Theoretical Informatics and Applications
Theoret. Informatics Appl. **36** (2002) 67–96
DOI: 10.1051/ita:2002006

ON THE DISTRIBUTION OF CHARACTERISTIC PARAMETERS OF WORDS *

Arturo Carpi[1] and Aldo de Luca[2]

Abstract. For any finite word w on a finite alphabet, we consider the basic parameters R_w and K_w of w defined as follows: R_w is the minimal natural number for which w has no right special factor of length R_w and K_w is the minimal natural number for which w has no repeated suffix of length K_w. In this paper we study the distributions of these parameters, here called characteristic parameters, among the words of each length on a fixed alphabet.

Mathematics Subject Classification. 68R15, 68R05.

Introduction

As is well known, a fundamental role in combinatorics on words is played by extendable and special factors (see, *e.g.* [1, 6, 8] and references therein). We recall that a factor u of a word w is (*right*) *extendable* if there exists a letter a such that ua is still a factor of w and it is (*right*) *special* if there exist two distinct letters a and b such that ua and ub are both factors of w.

Much information about the structure of a word w can be obtained by knowing some numerical parameters such as, for instance, the periods of the word (see, *e.g.* [10–12, 14]). Other parameters of this kind are the minimal natural number R_w such that w has no right special factor of length R_w and the length K_w of

Keywords and phrases: Special factor, characteristic parameter, repeated factor.

* *The work for this paper has been supported by the Italian Ministry of Education under Project COFIN 2001 – Linguaggi Formali e Automi: teoria ed applicazioni.*

[1] Dipartimento di Matematica e Informatica, Università di Perugia, via Vanvitelli 1, 06123 Perugia, Italy; e-mail: `carpi@dipmat.unipg.it`

[2] Dipartimento di Matematica dell'Università di Roma "La Sapienza", piazzale Aldo Moro 2, 00185 Roma, Italy and Centro Interdisciplinare "B. Segre", Accademia dei Lincei, via della Lungara 10, 00100 Roma, Italy; e-mail: `deluca@mat.uniroma1.it`

TABLE 1. $\frac{1}{2}D_R(i,n),\ 0 \le i \le n \le 20,\ n > 0.$

n \ i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0																			
2	1	1	0																		
3	1	2	1	0																	
4	1	2	4	1	0																
5	1	2	8	4	1	0															
6	1	2	10	14	4	1	0														
7	1	2	10	32	14	4	1	0													
8	1	2	10	53	43	14	4	1	0												
9	1	2	10	77	104	43	14	4	1	0											
10	1	2	10	97	215	125	43	14	4	1	0										
11	1	2	10	105	404	315	125	43	14	4	1	0									
12	1	2	10	105	683	720	340	125	43	14	4	1	0								
13	1	2	10	105	1042	1557	852	340	125	43	14	4	1	0							
14	1	2	10	105	1469	3172	2010	896	340	125	43	14	4	1	0						
15	1	2	10	105	1929	6103	4581	2230	896	340	125	43	14	4	1	0					
16	1	2	10	105	2407	11076	10121	5342	2281	896	340	125	43	14	4	1	0				
17	1	2	10	105	2887	19149	21631	12445	5602	2281	896	340	125	43	14	4	1	0			
18	1	2	10	105	3343	31762	44785	28330	13358	5672	2281	896	340	125	43	14	4	1	0		
19	1	2	10	105	3695	50857	89989	63158	31219	13732	5672	2281	896	340	125	43	14	4	1	0	
20	1	2	10	105	3823	78908	176030	137969	71721	32536	13807	5672	2281	896	340	125	43	14	4	1	0

the shortest unrepeated suffix of w. For instance, the maximal length G_w of a repeated factor of a non-empty word w is given [8] by

$$G_w = \max\{R_w, K_w\} - 1.$$

Moreover, as proved in [1], a word is uniquely determined by its factors up to length $G_w + 2$. This result suggested an algorithm for "sequence assembly" [5]. Moreover, some generalizations of the notion of periodic word, based on the previous parameters, have been recently considered in [2,3].

In the sequel we shall refer to R_w and K_w as the *characteristic parameters* of the word w. The aim of this paper is to study how the values of the characteristic parameters, as well as of some other related quantities, are distributed among the words of each length.

Fixed a d-letter alphabet A, for any pair of natural numbers i and n, we denote by $D_R(i,n)$ and $D_K(i,n)$ the number of words w of length n on the alphabet A such that, respectively, $R_w = i$ and $K_w = i$.

In the case of a binary alphabet, the values of $D_R(i,n)/2$ and $D_K(i,n)/2$ for small values of i and n are given in Tables 1 and 2, respectively. By inspecting these tables, one can recognize several regularities: for instance, the values of D_R on each column are initially increasing, and then constant at least on the first few columns. In both tables there are long diagonal segments where the values are constant.

In Section 2 we study the relations among the characteristic parameters of a given word and those of the words obtained by adding a letter on its right or on its left. For any non-empty word w we consider the set B_w of the letters extending on the right the longest repeated suffix of w in a factor of w. The main result of the section (*cf.* Prop. 2.1) states that for any letter $a \in B_w$ one has $R_{wa} = R_w$ and $K_{wa} = K_w + 1$, while, for any other letter b, $R_{wb} = \max\{R_w, K_w\}$ and $K_{wb} \le \min\{K_w, 1 + R_w\}$.

In Section 3 we establish some properties of the maximal length G_w of a repeated factor of a word w. In particular we show that the length $|w|$ of a word w cannot

TABLE 2. $\frac{1}{2}D_K(i,n),\ 0 \le i \le n \le 20,\ n > 0.$

n \ i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	1																			
2	0	1	1																		
3	0	1	2	1																	
4	0	1	4	2	1																
5	0	1	6	6	2	1															
6	0	1	9	13	6	2	1														
7	0	1	13	26	15	6	2	1													
8	0	1	19	49	35	15	6	2	1												
9	0	1	28	89	75	39	15	6	2	1											
10	0	1	42	158	160	88	39	15	6	2	1										
11	0	1	64	278	331	197	90	39	15	6	2	1									
12	0	1	99	486	671	428	210	90	39	15	6	2	1								
13	0	1	155	847	1338	922	464	216	90	39	15	6	2	1							
14	0	1	245	1475	2641	1948	1028	485	216	90	39	15	6	2	1						
15	0	1	390	2570	5167	4078	2233	1087	489	216	90	39	15	6	2	1					
16	0	1	624	4484	10037	8460	4818	2382	1104	489	216	90	39	15	6	2	1				
17	0	1	1002	7838	19387	17428	10281	5197	2434	1110	489	215	90	39	15	6	2	1			
18	0	1	1613	13730	37277	35679	21776	11225	5344	2459	1110	489	215	90	39	15	6	2	1		
19	0	1	2601	24106	71402	72672	45828	24078	11606	5419	2463	1110	489	215	90	39	15	6	2	1	
20	0	1	4199	42422	136336	147350	95948	51304	25055	11798	5444	2463	1110	489	215	90	39	15	6	2	1

exceed $G_w + d^{R_w}$ which implies $G_w \ge \lfloor \log_d |w| \rfloor - 1$ when $d > 1$. We also introduce the notion of *symmetric word*. A symmetric word of order m is any word w such that $R_w = K_w = m$. We show that there exists a symmetric word of order m and length n if and only if $2m \le n \le d^m + m - 1$.

In Section 4 we study the functions D_R and D_K, as well as some other related functions. We prove that for all $i, n > 0$ one has

$$D_R(i, n+1) = D_R(i, n) + (d-1)D_G(i-1, n),$$

where $D_G(i, n)$ denotes the number of the words of length n having repeated factors of maximal length i. Some further relations allow one to reduce the computation of D_R to the evaluation of $\mathrm{Card}(B_w)$ on symmetric words. More precisely, denote by $D^*_G(i, n)$ and $D^>_K(i, n)$ the number of the words of length n such that, respectively, $G_w \ge i$ and $K_w = i > R_w$, and let $D_S(i, n)$ be the sum of $\mathrm{Card}(B_w)$ extended to all symmetric words w of order i and length n. Then for $i \ge 0$ and $n > 1$ one has

$$D^*_G(i, n) = dD^*_G(i, n-1) + D^>_K(i+1, n),$$

and for $i, n \ge 0$,

$$D^>_K(i+1, n+1) = D^>_K(i, n) + D_S(i, n).$$

Thus, since $D_G(i, n) = D^*_G(i, n) - D^*_G(i+1, n)$, the values of any of the functions D_R, D_G, D^*_G, $D^>_K$, and D_S can be computed by knowing the values of only one of them. Moreover, for $i, n \ge 0$, one has

$$D_K(i, n+1) = dD_K(i, n) + D_P(i, n+1) - D_P(i+1, n+1),$$

where $D_P(i, n)$ denotes the number of periodic-like words (*cf.* [3]) w of length n such that $K_w = i$.

We also show that when i is fixed and n grows, $D_R(i, n)$ and $D_K(i, n)$ are non-decreasing. This is not true for $D_G(i, n)$, because one has $D_G(i, n) \ne 0$ if and only if $i < n \leq i + d^{i+1}$. Among other results, we prove that

$$D_G(i-1, n) \leq D_R(i, n) + D_K(i, n),$$

where equality holds if and only if $d = 1$ or $i > n/2$.

In Section 5 we study the "diagonal behaviour" of D_R, D_K, and D_G, *i.e.*, the behaviour of $D_R(i, n)$, $D_K(i, n)$, and $D_G(i, n)$ when variables i and n are simultaneously increased by 1. We show that, for any $i, n \geq 0$,

$$D_K(i, n) \leq D_K(i+1, n+1),$$

where equality holds if and only if $i > n/2$. In other terms, for any fixed $m \geq 0$, the values of D_K on the points of a diagonal line $(t, m+t)_{t \geq 0}$ are initially increasing and ultimately constant. Similar properties hold for D_G^* and $D_K^>$. Moreover, one has

$$D_R(t, m+t) \leq D_R(m, 2m) \quad \text{and} \quad D_G(t, m+t) \leq D_G(m, 2m),$$

where the "=" sign holds in the first equation if and only if $t \geq m$ and in the second one if and only if $t \geq m - 1$.

A consequence of these results is that when $i > n/2$ the values of $D_R(i, n)$, $D_K(i, n)$, and $D_G(i, n)$ depend uniquely on the difference $n - i$. In a forthcoming paper [4], we shall give the exact values of $D_G(i, n)$, $D_R(i, n)$, and $D_K(i, n)$ when $i > n/2$. In view of the diagonal behaviour, these values give upper bounds to the previous distributions in the general case. Moreover, we shall study the most frequent and the average values of the characteristic parameters and of the maximal length of a repeated factor over the set of all words of length n.

1. Preliminaries

Let A be a finite non-empty set, or *alphabet*, and A^* the set of all finite sequences of elements of A, including the empty sequence denoted by ϵ. The elements of A are usually called *letters* and those of A^* *words*. The word ϵ is called *empty word*. We set $A^+ = A^* \setminus \{\epsilon\}$. A word $w \in A^+$ can be written uniquely as a sequence of letters as

$$w = a_1 a_2 \cdots a_n,$$

with $a_i \in A$, $1 \leq i \leq n$, $n > 0$. The integer n is called the *length* of w and denoted by $|w|$. By definition, the length of ϵ is equal to 0. For any $n \geq 0$ we set $A^n = \{w \in A^* \mid |w| = n\}$. We shall denote by $w^\sim$ the *reversed word* of w, *i.e.*, $w^\sim = a_n a_{n-1} \cdots a_1$. Moreover, we set $\epsilon^\sim = \epsilon$.

Let $w \in A^*$. The word $u \in A^*$ is a *factor* (or *subword*) of w if there exist words λ, μ such that $w = \lambda u \mu$. A factor u of w is called *proper* if $u \neq w$. If $w = u\mu$, for some word μ (resp. $w = \lambda u$, for some word λ), then u is called a *prefix* (resp.

suffix) of w. For any word w we denote respectively by $\mathrm{Fact}(w)$, $\mathrm{Pref}(w)$, and $\mathrm{Suff}(w)$ the sets of its factors, prefixes, and suffixes.

Let $u \in \mathrm{Fact}(w)$. Any pair $(\lambda, \mu) \in A^* \times A^*$ such that $w = \lambda u \mu$ is called an *occurrence* of u in w. If $\lambda \neq \epsilon$ and $\mu \neq \epsilon$, then the occurrence of u is called *internal.* A factor u of w is *repeated* if it has at least two distinct occurrences in w, otherwise it is called *unrepeated.*

A factor u of w is *right extendable* (resp. *left extendable*) in w if there exists a letter $x \in A$ such that $ux \in \mathrm{Fact}(w)$ (resp. $xu \in \mathrm{Fact}(w)$). The factor ux (resp. xu) of w is called a *right* (resp. *left*) *extension* of u in w.

A word s is called a *right* (resp. *left*) *special factor* of w if there exist two letters $x, y \in A$, $x \neq y$, such that $sx, sy \in \mathrm{Fact}(w)$ (resp. $xs, ys \in \mathrm{Fact}(w)$). From the definition, one has that any suffix (resp. prefix) of a right (resp. left) special factor of w is right (resp. left) special.

With each word w one can associate a word k_w defined as the shortest suffix of w which is an unrepeated factor of w. This is also equivalent to say that k_w is the shortest factor of w which is not right extendable in w. In a symmetric way, one can define h_w as the shortest factor of w which is not left extendable in w.

One can remark that all proper suffixes of k_w and all proper prefixes of h_w are repeated factors, while k_w and h_w are unrepeated. In the following, we shall denote by k'_w (resp. h'_w) the longest repeated suffix (resp. prefix) of a non-empty word w.

For any word w we shall consider the parameters $K_w = |k_w|$ and $H_w = |h_w|$. Moreover, we shall denote by R_w the minimal natural number such that there is no right special factor of w of length R_w and by L_w the minimal natural number such that there is no left special factor of w of length L_w.

By definition, if $w \in A^+$, then

$$0 < H_w, K_w \leq |w|, \quad 0 \leq R_w, L_w < |w|. \tag{1}$$

Note that, in both equations, equality holds if and only if w is a power of a letter. For the empty word ϵ one has $R_\epsilon = L_\epsilon = H_\epsilon = K_\epsilon = 0$.

Let $w = a_1 a_2 \cdots a_n$ be a word, $a_i \in A$, $i = 1, \ldots, n$. A positive integer $p \leq n$ is called a *period* of w if for all $i, j \in [1, n]$ such that $i \equiv j \pmod{p}$, one has $a_i = a_j$.

Example 1.1. Let $A = \{a, b, c\}$ and $w = abccacbccabaab$. One has $|w| = 14$, $k_w = aab$, $k'_w = ab$, $h_w = abc$, $h'_w = ab$. Thus, $K_w = H_w = 3$. The right special factors of w are ϵ, a, b, c, ab, ca, cca, $bcca$. The left special factors of w are ϵ, a, b, c, ab, bcc, $bcca$. Hence, $R_w = L_w = 5$ and the periods of w are 12 and 14.

In the case of the word $v = abbbb$ on the alphabet $\{a, b\}$ one has $R_v = H_v = 1$ and $L_v = K_v = 4$.

2. Characteristic parameters of a word

As we have seen in the previous section, with any word w one can associate the integers R_w, L_w, K_w, and H_w. In subsequent sections we shall refer mainly to parameters R_w and K_w. However, any result admits a symmetric dual version in which "right" is replaced by "left", R_w by L_w, K_w by H_w, k_w by h_w and so on. The parameters R_w and K_w will be also called *characteristic parameters* of the word w. For any word $w \in A^+$ we set

$$B_w = \{a \in A \mid k'_w a \in \mathrm{Fact}(w)\}\cdot$$

Moreover, we define $B_\epsilon = A$. Thus, if $w \neq \epsilon$, B_w is the set of letters of A extending on the right k'_w in w. We remark that $B_w \neq \emptyset$, as k'_w is right extendable in w.

Proposition 2.1. *Let $w \in A^*$. For any $x \in B_w$ one has*

$$K_{wx} = K_w + 1 \quad \textit{and} \quad R_{wx} = R_w.$$

For any $x \in A \setminus B_w$ one has

$$K_{wx} \leq \min\{K_w, 1 + R_w\} \quad \textit{and} \quad R_{wx} = \max\{R_w, K_w\}\cdot$$

Proof. If $w = \epsilon$ the result is trivial. Thus, we assume $w \neq \epsilon$. If $x \in B_w$, then $k'_w x$ is a right extension of k'_w in w, so that $k'_w x$ is repeated in wx, whereas $k_w x$ is unrepeated in wx since it does not occur in w. Hence,

$$k_{wx} = k_w x$$

so that $K_{wx} = K_w + 1$.

Since any right special factor of w is also a right special factor of wx one has $R_{wx} \geq R_w$. Let us suppose that $R_{wx} > R_w$. This implies that there exists a right special factor s of wx of length R_w, *i.e.*, there are two distinct letters a and b such that $sa, sb \in \mathrm{Fact}(wx)$ and $|s| = R_w$. Since $a \neq b$, at least one of the words sa and sb is a factor of w. Since s is not a right special factor of w, one of these two words, say sb, does not occur in w and, therefore, it has to be a suffix of wx, that implies $x = b$ and $s \in \mathrm{Suff}(w)$. Since sa is a factor of w, s is a repeated suffix of w. This implies that s is a suffix of k'_w, so that sb is a suffix of $k'_w b$. Since $k'_w x$ is a factor of w, one has $sb \in \mathrm{Fact}(w)$, which is a contradiction. Thus, $R_{wx} = R_w$.

If $x \notin B_w$, then $k'_w x$ is an unrepeated suffix of wx. Hence, $K_w = |k'_w x| \geq K_{wx}$. Let us prove now that $K_{wx} \leq 1 + R_w$. This is trivial if $K_{wx} = 1$. If $K_{wx} \geq 2$, then $k'_{wx} = sx$ with $s \in \mathrm{Suff}(k'_w)$ as $K_{wx} \leq K_w$. Since k'_{wx} is repeated in wx, $sx = k'_{wx} \in \mathrm{Fact}(w)$. Moreover, since $B_w \neq \emptyset$ there exists a letter $a \in B_w$ such that $k'_w a \in \mathrm{Fact}(w)$ which implies $sa \in \mathrm{Fact}(w)$. We conclude that s is right special in w, so that $K_{wx} = |s| + 2 \leq R_w + 1$.

Since k'_w is right extendable in w but $k'_w x \in \mathrm{Fact}(wx) \setminus \mathrm{Fact}(w)$, it follows that k'_w is right special in wx. Thus $R_{wx} \geq K_w$. Moreover, trivially, $R_{wx} \geq R_w$, so that $R_{wx} \geq \max\{R_w, K_w\}$. Suppose that $R_{wx} > \max\{R_w, K_w\}$. This implies that there exists a right special factor s of wx of length $\max\{R_w, K_w\}$. Proceeding as in the first part of the proof, one derives that s is a repeated suffix of w, which is a contradiction, because $|s| > K_w$. □

Example 2.2. In the case of the word w of Example 1.1 one has $B_w = \{a, c\}$. Hence $K_{wa} = K_{wc} = 4 = K_w + 1$, whereas $R_{wa} = R_{wc} = 5 = R_w$. Moreover, $K_{wb} = 2 = K_w - 1$ and $R_{wb} = 5 = \max\{R_w, K_w\}$.

From Proposition 2.1, we derive the following noteworthy lemmas, which will be used in the sequel.

Lemma 2.3. *Let w be a word such that $R_w \geq K_w$. For any letter x one has*

$$R_{wx} = R_w.$$

Proof. By Proposition 2.1 one has either $R_{wx} = R_w$ or $R_{wx} = \max\{R_w, K_w\}$. Since $R_w \geq K_w$, in any case it follows $R_{wx} = R_w$. ☐

Lemma 2.4. *Let w be a word such that $R_w < K_w$. There exists a unique letter a such that*

$$K_{wa} = K_w + 1 \quad \textit{and} \quad R_{wa} = R_w.$$

For all letters $x \neq a$,

$$K_{wx} \leq 1 + R_w \leq K_w \quad \textit{and} \quad R_{wx} = K_w.$$

Proof. Since $R_w < K_w$, one has $|k'_w| \geq R_w$ so that k'_w is not a right special factor of w. However, as k'_w is right extendable in w, the set B_w contains a unique letter, say a. By Proposition 2.1, $K_{wa} = K_w + 1$ and $R_{wa} = R_w$. For all other letters $x \neq a$, as $x \notin B_w$ and $R_w < K_w$, by Proposition 2.1 it follows $K_{wx} \leq 1 + R_w$ and $R_{wx} = K_w$. ☐

Example 2.5. In the case of the word v of Example 1.1 one has $K_{vb} = 5 = K_v + 1$ and $R_{vb} = 1 = R_v$ whereas $K_{va} = 2 = 1 + R_v$ and $R_{va} = 4 = K_v$.

Lemma 2.6. *For any $w \in A^*$ and any $x \in A$ one has*

$$K_w \leq K_{xw} \leq 1 + K_w \quad \textit{and} \quad R_w \leq R_{xw} \leq 1 + R_w.$$

Proof. The result is trivial for $w = \epsilon$. For any $w \in A^+$ and any $x \in A$, one has $K_{xw} \geq K_w$ and $R_{xw} \geq R_w$. Indeed, k'_w is repeated in xw and any right special factor of w is a right special factor of xw.

If $K_{xw} = 1$, then certainly $K_{xw} \leq 1 + K_w$. Let us then suppose that $K_{xw} > 1$. In this case, we can write $k'_{xw} = yt$, with $y \in A$ and $t \in A^*$. Thus t is a repeated suffix of w and, therefore, $|t| \leq K_w - 1$. Hence, one derives $K_{xw} \leq 1 + K_w$.

Similarly, if $R_{xw} = 1$, then certainly $R_{xw} \leq 1 + R_w$. Let us then suppose that $R_{xw} > 1$. In this case, there is a right special factor u of xw of length $R_{xw} - 1$. We can write $u = yt$, with $y \in A$ and $t \in A^*$. Since t is a right special factor of w one has $|t| \leq R_w - 1$. Hence, one derives $R_{xw} \leq 1 + R_w$. ☐

A further consequence of Proposition 2.1 is the following proposition proved in [8] with a different technique:

Proposition 2.7. *For any word w,*

$$|w| \geq R_w + K_w.$$

Proof. The proof is by induction on the length of w. The statement is trivially true if $|w| \leq 1$. Let us then suppose that $|w| \geq 1$ and $|w| \geq R_w + K_w$. We shall prove that for any letter x one has $|wx| \geq R_{wx} + K_{wx}$. From Proposition 2.1, if $x \in B_w$ one has $K_{wx} = K_w + 1$ and $R_{wx} = R_w$, so that

$$R_{wx} + K_{wx} = R_w + K_w + 1 \leq |w| + 1 = |wx|.$$

If $x \notin B_w$ one has $K_{wx} \leq \min\{K_w, 1 + R_w\}$ and $R_{wx} = \max\{R_w, K_w\}$. Thus

$$\begin{aligned} R_{wx} + K_{wx} &\leq \min\{K_w, 1 + R_w\} + \max\{R_w, K_w\} \leq 1 + R_w + K_w \leq |w| + 1 \\ &= |wx|. \end{aligned}$$

This concludes the proof. □

Corollary 2.8. *Let w be a word. The following relations hold:*

$$\begin{aligned} 2K_w > |w| &\Rightarrow R_w < K_w, \qquad & 2R_w > |w| &\Rightarrow K_w < R_w, \\ 2K_w \geq |w| &\Rightarrow R_w \leq K_w, \qquad & 2R_w \geq |w| &\Rightarrow K_w \leq R_w. \end{aligned}$$

Proof. Let us suppose $2K_w > |w|$. By the preceding proposition, $2K_w > K_w + R_w$ that implies $R_w < K_w$. All other relations are proved in a similar way. □

The following lemma, whose proof is trivial, will be useful in the sequel:

Lemma 2.9. *For any word w one has $L_{w^\sim} = R_w$, $R_{w^\sim} = L_w$, $H_{w^\sim} = K_w$, and $K_{w^\sim} = H_w$.*

3. Length of repeated factors

In the sequel A will denote a fixed alphabet having cardinality $d > 0$. Let w be a non-empty word. We denote by G_w the maximal length of a repeated factor of w. We recall the following important relation between G_w and the characteristic parameters of a non-empty word w [8].

$$G_w = \max\{R_w, K_w\} - 1 = \max\{L_w, H_w\} - 1. \tag{2}$$

The *subword complexity* λ_w of a word w is the map $\lambda_w : \mathbb{N} \rightarrow \mathbb{N}$ defined by

$$\lambda_w(n) = \text{Card}(\{v \in \text{Fact}(w) \mid |v| = n\}), \quad \text{for all } n \in \mathbb{N}.$$

As proved in [7, 8], λ_w reaches its maximum value in $G_w + 1$ having

$$\lambda_w(G_w + 1) = \lambda_w(R_w) = |w| - \max\{R_w, K_w\} + 1 = |w| - G_w. \tag{3}$$

Lemma 3.1. *Let $w \in A^+$. Then*

$$G_w + 1 \le |w| \le G_w + d^{R_w}.$$

Proof. Since any repeated factor in a non-empty word w has a length smaller than $|w|$, one has $G_w \le |w| - 1$. Now, let λ_w be the subword complexity of the word w. Since $\lambda_w(R_w) \le d^{R_w}$, the result follows by equation (3). □

From the preceding lemma, one derives the following result proved in [8]:

Lemma 3.2. *Let $d > 1$. For any $w \in A^+$ one has*

$$G_w \ge \lfloor \log_d |w| \rfloor - 1.$$

Proof. Since, by equation (2), $R_w \le 1 + G_w$, by the preceding lemma and using $d > 1$, one has $|w| \le G_w + d^{R_w} \le G_w + d^{1+G_w} < d^{2+G_w}$. One derives $2 + G_w > \log_d |w| \ge \lfloor \log_d |w| \rfloor$, so that the result follows. □

We observe that the lower bounds in the previous lemmas are effectively reached (*cf.* Rem. 3.6).

Lemma 3.3. *Let n be a positive integer and $w \in A^n$. One has $G_w \ge n - 2$ if and only if w has one of the following forms:*

$$ab^{n-1}, \quad a^{n-1}b, \quad (ab)^{\lfloor n/2 \rfloor} a^{\delta}, \tag{4}$$

where $a, b \in A$ and δ is equal to 0 or 1, according to the parity of n. Moreover, one has $G_w = n - 2$ if and only if $n \ge 2$ and $a \ne b$.

Proof. If $n = 1$ the statement is trivially true. Thus we assume $n \ge 2$. Let us suppose that $G_w \ge n - 2$. Since w has a repeated factor v of length $n - 2$, there exist letters a, b, x, y such that one of the following three cases occurs:

$$w = vab = xvy, \qquad w = abv = vxy, \qquad w = avb = xyv.$$

Let us consider the first case. From the equation $w = vab = xvy$ one has $b = y$ and $va = xv$. This trivially implies $a = x$ and $v = a^{n-2}$. Therefore, $w = a^{n-1}b$. In a symmetric way one proves that in the third case $w = ab^{n-1}$. In the second case, from a classical result of Lyndon and Schützenberger [13] (see [11]), one derives $w = (ab)^{\lfloor n/2 \rfloor} a^{\delta}$.

Conversely, it is trivial to verify that for any of the words w in (4) one has $G_w \ge n - 2$, where equality holds if and only if $n \ge 2$ and $a \ne b$. □

Lemma 3.4. *Let $w \in A^+$ be a word. For any $x \in A$ one has*

$$G_w \leq G_{wx} \leq 1 + G_w \quad \textit{and} \quad G_w \leq G_{xw} \leq 1 + G_w.$$

Moreover, if $x \in A \setminus B_w$, then $G_w = G_{wx}$.

Proof. By Proposition 2.1, if $x \in B_w$, then $\max\{R_{wx}, K_{wx}\} = \max\{R_w, K_w + 1\}$, so that

$$\max\{R_w, K_w\} \leq \max\{R_{wx}, K_{wx}\} \leq 1 + \max\{R_w, K_w\}\cdot$$

By equation (2), $G_w \leq G_{wx} \leq 1 + G_w$. Now, let us suppose $x \in A \setminus B_w$. By Proposition 2.1, one has $K_{wx} \leq K_w \leq \max\{R_w, K_w\} = R_{wx}$, so that

$$\max\{R_{wx}, K_{wx}\} = R_{wx} = \max\{R_w, K_w\}\cdot$$

By equation (2), it follows $G_w = G_{wx}$.

From Lemma 2.6, for any $w \in A^+$ and any $x \in A$, one has

$$\max\{R_w, K_w\} \leq \max\{R_{xw}, K_{xw}\} \leq \max\{R_w, K_w\} + 1.$$

From equation (2) one derives $G_w \leq G_{xw} \leq 1 + G_w$. □

A word $w \in A^*$ is said to be a *de Bruijn word* (or *full cycle*) of *order* $m > 0$ if any word of A^m occurs exactly once in w. For instance, if $A = \{a, b\}$ and $m = 3$, the word *abaaabbbab* satisfies the previous condition. If $d \geq 2$ and w is a de Bruijn word of order m, then one has

$$R_w = K_w = m \quad \text{and} \quad G_w = m - 1.$$

Indeed, all factors of length m are unrepeated, so that $R_w, K_w \leq m$ and $G_w \leq m - 1$, while all factors of length $m - 1$ are right special and hence repeated, so that $R_w, K_w \geq m$ and $G_w \geq m - 1$.

As is well known (see [9]) the length of a de Bruijn word of order m on a d-letter alphabet is $d^m + m - 1$. In fact, if $d > 1$, since $G_w = m - 1$ and $\lambda_w(m) = d^m$, by equation (3), one derives $|w| = d^m + m - 1$. The case $d = 1$ is trivial.

Moreover, it is well known (see [9]) that for any word $v \in A^m$ the number of de Bruijn words of order m having v as a prefix (or suffix) is given by

$$(d-1)!^{d^{m-1}} d^{d^{m-1}-m}.$$

Thus the total number of de Bruijn words of order m on a d-letter alphabet is given by

$$(d-1)!^{d^{m-1}} d^{d^{m-1}}. \tag{5}$$

Lemma 3.5. *A word w is a de Bruijn word of order m if and only if*

$$|w| = d^m + m - 1 \quad \textit{and} \quad G_w = m - 1.$$

Proof. As we have previously seen, if w is a de Bruijn word of order m, then $|w| = d^m + m - 1$ and $G_w = m - 1$. Conversely, if a word w verifies the relation $G_w = m - 1$, then any factor of w of length m has a unique occurrence in w. If, moreover, the length of w is equal to $d^m + m - 1$, then by equation (3), w has d^m distinct factors of length m, so that all the words of A^m have to be factors of w, *i.e.*, w is a de Bruijn word of order m. □

Remark 3.6. The bounds in Lemmas 3.1 and 3.2 are effectively reached. Indeed, if $d = 1$ trivially $G_w = |w| - 1$ and $R_w = 0$, so that $|w| = G_w + d^{R_w}$.

Let us suppose $d > 1$ and $m \geq 1$. If w is a de Bruijn word of order m, then one has $|w| = d^m + m - 1$, $R_w = K_w = m$, and $G_w = m - 1$, so that $|w| = G_w + d^{R_w}$. Consequently, $\log_d |w| = \log_d(G_w + d^{R_w}) \geq R_w = 1 + G_w$ so that $G_w \leq \lfloor \log_d |w| \rfloor - 1$. Hence, by Lemma 3.2, $G_w = \lfloor \log_d |w| \rfloor - 1$.

For any $m, n \geq 0$ we consider the set of words $S(m, n)$ defined as

$$S(m, n) = \{w \in A^n \mid R_w = K_w = m\}\cdot$$

The words of this set will be called *symmetric words* of order m and length n.

For instance, if $d > 1$ any de Bruijn word of order m is a symmetric word of order m and length $d^m + m - 1$. However, there exist symmetric words such as $w = abbabaab$ which are not de Bruijn words. Indeed, in this case $R_w = K_w = 3$ but $|w| = 8 < 2^3 + 3 - 1 = 10$.

Proposition 3.7. *For any $m, n \geq 0$, one has*

$$S(m, n) \neq \emptyset \quad \textit{if and only if} \quad 2m \leq n \leq d^m + m - 1.$$

Proof. If there exists $w \in S(m, n)$, then by Proposition 2.7 one has

$$n = |w| \geq R_w + K_w = 2m$$

and, by Lemma 3.1, one derives

$$n = |w| \leq G_w + d^{R_w} = m - 1 + d^m.$$

This implies $2m \leq n \leq d^m + m - 1$.

To prove the converse, we suppose that $2m \leq n \leq d^m + m - 1$; we have to show the existence of a symmetric word of order m and length n. If $m = 0$, then $n = 0$ and $S(0, 0) = \{\epsilon\}$. If $d = 1$, then $m = n = 0$ and again $S(0, 0) = \{\epsilon\}$. Thus, we assume that $m > 0$ and $d > 1$. We consider a de Bruijn word of order m ending by a^m, $a \in A$, and its suffix w_0 of length n. We can write

$$w_0 = v_0 a^m, \text{ with } v_0 \in A^{n-m}.$$

Since in w_0 there is no repeated factor of length m one has

$$k'_{w_0} = a^{m-1} \quad \text{and} \quad K_{w_0} = m \geq R_{w_0}.$$

If $R_{w_0} = m$, we are done. Hence, we assume $R_{w_0} < m$. Thus, since k'_{w_0} is not right special, $B_{w_0} = \{a\}$. We select a letter $b \in A \setminus \{a\}$ and consider the sequence of words of length n

$$w_1 = v_1 a^m b, \quad w_2 = v_2 a^m b^2, \quad \ldots, \quad w_m = v_m a^m b^m,$$

where v_i, $1 \leq i \leq m$, is obtained from v_{i-1} by deleting its first letter. Let us observe that by Lemma 2.6 one has, for $1 \leq i \leq m$,

$$K_{w_i} \leq K_{w_{i-1}b} \quad \text{and} \quad R_{w_i} \leq R_{w_{i-1}b}. \tag{6}$$

Moreover, since a^{m-1} is a right special factor of all words w_i, $1 \leq i \leq m$, one obtains

$$R_{w_i} \geq m. \tag{7}$$

Let us denote by t the minimal positive integer such that $K_{w_t} \geq m$. Such an integer exists since clearly $K_{w_m} \geq m$. We shall show that w_t is a symmetric word of order m.

Let us first verify that $K_{w_t} = m$. If $t = 1$, by Proposition 2.1, since $b \notin B_{w_0}$, one has $K_{w_0 b} \leq K_{w_0} = m$ so that, by equation (6), $K_{w_1} \leq m$. Thus, $K_{w_1} = m$. If, on the contrary, $t > 1$, then, by the minimality of t, one has $K_{w_{t-1}} \leq m - 1$ and, by Proposition 2.1, $K_{w_{t-1}b} \leq K_{w_{t-1}} + 1 \leq m$. By equation (6), it follows $K_{w_t} \leq m$ and then $K_{w_t} = m$.

Now, let us verify that $R_{w_t} = m$. By Proposition 2.1, since $b \notin B_{w_0}$, one has $R_{w_0 b} = \max\{R_{w_0}, K_{w_0}\} = m$ and, by equations (6) and (7), it follows $R_{w_1} = m$. Moreover, for $1 \leq i < t$, by equation (7), $R_{w_i} \geq m > K_{w_i}$. Hence, by equation (6) and Lemma 2.3, one has $R_{w_{i+1}} \leq R_{w_i}$, so that

$$R_{w_t} \leq R_{w_1} = m.$$

By equation (7) one has $R_{w_t} \geq m$. Thus, $R_{w_t} = m$, so that w_t is a symmetric word of order m and length n, which concludes the proof. □

Corollary 3.8. *For any pair of positive integers m and n such that $m < n \leq d^m + m - 1$ there exists a word $u \in A^n$ such that $K_u \leq R_u = m$.*

Proof. Let $m < n \leq d^m + m - 1$. This condition implies that $d > 1$. If $n \geq 2m$ the conclusion follows from the preceding proposition. Then, let us suppose $m < n < 2m$. We consider the word $u = a^m b^{n-m}$, with a and b distinct letters. Since a^{m-1} is a right special factor of u one has $R_u = m$. Moreover, $K_u = n - m < m = R_u$, since $n < 2m$. This concludes the proof. □

4. Distributions

We introduce two functions D_R and D_K giving the distributions of the characteristic parameters among the words of any given length. These maps are defined as: for all $i, n \geq 0$,

$$D_R(i,n) = \mathrm{Card}(\{v \in A^n \mid R_v = i\}), \quad D_K(i,n) = \mathrm{Card}(\{v \in A^n \mid K_v = i\})\cdot$$

We remark that the values of D_R and D_K actually depend on the value of $d = \mathrm{Card}(A)$. However, as d is fixed, this dependence will not be explicitly written.

In the case $d = 2$, the values of $D_R(i,n)/2$ and $D_K(i,n)/2$ for $0 < n \leq 20$ and $0 \leq i \leq n$ are reported in Tables 1 and 2, respectively.

We observe that, in a similar way, one can introduce the distributions D_L and D_H. By Lemma 2.9 it follows easily that $D_L = D_R$ and $D_H = D_K$, so that we shall not consider them in the sequel.

Proposition 4.1. *The following relations hold:*

$$D_R(i,n) = D_K(i,n) = 0 \text{ for } i > n \geq 0,$$
$$D_R(n,n) = D_K(0,n) = 0 \text{ for } n > 0,$$
$$D_R(0,0) = D_K(0,0) = 1, \quad D_R(0,n) = D_K(n,n) = d \text{ for } n > 0,$$
$$D_R(n-1,n) = d(d-1) \text{ for } n \geq 2, \quad D_R(0,1) = d,$$
$$D_K(n-1,n) = 2d(d-1) \text{ for } n \geq 3, \quad D_K(1,n) = d(d-1)^{n-1} \text{ for } n > 1,$$
$$\sum_{m=0}^{n} D_K(m,n) = \sum_{m=0}^{n} D_R(m,n) = d^n \text{ for } n \geq 0.$$

Proof. The relations on the first two lines are trivial consequences of equation (1). Since $R_\epsilon = K_\epsilon = 0$, one has $D_R(0,0) = D_K(0,0) = 1$. Moreover, for a word $w \neq \epsilon$ one has $R_w = 0$ or $K_w = |w|$ if and only if w is a power of a letter. Since there are d words of length $n > 0$ which are powers of a letter, one obtains $D_R(0,n) = D_K(n,n) = d$.

Let w be a word of length $n \geq 2$ such that $R_w = n-1$ or $K_w = n-1$. This implies that $G_w \geq n-2$. By Lemma 3.3, w has one of the following forms: ab^{n-1}, $a^{n-1}b$, $(ab)^{\lfloor n/2 \rfloor}a^\delta$, with $a, b \in A$ and $\delta \in \{0,1\}$. If $a = b$, then $w = a^n$, $R_w = 0 < n-1$, and $K_w = n$. Thus $a \neq b$. Now let us suppose that $n \geq 3$. If $w = ab^{n-1}$ or $w = (ab)^{\lfloor n/2 \rfloor}a^\delta$, then $K_w = n-1$ and $R_w = 1$; if, on the contrary, $w = a^{n-1}b$, then $K_w = 1$ and $R_w = n-1$. As there are $d(d-1)$ words of each kind, we conclude that for $n \geq 3$,

$$D_R(n-1,n) = d(d-1) \quad \text{and} \quad D_K(n-1,n) = 2d(d-1).$$

If $n = 2$, then $w = ab$ and $R_w = 1$. The number of such words is again $d(d-1)$. We conclude that for $n \geq 2$, $D_R(n-1,n) = d(d-1)$. All the d letters $a \in A$ are such that $R_a = 0$. Thus $D_R(0,1) = d$.

Table 3. $\frac{1}{2}D_G(i,n),\ 0 \leq i \leq n \leq 20,\ n > 0.$

n \ i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0																			
2	1	1	0																		
3	0	3	1	0																	
4	0	4	3	1	0																
5	0	2	10	3	1	0															
6	0	0	18	10	3	1	0														
7	0	0	21	29	10	3	1	0													
8	0	0	24	61	29	10	3	1	0												
9	0	0	20	111	82	29	10	3	1	0											
10	0	0	8	189	190	82	29	10	3	1	0										
11	0	0	0	279	405	215	82	29	10	3	1	0									
12	0	0	0	359	837	512	215	82	29	10	3	1	0								
13	0	0	0	427	1615	1158	556	215	82	29	10	3	1	0							
14	0	0	0	460	2931	2571	1334	556	215	82	29	10	3	1	0						
15	0	0	0	478	4973	5540	3112	1385	556	215	82	29	10	3	1	0					
16	0	0	0	480	8073	11510	7103	3321	1385	556	215	82	29	10	3	1	0				
17	0	0	0	456	12613	23154	15885	7756	3391	1385	556	215	82	29	10	3	1	0			
18	0	0	0	352	19095	45204	34828	17861	8060	3391	1385	556	215	82	29	10	3	1	0		
19	0	0	0	128	28051	86041	74811	40502	18804	8135	3391	1385	556	215	82	29	10	3	1	0	
20	0	0	0	0	39503	160145	157840	90251	43329	19143	8135	3391	1385	556	215	82	29	10	3	1	0

Let w be a word of length $n > 1$ such that $K_w = 1$. This implies that the last letter of w, say a, does not occur elsewhere in w. Therefore, $w = ua$, with $u \in (A \setminus \{a\})^{n-1}$. Since for any $a \in A$ there are $(d-1)^{n-1}$ such words, it follows that $D_K(1,n) = d(d-1)^{n-1}$.

Finally, the last formula derives from the fact that D_R and D_K are distribution functions among the words of each length. □

Now let us introduce the distribution function D_G of the maximal length of a repeated factor in a word. It is defined as: for all $i \geq 0$ and $n > 0$,

$$D_G(i,n) = \mathrm{Card}(\{v \in A^n \mid G_v = i\})\cdot$$

In the case $d = 2$, the values of $D_G(i,n)/2$ for $n > 0$ and $0 \leq i \leq n \leq 20$ are reported in Table 3.

Proposition 4.2. *The following relations hold:*

$$D_G(n-1,n) = d \quad \textit{for } n \geq 1,$$
$$D_G(n-2,n) = 3d(d-1) \quad \textit{for } n \geq 3,$$
$$D_G(0,2) = d(d-1),$$
$$\sum_{i=0}^{n} D_G(i,n) = d^n \quad \textit{for } n \geq 1.$$

Proof. A word of length n has a repeated factor of length $n-1$ if and only if it is a power of a single letter. This proves that $D_G(n-1,n) = d$.

By Lemma 3.3 a word w of length $n \geq 2$ has a repeated factor of maximal length $n-2$ if and only if it has one of the forms of equation (4), with $a, b \in A$ and $a \neq b$. If $n \geq 3$, these words are all distinct and in number of $3d(d-1)$, so that $D_G(n-2,n) = 3d(d-1)$. If $n = 2$, then the preceding words reduce to the words ab, with $a, b \in A$ and $a \neq b$. The number of these words is then $d(d-1)$. This shows that $D_G(0,2) = d(d-1)$.

The last formula derives from the fact that D_G is a distribution function among the words of each length. □

Proposition 4.3. *Let $i, n > 0$. One has*

$$D_G(i-1,n) \neq 0 \quad \textit{if and only if} \quad i \leq n < i + d^i.$$

Proof. Let $w \in A^n$ be a word such that $G_w = i - 1$. Since $R_w \leq i$, by Lemma 3.1 one has:

$$i \leq n \leq i - 1 + d^i.$$

Hence, if $D_G(i-1,n) \neq 0$, then $i \leq n \leq i - 1 + d^i$. Let us prove now that if $i \leq n \leq i - 1 + d^i$, then $D_G(i-1,n) \neq 0$. If $i < n$, then by Corollary 3.8, there exists a word $u \in A^n$ such that $K_u \leq R_u = i$. Thus, by equation (2), $G_u = i - 1$ and this proves that $D_G(i-1,n) \neq 0$. If $i = n$, then by the previous proposition one has $D_G(i-1,n) = D_G(n-1,n) = d \neq 0$, which concludes the proof. □

Proposition 4.4. *For any $m > 0$, one has*

$$D_G(m-1, d^m + m - 1) = (d-1)!^{d^{m-1}} d^{d^{m-1}}.$$

Proof. By Lemma 3.5, $D_G(m-1, d^m + m - 1)$ is equal to the number of de Bruijn words of order m on a d-letter alphabet so that the result follows by equation (5). □

It is useful to introduce in the sequel the functions $D_R^{\geq}$ and $D_K^{>}$ defined as: for all $i, n \geq 0$,

$$D_R^{\geq}(i,n) = \mathrm{Card}(\{v \in A^n \mid R_v = i \geq K_v\})$$

and

$$D_K^{>}(i,n) = \mathrm{Card}(\{v \in A^n \mid K_v = i > R_v\})\cdot$$

We shall set also, for all $i, n \geq 0$,

$$D_R^{<}(i,n) = D_R(i,n) - D_R^{\geq}(i,n) \quad \text{and} \quad D_K^{\leq}(i,n) = D_K(i,n) - D_K^{>}(i,n).$$

Of course, one has $D_R^{\geq}(i,n) \leq D_R(i,n)$ and $D_K^{>}(i,n) \leq D_K(i,n)$.

Proposition 4.5. *Let $d > 1$, $i \geq 0$, and $n > 0$. One has*

$$D_K^{>}(i,n) = D_K(i,n) \quad \textit{if and only if} \quad i = 0 \textit{ or } i > \frac{n}{2}$$

and

$$D_R^{\geq}(i,n) = D_R(i,n) \quad \textit{if and only if} \quad i \geq \frac{n}{2}\cdot$$

Proof. If $i = 0$, then in view of Proposition 4.1, $D_K(i,n) = 0 = D_K^{>}(i,n)$. If $i > n/2$, then by Corollary 2.8 for any $w \in A^n$ such that $K_w = i$ one has $K_w > R_w$, so that $D_K^{>}(i,n) = D_K(i,n)$. Similarly, if $i \geq n/2$, for any $w \in A^n$ such that $R_w = i$ one has $R_w \geq K_w$, so that $D_R^{\geq}(i,n) = D_R(i,n)$.

Conversely, for $0 < i \leq n/2$, consider the word $v = a^{n-i}b^i$ where a and b are two distinct letters of the alphabet A. One has $|v| = n$, $K_v = i$, and $R_v = n - i \geq i$. This proves that $D_K^{>}(i,n) < D_K(i,n)$. Similarly, for any $i < n/2$, consider the word $w = a^i b^{n-i}$. One has $|w| = n$, $R_w = i$, and $K_w = n - i > i$. This proves that $D_R^{\geq}(i,n) < D_R(i,n)$. □

Lemma 4.6. *For all $i, n > 0$, one has*

$$D_G(i-1,n) = D_R^{\geq}(i,n) + D_K^{>}(i,n).$$

Proof. Let us verify that

$$\begin{aligned}\{w \in A^n \mid G_w = i-1\} = \\ \{w \in A^n \mid R_w = i \geq K_w\} \cup \{w \in A^n \mid K_w = i > R_w\}\cdot\end{aligned} \tag{8}$$

For any element v belonging to the right hand side of the previous equation, one has $G_v = \max\{R_v, K_v\} - 1 = i - 1$, so that $v \in \{w \in A^n \mid G_w = i-1\}$. Conversely, take an element v such that $G_v = i-1$. Thus either $R_v = i \geq K_v$ and $v \in \{w \in A^n \mid R_w = i \geq K_w\}$ or $K_v = i > R_v$ and $v \in \{w \in A^n \mid K_w = i > R_w\}$. This proves equation (8). Since the union in equation (8) is disjoint, one has $D_G(i-1,n) = D_R^{\geq}(i,n) + D_K^{>}(i,n)$. □

From the previous lemma one derives the following theorem which shows that the distribution D_R is determined by D_G and *vice versa*:

Theorem 4.7. *For all $i, n > 0$ one has*

$$D_R(i,n+1) = D_R(i,n) + (d-1)D_G(i-1,n). \tag{9}$$

Proof. Let $w \in A^n$. If $R_w = i \geq K_w$, then by Lemma 2.3 one has that for all $x \in A$, $R_{wx} = R_w$. If $R_w = i < K_w$, then by Lemma 2.4 there exists a unique letter a such that $R_{wa} = R_w$. If $R_w < i = K_w$, then from Lemma 2.4, for $d-1$ letters $x \in A$, $R_{wx} = K_w = i$. In this way, we obtain

$$dD_R^{\geq}(i,n) + D_R^{<}(i,n) + (d-1)D_K^{>}(i,n)$$

distinct words $v \in A^{n+1}$ such that $R_v = i$. Now, let us prove that these are the only words v of length $n+1$ such that $R_v = i$. Indeed, let $v \in A^{n+1}$ with $R_v = i$ and write $v = wa$ with $w \in A^n$ and $a \in A$. By Proposition 2.1, either $R_w = i$ or $K_w = i > R_w$. This ensures that any word $v \in A^{n+1}$ such that $R_v = i$ can be obtained by extending on the right a word $w \in A^n$ such that $R_w = i \geq K_w$ or

$R_w = i < K_w$ or $R_w < i = K_w$. Thus,

$$\begin{aligned} D_R(i, n+1) &= dD_R^{\geq}(i,n) + D_R^{<}(i,n) + (d-1)D_K^{>}(i,n) \\ &= dD_R^{\geq}(i,n) + D_R(i,n) - D_R^{\geq}(i,n) + (d-1)D_K^{>}(i,n) \\ &= D_R(i,n) + (d-1)(D_R^{\geq}(i,n) + D_K^{>}(i,n)), \end{aligned}$$

and the conclusion follows from Lemma 4.6. □

In the sequel, we follow the convention that a sum $\sum_{i=t}^{s} a_i$ holds 0 if $t > s$.

Corollary 4.8. *For all $i, n > 0$ one has*

$$D_R(i,n) = (d-1)\sum_{m=i}^{n-1} D_G(i-1, m).$$

Proof. If $n \leq i$, then by Proposition 4.1, $D_R(i,n) = 0$ and the result follows from our convention on the sums. Let us then suppose $n > i$. By iteration of equation (9) one has

$$D_R(i,n) = D_R(i,i) + (d-1)\sum_{m=i}^{n-1} D_G(i-1, m).$$

Since $D_R(i,i) = 0$, the result follows: □

Proposition 4.9. *Let $i, n > 0$. One has*

$$D_G(i-1, n) \leq D_R(i,n) + D_K(i,n)$$

where equality holds if and only if $d = 1$ or $i > n/2$.

Proof. The statement is trivially true if $d = 1$. Thus we suppose $d > 1$. If $i > n/2$, then by Lemma 4.6 and Proposition 4.5, one has

$$D_G(i-1,n) = D_R^{\geq}(i,n) + D_K^{>}(i,n) = D_R(i,n) + D_K(i,n).$$

Conversely, if $i \leq n/2$, by Proposition 4.5, $D_K^{>}(i,n) < D_K(i,n)$. Since $D_R^{\geq}(i,n) \leq D_R(i,n)$, by Lemma 4.6 one has

$$D_G(i-1,n) = D_R^{\geq}(i,n) + D_K^{>}(i,n) < D_R(i,n) + D_K(i,n)$$

and this proves our assertion. □

The following noteworthy relation among $D_R(i,n)$ and $D_K(i,n)$ holds when $i \geq n/2$.

Proposition 4.10. *For any $n > 0$ and any $i \geq n/2$, one has*

$$D_R(i,n) = dD_R(i,n-1) + (d-1)D_K(i,n-1).$$

Proof. By Theorem 4.7, one has

$$D_R(i,n) = D_R(i,n-1) + (d-1)D_G(i-1,n-1).$$

Since $i > (n-1)/2$, by Proposition 4.9, $D_G(i-1,n-1) = D_R(i,n-1) + D_K(i,n-1)$ so that $D_R(i,n) = dD_R(i,n-1) + (d-1)D_K(i,n-1)$. □

Proposition 4.11. *For all $i,n \geq 0$ one has*

$$D_R(i,n+1) \geq D_R(i,n),$$

where equality holds if and only if $d = 1$ or $n \geq d^i + i$ or $n < i$.

Proof. If $i = 0$ or $n = 0$ the result follows from Proposition 4.1. Let us then suppose $i,n > 0$. From equation (9) one has immediately $D_R(i,n) \leq D_R(i,n+1)$. Moreover equality holds if and only if

$$(d-1)D_G(i-1,n) = 0.$$

In view of Proposition 4.3 this occurs if and only if $d = 1$ or $n \geq d^i + i$ or $n < i$. □

Proposition 4.12. *For any integers $i \geq 0$ and $n > 0$ one has*

$$D_K(i,n+1) \geq (d-1)D_K(i,n).$$

Moreover, if $d > 1$, $i > 1$, and $n \geq i-1$, then

$$D_K(i,n+1) > (d-1)D_K(i,n).$$

Proof. Since by Proposition 4.1, $D_K(0,n) = D_K(0,n+1) = 0$, the statement is true in the case $i = 0$.

Let us suppose $i > 0$. Let $w \in A^n$ be such that $K_w = i$. Let us set $k_w = ak'_w$, with $a \in A$. For any $b \in A \setminus \{a\}$ we consider the word $v = bw$. By Lemma 2.6 one has $K_v \geq K_w = i$ and, since k_w is not repeated in v, it follows $K_v = K_w = i$. In this way one can obtain $(d-1)D_K(i,n)$ words of length $n+1$ with a minimal unrepeated suffix of length i. Hence,

$$D_K(i,n+1) \geq (d-1)D_K(i,n).$$

Now, let us suppose that $d > 1$, $i > 1$, and $n \geq i-1$. If $n > i$, we consider the word $u = ba^{n-i}b^i$, with a and b distinct letters. One has that $u \in A^{n+1}$ and $K_u = i$. Moreover, this word cannot be obtained by the previous procedure because its first letter is equal to the first letter of k_u. This proves that, in this case, $D_K(i,n+1) > (d-1)D_K(i,n)$. If $n = i$, then, as $n \geq 2$, by Proposition 4.1 one has $D_K(n,n+1) = 2d(d-1)$ and $D_K(n,n) = d$, so that $D_K(n,n+1) > (d-1)D_K(n,n)$. Finally, for $n = i-1$, the result follows since $D_K(n+1,n+1) = d$ and $D_K(n+1,n) = 0$. □

Now let us introduce the following function: for $i \geq 0$ and $n > 0$,

$$D_G^*(i,n) = \sum_{j \geq i} D_G(j,n) = \mathrm{Card}(\{w \in A^n \mid G_w \geq i\}\cdot$$

In other terms, $D_G^*(i,n)$ is the number of words of length n having at least one repeated factor of length i. From the definition, one has that for $i \geq 0$ and $n > 0$

$$D_G^*(i,n) = D_G(i,n) + D_G^*(i+1,n). \tag{10}$$

The following holds:

Theorem 4.13. *For $i \geq 0$ and $n > 1$ one has*

$$D_G^*(i,n) = dD_G^*(i,n-1) + D_K^>(i+1,n). \tag{11}$$

Proof. We shall prove that

$$\begin{aligned}\{v \in A^n \mid G_v \geq i\} = \\ \{w \in A^{n-1} \mid G_w \geq i\}A \cup \{v \in A^n \mid K_v = i+1 > R_v\}\cdot\end{aligned} \tag{12}$$

Moreover, the union will be disjoint. First of all, let us prove the inclusion "$\supseteq$". Indeed, if $w \in A^{n-1}$ and $G_w \geq i$, then, for any $x \in A$, one has $G_{wx} \geq i$ and if $v \in A^n$ and $K_v = i+1 > R_v$, then $G_v = i$.

Now, let us prove the inclusion "$\subseteq$". Let $v \in A^n$ be a word such that $G_v \geq i$. We can write v as $v = wx$, with $w \in A^{n-1}$ and $x \in A$. Now either $G_w \geq i$, and in this case, $v \in \{w \in A^{n-1} \mid G_w \geq i\}A$, or $G_w < i$. In this latter case, since $G_v \geq i$, v has a repeated suffix of length i, so that $K_v \geq i+1$. However, by Proposition 2.1, $K_v = K_{wx} \leq K_w+1 \leq G_w+2 \leq i+1$ and $R_v = R_{wx} \leq \max\{R_w, K_w\} = G_w+1 < i+1$. Hence, $K_v = i+1 > R_v$. This proves equation (12).

Let us prove now that the union in equation (12) is disjoint. Indeed, suppose that $v = wx$, $w \in A^{n-1}$, $x \in A$, is a word of A^n such that $G_w \geq i$ and $K_v = i+1 > R_v$. By Proposition 2.1, two cases may occur:

Case 1. $K_v = K_w + 1$, $R_v = R_w$. In this case, $K_w = i$, $R_w \leq i$. Thus $G_w = \max\{R_w, K_w\} - 1 < i$, which is a contradiction.

Case 2. $R_v = \max\{R_w, K_w\}$. In this case, one has $R_v = G_w + 1 \geq i+1$, which is a contradiction.

Since $\mathrm{Card}(A) = d$, by equation (12) and the fact that in that equation union is disjoint, the result follows: □

We observe that from the previous theorem one has that for $i \geq 0$ and $n > 1$, $D_G^*(i,n) \geq dD_G^*(i,n-1)$. If $n > i$, by iterating this relation one obtains $D_G^*(i,n) \geq d^{n-i-1}D_G^*(i,i+1)$. Since $D_G^*(i,i+1) = D_G(i,i+1) = d$, it follows

$$D_G^*(i,n) \geq d^{n-i}.$$

Corollary 4.14. *For $i \geq 0$ and $n > 0$ one has*

$$D_G^*(i,n) = \sum_{m=0}^{n-i-1} d^m D_K^{>}(i+1, n-m).$$

Proof. If $i = 0$, one has $D_G^*(0,n) = d^n = \sum_{m=0}^{n-1} d^m D_K^{>}(1, n-m)$ since, as one easily verifies, $D_K^{>}(1,1) = d$ and $D_K^{>}(1,j) = 0$ for $j > 1$. If $i \geq n$, the result is trivial. Thus we consider the case $0 < i < n$. By iterating equation (11) and taking into account that $D_G^*(i,i) = 0$, the result follows: □

Corollary 4.15. *For $i \geq 0$ and $n > 1$ one has*

$$D_G^*(i+1,n) = dD_G^*(i,n-1) - D_R^{\geq}(i+1,n).$$

Proof. From equation (10) and Lemma 4.6 one has

$$D_G^*(i,n) = D_G(i,n) + D_G^*(i+1,n) = D_R^{\geq}(i+1,n) + D_K^{>}(i+1,n) + D_G^*(i+1,n).$$

By Theorem 4.13, the result follows: □

Corollary 4.16. *For $i \geq 0$ and $n > 0$ one has*

$$D_G(i,n) = \sum_{m=0}^{n-i-1} d^m (D_K^{>}(i+1,n-m) - D_K^{>}(i+2,n-m)).$$

Proof. Since $D_G(i,n) = D_G^*(i,n) - D_G^*(i+1,n)$ and $D_K^{>}(i+2,i+1) = 0$, the statement follows easily from Corollary 4.14. □

By the previous corollary one derives the following iterative formula for D_G:

$$D_G(i,n+1) = dD_G(i,n) + D_K^{>}(i+1,n+1) - D_K^{>}(i+2,n+1).$$

Now, we introduce the map D_S defined for all $i, n \geq 0$ as

$$D_S(i,n) = \sum_{w \in S(i,n)} \mathrm{Card}(B_w),$$

where $S(i,n)$ denotes the set of symmetric words of order i and length n (*cf.* Sect. 3). In the case $d = 2$, the values of $D_S(i,n)/2$ for $0 \leq i \leq n \leq 20$ are reported in Table 4.

Theorem 4.17. *For any $i, n \geq 0$ one has*

$$D_K^{>}(i+1,n+1) = D_K^{>}(i,n) + D_S(i,n).$$

TABLE 4. $\frac{1}{2}D_S(i,n),\ 0 \le i \le n \le 20.$

n \ i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	1																				
1	0	0																			
2	0	2	0																		
3	0	0	0	0																	
4	0	0	6	0	0																
5	0	0	4	0	0	0															
6	0	0	0	11	0	0	0														
7	0	0	0	18	0	0	0	0													
8	0	0	0	28	21	0	0	0	0												
9	0	0	0	32	37	0	0	0	0	0											
10	0	0	0	16	83	25	0	0	0	0	0										
11	0	0	0	0	183	315	57	0	0	0	0	0									
12	0	0	0	0	291	151	44	0	0	0	0	0	0								
13	0	0	0	0	394	402	88	0	0	0	0	0	0	0							
14	0	0	0	0	442	937	240	51	0	0	0	0	0	0	0						
15	0	0	0	0	476	1907	588	107	0	0	0	0	0	0	0	0					
16	0	0	0	0	504	3561	1554	305	70	0	0	0	0	0	0	0	0				
17	0	0	0	0	560	6187	3758	834	164	0	0	0	0	0	0	0	0	0			
18	0	0	0	0	576	10155	8391	2132	410	75	0	0	0	0	0	0	0	0	0		
19	0	0	0	0	256	16279	18077	5492	1130	189	0	0	0	0	0	0	0	0	0	0	
20	0	0	0	0	0	25174	37357	13472	2862	501	118	0	0	0	0	0	0	0	0	0	0

Proof. Let us first verify that

$$D_K^>(i+1,n+1) = \sum_{w\in\{v\in A^n \mid K_v = i \ge R_v\}} \mathrm{Card}(B_w). \tag{13}$$

By Proposition 2.1, for any $w \in \{v \in A^n \mid K_v = i \ge R_v\}$ there are exactly $\mathrm{Card}(B_w)$ letters x which extend w in a word wx of length $n+1$ such that $K_{wx} = i+1 > R_{wx} = R_w$. To complete the proof of equation (13) we have to show that any word $v \in A^{n+1}$ such that $K_v = i+1 > R_v$ can be obtained by extending on the right a word of the set $\{v \in A^n \mid K_v = i \ge R_v\}$. In other terms, we prove that, for any $w \in A^n$ and $x \in A$, if $K_{wx} = i+1 > R_{wx}$, then $K_w = i \ge R_w$. Indeed, by Proposition 2.1, one has either

$$K_w = K_{wx} - 1 = i \quad \text{and} \quad R_w = R_{wx} \le i,$$

or

$$K_{wx} \le K_w \le R_{wx}.$$

Since this latter case gives a contradiction, equation (13) is proved.

We notice that if w is a word such that $K_w = i > R_w$, then by Lemma 2.4 $\mathrm{Card}(B_w) = 1$. Thus, we can rewrite the right hand side of equation (13) as

$$D_K^>(i,n) + \sum_{w\in S(i,n)} \mathrm{Card}(B_w),$$

which concludes the proof. □

Corollary 4.18. *For any i, n such that $n \ge i \ge 0$ one has*

$$D_K^>(i,n) = \sum_{m=1}^{i} D_S(i-m, n-m).$$

Proof. The proof is obtained by iteration from Theorem 4.17, taking into account that $D_K^{>}(0, n-i) = 0$. □

We recall that a word w is called *periodic-like* [3] if k'_w (or h'_w) has no internal occurrence in w. For instance, the word $w = abccbccab$ is periodic-like since $k'_w = ab$ has no internal occurrence in w.

Let P be the set of periodic-like words of A^*. We introduce the map D_P defined for all $i, n \geq 0$ as

$$D_P(i, n) = \mathrm{Card}(\{w \in P \cap A^n \mid K_w = i\})\cdot$$

In other terms, $D_P(i, n)$ gives the number of periodic-like words of length n having the shortest unrepeated suffix of length i. Since the minimal period of a periodic-like word w is equal to $|w| - K_w + 1$ [3], $D_P(i, n)$ gives the number of periodic-like words of length n and minimal period $n - i + 1$.

The following theorem shows that the distribution D_K is determined by D_P and *vice versa*:

Theorem 4.19. *For all* $i, n \geq 0$ *one has*

$$D_K(i, n+1) = dD_K(i, n) + D_P(i, n+1) - D_P(i+1, n+1).$$

Proof. We prove that

$$A\{v \in A^n \mid K_v = i\} = \\ \{w \in A^{n+1} \setminus P \mid K_w = i\} \cup \{w \in P \cap A^{n+1} \mid K_w = i+1\}\cdot \tag{14}$$

Indeed, let $v \in A^n$ be such that $K_v = i$ and x be a letter. By Lemma 2.6 either $K_{xv} = i$ or $K_{xv} = i + 1$. In the first case, $k'_{xv} = k'_v$, so that k'_{xv} is repeated in v and, consequently, it has an internal occurrence in xv. Thus, $xv \in A^{n+1} \setminus P$. In the second case, $k'_{xv} = k_v$, so that k'_{xv} is unrepeated in v and, consequently, it has no internal occurrence in xv. Thus, $xv \in P \cap A^{n+1}$. This proves the inclusion "$\subseteq$".

Conversely, suppose that $w \in \{u \in A^{n+1} \setminus P \mid K_u = i\} \cup \{u \in P \cap A^{n+1} \mid K_u = i + 1\}$ and write $w = xv$, with $x \in A$. By the previous argument, either $w \notin P$ and $K_w = K_v$ or $w \in P$ and $K_w = K_v + 1$. In both cases, one gets $K_v = i$ and therefore $w \in A\{v \in A^n \mid K_v = i\}$. This proves the inclusion "$\supseteq$".

By equation (14), since the union is disjoint, one derives

$$dD_K(i, n) = D_K(i, n+1) - D_P(i, n+1) + D_P(i+1, n+1),$$

from which the result follows. □

In this section we have considered several functions related to the structure of words of each length on a given alphabet, such as

$$D_R,\ D_G,\ D_G^*,\ D_K^{>},\ D_R^{\geq},\ \text{and } D_S.$$

It is worth noting that, as a consequence of the previous theorems and propositions, the values of any of these functions can be easily computed by knowing the values of only one of them, *e.g.*, D_S. Moreover, D_K is determined by D_P and *vice versa*. Therefore, all distributions and related functions depend on the class of symmetric words and on the class of periodic-like words, which are 'narrow' subclasses of the class of all words (*cf.* [4], Sect. 3).

5. Diagonal behaviour

In this section, we shall confine ourselves to consider only the case where $d = \mathrm{Card}(A) > 1$ even though some results hold true even for $d = 1$.

We study the "diagonal behaviour" of D_R, D_K, and D_G, *i.e.*, the behaviour, for any fixed $m \geq 0$ of the functions $D_R(t, m+t)$, $D_K(t, m+t)$, and $D_G(t, m+t)$ with respect to the variable t. We show that, for any $i, n \geq 0$, one has $D_K(i, n) \leq D_K(i+1, n+1)$, where equality holds if and only if $i > n/2$. In other terms, the values of D_K on the points of a diagonal line $(t, m+t)_{t\geq 0}$ are initially increasing and ultimately constant. Similar properties hold for D^*_G and $D^>_K$. Moreover, one has

$$D_R(t, m+t) \leq D_R(m, 2m) \quad \text{and} \quad D_G(t, m+t) \leq D_G(m, 2m),$$

where the "=" sign holds in the first equation if and only if $t \geq m$ and in the second one if and only if $t \geq m-1$.

Proposition 5.1. *For any $i, n \geq 0$ one has*

$$D_K(i+1, n+1) \geq D_K(i, n),$$

where equality holds if and only if $i > n/2$.

Proof. For $i > n/2$ equality holds. Indeed, from Propositions 4.5 one has that $D^>_K(i, n) = D_K(i, n)$ and $D^>_K(i+1, n+1) = D_K(i+1, n+1)$. Moreover, by Proposition 3.7, $S(i, n) = \emptyset$, so that the result follows from Theorem 4.17.

Now, suppose that $i \leq n/2$. In the case $i = 0$, from Proposition 4.1 one derives that $D_K(0, n) < D_K(1, n+1)$. Thus, we suppose $i > 0$. Let w be a word of A^n such that $K_w = i$. From Proposition 2.1, since $B_w \neq \emptyset$, there exists at least one letter $x \in A$ such that $K_{wx} = K_w + 1 = i + 1$. This proves that

$$\mathrm{Card}(\{v \in A^n \mid K_v = i\}) \leq \mathrm{Card}(\{v \in A^{n+1} \mid K_v = i+1\}),$$

i.e., $D_K(i, n) \leq D_K(i+1, n+1)$. In order to prove that inequality is strict, it suffices to show that there exists at least one word $w \in A^n$ such that $K_w = i$ and $\mathrm{Card}(B_w) > 1$; indeed, this implies the existence of at least two right extensions of w in the set $\{v \in A^{n+1} \mid K_v = i+1\}$. In fact, take the word $w = a^{n-i}ba^{i-1}$, where a and b are two distinct letters of A. In such a case, since $n - i > i - 1$, one has $k'_w = a^{i-1}$ and $B_w = \{a, b\}$. □

The following corollary shows that for $n/2 < i \leq n$, $D_K(i,n)$ depends only on the difference $n - i$:

Corollary 5.2. *For any integers i and n such that $n \geq i \geq 0$ one has*

$$D_K(i,n) \leq D_K(n-i+1, 2(n-i)+1),$$

where equality holds if and only if $i > n/2$.

Proof. Let us suppose first $i > n/2$. For any $t > 0$ one has

$$n - i + t > \frac{2(n-i)+t}{2}\cdot$$

Hence, since $i > n/2$, by an iterated application of Proposition 5.1 it follows

$$D_K(n-i+1, 2(n-i)+1) = D_K(n-i+2, 2(n-i)+2) = \cdots = D_K(i,n).$$

Now let us suppose that $i \leq n/2$. Then by Proposition 5.1 one has

$$\begin{aligned} D_K(i,n) &< D_K(i+1,n+1) \leq D_K(i+2,n+2) \\ &\leq \cdots \leq D_K(n-i+1, 2(n-i)+1), \end{aligned}$$

which proves our assertion. □

Proposition 5.3. *For any $i, n \geq 0$ one has*

$$D_K^{>}(i+1,n+1) \geq D_K^{>}(i,n),$$

where inequality is strict if and only if $2i \leq n \leq d^i + i - 1$.

Proof. From Proposition 3.7 for any $i, n \geq 0$ one has $S(i,n) \neq \emptyset$ if and only if $2i \leq n \leq d^i + i - 1$. Since, for any $w \in A^*$, $\mathrm{Card}(B_w) > 0$ the result follows from Theorem 4.17. □

Proposition 5.4. *For any $i \geq 0$ and $n > 0$ one has*

$$D_G^*(i+1,n+1) = D_G^*(i,n) + \sum_{m=0}^{n-i-1} d^m D_S(i+1, n-m).$$

Proof. By Corollary 4.14 and Theorem 4.17 one has

$$\begin{aligned} D^*_G(i+1,n+1) &= \sum_{m=0}^{n-i-1} d^m D^{>}_K(i+2,n+1-m) \\ &= \sum_{m=0}^{n-i-1} d^m D^{>}_K(i+1,n-m) + \sum_{m=0}^{n-i-1} d^m D_S(i+1,n-m) \\ &= D^*_G(i,n) + \sum_{m=0}^{n-i-1} d^m D_S(i+1,n-m). \end{aligned}$$

□

Proposition 5.5. *For any $i \geq 0$ and $n > 0$ one has*

$$D^*_G(i+1,n+1) \geq D^*_G(i,n),$$

where equality holds if and only if $i \geq \lfloor n/2 \rfloor$.

Proof. If $i \geq \lfloor n/2 \rfloor$, then $n < 2i+2$. Thus, by Proposition 3.7 one has that for $0 \leq m \leq n-i-1$, $S(i+1,n-m) = \emptyset$. Therefore, from Proposition 5.4, one has $D^*_G(i+1,n+1) = D^*_G(i,n)$.

If, on the contrary, $i < \lfloor n/2 \rfloor$, then $n \geq 2i+2$. Thus, by taking $m = n-2i-2$, by Proposition 3.7 one has $S(i+1,n-m) = S(i+1,2i+2) \neq \emptyset$. Since for any $w \in A^*$, $\mathrm{Card}(B_w) > 0$, from Proposition 5.4 one derives $D^*_G(i+1,n+1) > D^*_G(i,n)$. □

Now we introduce the functions D^*_R and D^*_K defined as follows: for $i, n \geq 0$,

$$D^*_R(i,n) = \sum_{m \geq i} D_R(m,n) = \mathrm{Card}(\{w \in A^n \mid R_w \geq i\}$$

and

$$D^*_K(i,n) = \sum_{m \geq i} D_K(m,n) = \mathrm{Card}(\{w \in A^n \mid K_w \geq i\}\cdot$$

In other terms, for $i > 0$, $D^*_R(i,n)$ is the number of words of length n having at least one special factor of length $i-1$ and $D^*_K(i,n)$ is the number of words of length n having a repeated suffix of length $i-1$.

Proposition 5.6. *For $i \geq 0$ and $n > 0$ one has*

$$D^*_R(i+1,n+1) \geq D^*_R(i,n),$$

where equality holds if and only if $i \geq n/2$ or in the case where $i = 0$, $n = 1$, and $d = 2$. Moreover, for $i, n \geq 0$ one has

$$D^*_K(i+1,n+1) \geq D^*_K(i,n),$$

where equality holds if and only if $i > n/2$.

Proof. If $i = 0$, then $D^*_R(0, n) = d^n$ and $D^*_R(1, n+1) = d^{n+1} - d \geq d^n$. Moreover, $d^{n+1} - d = d^n$ if and only if $n = 1$ and $d = 2$. Let us now suppose that $i > 0$. By Corollary 4.8 one has

$$D_R(i, n) = (d-1) \sum_{m=i}^{n-1} D_G(i-1, m),$$

so that

$$\begin{aligned} D^*_R(i, n) &= \sum_{j=i}^{n-1} D_R(j, n) = (d-1) \sum_{j=i}^{n-1} \sum_{m=j}^{n-1} D_G(j-1, m) \\ &= (d-1) \sum_{m=i}^{n-1} D^*_G(i-1, m). \end{aligned}$$

By replacing i and n by $i+1$ and $n+1$, respectively, one has

$$D^*_R(i+1, n+1) = (d-1) \sum_{m=i+1}^{n} D^*_G(i, m) = (d-1) \sum_{m=i}^{n-1} D^*_G(i, m+1).$$

By Proposition 5.5, for $i \leq m \leq n-1$ one has $D^*_G(i-1, m) \leq D^*_G(i, m+1)$. Moreover, the "=" sign holds in all these relations if and only if $i - 1 \geq \lfloor (n-1)/2 \rfloor$ or, equivalently, if and only if $i \geq n/2$.

To prove the second inequality, observe that

$$D^*_K(i, n) = \sum_{j=i}^{n} D_K(j, n) \quad \text{and} \quad D^*_K(i+1, n+1) = \sum_{j=i}^{n} D_K(j+1, n+1).$$

By Proposition 5.1, for $i \leq j \leq n$, one has $D_K(j, n) \leq D_K(j+1, n+1)$. Moreover, the "=" sign holds in all these relations if and only if $i > n/2$. One concludes that $D^*_K(i, n) \leq D^*_K(i+1, n+1)$, where equality holds if and only if $i > n/2$. □

Proposition 5.7. *For any* $i, n \geq 0$ *one has*

$$D^*_R(i+1, n+1) \leq d D^*_R(i, n) \quad \textit{and} \quad D^*_K(i+1, n+1) \leq d D^*_K(i, n).$$

For any $i \geq 0$ *and* $n > 0$ *one has*

$$D^*_G(i+1, n+1) \leq d D^*_G(i, n).$$

Proof. Let v be a word of A^{n+1} such that $R_v \geq i+1$. We can write $v = xw$ with $x \in A$ and $w \in A^n$. By Lemma 2.6 one has $R_w \geq R_v - 1 \geq i$. This proves that

$$\{v \in A^{n+1} \mid R_v \geq i+1\} \subseteq A\{w \in A^n \mid R_w \geq i\}\cdot$$

Therefore,

$$D_R^*(i+1,n+1) \le dD_R^*(i,n).$$

In a similar way one proves that $D_K^*(i+1,n+1) \le dD_K^*(i,n)$. By Corollary 4.15, it follows that $D_G^*(i+1,n+1) \le dD_G^*(i,n)$. □

Proposition 5.8. *For any $i > 0$ and $n > 1$, one has*

$$D_R(i,n) \le (d-1)D_G^*(i-1,n-1),$$

where equality holds if and only if $i \ge \lfloor n/2 \rfloor$.

Proof. By Corollary 4.8 one has

$$D_R(i,n) = (d-1)\sum_{m=i}^{n-1} D_G(i-1,m).$$

By equation (10), for $i \le m \le n-1$, one has

$$D_G(i-1,m) = D_G^*(i-1,m) - D_G^*(i,m).$$

Hence, we can write

$$\begin{aligned}\sum_{m=i}^{n-1} D_G(i-1,m) &= \sum_{m=i}^{n-1}(D_G^*(i-1,m) - D_G^*(i,m)) \\ &= D_G^*(i-1,n-1) - D_G^*(i,i) \\ &\quad + \sum_{m=i+1}^{n-1}(D_G^*(i-1,m-1) - D_G^*(i,m)).\end{aligned}$$

By Proposition 5.5, one has $D_G^*(i,m) \ge D_G^*(i-1,m-1)$. Moreover, if $i \ge \lfloor n/2 \rfloor$, then, for $i+1 \le m \le n-1$ one has $i-1 \ge \lfloor (m-1)/2 \rfloor$, so that by Proposition 5.5, $D_G^*(i,m) = D_G^*(i-1,m-1)$. On the contrary, if $i < \lfloor n/2 \rfloor$, then $D_G^*(i,n-1) > D_G^*(i-1,n-2)$. Since $D_G^*(i,i) = 0$, we have shown that

$$\sum_{m=i}^{n-1} D_G(i-1,m) \le D_G^*(i-1,n-1),$$

where equality holds if and only if $i \ge \lfloor n/2 \rfloor$. From this the assertion follows. □

From the preceding proposition and Corollary 4.8 one easily derives the following noteworthy proposition:

Proposition 5.9. *For any $i > 0$ and $n > 1$, one has*

$$\sum_{m=i}^{n-1} D_G(i-1,m) \le \sum_{m=i}^{n-1} D_G(m-1,n-1).$$

Proposition 5.10. *For any integers i and n such that $n \geq i \geq 0$ one has*

$$D_R(i,n) \leq D_R(n-i, 2(n-i)),$$

where equality holds if and only if $i \geq n/2$.

Proof. Let us suppose first $i \geq n/2$. By Proposition 5.8,

$$D_R(i,n) = (d-1)D_G^*(i-1, n-1)$$

and

$$D_R(n-i, 2(n-i)) = (d-1)D_G^*(n-i-1, 2(n-i)-1).$$

Since $n-i-1 \leq i-1$ and for $-1 \leq p \leq 2i-n-2$ one has $n-i+p \geq \lfloor (2(n-i)+p)/2 \rfloor$, by an iterated application of Proposition 5.5, one obtains

$$D_G^*(n-i-1, 2(n-i)-1) = D_G^*(n-i, 2(n-i)) = \cdots = D_G^*(i-1, n-1).$$

One derives

$$D_R(i,n) = D_R(n-i, 2(n-i)).$$

Now, let us suppose $i < n/2$. By Proposition 5.8,

$$D_R(i,n) \leq (d-1)D_G^*(i-1, n-1)$$

and

$$D_R(n-i, 2(n-i)) = (d-1)D_G^*(n-i-1, 2(n-i)-1).$$

Since $i-1 < n-i-1$ and $i-1 < \lfloor (n-1)/2 \rfloor$, one has $D_G^*(i-1, n-1) < D_G^*(i,n)$ and, by an iterated application of Proposition 5.5, $D_G^*(i,n) \leq D_G^*(n-i-1, 2(n-i)-1)$. It follows $D_R(i,n) < D_R(n-i, 2(n-i))$. □

Proposition 5.11. *For $0 \leq i < n$, one has*

$$D_G(i,n) \leq D_G(n-i-1, 2(n-i)-1),$$

where equality holds if and only if $i \geq \lfloor n/2 \rfloor$.

Proof. By Proposition 4.9 one has

$$D_G(i,n) \leq D_R(i+1, n) + D_K(i+1, n),$$

where equality holds if and only if $i+1 > n/2$ or, equivalently, $i \geq \lfloor n/2 \rfloor$. By Proposition 5.10 one has

$$D_R(i+1, n) \leq D_R(n-i-1, 2(n-i-1)) = D_R(n-i, 2(n-i)-1),$$

with equality if and only if $i+1 \geq n/2$ and by Corollary 5.2 one has

$$D_K(i+1,n) \leq D_K(n-i, 2(n-i)-1)$$

with equality if and only if $i \geq \lfloor n/2 \rfloor$. One derives, in view of Proposition 4.9,

$$\begin{aligned} D_G(i,n) &\leq D_R(n-i, 2(n-i)-1) + D_K(n-i, 2(n-i)-1) \\ &= D_G(n-i-1, 2(n-i)-1). \end{aligned}$$

Moreover, if $i \geq \lfloor n/2 \rfloor$ the equality holds, while if $i < \lfloor n/2 \rfloor$ the inequality is strict since $D_K(i+1,n) < D_K(n-i, 2(n-i)-1)$. □

By Proposition 4.10 and Corollary 5.2 one derives the following proposition, whose proof we omit for the sake of brevity:

Proposition 5.12. *For any $n > 0$ and any $i \geq n/2$, one has*

$$D_R(i,n) = (d-1)d^{n-i} \sum_{t=1}^{n-i} d^{-t} D_K(t, 2t-1).$$

As we have previously seen, when $i > n/2$ the functions D_R, D_K, and D_G are constant on the "diagonals", *i.e.*, the values of $D_R(i,n)$, $D_K(i,n)$, and $D_G(i,n)$ depend uniquely on the difference $n-i$. More precisely, from Propositions 5.10, 5.1, and 5.11 for any $n \geq 0$ and any $i \geq n/2$ one has

$$D_R(i,n) = D_R(i+1, n+1),$$

for any $n \geq 0$ and any $i > n/2$ one has

$$D_K(i,n) = D_K(i+1, n+1),$$

and for any $n > 1$ and any $i \geq \lfloor n/2 \rfloor$ one has

$$D_G(i,n) = D_G(i+1, n+1).$$

We have also shown (*cf.* Prop. 5.1) that D_K, as well as other functions like D_G^* and D_R^*, satisfies the stronger diagonal property: for all $i, n \geq 0$

$$D_K(i,n) \leq D_K(i+1, n+1).$$

In other words the value of the function on the pair (i,n) is less than or equal to the value of the function on the pair $(i+1, n+1)$. By using a computer, we verified that in the case $d = 2$ and $1 \leq n \leq 25$ (see Tab. 3 for $n \leq 20$) one has

$$D_G(i,n) \leq D_G(i+1, n+1).$$

We conjecture that this property is true for all d and n. We remark that if the conjecture is true, then, by using Corollary 4.8, one can easily derive that the same property is satisfied by D_R.

Acknowledgements. The authors are thankful to Istituto di Cibernetica "E. Caianiello" del C.N.R. for its valuable support to this research.

References

[1] A. Carpi and A. de Luca, Words and special factors. *Theoret. Comput. Sci.* **259** (2001) 145-182.

[2] A. Carpi and A. de Luca, Semiperiodic words and root-conjugacy. *Theoret. Comput. Sci.* (to appear).

[3] A. Carpi and A. de Luca, Periodic-like words, periodicity, and boxes. *Acta Informatica* **37** (2001) 597-618.

[4] A. Carpi and A. de Luca, On the distribution of characteristic parameters of words II. *RAIRO: Theoret. Informatics Appl.* **36** (2002) 97-127.

[5] A. Carpi, A. de Luca and S. Varricchio, Words, univalent factors, and boxes. *Acta Informatica* **38** (2002) 409-436.

[6] J. Cassaigne, Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc.* **4** (1997) 67-88.

[7] A. Colosimo and A. de Luca, Special factors in biological strings. *J. Theor. Biol.* **204** (2000) 29-46.

[8] A. de Luca, On the combinatorics of finite words. *Theoret. Comput. Sci.* **218** (1999) 13-39.

[9] H. Fredricksen, A survey of full length nonlinear shift register cycle algorithms. *SIAM Rev.* **24** (1982) 195-221.

[10] L.J. Guibas and A. M. Odlyzko, Periods in strings. *J. Comb. Theory (A)* **30** (1981) 19-42.

[11] M. Lothaire, *Combinatorics on Words*, 2nd Edition. Cambridge Mathematical Library, Cambridge University Press, Cambridge, UK (1997).

[12] M. Lothaire, *Algebraic Combinatorics on Words.* Cambridge University Press, Cambridge, UK (2002).

[13] R.C. Lyndon and M.P. Schützenberger, The equation $a^M = b^N c^P$ in a free group. *Mich. Math. J.* **9** (1962) 289-298.

[14] E. Rivals and S. Rahmann, *Combinatorics of periods in strings.* Springer, Berlin, *Lecture Notes in Comput. Sci.* **2076** (2001) 615-626.

Communicated by J. Berstel.
Received November 15, 2001. Accepted May 27, 2002.