Theoretical Computer Science 81 (1991) 147-153
Elsevier

Note

On the number of partially abelian square-free words on a three-letter alphabet

Robert Cori*
LaBRI, Université Bordeaux I, 351 Cours de la Libération, F-33405 Talence Cedex, France

Maria Rosaria Formisano
Istituto di Matematica, R. Caccioppoli, Università di Napoli, Italy

Communicated by D. Perrin
Received March 1990

0. Introduction

Generalizing the notion of square in the free monoid A^*, we have introduced in [4] the notion of a square with respect to a commutation relation $\sim_\theta$. These commutation relations are generated by the relations $ab \sim_\theta ba$ for the pairs (a, b) belonging to a subset θ of $A \times A$. They were introduced by Cartier and Foata [3] and extensively studied since Mazurkiewicz [9] pointed out their power as a formal model for concurrent computations.

Considering the combinatorial aspect of a commutation θ, we have defined a θ-square in A^* as a nonempty word fg such that $f \sim_\theta g$. Note that a θ-square is an ordinary square when $\theta = \emptyset$, and an abelian square when $\theta = A \times A$. We have considered the subset $L_2(A, \theta)$ of words not containing a θ-square as a factor and we have proved that $L_2(A, \theta)$ is infinite for almost all alphabets A and commutations θ. In this note we restrict the alphabet A to having three letters $A = \{a, b, c\}$ and the relation θ to have only one pair $ac \sim_\theta ca$. We prove that the number α_n of words of length n in $L_2(A, \theta)$ is bounded by a polynomial in n. This result is to be compared

* Work done while the author was visiting professor at the Università "La Sapienza", Rome, Italy.

0304-3975/91/$03.50

with the one obtained by Brandenburg [1] when $\theta=\emptyset$, proving that the number of square-free words of length n of $\{a, b, c\}^*$ is an exponential function of n.

Our proof is based on the bounds obtained by Restivo and Salemi [10] and by Kobayashi [7] for the number of overlapping-free words.

1. Notations and definitions

We follow Lothaire [8, 4]. We fix the alphabet A to be $A=\{a, b, c\}$ in all of the paper. The *length* of a word w of A^* will be denoted by $|w|$, and the number of occurrences of the letter $x\in A$ in w by $|w|_x$. A word u is a *factor* of w if $w=wuw''$ for some w', $w''\in A$. A word w is *square-free* if whenever uu is a factor of w then u is the empty word. We consider the subset $\theta=\{(a, c), (c, a)\}$ of $A\times A$, and the congruence $\sim_\theta$ of A^* generated by the relation $ac\sim_\theta ca$.

The following proposition, a consequence of [5, Proposition 1.1], characterizes the pair of words f, g equivalent under $\sim_\theta$. It may also be used as a definition of $\sim_\theta$.

Proposition 1. *Two words f, $g\in\{a, b, c\}^*$ are such that $f\sim_\theta g$ if and only if there exist decompositions of f and g*

$$f=f_1bf_2b\ldots f_pbf_{p+1},\qquad g=g_1bg_2b\ldots g_pbg_{p+1}$$

satisfying

$$\forall i=1,\ldots,p+1\quad f_i, g_i\in\{a, c\}^*\quad \textit{and}\quad |f_i|_a=|g_i|_a, |f_i|_c=|g_i|_c.$$

A nonempty word w is a *θ-square* if $w=fg$ $f\sim_\theta g$ and it is *θ-square*-free if none of its factors is a θ-square. Note that if w is θ-square-free then it is also square-free.

Carpi and De Luca [2] have defined a closely related but different notion of square-free trace.

We denote by $L_2(\theta)$ the set of θ-square-free words and by $L_2=L_2(\emptyset)$ the set of square-free words. Note that these two sets coincide for the words of length strictly less than 6, and *bacbca* is a square-free word of length 6 which is a θ-square thus not θ-square-free.

We denote by M_2 the subset of L_2 consisting of all the square-free words not containing *aba* or *cbc* as a factor

$$M_2=L_2\backslash A^*\{aba, cbc\}A^*.$$

We denote by α_n and β_n respectively the number of words of length n of $L_2(\theta)$ and M_2

$$\alpha_n=\mathrm{card}(L_2(\theta)\cap A^n),\qquad \beta_n=\mathrm{card}(M_2\cap A^n).$$

2. Relations between α_n and β_n

The following lemma is a consequence of [4, Proposition 3.2]; here we give a simple proof of it.

Lemma 1. *Let f be a word of M_2 of length $|f| \geqslant 2$, and let g be given by $f = xgy$, $x, y \in A$. Then $g \in L_2(\theta)$.*

Proof. Let f and g be as in the statement. We first show that if $g = g'bhbg''$ with $h \in \{a, c\}^*$ then $h \in \{a, c, aca, cac\}$. Since f is square-free so is h, hence $h \in \{a, c, ac, ca, aca, cac\}$. Thus we have only to rule out the cases $h = ac$ and $h = ca$. By symmetry we may only consider $h = ac$. Assume $g = g'bacbg''$; then f has $x'bacby'$ as a factor for some $x', y' \in A$. Since $f \in M_2$, $x' \neq a$ and $y' \neq c$, we have also $x' \neq b$ and $y' \neq b$ then $cbacba$ is a factor of f in contradiction with the assumption that f is square-free.

Assume now that g has a factor which is a θ-square; then $g = g'uvg''$ and $u \sim_\theta v$. By Proposition 1 we have

$$u = u_1bu_2 \ldots u_pbu_{p+1}, \qquad v = v_1bv_2 \ldots v_pbv_{p+1},$$

$u_i, v_i \in \{a, c\}^*$ and $|u_i|_x = |v_i|_x$ for $x \in \{a, c\}$. By the above remark we get $u_i = v_i$ for $i = 2, \ldots, p$, and $u_{p+1}v_1 \in \{a, c, aca, cac\}$. Since f is square-free $u_1 \neq v_1$ or $u_{p+1} \neq v_{p+1}$ (or both). If u_{p+1} or v_1 is empty then $u = v$ and f is not square-free. We may thus assume $u_1 = ac$ and $v_1 = ca$ or $u_{p+1} = ac$ and $v_{p+1} = ca$. Since $u_2 = v_2$, one of u_1bu_2 or v_1bv_2 has a factor in $\{aba, cbc\}$ in the first case and the same holds for $u_{p-1}bu_p$ or $v_{p-1}bv_p$ in the second case in contradiction with $f \in M_2$. $\square$

Lemma 2. *Let $f \in L_2(\theta)$ be of length greater than 7 and let $f = ugv$, $u, v \in A^2$; then $g \in M_2$.*

Proof. Assume aba is a factor of g; since f is square-free, one of *bcabac*, *cabacb*, *acabacab*, *bacabaca* is a factor of f contradicting $f \in L_2(\theta)$. $\square$

Proposition 2. *The numbers α_n and β_n satisfy the following inequalities:*
(i) $\beta_n \leqslant 4\alpha_{n-2}$, $n \geqslant 3$;
(ii) $\alpha_n \leqslant 9\beta_{n-4}$, $n \geqslant 7$.

Proof. (i): Consider the mapping from A^2A^* into A^* given by

$$xgy \to g, \quad x, y \in A.$$

By Lemma 1 the image of M_2 is a subset of $L_2(\theta)$. Any element g of A^* is the image of nine different elements xgy, if xgy is an element of M_2 then x must be different from the first letter of g and y must be different from the last. Thus any element of $L_2(\theta) \cap A^{n-2}$ is the image of at most four elements of $M_2 \cap A^n$ giving the first inequality.

(ii): Consider the mapping form $A^4A^* \to A^*$ given by

$$ugv \to g, \quad u, v \in A^2.$$

Let us show that, for any $g \in M_2$, $|g| \geqslant 3$ there are at most nine pairs (u, v) such that $ugv \in L_2(\theta)$. Since $L_2(\theta) \subset L_2$ it will be sufficient to show that, for any square-free word f, $|f|>2$, there are at most three words $u \in A^2$ such that uf is square-free. Let xy be the first two letters of f, $f=xyg$, and let z be the letter of A such that $z \neq x$, $z \neq y$; then, for u_1, $u_2 \in A$, $u_1u_2f \in L_2$ implies that u_1u_2 is one of the words xz, yz or zy.

Hence any word g of M_2 of length $n-4$ is the image of at most nine different words of $L_2(\theta)$ of length n, giving the second inequality. □

3. Square-free words versus overlapping-free words

Let A_0 be the alphabet consisting of the two letters $\{\underline{0}, \underline{1}\}$, a word w of A_0^* is overlapping-free if it contains no factor of the form $xuxux$ with $u \in A_0^*$ and $x \in A_0$. Let c_n be the number of overlapping-free words of length n. Kobayashi [7] proved the following bounds for c_n:

$$c_1 n^{1.155} < c_n < c_2 n^{1.587}.$$

We will reduce the enumeration of words in M_2 to that of overlapping-free words using the following morphism φ from A^* into A_0^*;

$$\varphi(a) = \underline{011}, \qquad \varphi(b) = \underline{01}, \qquad \varphi(c) = \underline{0}.$$

This morphism is well-known since it allows to build the Thue–Morse sequence on the alphabet A from that on A_0^*.

It is clear that φ is one-to-one and that the image $\varphi(A^*)$ consists of all words beginning with the latter $\underline{0}$ and which do not contain $\underline{111}$ as a factor.

The following proposition is contained in the paper of Thue [11] and is given as an exercise in [8]; here we give a proof of it in order to be complete.

Proposition 3. *For any word f of M_2, $\varphi(f)$ is overlapping-free. Conversely, if w is an overlapping-free word of length greater than 6 beginning with the letter $\underline{0}$, then there exists a unique word g of M_2 and letters x, y of A such that $\varphi(xgy) = w$.*

Proof. The first part is obtained using the following observation: if $u \in \varphi(A^*)$ is such that $u = v\underline{0}w$ then there exist a unique f and g such that $\varphi(f) = v$, $\varphi(g) = \underline{0}w$. This observation rules out from $\varphi(f)$ (for $f \in M_2$) the overlappings $\underline{0}u\underline{0}u\underline{0}$ since they would come from a square of f.

A similar argument holds for the overlappings of the form $\underline{1}u\underline{1}u\underline{1}$ when u begins with the letter $\underline{1}$, since two $\underline{1}$s must be preceded by a $\underline{0}$. When the overlapping $\underline{10}v\underline{10}v\underline{1}$ is a factor of $\varphi(f)$, the hypothesis that f is square-free implies that this overlapping is followed by a $\underline{1}$, and that v ends with a $\underline{0}$. We thus obtain the following decomposition for f:

$$f = f'gbgaf'' \quad \text{with } \varphi(f') \in \{\underline{0}, \underline{1}\}^*\underline{1}$$

Hence,

$$f' \in \{a, b, c\}^*\{a, b\}.$$

The contradiction is then obtained considering the cases $|g| = 0$, $|g| = 1$, and $|g \geq 2$ and using the fact that aba and cbc are not factors of f.

To prove the converse part, we first remark that any overlapping-free word u does not contain $\underline{111}$ as a factor. Assuming that it begins with $\underline{0}$ implies the existence of an $f \in A^*$ such that $\varphi(f) = u$. If u is of length greater than 6 then f is of length greater than 2 and $f = xgy$, $g \in A^*$, $x, y \in A$. Assume that g contains a square factor $g = g'hhg''$; then f contains hhz as a factor for a certain z of A and then $u = \varphi(f)$ contains an overlapping $\varphi(h)\varphi(h)\underline{0}$.

It remains to prove that g does not contain aba or cbc as a factor; the first is ruled out by remarking that 10101 is an overlapping factor of $\varphi(aba)$, the second possibility implies that f has $acbca$ as a factor and 1001001 is an overlapping factor of $\varphi(acbca)$. $\square$

Proposition 4. *The numbers β_n of words of M_2 of length n, and c_n of overlapping-free words of length n satisfy the following inequalities*:

(i) $\forall n\ \exists k \in \{2n-1, 2n, 2n+1\}$ *such that* $\beta_n \leq \frac{3}{2}c_\kappa$.

(ii) $c_{2n} \leq 2\beta_{n-2}$.

Proof. (i) We first prove that, for any $f \in M_2$, $|f|_a = |f|_c + \varepsilon$ where $\varepsilon \in \{-1, 0, 1\}$. This is done by induction on $|f|$. Of course, this identity is trivial for $|f| \leq 3$ since aa or cc is not a factor of f. Consider a word f, $|f| > 3$; then $f = f'u$ where $u \in \{b, ac, ca, abc, cba\}$. The inductive hypothesis applied to f' gives the result.

For any $f \in M_2$, $\varphi(f)$ is overlapping-free and

$$|\varphi(f)| = 3|f|_a + 2|f|_b + |f|_c = 2|f| + |f|_a - |f|_c.$$

By the above remark we obtain

$$|\varphi(f)| = 2|f| + \varepsilon \quad \text{where } \varepsilon = \{0, 1, -1\}.$$

Now φ is one-to-one and as $\varphi(f)$ begins with the letter $\underline{0}$ we get

$$\beta_n \leq \tfrac{1}{2}(c_{2n-1} + c_{2n} + c_{2n+1}).$$

The first inequality is then obtained by choosing k such that c_k is maximal among $\{c_{2n-1}, c_{2n}, c_{2n+1}\}$.

(ii) Let w be an overlapping-free word of length $2n$ beginning with a $\underline{0}$; then, by Proposition 3, there exists a unique $g \in M_2$ and letters $x, y \in A$ such that $\varphi(xgy) = w$. By the same remark above xgy is of length n, hence $|g| = n - 2$ and there are only two possible choices for x and y since $\varphi(w)$ is overlapping-free. $\square$

4. Consequences

The two following results are consequences of the propositions obtained in the preceding sections. The first one is an enumerative result on the number of θ-square-free words, the second one is a nonexistence theorem on θ-square-free morphisms.

Theorem 1. *The number α_n of θ-square-free words of length n satisfies*

$$c'_1 n^{1.15} \leqslant \alpha_n \leqslant c'_2 n^{1.587}.$$

Proof. Use Kobayashi's result for c_n, then apply Propositions 2 and 4. □

A morphism ψ from $\{a, b, c\}^*$ into itself is *θ-square-free* if, for any θ-square-free word f, $\psi(f)$ is also θ-square-free. Of course, the identity and the morphism ψ given by

$$\psi(b) = b, \qquad \psi(a) = c, \qquad \psi(c) = a$$

are θ-square-free; let us call them *elementary*. They are the only two permutations of the letters $\{a, b, c\}$ generating a square-free morphism, since for the other permutations *bacbca* is the image of a θ-square-free word.

Theorem 2. *Any θ-square-free morphism from $\{a, b, c\}^*$ into itself is elementary.*

Proof. Let ψ be a θ-square-free morphism, consider any square-free word f of length not greater than 5; f is also θ-square-free, hence $\psi(f)$ is square-free. By a characterization due to Crochemore [6] this proves that ψ is a square-free morphism, thus $\psi(g)$ is square-free for any square-free word g.

The next step in the proof is to show that if two letters among $\{a, b, c\}$ have images by ψ of length greater than 1 then neither *aba* nor *cbc* can be a factor of $\psi(f)$ for a θ-square-free word f. By the hypothesis on the lengths of images of letters we may assume that the square-free word f such that $\varphi(f)$ has *aba* or *cbc* as a factor is of length 2. Let $f = xy$, with $x \neq y$, be such a word; consider the letter z such that $x \neq z$, $z \neq y$ and the word $g = yzxyzy$. This word is θ-square-free for any x, y, z. Let $u = \psi(g)$ we have

$$u = \psi(y)\psi(z)\psi(xy)\psi(z)\psi(y),$$

$$u = u'abau'' \quad \text{with } |u'| \geqslant 3, |u''| \geqslant 3$$

and u contains a θ-square, a contradiction by Lemma 2.

Now, if two letters x, y have images of length 1 then either ψ is not a square-free morphism or ψ is elementary: if x' and y' are the images of x, y then $\psi(z)$ begins and ends with the same letter z'; if $\psi(z)$ is of length greater than 1, then $\psi(z) = z'u'z'v'u' \in \{x', y'\}^*$; consider $u \in \{x, y\}$ such that $\psi(u) = u'$; then uz is square-free and $\psi(uz)$ is not, a contradiction.

We have thus obtained that any θ-square-free morphism which is not elementary maps the set of square-free words L_2 into M_2 but this is not compatible with the following facts:

- the number of words of length n of L_2 has an exponential growth [1];
- the number β_n is bounded by a polynomial in n (Theorem 1). □

References

[1] F. Brandenburg, Uniformly growing kth power-free homomorphisms, *Theoret. Comput. Sci.* **23** (1983) 145–149.

[2] A. Carpi and A. De Luca, Square-free words on partially commmutative free monoids, *Inform. Process. Lett.* **22** (1986) 125–131.

[3] P. Cartier and D. Foata, *Propriétés Combinatoires de Commutations et Réarrangements*, Lecture Notes in Mathematics **5** (Springer, Berlin, 1969).

[4] R. Cori and M.R. Formisano, Partially abelian square-free words, *RAIRO Inform. Théor.*, to appear.

[5] R. Cori and D. Perrin, Automates et commutations partielles, *RAIRO Inform. Théor.* **19** (1985) 232.

[6] M. Crochemore, Sharp characterisations of square-free morphisms, *Theoret. Comput. Sci.* **18** (1982) 221–226.

[7] Y. Kobayashi, Enumeration of irreducible binary words, to appear.

[8] M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading, MA, 1983).

[9] A. Mazurkiewicz, Concurrent program schemes and their interpretations, Tech. Report, DIAMI-PB 76, Aarhus University, Aarhus, Denmark, 1977.

[10] A. Restivo and A. Salemi, On weakly square-free words, *Bull EATCS* **21** (1983) 49–56.

[11] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat-Nat. Kl., Christiana.* **1** (1912) 1–67.