

A Combinatorial Analysis of Tree-Like Sentences

Gilbert Labelle¹, Louise Laforest²

¹Département de Mathématiques, Université du Québec à Montréal, Montreal, Canada ²Département d'Informatique, Université du Québec à Montréal, Montreal, Canada Email: <u>labelle.gilbert@uqam.ca</u>, <u>laforest.louise@uqam.ca</u>

Received 3 June 2015; accepted 21 July 2015; published 24 July 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0/

Abstract

A *sentence* over a finite alphabet *A*, is a finite sequence of non-empty words over *A*. More generally, we define a *graphical sentence* over *A* by attaching a non-empty word over *A* to each arrow and each loop of a connected directed graph (digraph, for short). Each word is written according to the direction of its corresponding arrow or loop. Graphical sentences can be used to encode sets of sentences in a compact way: the *readable sentences* of a graphical sentence being the sentences corresponding to directed paths in the digraph. We apply combinatorial equations on enriched trees and rooted trees, in the context of combinatorial species and Pólya theories, to analyze parameters in classes of tree-like sentences. These are graphical sentences constructed on tree-like digraphs.

Keywords

Pólya Theory, Combinatorial Species, Digraphs, Tree-Like Sentences

1. Introduction

Figure 1 (left) shows a completely unlabelled¹ connected digraph. We define a graphical sentence over a finite alphabet *A* by attaching a non-empty word over *A* to each arrow and each loop of a completely unlabelled connected digraph. Each word must be written according to the direction of its corresponding arrow or loop, from source to target. **Figure 1** (middle) shows a graphical sentence over alphabet {A, C, G, T} and **Figure 1** (right) shows another over alphabet { Λ , J, Λ , J, Λ , J}.

Graphical sentences can be used to encode sets of ordinary sentences in a compact way: The *readable sentences* of a graphical sentence being the sentences corresponding to directed paths in its digraph. For example,

¹In the sense that both its vertices, arrows and loops are unlabelled.



Figure 1. Unlabelled digraph and graphical sentences over alphabets {A, C, G, T} and { $\{\lambda, \downarrow, \neg, \circ, \downarrow\}$

TTT C GCCTG CAT CAT GCAATT, is a readable sentence arising from the graphical sentence of Figure 1 (middle).

In the present paper we focus our attention on the *structure* of graphical sentences as *combinatorial objects* using methods from the theory of combinatorial species [1] [2] and classical Pólya theory [3]. We leave aside the generation of the readable sentences of a graphical sentence since this is easily done via the computation of powers of incidence matrices². Of course, special sentences among the readable sentences can be selected by adding extra structure to graphical sentences (such as *source points, sink points, STOP points, counters, extensions of the alphabet* by adding special characters such as \sqcup , !, ?, etc). We also leave aside this aspect in our analysis of graphical sentences.

Various descriptive parameters can be attached to each graphical sentence over a given alphabet *A*. For example, the graphical sentence of **Figure 1** (middle) is made of 7 vertices, 11 arrows, 2 loops, 52 letters, letter A appears 13 times, letter C appears 9 times, letter G appears 10 times and letter T appears 20 times.

As usual in enumerative combinatorics, families of parameters associated to structures are conveniently encoded by weight-monomials.

Definition 1.1. The *weight* of a graphical sentence *s* over an alphabet *A* is the (commutative) formal monomi al^3

$$\mathbf{w}(s) = x^{\text{#vertices in } s} y^{\text{#arrows in } s} z^{\text{#loops in } s} t^{\text{#letters in } s} \prod_{a \in A} a^{\nu_a(s)}$$
(1)

where $v_a(s)$ is the number of occurrences of letter *a* in *s*.

In (1), each letter $a \in A$ is reinterpreted as a formal variable. For example, the weight of the graphical sentence s of Figure 1 (middle) is given by

$$\mathbf{w}(s) = x^7 y^{11} z^2 t^{52} A^{13} C^9 G^{10} T^{20}.$$
 (2)

Definition 1.2. Let C be any class of totally unlabelled connected digraphs and $\mathbb{L} \subseteq \mathbb{N}^+ = \{1, 2, 3, \cdots\}$. Let \mathfrak{S} be the (countable) set of all graphical sentences over alphabet *A* arising from digraphs in C, the word on each arrow or loop having a length $\in \mathbb{L}$. The inventory of \mathfrak{S} is the formal sum of the weights of all graphical sentences in \mathfrak{S} :

$$\mathbf{w}\mathfrak{S} = \mathbf{w}\mathfrak{S}(x, y, z, t, (a)_{a \in A}) = \sum_{s \in \mathfrak{S}} \mathbf{w}(s).$$
(3)

As usual in enumeration problems, the (explicit or recursive) computation of an inventory of a class of structures provides a great deal of information about the structures to which it is associated. This information is extracted from the inventory through *expansion*, *collection of terms*, *specialization/confluence of variables*, *algebraic/differential manipulations and coefficient extraction*.

For example, in the present situation, expanding and collecting terms in (3) gives, of course,

$$\mathbf{w}\mathfrak{S}(x, y, z, t, (a)_{a \in A}) = \sum_{m, n, p, q, (v_a)_{a \in A}} c_{m, n, p, q, (v_a)_{a \in A}} x^m y^n z^p t^q \Pi a^{v_a},$$
(4)

²Entry (i, j) being defined as the word on the arrow or loop from vertex *i* to vertex *j*, etc.

³Variables *x*, *y*, *z* and *t* are taken distinct from the letters in alphabet *A*.

where the coefficient $c_{m,n,p,q,(v_a)_{a\in A}}$ is the total number of graphical sentences $s \in \mathfrak{S}$ having *m* vertices, *n* arrows, *p* loops, a total number of *q* letters, v_a of which are letter a, for each $a \in A$.

Assigning the value 1 to each letter *a* and collecting terms gives

$$\mathbf{w}\mathfrak{S}(x,y,z,t,(1)_{a\in A}) = \sum_{m,n,p,q} b_{m,n,p,q} x^m y^n z^p t^q,$$
(5)

where $b_{m,n,p,q}$ is the number of $s \in \mathfrak{S}$ having *m* vertices, *n* arrows, *p* loops and *q* letters. Letting x = y = z = 1, gives

$$\mathbf{w}\mathfrak{S}(1,1,1,t,(a)_{a\in A}) = \sum_{q,(v_a)_{a\in A}} d_{q,(v_a)_{a\in A}} t^q \Pi a^{v_a},$$
(6)

where $d_{q,(v_a)_{-1}}$ is the number of $s \in \mathfrak{S}$ made of q letters occurring with frequencies $(v_a)_{a \in A}$.

Letting x = y = z = 1 and assigning the value 1 to each letter *a* gives

$$\mathbf{w}\mathfrak{S}(1,1,1,t,(1)_{a\in A}) = \sum_{q} e_{q}t^{q},\tag{7}$$

where e_a is the number of $s \in \mathfrak{S}$ made of q letters.

Let $f_{p,q}$ be the number of graphical sentences $s \in \mathfrak{S}$ made of p words (*i.e.*, p is the total number of arrows and loops in s) and q letters. Then

$$\mathbf{w}\mathfrak{S}(1,z,z,t,(1)_{a\in A}) = \sum_{p,q} f_{p,q} z^p t^q.$$
(8)

Moreover, if we let t = 1 in (8) and if \mathfrak{S} is a finite set, then

$$\mathbf{w}\mathfrak{S}(1,z,z,1,(1)_{a\in A}) = \sum_{p} h_{p} z^{p},$$
(9)

is a polynomial, where h_p is the number of $s \in \mathfrak{S}$ made of p words. Differentiation gives

$$\left| \frac{\frac{\mathrm{d}}{\mathrm{d}z} \mathbf{w} \mathfrak{S}(1, z, z, 1, (1)_{a \in A})}{\mathbf{w} \mathfrak{S}(1, z, z, 1, (1)_{a \in A})} \right|_{z=1} = \frac{\sum_{p} p h_{p}}{\sum_{p} h_{p}} = \text{Expected number of words in a random } s \in \mathfrak{S}.$$
(10)

Of course, a variety of other similar manipulations of the inventory $\mathfrak{S}(x, y, z, t, (a)_{a \in A})$ are possible. In Sec-

tion 2 we apply methods from the theory of species and Pólya theory, to express inventories of general classes of graphical sentences in terms of cycle index series. Section 3 deals with specific classes of graphical sentences: *linear* sentences (corresponding to path-like digraphs) and general *tree-like* sentences (corresponding to classes of tree-like digraphs). We conclude (Section 4) by giving suggestions for possible extensions and generalizations of our results. Various explicit examples are given and to make the text easier to read, the proofs of the main results are collected in Section 5. In a previous paper, [4], we studied the distribution of *runs* in arborescent words.

We assume that the reader is familiar with Pólya theory [3] and with the basic concepts of the theory of combinatorial species [1] [2].

2. Inventory of Graphical Sentences via Cycle Index Series

In order to give a rigorous meaning to the notion of a totally unlabelled digraph and to be able to take into account the possible symmetries within graphical sentences, we must recall first some definitions concerning *labelled* digraphs. A *digraph* on (or labelled by) a finite set V of vertices, a finite set V_1 of arrows, and a finite set V_0 of *loops* is an ordered pair $g = (g_0, g_1)$ of injections,

$$g_0: V_0 \to V, \ g_1: V_1 \to V \times V \setminus \operatorname{diag}(V), \tag{11}$$

where diag $(V) = \{(v, v) : v \in V\}$ and $V \times V \setminus \text{diag}(V) = \{(v, w) : v \in V, w \in V, v \neq w\}$. Given any loop $i_0 \in V_0$ and any vertex $v \in V$, the equality $g_0(i_0) = v$ means that i_0 is a loop at v in the digraph g. Given any arrow

 $j_1 \in V_1$ and any ordered pair of distinct vertices $(v, w) \in V \times V \setminus \text{diag}(V)$, the equality $g_1(j_1) = (v, w)$ means that arrow j_1 is going from v to w in the digraph g. In other words

$$g_0(i_0) = v \Leftrightarrow \underset{i_0}{v} \text{ in } g, \quad g_1(j_1) = (v, w) \Leftrightarrow v \xrightarrow{j_1} w \text{ in } g.$$
(12)

Figure 2(a) shows a digraph on the sets $V = \{1, 2, 3, \dots, 8\}$ of vertices, $V_1 = \{1, 2, 3, \dots, 12_1\}$ of arrows and $V_0 = \{1_0, 2_0\}$ of loops, with loop 1_0 at vertex 4 and loop 2_0 at vertex 1. Let $g = (g_0, g_1)$ be a digraph on (V, V_1, V_0) and $g' = (g'_0, g'_1)$ be a digraph on (V', V'_1, V'_0) . An *isomorphism* $\theta: g \to g'$ from digraph g to digraph g', is an ordered triple, $\theta = (\alpha, \beta, \gamma)$, of bijections $\alpha: V \to V'$, $\beta: V_1 \to V'_1$, $\gamma: V_0 \to V'_0$, such that

$$\underset{i_{0}}{\overset{\circ}{_{i_{0}}}} \inf g \Leftrightarrow \alpha(v) \inf g', \tag{13}$$

$$v \xrightarrow{j_1} w \text{ in } g \Leftrightarrow \alpha(v) \xrightarrow{\beta(j_1)} \alpha(w) \text{ in } g'.$$
 (14)

If g = g' then θ is called an *automorphism* (or symmetry) of the digraph g. For example, the triple of permutations (α, β, γ) , defined (in cyclic notation) by

$$\alpha = (3,8)(7,6), \quad \beta = (4_1, 12_1)(9_1, 1_1)(2_1, 5_1), \quad \gamma = \mathrm{id}, \tag{15}$$

is an automorphism of the digraph of **Figure 2(a)**. Note that labelled digraphs are elastic and not considered as embedded in the plane. Only the incidence relations between vertices, arrows and loops are taken into account. A *totally unlabelled digraph* (see **Figure 2(b)**), is simply an isomorphism class of labelled digraphs. The class of a labelled digraph g is denoted [g]; so that [g] is a totally unlabelled digraph with representative g.

We now give a rigorous definition of the notion of a graphical sentence.

Definition 2.1. Let *A* be a finite alphabet and $A^+ = A + A^2 + A^3 + \cdots$, be the set of non-empty words over *A*. A *graphical sentence over A* is an equivalence class, *s*, of ordered triples, (g, σ_0, σ_1) , where *g* is a connected labelled digraph on (V, V_1, V_0) , and $\sigma_0 : V_0 \to A^+$, $\sigma_1 : V_1 \to A^+$ are arbitrary functions assigning a non-empty word to each loop and each arrow of *g*. Two such triples (g, σ_0, σ_1) and $(g', \sigma'_0, \sigma'_1)$ being equivalent if there exists an isomorphism $\theta = (\alpha, \beta, \gamma) : g \to g'$ such that $\sigma'_0 \circ \gamma = \sigma_0$ and $\sigma'_1 \circ \beta = \sigma_1$. We write $s = [(g, \sigma_0, \sigma_1)]$ to mean that *s* is a graphical sentence with representative (g, σ_0, σ_1) .

Now take any species \mathcal{G} of connected labelled digraphs⁴. Our goal is to compute the inventory (3) of the class \mathfrak{S} of all graphical sentences $s = [(g, \sigma_0, \sigma_1)]$ where $g \in \mathcal{G}$. To emphasize the fact that digraphs are made of three sorts of elements, *vertices*, *arrows* and *loops*, the given species \mathcal{G} of digraphs can be written in the form $\mathcal{G} = \mathcal{G}(X, Y, Z)$, where X is the sort of vertices, Y is the sort of arrows, and Z is the sort of loops⁵. Any digraph $g \in \mathcal{G}$ is called a \mathcal{G} -structure for short.

Following standard notations from the theory of species, the *set* of all \mathcal{G} -structures on a set V of vertices, a set V_1 of arrows and a set V_0 of loops is denoted by $\mathcal{G}[V, V_1, V_0]$ (note the square brackets). Given bijections $\alpha: V \rightarrow V', \beta: V_1 \rightarrow V_1', \gamma: V_0 \rightarrow V_0'$, we denote by

$$\mathcal{G}[\alpha,\beta,\gamma]:\mathcal{G}[V,V_1,V_0] \xrightarrow{\sim} \mathcal{G}[V',V_1',V_0'], \tag{16}$$

the bijection $(g \mapsto g')$ that *transforms* (or *transports*) each digraph $g \in \mathcal{G}[V, V_1, V_0]$ into a corresponding isomorphic digraph $g' \in \mathcal{G}[V', V'_1, V'_0]$, as described above. Note that if V = V', $V_1 = V'_1$ and $V_0 = V'_0$, then $\mathcal{G}[\alpha, \beta, \gamma]$ is a *permutation* of the set $\mathcal{G}[V, V_1, V_0]$ (*i.e.*, a bijection of $\mathcal{G}[V, V_1, V_0]$ into itself).

Many power series can be associated to any species \mathcal{F} . An important one is the Pólya-Joyal *cycle index series* $\mathbb{Z}_{\mathcal{F}}$. In the context of a species $\mathcal{G} = \mathcal{G}(X,Y,Z)$ of digraphs, this is a power series in a triple infinity of variables, $x_1, x_2, x_3, \dots; y_1, y_2, y_3, \dots; z_1, z_2, z_3, \dots$, defined by

$$\mathbf{Z}_{\mathcal{G}} = \mathbf{Z}_{\mathcal{G}} \left(x_1, x_2, x_3, \cdots; y_1, y_2, y_3, \cdots; z_1, z_2, z_3, \cdots \right)$$
(17)

$$= \sum_{m,n,p} \frac{1}{m!n!p!} \sum_{\alpha \in S_m, \beta \in S_n, \gamma \in S_p} \operatorname{fix} \mathcal{G}[\alpha, \beta, \gamma] x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} \cdots y_1^{\beta_1} y_2^{\beta_2} y_3^{\beta_3} \cdots z_1^{\gamma_1} z_2^{\gamma_2} z_3^{\gamma_3} \cdots,$$
(18)

⁴This means that \mathcal{G} is a class of connected digraphs on arbitrary finite sets, V of vertices, V_1 of arrows and V_0 of loops, which is closed under arbitrary isomorphisms.

⁵More precisely, X is the species of *singleton vertices*, Y is the species of *singleton arrows* and Z is the species of *singleton loops*.



Figure 2. (a) A labelled digraph g; (b) [g] = totally unlabelled g.

where, for each permutations $\alpha \in S_m$, $\beta \in S_n$ and $\gamma \in S_p$, fix $\mathcal{G}[\alpha, \beta, \gamma]$ is the number⁶ of all digraphs $g \in \mathcal{G}[V, V_1, V_0]$, on the sets $V = \{1, 2, \dots, m\}$ of vertices, $V_1 = \{1_1, 2_1, \dots, n_1\}$ of arrows and $V = \{1_0, 2_0, \dots, p_0\}$ of loops, for which (α, β, γ) is an automorphism (*i.e.*, fix $\mathcal{G}[\alpha, \beta, \gamma]$ is the cardinality (or total weight) of the set of fixed points of the permutation $\mathcal{G}[\alpha, \beta, \gamma]$). For a permutation $\sigma \in S_k$, the notation σ_i is used to denote the number of cycles⁷ of length *i* in the cyclic decomposition of σ .

The sequence of integers $(\sigma_1, \sigma_2, \sigma_3, \cdots)$ is called the *cyclic type of* σ and it is well known that the number of $\sigma \in S_k$ having cyclic type $\mathbf{k} = (k_1, k_2, k_3, \cdots)$, with $k_1 + 2k_2 + 3k_3 + \cdots = k$, is

$$\frac{k!}{\mathbf{k}\downarrow}, \quad \text{where } \mathbf{k}\downarrow = 1^{k_1} k_1! 2^{k_2} k_2! 3^{k_3} k_3! \cdots.$$
(19)

Note that each sequence **k** has a finite number of nonzero terms and will be considered, in the present text, as a finite sequence with $c = c(\mathbf{k}) = \max\{j: k_j \neq 0\}$ components. By this convention, **k** can be viewed as a partition of the integer $k = k_1 + 2k_2 + \dots + ck_c$, having k_i parts of length *i*, for $i = 1, 2, \dots, c$ and we use the classical notation⁸

$$\mathbf{k} \vdash k, \tag{20}$$

to express this fact. Note also that $\operatorname{fix} \mathcal{G}[\alpha, \beta, \gamma]$, in (18), only depends on the cyclic types of the three permutations α, β, γ . Hence, every three permutations α, β, γ having given cyclic types $\mathbf{m} = (m_1, m_2, m_3, \cdots)$, $\mathbf{n} = (n_1, n_2, n_3, \cdots)$, $\mathbf{p} = (p_1, p_2, p_3, \cdots)$ contribute to the *same* monomial

$$\mathbf{x}^{m} \mathbf{y}^{n} \mathbf{t}^{\mathbf{p}} = x_{1}^{m_{1}} x_{2}^{m_{2}} x_{3}^{m_{3}} \cdots y_{1}^{n_{1}} y_{2}^{n_{2}} y_{3}^{n_{3}} \cdots z_{1}^{p_{1}} z_{2}^{p_{2}} z_{3}^{p_{3}} \cdots,$$
(21)

in (18). In order to eliminate this redundancy, we regroup monomials which correspond to each of these types, and taking (19) into account we obtain the following more compact variant expression for the cycle index series of \mathcal{G} :

$$\mathbf{Z}_{\mathcal{G}} = \mathbf{Z}_{\mathcal{G}}\left(\mathbf{x}, \mathbf{y}, \mathbf{z}\right) = \sum_{\mathbf{m}, \mathbf{n}, \mathbf{p}} \operatorname{fix} \mathcal{G}\left[\mathbf{m}, \mathbf{n}, \mathbf{p}\right] \frac{\mathbf{x}^{\mathbf{m}}}{\mathbf{m} \downarrow} \frac{\mathbf{y}^{\mathbf{n}}}{\mathbf{n} \downarrow} \frac{\mathbf{z}^{\mathbf{p}}}{\mathbf{p} \downarrow},$$
(22)

in which each monomial appears only once and fix $\mathcal{G}[\mathbf{m},\mathbf{n},\mathbf{p}]$ is the number (or total weight) of all \mathcal{G} -structures for which any given (α,β,γ) of types $(\mathbf{m},\mathbf{n},\mathbf{p})$ is an automorphism.

The following proposition is a consequence of general principles from the theory of species and Pólya theory. It shows that the computation of \mathbf{Z}_{g} is an essential step in the determination of the inventory series (3) of classes of graphical sentences arising from digraphs $g \in \mathcal{G}$.

Proposition 2.1. Let $\mathcal{G} = \mathcal{G}(X, Y, Z)$ be any species of connected digraphs, let $\mathbb{L} \subseteq \mathbb{N}^+$, and let

 $\mathfrak{S} = \mathfrak{S}_{A,\mathcal{G},\mathbb{L}}$ be the set of all graphical sentences over alphabet A arising from \mathcal{G} , the word on each arrow or loop having a length $\in \mathbb{L}$. Then, the inventory of \mathfrak{S} is given by

$$\mathbf{w}\mathfrak{S}(x, y, z, t, (a)_{a \in A}) = \mathbf{Z}_{\mathcal{G}}(x, x^2, x^3, \dots; \lambda_1 y, \lambda_2 y^2, \lambda_3 y^3, \dots; \lambda_1 z, \lambda_2 z^2, \lambda_3 z^3, \dots),$$
(23)

⁶Or total weight, in the case of weighted digraphs.

⁷Not to be confused with $\sigma(i)$ which is the image of *i* under σ .

⁸In particular, the empty sequence, **0**, has no component and satisfies $\mathbf{0} \vdash \mathbf{0}$.

where $\lambda_k = \Lambda(t^k \Sigma a^k)$, $k = 1, 2, 3, \dots, \Lambda(u) = \sum_{i \in \mathbb{Z}} u^i$, and $\Sigma a^k = \sum_{a \in \mathbb{A}} a^k$ is the formal k-th power sum of the letters of the alphabet.

Proof. See Section 5.

Making use of the compact expression (22) for the cycle index series and collecting terms, inventory (23) can be rewritten in the following more explicit form.

Corollary 2.2. The inventory, $\mathbf{w} \hat{\mathfrak{S}} = \mathbf{w} \mathfrak{S}(x, y, z, t, (a)_{a \in A})$, of the set $\mathfrak{S} = \mathfrak{S}_{A,\mathcal{G},\mathbb{L}}$ of all graphical sentences arising from a species $\mathcal{G} = \mathcal{G}(X, Y, Z)$ of digraphs is given by

$$\mathbf{w}\mathfrak{S} = \sum_{m \ge 0, n \ge 0, p \ge 0} \omega_{m,n,p} \left(\lambda_1, \lambda_2, \lambda_3, \cdots \right) x^m y^n z^p,$$
(24)

where $\lambda_k = \Lambda(t^k \Sigma a^k)$, $k = 1, 2, 3, \dots, \Lambda(u) = \sum_{i \in \mathbb{T}} u^i$, and

$$\omega_{m,n,p}\left(\lambda_{1},\lambda_{2},\lambda_{3},\cdots\right) = \sum_{\mathbf{m}\vdash m,\mathbf{n}\vdash n,\mathbf{p}\vdash p} \frac{\operatorname{fix}\mathcal{G}[\mathbf{m},\mathbf{n},\mathbf{p}]}{\mathbf{m}\downarrow\mathbf{n}\downarrow\mathbf{p}\downarrow}\lambda_{1}^{n_{1}+p_{1}}\lambda_{2}^{n_{2}+p_{2}}\lambda_{3}^{n_{3}+p_{3}}\cdots.$$
(25)

Note. If $\mathbb{L} = \mathbb{N}^+$ in Proposition 2.1, then $\Lambda(u) = u/(1-u)$ and no restrictions are put on the lengths of the words in inventory (23). If $\mathbb{L} = \{1, 3, 5, \cdots\}$, then $\Lambda(u) = u/(1-u^2)$ and the lengths of the words are all odd. If $\mathbb{L} = \{1, 2, \cdots, N\}$, then $\Lambda(u) = \sum_{k=1}^{N} u^k = u(1-u^N)/(1-u)$ and the lengths of the words are bounded by *N*, etc.

3. Analysis of Classes of Tree-Like Sentences

As shown in the preceding section, the computation of the inventory of a class \mathfrak{S} of graphical sentences can be reduced to the computation of the cycle index series \mathbf{Z}_{g} provided that \mathfrak{S} arises from a 3-sort species $\mathcal{G} = \mathcal{G}(X, Y, Z)$ of connected digraphs. However, the explicit or recursive computation of the cycle index series of most species of graphical structures is a very difficult (or intractable) task. For example, even in the ordinary one-sort case9, the complete cycle index series of the species of all ordinary plane digraphs and all transitive di*graphs* are still unknown¹⁰.

For this reason, we focus our study on the following basic classes of graphical sentences:

- 1) *Linear sentences* (arising from the species of path-shaped digraphs).
- 2) General tree-like sentences (arising from various species of tree-like digraphs).

Note that linear sentences are special kinds of tree-like sentences. Due to their close relationship with ordinary sentences, we have chosen to present first a separate subsection devoted to their study. Our methods will use the fact that species of tree-like digraphs can be built from simpler species by making use of basic combinatorial operations and that cycle index series behave well with respect to these operations. For example, if F, Gand H are species, then

$$\mathbf{Z}_{F+G} = \mathbf{Z}_F + \mathbf{Z}_G, \ \mathbf{Z}_{F\cdot G} = \mathbf{Z}_F \cdot \mathbf{Z}_G, \ \mathbf{Z}_{F(G)} = \mathbf{Z}_F \circ \mathbf{Z}_G, \ \mathbf{Z}_{F(G,H)} = Z_F \circ (Z_G, Z_H), \ \text{etc},$$
(26)

where \circ denotes the classical plethystic substitution of cycle index series (see [1]).

3.1. Linear Sentences

We say that a digraph $g = (g_0, g_1)$ is *path-shaped* if its underlying simple graph is a simple path. A graphical sentence is *linear* if it comes from a path-shaped digraph. Figure 3 shows a path-shaped digraph, together with its underlying simple path and a linear sentence over alphabet $A = \{a,b,c\}$. Note that a path-shaped digraph can have non-trivial automorphisms. For example, the 180° rotation, (α, β, γ) , where the cyclic decompositions of the permutations α, β, γ are given by

$$\alpha = (3,8)(5,2)(6,4)(1,9), \ \beta = (3_0,4_0)(5_0,1_0), \tag{27}$$

$$\gamma = (7_1, 3_1)(9_1, 4_1)(1_1, 12_1)(5_1, 6_1)(10_1, 2_1)(8_1, 11_1),$$
(28)

is an automorphism of the path-shaped digraph of Figure 3.

⁹Where only the vertices are labelled.

¹⁰However, for the 3-sort species Dig(X,Y,Z) of *all* digraphs, \mathbf{Z}_{Dig} can be computed explicitly (see Section 4).

Special kinds of linear sentences over an alphabet include *ordinary sentences* (Figure 4 top), corresponding to *directed paths without loops*, and *ordinary sentences with* (*possible*) *loops* (Figure 4 bottom), corresponding to *directed paths with* (*possible*) *loops*.

For example, the sentence

MY TAYLOR IS RICH RICH AND MY COUSIN IS POOR POOR (29)

is one of the readable sentences in Figure 4 bottom.

Proposition 3.1 Let \mathcal{O} be the set of all ordinary sentences, \mathcal{O}° , the set of all ordinary sentences with loops, \mathfrak{L} , the set of all linear sentences without loops, and \mathfrak{L}° , the set of all linear sentences with loops over an alphabet A and a set $\mathbb{L} \subseteq \mathbb{N}^+$ of allowed word-lengths. Then, the following inventories hold

$$\mathbf{w}\mathcal{O} = \frac{x}{1 - \lambda_1 xy}, \ \mathbf{w}\mathcal{O}^{\circ} = \frac{(1 + \lambda_1 z)x}{1 - (1 + \lambda_1 z)\lambda_1 xy},$$
(30)

$$\mathbf{w}\mathfrak{L} = \frac{1}{2} \left(\frac{x}{1 - (2 + \lambda_1 y)\lambda_1 x y} + \frac{x + \lambda_2 x^2 y^2}{1 - (2 + \lambda_2 y^2)\lambda_2 x^2 y^2} \right),$$
(31)

$$\mathbf{w}\mathfrak{L}^{\circ} = \frac{1}{2} \left(\frac{(1+\lambda_{1}z)x}{1-(1+\lambda_{1}z)(2+\lambda_{1}y)\lambda_{1}xy} + \frac{(1+\lambda_{1}z)x+(1+\lambda_{2}z^{2})\lambda_{2}x^{2}y^{2}}{1-(1+\lambda_{2}z^{2})(2+\lambda_{2}y^{2})\lambda_{2}x^{2}y^{2}} \right),$$
(32)

where, $\lambda_1 = \Lambda(t\Sigma a), \lambda_2 = \Lambda(t^2\Sigma a^2), \quad \Lambda(u) = \sum_{i\in\mathbb{L}} u^i$. Proof. See Section 5.

In view of (30)-(32), the alternate general inventory formula (24) in the case of any class of *linear* sentences, does not involve $\lambda_3, \lambda_4, \cdots$. We have the following explicit expansions.

Corollary 3.2. For the classes $\mathcal{O}, \mathcal{O}^{\circ}, \mathfrak{L}, \mathfrak{L}^{\circ}$ of linear sentences, we have

$$\mathbf{w}\mathcal{O}^{\circlearrowright} = \sum_{m \ge 1, 0 \le p \le m} \binom{m}{p} \lambda_1^{m+p-1} x^m y^{m-1} z^p,$$
(33)

$$\mathbf{w}\mathcal{O} = \mathbf{w}\mathcal{O}^{\circ}\Big|_{z=0},\tag{34}$$

$$\mathbf{w}\mathfrak{L}^{\circlearrowright} = \sum_{(m,n,p)\in W} \left(a_{m,n,p} \lambda_1^{n+p} + b_{m,n,p} \lambda_2^{\frac{n+p}{2}} + c_{m,n,p} \lambda_1 \lambda_2^{\frac{n+p-1}{2}} \right) x^m y^n z^p,$$
(35)



Figure 3. Path-shaped digraph, its underlying simple path and a linear sentence.



Figure 4. An ordinary sentence and an ordinary sentence with loops.

$$\mathbf{w}\mathfrak{L} = \mathbf{w}\mathfrak{L}^{\circ}\Big|_{z=0},\tag{36}$$

where

$$W = \left\{ (m, n, p) \middle| m \ge 1, m - 1 \le n \le 2(m - 1), 0 \le p \le m \right\},\tag{37}$$

$$a_{m,n,p} = \binom{m-1}{n-m+1} \binom{m}{p} 2^{2m-n-3},$$
(38)

$$b_{m,n,p} = \left(\frac{\frac{m-1}{2}}{\frac{n-m+1}{2}}\right) \left(\frac{\frac{m-1}{2}}{\frac{p}{2}}\right) 2^{\frac{m-n}{2}-2} + \left(\frac{\frac{m}{2}-1}{\frac{n-m}{2}}\right) \left(\frac{\frac{m}{2}}{\frac{p}{2}}\right) 2^{\frac{m-n}{2}-2},$$
(39)

$$c_{m,n,p} = \left(\frac{\frac{m-1}{2}}{\frac{n-m+1}{2}}\right) \left(\frac{\frac{m-1}{2}}{\frac{p-1}{2}}\right) 2^{\frac{m-n}{2}-2},$$
(40)

using the convention $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 0$ if α and β are not both integers.

Proof. (Sketch) Formulas (34) and (36) are immediate since no loops are involved. Careful computations, starting from (30) and (32) using geometric series, the binomial theorem, and manipulation or indices lead to expansions (33) and (35).

Sample of explicit examples of computations.

Example 3.1. General shape of the inventory of the class $w \mathfrak{L}^{\circ}$.

The first few terms of $\mathbf{w}\mathfrak{L}^{\circ}$ read as follows

$$\begin{split} \mathbf{w}\mathcal{L}^{\circlearrowright} &= x + \lambda_{1}xz + \lambda_{1}x^{2}y + 2\lambda_{1}^{2}x^{2}yz + \lambda_{1}^{3}x^{2}yz^{2} + \left(\frac{1}{2}\lambda_{1}^{2} + \frac{1}{2}\lambda_{2}\right)x^{2}y^{2} + \lambda_{1}^{3}x^{2}y^{2}z \\ &+ \left(\frac{1}{2}\lambda_{1}^{4} + \frac{1}{2}\lambda_{2}^{2}\right)x^{2}y^{2}z^{2} + \left(2\lambda_{1}^{2} + \lambda_{2}\right)x^{3}y^{2} + \left(6\lambda_{1}^{3} + \lambda_{1}\lambda_{2}\right)x^{3}y^{2}z \\ &+ \left(6\lambda_{1}^{4} + \lambda_{2}^{2}\right)x^{3}y^{2}z^{2} + \left(2\lambda_{1}^{5} + \lambda_{1}\lambda_{2}^{2}\right)x^{3}y^{2}z^{3} + 2\lambda_{1}^{3}x^{3}y^{3} + 6\lambda_{1}^{4}x^{3}y^{3}z \\ &+ \left(6\lambda_{1}^{5}x^{3}y^{3}z^{2} + 2\lambda_{1}^{6}x^{3}y^{3}z^{3} + \left(\frac{1}{2}\lambda_{1}^{4} + \frac{1}{2}\lambda_{2}^{2}\right)x^{3}y^{4} + \left(\frac{3}{2}\lambda_{1}^{5} + \frac{1}{2}\lambda_{1}\lambda_{2}^{2}\right)x^{3}y^{4}z \\ &+ \left(\frac{3}{2}\lambda_{1}^{6} + \frac{1}{2}\lambda_{2}^{3}\right)x^{3}y^{4}z^{2} + \left(\frac{1}{2}\lambda_{1}^{7} + \frac{1}{2}\lambda_{1}\lambda_{2}^{3}\right)x^{3}y^{4}z^{3} + \cdots. \end{split}$$

$$(41)$$

The coefficients of $x^m y^n z^p$ in (41) are polynomials in λ_1 and λ_2 each having one or two terms, despite the fact that (35) suggests three terms. This is true for every m, n, p since $b_{m,n,p}$ and $c_{m,n,p}$ cannot be both non zero, due to the fact that $\frac{p-1}{2}$ and $\frac{p}{2}$ cannot be both integral in (39) and (40).

Example 3.2. Counting linear sentences with given parameters.

Corollary 3.2 is particularly useful when one wants to compute an individual term in the inventory of linear sentences. For example, consider the 3-letter alphabet $A = \{a,b,c\}$ and take $\mathbb{L} = \mathbb{N}^+$. This means that we impose no restrictions on the lengths of the words that are assigned to each arrow or loop in linear sentences. Suppose that we want to know the number of such linear sentences having m = 7 vertices, n = 10 arrows, p = 5 loops which are made of 7 times the letter a, 6 times the letter b and 12 times the letter c. In this case, the coefficient of $x^7 y^{10} z^5$ is given by

$$2\binom{6}{4}\binom{7}{5}\lambda_{1}^{15} + \binom{3}{2}\binom{3}{2}\lambda_{1}\lambda_{2}^{7} = 630\lambda_{1}^{15} + 9\lambda_{1}\lambda_{2}^{7},$$
(42)

where

$$\lambda_{1} = \frac{t(a+b+c)}{1-t(a+b+c)}, \quad \lambda_{2} = \frac{t^{2}(a^{2}+b^{2}+c^{2})}{1-t^{2}(a^{2}+b^{2}+c^{2})}.$$
(43)

The required number of linear sentences equals the coefficient of $t^{25}a^7b^6c^{12}$ in (42). Using Maple, this number is equal to

If we take $\mathbb{L} = \{1, 2\}$, then the words that are assigned to each arrow or loop in linear sentences are of length 1 or 2. In this case, (43) is replaced by

$$\lambda_{1} = t(a+b+c) + t^{2}(a+b+c)^{2}, \quad \lambda_{2} = t^{2}(a^{2}+b^{2}+c^{2}) + t^{4}(a^{2}+b^{2}+c^{2})^{2}$$
(45)

and (44) goes down to 16882686796464720.

Example 3.3. Manipulating the inventory of linear sentences.

As we have seen in Section 1, manipulations of inventories (specialization of variables, expansions, etc) can be made to analyze various parameters in graphical sentences. Proposition 3.1 is generally more suitable for such manipulations than Corollary 3.2. For example, let ℓ be the number of letters in alphabet A, $\mathbb{L} = \mathbb{N}^+$, and assign the value 1 to x, y and to each letter $a \in A$ in **w** \mathcal{L} given by (31). Then $\lambda_1 = \ell t/(1-\ell t)$, $\lambda_2 = \ell t^2/(1-\ell t^2)$, and, by (7)

$$\frac{1}{2} \left(\frac{\left(1 - \ell t\right)^2}{1 - 4\ell t + 2\ell^2 t^2} + \frac{1 - \ell t^2}{1 - 4\ell t^2 + 2\ell^2 t^4} \right) = \sum_q e_q t^q, \tag{46}$$

where e_q is the number of linear sentences without loops that are made of q letters. Note that (46) is a rational function of t, so that the sequence $(e_q)_{q\geq 0}$ satisfies a linear recurrence with constant coefficients and the asymptotic expansion of e_q , as $q \to \infty$, can be established using standard classical methods. The first few terms in expansion (46) are given by

$$\sum e_{q}t^{q} = 1 + \ell t + \frac{1}{2} (7\ell^{2} + 3\ell)t^{2} + 12\ell^{3}t^{3} + (41\ell^{4} + 5\ell^{2})t^{4} + 140\ell^{5}t^{5} + (478\ell^{6} + 17\ell^{3})t^{6} + 1632\ell^{7}t^{7} + (5572\ell^{8} + 58\ell^{4})t^{8} + 19024\ell^{9}t^{9} + (64952\ell^{10} + 198\ell^{5})t^{10} + 221760\ell^{11}t^{11} + (757136\ell^{12} + 676\ell^{6})t^{12} + \cdots$$

$$(47)$$

Example 3.4. Fibonacci numbers versus ordinary sentences with loops.

Consider now the class \mathcal{O}° of ordinary sentences with possible loops. Let y = z and assign the value 1 to x and to each letter $a \in A$ in $\mathbf{w}\mathcal{O}^{\circ}$ given by (30). Then we have

$$\mathbf{w}\mathcal{O}^{\circ}(1,z,z,t,(1)_{a\in A}) = \frac{1+\lambda_{1}z}{1-\lambda_{1}z-\lambda_{1}^{2}z^{2}} = \sum_{p\geq 0} F_{p+2}\lambda_{1}^{p}z^{p} = \sum_{p,q\geq 0} f_{p,q}z^{p}t^{q},$$
(48)

where F_k are the Fibonacci numbers¹¹ and $f_{p,q}$ is the number of ordinary sentences with possible loops made of p words and q letters. Now, fix $q \ge 1$ and consider the *finite* class $\mathcal{O}_q^{\circlearrowright}$ of all $s \in \mathcal{O}^{\circlearrowright}$ made of q letters. Since, $\lambda_1 = \ell t / (1 - \ell t)$, then collecting the coefficient of t^q in (48) we have the polynomial inventory

$$\mathbf{w}\mathcal{O}_{q}^{\bigcirc}(1,z,z,1,(1)_{a\in A}) = \sum_{1\leq p\leq q} f_{p,q} z^{p}, \ f_{p,q} = \binom{q-1}{p-1} F_{p+2} \ell^{q}.$$
(49)

Finally, making use of (10) and invoking Binet's formula

$$F_k = \frac{1}{\sqrt{5}} \left(\phi^k - \psi^k \right), \tag{50}$$

 $\overline{ {}^{11} F_0 = 0, F_1 = 1, F_k = F_{k-1} + F_{k-2} , \text{ for } k \ge 2 \text{ and } z/(1-z-z^2) = \sum_{k\ge 0} F_k z^k .$

where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden number, $\psi = \frac{1-\sqrt{5}}{2} = -1/\phi$, we find that the expected number of words in a random ordinary sentence with possible loops made of $q \ge 1$ letters is

$$1 + \left(q - 1\right) \frac{F_{2q}}{F_{2q+1}} \sim \frac{q}{\phi}, \ q \to \infty.$$

$$\tag{51}$$

The reader can check that if we do not allow loops, then (51) is replaced by $\frac{1}{2}(q+1)$.

A multitude of other similar examples can be obtained using Proposition 3.1 and Corollary 3.2.

3.2. General Tree-Like Sentences

We say that a digraph $g = (g_0, g_1)$ is *tree-like* if its underlying simple graph is a simple tree or a simple rooted tree. A graphical sentence is tree-like if it comes from a tree-like digraph. Figure 5 shows a tree-like digraph, its underlying simple tree and a tree-like sentence over alphabet $A = \{0,1\}$.

The tree-like structures of Figure 5 are free in the sense that they are not restricted to be embedded in the plane and no other constraints are assumed on the vertices, arrows and loops. More generally, by allowing such constraints, one can consider, for example, the above linear sentences (see Figure 3 and Figure 4), one way free binary rooted tree sentences (see Figure 6 left), one way free full binary rooted tree sentences (see Figure 6 right), plane tree sentences (see Figure 5 right) where, this time, the underlying tree is considered as being embedded in the plane), etc. We shall deal with these cases in a uniform manner by adding extra structure on the underlying trees or rooted trees. More precisely, the underlying trees or rooted trees will be *enriched* according to the following definition.

Definition 3.1. [5] Let R = R(X) be any given one-sort species.

1) A R-enriched rooted tree is a rooted tree in which the set of immediate descendants (away from the root) of every vertex is equipped with a *R*-structure (see Figure 7 left, in which each *dotted arc* represents a *R*-structure).



Figure 5. A tree-like digraph, its underlying simple tree, a tree-like sentence over $A = \{0, 1\}$



Figure 6. A one way free binary rooted tree sentence and a full one.



Figure 7. A R-enriched rooted tree and a R-enriched tree.

2) A *R*-enriched tree is a tree in which the set of immediate neighbors of each vertex is equipped with a *R*-structure (see Figure 7 right, in which each dotted circle represents a *R*-structure).

Lemma 3.3 [1] The species $A_R = A_R(X)$ of *R*-enriched rooted trees is characterized recursively by the combinatorial equation

$$A_{R} = XR(A_{R}), \tag{52}$$

and the species $a_R = a_R(X)$ of *R*-enriched trees satisfies the combinatorial equality¹²

$$XR(A_{R'}) + E_2(A_{R'}) = a_R + A_{R'}^2,$$
(53)

where $A_{R'} = A_{R'}(X)$ is the species of R'-enriched rooted trees (R' being the combinatorial derivative of the species R).

It is easy to see that the species of ordinary rooted trees (resp. ordinary trees) corresponds to the species A_R (resp. a_R) with the choice R = E, the species of all finite sets. The species of binary rooted trees (resp. full binary rooted trees) corresponds to the species A_R with the choice $R = 1 + X + E_2$ (resp. $R = 1 + E_2$), where 1 denotes, as usual, the species of the empty set and E_2 , the species of 2-element sets. The species of all plane trees corresponds to the species a_R with the choice R = 1 + C, where C is the species of cyclic permutations (see Example 3.8 below), etc.

For the computation of the inventories of various classes of enriched tree-like graphical sentences, we will make use of the following 3-sort extension of Lemma 3.3 which includes a new extension, (56) below, of the dissymmetry formula (53).

Lemma 3.4. The species $\vec{\mathcal{A}}_{R}^{\circ} = \vec{\mathcal{A}}_{R}^{\circ}(X,Y,Z)$ of one-way *R*-enriched rooted trees and $\mathcal{A}_{R}^{\circ} = \mathcal{A}_{R}^{\circ}(X,Y,Z)$ of *R*-enriched rooted trees on the sorts *X* of vertices, *Y* of arrows and *Z* of loops are characterized recursively by the combinatorial equations

a)
$$\vec{\mathcal{A}}_{R}^{\odot} = (1+Z) XR(Y\vec{\mathcal{A}}_{R}^{\odot}), \text{ b) } \mathcal{A}_{R}^{\odot} = (1+Z) XR(\Omega \mathcal{A}_{R}^{\odot}),$$
 (54)

where $\Omega = \Omega(Y) = 2Y + Y^2$. They can also be expressed explicitly in terms of the 1-sort species $A_R(X)$ as follows

a)
$$\vec{\mathcal{A}}_{R}^{\circ} = \frac{1}{Y} A_{R} \left(\left(1 + Z \right) XY \right), \text{ b) } \mathcal{A}_{R}^{\circ} = \frac{1}{\Omega} A_{R} \left(\left(1 + Z \right) X\Omega \right).$$
 (55)

The species $\mathfrak{a}_R^{\circ} = \mathfrak{a}_R^{\circ}(X,Y,Z)$ of *R*-enriched trees on sorts *X* of vertices, *Y* of arrows and *Z* of loops satisfies the combinatorial equality (extended dissymmetry formula)

$$(1+Z)XR(\Omega\mathcal{A}_{R'}^{\odot})+E_{2}(Y\mathcal{A}_{R'}^{\odot})=\mathfrak{a}_{R}^{\odot}+(Y+Y^{2})\cdot(\mathcal{A}_{R'}^{\odot})^{2},$$
(56)

where $\mathcal{A}_{R'}^{\odot} = \mathcal{A}_{R'}^{\odot}(X,Y,Z)$ is the species of R'-enriched rooted trees on sorts X, Y, Z.

Proof. See Section 5.

In our analysis of tree-like graphical sentences, we will use of the following useful compact "plethystic nota-

¹²Equality (53) is called the *dissymmetry formula for trees*. It is due P. Leroux [6] for R = E and to the first author [7] for general R.

tion" which is classical in the theory of species and cycle index series.

Notation 3.5. Let $\mathbf{S} = \mathbf{S}(v_1, v_2, v_3, \cdots)$ be a (formal) power series in the variables v_1, v_2, v_3, \cdots . For any integer $k \ge 1$, \mathbf{S}_k denotes the series \mathbf{S} in which each variable is raised to the power k:

$$\mathbf{S}_{k} = \mathbf{S}_{k} \left(v_{1}, v_{2}, v_{3}, \cdots \right) = \mathbf{S} \left(v_{1}^{k}, v_{2}^{k}, v_{3}^{k}, \cdots \right).$$

$$(57)$$

In particular, $\mathbf{S}_1 = \mathbf{S}$. Furthermore, given power series,

$$\mathbf{f} = \mathbf{f}(s_1, s_2, s_3, \cdots), \ \mathbf{g} = \mathbf{g}(s_1, s_2, s_3, \cdots; t_1, t_2, t_3, \cdots), \ etc,$$
(58)

$$\mathbf{S} = \mathbf{S}(v_1, v_2, v_3, \cdots), \ \mathbf{T} = \mathbf{T}(v_1, v_2, v_3, \cdots), \ etc,$$
(59)

then f[[S]], g[[S;T]], etc, denote the series

$$\mathbf{f}[[\mathbf{S}]] = f(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \cdots), \ \mathbf{g}[[\mathbf{S}; \mathbf{T}]] = g(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \cdots; \mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \cdots), \ etc,$$
(60)
$$v_1, v_2, v_2, \cdots.$$

in the variables v_1, v_2, v_3, \cdots .

For example, taking the variables $x, y, z, t, (a)_{a \in A}$ for the v_i 's, then formula (23) of Proposition 2.1 takes the compact form

$$\mathbf{w}\mathfrak{S}(x, y, z, t, (a)_{a \in A}) = \mathbf{Z}_{\mathcal{G}}[[x; \lambda y; \lambda z]],$$
(61)

where $\lambda = \lambda_1 = \Lambda(t\Sigma a)$, $\Lambda(u) = \sum_{i \in \mathbb{L}} u^i$, and $\Sigma a = \sum_{a \in A} a$ is the formal sum of the letters in *A*. We now describe how to compute the inventory of classes of tree-like sentences.

Proposition 3.6 *Given an arbitrary species* R = R(X), *let*

 $\vec{\mathfrak{T}}_{R}^{\circ}$ be the set of one way *R*-enriched rooted trees sentences with possible loops, (62)

 \mathfrak{T}_{R}° be the set of *R*-enriched rooted trees sentences with possible loops, (63)

 $\mathfrak{t}_{R}^{\bigcirc}$ be the set of *R*-enriched trees sentences with possible loops, (64)

over an alphabet *A* and a set $\mathbb{L} \subseteq \mathbb{N}^+$ of allowed word-lengths. Then, using *Notation* 3.5, the inventories $\mathbf{w} \hat{\mathfrak{T}}_R^{\odot}$ and $\mathbf{w} \mathfrak{T}_R^{\odot}$ can be computed recursively as follows

$$\mathbf{w}\vec{\mathfrak{I}}_{R}^{\odot} = (1+\lambda z) x Z_{R} \left[\left[\lambda y \left(\mathbf{w}\vec{\mathfrak{I}}_{R}^{\odot} \right) \right] \right], \tag{65}$$

$$\mathbf{w}\mathfrak{T}_{R}^{\circ} = (1+\lambda z) x Z_{R} \left[\left[(2+\lambda y) \lambda y \left(\mathbf{w}\mathfrak{T}_{R}^{\circ} \right) \right] \right], \tag{66}$$

They can also be expressed explicitly in terms of the cycle index series $Z_{A_R}(x_1, x_2, x_3, \cdots)$ of the 1-sort species $A_R(X)$ by

$$\mathbf{w}\vec{\mathfrak{T}}_{R}^{\circ} = \frac{1}{\lambda y} Z_{A_{R}} \left[\left[\left(1 + \lambda z \right) \lambda x y \right] \right], \tag{67}$$

$$\mathbf{w}\mathfrak{T}_{R}^{\odot} = \frac{1}{(2+\lambda y)\lambda y} Z_{A_{R}} \Big[\Big[(1+\lambda z)(2+\lambda y)\lambda xy \Big] \Big].$$
(68)

where $\lambda = \Lambda(t\Sigma a)$, $\Lambda(u) = \sum_{i \in \mathbb{L}} u^i$, $Z_R = Z_R(x_1, x_2, x_3, \cdots)$. Moreover, let *R'* be the combinatorial derivative of the species R. Then, the inventory $\mathbf{w}\mathbf{t}_R^{\odot}$ has the form

$$\mathbf{w}\mathfrak{t}_{R}^{\circ} = (1+\lambda z) x Z_{R} \left[\left[(2+\lambda y) \lambda y \left(\mathbf{w}\mathfrak{T}_{R'}^{\circ} \right) \right] \right] - \frac{1}{2} (2+\lambda y) \lambda y \left(\mathbf{w}\mathfrak{T}_{R'}^{\circ} \right)^{2} + \frac{1}{2} \lambda_{2} y^{2} \left(\mathbf{w}\mathfrak{T}_{R'}^{\circ} \right)_{2}, \tag{69}$$

In the case of the corresponding sets, \mathfrak{T}_{R} , \mathfrak{T}_{R} , \mathfrak{t}_{R} , in which no loops are allowed, we have

$$\mathbf{w}\vec{\mathfrak{I}}_{R} = \mathbf{w}\vec{\mathfrak{I}}_{R}^{\bigcirc}\Big|_{z=0}, \ \mathbf{w}\mathfrak{T}_{R} = \mathbf{w}\mathfrak{T}_{R}^{\bigcirc}\Big|_{z=0}, \ \mathbf{w}\mathfrak{t}_{R} = \mathbf{w}\mathfrak{t}_{R}^{\bigcirc}\Big|_{z=0}.$$
(70)

Proof. (sketch) Apply Proposition 2.1, taking into account Lemma 3.4.

Note. When written explicitly, (65) takes the form

$$\mathbf{w}\vec{\mathfrak{T}}_{R}^{\circ} = (1+\lambda_{1}z) x Z_{R} \Big(\lambda_{1} y \Big(\mathbf{w}\vec{\mathfrak{T}}_{R}^{\circ} \Big), \lambda_{2} y^{2} \Big(\mathbf{w}\vec{\mathfrak{T}}_{R}^{\circ} \Big)_{2}, \lambda_{3} y^{3} \Big(\mathbf{w}\vec{\mathfrak{T}}_{R}^{\circ} \Big)_{3}, \cdots \Big),$$
(71)

where $\mathbf{w}\vec{\mathfrak{T}}_{R}^{\odot} = \mathbf{w}\vec{\mathfrak{T}}_{R}^{\odot}(x, y, z, t, (a)_{a \in A}), \quad (\mathbf{w}\vec{\mathfrak{T}}_{R}^{\odot})_{2} = \mathbf{w}\vec{\mathfrak{T}}_{R}^{\odot}(x^{2}, y^{2}, z^{2}, t^{2}, (a^{2})_{a \in A}), \cdots, \quad \lambda_{1} = \lambda = \Lambda(t\sum_{a \in A}a),$

 $\lambda_2 = \Lambda \left(t^2 \sum_{a \in A} a^2 \right), \cdots$. Formulas (65) and (66) give rise to iterative schemes for the computation of the inventories $\mathbf{w} \vec{\mathfrak{T}}_R^{\odot}$, $\mathbf{w} \mathfrak{T}_R^{\odot}$. See, for example, [7]-[10] for descriptions of efficient ways to do such computations, including adaptations of quadratically convergent Newtonian methods. Formula (69) reduces the computation of $\mathbf{w} \mathfrak{t}_R^{\odot}$ to that of $\mathbf{w} \mathfrak{T}_{R'}^{\odot}$. Recall that $Z_{R'}(x_1, x_2, x_3, \cdots) = \frac{\partial}{\partial x_1} Z_R(x_1, x_2, x_3, \cdots)$.

Sample of explicit examples of computations.

Example 3.5. One way free binary rooted tree sentences.

Consider the set $\tilde{\mathfrak{B}}$ of one way free binary rooted tree sentences without loops (**Figure 6** left, shows such a tree-like sentence over the 26-letter alphabet $A = \{a, b, \dots, z\}$) and the set $\tilde{\mathfrak{B}}^{\circ}$ of such sentences where loops are allowed. These tree-like sentences correspond to R-enriched rooted trees with $R = 1 + X + E_2$. Since $Z_R = 1 + x_1 + (x_1^2 + x_2)/2$, formula (65) of Proposition 3.6 immediately gives the following recursive scheme for

 $Z_R = 1 + x_1 + (x_1^2 + x_2)/2$, formula (65) of Proposition 3.6 immediately gives the following recursive scheme for the computation of the inventory $\mathbf{w}\vec{\mathfrak{B}}^\circ$

$$\mathbf{w}\vec{\mathfrak{B}}^{\circ} = (1+\lambda z) x \left(1+\lambda y \mathbf{w}\vec{\mathfrak{B}}^{\circ} + \frac{1}{2} \left(\lambda^2 y^2 \left(\mathbf{w}\vec{\mathfrak{B}}^{\circ} \right)^2 + \lambda_2 y^2 \left(\mathbf{w}\vec{\mathfrak{B}}^{\circ} \right)_2 \right) \right),$$
(72)

and, since $\mathbf{w}\vec{\mathfrak{B}} = \mathbf{w}\vec{\mathfrak{B}}^{\circ}\Big|_{z=0}$, (70) leads to

$$\mathbf{w}\vec{\mathfrak{B}} = x \left(1 + \lambda y \mathbf{w}\vec{\mathfrak{B}} + \frac{1}{2} \left(\lambda^2 y^2 \left(\mathbf{w}\vec{\mathfrak{B}} \right)^2 + \lambda_2 y^2 \left(\mathbf{w}\vec{\mathfrak{B}} \right)_2 \right) \right).$$
(73)

Of course, as many terms as we want in (72) and (73) can be computed using a computer algebra system. For a more specific application, let ℓ be the number of letters in alphabet *A* and let $\vec{f}^{\circ}(t) = \mathbf{w}\vec{\mathfrak{B}}^{\circ}(1,1,1,t,(1)_{a\in A})$ and $\vec{f}(t) = \mathbf{w}\vec{\mathfrak{B}}(1,1,1,t,(1)_{a\in A})$ as in (7). Then $\lambda = \ell t/(1-\ell t)$, $\lambda_2 = \ell t^2/(1-\ell t^2)$ and (72), (73) give, after some symbolic manipulation,

$$\vec{f}^{\circ}(t) = \frac{1 - \ell t}{1 - 3\ell t + \ell^2 t^2} \left(1 + \frac{1}{2} \left(\frac{\ell^2 t^2}{\left(1 - \ell t\right)^2} \vec{f}^{\circ}(t)^2 + \frac{\ell t^2}{1 - \ell t^2} \vec{f}^{\circ}(t^2) \right) \right)$$
(74)

$$=1+2\ell t+\frac{1}{2}\ell(11\ell+1)t^{2}+\ell^{2}(17\ell+1)t^{3}+\frac{3}{2}\ell^{2}(37\ell^{2}+2\ell+1)t^{4}$$

$$+\frac{3}{2}\ell^{3}(125\ell^{2}+7\ell+2)t^{5}+\frac{1}{8}\ell^{2}(5197\ell^{4}+310\ell^{3}+73\ell^{2}+34\ell+2)t^{6}+\cdots,$$
(75)

$$\vec{f}(t) = \frac{1-\ell t}{1-2\ell t} \left(1 + \frac{1}{2} \left(\frac{\ell^2 t^2}{\left(1-\ell t\right)^2} \vec{f}(t)^2 + \frac{\ell t^2}{1-\ell t^2} \vec{f}(t^2) \right) \right)$$
(76)

$$=1+\ell t + \frac{1}{2}\ell (5\ell+1)t^{2} + \frac{1}{2}\ell^{2} (13\ell+1)t^{3} + \frac{1}{2}\ell^{2} (35\ell^{2}+3\ell+2)t^{4} + \frac{1}{2}\ell^{3} (97\ell^{2}+9\ell+2)t^{5} + \frac{1}{8}\ell^{2} (1101\ell^{4}+110\ell^{3}+25\ell^{2}+18\ell+2)t^{6} + \cdots.$$
(77)

The coefficient of t^q in (75) (resp. (77)) is the number of one way free binary rooted tree sentences with loops (resp. without loops) on a ℓ -letter alphabet A that are made of q letters. As an illustration, for the usual 26-letter alphabet $A = \{a, b, \dots, z\}$, series (75) and (77) read as follows up to q = 10:

$$\vec{f}^{\circ}(t) = 1 + 52t + 3731t^{2} + 299468t^{3} + 25415910t^{4} + 2232608976t^{5} + 201144059077t^{6} + 18486682363084t^{7} + 1726830972452898t^{8} (78) + 163481547244045680t^{9} + 15652457718105041990t^{10} + \cdots, \vec{f}(t) = 1 + 26t + 1703t^{2} + 114582t^{3} + 8024120t^{4} + 578320704t^{5} + 42679370357t^{6} + 3210985966578t^{7} + 245445450450824t^{8} (79) + 19011557820684416t^{9} + 1489108237113437148t^{10} + \cdots.$$

Example 3.6. Ordinary tree and rooted tree sentences.

Let R = R(X) = E(X), be the species of finite sets. Then, by Definition 3.1, the species $A_E = A_E(X)$ of *E*-enriched rooted trees coincides with the species A = A(X) of ordinary (free) rooted trees and the species $a_E(X)$ of *E*-enriched trees coincides with the species a = a(X) of ordinary (free) trees. Lemma 3.3 produces the familiar combinatorial equations,

$$A = XE(A), \ A + E_2(A) = a + A^2, \tag{80}$$

the second equation being the classical dissymmetry formula of Leroux. Taking cycle index series in (80) and using R' = E' = E and $Z_E = \exp\left(\sum_{k\geq 1} x_k/k\right)$, we obtain the classical formulas

$$Z_{A} = x_{1} \exp\left(Z_{A} + \frac{1}{2}(Z_{A})_{2} + \frac{1}{3}(Z_{A})_{3} + \cdots\right)$$
(81)

$$= x_{1} + x_{1}^{2} + \frac{1}{2} \left(3x_{1}^{2} + x_{2} \right) x_{1} + \frac{1}{3} \left(8x_{1}^{3} + 3x_{1}x_{2} + x_{3} \right) x_{1} + \frac{1}{24} \left(125x_{1}^{4} + 54x_{1}^{2}x_{2} + 16x_{1}x_{3} + 15x_{2}^{2} + 6x_{4} \right) x_{1} + \cdots,$$
(82)

$$Z_{a} = Z_{A} - \frac{1}{2}Z_{A}^{2} + \frac{1}{2}(Z_{A})_{2}$$
(83)

$$=1+\frac{1}{2}\left(x_{1}^{2}+x_{2}\right)+\frac{1}{2}\left(x_{1}^{2}+x_{2}\right)x_{1}+\frac{1}{6}\left(4x_{1}^{4}+3x_{1}^{2}x_{2}+2x_{1}x_{3}+3x_{2}^{2}\right)$$
(84)

$$+\frac{1}{24}\left(25x_1^4+18x_1^2x_2+8x_1x_3+15x_2^2+6x_4\right)x_1+\cdots,$$

from which $Z_A(x_1, x_2, x_3, \cdots)$ and $Z_a(x_1, x_2, x_3, \cdots)$ can be computed to arbitrary degree¹³. Now, let

 $\vec{\mathfrak{T}}^{\circ}$ be the set of one way free rooted tree sentences with possible loops, (85)

 \mathfrak{T}° be the set of free rooted tree sentences with possible loops, (86)

 \mathfrak{t}° be the set of free tree sentences with possible loops, (87)

Then, by (67)-(69),

$$\mathbf{w}\vec{\mathfrak{T}}^{\circ} = \frac{1}{\lambda y} Z_A \Big[\Big[\big(1 + \lambda z \big) \lambda x y \Big] \Big], \tag{88}$$

$$\mathbf{w}\mathfrak{T}^{\circ} = \frac{1}{(2+\lambda y)\lambda y} Z_{A} \Big[\Big[(1+\lambda z)(2+\lambda y)\lambda xy \Big] \Big], \tag{89}$$

$$\mathbf{w}\mathfrak{t}^{\circlearrowright} = \mathbf{w}\mathfrak{T}^{\circlearrowright} - \frac{1}{2}(2 + \lambda y)\lambda y (\mathbf{w}\mathfrak{T}^{\circlearrowright})^{2} + \frac{1}{2}\lambda_{2}y^{2} (\mathbf{w}\mathfrak{T}^{\circlearrowright})_{2}.$$
(90)

For more specific applications, let ℓ be the number of letters in the alphabet and consider the specializations

¹³Explicit expressions for the *individual* coefficients of Z_A and of Z_a have also been obtained, see [9].

 $F^{\circ}(t) = w\mathfrak{T}^{\circ}(1,1,1,t,(1)_{a\in A}) \text{ and } f^{\circ}(t) = \mathbf{w}\mathfrak{t}^{\circ}(1,1,1,t,(1)_{a\in A}). \text{ Then } \lambda_{k} = \ell t^{k} / (1-\ell t^{k}), \quad k = 1,2,\cdots, \text{ and by}$ (66) and (69) with R = E, we have

$$F^{\circ}(t) = \frac{1}{1 - \ell t} \exp\left(\sum_{k \ge 1} \frac{1}{k} \frac{(2 - \ell t^{k})\ell t^{k}}{(1 - \ell t^{k})^{2}} F^{\circ}(t^{k})\right)$$
(91)

$$= 1 + 3\ell t + \ell (14\ell + 1)t^{2} + \frac{1}{3}\ell (223\ell^{2} + 15\ell + 2)t^{3} + \frac{1}{6}\ell (2539\ell^{3} + 198\ell^{2} + 50\ell + 3)t^{4} + \frac{1}{30}\ell (75873\ell^{4} + 6790\ell^{3} + 1470\ell^{2} + 95\ell + 12)t^{5} + \cdots,$$
(92)

$$+ \frac{1}{30}\ell (75873\ell^{4} + 6790\ell^{3} + 1470\ell^{2} + 95\ell + 12)t^{5} + \cdots,$$
(93)

$$f^{\circ}(t) = F^{\circ}(t) - \frac{1}{2} \frac{(2 - \ell t)\ell t}{(1 - \ell t)^{2}} F^{\circ}(t)^{2} + \frac{1}{2} \frac{\ell t^{2}}{1 - \ell t^{2}} F^{\circ}(t^{2})$$
(93)

$$= 1 + 2\ell t + \frac{1}{2}\ell (13\ell + 3)t^{2} + \frac{1}{6}\ell (79\ell^{2} + 9\ell + 2)t^{3} + \frac{1}{2}\ell (241\ell^{3} + 28\ell^{2} + 18\ell + 1)t^{4}$$
(94)

$$+\frac{1}{30}\ell(17993\ell^4+2230\ell^3+760\ell^2+65\ell+12)t^5+\cdots.$$

The coefficient of t^q in (92) (resp. (94)) is the number of free rooted tree sentences (resp. free tree sentences) with loops on a ℓ -letter alphabet A that are made of q letters. For example, in the case of a 4-letter alphabet, say $A = \{A, C, G, T\}$, series (92) and (94) read as follows up to q = 9:

$$F^{\circ}(t) = 1 + 12t + 228t^{2} + 4840t^{3} + 110578t^{4} + 2650928t^{5} + 65889652t^{6} + 1683125528t^{7} + 43926578519t^{8} + 1166284622820t^{9} + \cdots,$$

$$f^{\circ}(t) = 1 + 8t + 110t^{2} + 1736t^{3} + 31890t^{4} + 634848t^{5} + 13484444t^{6} + 300435640t^{7} + 6950286407t^{8} + 165682482716t^{9} + \cdots.$$
(95)

Consider now the 2-letter alphabet $A = \{a, b\}$, and make the substitutions x = 1, y = 1, z = 0, t = 1, then $\lambda = \frac{a+b}{1-a-b}$, no loops are allowed, and (89)-(90) take the forms

$$\mathbf{w}\mathfrak{T} = \frac{(1-a-b)^2}{(2-a-b)(a+b)} Z_A \left[\left[\frac{(2-a-b)(a+b)}{(1-a-b)^2} \right] \right] = \sum_{i+j>0} g_{i,j} a^i b^j,$$
(97)

$$\mathbf{wt} = \mathbf{w}\mathfrak{T} - \frac{(2-a-b)(a+b)}{2(1-a-b)^2} (\mathbf{w}\mathfrak{T})^2 + \frac{(a^2+b^2)}{2(1-a^2-b^2)} (\mathbf{w}\mathfrak{T})_2 = \sum_{i+j>0} h_{i,j} a^i b^j,$$
(98)

where $g_{i,j}$ (resp. $h_{i,j}$) is the number of rooted tree sentences (respected tree sentences) without loops that contain exactly *i* times the letter *a* and *j* times the letter *b*.

If we choose $\mathbb{L} = \{1,2\}$ (only words of length 1 or 2 are allowed), then $\lambda = (a+b)+(a^2+b^2)$, and (89)-(90) begin with the terms

$$\mathbf{w}\mathfrak{T} = 1 + 2a + 2b + 10a^{2} + 18ab + 10b^{2} + 46a^{3} + 128a^{2}b + 128ab^{2} + 46b^{3} + 237a^{4} + 874a^{3}b + 1291a^{2}b^{2} + 874ab^{3} + 237b^{4} + 1258a^{5} + 5870a^{4}b$$
(99)
+ 11402a^{3}b^{2} + 11402a^{2}b^{3} + 5870ab^{4} + 1258b^{5} + \cdots,

$$\mathbf{wt} = 1 + a + b + 5a^{2} + 7ab + 5b^{2} + 15a^{3} + 39a^{2}b + 39ab^{2} + 15b^{3} + 66a^{4} + 218a^{3}b + 327a^{2}b^{2} + 218ab^{3} + 66b^{4} + 276a^{5} + 1234a^{4}b + 2366a^{3}b^{2} + 2366a^{2}b^{3} + 1234ab^{4} + 276b^{5} + \cdots,$$
(100)

etc. Again, all the above series, and many variants, can be expanded to arbitrary orders.

Example 3.7. Back to linear sentences.

Since linear sentences are special kinds of tree-like sentences, it is interesting to look at the dissymmetry formula (56) in the context of path-shaped graphs. Take the 1-sort species $R = R(X) = 1 + X + E_2(X)$. Then a *R*-structure is either void, a singleton, or an unordered pair of singletons. This means that a *R*-enriched tree is a simple path (see Definition 3.1, **Figure 7** right and **Figure 3** middle). Hence, the 2-sort species P(X,Y) of all path-shaped digraphs without loops (see **Figure 9**) coincides with the 2-sort species $a_R(X,Y) = a_R^{\circ}(X,Y,0)$ of *R*-enriched trees. Moreover, since R' = 1 + X, a *R'*-enriched rooted tree is a simple path pointed at an extremity (see Definition 3.1 and **Figure 7**, left). Hence, the 2-sort species K(X,Y) of all path-shaped digraphs without loops pointed at an extremity (see **Figure 9**) coincides with the 2-sort species $A_{R'}(X,Y) = A_{R'}^{\circ}(X,Y,0)$ of *R'*-enriched rooted trees. In this setting, the dissymmetry formula (56), with Z = 0, becomes

$$X \cdot \left(1 + \Omega K + E_2(\Omega K)\right) + E_2(YK) = P + \left(Y + Y^2\right)K^2, \tag{101}$$

where $\Omega = \Omega(Y) = 2Y + Y^2$. Now, using $K = X \cdot (1 + \Omega K)$, we can solve (101) for P as follows,

$$P = K + XE_2(\Omega K) + E_2(YK) - Y(1+Y)K^2.$$
(102)

This formula coincides with formula (124) which is used in the proof of Proposition 3.1.

Example 3.8. Plane tree sentences.

A plane tree is a (unrooted) tree that is embedded in a plane. Such tree-structures have fewer automorphisms than free trees. Take any vertex p of a plane tree τ and draw a vector starting at p which is perpendicular to the plane in which τ is embedded. This gives an orientation to that plane and the vertices that are adjacent to p are cyclically turning around p according to that orientation (see Figure 7). In other words, the set of immediate neighbors of p is equipped with a (1+C)-structure, where C is the species of non-empty oriented cycles (the empty set species, 1, corresponds to the special case where the tree is reduced to one point, $\tau = \{p\}$, for which the the set of immediate neighbors of p is empty). Since p is arbitrary, this shows that the species a_{plane} of plane trees coincides with the species $a_{\text{l+}C}$ of (1+C)-enriched trees. Take R = 1+C, then R' = 1' + C' = 1/(1-X) = L is the species of *linear orders*. The species A_L of L-

Take R = 1+C, then R' = 1'+C' = 1/(1-X) = L is the species of *linear orders*. The species A_L of Lenriched rooted trees coincides with the species A_o of *linearly ordered rooted trees* (the set of immediate descendants, away from the root, of every vertex is linearly ordered). Using the classical formulas,

$$Z_{R} = Z_{1+C} = 1 + \sum_{k \ge 1} \frac{\phi(k)}{k} \log\left(\frac{1}{1-x_{k}}\right) \text{ (where } \phi(k) \text{ denotes Euler function) and } Z_{R'} = Z_{L} = \frac{1}{1-x_{1}}, \text{ in Lem-$$

ma 3.3, we have

$$Z_{A_o} = \frac{1 - \sqrt{1 - 4x_1}}{2} = \sum_{m \ge 1} \frac{(2m - 2)!}{m!(m - 1)!} x_1^m,$$
(103)

$$Z_{a_{\text{plane}}} = x_1 \left(1 - \sum_{k \ge 1} \frac{\phi(k)}{k} \log\left(\frac{1 + \sqrt{1 - 4x_k}}{2}\right) \right) + \frac{1}{2}x_1 + \frac{1}{4} \left(\sqrt{1 - 4x_1} - \sqrt{1 - 4x_2}\right).$$
(104)

Using the expansion¹⁴ $-\log\left(\frac{1}{2}\left(1+\sqrt{1-4x}\right)\right) = \sum_{i\geq 1} \frac{(2i-1)!}{i!^2} x^i$, we get, from (68) and (69), the following *explicit expressions* for the inventory $\mathbf{w}\mathfrak{T}_o^{\bigcirc}$ (resp., $\mathfrak{t}_{\text{plane}}^{\bigcirc}$) of *linearly ordered rooted tree sentences with loops*

$$\mathbf{w}\mathfrak{T}_{o}^{\circ} = \frac{1 - \sqrt{1 - 4\theta}}{2(2 + \lambda_{1}y)\lambda_{1}y} = \frac{1}{(2 + \lambda_{1}y)\lambda_{1}y} \cdot \sum_{m \ge 1} \frac{(2m - 2)!}{m!(m - 1)!} \theta^{m}, \tag{105}$$

(resp., plane tree sentences with loops)

¹⁴Which can be proved by taking the derivative of both sides.

$$wt_{\text{plane}}^{\circ} = (1 + \lambda_1 z) x \left(1 - \sum_{j \ge 1} \frac{\phi(j)}{j} \log\left(\frac{1 + \sqrt{1 - 4\theta_j}}{2}\right) \right) - \frac{1 - \sqrt{1 - 4\theta} - 2\delta}{4(2 + \lambda_1 y)\lambda_1 y} + \frac{1 - \sqrt{1 - 4\theta_2}}{4(2 + \lambda_2 y^2)},$$
(106)

$$= (1+\lambda_{1}z)x\left(1+\sum_{i,j\geq 1}\frac{(2i-1)!\phi(j)}{i!^{2}j}\theta_{j}^{i}\right) - \frac{\sum_{i\geq 2}\frac{(2i-2)!}{i!(i-1)!}\theta^{i}}{2(2+\lambda_{1}y)\lambda_{1}y} + \frac{\sum_{i\geq 1}\frac{(2i-2)!}{i!(i-1)!}\theta_{2}^{i}}{2(2+\lambda_{2}y^{2})},$$
(107)

where

$$\theta = \theta_1 = (1 + \lambda_1 z) (2 + \lambda_1 y) \lambda_1 xy, \ \theta_j = (1 + \lambda_j z^j) (2 + \lambda_j y^j) \lambda_j x^j y^j.$$
(108)

Note that since linearly ordered rooted tree structures are asymmetric structures, no λ_k except λ_1 appear in (105). As before, let ℓ be the number of letters in alphabet A and let $F^{\circ}(t) = \mathbf{w}\mathfrak{T}_o^{\circ}(1,1,1,t,(1)_{a\in A})$ and $f^{\circ}(t) = \mathbf{w}\mathfrak{T}_o^{\circ}(1,1,1,t,(1)_{a\in A})$. Then

$$F^{\circ}(t) = 1 + 3\ell t + 16\ell^2 t^2 + 105\ell^3 t^3 + 765\ell^4 t^4 + 5951\ell^5 t^5 + \cdots,$$
(109)

$$f^{\circ}(t) = 1 + 2\ell t + \frac{1}{2}\ell(13\ell+3)t^{2} + \frac{1}{3}\ell(83\ell^{2}+3\ell+4)t^{3} + \frac{1}{6}\ell(859\ell^{3}+6\ell^{2}+53\ell+6)t^{4} + \frac{1}{30}\ell(25667\ell^{4}+30\ell^{3}+205\ell^{2}+30\ell+48)t^{5} + \cdots.$$
(110)

This time, the coefficient of t^q in (109) (resp. (110)) is the number of linearly ordered rooted tree sentences (resp. plane tree sentences) with loops on a ℓ -letter alphabet A that are made of q letters.

As a final illustration, fix $m \ge 1$ and consider the inventory $\mathbf{w}\mathfrak{T}_{o,m}$ of *linearly ordered rooted tree sentences* without loops having exactly m vertices. Letting z = 0 in (105) and (108), we have

$$\mathbf{w}\mathcal{I}_{o,m} = \frac{(2m-2)!}{m!(m-1)!} (2\lambda y + \lambda^2 y^2)^{m-1} x^m.$$
(111)

Now, let ℓ be the number of letters in the alphabet and assume that the length of the word on each arrow is at most *k*. Then, making the substitutions, x = y = 1, a = 1, $\forall a \in A$, we get

$$\mathbf{w}\mathfrak{T}_{o,m}(1,1,1,t,(1)_{a\in A}) = \frac{(2m-2)!}{m!(m-1)!} (2\lambda + \lambda^2)^{m-1} = \sum_{m-1 \le q \le 2m-2} v_q t^q,$$
(112)

which is a polynomial in *t*, since $\lambda = \frac{\ell t (1 - \ell^k t^k)}{1 - \ell t}$ has *k* terms. Differentiation gives

$$\left[\frac{\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{w}\mathfrak{T}_{o,m}(1,1,1,t,(1)_{a\in A})}{\mathbf{w}\mathfrak{T}_{o,m}(1,1,1,t,(1)_{a\in A})}\right]_{t=1} = \left[\frac{2(m-1)(1+\lambda)\frac{\mathrm{d}}{\mathrm{d}t}\lambda}{(2+\lambda)\lambda}\right]_{t=1} = \frac{\sum_{q}qv_{q}}{\sum_{q}v_{q}} = e_{m,\ell,k},$$
(113)

where $e_{m,\ell,k}$ is the expected total number of letters in random *m*-vertex linearly ordered rooted tree sentence without loops on a ℓ -letter alphabet in which the word on each arrow has at most *k* letters. Further computations give

$$e_{m,\ell,k} = 2(m-1)\frac{\left(\ell^{k+1}-1\right)\left(k\ell^{k+1}-\left(k+1\right)\ell^{k}+1\right)}{\left(\ell-1\right)\left(\ell^{k}-1\right)\left(\ell^{k+1}+\ell-2\right)} \sim 2k(m-1)\left(1-\frac{1}{\ell}\right), \quad \ell \to \infty.$$
(114)

Again, all the above inventories can be manipulated in a great number of ways.

4. Concluding Remarks

It would be interesting to extend the above analysis to other classes of graphical sentences arising from other

families of 3-sort species of connected digraphs. As said before, this is generally a very difficult task. However, the analysis can be done, for example, for the class of *cyclic graphical sentences* (for which the underlying simple graphs are unoriented cycles) by making use of (3-sort) cycle index series related to subgroups of the dihedral groups. The analysis can also be done for the whole class of all graphical sentences since the cycle index series of the 3-sort species, $\mathcal{D}^{\circ} = \mathcal{D}^{\circ}(X, Y, Z)$, of all digraphs (with labelled vertices of sort *X*, arrows of sort *Y* and loops of sort *Z*) turns out to be tractable¹⁵. In fact,

$$\mathcal{D}^{\diamond}(X,Y,Z) = \mathcal{D}((1+Z)X,Y), \tag{115}$$

where $\mathcal{D} = \mathcal{D}(X, Y)$ is the 2-sort species of all digraphs with vertices of sort *X*, arrows of sort *Y* and no loop, which, in the spirit of [1] [11], can be expressed in terms of simpler species by making use of a 2-sort version of the more advanced operation, \Box , called *functorial composition of species*.

Another direction of investigation would be to replace digraphs by dimultigraphs (directed multigraphs) and study associated inventories of classes of multigraphical sentences. For example, in the case of the 3-sort (resp. 2-sort) species $\mathcal{M}^{\circ}(X,Y,Z)$ (resp. $\mathcal{M}(X,Y)$) of all dimultigraphs, equation (115) must be replaced by $\mathcal{M}^{\circ}(X,Y,Z) = \mathcal{M}(E(Z)X,Y)$ and operation \Box can be used.

5. Proofs of the Main Results

Proof of Proposition 2.1. First, consider the weighted species $\mathcal{G}^*(X, Y, Z) = \mathcal{G}(xX, yY, zZ)$ of all \mathcal{G} -structures in which each vertex is given a weight *x*, each arrow a weight *y* and each loop a weight *z*. Then

$$\mathbf{Z}_{g^{\star}}(x_{1}, x_{2}, \cdots; y_{1}, y_{2}, \cdots; z_{1}, z_{2}, \cdots) = \mathbf{Z}_{g}(xx_{1}, x^{2}x_{2}, \cdots; yy_{1}, y^{2}y_{2}, \cdots; zz_{1}, z^{2}z_{2}, \cdots).$$
(116)

Next, let $\Lambda(Y) = \sum_{k \in \mathbb{L}} Y^k$ be the species of all *k*-lists of arrows, where $k \in \mathbb{L}$. Figure 8(a) shows an unlabelled $\Lambda(Y)$ -structure with k = 5. Now, define the 2-sort species, $\widehat{\mathcal{G}}^*(X,Y)$ by substituting $\Lambda(Y)$ for Y and for Z in $\mathcal{G}^*(X,Y,Z)$:

$$\mathcal{G}^{\star}(X,Y) = \mathcal{G}^{\star}(X,\Lambda(Y),\Lambda(Y)) = \mathcal{G}(xX,y\Lambda(Y),z\Lambda(Y)).$$
(117)

Figure 8(b) shows an unlabelled $\widehat{\mathcal{G}^{\star}}(X,Y)$ -structure of weight $x^4y^6z^2$ on a set of vertices and arrows.

Taking the cycle index series of (117) we get

$$\mathbf{Z}_{\widehat{\mathcal{G}}^{\star}}\left(x_{1}, x_{2}, \cdots; y_{1}, y_{2}, \cdots\right) = \mathbf{Z}_{\mathcal{G}}\left(xx_{1}, x^{2}x_{2}, \cdots; y\Lambda\left(y_{1}\right), y^{2}\Lambda\left(y_{2}\right), \cdots; z\Lambda\left(y_{1}\right), z^{2}\Lambda\left(y_{2}\right), \cdots\right).$$
(118)

Next, assigning a weight ta, $a \in A$, to every arrow of every $\widehat{\mathcal{G}}^*$ -structure, gives the species

$$\mathcal{S}(X,Y) = \widehat{\mathcal{G}^{\star}}\left(X, \left(\sum_{a \in A} ta\right)Y\right),\tag{119}$$

whose cycle index, $\mathbf{Z}_{\mathcal{S}}(x_1, x_2, \dots; y_1, y_2, \dots)$, is obtained by the substitutions $y_k := \left(\sum_{a \in A} t^k a^k\right) y_k = t^k \left(\sum a^k\right) y_k$, $k = 1, 2, \dots$, in (118). Note that an unlabelled \mathcal{S} -structure can be canonically identified with a graphical sentence. So that (23) follows by the substitutions $x_k := 1, y_k := 1, k = 1, 2, \dots$, in $\mathbf{Z}_{\mathcal{S}}$ that unlabels the arrows and vertices in \mathcal{S} -structures.

Proof of Proposition 3.1. The inventories (30) of the two special kinds of linear graphical sentences, \mathcal{O} and \mathcal{O}° , are very easy to compute since directed paths are sequences of very simple structures with trivial automorphism groups. More precisely, let D(X,Y) be the species of dipaths without loops and $D^{\circ}(X,Y,Z)$ be the species of dipaths with possible loops. Since *XY*-structures are of the form \rightarrow and ZXY-structures are of the form \rightarrow , we have

$$D(X,Y) = \sum_{k\geq 0} (XY)^k X = \frac{X}{1 - XY},$$
(120)

¹⁵The cycle index series of the 1-sort species, Dig(X), of *ordinary* digraphs (labelled vertices only) is well known:

 $Z_{\text{Dig}} = \sum_{m} \left(\sum_{i \ i j \in \mathbb{Z}} 2^{\gcd(i,j)m_{m_{j}}} \right) \mathbf{x}^{m} / \mathbf{m} \downarrow \text{, where } \gcd(i,j) = \text{greatest common divisor of } i \text{ and } j \text{ (see [1], for example).}$

$$D^{\circ}(X,Y,Z) = \sum_{k\geq 0} (XY + ZXY)^{k} (1+Z)X = \frac{(1+Z)X}{1-(1+Z)XY},$$
(121)

and, since $\mathbf{Z}_X = x_1, \mathbf{Z}_Y = y_1, \mathbf{Z}_Z = z_1$, Proposition 2.1 immediately gives (30).

On the other hand, the inventories of the sets \mathfrak{L} and \mathfrak{L}° of linear graphical sentences are more difficult to compute since a path-shaped digraph can have a nontrivial automorphism (as we saw above). We first analyze the species P(X,Y) of all path-shaped digraphs without loops (see Figure 9 top). Introduce the auxiliary species $\Omega = \Omega(Y) = Y + Y + Y^2 = Y(2+Y)$. An ΩX -structure is of the form

$$\rightarrow \text{ or } \leftarrow \text{ or } \rightleftharpoons$$
. (122)

Let K = K(X,Y) be the species of all *P*-structures pointed at an extremity (see Figure 9 bottom).

This pointing induces a global orientation to these pointed structures (see dotted arrow) and implies that the species K is a species of sequences:

$$K = X \sum_{k \ge 0} (\Omega X)^{k} = \frac{X}{1 - \Omega X} = \frac{X}{1 - Y(2 + Y) X}.$$
(123)

As a consequence of the general dissymmetry formula (56) the species *P* can be expressed in terms of *K* and Ω as follows (see details in Example 3.7)

$$P = K + XE_{2}(\Omega K) + E_{2}(YK) - Y(1+Y)K^{2}, \qquad (124)$$

where E_2 denotes the species of 2-element sets. Formula (31) then follows from Proposition 2.1 by taking the cycle index series of (124) and using the fact that $\mathbf{Z}_{E_2(X)} = (x_1^2 + x_2)/2$. Finally, let $P^{\circ} = P^{\circ}(X, Y, Z)$ be the species of all path-shaped digraphs possible with loops, then the following combinatorial equation holds

$$P^{\bigcirc}(X,Y,Z) = P((1+Z)X,Y), \qquad (125)$$

since every P° -structure is obtained from a *P*-structure by adding a loop to each vertex (that is, X := ZX) or doing nothing to the vertex (that is, X := X). So that (32) follows by substituting (1+Z)X for X in (124). The computations are elementary but long and are left to the reader.

Proof of Lemma 3.4. Consider an A_R^{\odot} -structure τ and look at its underlying R-enriched rooted tree t (see **Figure 7** left). To reconstruct τ from *t*, one must replace the root of t by adding a possible loop (that is by replacing the root by an (1+Z)X-structure) and by replacing each edge adjacent to the root of t by an (outward)



Figure 8. (a) Unlabelled $\Lambda(Y)$ -structure, (b) Unlabelled $\widehat{\mathcal{G}}^*(X,Y)$ -structure (weight: $x^4y^6z^2$).

$$\underbrace{3\overset{9_1}{\leftarrow} \underbrace{5}\overset{7_1}{\leftarrow} \underbrace{6}\overset{5_1}{\leftarrow} \underbrace{10_1} \underbrace{4}\overset{7_1}{\leftarrow} \underbrace{9}\overset{2_1}{\leftarrow} \underbrace{4}_1 \underbrace{2}_3 \underbrace{4}\overset{8_1}{\leftarrow} \underbrace{9}\overset{2_1}{\leftarrow} \underbrace{4}_1 \underbrace{2}_3 \underbrace{4}\overset{8_1}{\leftarrow} \underbrace{9}\overset{2_1}{\leftarrow} \underbrace{6}_1 \underbrace{4}\overset{8_1}{\leftarrow} \underbrace{2}_3 \underbrace{4}\overset{8_1}{\leftarrow} \underbrace{8}\overset{8_1}{\leftarrow} \underbrace{9}\overset{2_1}{\leftarrow} \underbrace{8}\overset{8_1}{\leftarrow} \underbrace{8}\overset{8_1}{\leftarrow} \underbrace{9}\overset{8_1}{\leftarrow} \underbrace{8}\overset{8_1}{\leftarrow} \underbrace{8}\overset{8}{\leftarrow} \underbrace{8}\overset{8$$



arrow (that is, by replacing each such edge by an *Y*-structure). This establishes (54a). The proof of the combinatorial Equation (54b) is similar, where, this time, each edge adjacent to the root of t is replaced by an outward arrow, an inward arrow or a double arrow (that is, by replacing each such edge by a $(\Omega = 2Y + Y^2)$ -structure). To obtain the explicit formula (55a), multiply first both sides of (54a) by *Y*. This gives

$$Y\vec{\mathcal{A}}_{R}^{\circlearrowright} = (1+Z)XYR(Y\vec{\mathcal{A}}_{R}^{\circlearrowright}), \qquad (126)$$

But, by (52), the species $A_R((1+Z)XY)$ also satisfies (126). Hence, by the unicity of solution in the implicit species Theorem of Joyal [2], we must have $Y\vec{A}_R^{\odot} = A_R((1+Z)XY)$, and (55a) follows by factoring out Y. A similar argumentation can be used to prove (55b) from (54b).

The dissymmetry formula (56) is much more difficult to establish since more automorphisms are involved in enriched trees. To prove this combinatorial equality, we express in two ways the auxiliary species

 $\mathcal{F}_R = \mathcal{F}_R(X, Y, Z)$ of \mathfrak{a}_R^{\odot} -structures which are pointed either at a single vertex or at two adjacent vertices:

$$\mathcal{F}_{R} = \left(\mathfrak{a}_{R}^{\circlearrowright}\right)^{\bullet} + \left(\mathfrak{a}_{R}^{\circlearrowright}\right)^{\operatorname{adj}\{\bullet\bullet\}}.$$
(127)

• The first expression for \mathcal{F}_R reads as follows

$$\mathcal{F}_{R} = (1+Z) XR \left(\Omega \mathcal{A}_{R'}^{\circ} \right) + \left(E_{2} \left(Y \mathcal{A}_{R'}^{\circ} \right) + Y \left(\mathcal{A}_{R'}^{\circ} \right)^{2} \right).$$
(128)

To prove it, consider a \mathcal{F}_{R} -structure ϕ and look at its underlying pointed or bipointed *R*-enriched tree, *f*. We have two cases to consider:

1) If ϕ is a $(\mathfrak{a}_R^\circ)^{\bullet}$ -structure, then by **Figure 10** we see that ϕ is canonically equivalent to a $(1+Z)XR(\Omega A_{R'}^\circ)$ -structure since to recover ϕ from *f*, the vertices of *f* must be replaced by (1+Z)X-structures and the edge adjacent to the pointed vertex (and subsequently, all other edges) must be replaced by an Ω -structure.

2) If ϕ is a $(\mathfrak{a}_R^{\circ})^{\operatorname{adj}(\bullet)}$ -structure, then **Figure 11** shows that ϕ is canonically equivalent to a $(E_2(Y\mathcal{A}_R^{\circ})+Y(\mathcal{A}_R^{\circ})^2)$ -structure since to recover ϕ from *f*, the edge of *f* between the two adjacent pointed vertices must be replaced either by a double arrow (*i.e.*, ϕ is equivalent to a $E_2(Y\mathcal{A}_{R'}^{\circ})$ -structure) or by a single arrow (*i.e.*, ϕ is equivalent to a $Y(\mathcal{A}_{R'}^{\circ})^2$ -structure). This establishes (128).

• The second expression for \mathcal{F}_{R} reads as follows

$$\mathcal{F}_{R} = \mathfrak{a}_{R}^{\circlearrowright} + \left(2Y + Y^{2}\right) \cdot \left(\mathcal{A}_{R'}^{\circlearrowright}\right)^{2}.$$
(129)

To prove it, we first split the species \mathcal{F}_{R} into two subspecies according to whether the pointing(s) coincides exactly with the center or not:

$$\mathcal{F}_{R} = \mathcal{F}_{R}^{\text{pointing}(s)=\text{center}} + \mathcal{F}_{R}^{\text{pointing}(s)\neq\text{center}}.$$
(130)

Since the center of a tree is a canonical object, pointing a tree exactly at its center is naturally equivalent to doing nothing to the tree and we have

$$\mathcal{F}_{R} = \mathfrak{a}_{R}^{\bigcirc} + \mathcal{F}_{R}^{\text{pointing}(s)\neq\text{center}}$$
(131)

$$=\mathfrak{a}_{R}^{\circlearrowright} + \left(\left(\mathfrak{a}_{R}^{\circlearrowright}\right)^{\bullet\neq\text{center}} + \left(\mathfrak{a}_{R}^{\circlearrowright}\right)^{\mathrm{adj}\{\bullet\bullet\}\neq\text{center}}\right).$$
(132)

Now, consider a \mathcal{F}_{R} -structure ϕ that is *not* pointed at its center. We have two cases to consider:

1) If ϕ is an $(\mathfrak{a}_R^{\circ})^{\bullet\neq\text{center}}$ -structure, then **Figure 12** shows that ϕ is canonically equivalent to a $(2Y + Y^2) \cdot (\mathcal{A}_{R'}^{\circ})^2$ -structure ϕ_1 . Indeed, the pointing induces an orientation on the edge from the pointed vertex of f in the direction of the center (see dotted arrow) giving rise to an ordered pair of rooted trees. Moreover, to



recover ϕ from *f*, that edge must be replaced by an arrow going in same direction, or in the opposite direction of the dotted arrow, or by a double arrow. That is, the edge must be replaced by a Ω -structure. Note that ϕ_1 is not an arbitrary $(2Y + Y^2) \cdot (\mathcal{A}_{R'}^{\odot})^2$ -structure, since the global center is on the side *pointed* by the dotted arrow. 2) If ϕ is an $(\mathfrak{a}_R^{\odot})^{\operatorname{adj}(\bullet) \neq \operatorname{center}}$ -structure, then **Figure 13** shows that ϕ is canonically equivalent to a $(2Y + Y^2) \cdot (\mathcal{A}_{R'}^{\odot})^2$ -structure $\phi_2 \neq \phi_1$. Indeed, the bi-pointing induces an orientation on the edge between the pointed vertices of f in the direction *opposite* to the center (see dotted arrow) giving rise to an ordered pair of rooted trees. To recover ϕ from *f*, that edge must be replaced, as above, by an Ω -structure. Note that ϕ_2 is not



an arbitrary $(2Y + Y^2) \cdot (\mathcal{A}_{R'}^{\odot})^2$ -structure, since the global center is now on the side of the *source* of the dotted arrow.

This establishes (129) since any $(2Y + Y^2) \cdot (\mathcal{A}_{\mathcal{R}}^{\odot})^2$ -structure is either of the form ϕ_1 or of the form ϕ_2 . The general 3-sort dissymmetry formula (56) follows by cancelleing the common term $Y \cdot (\mathcal{A}_{\mathcal{R}}^{\odot})^2$ in the right-hand-sides of the expressions (128) and (129) for $\mathcal{F}_{\mathcal{R}}$.

References

- Bergeron, F., Labelle, G. and Leroux, P. (1998) Combinatorial Species and Tree-Like Structures (Encyclopedia of Mathematics and Its Applications). Cambridge University Press, Cambridge.
- Labelle, G. (1992) Counting Asymmetric Enriched Trees. Journal of Symbolic Computation, 14, 211-242. http://dx.doi.org/10.1016/0747-7171(92)90037-5
- [3] Pólya, G. and Read, R.C. (1987) Combinatorial Enumeration of Groups, Graphs and Chemical Compounds. Springer-Verlag, Berlin, Heidelberg, and New-York.
- [4] Décoste, H., Labelle, G. and Leroux, P. (1982) Une approche combinatoire pour l'itération de Newton-Raphson. Advances in Applied Mathematics, 3, 407-416. <u>http://dx.doi.org/10.1016/S0196-8858(82)80013-4</u>
- [5] Joyal. A. (1981) Une théorie combinatoire des séries formelles. Advances in Mathematics, 42, 1-82. http://dx.doi.org/10.1016/0001-8708(81)90052-9
- [6] Labbé, J.-P. and Labelle, G. (2013) Counting Types of Runs in Classes of Arborescent Words. Open Journal of Discrete Mathematics, 3, 7-15. <u>http://dx.doi.org/10.4236/ojdm.2013.31002</u>
- [7] Leroux, P. and Miloudi, B. (1992) Généralisations de la formule d'Otter. Annales des Sciences Mathématiques du Québec, 16, 53-80.
- [8] Labelle, G. (1986) Some New Computational Methods in the Theory of Species. In: Labelle, G. and Leroux, P., Eds., *Combinatoire Enumérative, Lecture Notes in Mathematics* 1234, Springer Berlin Heidelberg, Berlin, 192-209. <u>http://dx.doi.org/10.1007/bfb0072517</u>
- [9] Labelle, G. (1981) Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange. Advances in Mathematics, 42, 217-247. <u>http://dx.doi.org/10.1016/0001-8708(81)90041-4</u>
- [10] Pivoteau, C., Salvy, B. and Soria, M. (2012) Algorithms for Combinatorial Structures: Well-Founded Systems and Newton Iterations. *Journal of Combinatorial Theory, Series A*, **119**, 1711-1773. http://dx.doi.org/10.1016/j.jcta.2012.05.007
- [11] Décoste, H., Labelle, G. and Leroux, P. (1992) The Functorial Composition of Species, a Forgotten Operation. Discrete Mathematics, 99, 31-48. <u>http://dx.doi.org/10.1016/0012-365X(92)90363-K</u>