

Counting Types of Runs in Classes of Arborescent Words

Jean-Philippe Labbé¹, Gilbert Labelle²

¹Mathematics Institute, Free University of Berlin, Berlin, Germany

²University of Quebec at Montreal, Montreal City, Canada

Email: labbe@math.fu-berlin.de, labelle.gilbert@uqam.ca

Received November 16, 2012; revised December 16, 2012; accepted December 25, 2012

ABSTRACT

An *arborescence* is a directed rooted tree in which all edges point away from the root. An *arborescent word* is obtained by replacing each element of the underlying set of an arborescence by an arbitrary letter of a given alphabet (with possible repetitions). We define a *run* in an arborescent word as a maximal sub-arborescent word whose letters are all identical. Various types of runs (e.g., runs of size $\leq k$, linear runs, etc.) are studied in the context of R -enriched arborescent words, where R is a given species of structures.

Keywords: Runs; Structured Words; Treelike Structures; Combinatorial Species; Generating Series; Cycle Index Series

1. Introduction

Classically, a *word* over an alphabet Ω is a finite sequence w of letters taken from Ω and a *run* in a word w is a maximal subsequence of consecutive identical letters in w , see, for example, Feller [1], or Mood [2]. A slightly different point of view consists of interpreting a word over Ω as the result of the replacement of each element of a linear digraph by an arbitrary letter of Ω . A run in a word is then interpreted as a maximal linear sub-digraph whose letters are all identical. Here are four runs, $\clubsuit\clubsuit, \heartsuit\heartsuit\heartsuit, \clubsuit\clubsuit\clubsuit, \diamond$, in a 9-letter word over the alphabet $\Omega = \{\heartsuit, \spadesuit, \diamond, \clubsuit\}$ obtained under this interpretation (see below).

Using this point of view, words can now be easily “structured” in various ways. For example, starting from the digraph of a circular permutation (instead of a linear digraph, as above) gives rise to the concept of a *circular word* studied probabilistically in the basic paper of Koutras *et al.* [3]. In a similar way, any combinatorial species F , in the sense of Joyal [4], leads to a corresponding notion of *F-structured word* (or *F-word*, for short). Technically speaking, a *F-word of length n over an alphabet Ω* is an orbit of the following natural action of the symmetric group S_n ,

$$S_n \times (F[n] \times \Omega^n) \rightarrow F[n] \times \Omega^n, \quad (2)$$

defined by

$$\sigma \cdot (s, w_1 \cdots w_n) = (\sigma \cdot s, w_{\sigma^{-1}(1)} \cdots w_{\sigma^{-1}(n)}),$$

where $F[n]$ is the set of F -structures on

$$[n] = \{1, 2, \dots, n\},$$

$\sigma \cdot s$ is the result of relabelling the F -structure s along the permutation σ , and $w_1 \cdots w_n$ is an ordinary word of length n over the alphabet Ω . **Figure 1** illustrates this action for $n=5$, $\sigma = (132)(45)$: the left drawing shows the ordered pair $(s, w_1 \cdots w_5)$ where s is a graph on $[5]$ and w_i , the i -th letter of the word $w_1 \cdots w_5$, is written just under its associated vertex i in the graph, for $i=1, \dots, 5$; while the right drawing shows the resulting $\sigma \cdot (s, w_1 \cdots w_5)$ under the action of σ .

It is important to note that under this action, the labels $1, 2, \dots, 5$ of the underlying set of the structure s are permuted while each letter is kept in a *fixed* position. This is a general fact and it constitutes the reason why the orbits of action (2) can be thought of as unlabelled F -structures whose nodes are replaced (or filled) by letters of the alphabet, that is, F -words. Summing the orbits of (2) over n , gives rise to the concept of an *analytic functor* in the sense of Joyal [5],

$$\begin{array}{cccccccccc}
 1 & \rightarrow & 2 & \rightarrow & 3 & \rightarrow & 4 & \rightarrow & 5 & \rightarrow & 6 & \rightarrow & 7 & \rightarrow & 8 & \rightarrow & 9 \\
 \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
 \clubsuit & \rightarrow & \clubsuit & \rightarrow & \heartsuit & \rightarrow & \heartsuit & \rightarrow & \heartsuit & \rightarrow & \clubsuit & \rightarrow & \clubsuit & \rightarrow & \clubsuit & \rightarrow & \diamond
 \end{array} \quad (1)$$

$$\Omega \mapsto \sum_{n \geq 0} \frac{F[n] \times \Omega^n}{S_n},$$

of the category of sets into itself, sending each set Ω (= alphabet) to the set of all F -words over Ω .

There exist special classes of F -words, for which a natural notion of “run” can be defined. This is the case when F is a *graphical species*, whose structures are of the form of a simple graph possibly equipped with extra structures. Standard examples of graphical species include: the species of digraphs, of circular permutations, of tree-like structures (e.g., when F is the species of R -enriched rooted trees or the species of R -enriched trees in the sense of Labelle [6]), etc.

Definition 1 Let F be a graphical species and w be a F -word over an alphabet Ω . A run in w is a maximal connected subgraph of w having identical letters.

This definition makes sense since the relabellings of F -structures under the action (2) is letter-invariant. **Figure 2** shows the runs in a circular word and in a tree-like word over the alphabet $\{\heartsuit, \spadesuit, \diamondsuit, \clubsuit\}$.

The study of runs in graphical words appears far from being trivial, under such a general setting. For example, a simple graph word whose runs are all singletons is equivalent to a proper coloring of the graph. So that, even in this special case, the whole theory of chromatic polynomials must be used. In the present text, our run analysis will be concentrated on classes of arborescent words.

In Section 2, we study various types of runs in F -words over an alphabet of k letters, where F is the graphical species of R -enriched rooted trees, called here R -arborescences, for short. To do so, we make use of functional equations on multisort species and introduce a special 2-sort species, that we call a *runs selector species*, to classify the types of runs under study. Various enu-

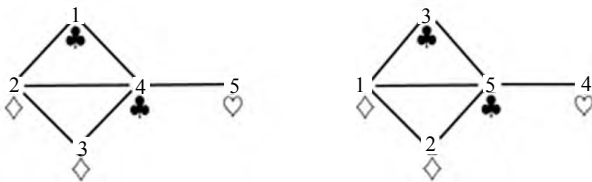


Figure 1. The pairs $(s, w_1 \cdots w_5)$ and $\sigma \cdot (s, w_1 \cdots w_5)$ under the action of $\sigma = (132)(45)$.

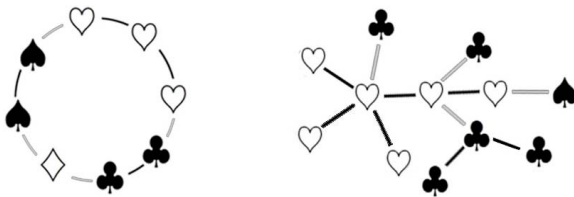


Figure 2. Four runs in a circular word and five runs in a tree-like word.

merative results about runs (e.g., number of F -words whose runs are all of a given type, distribution of the biggest run, etc) then follow by making use of appropriate underlying cycle index series. Section 3 illustrates the theory of R -arborescent words by analysing specific examples and their associated series. Section 4 indicates some extensions to be developed in the future. For an introduction to species, the reader can consult the basic paper of Joyal [4] or the book by Bergeron, Labelle, and Leroux [7].

2. Runs in Enriched Arborescent Words

An *arborescence* is a directed rooted tree in which all edges point away from the root. Let $R = R(X)$ be a given species of structures having at least one structure on the empty set. Recall that an R -arborescence is an arborescence in which the (possibly empty) set of direct successors of each vertex is equipped with an R -structure. **Figure 3(a)** shows an R -arborescence, where the black dots represent the elements of its underlying set. The small arc at each vertex represents the R -structure placed on the set of successors of the vertex. The species $A = A(X)$ of R -arborescences is characterized by the combinatorial equation,

$$A = XR(A), \tag{3}$$

and, depending on the choice of the “enriching” species, R , various species of tree-like structures are obtained. Now, take a k -letter alphabet $\Omega = \{\ell_1, \dots, \ell_k\}$ and associate a sort X_i of singletons to letter ℓ_i , for $i = 1, 2, \dots, k$. The structures belonging to the k -sort species $A(X_1 + \dots + X_k)$ are R -arborescences whose vertices are of arbitrary sorts taken among the X_i ’s. The first step in our run analysis is to classify these multisort structures according to the *runs of sorts* they contain. From (3), we obviously have,

$$\begin{aligned} & A(X_1 + \dots + X_k) \\ &= (X_1 + \dots + X_k)R(A(X_1 + \dots + X_k)) \\ &= X_1R(A(X_1 + \dots + X_k)) + \dots \\ &+ X_kR(A(X_1 + \dots + X_k)) \\ &= A^{[1]}(X_1, \dots, X_k) + \dots + A^{[k]}(X_1, \dots, X_k), \end{aligned} \tag{4}$$

where $A^{[i]}(X_1, \dots, X_k) = X_iR(A(X_1 + \dots + X_k))$ is the species of the $A(X_1 + \dots + X_k)$ -structures having a root of sort X_i . Note that this root of sort X_i is part of a run of vertices of sort X_i . To reflect this fact, we now give an alternate expression for $A^{[i]}$ in terms of an auxiliary two-sort species, $B = B(X, T)$, where T is an extra sort of singletons.

Lemma 2 Let $B = B(X, T)$ be the two-sort species defined by

$$B = XR(B + T). \tag{5}$$

Then, the species $A^{[i]} = A^{[i]}(X_1, \dots, X_k)$ satisfy the combinatorial equations,

$$A^{[i]} = B\left(X_i, \sum_{j \neq i} A^{[j]}\right), \quad i = 1, \dots, k. \tag{6}$$

Proof. **Figure 3(b)** shows a B -structure, where singletons of sort X (resp., T) are drawn as black dots (resp., white triangles). The result follows from the fact that an $A^{[i]}$ -structure is a B -structure in which the blackdots are replaced by singletons of sort X_i and the white triangles are replaced by arbitrary $A^{[j]}$ -structures, $j \neq i$, (see **Figure 4(a)**, where $n = 3$, $i = 2$, X_1 -singletons are white dots, X_2 -singletons are black dots, X_3 -singletons are gray dots).

Definition 3 Any subspecies $\hat{B}(X, T) \mapsto B(X, T)$ is called a runs selector species. The species

$$\hat{A} = \hat{A}(X_1, \dots, X_k)$$

of R -arborescences having only runs of sorts of type \hat{B} is defined by

$$\begin{aligned} \hat{A} &= \hat{A}^{[1]} + \dots + \hat{A}^{[i]} + \dots + \hat{A}^{[k]}, \\ \hat{A}^{[i]} &= \hat{B}\left(X_i, \sum_{j \neq i} \hat{A}^{[j]}\right), \quad i = 1, \dots, k. \end{aligned} \tag{7}$$

An R -arborescent word having only runs of type \hat{B} over the alphabet $\Omega = \{\ell_1, \dots, \ell_k\}$ is the result of the replacement of each vertex of sort X_i in a \hat{A} -structure, by the letter ℓ_i , for $i = 1, \dots, k$.

It is worthwhile to note that a runs selector species $\hat{B}(X, T)$ does not only select the *shape* of the runs under interest but also specifies how these runs should be *interconnected*. For example,

- $\hat{B}(X, T) = B(X, T)$ selects *unrestricted runs* (see **Figure 3(b)**),
- $\hat{B}(X, T) = XR(T)/(1 - XR'(T))$ selects *linear runs* (see **Figure 5(a)**),
- $\hat{B}(X, T) = XR(T)(1 + XR'(T) + \dots + (XR'(T))^{m-1})$ selects *linear runs of length $\leq m$* ,
- $\hat{B}(X, T) = XR(T)$ selects *singleton runs* (see **Figure 5(b)**),
- $\hat{B}(X, T) = A(X) + XR'(A(X))T/(1 - XR'(A(X)))$ selects *maximal runs* (see **Figure 5(c)**), etc.

The main interest in Lemma 2 and Definition 3 is that they give rise to a variety of enumerative results about runs of various types in R -arborescent words by the introduction of special weight counters in (5)-(7). For example, taking $R = (1 + C)$, where C is the species of cyclic permutations, Equation (3) takes the form

$A = X(1 + C(A))$ whose solution is the species M of

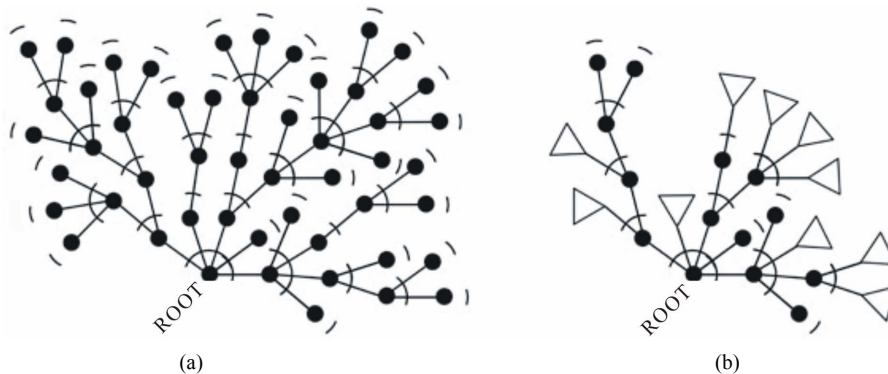


Figure 3. (a) R -arborescence; (b) B -structure.

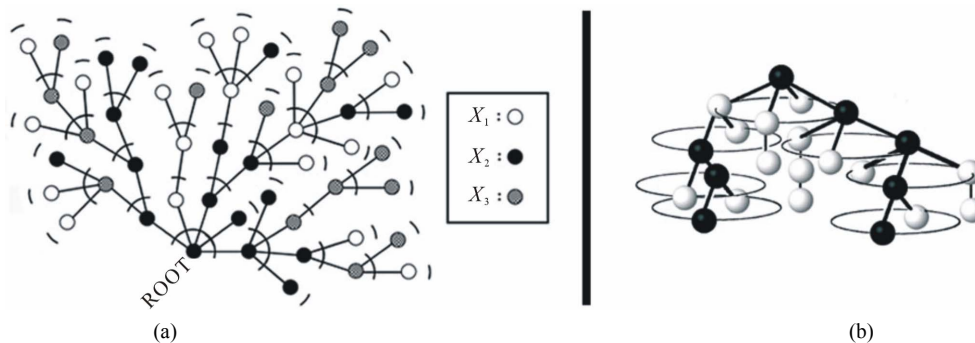


Figure 4. (a) $A_2(X_1, X_2, X_3)$ -structure; (b) Mobile word with linear runs.

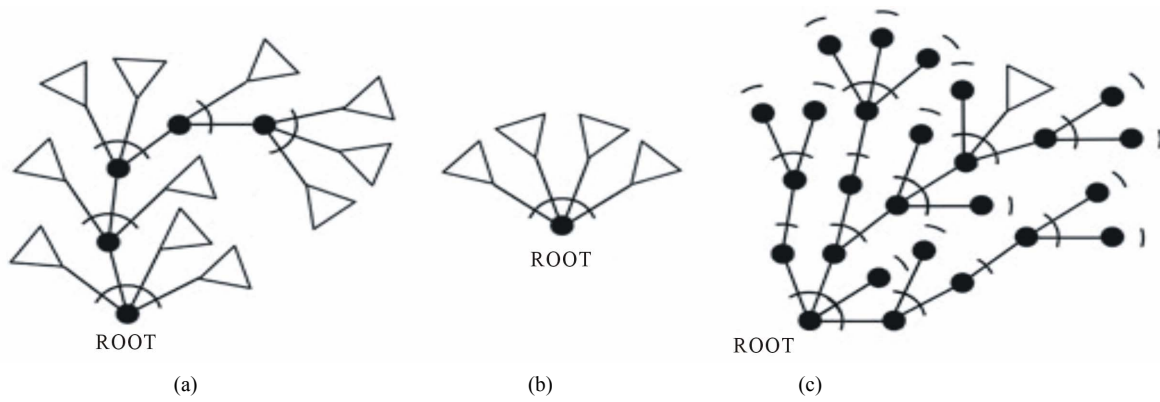


Figure 5. Selectors for (a) Linear runs; (b) Singleton runs; and (c) Maximal runs.

mobiles in the terminology of Bergeron, Labelle, and Leroux [7]. The reason is the following. Putting the root on top, the descendants of each node can be thought as turning around along horizontal circles. In this case, M -words are called mobile words. Figure 4(b) illustrates a mobile word on a 2-letter alphabet $\Omega = \{\text{black ball, white ball}\}$ having only linear (type (ii)) runs.

The introduction of a run counter u and the simultaneous substitutions $X_i = zX, i = 1, 2, \dots, n$, in (5)-(7), where z is a letter counter lead to the following general proposition.

Proposition 4

Let $A = A(X)$ be the species of R -arborescences and $\hat{B} = \hat{B}(X, T) \rightarrow B$ be a runs selector species. Give a weight u^h to each A -word with h runs and let $f_n(u)$ be the total weight of A -words of length n having only runs of type \hat{B} over the k -letter alphabet $\Omega = \{\ell_1, \dots, \ell_k\}, k > 1$. Then the generating series

$$f(u, z) = \sum_{n \geq 0} f_n(u) z^n$$

is given by,

$$f(u, z) = \frac{k}{k-1} g(u, z),$$

$$g(u, z) = (k-1)uZ_{\hat{B}}(z, g(u, z), z^2, g(u^2, z^2), z^3, g(u^3, z^3), \dots), \quad (8)$$

where $Z_{\hat{B}}(x_1, t_1, x_2, t_2, x_3, t_3, \dots)$ is the cycle index series of $\hat{B}(X, T)$. In the special case, $\hat{B} = B$, corresponding to unrestricted runs, the generating series $f(u, z)$ is explicitly given by,

$$f(u, z) = \frac{ku}{1+(k-1)u} Z_A\left(\left(1+(k-1)u\right)z, \left(1+(k-1)u^2\right)z^2, \left(1+(k-1)u^3\right)z^3, \dots\right), \quad (9)$$

where $Z_A(x_1, x_2, x_3, \dots)$ is the cycle index series of A . Proof. The series $f(u, z)$ is the total weight of all

unlabelled \hat{A} -structures in which each node is replaced by the variable counter z and each run is given a weight u . Now, consider the species

$$F(X) = \hat{A}(X, \dots, X)$$

obtained by the substitutions $X_i := X, i = 1, \dots, k$, in the species $\hat{A}(X_1, \dots, X_k)$ weighted as above. Then, from Pólya Theory in the context of species, $f(u, z) = Z_F(z, z^2, z^3, \dots)$, where $Z_F(x_1, x_2, x_3, \dots)$ is the cycle index series of $F(X)$. By symmetry, all the species $\hat{A}^{[i]}(X, \dots, X), i = 1, \dots, k$, are equal. So that they can all be identified with a single species, $H(X)$, say. By (4), taking into account the run counter, u ,

$$F(X) = \hat{A}(X, \dots, X) = \sum_{1 \leq i \leq k} \hat{A}^{[i]}(X, \dots, X) = kH(X), \quad (10)$$

$$H(X) = u\hat{B}\left(X, \sum_{j \neq i} \hat{A}^{[j]}(X, \dots, X)\right) = u\hat{B}(X, (k-1)H(X)). \quad (11)$$

Define the species $G(X)$ by $G(X) = \frac{k-1}{k} F(X)$.

Then, $G(X) = (k-1)u\hat{B}(X, G(X))$, and the recursive scheme (8) follows by taking the cycle index series of both sides of this last equation. Now, consider the special choice, $\hat{B} = B$. Then, using the fact that

$$B(X, T) = XR(B(X, T) + T),$$

we have,

$$G(X) = (k-1)uXR(B(X, G(X)) + G(X)) = (k-1)uXR\left(\frac{1}{(k-1)u}G(X) + G(X)\right) = (k-1)uXR\left(\left(1 + \frac{1}{(k-1)u}\right)G(X)\right). \quad (12)$$

Hence,

$$\begin{aligned} & \left(1 + \frac{1}{(k-1)u}\right)G(X) \\ &= (1+(k-1)u)XR\left(\left(1 + \frac{1}{(k-1)u}\right)G(X)\right). \end{aligned}$$

This implies that the species

$$K(X) = \left(1 + \frac{1}{(k-1)u}\right)G(X)$$

satisfies the equation

$$K(X) = (1+(k-1)u)XR(K(X)).$$

Since $A = XR(A)$, we must have

$$K(X) = A((1+(k-1)u)X)$$

by the implicit species theorem [4]. So that,

$$F(X) = \frac{ku}{1+(k-1)u}A((1+(k-1)u)X),$$

and (9) follows by taking the cycle index series on both sides of this combinatorial equation.

The following corollary is immediate from Proposition 4, via the substitution $u = 1$.

Corollary 5 *Let $A = A(X)$ be the species of R -arborescences and $\hat{B} = \hat{B}(X, T) \mapsto B$ be a runs selector species. Let f_n be the number of A -words of length n having only runs of type \hat{B} over the k -letter alphabet $\Omega = \{\ell_1, \dots, \ell_k\}$, $k > 1$. Then the generating series, $f(z) = \sum_{n \geq 0} f_n z^n$, is given by,*

$$\begin{aligned} f(z) &= \frac{kg(z)}{k-1}, \\ g(z) &= (k-1)Z_{\hat{B}}(z, g(z), z^2, g(z^2), z^3, g(z^3), \dots). \end{aligned} \tag{13}$$

In the special case, $\hat{B} = B$, corresponding to unrestricted runs, the series $f(z)$ is explicitly given by,

$$f(z) = Z_A(kz, kz^2, kz^3, \dots). \tag{14}$$

The use of cycle index series above comes from the fact that R -enriched arborescences have non-trivial automorphisms, in general. This occurs when R -structures have non-trivial automorphisms. In the *asymmetric* case, however, that is when the automorphism group of each R -structure is trivial, it is well-known that cycle index series reduce to exponential generating series.

Corollary 6 *If the species, R , is asymmetric, then*

$$f(u, z) = \frac{k}{k-1}g(u, z) \text{ and } f(z) = \frac{k}{k-1}g(z) \text{ with}$$

$$\begin{aligned} g(u, z) &= (k-1)u\hat{B}(z, g(u, z)), \\ g(z) &= k\hat{B}(z, g(z)), \end{aligned} \tag{15}$$

where $\hat{B}(x, t)$ is the exponential generating series of \hat{B} . Moreover, if $\hat{B} = B$, then

$$\begin{aligned} f(u, z) &= \frac{ku}{1+(k-1)u}A((1+(k-1)u)z), \\ f(z) &= A(kz), \end{aligned} \tag{16}$$

where $A(x)$ is the exponential generating series of A .

Various computational general methods exist to obtain successive approximations (or explicit expressions, sometimes) for $f(z)$ and $f(u, z)$. Simple iteration is the most obvious one, but more sophisticated methods include Good-Lagrange inversion, the use of eclosion operators and Newton-Raphson iteration adapted to cycle index series (see Labelle [8,9]). There is not enough space here to describe these methods in detail in the present context. Instead, we have chosen to illustrate our analysis of runs by focusing on specific examples.

3. Specific Examples

3.1. Runs in Ordinary Words

With $R = 1 + X$, Equation (3) takes the form

$A = X(1 + A)$ whose solution is the *asymmetric* species $A = L_+ = X/(1 - X)$ of non empty linear orders. In this case, L_+ -words are simply (ordinary non empty) words and Equation (5) takes the form $B = X(1 + B + T)$ whose solution is

$$B(X, T) = \frac{X}{1-X}(1+T). \tag{17}$$

- In the case $\hat{B} = B$, of unrestricted runs, (16) immediately gives,

$$f(u, z) = \frac{kuz}{1-(1+(k-1)u)z}, f(z) = \frac{kz}{1-kz}. \tag{18}$$

- Take the runs selector

$$\begin{aligned} \hat{B} &= (X + X^2 + \dots + X^m)(1+T) \\ &= \frac{X(1-X^m)}{1-X}(1+T) \mapsto B \end{aligned}$$

of linear runs of length $\leq m$. Then recurrences (15) take the form,

$$\begin{aligned} g(u, z) &= (k-1)u \frac{z(1-z^m)}{1-z}(1+g(u, z)), \\ g(z) &= (k-1) \frac{z(1-z^m)}{1-z}(1+g(z)). \end{aligned} \tag{19}$$

Solving for g , we get the well known g.f.'s of words

whose runs are of size $\leq m$,

$$f(u, z) = \frac{kuz(1-z^m)}{1-(1+(k-1)u)z+(k-1)uz^{m+1}}, \tag{20}$$

$$f(z) = \frac{kz(1-z^m)}{1-kz+(k-1)z^{m+1}}.$$

Note that taking $k = m = 2$ in (20) can be called the *Fibonacci* case since gives,

$$f(z) = 2 \frac{z+z^2}{1-z-z^2} = 2 \sum_{n \geq 1} F_{n+1} z^n, \tag{21}$$

where F_m is the m -th Fibonacci number. This corresponds to twice the well known fact that a sequence of n adjacent squares can be tiled by monominos or dominos in F_{n+1} ways. Taking $u = 1$ in (20) and subtracting gives the g.f. of words whose maximal runs are of size m ,

$$\frac{kz(1-z^m)}{1-kz+(k-1)z^{m+1}} - \frac{kz(1-z^{m-1})}{1-kz+(k-1)z^m}, \tag{22}$$

from which one can compute or estimate the expected size of the largest run (see, for example, Erdos and Revesz [10], Guibas and Odlyzko [11], Kong [12]).

- Take any subset $S \subseteq \mathbb{N}_+$ and the selector

$$\hat{B} = \left(\sum_{i \in S} X^i \right) (1+T) \mapsto B,$$

of runs of length in S . Then,

$$f(u, z) = \frac{ku \sum_{i \in S} z^i}{1-(k-1)u \sum_{i \in S} z^i}, \tag{23}$$

$$f(z) = \frac{k \sum_{i \in S} z^i}{1-(k-1) \sum_{i \in S} z^i}.$$

3.2. Runs in Binary Arborescent Words

Planar case. With $R = 1 + X^2$, where X^2 is the species of linearly ordered 2-point sets, Equation (3) takes the form $A = X(1 + A^2)$ whose solution is the *asymmetric* species $\text{Bin}^{\text{plane}}$ of binary plane arborescences. Accordingly, $\text{Bin}^{\text{plane}}$ -words are binary plane arborescent words and (5) takes the form $B = X(1 + (B+T)^2)$.

- Take the *linear* runs selector

$$\hat{B} = \frac{X(1+T^2)}{1-2XT} \mapsto B$$

corresponding to (ii) above. Then by (16), the g.f.'s of plane binary arborescent words having only linear runs are given by,

$$f(u, z) = \frac{k(1 - \sqrt{1 - 4u(k-1)(2 + (k-1)u)z^2})}{2(k-1)(2 + (k-1)u)z}, \tag{24}$$

$$f(z) = \frac{k(1 - \sqrt{1 - 4(k^2-1)z^2})}{2(k^2-1)z}.$$

Coefficient extractions give, for n odd,

$$f_n(u) = k((k-1)(2+(k-1)u))^{(n-1)/2} C_{(n-1)/2} u^{(n+1)/2},$$

$$f_n = k(k^2-1)^{(n-1)/2} C_{(n-1)/2},$$

where $C_m = \frac{1}{m+1} \binom{2m}{m}$, $m = 0, 1, 2, \dots$, are the Catalan numbers.

- The *Fibonacci* case, $k = 2, m = 2$, corresponds, in the present context, to case (iii) above with a 2-letter alphabet. Here, $\hat{B} = X(1+T^2)(1+2XT) \mapsto B$. So that,

$$f(u, z) = 2uz \left(1 + \frac{1}{4} f(u, z)^2 \right) (1 + zf(u, z)). \tag{25}$$

The first terms of $f(u, z)$ and of $f(z)$ are given by,

$$f(u, z) = 2uz + (4u^2 + 2u^3)z^3$$

$$+ (8u^3 + 16u^4 + 4u^5)z^5$$

$$+ (16u^4 + 80u^5 + 60u^6 + 10u^7)z^7$$

$$+ (32u^5 + 320u^6 + 504u^7 + 224u^8 + 28u^9)z^9 + \dots$$

$$f(z) = 2z + 6z^3 + 28z^5 + 166z^7 + 1108z^9 + 7932z^{11}$$

$$+ 59512z^{13} + 461830z^{15} + 3676324z^{17} + \dots$$

Non-planar case. With $R = 1 + E_2$, where E_2 is the species of 2-element sets, Equation (3) takes the form $A = X(1 + E_2(A))$ whose solution is the species Bin of (free) binary arborescences. Accordingly, Bin -words are binary arborescent words and (5) takes the form $B = X(1 + E_2(B+T))$. Note that in this case, Bin -structures have non trivial symmetries in general, so that cycle-index series must be used instead of exponential generating series.

- Take the *linear* runs selector

$$\hat{B} = \frac{X(1+E_2(T))}{1-XT} \mapsto B$$

corresponding to (ii) above. Then

$$Z_{\hat{B}} = x_1 \left(1 + \frac{1}{2} (t_1^2 + t_2) \right) / (1 - x_1 t_1)$$

and by (8), $f(u, z) = \frac{k}{k-1} g(u, z)$ where,

$$g(u, z) = (k-1)uz \left(1 + \frac{1}{2} \left(g(u, z)^2 + g(u^2, z^2) \right) \right) / (1 - zg(u, z)) \quad (26)$$

For example, the first terms of $f(z)$ are given by,

$$f(z) = kz + \frac{1}{2}(k+2)(k-1)kz^3 + \frac{1}{2}(k+2)(k+1)(k-1)^2 kz^5 + \frac{1}{8}(k+2)(5k^3 + 6k^2 - 5k - 4)(k-1)^2 kz^7 + \dots$$

- In the *Fibonacci* case, $k = 2, m = 2$, the runs selector (iii) above is given by

$$\hat{B} = X(1 + E_2(T))(1 + XT) \mapsto B.$$

So that,

$$f(u, z) = uz \left(2 + zf(u, z) \cdot \left(1 + \frac{1}{8} f(u, z)^2 + \frac{1}{4} f(u^2, z^2) \right) \right). \quad (27)$$

The first terms of $f(u, z)$ and of $f(z)$ are given by,

$$f(u, z) = 2uz + (2u^2 + 2u^3)z^3 + (2u^3 + 6u^4 + 2u^5)z^5 + (2u^4 + 14u^5 + 14u^6 + 4u^7)z^7 + (2u^5 + 26u^6 + 54u^7 + 34u^8 + 6u^9)z^9 + \dots$$

$$f(z) = 2z + 4z^3 + 10z^5 + 34z^7 + 122z^9 + 482z^{11} + 1976z^{13} + 8436z^{15} + 36890z^{17} + 164850z^{19} + \dots$$

3.3. Runs in Cayley Arborescent Words

With $R = E$, where E is the species of finite sets, Equation (3) takes the form $A = XE(A)$ whose solution is the species A of standard Cayley arborescences. In this case, A -words are Cayley arborescent words and equation (5) takes the form $B = XE(B+T)$. Since $E(B+T) = E(B)E(T)$, this equation can be rewritten as $B = (XE(T))E(B)$. Hence, from the unicity of solution in the implicit species theorem, we have,

$$B(X, T) = A(XE(T)), \quad (28)$$

$$Z_B = Z_A(x_1 e^{t_1+t_2/2+t_3/3+\dots}, x_2 e^{t_2+t_4/2+t_6/3+\dots}, \dots).$$

- Take the runs selector $\hat{B} = B$. Then since $Z_A = x_1 Z_E(Z_A)$, (9) and (14) can be rewritten as,

$$f(u, z) = kuz \exp \left(\sum_{i \geq 1} \frac{1+(k-1)u^i}{iku^i} f(u^i, z^i) \right), \quad (29)$$

$$f(z) = kz \exp \left(\sum_{i \geq 1} \frac{1}{i} f(z^i) \right).$$

For example, the first terms of $f(u, z)$ and $f(z)$ are given by,

$$f(u, z) = kuz + (ku + k(k-1)u^2)z^2 + \left(2ku + 3k(k-1)u^2 + \frac{1}{2}k(3k-2)(k-1)u^3 \right)z^3 + \left(4ku + 9k(k-1)u^2 + k(8k-7)(k-1)u^3 + \frac{1}{3}k(k-1)(8k^2-13k+6)u^4 \right)z^4 + \dots$$

$$f(z) = kz + k^2z^2 + \frac{1}{2}k^2(1+3k)z^3 + \frac{1}{3}k^2(1+3k+8k^2)z^4 + \frac{1}{24}k^2(6+31k+54k^2+125k^3)z^5 + \frac{1}{15}k^2(3+10k+45k^2+80k^3+162k^4)z^6 + \dots$$

- Take the *linear* runs selector $\hat{B} = \frac{XE(T)}{1-XE(T)} \mapsto B$ corresponding to (ii) above. Then,

$$f(u, z) = \frac{k}{k-1} g(u, z),$$

$$g(u, z) = (k-1)u \frac{z \exp \left(\sum_{i \geq 1} \frac{1}{i} g(u^i, z^i) \right)}{1 - z \exp \left(\sum_{i \geq 1} \frac{1}{i} g(u^i, z^i) \right)}. \quad (30)$$

More generally, the runs selector

$$\hat{B} = \frac{X^m E(mT)}{1-XE(T)} \mapsto B$$

corresponds to linear runs of size $\geq m$. Here,

$$f(u, z) = \frac{k}{k-1} g(u, z),$$

$$g(u, z) = (k-1)u \frac{z^m \exp \left(m \sum_{i \geq 1} \frac{1}{i} g(u^i, z^i) \right)}{1 - z \exp \left(\sum_{i \geq 1} \frac{1}{i} g(u^i, z^i) \right)}. \quad (31)$$

- For Cayley arborescences, the *Fibonacci* case, $k = 2, m = 2$, corresponds to the runs selector $\hat{B} = XE(T)(1 + XE(T)) \mapsto B$. So that,

$$f(u, z) = 2uz \exp \left(\frac{1}{2} \sum_{i \geq 1} \frac{1}{i} f(u^i, z^i) \right) \cdot \left(1 + z \exp \left(\frac{1}{2} \sum_{i \geq 1} \frac{1}{i} f(u^i, z^i) \right) \right). \quad (32)$$

- For singleton runs (case (iv) above),

$$\hat{B} = XE(T) \mapsto B. \text{ So that,}$$

$$f(u, z) = \frac{k}{k-1} g(u, z),$$

$$g(u, z) = (k-1)uz \exp\left(\sum_{i \geq 1} \frac{1}{i} g(u^i, z^i)\right). \tag{33}$$

- For maximal runs (case (v) above), $\hat{B} = A(X) + A(X)T / (1 - A(X)) \mapsto B$, since $XR'(A) = XE(A) = A$. So that,

$$f(u, z) = ku \frac{\tilde{A}(z)(1 - \tilde{A}(z))}{1 - (1 + (k-1)u)\tilde{A}(z)},$$

$$f(z) = k \frac{\tilde{A}(z)(1 - \tilde{A}(z))}{1 - k\tilde{A}(z)}, \tag{34}$$

where $\tilde{A}(z) = Z_A(z, z^2, z^3, \dots)$ is the ordinary generating series of unlabelled Cayley arborescences.

3.4. Runs in Mobile Words

Recall that with $R = (1 + C)$, where C is the species of cyclic permutations, Equation (3) takes the form $A = X(1 + C(A))$ whose solution is the species M of mobiles. In this case, M -words are called *mobile words* (see **Figure 4(b)**). Since

$$R' = (1 + C)' = C' = L = 1/(1 - X) = \text{linear orders,}$$

the above case (ii) of linear runs, for example, corresponds to the runs selector,

$$\hat{B} = X(1 + C(T))L(XL(T)), \tag{35}$$

and the generating series, $f(u, z)$, of mobile words having only linear runs, weighted by the run counter u is given by,

$$f(u, z) = \frac{k}{k-1} g(u, z),$$

$$g(u, z) = (k-1)uz \frac{1 - g(u, z)}{1 - z - g(u, z)} \cdot \left(1 - \sum_{i \geq 1} \frac{\phi(i)}{i} \log(1 - g(u^i, z^i))\right) \tag{36}$$

since $Z_C(t_1, t_2, \dots) = -\sum_{i \geq 1} \frac{\phi(i)}{i} \log(1 - t_i)$, where ϕ is the Euler function. For example, the first terms of $f(z)$, for a 2-letter alphabet ($u = 1, k = 2$) are given by,

$$f(z) = 2z + 4z^2 + 12z^3 + 40z^4 + 150z^5 + 596z^6 + 2522z^7 + 11012z^8 + 49536z^9 + \dots$$

4. Some Extensions

The present approach to runs in structured words can be

extended in various directions. For example, a probabilistic analysis of runs can be made by adding a probability of occurrence π_i to letter ℓ_i in the alphabet $\Omega = \{\ell_1, \ell_2, \dots\}$. Runs in other classes of graphical words can also be considered. For example, the class of *R-enriched tree words* is a tractable one. Here, the species $A = XR(A)$ of *R-enriched arborescences* is replaced by the species \mathcal{A} of *R-enriched trees*. This case needs a more complex analysis since extra symmetries of the structures have to be taken into account (the structures being unrooted). The (enriched) dissymmetry formula,

$$\mathcal{A}_R = XR(\mathcal{A}_{R'}) + E_2(\mathcal{A}_{R'}) - \mathcal{A}_{R'}^2, \tag{37}$$

which expresses the species \mathcal{A}_R of *R-enriched trees* in terms of the species $\mathcal{A}_{R'}$ of *R'-enriched arborescences* (see Bergeron, Labelle, and Leroux [7]), is the tool to be used for the analysis of the multisort species $\mathcal{A}_R(X_1 + \dots + X_k)$ leading to an analogue of Proposition 4 for *R-enriched tree words*. These extensions will be developed elsewhere.

5. Acknowledgements

The authors would like to thank Jérôme Tremblay for his LaTeX and Maple help.

REFERENCES

- [1] W. Feller, "An Introduction to Probability Theory and Its Applications: Vol. I," 3rd Edition, John Wiley & Sons Inc., New York, 1968.
- [2] A. M. Mood, "The Distribution Theory of Runs," *Annals of Mathematical Statistics*, Vol. 11, No. 4, 1940, pp. 367-392. [doi:10.1214/aoms/1177731825](https://doi.org/10.1214/aoms/1177731825)
- [3] M. V. Koutras, G. K. Papadopoulos and S. G. Papastavridis, "Runs on a Circle," *Journal of Applied Probability*, Vol. 32, No. 2, 1995, pp. 396-404. [doi:10.2307/3215295](https://doi.org/10.2307/3215295)
- [4] A. Joyal, "Une Théorie Combinatoire des Séries Formelles," *Advances in Mathematics*, Vol. 42, No. 1, 1981, pp. 1-82. [doi:10.1016/0001-8708\(81\)90052-9](https://doi.org/10.1016/0001-8708(81)90052-9)
- [5] A. Joyal, "Foncteurs Analytiques et Espèces de Structures," In: *Combinatoire Énumérative (Montreal, Quebec, 1985)*, Vol. 1234 of *Lecture Notes in Mathematics*, Springer, Berlin, 1986, pp. 126-159.
- [6] G. Labelle, "Une Nouvelle Démonstration Combinatoire des Formules D'inversion de Lagrange," *Advances in Mathematics*, Vol. 42, No. 3, 1981, pp. 217-247. [doi:10.1016/0001-8708\(81\)90041-4](https://doi.org/10.1016/0001-8708(81)90041-4)
- [7] F. Bergeron, G. Labelle and P. Leroux, "Combinatorial species and tree-like structures, Vol. 67 of *Encyclopedia of Mathematics and its Applications*," Cambridge University Press, Cambridge, 1998.
- [8] G. Labelle, "Éclisions Combinatoires Appliquées à L'inversion Multidimensionnelle des Séries Formelles," *Journal of Combinatorial Theory, Series A*, Vol. 39, No. 1,

- 1985, pp. 52-82. [doi:10.1016/0097-3165\(85\)90083-4](https://doi.org/10.1016/0097-3165(85)90083-4)
- [9] G. Labelle, "Some New Computational Methods in the Theory of Species," In *Combinatoire Énumérative Lecture Notes in Mathematics Vol. 1234*, Springer, Berlin, 1986, pp. 192-209.
- [10] P. Erdős and P. Révész, "On the Length of the Longest Head-Run," In *Topics in Information Theory (Second Colloq., Keszthely, 1975), Colloquia Mathematica Societatis János Bolyai, Vol. 16*, North-Holland, Amsterdam, 1977, pp. 219-228.
- [11] L. J. Guibas and A. M. Odlyzko, "Long Repetitive Patterns in Random Sequences," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol. 53, No. 3, 1980, pp. 241-262. [doi:10.1007/BF00531434](https://doi.org/10.1007/BF00531434)
- [12] Y. Kong, "Distribution of Runs and Longest Runs: A New Generating Function Approach," *Journal of the American Statistical Association*, Vol. 101, No. 475, 2006, pp. 1253-1263. [doi:10.1198/016214505000001401](https://doi.org/10.1198/016214505000001401)