i^m y_i \mid n \neq m\}$. By convention, if L_i is a single loop, then $A_{1,i} = \emptyset$. By (2), we write

$$\overline{L} = \bigcup_{i_j \in \{0,1\}} A_{i_1,1} \cap A_{i_2,2} \cap \ldots \cap A_{i_m,m}.$$

Now, all sets $A_{0,i}, 1 \leq i \leq m$, and $A_{1,i}$, for L_i single loop, are regular and so is any intersection of those. Hence, the problem is to show that any intersection of sets $A_{1,i}, L_i$ pure paired loops, is linear. We restrict ourselves to those intersections that are infinite and study first an infinite intersection of two such sets; for simplifying the notations, we consider

$$\{uv^n wx^m y \mid n \neq m\} \cap \{u'v'^p w'x'^q y' \mid p \neq q\}. \tag{3}$$

As (3) is infinite, we have some identity

$$uv^n wx^m y = u'v'^p w'x'^q y'. \tag{4}$$

We notice that we can omit (and we do) finitely many values of n, m, p, and q in (3) and assume n, m, p, q large enough for the considerations below. Indeed, for some particular values of, say, n, only some single loops are involved. Moreover, if we put an upper bound on, say, n, this implies un upper bound on p since in (4) the overlapping between x^m and v'^p cannot be longer than $|xv'|$, as it would imply that we no longer have pure paired loops.

Without loss of generality, we may assume $u' = uu'', y' = y''y$. As the powers in (4) are assumed large enough, we obtain

$$\begin{array}{llll} v = (t_1t_2)^r, & v' = (t_2t_1)^s, & u'' = t_1(t_2t_1)^h, & t_1t_2 \text{ primitive,} \\ x = (z_1z_2)^i, & x' = (z_2z_1)^j, & y'' = z_2(z_1z_2)^k, & z_1z_2 \text{ primitive.} \end{array}$$

Thus, (4) becomes

$$(t_1t_2)^{rn} w (z_1z_2)^{im} = t_1(t_2t_1)^{h+sp} w' (z_2z_1)^{k+jq} z_2$$

and, assuming $|w| \geq |w'|$, we get for w and w' an identity of one the forms

$$\begin{array}{rcl} t_2(t_1t_2)^l w & = & w'(z_2z_1)^o z_2, \\ w(z_1z_2)^l z_1 & = & t_1(t_2t_1)^o w', \\ w & = & t_1(t_2t_1)^l w'(z_2z_1)^o z_2. \end{array} \tag{5}$$

Moreover, simple combinatorial arguments give that the existence of two different such identities for w and w' (of the same or of different forms) would imply that our paired loops are single ones. Therefore, there is a unique such identity. It means that no matter what values are there for n, m, p, q in (4), the relative position of w and w' is the same. To be more precise, assume the unique identity is of the form in (5). (A similar reasoning holds for the other cases.) Then, all 4-tuples (n, m, p, q) satisfying (4) are given by the system

$$\begin{cases} rn + o = sp + h \\ im - l - 1 = jq + k \\ n \neq m \\ p \neq q \end{cases} \tag{6}$$

In the general case, when intersections of several sets $A_{1,i}$ (L_i pure paired loop) are considered, (6) becomes

$$\begin{cases} \alpha_1 n_1 + \beta_1 = \alpha_2 n_2 + \beta_2 = \cdots = \alpha_k n_k + \beta_k \\ \gamma_1 m_1 + \delta_1 = \gamma_2 m_2 + \delta_2 = \cdots = \gamma_k m_k + \delta_k \\ n_i \neq m_i, 1 \leq i \leq k \end{cases} \tag{7}$$

and the solutions $(n_1, m_1, n_2, m_2, \ldots, n_k, m_k)$ of (7) give us the intersection set. The set of solutions of (7) has a form

$$\{(a_1 p + b_1, c_1 q + d_1, \ldots, a_k p + b_k, c_k q + d_k) \mid p, q \geq 0, a_i p + b_i \neq c_i q + d_i, 1 \leq i \leq k\}. \tag{8}$$

Thus, important for us are the pairs of non-negative integers (p, q) which give the tuples of (8). Represented in coordinates (p, q), these are all points of non-negative integer coordinates which are not lying on any of the lines with the equations $a_i p + b_i = c_i q + d_i$. These lines share the first quadrant into finitely many disjoint areas. The words corresponding to the finite areas can be generated directly. For the infinite ones, it is easy to reduce them to angles, the corresponding words of which are not difficult to generate.

In order to avoid cumbersome notations, let us take an example. It will be clear that everything works similarly in the general case. Consider the language

$$L = \{a^n b^n \mid n \geq 0\} \cup \{a^n b^{2n} \mid n \geq 0\} \cup \{a^{3n+2} b^{2n+4} \mid n \geq 0\}.$$

In constructing $\overline{L}$, the complicated thing is the intersection

$$A_{1,1} \cap A_{1,2} \cap A_{1,3} = \{a^n b^m \mid n \neq m\} \cap \{a^r (b^2)^s \mid r \neq s\} \cap \{a^2 (a^3)^k (b^2)^l b^4 \mid k \neq l\}.$$

The system (7) is here

$$\begin{cases} n = r = 3k + 2 \\ m = 2s = 2l + 4 \\ n \neq m, r \neq s, k \neq l \end{cases}$$

with the general solution (corresponding to (8); the order is (n, m, r, s, k, l))

$$S = \{(3p + 2, 2q + 4, 3p + 2, q + 2, p, q) \mid p, q \geq 0, 3p \neq 2q + 2, 3p \neq q, p \neq q\}.$$

If we put $S' = \{(p, q) \mid (3p + 2, 2q + 4, 3p + 2, q + 2, p, q) \in S\}$, then

$$A_{1,1} \cap A_{1,2} \cap A_{1,3} = \{a^{3p+2}b^{2q+4} \mid (p, q) \in S'\},$$

hence we are interested in the set S'. Represented in coordinates (p, q), S' is the set of all poins of integer coordinates in the first quadrant not lying on any of the lines $d_1 : 3p = q$, $d_2 : 3p = 2q + 2$, and $d_3 : p = q$ (see Figure 1(a)). There are five disjoint areas, $A_i, 1 \leq i \leq 5$. A_1 is finite. (In fact, there is no word corresponding to it.) Consider the two angles A_2 and A_4. The words corresponding to A_4 are generated by the rules

$$\begin{array}{l} A \longrightarrow a^2Bb^4 \\ B \longrightarrow (a^3)^5C(b^2)^6 \\ C \longrightarrow a^3Cb^2 \mid (a^3)^2C(b^2)^3 \mid \varepsilon \end{array}$$

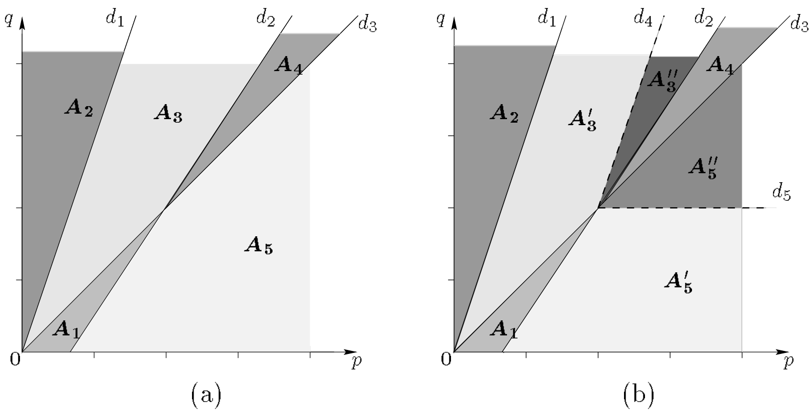

Figure 1

For the angle A_2 it is simpler as it has one restriction only. Consider next A_3. It is not an angle, but it can be reduced to one. We split it using the line d_4 of equation $q = 3p - 4$ (it is parallel with d_1) into $A'_3 \cup d_4$ and the angle A''_3 (see Figure 1(b)). The part $A'_3 \cup d_4$ is easy to handle; the words corresponding to it are

$$\{a^5b^8, a^8b^{10}, a^8b^{12}, a^8b^{14}\} \cup \{a^{11}(a^3)^n(b^6)^n \mid n \geq 0\}\{b^{14}, b^{16}, b^{18}, b^{20}\}.$$

The angle A''_3 is handled as A_4. Finally, for A_5, things are again simpler; it is splitted by the line $d_5 : q = 2$.

It should be clear that in the general case everything works the same way. All areas are reduced to three basic ones: finite, in between two parallel lines, angles. All are delt with as above. It is worth noticing that the infinite areas which are not angles can still be treated directly, in a similar way as for angles. We preferred the reduction in order to simplify the proof.

We therefore proved

Theorem 3. *Any slender context-free language L is strongly linear. Moreover, given a context-free grammar for L, a linear grammar for $\overline{L}$ can be effectively constructed.*

References

[1] Ilie, L., On a conjecture about slender context-free languages, *Theoret. Comput. Sci.* **132** (1994) 427 – 434.

[2] Ilie, L., On lengths of words in context-free languages, *Theoret. Comput. Sci.*, to appear.

[3] Păun, Gh., Rozenberg, G., Salomaa, A., On strongly context-free languages, submitted, 1998.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**