Theoretical Informatics and Applications
Theoret. Informatics Appl. **35** (2001) 437–452

A TEST-SET FOR k-POWER-FREE BINARY MORPHISMS

F. Wlazinski[1]

Abstract. A morphism f is k-power-free if and only if $f(w)$ is k-power-free whenever w is a k-power-free word. A morphism f is k-power-free up to m if and only if $f(w)$ is k-power-free whenever w is a k-power-free word of length at most m. Given an integer $k \geq 2$, we prove that a binary morphism is k-power-free if and only if it is k-power-free up to k^2. This bound becomes linear for primitive morphisms: a binary primitive morphism is k-power-free if and only if it is k-power-free up to $2k+1$

Mathematics Subject Classification. 68R15.

1. Introduction

Research on repetition in words has been initiated by Thue [19, 20] (see also [1, 2]). Thue has shown the existence of infinite words over a binary alphabet without overlap, *i.e.*, without factors of the form $xuxux$ where x is a letter and u is a word. Overlap is one of the elementary patterns with square (uu), cube (uuu) or, more generally, k-power (u^k). Research on avoidable patterns in words is still active (see for instance [12]).

One way to show the avoidability of a pattern is to generate, by iterating a morphism, an infinite word that do not contain this pattern. Thue obtained an infinite overlap-free word over a two-letter alphabet (called Thue–Morse word since the works of Morse [14]) by iterating a morphism μ ($\mu(a) = ab$ and $\mu(b) = ba$). Séébold [18] has shown that the Thue–Morse sequence is the only infinite binary overlap-free sequence starting with a that can be generated by an iterated morphism. Karhumäki [5] and Richomme and Wlazinski [17] have given characterizations of endomorphisms defined on a binary alphabet that generate infinite cube-free words.

Keywords and phrases: Combinatorics on words, k-power-free words, morphisms, test-sets.

[1] LaRIA, Université de Picardie Jules Verne, 5 rue du Moulin Neuf, 80000 Amiens, France; e-mail: `wlazinsk@laria.u-picardie.fr`

Two sufficient conditions for an endomorphism f to generate a k-power-free word when iteratively applied from a letter x, are that $f(x)$ starts with x and that f is a k-power-free morphism, that is, the image by f of any k-power-free word is also k-power-free. In the case of overlap-freeness, Berstel and Séébold [3,18] have shown that, on a two letter alphabet, endomorphisms that generate overlap-free words and endomorphisms that preserve the property of overlap-freeness (called overlap-free morphisms) are the same. This property is no longer true for the k-power-free endomorphism with k at least 3. Richomme and Wlazinski [17] have given an example of a morphism that is not cube-free but generates a cube-free word. For $k \geq 4$, one has just to consider the Fibonnaci word: it is 4-power-free (see [5]) and thus k-power-free for any $k > 4$, but it is generated by the morphism φ ($\varphi(a) = ab$ and $\varphi(b) = a$) which is not k-power-free since $\varphi(b^{k-1}a) = a^k b$.

Many papers deal with the characterization of k-power-free morphisms. For instance, Crochemore [4] has shown that on a three-letter alphabet, a morphism is square-free if and only if the images of all square-free words of length at most 5 are square-free. When k is greater than 3, the results are partial. The particular case of the length-uniform morphisms has been treated by Keränen [6,7]. Some characterizations that depend on the length of morphisms have been given by Leconte [8,9].

In this paper, we are interested in test-sets for k-power-free morphisms. Such a test-set is a set of words which has the particular property that a morphism is k-power-free if and only if the image by this morphism of all the words in the set are k-power-free. In the case of overlap-free endomorphisms on a binary alphabet, a complete characterization of test-sets for such morphisms has been given by Richomme and Séébold [15]. We already have test-sets for cube-free morphisms when the starting alphabet only contains but 2 letters [9,16,17]. We also know [17] that, given two alphabets A and B such that $\mathrm{Card}(A) \geq 3$ and $\mathrm{Card}(B) \geq 2$, and given an integer $k \geq 3$, there is no finite test-set for k-power-free morphisms from A to B. But the question remained open when $\mathrm{Card}(A)$ is 2. We give a positive answer here.

In our approach, we need to study the notion of primitiveness. Section 3 is devoted to primitive morphisms with a particular attention to morphisms on binary alphabet. We revisit a result by Leconte who stated that a binary morphism which is k-power-free up to $\frac{k(k+1)}{2}$ is primitive. We improve this bound by giving an optimal one as well as morphisms which show this optimality (Prop. 3.1).

Section 4 contains results on words equations. The properties given in this section have their own interest, but, we mainly use them in Section 5 to demonstrate the main result of this paper (Th. 5.1) a binary primitive morphism is k-power-free if and only if it is k-power-free up to $2k + 1$. Using Proposition 3.1 and Theorem 5.1, we obtain that a binary morphism is k-power-free if and only if it is k-power-free up to k^2.

2. Definitions and notations

In this section, we recall and introduce some basic notions on words and morphisms.

Let A be an *alphabet*, that is a finite non-empty set of abstract symbols called *letters*. The *Cardinal* of A, *i.e.*, the number of elements of A, is denoted by $\mathrm{Card}(A)$. When $\mathrm{Card}(A) = 2$, A is called a *binary* alphabet. A *word* over A is a finite sequence of letters from A. We denote by ε the *empty word*. The set of words over A equipped with the concatenation of words is a free monoid denoted by A^*.

Let $u = a_1a_2 \dots a_n$ be a non-empty word over A, with $a_i \in A$ $(1 \leq i \leq n)$. The number n of letters of u is called its *length* and is denoted by $|u|$. The length of the empty word is $|\varepsilon| = 0$. The *mirror image* of u, denoted by $\tilde{u}$, is the word $a_n \dots a_2a_1$. In the particular case of the empty word, $\tilde{\varepsilon} = \varepsilon$.

A word u is a *factor* of a word v if $v = v_1uv_2$ for some words v_1, v_2. If $v_1 = \varepsilon$, u is a *prefix* of v. If $v_2 = \varepsilon$, u is a *suffix* of v. If $v_1 \neq \varepsilon$ and $v_2 \neq \varepsilon$, u is an *internal* factor of v.

Let us consider a non-empty word w and its letter-decomposition $x_1 \dots x_n$. For any integers i, j, $1 \leq i \leq j \leq n$, we denote by $w[i \dots j]$ the factor $x_i \dots x_j$ of w. We extend this notation when $i > j$: in this case, $w[i \dots j] = \varepsilon$. We abbreviate $w[i \dots i]$ in $w[i]$. This notation denotes the i^{th} letter of w. For an integer $k \geq 2$, we denote by u^k the concatenation of k occurrences of the word u; $u^0 = \varepsilon$ and $u^1 = u$. A *k-power* is a word of the form u^k with $u \neq \varepsilon$. A word w contains a k-power if at least one of its factors is a k-power. A word is called *k-power-free*, if it does not contain any k-power as a factor. A set of k-power-free words is said *k-power-free.*

A word w is said *primitive* if, for any word z, the equality $w = z^n$ implies $n = 1$. In particular, the empty word ε is not primitive.

The following proposition gives the well-known solutions (see [11]) to three elementary equations in words and will be widely used in the following sections:

Proposition 2.1. *Let A be an alphabet and u, v, w three words over A.*

1. *If $vu = uw$ and $v \neq \varepsilon$ then there exist two words r and s over A and an integer n such that $u = r(sr)^n$, $v = rs$ and $w = sr$.*
2. *If $vu = uv$, then there exists a word w over A and two integers n and p such that $u = w^n$, $v = w^p$.*
3. *If $u^n = v^m$ for two integers n and m such that $(n, m) \neq (0, 0)$ then there exist a word t and two integers p and q such that $u = t^p$ and $v = t^q$.*
4. *Any non-empty word is a power of a unique primitive word.*

We also need three other properties on words. The first one is an immediate consequence of Proposition 2.1(2).

Lemma 2.2. [7, 9] *If a non-empty word v is an internal factor of vv, i.e., if there exist two non-empty words x and y such that $vv = xvy$ then there exist a non-empty word t and two integers $i, j \geq 1$ such that $x = t^i$, $y = t^j$ and $v = t^{i+j}$.*

The following proposition is a well-known result on combinatorics on words:

Proposition 2.3 (Fine and Wilf)**.** [11,12] *Let x and y be two words. If a power of x and a power of y have a common prefix of length at least equal to $|x| + |y|$ $-gcd(|x|, |y|)$ then x and y are powers of the same word.*

As a consequence of Proposition 2.3, we get:

Corollary 2.4 (Keränen)**.** [7] *Let x and y be two words. If a power of x and a power of y have a common factor of length at least equal to $|x| + |y| - gcd(|x|, |y|)$ then there exist two words t_1 and t_2 such that x is a power of t_1t_2 and y is a power of t_2t_1 with t_1t_2 and t_2t_1 primitive words. Furthermore, if $|x| > |y|$ then x is not primitive.*

A *morphism* f from an alphabet A to another alphabet B is a mapping from A^* to B^* such that given any words u and v over A, we have $f(uv) = f(u)f(v)$. When $B = A$, f is called an *endomorphism* on A. When B has no importance, we say that f is defined on A or that f is a morphism on A (in particular, this does not mean that f is an endomorphism). A morphism defined on a binary alphabet is said to be a *binary* morphism. Observe that for a morphism f on A, we necessarily get $f(\varepsilon) = \varepsilon$, and f is uniquely defined by the values of $f(x)$ for all x in A. When A is a binary alphabet, say $A = \{a, b\}$, the *Exchange* endomorphism E is defined by $E(a) = b$ and $E(b) = a$. If X is a set of words, $f(X)$ denotes the set of all the images of words in X.

A morphism f on A is said *k-power-free* (resp. *primitive*) if for every k-power-free (resp. primitive) word w over A, $f(w)$ is k-power-free (resp. primitive). Given a morphism f on A, the *mirror morphism* $\tilde{f}$ of f is defined for all words w over A, by $\tilde{f}(w) = \widetilde{f(\tilde{w})}$. In particular, $\tilde{f}(a) = \widetilde{f(a)}$, $\forall a \in A$. Note that f is k-power-free if and only if $\tilde{f}$ is k-power-free. A morphism f is k-power-free up to m if and only if $f(w)$ is k-power-free whenever w is a k-power-free word of length at most m.

A morphism f is *prefix* (resp. *suffix*) if for all letters x, y with $x \neq y$, $f(x)$ is not a prefix (resp. not a suffix) of $f(y)$. A morphism both prefix and suffix is said *biprefix*.

The property of being biprefix is a necessary condition for a morphism to be k-power-free:

Lemma 2.5. [9] *If a morphism is k-power-free up to $k + 1$, then it is a biprefix morphism.*

The proofs of the two following lemmas are left to the reader:

Lemma 2.6. *Let f be a prefix morphism on an alphabet A, let u and v be words over A and let p_1 and p_2 be prefixes of images by f of some letters in A. If p_1 and p_2 are non-empty or if p_1 and p_2 are not images of a letter then the equality $f(u)p_1 = f(v)p_2$ implies $u = v$ and $p_1 = p_2$.*

Lemma 2.7. *Let f be a suffix morphism on an alphabet A, let u and v be words over A and let s_1 and s_2 be suffixes of images by f of some letters in A. If s_1 and s_2 are non-empty or if s_1 and s_2 are not images of a letter then the equality $s_1f(u) = s_2f(v)$ implies $u = v$ and $s_1 = s_2$.*

3. Primitive morphisms

Let $(t_k)_{k\geq 2}$ be defined by $t_2 = 3$, $t_3 = 4$, $t_k = \frac{k^2}{2}$ if $k \geq 4$ is even and $t_k = \frac{k\times(k-1)}{2} + 2$ if $k \geq 5$ is odd. Note that, when $k \geq 4$, we have $t_k = k\lfloor\frac{k}{2}\rfloor + 2(k \bmod 2)$.

In this section, we prove:

Proposition 3.1. *Let $k \geq 2$ be an integer. If a binary morphism is k-power-free up to t_k then it is primitive.*

This proposition improves the bound $\frac{k\times(k+1)}{2}$ which was given by Leconte in his Ph.D. Thesis [9] (however, the proof we make is similar to his proof). Moreover, the bounds t_k are optimal. Indeed, in what follows, we give for each integer $k \geq 2$ a non-primitive morphism f_k and a shortest k-power-free word u such that $f_k(u)$ contains a k-power. In each case, we have $|u| = t_k$.

- $f_2(a) = a$,$f_2(b) = \varepsilon$ and $u = aba$.
- $f_3(a) = a$,$f_3(b) = baab$ and $u = baab$.
- If $k \geq 4$ and k even, $f_k(a) = a$, $f_k(b) = ba^{k-1}b$ and $u = (ba^{k-1})^{k/2}$.
- If $k \geq 5$ and k odd, $f_k(a) = aba$, $f_k(b) = ba(aba)^{k-3}ab$ and $u = (ba^{k-1})^{(k-1)/2}ba$.

The proof of Proposition 3.1 is based on a previous result due to Lentin and Schützenberger:

Lemma 3.2. [10] *A morphism f on $\{a,b\}$ is primitive if and only if $f(w)$ is primitive for all words $w \in a^*b \cup ab^*$.*

Proof of Proposition 3.1. For an integer $k \geq 2$, let F_1 be the set of all the k-power-free words over $\{a,b\}$ of length at most t_k.

By contradiction, assume that f is not primitive. We are going to show that the image of at least one word in F_1 contains a k-power.

By Lemma 3.2, there exist a non-empty word u, an integer $q \geq 0$ and an integer $n \geq 2$ such that $f(a^qb) = u^n$ or $f(ab^q) = u^n$.

But $f(ab^q)$ is the mirror image of $(\tilde{f} \circ E)(a^qb)$ and $(\tilde{f} \circ E)(F_1)$ k-power-free is equivalent to $f(F_1)$ k-power-free. So, $f(ab^q) = u^n$ implies that $f(F_1)$ contains a k-power is equivalent to $f(a^qb) = u^n$ implies $f(F_1)$ contains a k-power: we may assume that $f(a^qb) = u^n$ without loss of generality.

If $q = 0$ or $f(a) = \varepsilon$, $f(b^{\lceil\frac{k}{2}\rceil})$ contains a k-power and $b^{\lceil\frac{k}{2}\rceil} \in F_1$. Thus we may assume $q \geq 1$ and $f(a) \neq \varepsilon$.

Now, if $|f(a^q)| \geq |f(a)| + |u|$ then, by Proposition 2.3, $f(a)$ and u are powers of the same word. This implies that $f(a)$ and $f(b)$ are powers of the same word: f is not biprefix. Since F_1 contains all the k-power-free words of length at most $k+1$ and by Lemma 2.5, the image of at least one word in F_1 contains a k-power.

Thus, we may assume $|f(a^q)| < |f(a)|+|u|$. We get $|f(a^{q-1})| < |u|$, *i.e.*, $f(a^{q-1})$ is a prefix of u. On the other hand, $|u^{n-1}| = |f(a^q)| + |f(b)| - |u| < |f(ab)|$. Thus u^{n-1} is a suffix of $f(ab)$. Three different cases may occur:

Case 1. $q > k$

Since $f(a^{q-1})$ is a factor of u itself a factor of $f(ab)$, thus $f(ab)$ contains $f(a)^k$ as factor and $ab \in F_1$.

Case 2. $q = k$

Let us consider two cases $|f(a^q)| \leq |u|$ and $|f(a^q)| > |u|$. In the first case, since $f(a^q b) = u^n$, $f(a)^k$ is a factor of u itself a factor of $f(ab)$ with $ab \in F_1$. In the second case, $|f(a^{k-1})| < |u| < |f(a^k)|$ and there exist three words v_1, v_2 and v_3 such that $f(a) = v_1v_2$, $u = f(a^{k-1})v_1 = v_2v_3$ and $f(b) = v_3u^{n-2}$. Since $k \geq 2$, we can define p as the prefix of v_3 such that $v_1v_2 = v_2p$. We get $f(aba) = v_1v_2v_3(v_2v_3)^{n-2}v_1v_2 = v_1(v_2v_3)^{n-2}v_2v_3v_1v_2 = v_1u^{n-2}(f(a))^{k-1}v_1v_2p$ $= v_1u^{n-2}(f(a))^kp$ with $aba \in F_1$.

Case 3. $q < k$

Let us consider three subcases: k even, k odd with $k \geq 5$ and $k = 3$. If k is even $f((a^qb)^{\frac{k}{2}}) = u^{\frac{nk}{2}}$ with $\frac{nk}{2} \geq k$ and $(a^qb)^{\frac{k}{2}} \in F_1$. If k is odd with $k \geq 5$, let us recall that u^{n-1} is a suffix of $f(ab)$. Thus $f(ab(a^qb)^{\frac{k-1}{2}}))$ contains $u^{n-1+\frac{n(k-1)}{2}}$ with $n - 1 + \frac{n(k-1)}{2} \geq k$ and $ab(a^qb)^{\frac{k-1}{2}} \in F_1$. If $k = 3$, let us recall that $f(a^qb) = u^n$ with $q \in \{1, 2\}$, $n \geq 2$ and $|u^{n-1}| < |f(ab)|$. If $n \geq 3$, $f(a^qb)$ contains a 3-power and $a^qb \in F_1$. If $q = 1$ and $n = 2$, $f(abab) = u^4$ and $abab \in F_1$. If $n = 2$, $q = 2$ and $|u| \leq |f(b)|$, $f(b)$ ends with u, thus $f(baab)$ ends with u^3 with $baab \in F_1$. Thus it remains the case $n = 2$, $q = 2$ and $|f(b)| < |u| < |f(ab)|$. There exist two words α_1 and α_2 such that $f(a) = \alpha_1\alpha_2$, $u = \alpha_1\alpha_2\alpha_1 = \alpha_2 f(b)$. Let p be the prefix of $f(b)$ such that $\alpha_1\alpha_2 = \alpha_2 p$. We have $f(b) = p\alpha_1$ and $f(bab) = p\alpha_1\alpha_1\alpha_2p\alpha_1 = p(\alpha_1)^3\alpha_2\alpha_1$ with $bab \in F_1$. □

4. Equations

In this section, we extend the result of Proposition 2.1(1) to an arbitrary number of equations.

Proposition 4.1. *Let $x_1, x_2, \cdots, x_p, y_1, y_2, \cdots, y_p, \alpha$ be $2p+1$ words ($p \geq 2$) such that $y_1 = x_p = \varepsilon$ and $|\alpha| > |y_p| > |y_{p-1}| > \cdots > |y_1|$.*

If $x_1\alpha y_1 = x_2\alpha y_2 = \cdots = x_p\alpha y_p$, then there exist two words r, s and two integers $m, n \geq p - 1$ such that $rs \neq \varepsilon$, $x_1 = (rs)^m$, $y_p = (sr)^m$ and $\alpha = (rs)^n r$.

The situation described in this proposition can be summed up by Figure 1.

Proof. The proof of this proposition is done by induction.

If $p = 2$, we are in the situation of Proposition 2.1(1). So, we have $m = 1$ and $n \geq 1$ since $|\alpha| > |y_2|$. We also have $rs \neq \varepsilon$ since $|rs| = |y_2| > |y_1| = 0$

Assume that this property is true up to an order $p - 1 \geq 2$ and that $x_1\alpha y_1$, $x_2\alpha y_2$, $\cdots$, $x_p\alpha y_p$ are p equal words such that $y_1 = x_p = \varepsilon$ and $|\alpha| > |y_p|$ $> |y_{p-1}| > \cdots > |y_1|$. If we take out the common prefix x_{p-1} from each words $x_q\alpha y_q$, $1 \leq q \leq p - 1$, we are in the situation of the hypotheses of the induction. Thus there exist two words r and s and two integers $m, n \geq p - 2 \geq 1$ such that $rs \neq \varepsilon$, $y_{p-1} = (sr)^m$, $\alpha = (rs)^n r$ and $x = (rs)^m$ where x is the word such that $x_1 = x_{p-1}x$.

The equality $x_{p-1}\alpha y_{p-1} = \alpha y_p$ implies that there exists a word y such that $y_p = yy_{p-1}$ and $x_{p-1}\alpha = \alpha y$. By Proposition 2.1(1), there exist two words u

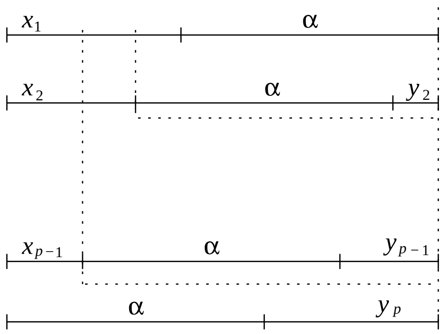

FIGURE 1. Induction cases.

and v and an integer l such that $x_{p-1} = uv$, $y = vu$ and $\alpha = (uv)^l u$. Since $|\alpha| > |x_{p-1}| > |x_p| = 0$, we have $l \geq 1$ and $uv \neq \varepsilon$.

An important fact is that α is a common prefix of $(rs)^{n+1}$ and $(uv)^{l+1}$ of length greater than $|uv| + |rs|$. Indeed $|\alpha| > |x_1| = |x_{p-1}x| = |uv(rs)^m| > |uv| + |rs|$.

Thus, by Proposition 2.3, there exist a word t and two integers n_1 and n_2 such that $rs = t^{n_1}$ and $uv = t^{n_2}$. Since $rs \neq \varepsilon$ and $uv \neq \varepsilon$, we have $n_1, n_2 \geq 1$ and $t \neq \varepsilon$.

Moreover $\alpha = (t^{n_1})^n r = (t^{n_2})^l u$. Since r and u are both prefixes of a power of t, there exist two words t_1, t_2 and two integers n_3, n_4 such that $t = t_1 t_2$, $r = t^{n_3} t_1$, $u = t^{n_4} t_1$. Consequently, $s = t_2 t^{n_1 - n_3 - 1}$ and $v = t_2 t^{n_2 - n_4 - 1}$. It follows that $x_1 = x_{p-1}x = uv(rs)^m = (t_1 t_2)^{n_2 + mn_1}$, $y_p = yy_{p-1} = vu(sr)^m = (t_2 t_1)^{n_2 + mn_1}$ and $\alpha = (rs)^n r = (t_1 t_2)^{nn_1 + n_3} t_1$. Observe that $|\alpha| > |x_1|$ implies $nn_1 + n_3 \geq n_2 + mn_1$. This ends the proof since $n_2 + mn_1 \geq m + 1 \geq p - 1$. $\square$

When we are working with equations on words, they are not necessarily ordered by the length of some of their components. Consequently, the hypotheses and conclusions of the previous proposition are not sufficient. So we generalize the previous result by the two following corollaries. The proof of the first corollary is an immediate consequence of Proposition 4.1: it is done by ordering the terms in an increasing way.

Corollary 4.2. *Let $x_1, x_2, \cdots, x_p, y_1, y_2, \cdots, y_p, \alpha$ be $2p+1$ words ($p \geq 2$) verifying $|y_i| \neq |y_j|$ when $i \neq j$ and $|y_i| < |\alpha|$ ($\forall\ 1 \leq i, j \leq p$) and such that $y_{i_1} = \varepsilon$ and $x_{i_2} = \varepsilon$ for some different integers i_1 and i_2 between 1 and p.*

If $x_1 \alpha y_1 = x_2 \alpha y_2 = \cdots = x_p \alpha y_p$, then there exist two words r, s and two integers $m, n \geq p-1$ such that $rs \neq \varepsilon$, $x_{i_1} = (rs)^m$, $y_{i_2} = (sr)^m$ and $\alpha = (rs)^n r$.

Corollary 4.3. *Let $x_1, x_2, \cdots, x_p, y_1, y_2, \cdots, y_p, y_{p+1}, \alpha$ be $2p+2$ words ($p \geq 2$) verifying $|y_i| \neq |y_j|$ when $i \neq j$ and $|y_i| < |\alpha|$ ($\forall\ 1 \leq i, j \leq p+1$) and such that $y_{i_1} = \varepsilon$ and $x_{i_2} = \varepsilon$ for some different integers i_1 and i_2 between 1 and p. Let α_0 be a suffix of α.*

If $x_1\alpha y_1 = x_2\alpha y_2 = \cdots = x_p\alpha y_p = \alpha_0 y_{p+1}$*, then there exist two words* r*,* s *and three integers* $n, q \geq p$ *and* $m \geq p-1$ *such that* $rs \neq \varepsilon$*,* $x_{i_1} = (rs)^m$*,* $y_{p+1} = (sr)^q$ *and* $\alpha = (rs)^n r$*.*

Proof. By Corollary 4.2, there exist two words r and s and two integers $m, n \geq p-1$ such that $rs \neq \varepsilon$, $x_{i_1} = (rs)^m$, $y_{i_2} = (sr)^m$ and $\alpha = r(sr)^n$.

By hypotheses, $|\alpha| > |\alpha_0| > |x_{i_1}|$. Thus we have $\alpha_0 = v(sr)^k$ where v is a suffix of sr and k is an integer with $p-1 \leq k \leq n$. We also have $\alpha y_{i_2} = (rs)^{m+n} r$.

If $v = r$ then $k < n$ and $r(sr)^k y_{p+1} = \alpha_0 y_{p+1} = \alpha y_{i_2} = r(sr)^{n+m}$. We have $y_{p+1} = (sr)^{n+m-k}$ and $n + m - k \geq m + 1 \geq p$. Since $|y_{p+1}| < |\alpha|$, we also get $n \geq p$.

If $|v| < |r|$ then $r = uv$ for a non-empty word u. Using the prefix of length $|rs|$ of the word $x_{i_1}\alpha = \alpha_0 y_{p+1}$, we obtain $u(vs) = (vs)u$. By Proposition 2.1(2), u and vs are powers of the same word. If $vs = \varepsilon$, $r^k y_{p+1} = \alpha_0 y_{p+1} = \alpha y_{i_2} = r^{n+m+1}$. We have $y_{p+1} = r^{n+m+1-k}$ and $n + m + 1 - k \geq m + 1 \geq p$. Since $|y_{p+1}| < |\alpha|$, we also get $n \geq p$. Now, assume vs non-empty. Since u is non-empty, there exist two words t_1, t_2 and three integers $i, j \geq 1$ and $l \geq 0$ such that $u = (t_1t_2)^i$, $vs = (t_1t_2)^j$, $rs = (t_1t_2)^{i+j}$ and $v = (t_1t_2)^l t_1$, $r = (t_1t_2)^{i+l} t_1$, $s = t_2(t_1t_2)^{j-l-1}$. Thus $y_{p+1} = (t_2t_1)^{(m+n-k)(i+j)+i}$, $x_{i_1} = (t_1t_2)^{m(i+j)}$, $\alpha = (t_1t_2)^{n(i+j)+i+l} t_1$ and each power is greater than $p (\geq 2)$.

If $|v| > |r|$ then $k < n$, $v = ur$ and $s = wu$ for some words u and w. From $x_{i_1}\alpha = \alpha_0 y_{p+1}$, we get $(rw)u = rs = u(rw)$. This case can be solved as the previous one. □

5. Test-sets

This section is devoted to the following theorem, its consequences and its proof:

Theorem 5.1. *Let* $k \geq 2$ *be an integer. A binary primitive morphism* f *is* k*-power-free if and only if it is* k*-power-free up to* $2k+1$*.*

The only seven square-free (*i.e.*, 2-power-free) words over $\{a, b\}$ are ε, a, b, ab, ba, aba and bab, each length being at most 3. Thus actually, a binary morphism (primitive or not) is square-free if and only if it is square-free up to 3.

The case $k = 3$ was already treated in [9, 16, 17]. Thus, we are only interested in the case $k \geq 4$ and our proof is given for this case.

The bound given in Theorem 5.1 is optimal for any integer $k \geq 3$ as we can see using the morphism f_k defined by:

$$\begin{cases} f_k(a) = x(zyb^{k-1}xyb^{k-1}x)^{k-1}zy, \\ f_k(b) = b. \end{cases}$$

In [17], it is shown that, for a word w, the word $f_k(w)$ contains a k-power if and only if w contains $ab^{k-1}ab^{k-1}a$ as a factor. Using Lemma 3.2, we can see that f_k is primitive.

As we have previously noticed when $k = 2$ and as an immediate consequence of Proposition 3.1 and Theorem 5.1 when $k \geq 3$, we get:

Corollary 5.2. *Let $k \geq 2$ be an integer. A binary morphism is k-power-free if it is k-power-free up to t'_k with $t'_2 = 3$, $t'_3 = 7$, $t'_4 = 9$ and $t'_k = t_k$ when $k \geq 5$.*

If we want a more general bound, we can see for instance:

Corollary 5.3. *Let $k \geq 2$ be an integer and let f be a binary morphism. If f is k-power-free up to k^2 then f is k-power-free.*

From an algorithmic point of view, Corollary 5.2 provides a method to verify the k-power-freeness of a given binary morphism. This can also be done in an other way. We can first establish whether the morphism is primitive (see [10,13]). After that, we have to determine if it is k-power-free that is to verify the k-power-freeness of the images of words whose lengths grow linearly with k.

Proof of Theorem 5.1. Let us first recall that the case $k = 2$ is trivial and that the case $k = 3$ was already treated in [9,16,17].

Let $k \geq 4$ be an integer. By definition of k-power-free morphisms, we only have to prove the "if" part of Theorem 5.1.

Let f be a primitive morphism on $\{a, b\}$. We assume:

Assumption 1. f is k-power-free up to $2k + 1$
Assumption 2. f is not k-power-free.

Note that, by Lemma 2.5 and Assumption 1, since $2k + 1 > k + 1$, we have f biprefix.

We are going to show that the two assumptions above are contradictory. For this, by successive contradictions, we will reduce the field of investigation. We end by a final contradiction. We alternate steps of reduction and definitions that describe the combinatoric situation in which we are.

Preliminary definitions

By Assumption 2, there exists a shortest k-power-free word w (not necessarily unique) such that $f(w)$ contains a k-power u^k with $u \neq \varepsilon$. First, note that $|w| \geq 2k + 2$ by Assumption 1. Moreover, since the length of w is minimal, we may assume that $f(w) = \pi u^k \sigma$ where π is a prefix of $f(w[1])$ different from $f(w[1])$ and σ is a suffix of $f(w[|w|])$ different from $f(w[|w|])$.

Reduction 1. $|u| > \max\{|f(a)|, |f(b)|\}$.

Since $|w| \geq 2k + 2$, we have $|w[2 \ldots n-1]| \geq 2k$. Since w is k-power-free, it follows that $w[2 \ldots n-1]$ contains at least two occurrences of the letter a. That is $w[2 \ldots n-1]$ contains a factor of the form $ab^j a$ for some integer $j \geq 0$. Thus $f(ab^j a)$ is a common factor of $(f(ab^j))^2$ and of u^k. If $|f(a)| > |u|$, $|f(ab^j a)| = |f(ab^j)| + |f(a)| > |f(ab^j)| + |u|$. By Corollary 2.4, $f(ab^j)$ is not primitive, *i.e.*, f is not primitive: a contradiction.

Thus $|f(a)| \leq |u|$. Another consequence of $|w[2 \ldots n-1]| \geq 2k$ is that ab or ba is a factor of $w[2 \ldots n-1]$. Consequently, $f(ab)$ or $f(ba)$ is a factor of u^k. If $|f(a)| = |u|$, there exist two words u_1 and u_2 such that $u = u_1 u_2$ and $f(a) = u_2 u_1$. Moreover $f(b)$ is a prefix of $u_2 u^l$ or a suffix of $u^l u_1$ for an integer $l \geq 0$. This implies that f is not a biprefix morphism: a contradiction. We get $|f(a)| < |u|$. In the same way, we obtain $|f(b)| < |u|$.

Intermediate definitions

For each integer j with $0 \leq j \leq k$, we define i_j as the smallest integer such that $|f(w[1 \ldots i_j])| \geq |\pi u^j|$. In particular, we have $i_0 = 1$ and $i_k = |w|$. By Reduction 1, we have $1 = i_0 < i_1 < i_2 < \cdots < i_{k-1} < i_k = |w|$. For any integer q with $0 \leq q \leq k$, there exist some words p_q and s_q such that $f(w[i_q]) = p_q s_q$ and, for all $1 \leq q \leq k$, $u = s_{q-1} f(w[i_{q-1}+1 \ldots i_q - 1]) p_q$ with $p_q \neq \varepsilon$ (by definition of i_q). Furthermore $s_0 \neq \varepsilon$, $p_0 = \pi$ and $s_k = \sigma$.

The previous situation can be summed up by Figure 2.

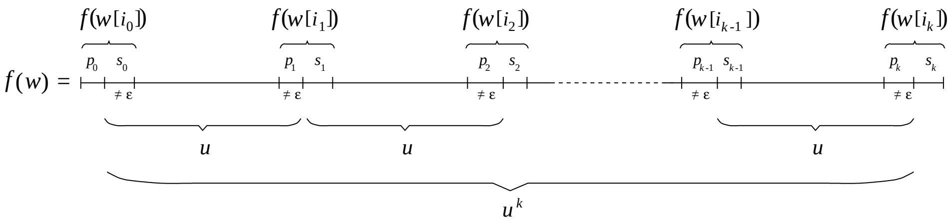

FIGURE 2. Decomposition of a k-power.

Reduction 2. $|p_l| \neq |p_m|$ for all integers $1 \leq l < m \leq k$ and $|s_l| \neq |s_m|$ for all integers $0 \leq l < m \leq k-1$.

Let us first remark that $|s_l| = |s_m|$ for two integers $0 \leq l, m \leq k-1$ implies that $|p_{l+1}| = |p_{m+1}|$ (the converse also holds). Indeed, we know that $u = s_l f(w[i_l+1 \ldots i_{l+1}-1]) p_{l+1} = s_m f(w[i_m+1 \ldots i_{m+1}-1]) p_{m+1}$. Since f is biprefix and by Lemma 2.6, we get that $p_{l+1} = p_{m+1}$.

Thus, we only have to prove that $|p_l| \neq |p_m|$ for all integers $1 \leq l < m \leq k$. By contradiction, assume that there exist two integers l and m such that $1 \leq l < m \leq k$ and $|p_l| = |p_m|$, *i.e.*, $p_l = p_m$.

We first show that we can assume $m = l+1$.

Since $i_l < i_m$, $w[i_l \ldots i_m - 1] \neq \varepsilon$. By Proposition 2.1(4), there exist a unique primitive word z and an integer $q_0 \geq 1$ such that $w[i_l \ldots i_m - 1] = z^{q_0}$. Let v be the word such that $u = v p_l$. We have $u^{m-l} = s_l f(w[i_l+1 \ldots i_m - 1]) p_m$ and so $(p_l v)^{m-l} = f(w[i_l \ldots i_m - 1]) = f(z)^{q_0}$ with $m - l \geq 1$. By Proposition 2.1(3), $f(z)$ and $p_l v$ are powers of the same word. Since f is primitive, $f(z)$ is a primitive word. This implies that $\frac{q_0}{m-l}$ is an integer and $p_l v = f(z)^{q_0/m-l}$. Let us denote $t = z^{q_0/m-l}$. We have $f(w[i_l \ldots i_{l+1}-1]) p_{l+1} = p_l u = p_l v p_l = f(t) p_l$. Since $p_{l+1} \neq \varepsilon$ and $p_l \neq \varepsilon$, by Lemma 2.6, $p_{l+1} = p_l$.

Thus for an integer l, $1 \leq l < k$, $p_l = p_{l+1}$. We will now show that for any integer r such that $1 \leq r \leq l$, we necessarily have $p_r = p_l$. By contradiction, assume that there exist an integer r verifying $1 \leq r \leq l$ and $p_r \neq p_l$. Since $p_l = p_{l+1}$, we can choose r such that $p_{r+1} = p_{r+2} = p_l$. We get $s_r f(w[i_r + 1 \dots i_{r+1} - 1]) = s_{r+1} f(w[i_{r+1} + 1 \dots i_{r+2} - 1])$. Since $s_r \neq f(w[i_r])$ and $s_{r+1} \neq f(w[i_{r+1}])$, by Lemma 2.7, we get $s_r = s_{r+1}$. But one of the two different words p_r or p_{r+1} is a suffix of the other. Thus one of the two different words $f(w[i_r])$ or $f(w[i_{r+1}])$ is a suffix of the other: a contradiction with f biprefix.

In a similar way, we can prove that $p_r = p_l$ for all integer r such that $l + 1 \leq r \leq k$.

Thus we have $p_r = p_l$ for all $1 \leq r \leq k$. Since $u = s_{q-1} f(w[i_{q-1} + 1 \dots i_q - 1])p_q$ for all $1 \leq q \leq k$ and f biprefix, by Lemma 2.7, we get that $s_{q_1-1} = s_{q_2-1}$ and $w[i_{q_1-1} + 1 \dots i_{q_1} - 1] = w[i_{q_2-1} + 1 \dots i_{q_2} - 1]$ for all $1 \leq q_1, q_2 \leq k - 1$.

It follows that $w = w[i_0](w[i_0 + 1 \dots i_1])^{k-1} w[i_0 + 1 \dots i_1 - 1]w[i_k]$. Consequently, since w is k-power-free, we have $w[i_k] \neq w[i_1]$ and $w[i_0] \neq w[i_1]$. Thus the word $w[i_0](w[i_1])^{k-1}w[i_k]$ is k-power-free. But $f(w[i_0](w[i_1])^{k-1}w[i_k]) = p_0(s_0p_1)^k s_k$ with $p_1 \neq \varepsilon$ and $|w[i_0](w[i_1])^{k-1}w[i_k]| \leq 2k + 1$: a contradiction with Assumption 1.

Intermediate definitions

Let us now consider two sets: $I_a = \{0 < r < k \mid w[i_r] = a\}$ and $I_b = \{0 < r < k \mid w[i_r] = b\}$. We have $\max\{\mathrm{Card}(I_a), \mathrm{Card}(I_b)\} \geq \lceil \frac{k-1}{2} \rceil$. Without loss of generality, we may assume $\mathrm{Card}(I_a) \geq \mathrm{Card}(I_b)$. Indeed, the proof of the case $\mathrm{Card}(I_b) \geq \mathrm{Card}(I_a)$ is obtained by exchanging the roles of a and b.

Reduction 3. $\mathrm{Card}(I_a) < \frac{k+3}{3}$.

Let r_1 be the integer in I_a such that $|s_{r_1}| = \max\{|s_r| \mid r \in I_a\}$ and r_2 be the integer in I_a such that $|p_{r_2}| = \max\{|p_r| \mid r \in I_a\}$. Let us remark that $|s_{r_1}| < |u|$ and $|p_{r_2}| < |u|$. For all $1 \leq q \leq k$, we have $u = s_{q-1} f(w[i_{q-1} + 1 \dots i_q - 1])p_q$. Thus, for all $j \in I_a$, there exist two words x_j and y_j such that $x_j p_j = p_{r_2}$ and $s_j y_j = s_{r_1}$. We have $y_{r_1} = \varepsilon$, $x_{r_2} = \varepsilon$.

Since $f(a) = p_j s_j$, all the $\mathrm{Card}(I_a)$ (≥ 2) terms of the form $x_j f(a) y_j$ are equal. Moreover, they fulfill assumptions of Corollary 4.2 with $\alpha = f(a)$. Thus there exist two words r, s and two integers $m, n \geq \mathrm{Card}(I_a) - 1$ such that $rs \neq \varepsilon$, $x_{r_1} = (rs)^m$, $y_{r_2} = (sr)^m$ and $f(a) = (rs)^n r$.

But x_{r_1} is a suffix of $s_{r_1-1} f(w[i_{r_1-1} + 1 \dots i_{r_1} - 1])$ and we have $|x_{r_1}| < |f(a)|$, thus x_{r_1} is a suffix of the image of a k-power-free word ab^{l_1} with $0 \leq l_1 < k$. In the same way, y_{r_2} is a prefix of the image of a k-power-free word $b^{l_2}a$ with $0 \leq l_2 < k$. Thus $|ab^{l_1}ab^{l_2}a| \leq 2k + 1$ and $f(ab^{l_1}ab^{l_2}a)$ contains $(rs)^{2m+n}$ with $2m + n \geq 3 \times \mathrm{Card}(I_a) - 3$. When $3 \times \mathrm{Card}(I_a) - 3 \geq k$, we obtain a contradiction with Assumption 1.

Reduction 4. $k = 4$ and $\mathrm{Card}(I_a) = 2$.

Since $\mathrm{Card}(I_a) \geq \lceil \frac{k-1}{2} \rceil$, Reduction 3 implies that only three cases are possible: $k = 7$ with $\mathrm{Card}(I_a) = 3$, $k = 5$ with $\mathrm{Card}(I_a) = 2$ and $k = 4$ with $\mathrm{Card}(I_a) = 2$. We are going to show that the first two cases lead to a contradiction.

In these cases, we have $\mathrm{Card}(I_a) = \mathrm{Card}(I_b) \geq 2$. Once again, a and b play symmetrical roles. Without loss of generality, we may assume $\min\{|p_i| \mid i \in I_a\}$

$> \min\{|p_i| \mid i \in I_b\}$. For $i \in I_a$, let X_i be the word such that $X_i p_i$ is the suffix of u of length $M_1 = \max\{|p_i| \mid i \in I_a\}$ and let Y_i be the word such that $s_i Y_i$ is the prefix of u of length $M_2 = \max\{|s_i| \mid i \in I_a\}$. Of course, there are two different integers j_1 and j_2 in I_a such that $|p_{j_1}| = M_1$ and $|s_{j_2}| = M_2$. That is $X_{j_1} = \varepsilon$ and $Y_{j_2} = \varepsilon$.

Now, let j be the integer in I_b such that $|p_j| = \min\{|p_i| \mid i \in I_b\}$. For any l in $I_b \setminus \{j\}$, $|p_l| > |p_j|$ and $|s_l| < |s_j|$. Since p_l and $f(w[i_j - 1])p_j$ are suffixes of uu and since s_j and s_l are prefixes of u, if $w[i_j - 1] = b$, then $f(b) = p_l s_l$ is an internal factor of $f(w[i_j - 1])p_j s_j = f(bb)$. By Lemma 2.2, $f(b)$ is not primitive: a contradiction with f primitive.

Thus $w[i_j - 1] = a$. Moreover, by definition of j, there exists a word α_0 such that $\alpha_0 p_j$ is the suffix of u of length M_1. Since $|\alpha_0| < M_1 < |f(a)|$, α_0 is a suffix of $f(a)$. The word $\alpha_0 p_j s_{j_2}$ equals any of the $\mathrm{Card}(I_a)$ words $X_i p_i s_i Y_i = X_i f(a) Y_i$ where $i \in I_a$.

For $i \in I_a$, $|s_i| \le |s_{j_2}|$. It follows that $|p_i| \ge |p_{j_2}|$, *i.e.* $|p_{j_2}| = \min\{|p_i| \mid i \in I_a\}$. Consequently $|X_i| \le |X_{j_2}|$ and $|p_j s_{j_2}| < |p_{j_2} s_{j_2}| = |f(a)|$. By Corollary 4.3, there exist two words r and s and three integers $n, q \ge \mathrm{Card}(I_a)$ and $m \ge \mathrm{Card}(I_a) - 1$ such that $rs \neq \varepsilon$, $X_{j_2} = (rs)^m$, $p_j s_{j_2} = (sr)^q$ and $f(a) = (rs)^n r$.

The word X_{j_2} is a suffix of the k-power-free word $f(w[1 \dots i_{j_2} - 1])$. Moreover $|X_{j_2}| < M_1 < |f(a)|$: there exists an integer $n_1 < k$ such that X_{j_2} is a suffix of $f(ab^{n_1})$. The word $p_j s_{j_2}$ is a prefix of the k-power-free word $f(w[i_j \dots |w|])$. Since $|p_j s_{j_2}| < |f(a)|$, there exists an integer $n_2 < k$ such that $p_j s_{j_2}$ is a prefix of $f(b^{n_2} a)$. Let $x = ab^{n_1} ab^{n_2} a$. We have $|x| \le 2k + 1$ and $f(x)$ contains $(rs)^{m+n+q}$. Since $m + n + q \ge 3 \times \mathrm{Card}(I_a) - 1 \ge 2 \times \mathrm{Card}(I_a) + 1 = k$, $f(x)$ contains a k-power: a contradiction with Assumption 1.

Intermediate definitions

According to Reduction 4, $k = 4$ and $\mathrm{Card}(I_a) = 2$. So $\mathrm{Card}(I_b) = 1$. Let us call j_1, j_2 and j_3 the integers such that $I_a = \{j_1, j_2\}$, $I_b = \{j_3\}$ and $|p_{j_2}| > |p_{j_1}|$.

Since $u = s_{q-1} f(w[i_{q-1} + 1 \dots i_q - 1]) p_q$ for all $1 \le q \le 4$, we will work with three equal terms of the form uu (see Fig. 3).

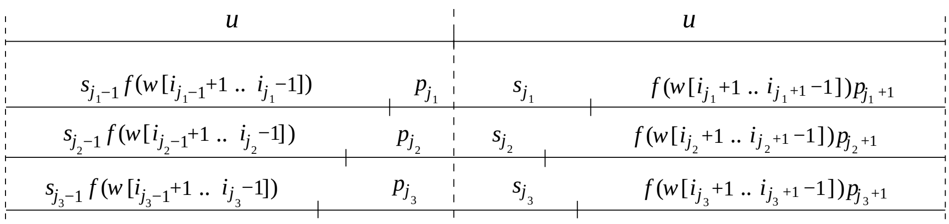

FIGURE 3. Equations.

In what follows, we will very often have to use some of the prefixes of the equalities $u = s_{j_1} f(w[i_{j_1} + 1 \dots i_{j_1+1} - 1]) p_{j_1+1} = s_{j_2} f(w[i_{j_2} + 1 \dots i_{j_2+1} - 1]) p_{j_2+1} =$

$s_{j_3}f(w[i_{j_3}+1\dots i_{j_3+1}-1])p_{j_3+1}$ as well as some of the suffixes of the equalities $u = s_{j_1-1}f(w[i_{j_1-1}+1\dots i_{j_1}-1])p_{j_1} = s_{j_2-1}f(w[i_{j_2-1}+1\dots i_{j_2}-1])p_{j_2} = s_{j_3-1}f(w[i_{j_3-1}+1\dots i_{j_3}-1])p_{j_3}$.

Since $|p_{j_2}| > |p_{j_1}|$, let x be the word such that $xp_{j_1} = p_{j_2}$. We know that x is a suffix of $s_{j_1-1}f(w[i_{j_1-1}+1\dots i_{j_1}-1])$ and $|x| < |f(a)|$, thus x is a suffix of the image of a 4-power-free word of the form ab^{l_3} with $0 \le l_3 < 4$. In the same way, let y be the word such that $s_{j_1} = s_{j_2}y$: y is a prefix of the image of a 4-power-free word of the form $b^{l_4}a$ with $0 \le l_4 < 4$. Thus we have $|ab^{l_3}ab^{l_4}a| \le 9$ and $xf(a) = xp_{j_1}s_{j_1} = p_{j_2}s_{j_2}y = f(a)y$. Thus, by Proposition 2.1(1), there exist two words r and s and an integer i such that $rs \neq \varepsilon$, $x = rs$, $y = sr$ and $f(a) = (rs)^i r$. Since $|x| < |p_{j_2}| < |f(a)|$, we have $i \ge 1$.

Reduction 5. $i = 1$, $r \neq \varepsilon$ and $s \neq \varepsilon$.

If $i \ge 2$, we have $|ab^{l_3}ab^{l_4}a| \le 9$ and $f(ab^{l_3}ab^{l_4}a)$ contains $xf(a)y$ and thus $(rs)^4$: a contradiction with Assumption 1. Thus $i = 1$. Since $f(a) = rsr$, $x = rs$ and $|x| < |f(a)|$, we get $r \neq \varepsilon$. Since f is primitive, we have $s \neq \varepsilon$.

Reduction 6. rs (resp. sr) is not an internal factor of $(rs)^2$ (resp. of $(sr)^2$).

For instance, if rs is an internal factor of $(rs)^2$, by Lemma 2.2, $rs = t^{i_0}$ for a non-empty word t and an integer $i_0 \ge 2$. We have $|ab^{l_3}a| \le 9$ and $f(ab^{l_3}a)$ contains $(rs)^2$ and thus t^4: a contradiction with Assumption 1.

Reduction 7. $w[i_{j_1}-1] = b$ and $w[i_{j_2}+1] = b$.

If $w[i_{j_1}-1] = a$ or $w[i_{j_2}+1] = a$, $f(a)$ is an internal factor of $f(a)f(a)$. By Lemma 2.2, $f(a)$ is not primitive: a contradiction with f primitive.

Reduction 8. $|f(b)| > |sr|$.

In the case $|f(b)| = |sr|$, we have $rs = x = f(b)$: a contradiction with f biprefix.

Let us assume that $|f(b)| < |sr|$. We have $|s_{j_1}| \ge |y| = |sr| > |f(b)| > |s_{j_3}|$. Let z be the word such that $s_{j_1} = s_{j_3}z$: z is a prefix of $f(w[i_{j_3}+1\dots i_{j_3+1}-1])p_{j_3+1}$. Since $|z| < |f(a)|$, z is a prefix of a 4-power-free word of the form $f(b^{l_5}a)$ for an integer $0 \le l_5 < 4$.

If $|s_{j_3}f(b^{l_5})| \ge |s_{j_2}f(b)|$, $s_{j_2}f(w[i_{j_2}+1]) = s_{j_2}f(b)$ is a prefix of $s_{j_3}f(b^{l_5})$. We have $|s_{j_3}| < |f(b)|$ and $s_{j_3} \neq s_{j_2}$. Two cases are possible: $f(b)$ is a suffix of s_{j_2} or $f(b)$ is an internal factor of $f(b)f(b)$. The first case is in contradiction with f biprefix. By Lemma 2.2, the second case implies that $f(b)$ is not primitive: a contradiction with f primitive.

Thus $|s_{j_3}f(b^{l_5})| < |s_{j_2}f(b)|$. Since $|f(b)| < |sr|$, $|s_{j_3}f(b^{l_5})| < |s_{j_1}|$. There exists a prefix α' of $f(a)$ such that $z = f(b^{l_5})\alpha'$. Note that $|s_{j_1}| = |s_{j_3}f(b^{l_5})\alpha'| < |s_{j_2}f(b)\alpha'|$, so $|f(b)\alpha'| > |s_{j_1}| - |s_{j_2}| = |rs|$. Consequently, the suffix sr of s_{j_1} is a suffix of $f(b)\alpha'$. Since $w[i_{j_1}-1] = b$, $f(b)$ is a suffix of $x = rs$. From α' prefix of $f(a) = rsr$, it follows that sr is a factor of $(rs)^2r$. If $\alpha' \neq r$, sr is an internal factor of $(sr)^2$: a contradiction with Reduction 6. Thus $\alpha' = r$. From $rsrsr = rsf(a) = xp_{j_1}s_{j_1} = p_{j_2}s_{j_3}f(b^{l_5})r$, we get $p_{j_2}s_{j_3}f(b^{l_5}) = rsrs$. Moreover $p_{j_2}s_{j_3}f(b^{l_5})$ is a suffix of $w[1\dots i_{j_3}+l_5]$ which is 4-power-free. We have $w[i_{j_3}] = b$

and $|p_{j_2}| \leq |f(a)|$. Thus there exists an integer $l_6 \geq 0$ verifying $l_5+l_6 < 4$ and such that $(rs)^2$ is a suffix of $f(ab^{l_5+l_6})$. We have $|ab^{l_5+l_6}ab^{l_4}a| \leq 9$ and $f(ab^{l_5+l_6}ab^{l_4}a)$ contains $(rs)^4$: a contradiction with Assumption 1.

Reduction 9. $|s_{j_3}| > |s_{j_1}|$.

By Reduction 2, $|s_{j_3}| \neq |s_{j_1}|$. Let us assume that $|s_{j_3}| < |s_{j_1}|$. Since $w[i_{j_1}-1] = b$ and $|f(b)| > |rs|$, the word $x = rs$ is a suffix of $f(b)$. Let z be the word such that $s_{j_1} = s_{j_3}z$. We have $|z| \leq |s_{j_1}| < |f(a)| = |rsr|$. The word rsz is a suffix of $f(b)z = p_{j_3}s_{j_3}z$. Furthermore $|p_{j_2}| > |x| = |rs|$. Since p_{j_2} and p_{j_3} are both suffixes of u, rsz is a suffix of $p_{j_2}s_{j_3}z = xp_{j_1}s_{j_1} = xf(a) = (rs)^2r$.

If $z \neq r$, rs is an internal factor of $(rs)^2$: a contradiction with Reduction 6.

If $z = r$, $p_{j_2}s_{j_3} = (rs)^2$ is a suffix of $w[1\ldots i_{j_3}]$. Since $w[i_{j_3}] = b$, $|f(b)| > |sr|$ and $|f(a)| > |sr|$, $(rs)^2$ is a suffix of $f(ab)$ or of $f(bb)$. Thus $f(bbab^{l_4}a)$ or $f(abab^{l_4}a)$ contains $(rs)^2f(a)y$ and so $(rs)^4$. But $|bbab^{l_4}a| \leq 9$ and $|abab^{l_4}a| \leq 9$: a contradiction with Assumption 1.

Reduction 10. $|p_{j_3}| > |p_{j_2}|$.

When $|p_{j_3}| < |p_{j_2}|$, beginning with $w[i_{j_2}+1] = b$ and considering prefixes instead of suffixes, by a proof similar to Reduction 9, we get that $f(ab^{l_3}aba)$ or $f(ab^{l_3}abb)$ contains $(sr)^4$ with $|ab^{l_3}aba| \leq 9$ and $|ab^{l_3}abb| \leq 9$.

Intermediate definitions

We have $|f(b)| > |p_{j_2}s_{j_1}| = |rsrsr|$.

Let z_1 be the word such that $z_1p_{j_2} = p_{j_3}$. We have $|z_1| < |f(b)|$ and z_1 is a suffix of $s_{j_2-1}f(w[i_{j_2-1}+1\ldots i_{j_2}-1])$. Thus z_1 is a suffix of a word of the form $f(ba^{l_7})$ for an integer $0 \leq l_7 < 3$. Let z_2 be the word such that $s_{j_1}z_2 = s_{j_3}$. We have $|z_2| < |f(b)|$ and z_2 is a prefix of $f(w[i_{j_1}+1\ldots i_{j_1+1}-1])p_{j_1+1}$. Thus z_2 is a prefix of a word of the form $f(a^{l_8}b)$ for an integer $0 \leq l_8 < 3$.

Let β_1 and β_2 be the non-empty words such that $s_{j_2}\beta_1 = s_{j_3}$ and $p_{j_3} = \beta_2p_{j_1}$. We have $f(b) = \beta_2p_{j_1}s_{j_2}\beta_1$. Since $p_{j_1}s_{j_2}sr = p_{j_1}s_{j_2}y = p_{j_1}s_{j_1} = f(a) = rsr$, we have $p_{j_1}s_{j_2} = r$ and $f(b) = \beta_2r\beta_1$. Now, observe that β_1 is a prefix of $f(w[i_{j_2}+1\ldots|w|])$. Since $w[i_{j_2}+1] = b$, β_1 is a prefix of $f(b)$. In a similar way, since β_2 is a suffix of $f(w[1\ldots i_{j_1}-1])$ and since $w[i_{j_1}-1] = b$, β_2 is a suffix of $f(b)$.

It follows that there exists a word β_0 such that $f(b) = \beta_1\beta_0\beta_2$.

Note that $\beta_1 = yz_2 = srz_2$ and $\beta_2 = z_1x = z_1rs$.

Reduction 11. $|\beta_1| < |r| + |\beta_2|$ and $|\beta_2| < |r| + |\beta_1|$.

If $|\beta_1| = |r| + |\beta_2|$, then $\beta_1 = \beta_2r$ and $f(b) = (\beta_1)^2$: a contradiction with f primitive.

If $|\beta_1| > |r| + |\beta_2|$, by Proposition 2.1(1), the equality $(\beta_2r)\beta_1 = \beta_1(\beta_0\beta_2)$ implies that there exist two words v_1, v_2 and an integer $j \geq 1$ such that $\beta_2r = v_1v_2$, $\beta_1 = (v_1v_2)^jv_1$ and $\beta_0\beta_2 = v_2v_1$.

We have $|ba^{l_7}ab| \leq 9$ and $f(ba^{l_7}ab)$ contains the word $z_1 f(ab) = z_1 p_{j_2} s_{j_2} f(b) = p_{j_3} s_{j_2} f(b) = \beta_2 p_{j_1} s_{j_2} f(b) = \beta_2 r \beta_1 \beta_0 \beta_2 = (v_1 v_2)^{j+2}$. If $j \geq 2$, we get a contradiction with Assumption 1. Thus we may assume $j = 1$.

Note that β_1 starts with $y = sr$ and with v_1.

If $|v_1| < |sr|$, v_1 is a prefix of sr. We also have that rsr is a suffix of $\beta_2 r = v_1 v_2$. If $v_1 = s$, $v_1 v_2 v_1$ ends with $(rs)^2$ and $f(bab)$ contains $(rs)^4$. If $v_1 \neq s$, since $\beta_1 = v_1 v_2 v_1$ ends with $rsrv_1$ and also with β_2 thus with rs, we obtain that rs is an internal factor of $(rs)^2$: a contradiction with Reduction 6.

Thus $|v_1| \geq |sr|$. Since we have $|u| > |f(b)|$, we can consider the prefix v_0 of $f(w[i_{j_3} + 1 \ldots i_{j_3+1} - 1])p_{j_3+1}$ of length $|f(b)| - |s_{j_3}|$, *i.e.*, $s_{j_3} v_0$ is the prefix of u of length $|f(b)|$. There exists an integer $0 \leq l_9 < 4$ such that v_0 is a prefix of a word of the form $f(a^{l_9} b)$. Since $s_{j_3} v_0$ is a prefix of $s_{j_2} f(w[i_{j_2+1}]) = s_{j_2} f(b) = s_{j_2} \beta_1 \beta_0 \beta_2 = s_{j_3} \beta_0 \beta_2$, v_0 is a prefix of $\beta_0 \beta_2 = v_2 v_1$. We have $|f(b)| = |\beta_2 p_{j_1} s_{j_2} \beta_1| = |s_{j_3} v_0| = |s_{j_2} \beta_1 v_0|$. Thus $|v_0| = |\beta_2 p_{j_1}| = |\beta_2 p_{j_1} s_{j_2}| - |s_{j_2}| = |\beta_2 r| - |s_{j_2}| = |v_1 v_2| - |s_{j_2}| \geq |sr| + |v_2| - |s_{j_2}| = |s| + |p_{j_1}| + |v_2| \geq |v_2|$. So v_0 starts with v_2. We have $|ba^{l_7} aba^{l_9} b| \leq 9$ and $f(ba^{l_7} aba^{l_9} b)$ contains $z_1 f(ab) v_0 = z_1 rsr f(b) v_0 = \beta_2 r f(b) v_0 = (v_1 v_2)^3 v_1 v_0$ which contains $(v_1 v_2)^4$: a contraction with Assumption 1.

The cases $|\beta_1| < |r| + |\beta_2|$ and $|\beta_2| < |r| + |\beta_1|$ are symmetrical. In the same way that the case $|\beta_1| < |r| + |\beta_2|$, considering suffixes instead of prefixes and prefixes instead of suffixes, the case $|\beta_2| < |r| + |\beta_1|$ leads to a contradiction with the assumptions.

Reduction 12. $|\beta_1| \neq |\beta_2|$.

If $|\beta_1| = |\beta_2|$, $z_1 rs = \beta_2 = \beta_1 = srz_2$ and $\beta_0 = r$ that is $f(b) = \beta_1 r \beta_2 = z_1 rsrz_1 rs$. Let us recall that z_1 is a suffix of a word of the form $f(ba^{l_7})$ for an integer $0 \leq l_7 < 3$. Thus $f(ba^{l_7} ab)$ has $(z_1 rsr)^2 rs$ as suffix. We have $|u| > |f(b)| > |\beta_2| + |s_{j_3}| > |r| + |s_{j_3}| = |s_2 \beta_1 r|$.

Since $w[i_{j_2} + 1] = b$ and since r is a prefix of $f(b)$, $s_{j_2} \beta_1 r$ is a prefix of u. Since $s_{j_2} \beta_1 r = s_{j_2} srz_2 r = s_{j_1} z_2 r$, $u = s_{j_1} f(w[i_{j_1} + 1 \ldots i_{j_1+1} - 1])p_{j_1+1}$ and $w[i_{j_1}] = a$, $z_2 r$ is a prefix of a word of the form $f(a^{l_{10}} b)$ for an integer $0 \leq l_{10} < 3$. It follows that $f(ba^{l_7} abaa^{l_{10}} b)$ contains $(z_1 rsr)^2 z_1 rsrsrz_2 r = (z_1 rsr)^4$: a contradiction with Assumption 1 since $|ba^{l_7} abaa^{l_{10}} b| \leq 9$.

Final Contradiction

If $|\beta_2| < |\beta_1| < |r| + |\beta_2|$, since $\beta_1 \beta_0 \beta_2 = \beta_2 r \beta_1$, we have $\beta_1 = \beta_2 r'$ for a non-empty prefix r' of r different from r. Let us recall that rs is both a suffix of β_2 and of β_1. It follows that rsr' has rs as a suffix, that is, rs is an internal factor of $(rs)^2$: a contradiction with Reduction 6.

The case $|\beta_1| < |\beta_2| < |r| + |\beta_1|$ is symmetrical to the case $|\beta_2| < |\beta_1| < |r| + |\beta_2|$ and leads to a final contradiction considering suffixes instead of prefixes and prefixes instead of suffixes.

Acknowledgements. The author wishes to thank Gwénaël Richomme for his suggestions. His many rereadings and comments have helped to improve the results and the presentation of this paper. The author would like also to thank Patrice Séébold for his remarks.

References

[1] J. Berstel, Axel thue's work on repetitions in words, edited by P. Leroux and C. Reutenauer, *Séries formelles et combinatoire algébrique*. LaCIM, University of Québec at Montréal (1992) 65-80.

[2] J. Berstel, *Axel thue's papers on repetitions in words: A translation*, Technical Report 20. LaCIM, University of Québec at Montréal (1995).

[3] J. Berstel and P. Séébold, A characterization of overlap-free morphisms. *Discrete Appl. Math.* **46** (1993) 275-281.

[4] M. Crochemore, Sharp characterizations of squarefree morphisms. *Theoret. Comput. Sci.* **18** (1982) 221-226.

[5] J. Karhumäki, On cube free ω-words generated by binary morphisms. *Discrete Appl. Math.* **5** (1983) 279-297.

[6] V. Keränen, On the k-repetition freeness of length uniform morphisms over a binary alphabet. *Discrete Appl. Math.* **9** (1984) 297-300.

[7] V. Keränen, On the k-freeness of morphisms on free monoids. *Ann. Acad. Sci. Fenn.* **61** (1986).

[8] M. Leconte, A characterization of power-free morphisms. *Theoret. Comput. Sci.* **38** (1985) 117-122.

[9] M. Leconte, *Codes sans répétition*, Ph.D. Thesis. LITP Université P. et M. Curie (1985).

[10] A. Lentin and M.P. Schützenberger, A combinatorial problem in the theory of free monoids, edited by R.C. Bose and T.E. Dowling. Chapell Hill, North Carolina Press edition, *Comb. Math.* (1945) 112-144.

[11] M. Lothaire, *Combinatorics on words*, Vol. 17 of *Encyclopedia of Mathematics*, Chap. 9, Equations on words. Addison-Wesley (1983). Reprinted in 1997 by Cambridge University Press in the Cambridge Mathematical Library.

[12] M. Lothaire, *Algebraic combinatorics on words*. Cambridge University Press (to appear).

[13] V. Mitrana, Primitive morphisms. *Inform. Process. Lett.* **64** (1997) 277-281.

[14] M. Morse, Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.* **22** (1921) 84-100.

[15] G. Richomme and P. Séébold, Characterization of test-sets for overlap-free morphisms. *Discrete Appl. Math.* **98** (1999) 151-157.

[16] G. Richomme and F. Wlazinski, About cube-free morphisms, edited by H. Reichel and S. Tison, *STACS 2000*. Springer-Verlag, *Lecture Notes in Comput. Sci.* **1770** (2000) 99-109.

[17] G. Richomme and F. Wlazinski, Some results on k-power-free morphisms. *Theoret. Comput. Sci.* **273** (2002) 119-142.

[18] P. Séébold, Sequences generated by infinitely iterated morphisms. *Discrete Appl. Math.* **11** (1985) 255-264.

[19] A. Thue, Uber unendliche Zeichenreihen. *Videnskapsselskapets Skrifter*, I. Mat.-naturv. Klasse, Kristiania (1906) 1-22.

[20] A. Thue, Uber die gegenseitige Lage gleigher Teile gewisser Zeichenreihen. *Videnskapsselskapets Skrifter*, I. Mat.-naturv. Klasse, Kristiania (1912) 1-67.

Communicated by J. Karhumäki.
Received May, 2001. Accepted November, 2001.