Enumeration of bordered words

Mireille Regnier

INRIA

UNITE DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports de Recherche

N° 956

Programme 1

ENUMERATION OF BORDERED WORDS

Mireille REGNIER

Décembre 1988

ENUMERATION OF BORDERED WORDS
ENUMÉRATION DES MOTS À BORDS

Mireille REGNIER

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 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 étudions dans ce papier la famille des mots à 1 bord sur un q-alphabet, i.e. les mots de la forme bwb. Nous considérons aussi les mots à k bords. Nous dénombrons ces mots, en calculant les séries génératrices associées. Puis nous obtenons des estimations asymptotiques par la méthode de Darboux. En particulier, nous prouvons que la densité des mots à k bords est non nulle.

ENUMERATION OF BORDERED WORDS
Le langage de la vache-qui-rit †

Mireille REGNIER
INRIA-Rocquencourt
78 153-Le Chesnay (FRANCE)

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 derive asymptotic estimates by the Darboux method. In particular, we prove that the density of k-bordered words is α_k, $\alpha_k \neq 0$.*

I. INTRODUCTION:

This note is devoted to 1-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 side 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. This part makes use of the general theorems in combinatorics on 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.

In Section II, we list some general theorems in combinatorics to be used in the following sections. In Section III, we introduce our techniques on the set B_1. In Section IV, we consider the more general and intricate case of the set B_2. In Section V, we extend the methods and results of Section IV to the general case of sets B_k. In Section VI, we deal with the asymptotics of b_n^k.

---

† *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...*

II. SOME COMBINATORIAL THEOREMS ON WORDS:

In all this paper we will use the:

Definition 1: *The length of a word w on a q-ary alphabet is the number of letters w is a product of and is denoted by $|w|$.*

Definition 2: *A word $v \in A^*$ is said to be a* **factor** *of a word $x \in A^*$ if there exist words $u, w \in A^*$ such that:*

$$x = uvw.$$

A word $v \in A^*$ is said to be a **left** *(resp.a* **right** *)* **factor** *of a word $x \in A^*$ if there exists a word $w \in A^*$ such that:*

$$x = vw \ (\textbf{resp.} x = wv) \ .$$

We will note in the following:

$$v \preceq x (\textbf{resp.} v \subseteq x)$$

or, when the factors are strict:

$$v \prec x (\textbf{resp.} v \subset x).$$

The following theorems are stated and proved in [LO83,Sect. 1.3.] in the general case of free monoids. To introduce them, we first remark that if a given word w has two different sides b and b', with $|b'| > |b|$, there exist two words u and v, with $|u| = |v|$, such that: $b' = bu = vb$.

$$\begin{aligned} b' &= |--|----|--|----|--| \\ b \preceq b' &= |--|----|--| \\ b \subseteq b' &= \qquad\qquad |--|----|--| \\ &\qquad\qquad\quad u \quad\ v \quad\ u \end{aligned}$$

For example, if $w = 01001001001$, then: $b = 01$ and $b' = 01.001 = 010.01$.

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

Theorem A: *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:

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

The proof is given in [LO83], for free monoids. A direct proof will be given in Section III.

To determine a unique representation of 1-bordered words, we will use *primitive* words.

Definition 4: *A word $x \in A^*$ is said to be* **primitive** *if it is not a power of another word. Let P be the set of primitive words.*

Theorem B: *If*

$$x^n = y^n,\ x, y \in A^*, n, m > 0,$$

there exists a word z such that $x, y \in z^*$.

In particular, for each word $w \in A^+$, there exists a unique primitive word x such that $w \in x^*$.

Theorem C: *Two words $x, y \in A^+$ commute iff they are powers 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.*

III 1-BORDERED WORDS:

We first state our definition of 1-bordered words.

Definition 5: *Let A be a q-ary alphabet. Let B_1 be the set of words w of the form: bvb where $v \in A^*, b \in A^+$. One says that w is a bordered word and that b is a border of w. One notes $S = A^* - B_1$ its complement, 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 may have several borders. In order to get a *unique representation* of words in B_1 (and more generally in B_k), we need to determine the relationship between different sides of a given word. This problem of deciding a unique representation is fairly general in combinatorics on words. Some interesting examples can be found in [CP88], [DA1878] or [OD85].

Theorem 1: *Let b and b' be two borders of a same word w, with $|b'| > |b|$. Then there exists a primitive word z in A^*, and a factorization: $z = ae$ such that:*

$$\begin{cases} b = e.z^p, p \geq 0 \\ b' = e.z^q, q > p \end{cases}$$

Example: Consider the 1-bordered word $w = 1011101$. The words $b' = 101$ and $b = 1$ are two different borders. They satisfy the relationship above with: $z = 01, e = 1, a = 0$.

Proof: This is a consequence of Theorem A. As:$w = bxb = b'x'b'$, one has: $b' = bu = vb$, and u and v are conjugate. It is also worth giving a direct (and therefore intuitive) proof. We note:

$$\begin{cases} b = b_0 \\ (n-1)|u| < |b| \leq n|u|. \end{cases}$$

$$\text{If n=1 then} \quad \begin{cases} |b| \leq |u| = |v| \\ v = bv' \\ b' = b.(v'b) = b.z^q = e.z^p.z^q \end{cases}$$

$$\text{else} \quad \begin{cases} b = e_u.u^{n-1} = e_z.z^p \\ b' = e.z^p.u = e.z^q \end{cases}$$

Proposition 2: *Any word w in B_1 can be written, in a single way, as: $sw's$, where s is in S^* and w' in A^**

Proof: The smallest border is in S, and is the only border in S.

As a corollary of Proposition 2, we get the functional equation:

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

We use the methods developed in [GJ83] and [FL84]. The concatenation $s.w'$ of two different words translates in 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:

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

As: $B_1(z) + S(z) = W(z)$, Proposition 3 follows.

Proposition 3: *The series B_1 and S satisfy the functional equations:*

$$\begin{cases} S(z) = \dfrac{2 - S(z^2)}{1-qz} \\ B_1(z) = \dfrac{1}{1-qz}\left[\dfrac{qz^2}{1-qz^2} - B_1(z^2)\right] \end{cases}$$

Remark: Such an approach where combinatorial constructions translate into functional properties of generating functions is quite powerful. A general framework can be found in [FL84]. An example, related to the analysis of the Knuth-Morris-Pratt algorithm, can be found in [RE88].

In Section VI, we will derive asymptotics for the coefficients of these functions.

IV. 2-BORDERED WORDS:

In this Section we generalize the scheme of the previous Section. We have seen that any word in B_1 can be written, in a unique manner, under the form: sws. We define here a minimal set G_2 such that any word in B_2 can be written, in a unique manner: $g_2 w g_2$. Then, denoting $G_2(z)$ (resp. $B_2(z)$) the generating function counting the words in G_2 (resp. in B_2), one has:

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

Thus, we first state the definition of the minimal set G_2, extended for all the sets B_k. Then, we will characterize the elements of G_2. Finally, we will derive a functional equation on an associated generating function.

Definition 5: *A bordered word x is said to be k-minimal if:*

(i) $x \in B_k$

(ii) $w \preceq x, w \subseteq x, w \in B_k \Leftrightarrow w = x$

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

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

Remark: A word in G_k has no side in B_{k-1}. And any word x in B_{k-1} contains exactly one side in G_k . Hence, we define:

Definition 6: *We note $g(b_{k-1})$ the unique k-minimal word such that:*

$$g(b_{k-1}) \preceq b_{k-1} \text{ and } g(b_{k-1}) \subseteq b_{k-1}.$$

We can get now an equivalent to Proposition 2:

Theorem 4: *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 $g_k \in G_k$.

Hint: Let b_{k-1} be the largest border of b_k and choose $g_k = g(b_{k-1})$.

Our aim is now to *enumerate* the set G_k. To do so, we first derive some characterizations of the words g_k. For a sake of clarity, we first consider the case of G_2.

Our characterization of G_2 will rely on the Lemmas of 1-factorization and 2-factorization.

Lemma 5 (Lemma of 1-factorization:) *Let $x = z^*, z \in P$, be a word in L^*. Then, if a word $s \in S$ is a right (resp. left) factor of x, then s is a right (resp. left) factor of x.*

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

Remark 6: *A word in S cannot overlap with himself.*

As $s = zu = vz$ implies (Theorem A): $s = ea.e.(ae)^*$ which is not in S.

We know that any word in B_1 can be written *sas*. Proposition 8 and Lemma L, a generalization of Lemma 5, characterize the words *sas* that are in G_2.

Lemma 7 of 2-factorization *Let $g_2 = sas$ be a word in G_2, and $x = wsas \in z^*$, with $z \in P$. Then:*

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

Example: Les 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 $sas \not\subseteq z$.

Proposition 8 : *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: $g(b_1) = sas$.

Proof:

(i) From Proposition 2, one can write $g_2 = sa's$. If $as \notin P$, the Lemma of 1-factorization 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 important Remark 6 and Lemma 7 of 2-factorization, 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 1-bordered words. We prove:

Theorem 9: *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:*

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

and we have:

$$B_2(z) \;=\; G_2(z^2,z^2)W(z). \tag{3}$$

Proof: We define a 3-variate generating function associated to $H_1 - B_2$:

$$\phi(x,z,t_3) = \sum x^{|s|}z^{|as|}t_3^{|usas|}1_{sas\in G_2}1_{usas\in P}$$

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

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

From Proposition 8, this is also:

$$\sum_{g_2\in G_2}\Big(\sum_{m\geq 1} x^{|s|}z^{m|as|} + \sum_{usas\in P}\sum_{m\geq 1} x^{|s|}z^{|as|}z^{m|usas|}\Big)$$
$$= \sum_{m\geq 1} G_2(x,z^m) \;+\; \sum_{m\geq 1}\phi(x,z,z^m).$$

Hence:

$$S(zx)W(z) \;=\; \sum_{m\geq 1} G_2(x,z^m) \;+\; \sum_{m\geq 1}\phi(x,z,z^m) \tag{4}$$

To get the equation (5), we consider the words $g_2wg_2 = sas.w.sas \in B_2$. From the Lemma 7 of 2-factorization, we get:

$$\sum x^{|s|}z^{|as|}t^{wsas}1_{sas\in G_2}1_{w\in A^*} = \sum_{m\geq 3} G_2(x,zt^m) + \sum_{m\geq 1}\phi(x,z,t^m).$$

Hence:

$$G_2(xt_3,zt_3)W(t_3) = \sum_{m\geq 3} G_2(x,zt_3^m) + \sum_{m\geq 1}\phi(x,z,t_3^m). \tag{5}$$

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

V. PROPERTIES OF k-BORDERED WORDS

In this section, we extend the scheme of Section IV to the general case of k-bordered words. We first study the set G_k, and prove a Lemma of k-factorization. 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.

Proposition 10: *Let g_k be in G_k. It can be factorized as:*

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

with $g_i \in G_i, p_i \in P$.

Theorem 11 (k-overlapping): *Let g_k be a word in G_k and $x = z^*, z \in P$ a word in L^* such that:*

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

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

Theorem 12: *For any i, this decomposition is unique and has the following property:*

$$\text{If } g_i \subseteq p_i \text{ then } \begin{cases} \exists j \geq i : g_i.p_i \in B_j - B_{j+1} \\ \theta_i = 2^{k+1-j} - 1. \end{cases}$$

$$\text{else } \begin{cases} \exists m, m \leq i : g_m \subseteq p_i = p_m. \\ \theta_m = 2^{k+1-j} - 2^{i+1-j} \end{cases}$$

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 of Proposition 10: As $g_k \in B_{k-1}$, its largest border b satisfies:

$$b \in B_{k-2} \subset B_{k-1} \subset \ldots \subset B_1 \;.$$

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.

Proof of Theorem 11: We know by Remark K that Theorem 11 holds for g_1 in G_1. Assume now that it holds true for $G_1, \ldots, G_{k-1}$. If $z \neq wg_k$ we may assume $x = z$: as we have then: $z \subset g_k$, thus $g_k = e_z.z^p$ with $e_z \subseteq z$ and $p \geq 1$. We get the equation:

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

From Remark K, we know that $g_1 \subseteq x$. Thus, we may define $i, 1 \leq i \leq k-1$ by:

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

We have then: $g_{i+1}x = y'g_{i+1}$, with $|x| < |g_{i+1}|$. Applying Theorem 11, we have:

$$\begin{cases} g_{i+1} & = g'_m.p'^*_m \\ x & = p'_m \end{cases}$$

and

$$\left\{ \begin{array}{l} g'_m \subseteq g_{i+1} \subseteq g_k \\ g'_m \preceq g_{i+1} \preceq g_k \end{array} \right\} \Rightarrow g_m = g'_m.$$

Now, if $i = k-1$, we have: $g_{k-1} \subseteq x \subseteq g_k$. From above, we know that: $g_k = g_m p^*_m = g_m.x^*$. To prove in all other cases that $p'_m \in \{p_1, \ldots, p_{k-1}\}$, we remark that:

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

Moreover, we have $p \geq 1$(as $x \subseteq g_{k-1}$) and $m \geq 1$ (as $g_{i+1} \subseteq g_{k-1} \Rightarrow 2|g_{i+1}| \leq |g_k|$) . Then:

$$xa = ax$$

or (Theorem D): $a = \epsilon$ and $g_k = g_i.x^p = g_i.p_i^{\theta_i}$, hence: $x = p_i$.

Proof of Theorem 12: If $p_i \not\subseteq g_i$ then $|p_i| > |g_i|$. As $g_i \subseteq g_i.p_i^{\theta_i}$, we get $g_i \subseteq p_i$. Then:

$$g_i.p_i = g_i.w_i g_i \in B_i \subset B_{i+1} \subset \ldots \subset B_j \subset B_{j+1} \subset \ldots$$

The expression of θ_i follows immediately from the remark that the largest border of $g_i(w_ig_i)^m, m \geq 1$ is $g_i(w_ig_i)^{\lfloor \frac{m-1}{2} \rfloor}$.

Now, if $p_i \subset g_i$ the equation: $g_k = g_ip_i^* = yg_i$ implies:

$$\begin{cases} g_i = g'_m.p'^*_m = g_m p^\theta_m, \theta = 2^{i+1-j} - 1 \\ p_i = p'_m. \end{cases}$$

Hence: $g_k = g_i.p_i^* = g_m.p^\theta_m.p^{\theta_i}_m$ with: $\theta_i = (2^{k+1-j} - 1) - \theta$.

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 $H_{i,k} = \{g_i.p_i | g_i \in G_i, p_i \in P, g_i \subseteq p_i, g_i.p_i \in B_k - B_{k+1}\}$. And $\phi_{i,k+1}(t_1, \ldots, t_i, t_{i+1}, \ldots, t_k)$ counts $L_{i,k+1} = \{g_i.p_i | g_i \in G_i, p_i \in P, g_i \subseteq p_i, g_i.p_i \in B_{k+1}\}$. One has: $\psi_{i,j} = \phi_{i,j} - \phi_{i,j+1}$. Then: $G_k(t_1, \ldots, t_k) = \sum_{i=1}^k \psi_{i,k}(t_1, \ldots, t_k) + f(\{\phi_{i,j}\}_{1 \leq j \leq k-1})$. For any k, counting $\{g_iwg_i\}$ and applying Theorem 12 yields k functional equations. These equations involve: $2k$ unknown, but dependent, functions $(\phi_{i,k+1})_{1 \leq i \leq k+1}$ and $(\psi_{i,k})_{1 \leq i \leq k}$. One can derive $(\phi_{i,k+1})_{1 \leq i \leq k+1}$ from these equations and hence: $\psi_{i,k} = \phi_{i,k} - \phi_{i,k+1}$. Finally, we get G_{k+1}. We can also relate these notations to the ones in Section IV:

$$\phi(t_1, t_2, t_3) = \phi_{2,2}(t_1, t_2, t_3)$$
$$\phi(t_1, t_2, t_2) = \phi_{1,2}(t_1, t_2) = \phi_{2,2}(t_1, t_2, t_3)$$

We also have: $\phi_{1,1}(t_1, t_2) = G_2(t_1, t_2)$. Equation (4) counts $\{g_1.wg_1\}$ and Equation (5) counts $\{g_2.wg_2\}$.

VI. ASYMPTOTICS ON k-BORDERED WORDS

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 [DA1878]. Examples of this approach are developed in [ST84].

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:*

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

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

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

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

Theorem 13: *Let $B_1(z) = \sum b_n^1 z^n$ be the generating function of the bordered words on a q-ary alphabet A. Then:*

$$b_n^1 \sim \alpha_1 q^{-n}$$

where: $\alpha_1 = S(\frac{1}{q^2})$.

Proof: We know that:$W(z) = S(z) + B_1(z)$. Thus, S and B_1 are both defined at least for $|z| < \frac{1}{q}$. Hence $S(z^2)$ is analytic for $|z| < \frac{1}{\sqrt{q}}$, notably around $z = 1/q$. We apply the Darboux theorem to equation (1). Note that $S(\frac{1}{q^2})$ can be numerically computed as:

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

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

q	α_1
2	0.8638659
3	0.4617496
10	0.1101101

Table 1

Theorem 14: *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: $\alpha_k = G_k(\frac{1}{q^2}, \ldots, \frac{1}{q^2})$.

Proof: 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 $z = \frac{1}{q^2}$. It allows for the application of the Darboux theorem with $s = 1$ and $g(z) = G_k(z^2, \ldots, z^2)$.

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 $k = 1$. Such methods also apply to other combinatorial problems on words.

References:

[CP88] M. CROCHEMORE AND D. PERRIN "Pattern Matching in Strings" in LITP (Paris) Research Report 88-5.

[DA1878] G. DARBOUX "Mémoire sur l'approximation des fonctions de très grands nombres, et sur une classe étendue de développements en série" in *J. de Mathématiques pures et appliquées 3-ème série* **VI**, (1878) 1–56.

[FL84] PH. FLAJOLET *"Mathematical Methods in the Analysis of Algorithms and Data Structures"*, Lecture Notes for *A Graduate Course on Computation Theory*, Udine(Italy) 225–304(1984).

[GJ83] I. GOULDEN AND D. JACKSON *"Combinatorial Enumerations"*, Wiley, New York, (1983).

[GO81] L.J. GUIBAS AND A.M.ODLYZKO "String overlaps, pattern matching, and nontransitive games" in *J. Comb. Theory (A)*, **30** (1981) 183–208.

[LO83] LOTHAIRE *"Combinatorics on Words"*, Addison-Wesley, Reading,Mass., (1983).

[OD85] A.M. ODLYZKO *"Periodicities in Strings"* in Combinatorial Algorithms on Words, Springer NATO ASI SEr. F12, (1985) 241–254.

[RE88] M. RÉGNIER *"Knuth-Morris-Pratt Algorithm: an Analysis"* Actes des Journées Informatique et Mathématique, Marseille, (1988).

[ST84] J.M. STEYAERT *"Structure et Complexité des Algorithmes"* Thèse d'Etat, Paris-Orsay, (1984).

APPENDIX

Proof of Lemma 7: Proposition E implies $s \subseteq z$. From Remark K: $s = z \Rightarrow a \in z^*$, and the mimimality constraint implies $\alpha = \epsilon$, hence $z = s = as$. The case $as \subset z \subset sas$ contradicts also Remark K. Now, if $s \subset z \subset as$ then (minimality condition) $as \notin z^*$, and there exists a factorization

$$z = uv,\ v \preceq as,\quad u \subseteq s \text{ or } s \subseteq u$$

From Remark K (again!): $s \subseteq u$ and $s \subseteq v$. Then: $u = bs$ and v satisfies: $ds \preceq as$ and $ds \subseteq as$, hence $d = a$. Thus $z = u.v = bs.as$ which contradicts our assumption ($z \subset as$).

Proof of Proposition 8: We have already proved (i). Let $h_1 = s'a's'$. From Proposition E and Remark K, one has: $s = s'$ and: $as \preceq a's \Rightarrow as \preceq a'$. If we note: $a' = asb = csa$, we have (Remark K):

$$\begin{cases} b \subseteq sa \Rightarrow b \subseteq a \\ b \not\subseteq a \Rightarrow sb \not\subseteq sa \ \Rightarrow sb \subseteq a. \end{cases}$$

Thus, if $b \not\subseteq a$, we can factorize: $a = dsb$ and we have: $asb = csa = cs.dsb = csd.sb$. Hence:

$$\left\{\begin{array}{l} sd \subseteq a \\ ds \preceq a \end{array}\right\} \quad \Rightarrow \left\{\begin{array}{l} sds \subseteq as \subseteq sas \\ sds \preceq sa \preceq sas \end{array}\right\} \quad \Rightarrow d = a.$$

From this contradiction, it follows that:

$$|b| \prec |sa| \ \Rightarrow \ b = a.$$

If $|b| \geq |sa|$, 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:

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

Applying Lemma 7 of 2-factorization in the second case yields $sasxsas = sas(usas)^m = b_1$.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique