Unavoidable sets of constant length

Jean-Marc Champarnaud, Georges Hansel, Dominique Perrin

Unavoidable sets of constant length

Jean-Marc Champarnaud Georges Hansel
Dominique Perrin

October 28, 2003

Abstract

A set of words X is called unavoidable on a given alphabet A if every infinite word on A has a factor in X. For $k, q \geq 1$, let $c(k, q)$ be the number of conjugacy classes of words of length k on q letters. An unavoidable set of words of length k on q symbols has at least $c(k, q)$ elements. We show that for any $k, q \geq 1$ there exists an unavoidable set of words of length k on q symbols having $c(k, q)$ elements.

1 Introduction

A word t is said to avoid a word p if p is not a factor of t, i.e. if the pattern p does not appear in the text t. For example the word `abracadabra` avoids `baba`. The set of all words avoiding a given set X of words has been of interest in several contexts including the notion of a system of finite type in symbolic dynamics (see [7] for example) or in enumerative combinatorics (see [16] for example).

A set X of words on the alphabet A is called *unavoidable* on A if any long enough word on the alphabet A has a factor in X. The notion of an unavoidable set has been the subject of several results since its introduction by M.P. Schützenberger in [15]. In particular L. Rosaz has shown in [13] that there is a finite number of types of unavoidable sets having a given number of elements.

It is easy to see that an unavoidable set of words of constant length k on some alphabet A has to contain at least one word of each conjugacy class of words of length k on this alphabet. Thus the minimal number $m(k, q)$ of elements of an unavoidable set of words of length k on q letters is greater than or equal to the number $c(k, q)$ of conjugacy classes of words of length k on q letters. It has been shown by Schützenberger that it is asymptotically true that $m(k, q) \sim c(k, q)$ (actually both numbers are asymptotically equivalent to q^k/k). Later on, answering a conjecture of S. W. Golomb, J. Mykkeltveit proved that actually, $m(k, q) = c(k, q)$ [11]. His proof uses trigonometric sums (see the last section).

Our main result here is a new proof of this equality, namely that for every $k \geq 1$ and $q \geq 1$, there exists an unavoidable set with $c(k, q)$ elements (Theorem 1). We actually obtained this result without being aware of J. Mykkeltveit's work (see below the script of the story).

It may be convenient for the reader to formulate the statement in terms of graphs. A *feedback vertex set* in a directed graph G is a set F of vertices containing at least one vertex from every cycle in G. Consider, for $k \geq 1$, the De Bruijn graph B_k of order k on the alphabet A whose vertices are the words of length k on A and the edges are the pairs (au, ub) for all $a, b \in A$ and $u \in A^{k-1}$. It is easy to see that a set of words of length k is unavoidable if the corresponding set of vertices is a feedback vertex set of the graph B_k. Thus, the problem of determining an unavoidable set of words of length k of minimal size is the same as determining the minimal size of a feedback vertex set in B_k. The problem is, for general directed graphs, known to be NP-complete (see [4] for example).

The genesis of our work has a rich background and it has benefited from the help of several colleagues that we would like to thank. First of all, it was Christopher Saker who pointed out to the third author a mistake in [8] where it is asserted erroneously that $m(2, 5) = 9$ although $c(2, 5) = 8$ (Exercise 5.1.4 page 99). He was able to verify that $m(k, n) = c(k, n)$ for $k = 2$ and $n \leq 7$. It was conjectured then by P. Higgins and C. Saker (see [14]) that one has actually $m(k, q) = c(k, q)$. In a joint work with Guoniu Han [5], the third author was able to reach $n = 9$ and to discuss some generalizations on systems of finite type (see at the end of this paper). The first two authors found a first proof of the main result (see [1] and the last section) and the final form given here is the result of this extended cooperation. We learnt after completing a first version that the conjecture had already been proved by J. Mykkelveit [11]. We would like to thank David Penman and Christopher Saker again for pointing out this reference. The authors wish also to express their thanks to Donald Knuth for reading the paper and suggesting several improvements.

We begin this paper by some preliminaries on words (see [8] for a more general introduction). We state in particular elementary properties of Lyndon words used in the sequel. The following section contains the proof of the main result.

2 Preliminaries

Let A be a finite set with a linear order $<$. A *word* on the alphabet A is a finite sequence of elements of A, including the empty sequence called the *empty word*. We denote by A^* the set of all words on the alphabet A. The length of a word $w \in A^*$ is denoted by $|w|$. A word p is said to be a *prefix* of a word w if there exists a word u such that $w = pu$. The prefix p is *proper* if $p \neq w$. The definition of a *suffix* is symmetrical. A word x is a *factor* of a word w if there exist words p, q such that $w = pxq$.

We shall also use infinite words. A two-sided infinite word on A is a sequence $(a_n)_{n\in\mathbb{Z}}$. A word x is a factor of a two-sided infinite word $w = (a_n)_{n\in\mathbb{Z}}$ if there exists an index $n \in \mathbb{Z}$ such that $x = a_n a_{n+1} \cdots a_{n+k-1}$ where $k = |x|$. A two-sided infinite word $(a_n)_{n\in\mathbb{Z}}$ is said to be *periodic* if there is an integer $p \geq 1$ such that $a_{n+p} = a_n$ for all $n \in \mathbb{Z}$. Thus a two-sided infinite word is periodic if it is made of the repetition of a finite word u. We denote by u^ζ the set of periodic words of this form.

The same notions hold for one-sided infinite words which are sequences $(a_n)_{n\in\mathbb{N}}$. For $x, y \in A^*$ with y non-empty, we denote by xy^ω the infinite word $xyyy\cdots$.

A *border* of a word w is a non-empty word which is both a prefix and a suffix of w. A word is called *unbordered* if its only border is itself.

The set A^* of all words on the alphabet A is linearly ordered by the alphabetic order induced by the order $<$ on A. By definition, one has $x < y$ either if x is a proper prefix of y or if $x = uav, y = ubw$ with $u, v, w \in A^*$, $a, b \in A$ and $a < b$. A basic property of the alphabetic order is that if $x < y$ and if x is not a prefix of y, then for all words u, v, $xu < yv$.

Two words x, y are *conjugate* if there exist words u, v such that $x = uv$ and $y = vu$. Conjugacy is an equivalence class in A^*. A word is *primitive* if it is not a proper power, i.e. if it is not of the form r^n for $r \in A^*$ and $n > 1$. The number $p(k, q)$ of conjugacy classes of primitive words of length k on q symbols is given by the well-known Witt's formula.

$$p(k, q) = \frac{1}{k}\sum_{d|k}\mu(k/d)q^d$$

where μ is the Möbius function. The number $c(k, q)$ is given by

$$c(k, q) = \frac{1}{k}\sum_{d|k}\varphi(k/d)q^d$$

where φ is the Euler function. The values of $p(k, 2)$ and $c(k, 2)$ for $k \le 13$ are tabulated below.

k	1	2	3	4	5	6	7	8	9	10	11	12	13
$p(k,2)$	2	1	2	3	6	9	18	30	56	99	186	335	630
$c(k,2)$	2	3	4	6	8	14	20	36	60	108	188	352	632

A word is said to be *minimal* if it the least one in its conjugacy class. We denote by M the set of minimal words and by P the set of prefixes of minimal words. A *Lyndon word* is a word which is both primitive and minimal. We denote by L the set of Lyndon words.

The following propositions give elementary and well-known properties of Lyndon words used in the sequel. We include a proof for the sake of completeness.

The first one gives equivalent definitions of Lyndon words. It shows in particular that Lyndon words are unbordered.

Proposition 1 *The following conditions are equivalent for any non-empty word w.*

(i) *w is a Lyndon word.*

(ii) *for any non-empty words u, v such that $w = uv$, we have $w < vu$.*

(iii) *w is strictly less than any of its proper non-empty suffixes.*

Proof. If u, v are non-empty and if uv is primitive, then $uv \neq vu$ as it is well-known. Thus (i)$\Rightarrow$ (ii).
(ii)$\Rightarrow$ (iii). Let $w = vq$ with v, q both non-empty. It is not possible that q be a prefix of w. Indeed, otherwise, we have $w = qu$ for some non-empty word u. Since $vq = qu < uq$, we obtain $v < u$ and thus $qv < qu = vq$, a contradiction. Since q is not a prefix of w, it follows from $w < qv$ that $w < q$.
(iii)$\Rightarrow$ (i). It is clear that w is primitive. Next, if $w = uv$, with u, v non-empty, if follows from $w < v$ that $w < vu$. Thus w is a Lyndon word. □

The following proposition is clear since every non-empty word is in a unique way a power of a primitive word.

Proposition 2 *A word is minimal if and only if it is a power of a Lyndon word. This Lyndon word is uniquely determined.*

The following proposition is also straightforward to prove. Its converse is actually also true.

Proposition 3 *Let w be a prefix of a minimal word. Then any suffix of w either is a prefix of w or is greater than w. Equivalently, any prefix of w is less than or equal to the suffix of the same length of w.*

A word $w \in A^*$ is said to be a *sesquipower* of a word x if it is of the form $w = x^n p$ where $n \geq 1$ and p is a proper prefix of x. We will be especially interested in sesquipowers of Lyndon words. For example, if $a < b$, the word $w = aabbaabbaa$ is a sesquipower of the Lyndon word $l = aabb$.

The following property is linked to a method of generation of Lyndon words due independently to Fredericksen and Maiorana [3] and to Duval [2]. It is indeed relatively easy to generate the elements of P in alphabetic order and this gives a method to generate also either the elements of M or those of L. This generation problem has been considered in several contexts (see [12], [10] or [6] in particular).

Proposition 4 *The following conditions are equivalent for any word $w \in A^*$.*

(i) w is a non-empty prefix of a minimal word.

(ii) w is a sesquipower of a Lyndon word.

Proof. It is obvious that (ii) implies (i). Let us show the converse. By hypothesis there is a Lyndon word s such that w is a prefix of s^n. If $|w| \geq |s|$, then w is clearly a sesquipower of s. It is therefore enough to consider the case where w is a non-empty prefix of s.

We use an induction on the length of w. If $|w| = 1$, then w is a Lyndon word and the property is true. Let us then suppose that $|w| > 1$ and let $w = va$ with

$a \in A$. By the induction hypothesis, we have a factorization $v = l^n p$ where l is a Lyndon word, $n \geq 1$ and p is a proper prefix of l.

Let $l = pbx$ with $b \in A$ and $x \in A^*$. We are going to show that either $a = b$ and thus $w = va$ is a sequipower of l or w itself is a Lyndon word.

Indeed, since $w \in P$, we have by Proposition 3, $pb \leq pa$ and thus $b \leq a$. We therefore have to prove that if $b < a$, then w is a Lyndon word. For this, let $s = ta$ be a proper suffix of $w = l^n pa$. The word $u = tb$ is a proper suffix of $z = l^n pb$. Since $z \in P$, the word u is larger than or equal to the prefix of the same length of z and thus also of w. But since $s > u$, the word s is larger than w and thus w is a Lyndon word. □

It follows from the proof above that P is actually equal to the set of prefixes of Lyndon words.

A *division* of a word $w \in P$ is a pair (l^n, u) such that $w = l^n u$ where $l \in L$, $n \geq 1$ and $u \in A^*$ with $|u| < |l|$.

By Proposition 4 each word in P admits at least one division. We say that a Lyndon word $l \in L$ *meets* the word w if there is a division of w of the form (l^n, u). It is clear that for any $l \in L$ there is at most one such division of w.

The *main division* of $w \in P$ is the division (l^n, u) where l is the shortest Lyndon word which meets w. The word l^n is the *principal part* of w, denoted by $p(w)$, and u is the *rest*, denoted by $r(w)$.

For example, with $a < b$, the word $aabaabbba$ admits two divisions which are $(aabaabbb, a)$ and $(aabaabb, ba)$. The first one corresponds to its decomposition as a sesquipower of a Lyndon word. The second one is its main division.

3 Unavoidable sets

Let A be a finite alphabet. An *unavoidable* set on A is a set $I \subset A^*$ of words on the alphabet A such that any two-sided infinite word $(a_n)_{n \in Z}$ on the alphabet A admits at least one factor in I. It is of course equivalent to ask that any one-sided infinite word has a factor in I or also, since the alphabet is finite, that the set of words that avoids I is finite (see [9] for an exposition of the properties of unavoidable sets).

Example Let $A = \{a, b\}$. The set $U = \{a, b^{10}\}$ is unavoidable since any word of length 10 either has a letter equal to a or is the word b^{10}. On the contrary, the set $V = \{aa, b^{10}\}$ is avoidable. Indeed, the infinite word $(ab)^\omega = ababababab\ldots$ has no factor in V.

In the sequel, we will be interested in unavoidable sets made of words having all the same length k. The following proposition is easy to prove.

Proposition 5 *Let A be a finite alphabet and let I be an unavoidable set of words of length k on A. The cardinality of I is at least equal to the number of conjugacy classes of words of length k on the alphabet A.*

Proof. Let $u \in A^*$ be a word of length k. The factors of length k of the word u^ω are the elements of the conjugacy class of u. Thus I must contain at least one element of this class. □

Let $k \geq 1$ be an integer and let $\mathcal{M}_k$ be the set of minimal words of length k. For each $m \in \mathcal{M}_k$, let $p(m)$ be its principal part and $r(m)$ its rest. Let I_k be the set

$$I_k = \{r(m)p(m) | m \in \mathcal{M}_k\}$$

We remark that any minimal word which is not primitive appears in I_k.

Example. Table 1 gives the sets $\mathcal{M}_7$ and I_7.

The object of what follows is to show that I_k is an unavoidable set. By Proposition 5, the number of elements of I_k is the minimal possible number of elements of an unavoidable set of words of length k.

aaaaaaa	aaaaaaa
aaaaaab	aaaaaab
aaaaabb	baaaaab
aaaabab	abaaaab
aaaabbb	bbaaaab
aaabaab	aabaaab
aaababb	abbaaab
aaabbab	babaaab
aaabbbb	bbbaaab
aabaabb	baabaab
aababab	abaabab
aababbb	bbaabab
aabbabb	abbaabb
aabbbab	babaabb
aabbbbb	bbbaabb
abababb	bababab
ababbbb	bbababb
abbabbb	babbabb
abbbbbb	bbbabbb
bbbbbbb	bbbbbbb

Table 1. The sets $\mathcal{M}_7$ and I_7

4 Main result

We are going to prove the following result which shows that the lower bound $c(k, q)$ on the size of unavoidable sets of words of length k on q symbols is reached for all $k, q \geq 1$. This has, as said in the Introduction, already been obtained by J. Mykkelveit [11].

Theorem 1 *For all $k, q \geq 1$, there exists an unavoidable set formed of $c(k, q)$ words of length k on q symbols.*

The theorem will be a consequence of the following one, giving a construction of the minimal unavoidable sets.

Theorem 2 *Let A be a finite alphabet and let $k \geq 1$. Let $\mathcal{M}_k$ be the set of words on the alphabet A of length k and which are minimal in their conjugacy class. For every word $m \in \mathcal{M}_k$, let $p(m)$ be the principal part of m and let $r(m)$ be its rest. Then the set*

$$I_k = \{r(m)p(m) | m \in \mathcal{M}_k\}$$

is an unavoidable set.

To prove Theorem 2, we need some preliminary results. The first one is a simple equivalent definition of finite unavoidable sets.

Proposition 6 *Let $I \subset A^*$ be a finite set of words. The following conditions are equivalent.*

(i) *The set I is unavoidable.*

(ii) *Each two-sided infinite periodic word has at least one factor in I.*

Proof. It is enough to show that (ii) $\Rightarrow$ (i). Let $(a_n)_{n \in Z}$ be a two-sided infinite sequence of letters. Let $u \in A^*$ be a word longer than any word in I and having an infinite number of occurrences in the sequence $(a_n)_{n \in \mathbb{Z}}$. This sequence has at least one factor of the form uvu. By the hypothesis, the infinite periodic word $\ldots uvuvuvuv \ldots$ has a factor $w \in I$. The word w is a factor of at least one of the words uv and vu. It is also a factor of the sequence $(a_n)_{n \in \mathbb{Z}}$ and thus I is unavoidable.□

Proposition 7 *Let λ and l be two Lyndon words, with λ a prefix of l. Let $s \in A^*$ be a proper suffix of l, with $|s| < |\lambda|$. Then for all $n > 0$, the word $w = \lambda^n s$ is a Lyndon word.*

Proof. Let t be a proper suffix of w. Three cases may arise.
1. One has $|t| \leq |s|$. Then t is a proper suffix of the Lyndon word l and thus $t > l \geq \lambda$ and since $|t| < |\lambda|$, we have $t > \lambda^n s = w$.
2. One has $|t| > |s|$ and the word t factorizes as $t = \lambda^i s$, with $0 \leq i < n$. Since s is a proper suffix of l, we have $s > l \geq \lambda$. Consequently $t = \lambda^i s > \lambda^{i+1}$ and since $|s| < |\lambda|$, we have $t > \lambda^n s = w$.
3. One has $|t| > |s|$ and the word t factorizes as $t = s't'$, where s' is a proper suffix of λ. Since $\lambda \in \mathcal{L}$, one has $s' > \lambda$, and consequently $t = s't' > \lambda^n s = w$.

In all cases $t > w$ and thus w is a Lyndon word.□

Proposition 8 *Let w be a prefix of a minimal word and let (λ^n, u) be its main division. Let $u' \in A^*$ be a word of the same length as u and such that the word $w' = \lambda^n u'$ is also a prefix of a minimal word. Then the main division of w' is the pair (λ^n, u').*

Proof. Let (μ^i, v) be the main division of w'. We have $w' = \mu^i v$ with $|v| < |\mu|$. Since (λ^n, u') is a division of w', the word μ is a prefix of λ. We are going to show by contradiction that μ cannot be a proper prefix of λ.

Suppose that μ is a proper prefix of λ. Since the factorization of a minimal word as a power of a Lyndon word is unique, we cannot have the equality $\mu^i = \lambda^n$. Suppose first that $|\mu^i| < |\lambda^n|$. Since $w' = \mu^i v = \lambda^n u'$, the word μ^i is a proper prefix of the word λ^n. Thus there exists a non-empty word $x \in A^*$ such that $\mu^i x = \lambda^n$ et $xu' = v$. We thus have

$$w = \lambda^n u = \mu^i x u$$

Since $|xu| = |xu'| = |v| < |\mu|$, the pair (μ^i, xu) is a division of w, which is a contradiction since μ is a proper prefix of λ and that (λ^n, u) is the main division of w.

Let us now suppose that $|\mu^i| > |\lambda^n|$. Since $w' = \mu^i v = \lambda^n u'$, the word λ^n is a proper prefix of the word μ^i. Since μ is a proper prefix of λ, there exists an integer $j > 0$ and a prefix μ' of μ such that $\lambda = \mu^j \mu'$. Since λ is a primitive word, μ' is non-empty. As a consequence, λ admits μ' both as a non-empty prefix and and suffix, which is contradictory since λ is a Lyndon word.□

The final property needed to prove Theorem 2 is the following.

Proposition 9 *Let l be a Lyndon word. Let $n \in \mathbb{N}$ be the smallest integer such that $|l^n| > k$. Then the word l^{n+1} has a factor in I_k.*

Proof. Let w be the prefix of length k of l^n. Let (λ^i, u) be the main division of w. If u is the empty word, then, by construction, $w \in I_k$ and the proposition is true. Suppose that u is not empty.

The word λ is a prefix of l since either $|w| < |l|$ or w admits a division of the form (l^j, l'). Let s be the suffix of l having the same length as u. By Proposition 7, the word $\lambda^i s$ is a Lyndon word. Thus, by Proposition 8, the main division of $\lambda^i s$ is the pair (λ^i, s). Consequently, the word $s\lambda^i$ belongs to I_k. But this word is a factor of l^{n+1}. Thus l^{n+1} has a factor in I_k.□

We are now able to prove Theorem 2. By Proposition 6, it is enough to show that every periodic two-sided infinite word of the form $\ldots uuuuu \ldots$ has at least one factor in I_k. We may suppose without loss of generality that u is a Lyndon word. Let n be the least integer such that $n|u| > k$. Then, by Proposition 9, the word u^{n+1} has a factor in I_k. Thus I_k is unavoidable.

5 Conclusion

The proof of J. Mykkeltveit in [11] is based on the following principle, presented in the case of a binary alphabet. Let us associate to a word $w = a_0 a_1 \cdots a_{n-1}$ on the alphabet $\{0, 1\}$ the sum $s(w) = \sum a_j \omega^j$ where $\omega = e^{2i\pi/n}$. We denote by $Is(w)$ the imaginary part of $s(w)$. It can be shown that for each conjugacy class of words, only two cases occur:

(i) either all words w are such that $Is(w) = 0$ (and then, for $n > 2$ one has actually $s(w) = 0$ for each of them)

(ii) or there is, in clockwise order, one block of words w such that $Is(w) > 0$ followed by one block of words w such that $Is(w) < 0$ separated by at most two words w such that $Is(w) = 0$.

Consider the set S_n of words of length n formed of

(i) a representative of each conjugacy class of words w of length n such that $Is(w) = 0$ for all the conjugates.

(ii) the words $w = a_0a_1 \cdots a_{n-1}$ of length n such that $Is(w) > 0$ for the first time clockwise.

It is shown in [11] that this set is unavoidable for all $n > 2$. The comparison with our definition of a minimal unavoidable set shows that the definitions have nothing in common. Moreover, the sets obtained are indeed different. The sets defined by J. Mykkeltveit have a slight advantage on ours in the sense that the maximal length of words avoiding the set is less. For example, for $n = 20$, there are 256 words of length 2579 that avoid I_n, but none of length 563 that avoid all of S_n (and there is a unique way to avoid S_n with length 562). This computation has been performed using D. Knuth's program UNAVOIDABLE2 (see `http://www-cs-faculty.stanford.edu/~knuth/programs.html`). Our proof has the advantage of using only elementary concepts and in particular no real or complex arithmetic.

The first proof of Theorem 1 obtained by the first two authors (see the Introduction) is a construction working by stages. To explain these stages, let us consider to simplify the notation the case of a binary alphabet $A = \{a, b\}$. Given a set X of two-sided infinite words, we say that a set Y of words is unavoidable in X if every word of X has a factor in Y.

For $n \geq 1$, let X_n be the set of two-sided infinite words on A which avoid a^n. Let $c_n(k, q)$ be the number of conjugacy classes of words x of length k on q symbols such that the words of the form $x^\zeta = \cdots xxx \cdots$ are in X_n. It is thus also equal to the number of orbits of period k in X_n. The table below gives the values of $c_n(k, 2)$ for $1 \leq n \leq 10$ and $1 \leq k \leq 10$.

	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	2	2	3	3	5	5	8	10	15
3	1	2	3	4	5	9	11	19	29	48
4	1	2	3	5	6	11	15	27	43	59
5	1	2	3	5	7	12	17	31	51	91
6	1	2	3	5	7	13	18	33	55	99
7	1	2	3	5	7	13	19	34	57	103
8	1	2	3	5	7	13	19	35	58	105
9	1	2	3	5	7	13	19	35	59	106
10	1	2	3	5	7	13	19	35	59	107

The idea of the step by step construction of a minimal unavoidable set of words of length k is to construct a sequence $Y_1 \subset Y_2 \subset \ldots \subset Y_k$ of sets of words of length k such that for $1 \leq n \leq k$, the set Y_n is unavoidable in X_n with $c_n(k,q)$ elements. This can be stated as the following result.

Theorem 3 *For each $q \geq 1$, and $k \geq n \geq 1$, there exists a set of $c_n(k,q)$ words of length k on q symbols which is unavoidable in X_n.*

The proof consists in showing that the $c_n(k,q)$ last elements of I_n form a set unavoidable in X_n.

It is interesting to remark that not all minimal unavoidable sets are build in this way. Indeed, there are sets which are minimal unavoidable in X_{n+1} which do not contain a minimal unavoidable set in X_n.

For example, let $Y_1 = \{bbb\}$, $Y_2 = \{bbb, bab\}$, $Y_3 = \{bbb, bab, aab\}$. Then each Y_n for $1 \leq n \leq 3$ is unavoidable in X_n of size $c_n(3,2)$ and $I_3 = Y_3 \cup \{aaa\}$. In particular, the set I_3 contains an unavoidable set in X_2 with 2 elements, namely Y_2. However, the set $J_3 = \{aaa, aba, bba, bbb\}$ obtained from I_3 by exchanging a and b does not contain a two element set unavoidable in X_2.

A set of the form X_n is a particular case of what is called a system of finite type. This is, by definition the set of all two-sided infinite words avoiding a given finite set of words (see [7]). We do not know in general in which systems of finite type it is true that for each k there exists an unavoidable set having no more elements than the number of orbits of period k.

References

[1] Jean-Marc Champarnaud and Georges Hansel. Ensembles inévitables et classes de conjugaison. *Bulletin of the Belgian Mathematical Society*, 2003. to appear.

[2] Jean-Pierre Duval. Factorizing words over an ordered alphabet. *J. Algorithms*, 4(4):363–381, 1983.

[3] Harold Fredricksen and James Maiorana. Necklaces of beads in k colors and k-ary de Bruijn sequences. *Discrete Math.*, 23(3):207–210, 1978.

[4] Michael R. Garey and David S. Johnson. *Computers and intractability.* W. H. Freeman and Co., San Francisco, Calif., 1979. A guide to the theory of NP-completeness, A Series of Books in the Mathematical Sciences.

[5] Guo-Niu Han and Dominique Perrin. Ensembles inévitables. *Sém. Lothar. Combin.*, 47:Article B47e, 16 pp. (electronic), 2001.

[6] Donald E. Knuth. *The Art of Computer Programming*, volume 4. Addison Wesley, 2002. to appear, see `http://www-cs-faculty.stanford.edu/~knuth/news01.html`.

[7] Douglas Lind and Brian H. Marcus. *An Introduction to Symbolic Dynamics and Coding.* Cambridge, 1995.

[8] M. Lothaire. *Combinatorics on words.* Cambridge University Press, Cambridge, 1997.

[9] M. Lothaire. *Algebraic Combinatorics on Words.* Cambridge University Press, 2002.

[10] Eduardo Moreno. A note on the theorem of Fredericksen and Maiorana. *Advances in Applied Mathematics*, 2003. to appear.

[11] Johannes Mykkeltveit. A proof of Golomb's conjecture for the de Bruijn graph. *J. Combinatorial Theory Ser. B*, 13:40–45, 1972.

[12] Christophe Reutenauer. *Free Lie algebras.* The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.

[13] Laurent Rosaz. Inventories of unavoidable languages and the word-extension conjecture. *Theoret. Comput. Sci.*, 201(1-2):151–170, 1998.

[14] C. J. Saker and P. M. Higgins. Unavoidable sets of words of uniform length. *Inform. and Comput.*, 173(2):222–226, 2002.

[15] Marcel-Paul Schützenberger. On the synchronizing properties of certain prefix codes. *Inform. and Control*, 7:23–36, 1964.

[16] Richard P. Stanley. *Enumerative combinatorics. Vol. 1.* Cambridge University Press, Cambridge, 1997.

Jean-Marc Champarnaud and Georges Hansel
LIFAR, Université de Rouen
B.P. 67, 76 130 Mont Saint Aignan

Dominique Perrin
Institut Gaspard Monge
Université de Marne-la-Vallée
77454 Marne-la-Vallée cedex 2