Atti Accademia Nazionale Lincei Classe Scienze Fisiche Matematiche Naturali

Rendiconti Lincei
Matematica e Applicazioni

Claudio Baiocchi

On some properties of doubly-periodic words

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti Lincei. Matematica e Applicazioni, Serie 9, Vol. **8** (1997), n.1, p. 39–47.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLIN_1997_9_8_1_39_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti Lincei. Matematica e Applicazioni, Accademia Nazionale dei Lincei, 1997.

Rend. Mat. Acc. Lincei
s. 9, v. 8:39-47 (1997)

Scienza dell'informazione. — *On some properties of doubly-periodic words.* Nota (*) del Corrisp. CLAUDIO BAIOCCHI.

ABSTRACT. — We study the functional equation:

$$ABC = CDA \tag{1}$$

where A, B, C and D are *words* over an *alphabet* $\mathcal{A}$. In particular we prove a «structure result» for the inner factors B, D: for suitably chosen words X, Y, Z one has:

$$B = XYZ, \qquad D = ZYX. \tag{2}$$

It is a generalization of the Lyndon-Schützenberger's Theorem (see [7]): if in (1) A or C is empty, formula (2) holds true with one among X, Y, Z which can be chosen empty.

KEY WORDS: Words; Periodicity; Palindromy.

RIASSUNTO. — *Su alcune proprietà delle parole doppiamente periodiche.* Si studia l'equazione funzionale:

$$ABC = CDA \tag{1}$$

in cui A, B, C e D sono *parole* su un *alfabeto* $\mathcal{A}$. In particolare si ottiene una «formula di struttura» per i fattori centrali B e D: per opportune parole X, Y, Z vale:

$$B = XYZ, \qquad D = ZYX. \tag{2}$$

Si tratta di una generalizzazione del Teorema di Lyndon-Schützenberger (cfr. [7]): con due soli fattori nella (1) (cioè se una delle parole A, C è vuota) in (2) bastano due fattori (cioè: una almeno tra X, Y e Z è vuota).

1. NOTATIONS AND STATEMENTS

Let $\mathcal{A}$ be a non empty set, whose elements will be called letters; $\mathcal{A}^*$ will denote the free monoid over $\mathcal{A}$; the elements of $\mathcal{A}^*$ will be called words; for any word W the length of W is denoted by $|W|$; the empty word (*i.e.* the identity of $\mathcal{A}^*$) will be denoted by Θ (and, of course, $|\Theta| = 0$).

Let p be a positive integer; a word $W = \alpha_1 \ldots \alpha_n$ is p-periodic (and p is a period for W) if the relation $\alpha_j = \alpha_{j+p}$ holds true for $1 \leq j \leq n - p$ (1).

We will deal with the following problem:

PROBLEM A. *We are given three positive integers, w, p, q such that:*

$$p, q < w < p + q - \gcd(p, q). \tag{3}$$

We ask for words W of length w which are both p-periodic and q-periodic. ∎

REMARK 1.1. *We will in fact work in a more general setting, but the most interesting results will hold true under the restriction* (3). *The reason for such a restriction is that we would work with «truly-double-periodic» words; this is obviously not the case if $p = q$; and,*

(*) Pervenuta all'Accademia il 30 luglio 1996.

(1) As usual, we do not require $p < n$; any word W is p-periodic for $p \geq |W|$.

more generally, if $w \geqslant p + q - \gcd(p, q)$: *the well known Fine-Wilf's Theorem* (see [5]) *says that, in such framework, W is simply a* $\gcd(p, q)$*-periodic word. On the other hand, as pointed out in footnote* ([1]), *if p or q reaches w the corresponding periodicity imposes no restrictions.* ■

REMARK 1.2. *A special case of particular interest is the case of* $w = p + q - 2$; (3) *then reads*:

$$p, q \text{ are coprime}; \quad |w| = p + q - 2 \tag{4}$$

and the solutions W of Problem A are strictly related to the «Sturmian Words», as proved in recent papers [2-4]; *in particular any such W is a palindrome word. Our results will give a new proof of this palindromy.* ■

Let us define the quantities a, b and c by setting:

$$a := w - q; \quad b := p + q - w; \quad c := w - p \tag{5}$$

so that, in the framework of (3), a, b and c are strictly positive. Because of $a + b + c = w$, for the solutions W of Problem A the formula:

$$W = ABC; \quad |A| = a, \ |B| = b, \ |C| = c \tag{6}$$

defines the words A, B, C; and the double periodicity of W holds true if and only if there exists a word D (of course: with length $d \equiv b$) such that (1) holds true. Of course the inverse formula of (5) is given by:

$$p = a + b; \quad q = b + c; \quad w = a + b + c \tag{7}$$

and (3), in terms of a, b and c, implies:

$$a, b, c > 0; \quad b > \gcd(a + b, b + c). \tag{8}$$

However problem (1) (with prescribed lengths a, b, c for A, B and C) could be studied under the more general assumption $a, b, c \in \mathbf{N}$. Let us recall some known results in such a framework.

1) If c vanishes, C becomes the empty word, thus disappearing from (1). The corresponding equation is the so called *Lyndon-Schützenberger's equation*. For any triple $\{j, R, S\}$ with $j \in \mathbf{N}$ and $R, S \in \mathcal{A}^*$, setting:

$$A = (RS)^j R, \quad B = SR; \quad D = RS \tag{9}$$

we get a solution of (1) with $C = \Theta$. Conversely, (see [7]), for any triple $\{A, B, D\}$ such that (1) holds true with $C = \Theta$, one has (9) for suitably chosen j, R, S ([2]). In particular (2) holds true with one less factor. Of course a similar result holds true when a vanishes, say $A = \Theta$; the case $b = 0$ can be treated by an obvious change of names in (1).

2) If $a, b, c > 0$, and $b \leqslant \gcd(a + b, b + c)$, the Fine-Wilf's theorem (see [5])

([2]) The value of $|R|$, $|S|$ and j can be calculated by some obvious «modular» operations on $|A|$ and $|B|$.

implies that, for a suitable word E of length $\gcd(a+b, b+c) - b$, one has

$$A = (EB)^j E\,, \quad C = (EB)^k E\,, \quad D = B \tag{10}$$

with j, k defined by the obvious modular operations. Conversely, for any $j \in \mathbf{N}$ and for any choice of $B, E \in \mathcal{A}^*$, formula (10) gives a solution of (1). Of course (2) still holds true, with *two* factors vanishing [3].

3) The remaining case can be described by:

$$b > \gcd(a+b, b+c) \tag{11}$$

which in turn (because of $a, b, c \geqslant 0$) implies $a, b, c > 0$; thus we are in the framework of (8). As far as we know, no results are known. As already said in the Abstract, we shall prove a formula of type (2); for the moment being, let us remark that in general three (non empty) factors could be needed. Fix any triple $X, Y, Z \in \mathcal{A}^*$ and set $B := XYZ$. Then, for any choice of $j \in \mathbf{N}$, formulae:

$$A = (YZ)^{j+1} Y; \quad C = (YZ)^j Y; \quad D = ZYX \tag{12}$$

give a solution of (1), the «symmetric» one (see footnote [3]) being given by:

$$A = (YX)^j Y; \quad C = (YX)^{j+1} Y; \quad D = ZYX\,. \tag{13}$$

The results we will given in Section 2, together the ones we just recalled, can be grouped into the following Theorems 1.1 and 1.2.

THEOREM 1.1. *For any choice of a, B, c, with $a, c \in \mathbf{N}$ and $B \in \mathcal{A}^*$, there exists at least a triple A, C, D of words, with*

$$|A| = a\,, \quad |C| = c \tag{14}$$

such that (1) *holds true. All solutions have the same D; the condition*

$$b \geqslant \gcd(a+b, b+c) \tag{15}$$

is necessary and sufficient for the uniqueness of A, C. ■

Let us be more precise about the map $\{a, B, c\} \mapsto D$. We will construct a map from $\mathbf{N}^3$ into itself:

$$\mathbf{N}^3 \ni \{a, b, c\} \mapsto \{x, y, z\} \in \mathbf{N}^3 \tag{16}$$

such that for any solution of (1) (with lengths a, b, c for A, B, C) one has (2) with:

$$|X| = x(a, b, c)\,, \quad |Y| = y(a, b, c)\,, \quad |Z| = z(a, b, c)\,; \tag{17}$$

of course our map will satisfy:

$$x(a, b, c) + y(a, b, c) + z(a, b, c) = b \quad \text{for all } a, b, c \in \mathbf{N} \tag{18}$$

so that for any B of length b there exists a unique triple $\{X, Y, Z\}$ satisfying (17) and $B = XYZ$. Let us summarize the corresponding result:

[3] We could *e.g.* choose $X = Z = \Theta$, $Y = B$, in order to respect the «symmetry» of the problem: equation (1) is symmetric with respect to the swaps $a \leftrightarrow c$, $A \leftrightarrow C$; swaps that in (2) just require $X \leftrightarrow Z$.

THEOREM 1.2. *For any choice of* a, B, c, *with* $a, c \in \mathbf{N}$ *and* $B \in \mathcal{A}^*$, *let* X, Y, Z *be defined through* (17) *and*:

(19) $$B = XYZ\,.$$

For any solution of (1) *with* (14), *one has* $D = ZYX$; *furthermore, if* (15) *holds true, also the words* A *and* C *can be uniquely reconstructed by suitably combining powers of* X, Y, Z. ■

These results (that of course also apply to Problem A) will be proved in § 2; let us remark, however, that they must be proved only in the framework of (8), the remaining cases being already known. In § 3 we will investigate some interconnections between double periodicity and a generalization of the notion of palindromy (see Remark 1.2).

2. PROOFS

In this Section we will use «mixed» notations: A, B, C, D will denote solutions of (1) of lengths a, b, c and $d \equiv b$; p, q, w will denote the quantities given by (7); $W = ABC$ will denote the corresponding word of length w and periods p, q. We will assume that (3) holds true; say, in «mixed» notations:

(20) $$b > \gcd(p, q)\,;$$

as we already remarked, we will also have:

(21) $$a, b, c > 0\,; \qquad a \neq c\,.$$

Let us first work under the assumption:

(22) $$|b| > |a - c|$$

and show that the corresponding solutions of (1) must be described through (12) or (13); we detail only the case of

(23) $$c < a < b + c$$

corresponding to (12), the other case being quite similar. We set [4]:

(24) $$x(a, b, c) := b + c - a\,; \quad y(a, b, c) := a \operatorname{Mod} c\,; \quad z(a, b, c) := a - c - y$$

so that, because of (18), for any B of length b formulae (19), (17) *define* the words X, Y, Z. Setting also:

(25) $$j := c \operatorname{Div} (a - c)$$

we remark that formula (12) provides a solution of (1) with $|A| = a$, and $|C| = c$. Let us prove that this is the only solution:

LEMMA 2.1. *Let* A, B, C, D *be given with* (1) *and such that, for the corresponding lengths,* (23) *holds true. Then formula* (12) *holds true with* X, Y, Z, j *defined through* (24), (19), (17), (25).

[4] As usual, we denote by «a Mod b», «a Div b» the *reminder* and the *integer quotient* between the positive integers a and b.

PROOF. Using our definitions for X, Y, Z, we also set $U := YZ$, so that $B = XU$; remark that, if we can prove the decomposition:

$$U = SR\,; \qquad |R| = z\,, \qquad |S| = y \tag{26}$$

we will have $R = Z$, $S = Y$. Replacing $B = XU$ into (1) we get $(AX)\,UC = (CD)A$ where, in both sides, the parentheses denote words of length $b + c$. It follows that $AX = CD$, $UC = A$ so that D (which is longer than X) can be factorized in the form $D = VX$. Now $AX = CD$ becomes $A = CV$, so that we have a double representation for A, say $A = CV = UC$. The Lyndon-Schützenberger's theorem applied to $CV = UC$ then implies $V = RS$, $U = SR$, $C = (SR)^j S$, for suitably chosen words R, S, the length of S being given by $|UC| \operatorname{Mod} |U|$. The lemma then follows immediately by remarking that the values for $|R|$, $|S|$ coincide with the values z and y, so that (26) holds true. ∎

REMARK 2.1. *Let us point out that* (23) *implies* $0 < x < b$, *so that* (*because of* $x + y + z = b$):

(27) two at least among X, Y, Z are non-empty.

On the other hand, due to the symmetry of the problem (see ([3])), *formula* (27) *holds true under the general assumption* (22), *and not only in the framework of* (23). ∎

The following remark will be useful: starting from a solution $\{A, B, C, D\}$ of (1), we can construct a «longer» solution by replacing A or C by the whole $W = ABC$; in other words, both the quadruplets $\{ABC, B, C, D\}$ and $\{A, B, ABC, D\}$ still satisfy (1) (the proof is immediate). Remark that for such «expanded» solution the (new) lengths a, b, c satisfy respectively $a \geqslant b + c$ and $c \geqslant a + b$; so that in any case one has:

$$|a - c| \geqslant b \tag{28}$$

and in particular (22) fails. Conversely, let us start with a quadruplet satisfying (28) and let us prove that it is an «expansion» of a shorter quadruplet. We detail the case $a \geqslant b + c$, the other one being similar: the word A is longer than CD; so that from (1) follows that, for a suitably chosen (possibly empty) word A_0, it is $A = CDA_0$. By substituting such a formula for A into (1) we get $(CDA_0)\,BC = CD(CDA_0)$; thus, simplifying, we get $A_0BC = CDA_0$. The new quadruplet $\{A_0, B, C, D\}$ still satisfies (1); and of course the starting solution can be reconstructed by means of the formula $\{CDA_0, B, C, D\}$; so that, if for the shorter quadruplet one has uniqueness, there is uniqueness also for the starting one.

Because of (20) (which implies $b > 0$) the new word $W_0 = A_0BC$ is definitly shorter than the original W; and (if (22) still fails) we can iterate the procedure. Let us point out what happens concerning the periods of the (shorter and shorter) word W:

$$\text{(29)} \qquad \begin{cases} \text{while } |p-q| \geqslant b: & \\ \qquad\qquad \text{if } p \geqslant q: \quad p \text{ becomes } p-q \\ \qquad\qquad \text{else:} \qquad q \text{ becomes } q-p \\ \text{end while} \end{cases}$$

which is nothing else but the Euclide's Algorithm for the evaluation of $\gcd(p,q)$, with an unusual stop-criterion. It is a «true» algorithm, that ends after a finite number of steps (at most as many as for the standard criterion, which ends when $p-q=0$; here we stop when $|p-q|<b$, with $b>0$).

Of course, when the algorithm ends, one has $|p-q|<b$; so that (22) holds true. The existence and uniqueness theorems proved in this case will remain valid for all expansions; and (the words B and D being unchanged) the representation formula (2), as well as the assertion (27), still holds true.

Theorems 1.1 and 1.2 are thus completely proved; let us end this Section with two remarks:

REMARK 2.2. *In the framework of Remark* 1.2, *we have* $b=2$; *so that* (27) *says that exactly two among* X, Y, Z *have length* 1 (*the third one being empty*). *In particular it is*

$$\text{(30)} \qquad B=\alpha\beta\,; \qquad D=\beta\alpha$$

for (*possibly coinciding*) *letters* α, β. ■

REMARK 2.3. *In the framework of* $a,b,c \geqslant 0$ *one has the implication*:

$$\text{(31)} \qquad B \neq D \Rightarrow \gcd(a+b,b+c) < b$$

because from $\gcd(a+b,b+c)=b$ *it follows* $B=D$. ■

3. Biperiodicity and palindromy

In this Section we will denote by σ an involution of $\mathcal{A}$; to any σ we associate a map $\mathcal{M}_\sigma$ of $\mathcal{A}^*$ into itself by setting $\mathcal{M}_\sigma(\Theta)=\Theta$ and, by induction on the length of the word A:

$$\mathcal{M}_\sigma(A\alpha) := \sigma(\alpha)\,\mathcal{M}_\sigma(A) \qquad \text{for all } \alpha \in \mathcal{A}\,.$$

One easily verifies that such $\mathcal{M}_\sigma$ is an involution which is also an antimorphism:

$$\text{(32)} \qquad \mathcal{M}_\sigma(AB) = \mathcal{M}_\sigma(B)\,\mathcal{M}_\sigma(A) \qquad \text{for all } A,B \in \mathcal{A}^*\,.$$

REMARK 3.1. *Conversely, let* $\mathcal{M}$ *be an involution of* $\mathcal{A}^*$ *which satisfies* (32); *then* $\mathcal{M}$ *cannot modify the lengths so that we can define* σ: $\mathcal{A} \to \mathcal{A}$ *by setting* $\sigma(\alpha) := \mathcal{M}(\alpha)$ *for all* $\alpha \in \mathcal{A}$; *and one easily checks that* $\mathcal{M} = \mathcal{M}_\sigma$. *In particular, in order to describe the* fixed points *of an involutory antimorphism, we will confine ourselves to work with fixed points of an* $\mathcal{M}_\sigma$. ■

If $F \in \mathcal{A}^*$ is a *fixed point* for $\mathcal{M}_\sigma$, say $\mathcal{M}_\sigma(F) = F$, we will say that F is σ-palindrome, and we will write $F \in \sigma$-PAL; of course, if σ is the identity, the σ-palindromy coincides with the palindromy; and we will write simply $F \in$ PAL.

There exist words which, for any σ, cannot be in σ-PAL; *e.g.* the word $\alpha\beta\beta$ with $\alpha, \beta \in \mathcal{A}$, $\alpha \neq \beta$; so that it does make sense to ask whether a given word is in σ-PAL for some σ. We will deal with the following problem:

PROBLEM B. *Let W be a doubly periodic word. Find the involutions σ (if any!) such that $W \in \sigma$-PAL.* ∎

By slightly enlarging the restriction (3) we will assume:

$$p, q \leq w \leq p + q + \gcd(p, q) \tag{33}$$

so that, with respect to the decomposition $W = ABC = CDA$, the knowledge of a, B, c will uniquely determine the whole word W; see (10) where (because of (33)) it is $E = \Theta$. In particular we will prove:

THEOREM 3.1. *In the framework of* (4) *one has always* $W \in$ PAL. *Under the weaker assumption* (33) *there exists at most one involution σ of* Alph(W) [5] *such that* $W \in \sigma$-PAL; *with respect to the decomposition* (2), *one has*

$$W \in \sigma\text{-PAL} \Leftrightarrow X, Y, Z \in \sigma\text{-PAL}\,; \tag{34}$$

moreover, the condition:

$$\text{the letters of } B \text{ are all different} \tag{35}$$

is sufficient to guarantee the existence of σ. ∎

REMARK 3.2. *In the framework of $b \leq \gcd(a + b, b + c)$, from* (10) *we get*:

$$W \in \sigma\text{-PAL} \Leftrightarrow B, E \in \sigma\text{-PAL}\,.$$

Of course in the limit case $b = \gcd(p, q)$ one has $E = \Theta$ and the result of Theorem 3.1 *will be a consequence of the following Lemma* 3.1; *while if $|E| > 0$ the knowledge of a, B, c does not suffice to characterize W.* ∎

For the proof of Theorem 3.1 we will use the following (obvious) lemma:

LEMMA 3.1. *For any word F, there exists a most one σ, involution of* Alph(F), *such that $F \in \sigma$-PAL. If the letters of F are all different, such a σ does exist.*

PROOF. Let F have the form $F = \alpha_1 \dots \alpha_n$. The involution σ must obviously satisfy $\sigma(\alpha_j) = \alpha_{n-j}$ for $j = 1, \dots, n$; and such a formula uniquely determines σ on Alph(F). However, in general, our formula defines a *multi-valued* map. If the letters of F are all different, the map is single-valued and is obviously an involution. ∎

PROOF OF THEOREM 3.1. Let us firstly remark that, because of (32), one has:

$$W \in \sigma\text{-PAL} \Leftrightarrow A, C \in \sigma\text{-PAL}\,; D = \mathcal{M}_\sigma(B)$$

[5] Alph(F) will denote the set of letters which appear in F.

and the condition $D = \mathcal{M}_\sigma(B)$ can obviously (see (32)) be rewritten in the form $X, Y, Z \in \sigma$-PAL. In particular, we need only to prove that this last property implies $A, C \in \sigma$-PAL. In the framework of (22), this follows immediately from the corresponding formula (which is (12) or (13)); in the general case one will use the characterization of the solutions as «extensions» of shorter ones, and the result follows by induction. The uniqueness of σ follows from Lemma 3.1; the case of (4) follows from (30). ■

We conclude with a problem posed by Robinson (see [9, 1]) and solved by Pedersen (see [8]). In the framework of a binary alphabet, say $\mathcal{A} = \{\alpha, \beta\}$, we are given a palindrome word T such that, for suitably chosen palindrome words R, S one has $RS = T\alpha\beta$. What can be said about the lengths r, s, t of such words? The Pedersen's answer is that

(36) $$r + 2 \text{ and } t + 2 \text{ must be coprime}$$

so that, because of $r + s = t + 2$, if $s > 2$ also $r + 2$ and $s - 2$ are coprime (6). On the other hand it was proved by de Luca and Mignosi [4] that the set of words $W = RS = T\alpha\beta$ coincides with the set of all the finite standard sturmian words of length > 1; moreover the set of the words T satisfying the above equation coincides with the set of the words having two periods p and q coprime, whose length is $p + q - 2$.

Independently from the cardinality of $\mathcal{A}^*$, let us assume that we are given R, S, T, U σ such that:

$$R, S, T \in \sigma\text{-PAL}\,; \qquad U \notin \sigma\text{-PAL}\,; \qquad RS = TU\,.$$

Setting $W := TUR$ (so that $W = RSR$), from (32) and $R, S \in \sigma$-PAL we derive $W \in \sigma$-PAL; again from (32), because of $R, T, TUR \in \sigma$-PAL, we derive $TUR = R\mathcal{M}_\sigma(U)T$, say an equation of type (1) with $B \neq D$ (because $U \notin \sigma$-PAL). From (31) we then get

(37) $$\gcd(r + u, t + u) < u\,.$$

Let us now assume $u = 2$; (37) then coincides with (36); furthermore, from Remark 2.2, we get that the map σ must be the identity and (see (30)) one has $U = \alpha\beta$, $\mathcal{M}_\sigma(U) = \beta\alpha$. Finally, from $U \neq \mathcal{M}_\sigma(U)$, we get $\alpha \neq \beta$.

Acknowledgements

I want to warmly thank Aldo de Luca for the very stimulating discussions.

References

[1] A. de Luca, *A combinatorial property of the Fibonacci word*. Information Processing Letters, 12, 1981, 193-195.

(6) [8] also proved that, for any T, there exists at most one decomposition; we will not deal with such uniqueness.

[2] A. DE LUCA, *Sturmian words: new combinatorial results*. In: J. ALMEIDA - G. M. S. GOMES - P. V. SILVA (eds.), *Semigroups, Automata and Languages*. World Scientific, 1996, 67-83.
[3] A. DE LUCA, *Sturmian words: Structure, Combinatorics, and their Arithmetics*. Theoretical Computer Science, special issue on *Formal Language*, to appear.
[4] A. DE LUCA - F. MIGNOSI, *Some combinatorial properties of sturmian words*. Theoretical Computer Science, 136, 1994, 361-385.
[5] N. J. FINE - S. H. WILF, *Uniqueness theorems for periodic functions*. Proc. Amer. Math. Soc., 16, 1965, 109-114.
[6] M. LOTHAIRE, *Combinatorics on Words*. Addison-Wesley, Reading, MA, 1983.
[7] R. C. LYNDON - M. P. SCHÜTZENBERGER, *On the equation* $a^M = b^N c^P$ *in a free group*. Michigan Math. J., 9, 1962, 289-298.
[8] A. PERDERSEN, *Solution of Problem* E 3156. The American Mathematical Monthly, 95, 1988, 954-955.
[9] P. M. ROBINSON, *Problem* E 3156. The American Mathematical Monthly, 93, 1986, 482.

Dipartimento di Matematica
Università degli Studi di Roma «La Sapienza»
Piazzale A. Moro, 5 - 00185 ROMA