INFORMATIQUE THÉORIQUE ET APPLICATIONS

M. RÉGNIER

Enumeration of bordered words, le langage de la vache-qui-rit

Informatique théorique et applications, tome 26, nº 4 (1992), p. 303-317

<http://www.numdam.org/item?id=ITA_1992__26_4_303_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 26, n° 4, 1992, p. 303 à 317)

ENUMERATION OF BORDERED WORDS LE LANGAGE DE LA VACHE-QUI-RIT (1) (*)

by M. RÉGNIER (2)

Communicated by Jean BERSTEL

Abstract. – *We consider here the family of bordered words on a q-ary alphabet,* i.e. *the words bwb. We also consider the k-bordered words. We enumerate such words, using generating functions and combinatorics on words, and derive asymptotic estimates by the Darboux method. In particular, we prove that the density of k-bordered words is* α_k, $\alpha_k \neq 0$.

Résumé. – *Nous considérons la famille des mots avec bord* (i.e. *de la forme bwb*) *sur un alphabet à q lettres, et plus généralement des mots à k bords imbriqués. Nous énumérons ces mots à l'aide de fonctions génératrices et de combinatoire des mots, et obtenons des résultats asymptotiques par la méthode de Darboux. En particulier nous prouvons que la densité des mots à k bords est un nombre* α_k *non nul.*

1. INTRODUCTION

This note is devoted to bordered words on a q-ary alphabet A, which are to be counted. A 1-bordered word w is defined as a word: $bw'b$ where b and w' are in A^*, and b, the border, is non empty. For example, $w = 101.1.101$. One can define recursively the set B_{k+1} of the $k+1$-bordered words: w is in B_{k+1} if w is in B_1, and if its largest border is in B_k. For example, w is in B_2 as $b = 101$ is in B_1. To count the words in sets B_k, we make use of the associated generating functions $B_k(z)$. The scheme is the following: we first establish functional equations satisfied by the series $B_k(z)$. To do so, we need a unique representation of the words in B_k. Hence, we define a notion of k-minimality. This part makes use of general theorems in combinatorics on

(*) Received March 1990, final version October 1991.
(1) This subtitle refers to the famous Rabier's commercial drawing: a cow with two identical © "Vache-Qui-Rit" boxes as ear-rings. Inside each cheese box, a cow with two earrings...
This work was partially supported by the ESPRIT II Basic Research Actions Program of the EC under contract No. 3075 (project ALCOM).
(2) I.N.R.I.A., 78153 Le Chesnay, France.

words. To get asymptotics for the coefficients, we do not need to solve the functional equations. We study the singularities of $B_k(z)$, that appear to be polar singularities. Then one can apply the Darboux theorem, and prove that the number of k-bordered words of length n, b_n^k, satisfies $b_n^k \sim \alpha_k . q^n$, where α_k is computable, with any given precision, from the functional equation.

First, we list some general theorems in combinatorics to be used in the following sections. Then, we introduce our techniques on the set B_1. In the following Section, we consider the more general and intricate case of the set B_2. Finally, we extend these methods and results to the general case of sets B_k. In the last section, we deal with the asymptotics of b_n^k and numerical computations.

2. k-MINIMALITY

The aim of this section is to determine a unique representation of k-bordered words. We first state our definition.

DEFINITION 1: *Let A be a q-ary alphabet. A word $w \in A^+$ is a 1-bordered word if there exist two words $b \in A^+$ and $x \in A^*$ such that:*

$$w = bxb.$$

Then, b is a border of w.

A word $w \in A^+$ is a k-bordered word, $k \geqq 2$, if it is a 1-bordered word and if its largest border is a $(k-1)$-bordered word.

One notes B_k the set of k-bordered words and $S = A^* - B_1 = B_0$ the set of unbordered words.

Remarks:

(i) The empty word and the words of length 1 are in S.

(ii) This definition does not allow overlaps. Accordingly, 0101 is a side of 01010101 but 010101 is not.

Example: $w = 1011101$ is a 1-bordered word on a binary alphabet. The words $b_1 = 1$ and $b_2 = 101$ are both borders of w.

We see on this example that a word in B_1 (more generally in B_k) may have several borders. In order to *enumerate* G_k, we need a *unique representation* of words in B_k. This problem of deciding a unique representation is fairly general in combinatorics on words. Some interesting examples can be found in [CP, Dar78, Odl85] and [Lot83, ch. 5]. Hence, we define below a *k-minimality* notion in B_k, *i.e.* a subset G_k of B_{k-1} such that any k-bordered word may be written, in a unique manner:

$$b_k = g_k w g_k,$$

which yields the functional equation on the corresponding generating functions:

$$B_k(z) = G_k(z^2) W(z). \tag{1}$$

Thus, the problem reduces to the enumeration of k-minimal words.

DEFINITION 2: *A bordered word x is said to be k-minimal if:*

(i) $x \in B_k$

(ii) $w \underline{\underline{\prec}} x,\ w \subseteqq x,\ w \in B_k \Leftrightarrow w = x$

This subset of B_k is noted G_{k+1}.

Example: $x = 1001001 \in B_1$. As $1001 \subseteqq x$ and $1001 \prec x$, $x \notin G_2$, but $1001 \in G_2$.

Remark: A word in G_k has no side in B_{k-1}.

THEOREM 2.1: *Any k-bordered word b_k in B_k can be written, in a single way, as:*

$$b_k = g_k w g_k$$

with $g_k \in G_k$.

Proof: From Definition 2, any word x in B_{k-1} contains exactly one side in G_k. Hence, if $k \geqq 2$, let b_{k-1} be the largest border of b_k and choose g_k as its only side in G_k. Now, remark that the smallest border of a word is always in $S = B_0$ and choose g_0 as its only side in its subset G_1.

Our aim is now to *enumerate* the set G_k. To do so, we need some characterizations of the words g_k. For a sake of clarity, we first consider the case of 1-bordered words and 2-bordered words.

3.1 BORDERED WORDS

We show here that $G_1 = S - \{\varepsilon\} = B_0 - \{\varepsilon\}$, and then derive the generating functions $B_1(z)$ and $S(z)$. We need some basic results from combinatorics on words.

DEFINITION 3: *Two words u and v are said to be* conjugate *if there exist words a, $e \in A^*$ such that: $u = ae$, $v = ea$.*

From the Defect Theorem [Lot83, ch. 1], one easily deduces the conjugacy theorem in [Lot83, ch. 1]:

THEOREM 3.1: *Two words u and $v \in A^*$ are conjugate iff there exists some $z \in A^*$ such that*:

$$zu = vz.$$

This equality holds iff there exist a, $e \in A^*$ such that:

$$\begin{cases} v = ea, \qquad u = ae \\ \quad z \in e(ae)^* \end{cases}$$

The proof is given in [Lot83]. Now, we prove:

LEMMA 3.1: *The set of* 0-*minimal words is:*

$$G_1 \cup \{\varepsilon\} = S.$$

Moreover, a word in S cannot overlap with himself.

Proof: Let s be in S, and not 0-minimal. Let w be a word in S satisfying $w \underline{\underline{\prec}} s$ and $w \subseteqq s$. Then, $s = wu = vw$. From conjugacy theorem, this would imply: $s = ea.e.(ae)^*$, which is not in S. Hence, any word in S is 0-minimal. Similarly, a word in S cannot overlap with himself, as it would also imply: $s = wu = vw$.

As a corollary of Lemma 3.1, we get the functional equation:

$$B_1(z) = [S(z^2) - 1] . W(z). \tag{2}$$

We use the methods developed in [GJ83] and [Fla84]. The concatenation $s.w'$ of two different words translates into the product of the generating functions counting these words. The repetition of the word s is taken into account by squaring z in the corresponding generating function S. Moreover, we have exactly q^n words of length n. Thus:

$$W(z) = \sum_{n \geqq 0} q^n z^n = \frac{1}{1 - qz}.$$

As: $B_1(z) + S(z) = W(z)$, Proposition 3.1 follows.

THEOREM 3.2: *The series B_1 and S satisfy the functional equations:*

$$S(z)=\frac{2-S(z^2)}{1-qz}$$

$$B_1(z)=\frac{1}{1-qz}\left[\frac{qz^2}{1-qz^2}-B_1(z^2)\right]$$

which yields:

$$S(z)-(1+qz)=q(q-1)\sum_{j=0}^{\infty}(-1)^j z^{2^{j+1}}\prod_{i=0}^{j}W(z^{2^i}).$$

Remark: Such an approach where combinatorial constructions translate into functional properties of generating functions is quite powerful. A general framework can be found in [Fla84]. An example, related to the analysis of the Knuth-Morris-Pratt algorithm, can be found in [Rég89].

4.2 BORDERED WORDS

In this section, we generalize the scheme of the previous section. Our characterization of G_2 will rely on the Lemmas of 1-factorization and 2-factorization.

LEMMA 4.1 (1-factorization lemma): *Let $x \in z^*$, $z \in P$, be a word in A^*. Then, if a word $s \in S$ is a right (resp. left) factor of x, then s is a right (resp. left) factor of z.*

Proof: If $s \nsubseteq z$, then there exists $\alpha \subseteqq z$ (or $z=\beta\alpha$), $\alpha \neq \varepsilon$ such that: $s=\alpha . z^m$, $m \geqq 1$. Thus: $s=\alpha.(\beta\alpha)^{m-1}\beta.\alpha$ is not in S.

LEMMA 4.2 (2-factorization lemma): *Let $g_2=sas$ be a word in G_2, and $x=wsas \in z^*$, with $z \in P$. Then:*

$$\begin{cases} z=as \\ \text{or} \\ z=usas \end{cases}$$

Example: Let us show what happens for a word in B_1, but not in G_2, such as: 1.0100101. We have: $s=1$ and $as=0100101 \in P$. Let $w=00$. Then $x=wsas=00101.00101$ and $z=00101$; here $z \neq as$ and $sas \nsubseteq z$.

PROPOSITION 4.1: *Let* H_1 *be the subset of* B_1*:*

$$H_1 = \{sas;\ s \in S,\ as \in P\}.$$

Then:

(i) $G_2 \subset H_1 \subset B_1$

(ii) *Any* $b_1 \in B_1$ *can be written:*

$$\begin{cases} b_1 = s(as)^m \\ b_1 = sas.(usas)^m, usas \in P \end{cases}$$

where s and a are defined by: $g(b_1) = sas$, *and* $m \geqq 1$.

Proof: (i) From Proposition 3.1, one can write $g_2 = sas$. If $as \notin P$, the 1-factorization Lemma implies: $g_2 = s(as)^m$, and g_2 is not minimal.

(ii) The proof of (ii) is deferred to the Appendix. It uses the 0-minimality and 2-factorization lemma, whose proof is also in Appendix. Note that the second form implies that b_1 is in B_2, as $sas.usas$ is.

We can turn now to the study of the generating function of 2-bordered words. We prove:

PROPOSITION 4.2: *Let* $G_2(x, z) = \sum_{g_2 = sas \in G_2} x^{|s|} z^{|as|}$ *be the bivariate generating function of the set* G_2. *It satisfies the functional equation:*

$$[S(zx) - 1 - G_2(zx, z^2)].W(z) = G_2(x, z) + G_2(x, z^2). \tag{3}$$

Proof: We define a 3-variate generating function associated to $H_1 - G_2$:

$$\varphi(x, z, t) = \sum x^{|s|} z^{|as|} t^{|usas|} 1_{sas \in G_2} 1_{usas \in P}$$

We shall write two equations in φ and G_2 and eliminate φ from them. We have:

$$\sum_{\substack{b_1 \in B_1 \\ b_1 = sws}} x^{|s|} z^{|ws|} = [S(zx) - 1] W(z).$$

From Proposition 4.1, this is also:

$$\sum_{g_2 \in G_2} \Big(\sum_{m \geqq 1} x^{|s|} z^{m|as|} + \sum_{usas \in P} \sum_{m \geqq 1} x^{|s|} z^{|as|} z^{m|usas|}\Big)$$

$$= \sum_{m \geqq 1} G_2(x, z^m) + \sum_{m \geqq 1} \varphi(x, z, z^m).$$

Hence:

$$[S(zx)-1]\,W(z)=\sum_{m\geqq 1} G_2(x,z^m)+\sum_{m\geqq 1}\varphi(x,z,z^m) \tag{4}$$

To get the equation (5), we consider the words $g_2\,wg_2=sas\,.\,w\,.\,sas\in B_2$. From the 2-factorization lemma, we get:

$$\sum x^{|s|}z^{|as|}t^{wsas}\,1_{sas\in G_2}\,1_{w\in A^*}=\sum_{m\geqq 2} G_2(x,zt^m)+\sum_{m\geqq 1}\varphi(x,z,t^m).$$

Hence:

$$G_2(xt,zt)\,W(t)=\sum_{m\geqq 2} G_2(x,zt^m)+\sum_{m\geqq 1}\varphi(x,z,t^m). \tag{5}$$

Eliminating φ from (4) and (5) yields G_2.

Finally, unwinding (3), we get a closed formula for G_2.

THEOREM 4.1: *The bivariate generating function of the set G_2 is:*

$$G_2(z,x)=\sum_{m=0}^{\infty}(-1)^m\,W(z^{2^m})\sum_{n=0}^{2^m-1}[S(xz^{2^m+n})-1]\,\theta(n)$$

where $\theta(n)$ is defined, for a binary decomposition $n=\sum_i b_i 2^i$ as:

$$\theta(n)=\prod_{b_i\neq 0} W(z^{2^i}).$$

Moreover:

$$B_2(z)=G_2(z^2,z^2)\,W(z).$$

5. PROPERTIES OF k-BORDERED WORDS

In this section, we extend the scheme of Section 4 to the general case of k-bordered words. We first study the set G_k, and prove a k-factorization Lemma. Then, we associate to G_k a k-variate generating function $G_k(t_k,\ldots,t_1)$. We prove that $G_k(t_k,\ldots,t_1)$ and $G_k(z)$ satisfy equations similar to (2) and (3).

We first derive some properties of G_k.

THEOREM 5.1: *Let g_k be in G_k. It can be factorized as:*

$$g_k = g_{k-1} . p_{k-1}^{\theta_{k-1}} = \ldots = g_i . p_i^{\theta_i} = \ldots = g_1 . p_1^{\theta_1}$$

with $g_i \in G_i$, $p_i \in P$. Moreover, this decomposition satisfies either one of the two cases:

(i) $\forall i : \theta_i = 1$ *and* $g_i \subseteq p_i$.

(ii) $\exists ! m$ *such that* $g_m \subseteq p_m$ *and* $g_m . p_m \in G_j$, $m+1 \leqq j < k-1$.

Then:

$$\begin{cases} p_m = p_j = p_{j+1} = \ldots = p_{k-1} \\ \theta_m = 2^{k+1-j} - 1 \\ \theta_{j+l} = 2^{k+1-j} - 2^{l+1}, \qquad 0 \leqq l \leqq k-1-j \end{cases}$$

Examples:

$k=1 : g_1 = s \in S$.

$k=2 : g_2 = s . as$, $s \in S$, $as \in P$.

$k=3 : g_3 = s(as)^3$ or $sas . usas$. We can rewrite:

$$\begin{cases} sas . (as)^2 = s . (as)^3 \\ sas . usas = s . (asusas). \end{cases}$$

$k=4 : g_4 = s(as)^7$ or $s(as)^3 . ws(as)^3$ or $sasusas.wsasusas$ or $sas . (usas)^3$ or $s . (asusas)^3$. We can rewrite, for example:

$$s . (asusas)^3 = sas . (usas . asusas . asusas) = sasusas . asasusasusas$$

We have: $g_1 . p_1 \in B_2 - B_3$ and $g_2 . p_2$, $g_3 . p_3 \in B_3$.

Proof: As $g_k \in B_{k-1}$, its largest border b satisfies:

$$b \in B_{k-2} \subset B_{k-3} \subset \ldots \subset B_1.$$

Thus, there exists a sequence: $g_{k-1}, g_{k-2}, \ldots, g_1$ such that:

$$\begin{cases} g_1 \preceq g_2 \preceq \ldots \preceq g_{k-2} \preceq g_{k-1} \preceq g_k \\ g_1 \subseteq g_2 \subseteq \ldots \subseteq g_{k-2} \subseteq g_{k-1} \subseteq g_k \end{cases}$$

and we get the factorization. We then rely upon the overlapping lemma:

LEMMA 5.1 (k-overlapping theorem): *Let g_k be a word in G_k and $x \in z^*$, $z \in P$ a word in L^* such that:*

$$g_k x = y g_k, y \in L^*.$$

Then: $z = w g_k$ or $z \in \{p_1, \ldots, p_{k-1}\}$.

Our proof will make use of the following theorem, also a consequence of [Lot83, ch. 1].

THEOREM 5.2: *Two words $x, y \in A^+$ commute iff they are power of the same word. More precisely, the set of words commuting with a word $x \in A^+$ is a monoid generated by a single primitive word.*

Proof of Lemma 5.1: We know from the non-overlapping property and 1-factorization lemma that the property holds true for g_1 in G_1. Assume now that it holds true for $G_1, \ldots, G_{k-1}$. If $z \neq w g_k$, then $z \subset g_k$ and $g_k = e_z . z^p$ with $e_z \subseteqq z$ and $p \geqq 1$. Noticing that:

$$g_k . z = e_z . z^{p+1} = y' . e_z . z^p = y . g_k,$$

we restrict the problem to $x = z$. As non-overlapping property implies $g_1 \subseteqq x$, we may define i, $1 \leqq i \leqq k-1$ by:

$$g_i \subseteqq x \subset g_{i+1}.$$

If $i < k-1$, we deduce from our equation:

$$\begin{cases} x \subseteqq g_k \Rightarrow g_k = e . x^p, & e \subset x \\ g_i a g_i = g_i . x = g_{i+1} \underline{\prec} g_k \Rightarrow g_k = g_i . x^m a, & a \prec x. \end{cases}$$

Moreover, we have

$$p \geqq 1 \quad (\text{as } x \subseteqq g_{k-1}) \qquad \text{and} \qquad m \geqq 1 \quad (\text{as } g_{i+1} \subseteqq g_{k-1} \Rightarrow 2|g_{i+1}| \leqq |g_k|).$$

Then:

$$a \subseteqq x, \qquad x \subseteqq xa \qquad \text{or} \qquad xa = \alpha x \qquad \text{with} \quad |\alpha| = |a| \quad \text{and} \quad \alpha \underline{\prec} x.$$

Hence:

$$xa = ax$$

or, from Theorem 5.2: $a = \varepsilon$ and $g_k = g_i . x^p = g_i . p_i^{\theta_i}$, hence: $x = p_i$.

Now, if $i=k-1$, we may prove $x=p_{k-1}$. If $p_{k-1} \subseteq x \subseteq g_k$, the two equations $yg_k=g_k x$ and $g_k=g_{k-1}p_{k-1}^*$ imply that:

$$x=\beta p_{k-1}^*, \qquad \beta \subseteq p_{k-1} \subseteq g_k.$$

Hence: $\beta x=x\beta$ and by Theorem 5.2 one has $x=\beta$ which implies $\beta=p_{k-1}=x$. If $x \subseteq p_{k-1} \subseteq g_k$, one has:

$$p_{k-1}=\alpha x^*, \qquad \alpha \subseteq x \subseteq g_k.$$

As $|x|+|g_{k-1}|<|g_k|$ and $|2g_{k-1}|=|g_k|$, we get: $p_{k-1}x=xp_{k-1}$ and again $x=p_{k-1}$.

Continuation of the proof: As Theorem 5.2 and k-overlapping Lemma imply that the largest border of $g_i(w_i g_i)^l$, $l \geqq 1$ is $g_i(w_i g_i)^{[(l-1)/2]}$, the expression of θ_i follows immediately, as well as the equation: $p_j=\ldots=p_{k-1}$ and the expressions of $\theta_j, \ldots, \theta_{k-1}$. Assume now that $\theta_{m'}>1$, with $m' \neq m$. The reasoning above applies, hence $p_{m'}=p_{k-1}=p_m$. From $\theta_{k-1}=2^{k-j}=2^{k-j'}$, we get $j=j'$. Finally, $g_j=g_m \cdot p_m=g_{m'} \cdot p_m$ implies $g_m=g_{m'}$ and $m=m'$.

We can now draw a scheme that generalizes the derivation of G_1 and G_2 in the previous sections. We define some multivariate generating functions. $\psi_{i,k}(t_1, \ldots, t_i, t_{i+1}, \ldots, t_k)$ counts

$$H_{i,k}=\{g_i \cdot p_i \mid g_i \in G_i,\ p_i \in P,\ g_i \subseteq p_i,\ g_i \cdot p_i \in B_k - B_{k+1}\}.$$

And $\varphi_{i,k+1}(t_1, \ldots, t_i, t_{i+1}, \ldots, t_k)$ counts

$$L_{i,k+1}=\{g_i \cdot p_i \mid g_i \in G_i,\ p_i \in P,\ g_i \subseteq p_i,\ g_i \cdot p_i \in B_{k+1}\}.$$

One has: $\psi_{i,j}=\varphi_{i,j}-\varphi_{i,j+1}$. Then:

$$G_k(t_1, \ldots, t_k)=\sum_{i=1}^{k} \psi_{i,k}(t_1, \ldots, t_k)+f(\{\varphi_{i,j}\}_{1 \leqq j \leqq k-1}).$$

For any k, counting $\{g_i w g_i\}$ and applying Theorem 5.1 yields k functional equations. These equations involve: $2k$ unknown, but dependent, functions $(\varphi_{i,k+1})_{1 \leqq i \leqq k+1}$ and $(\psi_{i,k})_{1 \leqq i \leqq k}$. One can derive $(\varphi_{i,k+1})_{1 \leqq i \leqq k+1}$ from these equations and hence: $\psi_{i,k}=\varphi_{i,k}-\varphi_{i,k+1}$. Finally, we get G_{k+1}. We can also relate these notations to the ones in the sections above. Notably: $\varphi_{1,1}(t_1, t_2)=G_2(t_1, t_2)$.

6. ASYMPTOTICS ON k-BORDERED WORDS

6.1 Asymptotic order

The equations derived in the previous sections are rather involved, and cannot be solved explicitly. To derive asymptotics on the coefficients, we study the singularities of the generating functions and use the Darboux Theorem [Dar78]. Examples of this approach are developed in [Ste84].

THEOREM 6.1 (Darboux Theorem): *Let $f(z)$ be some complex function, analytic for $|z|<\rho$, with a single singularity, $z=\rho$, on its circle of convergence. If it can be continued as:*

$$f(z)=g(z).\left(1-\frac{z}{\rho}\right)^{-s}+h(z)$$

where g and h are analytic and s ranges in $\mathscr{R}-\{0,-1,-2,\ldots\}$, then:

$$f_n=[z^n]f(z)=\rho^{-n}n^{s-1}\frac{g(\rho)}{\Gamma(s)}\left(1+O\left(\frac{1}{n}\right)\right).$$

In our case, $W(z)=\dfrac{1}{1-qz}$ has a unique singularity around $z=\dfrac{1}{q}$. Moreover, each $B_k(z)$ is the product of $W(z)$ by a, possibly intricate, generating function analytic around $z=1/q$ (precisely, for $|z|\leq 1/q^2$). This can be seen in Equation (2) and is precised notably in (3).

THEOREM 6.2: *Let $B_k(z)=\sum_n b_n^k z^n$ be the generating function of the k-bordered words. Then:*

$$b_n^k \sim \alpha_k q^{-n}$$

where: $\alpha_k=G_k(1/q^2,\ldots,1/q^2)$.

Moreover,

$$\alpha_k=\frac{1}{q^{2^k-1}}\left(1+o\left(\frac{1}{q}\right)\right).$$

Proof: We know that: $W(z)=S(z)+B_1(z)$. Thus, S and B_1 are both defined at least for $|z|<1/q$. Hence $S(z^2)$ is analytic for $|z|<1/\sqrt{q}$, notably around $z=1/q$. We may apply the Darboux theorem to equation (2). More generally, the definition of G_k implies:

$$B_k(z) = G_k(z^2, \ldots, z^2) \,.\, W(z).$$

A priori, $G_k(z, \ldots, z)$ is analytic around $z = 1/q^2$. It allows for the application of the Darboux theorem with $s = 1$ and $g(z) = G_k(z^2, \ldots, z^2)$.

To prove the second assertion, we just remark that the smallest word in B_{k-1} is a^k, $a \in A$. It is associated to words $a^k . w . a^k$ in B_k, and 1 word out of $q^{2^k - 1}$ has this type.

6.2 Numerical Computation

Note that $S(1/q^2)$ can be numerically computed from the sum in Theorem 3.2.

$$S(z) = 1 + \frac{1}{q} + q(q-1) \sum_{j \geq 0} \frac{(-1)^j q^{-2^{j+2}}}{(1-1/q) \ldots (1-q^{1-2^{j+1}})}.$$

The different values of α_1 for $q = 2$, 3, 10 are given in Table I.

TABLE I

q	α_1
2.	0.732,2
3.	0.443,0
10.	0.110,0

Note that the convergence to $1/q$ is fast. As a matter of fact, $|s(w)| = 1$ for one word out of q: whenever some character is a prefix and a suffix. Moreover, $|s(w)| = 2$ is associated to the configuration: $ab \underline{\preceq} w$, $ab \subseteqq w$, $a \neq b$.

The contribution is

$$q(q-1) . \left(\frac{1}{q^2}\right)^2 = \frac{q-1}{q^3} = o\left(\frac{1}{q}\right).$$

Now, Theorem 4.1 yields a closed formula for $G_2(x, z)$. Nevertheless, it converges quite slowly for small q. As a matter of fact, the first term is: $W(z)[S(zx) - 1]$ that counts words sws, $s \neq \varepsilon$. Now, for $|s| = 1$ and $w = s$, $sws \notin G_2$. This yields a (relative) error $1/q$ on the first term, which is also the more important. Hence, it appears more efficient to compute α_2 from a truncation of the entire series $G_2(x, z)$, i.e.

$G_2(x,y)=$

q*x*z+(q*(q−1)*x**2+q**2*x−q*x)*z**2+
(−q*x+q**2*(q−1)*x**3+q**2*(q−1)*x**2+q**3*x)*z**3+
(−q*(q−1)*x**2−2*q**2*x+q**2*(q−1)*(q**2−1)*x**4+
q**3*(q−1)*x**3+q**3*(q−1)*x**2+q**4*x)*z**4+
(−q**3*x−q**2*x+q**3*(q−1)*(q**2−1)*x**4+q**4*(q−1)*x**3+q**5*x+
q**4*(q−1)*x**2)*z**5+0(z**6);

$G_2\left(\frac{1}{q^2}, \frac{1}{q^2}\right)=$ 1/q**3+1/q**4+2/q**6−2/q**8−3/q**10+2/q**11+0(1/q**12);

which yields the table:

TABLE II

q	α_2
2.	0.209
3.	0.051,8
10.	0.001,1

For $k \geqq 3$, one can derive closed formulae similar to (3). Nevertheless, a truncated development of G_k provides again a good numerical approximation. It is derived by a word enumeration based on Theorem 5.1. Typically, words $g_k = s.(as)^{2^{k-1}}$ are counted by $G_2(x, z^{2^{k-1}-1})$. A bound is available for other words $g_i.p_i^{\theta_i}:s(as)$, $a \neq \varepsilon$ occur 2^{k-1} times, while some u satisfying $u \neq \varepsilon$, $usas \in P$, occurs 2^{k-2} times. Hence, the contribution is upper bounded by

$$\frac{1}{q^{3.(2^{k-1}-1)}} = o\left(\frac{1}{q^9}\right)$$

and this approximation seems good enough.

7. CONCLUSION

In this paper, we have considered the k-bordered words. In an algebraic part, we use general results in combinatorics on words to define a unique representation of k-bordered words. In particular, we introduce and characterize *minimal* k-bordered words. Then, we show how these constructions translate into functional equations satisfied by the associated generating functions. Finally, we show that these equations need not to be solved (the solutions are intricate) and get directly asymptotic estimates on the number of k-bordered words. We prove that there are $\alpha_k q^n$ k-bordered words, or equivalently, that the density of the family of k-bordered words is always non-zero. The constant α_k can be computed for any k from the functional equation, and is explicitly given, for various q, when $k=1$, 2. Such methods also apply to other combinatorial problems on words.

REFERENCES

[CP] M. Crochemore and D. Perrin, Pattern matching in strings, L.I.T.P., Paris, *Research Report*, 88-5.

[Dar78] G. Darboux, Mémoire sur l'approximation des fonctions de très grands nombres, et sur une classe étendue de développements en série, *J. Math. Pures Appl.*, février 1978, pp. 5-56, 377-416.

[Fla84] Flajolet, *Algorithmique*, in *Encyclopedia Universalis*, *1*, pp. 758-763. Éditions de l'E. U., 1984.

[GJ83] I. P. Goulden and D. M. Jackson, Combinatorial Enumeration, *John Wiley*, New York, 1983.

[Lot83] Lothaire, Combinatorics on Words, *Addison-Wesley*, Reading, Mass., 1983.

[Odl85] A. Odlyzko, *Enumeration of strings*, in A. Apostolico and Z. Galil Ed., *Combinatorial Algorithms on Words*, *12* of NATO *Advance Science Institute Series*, Series F: Computer and Systems Sciences, pp. 203-228, Springer Verlag, 1985.

[Rég89] M. Régnier, Knuth-Morris-Pratt algorithm: an analysis, in MFCS'89, *379*, *Lecture Notes in Comput. Sci.*, pp. 431-444, Springer-Verlag, 1989. Proc. Mathematical Foundations for Computer Science 89, Porubka, Poland.

[Ste84] J.-M. Steyaert, Structure et complexité des algorithmes. *Thèse d'État*, Université de Paris-VII, 1984.

APPENDIX

Proof of 2-factorization Lemma: 1-factorization lemma implies $s \subseteqq z$. If $s=z$, the non-overlapping property proved in 3.1 shows that $a \in z^*$ and the minimality constraint on g_2 implies $\alpha=\varepsilon$, hence $z=s=as$. Now, the case $as \subset z \subset sas$ also implies an overlap for s. Finally, if $s \subset z \subset as$, the minimality constraint on g_2 implies $as \notin z^*$. Hence, depending whether $z^2 \subset sas$ or

$sas \subset z^2$, there exists a factorization

$$as = vz, \qquad z = uv, \qquad \text{with} \quad u \subseteqq s \quad \text{or} \quad s \subseteqq u.$$

Again, $u \subset s$ implies $s = v' u = tv'$, $v' \subset v$, which contradicts 0-minimality. Similarly, $v \subset s$ implies $s = u' v = tu'$, $u' \subset u$ which yields the same contradiction. Then: $u = bs$ and $v = ds$. Hence: $sas = s \,.\, vuv = s \,.\, dsbsds = sds \,.\, b \,.\, sds$ which contradicts the minimality assumption.

Proof of Proposition 4.1: We have already proved (i). Let $h_1 = s' a' s'$. If h_1 is not minimal, define $g_1 = sas$, $g_1 \underline{\underline{\prec}} h_1$ *and* $g_1 \subseteqq h_1$. If $|s| \neq |s'|$, the largest of the two overlaps with himself, a contradiction. Again,

$$sas \underline{\underline{\prec}} sa's \;\Leftrightarrow\; as \underline{\underline{\prec}} a's \;\Rightarrow\; as \underline{\underline{\prec}} a'$$

when $a \neq a'$. Similarly, $sa \subseteqq a'$ and we note: $a' = asb = csa$. Now, first:

$$|sb| \leqq |sa| \;\Leftrightarrow\; sb \subseteqq sa \;\Leftrightarrow\; sb \subseteqq a \quad \text{or} \quad b = a.$$

If $sb \subseteqq a$, we can factorize: $a = dsb$ and we have: $asb = csa = cs \,.\, dsb = csd \,.\, sb$. Hence:

$$\begin{Bmatrix} sd \subseteqq a \\ ds \underline{\underline{\prec}} a \end{Bmatrix} \;\Rightarrow\; \begin{Bmatrix} sds \subseteqq as \subseteqq sas \\ sds \underline{\underline{\prec}} sa \underline{\underline{\prec}} sas \end{Bmatrix} \;\Rightarrow\; d = a.$$

From this contradiction, it follows that:

$$b = a.$$

Second, if $|b| \geqq |sa|$, then $b = wsa$ and $h_1 = sa's = saswsas$ and we apply the Lemma of 2-factorization. Now, let b_1 be in B_1, not necessarily in H_1. Then:

$$b_1 = s \,.\, ws = s \,.\, (a's)^p = \begin{cases} s \,.\, (as)^q & \text{if} \quad a's \in (as)^* \\ s \,.\, (as\,(usas)^m)^q = sas \,.\, xsas & \text{otherwise} \end{cases}$$

In the second case, $b_1 = sasxsas$. Applying 2-factorization Lemma yields $xsas = (usas)^m$.