Fundamenta Informaticae XX (2005) 1–23
IOS Press

Shuffle of words and araucaria trees

René Schott
LORIA and IECN, Université Henri Poincaré, 54506 Vandoeuvre-lès-Nancy, France,
schott@loria.fr

Jean-Claude Spehner [C]
Laboratoire MIA, Equipe MAGE, FST, Université de Haute Alsace, 68093, Mulhouse, France,
JC.Spehner@uha.fr

Abstract. The shuffle of k words u_1,..., u_k is the set of words obtained by interleaving the letters of these words such that the order of appearance of all letters of each word is respected. The study of the shuffle product of words leads to the construction of an automaton whose structure is deeply connected to a family of trees which we call araucarias. We prove many structural properties of this family of trees and give some combinatorial results. We introduce a family of remarkable symmetrical polynomials which play a crucial role in the computation of the size of the araucarias. We prove that the minimal partial automaton which recognizes the shuffle of a finite number of special words contains an araucaria for each integer $k > 0$. [1]

Keywords: Automaton, shuffle of words, remarkable polynomials, trees.

1. Introduction

If u and v are words of the free monoid A^*, the language whose words are of the form $u_1v_1u_2v_2 \dots u_mv_m$ where $u_1u_2 \dots u_m$ is a factorisation of u, $v_1v_2 \dots v_m$ a factorisation of v and the factors u_1 and v_m are possibly empty, is called the shuffle of the words u and v and is denoted $u \sqcup\!\sqcup v$. More generally, if I and J are two languages of A^*, the union of the sets $u \sqcup\!\sqcup v$ for $u \in I$ and $v \in J$ is called the shuffle of the languages I and J and is denoted $I \sqcup\!\sqcup J$. Hence the shuffle of k words $u_1, \dots, u_k$ can be defined by associativity.

[C] Corresponding author
[1] A preliminary version of this paper has been presented at MCU-2004 (see [16]).

Some theoretical results on shuffle products exist (see [8, 9, 12]) but many problems remain open. The shuffle product has been proposed as model of parallel processing [7, 13]. Few algorithmic results exist: Spehner [17] designed an algorithm which determines the shuffle of two words u and v without repetition in time $O((|u|+|v|^2)\times m)$ where m is the size of the shuffle $u \sqcup\!\sqcup v$. More recently, Allauzen [1] used a suffix tree for computing the shuffle product of k words $u_1,\ldots,u_k$ in time $O\big(\binom{|u_1|+\ldots+|u_k|}{|u_1|,\ldots,|u_k|}\big)$. Efficient parallel algorithms which test if a word can be written as shuffle of two words are given in [10, 11]. Berstel and Boasson [2] solve the shuffle factorization problem: given a finite language L, do there exist words $u_1,\ldots,u_k$ such that $L = u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$? But to the best of our knowledge, no result on the minimal automaton of the shuffle product of words has been published. The aim of this paper is to fill partially this gap.
In Section 2 we give a definition of a family of trees which we call araucarias by using properties of their maximal paths. Then we prove the existence and the unicity, up to isomorphism, of an araucaria defined by its arity k which is a positive integer and by an ordered sequence of k positive integers $(p_1,\ldots,p_k)$ called its type. We prove that the number of successors of the root of an araucaria is an exponential function of its arity. Finally we give a construction based on the properties of the terminal sections of the maximal paths. Then we introduce the following family of symmetric polynomials

$$\Upsilon_k(X_1,\ldots,X_k) = \sum_{m=0}^{m=k} m! \times \Psi_m(X_1,\ldots,X_k)$$

where $\Psi_m(X_1,\ldots,X_k)$ is the elementary symmetric polynomial of degree m on variables $X_1,\ldots,X_k$ and we prove that the size of an araucaria of type $(p_1,\ldots,p_k)$ is equal to $\Upsilon_k(p_1,\ldots,p_k)$.
Section 3 is devoted to the study of the minimal automaton of the shuffle product of words. We prove that, for every given type $(p_1,\ldots,p_k)$, there exists a minimal automaton of the shuffle of k words which contains an araucaria of this type.
In a short section 4, we compare shuffle products of words with partial commutations. If Θ is a set of pairs of letters of an alphabet A, the replacement of a factor ab in a word by ba, where $(a,b)\in\Theta$ or $(b,a)\in\Theta$ and $a\neq b$ is called a partial commutation relative to Θ. Partial commutations have been intensively investigated over the two last decades in connection with parallel processing [3, 4, 6, 8, 18].

2. Araucarias

2.1. Basic definitions and properties

Below we give a direct definition which is independent of the minimal automaton which recognizes the shuffle of words.

Definition 2.1. (i) Every pair $G=(V,E)$ where $E\subset V\times V$ is called a directed graph.
Every $v\in V$ is called a vertex of G and every $(s,t)\in E$ is called an (oriented) edge of G. For every edge (s,t) of E, t is called a successor of s and s is called a predecessor of t.
The graph $G^{op}=(V,E^{op})$ where $(t,s)\in E^{op}$ if and only if $(s,t)\in E$ is called the opposite of the graph $G=(V,E)$.
(ii) Every sequence $\sigma=(s_0,\ldots,s_f)$ of vertices of V such that $(s_0,s_1),\ldots,(s_{f-1},s_f)$ are edges of E is called a path of G from s_0 to s_f.

The integer f is called the length of σ and is denoted by $|\sigma|$.
If $\sigma = (s_0, \ldots, s_f)$ and $\tau = (s_f, \ldots, s_g)$ are paths of G where the last vertex of σ is the first vertex of τ, the path $\lambda = (s_0, \ldots, s_{f-1}, s_f, s_{f+1}, \ldots, s_g)$ is called the product of σ and τ and we write $\lambda = \sigma.\tau$.
(iii) A directed graph $G = (V, E)$ is called a directed tree if there exists a vertex $r \in V$ without predecessor and such that, for every vertex $s \in V \setminus \{r\}$, there exits a unique path from r to s.
The vertex r is called the root of the directed tree G.
The length of the path from the root r to s is called the heigth of s.
Every vertex s of G without successor is called a leaf of G.
Each path issued from the root r and whose last vertex is a leaf of G is said to be maximal.

Definition 2.2. Let $\sigma = (s_0, \ldots, s_f)$ be a path of a directed tree A.
(i) For every integers $i_1, \ldots, i_{h-1}$ such that $0 < i_1 < \ldots < i_{h-1} < f$, the subpaths $\sigma_1 = (s_0, \ldots, s_{i_1})$, $\sigma_2 = (s_{i_1}, \ldots, s_{i_2})$, ..., $\sigma_h = (s_{i_{h-1}}, \ldots, s_f)$ are called sections of σ and the product $\sigma = \sigma_1 ... \sigma_h$ is called a factorisation of σ into sections.
(ii) Let $\{p_1, \ldots, p_k\}$ be a set of positive integers and $\sigma_1 ... \sigma_h$ a factorisation of a maximal path σ into sections.
Every one-to-one mapping η from the set $\{1, \ldots, h\}$ into $\{1, \ldots, k\}$ such that $\forall i \in \{1, \ldots, h-1\}$, $|\sigma_i| \leq p_{\eta(i)}$ and $|\sigma_h| = p_{\eta(h)}$, is called an attribution function for $\sigma_1 ... \sigma_h$.
For each $i \in \{1, \ldots, h\}$, the section σ_i is said to be attributable to the integer $p_{\eta(i)}$.
If $|\sigma_i| = p_{\eta(i)}$, then σ_i is said to be maximal (relatively to η).
(iii) If η is a one-to-one mapping from the set $\{1, \ldots, h\}$ into $\{1, \ldots, k\}$, a sequence $\pi = (l_1, \ldots, l_h)$ of positive integers such that $\forall i \in \{1, \ldots, h-1\}, 1 \leq l_i \leq p_{\eta(i)}$ and $l_h = p_{\eta(h)}$ is said to be linked to η.
If π is linked to η, then the pair (η, π) is called an attribute of $(p_1, \ldots, p_k)$.
(iv) If (η, π) is an attribute of $(p_1, \ldots, p_k)$, a maximal path σ of A of length $|\sigma| = l_1 + \ldots + l_h$ is said to be associated to (η, π) and the factorisation $\sigma_1 ... \sigma_h$ of σ such that $|\sigma_1| = l_1, \ldots, |\sigma_h| = l_h$ is called the canonical decomposition of σ defined by (η, π).

Remark 2.1. If $\sigma_1 ... \sigma_h$ is the canonical decomposition of a maximal path σ defined by an attribute (η, π) of $(p_1, \ldots, p_k)$, then η is an attribution function for $\sigma_1 ... \sigma_h$.

Definition 2.3. (i) Let s be a vertex of a directed tree A. Let $\lambda(s)$ be the maximum length of the paths of A whose first vertex is s. If (s, t) is an edge of A such that $\lambda(s) > \lambda(t) + 1$ then s is called a breaking vertex for the edge (s, t).
If $\sigma = (s_0, \ldots, s_f)$ is a path of A, then every nonterminal vertex s_j of σ which is a breaking vertex of the edge (s_j, s_{j+1}) is called a breaking vertex of σ.
(ii) Let $\sigma_1 ... \sigma_h$ be a canonical decomposition of a maximal path $\sigma = (s_0, \ldots, s_f)$ and $\sigma_{j_1}, \ldots, \sigma_{j_{t-1}}$ the sections of $\{\sigma_1, \ldots, \sigma_h\}$ whose first vertex is a breaking vertex of σ with $0 < j_1 < \ldots < j_{t-1} < f$.
The paths $\tau_1 = \sigma_1 ... \sigma_{j_1 - 1}$, $\tau_2 = \sigma_{j_1} ... \sigma_{j_2 - 1}$, ..., $\tau_t = \sigma_{j_{t-1}} ... \sigma_h$ are called the truncations associated to $\sigma_1 ... \sigma_h$ and the product $\tau_1 ... \tau_t$ is called the decomposition of σ into truncations associated to $\sigma_1 ... \sigma_h$.
(iii) Let (η, π) and (η', π') be two attributes of $\{p_1, \ldots, p_k\}$ such that there exists a path σ associated simultaneously to (η, π) and to (η', π') and let $\sigma_1 ... \sigma_h$, $\sigma'_1 ... \sigma'_{h'}$ be the canonical decompositions respectively defined by (η, π) and (η', π') and $\tau_1 ... \tau_t$ and $\tau'_1 ... \tau'_{t'}$ the decompositions into truncations respectively associated to $\sigma_1 ... \sigma_h$ and $\sigma'_1 ... \sigma'_{h'}$.
If $t' = t$ and $\tau_i = \tau'_i$ for every $i \in \{1, \ldots, t\}$, then the canonical decompositions $\sigma_1 ... \sigma_h$ and $\sigma'_1 ... \sigma'_{h'}$ are said to be equivalent.

Remark 2.2. The definition of the notion of pseudo-permutation given in [16] (Definition 2) is incomplete and is replaced by the more general notion of equivalence of Definition 2.3.

Definition 2.4. Let k be a positive integer, $(p_1, \ldots, p_k)$ a sequence of k positive integers and A a directed tree.
(i) If there exists a mapping $\zeta : (\eta, \pi) \to \sigma$ from the set $Att(p_1, \ldots, p_k)$ of attributes of $(p_1, \ldots, p_k)$ onto the set $MP(A)$ of maximal paths of A such that for every $(\eta, \pi) \in Att(p_1, \ldots, p_k)$, the maximal path $\zeta(\eta, \pi)$ is associated to (η, π) then A is called complete for the canonical decomposition.
(ii) If, moreover, the mapping ζ is such that $\zeta(\eta, \pi) = \zeta(\eta', \pi')$ if and only if the canonical decompositions defined by (η, π) and (η', π') are equivalent, then A is called araucaria of type $(p_1, \ldots, p_k)$ and arity k. [2]
Any araucaria of arity 1 is called elementary.

Example 2.1. The directed tree A given in Figure 1.c is an araucaria of type $(3, 2)$ since the maximal path $\sigma = (r, s_1, s_2, s_3, s_4, s_5)$ admits two equivalent canonical decompositions $\sigma_1.\sigma_2$ and $\sigma'_1.\sigma'_2$ where $\sigma_1 = (r, s_1, s_2, s_3)$ and $\sigma'_2 = (s_2, s_3, s_4, s_5)$ [resp. $\sigma_2 = (s_3, s_4, s_5)$ and $\sigma'_1 = (r, s_1, s_2)$] are attributable to 3 [resp. 2] and every other maximal path admits only one canonical decomposition. For example the maximal path $\lambda = (r, s_1, t_1, t_2, t_3)$ admits only the canonical decomposition $\sigma''_1.\sigma''_2$ with $\sigma''_1 = (r, s_1)$ attributable to 2 and $\sigma''_2 = (s_1, t_1, t_2, t_3)$ attributable to 3.
Moreover σ is a truncation and $\sigma''_1.\sigma''_2$ is the decomposition of λ into truncations.

Theorem 2.1. (i) For every positive integer k and every sequence of k positive integers $(p_1, \ldots, p_k)$, there exists a unique araucaria $A(p_1, \ldots, p_k)$ of type $(p_1, \ldots, p_k)$ up to an isomorphism.
(ii) The araucaria A of type $(p_1, \ldots, p_k)$ and arity k is such that:
- A admits a unique path τ of length $p = p_1 + \ldots + p_k$;
- for each part I of $\{1, \ldots, k\}$ whose cardinality $|I|$ verifies $0 \leq |I| < k-1$, for each $i \in \{1, \ldots, k\} \setminus I$ and each h such that $\sum_{j \in I} p_j \leq h < \sum_{j \in I} p_j + p_i$, A admits a subtree $A_{I \cup \{i\}, h}$ whose root is the vertex s_h of τ of height h and which is an araucaria of type $(p_{i_1}, \ldots, p_{i_m})$ where $i_1 < \ldots < i_m$ and $\{i_1, \ldots, i_m\} = \{1, \ldots, k\} \setminus (I \cup \{i\})$;
- for each part I of $\{1, \ldots, k\}$ such that $0 \leq |I| < k-2$, for all $i, j \in \{1, \ldots, k\} \setminus I$ $(i < j)$ and for each integer h such that $\sum_{j \in I} p_j \leq h < \sum_{j \in I} p_j + min(p_i, p_j)$, the common subtree of $A_{I \cup i, h}$ and $A_{I \cup j, h}$ is a subaraucaria of type $(p_{j_1}, \ldots, p_{j_{m-1}})$ whose root is s_h and where $j_1 < \ldots < j_{m-1}$ and $\{j_1, \ldots, j_{m-1}\} = \{1, \ldots, k\} \setminus (I \cup \{i, j\})$.

Proof:

If $k = 1$, for every positive integer p_1, the directed tree reduced to a path σ of length p_1 is an araucaria of arity 1 and type p_1 since σ contains a unique section which is attributable to p_1.

Let us assume the existence of an araucaria of any arity $l \in \{1, \ldots, k-1\}$ and any type $(p_1, \ldots, p_l)$, unique up to an isomorphism and having the properties stated in Theorem 2.1.
Let $(p_1, \ldots, p_k)$ be a sequence of k positive integers.
Let B be the directed tree such that:
- B admits a unique path τ of length $p = p_1 + \ldots + p_k$ (called the trunk);

[2] Araucarias are trees growing in South-America. They are also called monkey puzzle trees. Perhaps the terminology explains the complexity of their definition!

- for each part I of $\{1,\ldots,k\}$ such that $0 \leq |I| < k-1$, for each $i \in \{1,\ldots,k\} \setminus I$ and each integer h such that $\sum_{j\in I} p_j \leq h < \sum_{j\in I} p_j + p_i$, B admits a subtree $A_{I\cup\{i\},h}$ whose root is the vertex s_h of τ of height h and which is an araucaria of type $(p_{i_1},\ldots,p_{i_m})$ where $\{i_1,\ldots,i_m\} = \{1,\ldots,k\} \setminus (I\cup\{i\})$ and $i_1 < \ldots < i_m$;
- for each such subtree, τ and $A_{I\cup\{i\},h}$ have only the vertex s_h in common;
- two such subtrees $A_{I\cup\{i\},h}$ and $A_{J\cup\{j\},h'}$ are disjoint if $h \neq h'$ and have only the vertex s_h in common if $h = h'$.

Let A be the subtree of B obtained by removing for each part I of $\{1,\ldots,k\}$ such that $0 \leq |I| < k-2$, for all $i,j \in \{1,\ldots,k\} \setminus I$ such that $i < j$ and h such that $\sum_{j\in I} p_j \leq h < \sum_{j\in I} p_j + min(p_i,p_j)$, the subtree $T_{(i,j)}$ of $A_{I\cup j,h}$ which is an araucaria of type $(p_{j_1},\ldots,p_{j_{m-1}})$ whose root is s_h and where $\{j_1,\ldots,j_{m-1}\} = \{1,\ldots,k\}\setminus(I\cup\{i,j\})$ and $j_1 < \ldots < j_{m-1}$. Since, by induction hypothesis, $A_{I\cup i,h}$ admits a subtree $T'_{(i,j)}$ isomorphic to $T_{(i,j)}$, $A'_{I\cup j,h} = (A_{I\cup j,h} \setminus T_{(i,j)}) \cup T'_{(i,j)}$ is a subaraucaria of A isomorphic to $A_{I\cup j,h}$ and this subaraucaria is unique.
Hence A has the properties stated in part (ii) of Theorem 2.1.

(i) We prove first that A is complete for the canonical decomposition.
Let $\sigma = (s_0,s_1,\ldots,s_f)$ be a maximal path of A.
If the trunk τ of A does not contain the edge (s_0,s_1), there exists $i \in \{1,\ldots,k\}$ such that σ is a path of the subaraucaria $A_{\{i\},0}$ of arity $k-1$ and the property follows by induction since every attribute of $(p_1,\ldots,p_{i-1},p_{i+1},\ldots,p_k)$ is also an attribute for $(p_1,\ldots,p_k)$.
If $\sigma = \tau$, let η be the identity mapping from $\{1,\ldots,k\}$ onto itself and $\pi = (p_1,\ldots,p_k)$. Then (η,π) is an attribute of $(p_1,\ldots,p_k)$ and τ is associated to (η,π).
If τ contains (s_0,s_1) but $\sigma \neq \tau$, by construction of A, there exist a part I of $\{1,\ldots,k\}$ and h such that $0 \leq |I| < k-1$, $i \in \{1,\ldots,k\}\setminus I$, $\sum_{j\in I} p_j \leq h < \sum_{j\in I} p_j + p_i$ and $\sigma' = (s_h,\ldots,s_f)$ is a path of $A_{I\cup\{i\},h}$. Since, by induction, $A_{I\cup\{i\},h}$ is an araucaria of type $(p_{i_1},\ldots,p_{i_m})$ where $i_1 < \ldots < i_m$ and $\{i_1,\ldots,i_m\} = \{1,\ldots,k\}\setminus(I\cup\{i\})$, there exist an attribute (η',π') of $(p_{i_1},\ldots,p_{i_m})$ where $n \in \{1,\ldots,m\}$, η' is a one-to-one mapping from $\{1,\ldots,n\}$ into $\{1,\ldots,m\}$ and $\pi' = (l_1,\ldots,l_n)$ verifies, $\forall j \in \{1,\ldots,n\}$, $0 < l_j \leq p_{i_{\eta'(j)}}$. Let $J = I$ if $h = \sum_{j\in I} p_j$ and $J = I\cup\{i\}$ in the opposite case. Then $J \neq \emptyset$ and, if $J = \{j_1,\ldots,j_{m'}\}$ where $j_1 < \ldots < j_{m'}$, let η be the one-to-one mapping from $\{1,\ldots,n+m'\}$ into $\{1,\ldots,k\}$, $\pi = (l'_1,\ldots,l'_{n+m'})$ such that $\forall r \in \{1,\ldots,m'\}$, $\eta(r) = j_r$ and $l'_r = p_{j_r}$ if $j_r \neq i$, $l'_r = h - \sum_{j\in I} p_j$ in the opposite case and such that $\forall j \in \{1,\ldots,n\}$, $\eta(m'+j) = \eta'(j)$ and $l'_{m'+j} = l_j$. Then π is linked to η and (η,π) is an attribute of $(p_1,\ldots,p_k)$. Since $|\sigma| = h + |\sigma'|$, $|\sigma'| = \sum_{j\in\{1,\ldots,n\}} l'_{m'+j}$ and $h = \sum_{r\in\{1,\ldots,m'\}} l'_r$, σ is associated to (η,π).
Hence there exists a mapping ζ from $Att(p_1,\ldots,p_k)$ onto $MP(A)$ such that $\zeta(\eta,\pi) = \sigma$ is associated to (η,π) and A is complete for the canonical decomposition.

(ii) We prove now that in A, $\zeta(\eta,\pi) = \zeta(\eta',\pi')$ if and only if the canonical decompositions defined respectively by (η,π) and (η',π') are equivalent.
If we assume that, for any attributable section $\lambda = (s_l,\ldots,s_m)$ of a maximal path of a subaraucaria of arity $k-1$, s_j is not a breaking vertex of λ for each $j \in \{l+1,\ldots,m-1\}$, then every attributable section λ of A which is not supported by the trunk τ belongs to a subaraucaria of arity $l < k$ and has the same property by induction. The same property holds also for any attributable section λ supported by the trunk $\tau = (r_0,\ldots,r_f)$ since τ is the longest path of A and $\forall i \in \{1,\ldots,f-1\}$, $\lambda(r_i) = f - i = \lambda(r_{i+1}) + 1$.
Let $\sigma = (s_0,\ldots,s_f)$ be a maximal path of A. Suppose that there exist attributes (η,π), (η',π') in $Att(p_1,\ldots,p_k)$ such that $\sigma = \zeta(\eta,\pi) = \zeta(\eta',\pi')$. Let $\sigma_1 ... \sigma_h$, $\sigma'_1 ... \sigma'_{h'}$ be the canonical decompositions respectively defined by (η,π), (η',π') and $\tau_1 ... \tau_t$, $\tau'_1 ... \tau'_{t'}$ the decompositions into truncations associated

respectively to $\sigma_1 ... \sigma_h$ and $\sigma'_1 ... \sigma'_{h'}$. Hence the only breaking vertices of $\zeta(\eta, \pi)$ are the first vertices of $\tau_2, \ldots, \tau_t$ and, possibly, the first vertex of τ_1 and the only breaking vertices of $\zeta(\eta', \pi')$ are the first vertices of $\tau'_2, \ldots, \tau'_{t'}$ and, possibly, the first vertex of τ'_1. Since $\zeta(\eta, \pi) = \zeta(\eta', \pi')$, these breaking vertices are the same. This proves that $t = t'$ and $(\tau_1, \ldots, \tau_t) = (\tau'_1, \ldots, \tau'_t)$ and that the canonical decompositions $\sigma_1 ... \sigma_h$ and $\sigma'_1 ... \sigma'_{h'}$ are equivalent.
We have therefore proved that A is an araucaria of arity k and type $(p_1, \ldots, p_k)$.

(iii) It remains to prove the unicity of such an araucaria up to an isomorphism.
Suppose that A' is an other araucaria of arity k and type $(p_1, \ldots, p_k)$. Since A and A' admit only one path of length $p_1 + \ldots + p_k$ respectively the trunk τ and the trunk τ', we can define a bijective morphism θ_0 from τ to τ'. Since for each subset I of $\{1, \ldots, k\}$ such that $0 \leq |I| < k-1$, for each $i \in \{1, \ldots, k\} \setminus I$ and for each h such that $\sum_{j \in I} p_j \leq h < \sum_{j \in I} p_j + p_i$, the subaraucarias $A_{I \cup \{i\}, h}$ and $A'_{I \cup \{i\}, h}$ are isomorphic by induction, θ_0 can be extended to an isomorphism from A onto A'. □

Example 2.2. If $A(p)$ and $A(q)$ are two elementary araucarias of type (p) and (q) respectively, the directed tree $A(p, q)$ formed with the trunk τ of length $p+q$, p paths isomorphic to $A(q)$ issued from the p first vertices of τ and q paths isomorphic to $A(p)$ issued from the q first vertices of τ is an araucaria of type (p, q) by Theorem 2.1.

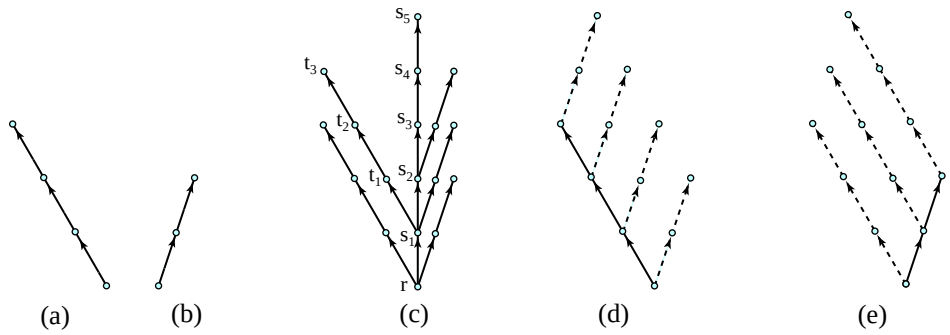

Figure 1. The araucaria of type $(3, 2)$ of Figure (c) admits a trunk of length 5, 2 subtrees isomorphic to $A(3)$ (Figure (a)) and 3 subtrees isomorphic to $A(2)$ (Figure (b)). The directed trees of Figures (d) and (e) are isomorphic to subtrees of the araucaria $A(3, 2)$.

Corollary 2.1. For each permutation φ of $\{1, \ldots, k\}$, the araucaria $A(p_{\varphi(1)}, \ldots, p_{\varphi(k)})$ is isomorphic to $A(p_1, \ldots, p_k)$.

Proof:

For each one-to-one mapping $\eta : \{1, \ldots, h\} \to \{1, \ldots, h\}$ where $h \leq k$, $\varphi \circ \eta$ is also one-to-one. Moreover the mapping which associates to the attribute (η, π) of $Att(p_1, \ldots, p_k)$ the attribute $(\varphi \circ \eta, \pi)$ of $Att(p_{\varphi(1)}, \ldots, p_{\varphi(k)})$ is a bijection. It follows that the araucarias $A(p_{\varphi(1)}, \ldots, p_{\varphi(k)})$ and $A(p_1, \ldots, p_k)$ are isomorphic. □

Hence the type $(p_1, \ldots, p_k)$ of an araucaria can be given in the canonical form such that

$$p_1 \geq p_2 \geq \ldots \geq p_k.$$

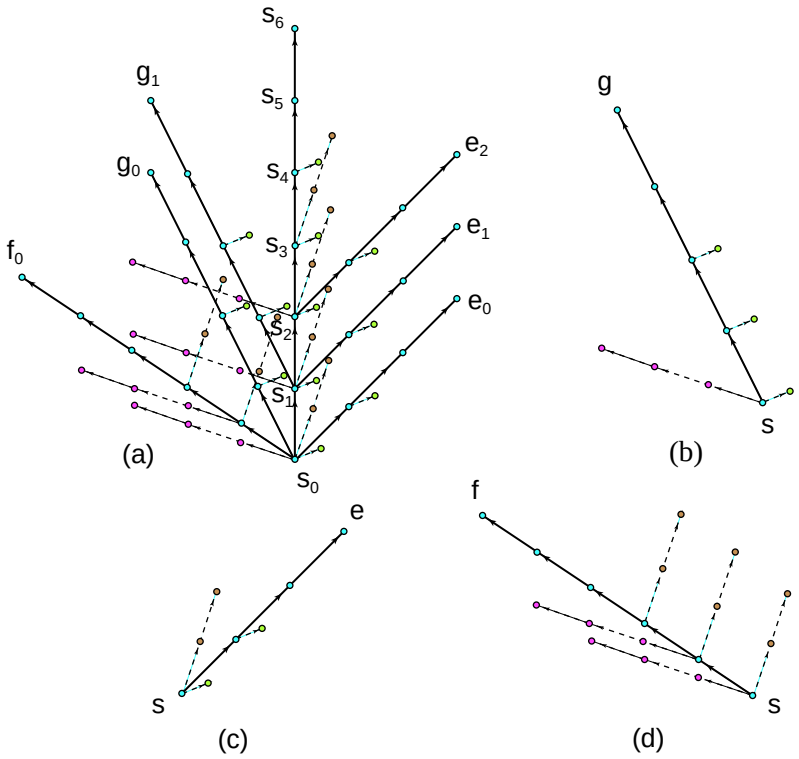

Figure 2. The araucaria $A(3,2,1)$ (Figure (a)) has 3 subaraucarias isomorphic to $A(2,1)$ (Figure (c)) with roots respectively s_0, s_1 and s_2, 2 subaraucarias isomorphic to $A(3,1)$ (Figure (b)) with roots respectively s_0 and s_1 and 1 subaraucaria isomorphic to $A(3,2)$ (Figure (d)) with root s_0. The six canonical decompositions of the trunk $\tau = (s_0, \ldots, s_6)$ of the araucaria $A(3,2,1)$ are deduced each from the other by equivalence as, for example $\sigma_1.\sigma_2.\sigma_3$ and $\sigma'_1.\sigma'_2.\sigma'_3$ with $\sigma_1 = (s_0, \ldots, s_3)$ and $\sigma'_3 = (s_3, \ldots, s_6)$ attributable to 3, $\sigma'_1 = (s_0, \ldots, s_2)$ and $\sigma_2 = (s_3, \ldots, s_5)$ attributable to 2 and $\sigma_3 = (s_5, s_6)$ and $\sigma'_2 = (s_2, s_3)$ attributable to 1. If $\tau_1 = (s_0, s_1, s_2)$ and $\tau_2 = (s_2, \ldots, e_2)$, $\sigma = \tau_1.\tau_2$ is a maximal path of $A(3,2,1)$, s_2 is a breaking vertex of σ and $\tau_1.\tau_2$ is its decomposition into truncations.

Definition 2.5. Let A be an araucaria of arity k and type $(p_1, \ldots, p_k)$, s an internal vertex of A distinct from the root r of A, ν the path from r to s in A, σ a maximal path of A which contains s and $\sigma_1 \ldots \sigma_h$ a canonical decomposition of σ.
If m is the largest integer of $\{1, \ldots, h\}$ such that σ_m contains at least an edge of ν, for every attribution function η of the canonical decomposition $\sigma_1 ... \sigma_m$, $\eta(\{1, \ldots, m\})$ is called an attribution set of ν.

Proposition 2.1. Let A be an araucaria of type $(p_1, \ldots, p_k)$, s a vertex distinct from the root r of A and which is not a leaf of A and ν the path from r to s.
For every attribution set $\{i_1, \ldots, i_m\}$ of ν there exists a subaraucaria $A(s)$ of A whose root is s, whose leaves are leaves of A and whose type $(p_{j_1}, \ldots, p_{j_m})$ is such that

$$\{j_1, \ldots, j_m\} = \{1, \ldots, k\} \setminus \{i_1, \ldots, i_m\}.$$

Proof:

Since, by Definition 2.4, A is complete for the canonical decomposition of maximal paths, for ev-

ery subset $\{h_1, \ldots, h_m\}$ of $\{1, \ldots, k\} \setminus \{i_1, \ldots, i_l\}$ and every sequence $\pi = (p'_{h_1}, \ldots, p'_{h_m})$ such that $0 < p'_{h_r} \leq p_{h_r}$ where $1 \leq r < m$ and $p'_{h_m} = p_{h_m}$, there exists a maximal path σ having a canonical decomposition $\sigma_1 \ldots \sigma_i . \sigma_{i+1} \ldots \sigma_{i+m}$ such that $\nu = \sigma_1 \ldots \sigma_i$ and, for every $r \in \{1, \ldots, m\}$, $\eta(\sigma_{i+r}) = h_r$ and $|\sigma_{i+r}| = p'_{h_r}$ and this decomposition is unique up to equivalence. It follows that $\sigma' = \sigma_{i+1} \ldots \sigma_{i+m}$ is a canonical decomposition of a maximal path whose first vertex is s. Hence, by Definition 2.4, there exists a subaraucaria $A(s)$ of A whose root is s and whose type $(p_{j_1}, \ldots, p_{j_m})$ is such that $\{j_1, \ldots, j_m\} = \{1, \ldots, k\} \setminus \{i_1, \ldots, i_l\}$. Moreover, since σ' is maximal in $A(s)$, its last vertex is a leaf of $A(s)$. □

Definition 2.6. A truncation whose last vertex is a leaf is said to be terminal.

Lemma 2.1. Every edge of an araucaria belongs to a unique terminal truncation.

Proof:

This result is trivial for $k = 1$ since every elementary araucaria is reduced to its trunk. Assume that the result holds true for every arity smaller than k with $k > 0$ and let A be an araucaria of arity k. Then the property is verified by every edge of the trunk τ of A and, by Theorem 2.1, for every edge of A which does not belong to τ, there exists a maximal arity $l \in \{1, \ldots, k-1\}$ and a unique subaraucaria B of arity l whose trunk contains (s, t). Hence, the result follows by induction. □

Corollary 2.2. Each maximal path of an araucaria admits a unique decomposition into truncations.
A section μ of a maximal path σ of an araucaria A is a truncation of σ if and only if μ is a maximal section of σ supported by a trunk of a subaraucaria of A.

Proof:

First part of this corollary is proved in part (ii) of the proof of Theorem 2.1.
If $\tau = (s_i, \ldots, s_f)$ is the trunk of a subaraucaria B of A, then, for each $j \in \{i+1, \ldots, f-1\}$, $\lambda(s_j) = f - j = \lambda(s_{j+1}) + 1$ and s_j is not a breaking vertex of τ. But if (s_j, t) is an edge with $t \neq s_{j+1}$, then $\lambda(s_j) > \lambda(t) + 1$ since, by the proof of Proposition 2.1, every path wich admits (s_j, t) as first edge is a path of B and for each maximal path $\sigma \neq \tau$ of B, $|\sigma| < |\tau|$. Hence s_j is a breaking vertex for the edge (s_j, t) and for each path which contains (s_j, t). It follows that each maximal section μ of a maximal path σ contained in a trunk of a subaraucaria of A is a truncation of σ.
Conversely, if μ is a truncation of a maximal path σ, there exists, by Lemma 2.1, a terminal truncation τ' which contains the first edge of μ and, by the preceding argument, μ is a maximal section of σ contained in τ'. □

Definition 2.7. (i) If $\tau_1 ... \tau_t$ is the decomposition of a maximal path σ into truncations, the number t of truncations is called the rank of σ.
(ii) If s is a vertex of a directed tree A, the number of successors of s is called the degree of s and is denoted $deg(s)$.
The maximum of all $deg(s)$ for $s \in A$ is called the degree of A.

Proposition 2.2. Let A be an araucaria of arity k.
(i) For each edge (s, t) of A, $deg(s) \geq deg(t)$.
(ii) The degree of A is equal to the degree $deg(r)$ of its root r and $deg(r) = 2^k - 1$.

Proof:

(i) The result is trivial for every edge whose extremity is a leaf.

Let (s,t) be an edge of A such that t is not a leaf. By Lemma 2.1, there exists a unique terminal truncation τ of A which contains (s,t). Let ν' be the unique path from the root r of A to the vertex s and $\nu = \nu'.(s,t)$. Let q be the successor of t in τ, $\Gamma(t)$ the set of successors of t distinct from q and $\Gamma(s)$ the set of successors of s distinct from t.

By Lemma 2.1, $\forall p \in \Gamma(t)$, there exits a unique terminal truncation τ_p wich contains the edge (t,p) and, since $p \neq q$, $\tau_p \neq \tau$ and (t,p) is the first edge of τ_p. Then $\sigma = \nu'.\tau_p$ is a maximal path of A and, by Definition 2.4, σ admits a canonical decomposition $\sigma_1 \dots \sigma_h$. Since (t,p) is the first edge of τ_p, there exists $i \in \{1,\dots,h-1\}$ such that $\tau_p = \sigma_{i+1} \dots \sigma_h$ and $\nu = \sigma_1 \dots \sigma_i$. Let σ'_i be the section of σ_i whose last vertex is s. Since A is complete for the canonical decomposition, there exists a maximal path σ' of A whose canonical decomposition is isomorphic to $\sigma_1 \dots \sigma_{i-1}.\sigma'_i.\sigma_{i+1} \dots \sigma_h$. Then ν' is an initial section of σ' and the successor p' of s in σ' is contained in $\Gamma(s)$ and (s,p') is the first edge of the terminal truncation τ'_p of σ' and τ'_p is isomorphic to τ_p. This proves the unicity of the vertex p' and the existence of an injective mapping from $\Gamma(t)$ into $\Gamma(s)$. Hence $deg(t) = |\Gamma(t)| + 1 \geq |\Gamma(s)| + 1 = deg(s)$.

(ii) It follows from (i) that for every path $\sigma = (c_0,\dots,c_m)$ of A, $deg(c_0) \geq \dots \geq deg(c_m)$. Hence the degree of A is equal to the degree of its root.

By Lemma 2.1, for every successor p of the root r there exits a unique terminal truncation τ_p of A which contains the edge (r,p) and (r,p) is the first edge of τ_p. By Definition 2.4 and Corollary 2.2, there exists a unique part $I_p = \{i_1,\dots,i_h\}$ of $\{1,\dots,k\}$ distinct from $\{1,\dots,k\}$ such that τ_p admits a canonical decomposition $\sigma_{i_1} \dots \sigma_{i_h}$ where $\sigma_{i_1},\dots,\sigma_{i_h}$ are maximal sections respectively attributable to $p_{i_1},\dots,p_{i_h}$. Hence the degree $deg(r)$ is equal to the number of nonempty parts of $\{1,\dots,k\}$.

Thus $deg(r) = \sum_{h=1}^{h=k} \binom{k}{h} = 2^k - 1$. □

2.2. Shuffle product of elementary araucarias

The proof of Theorem 2.1 uses the first attributable sections although the proof of Theorem 2.2 below uses the last attributable section. This new characterization of the auracarias will be used in the next section.

Definition 2.8. Let $A = (S,U)$ be a directed tree and $\sigma = (c_0,\dots,c_r)$ a path of length r.

For each vertex s of A, let $\sigma(s) = (s_0,\dots,s_r)$ be a path isomorphic to σ such that $s_0 = s$, S and $\{s_1,\dots,s_r\}$ are disjoint and such that, for each vertex t of A distinct from s, $\sigma(s) \cap \sigma(t) = \emptyset$.

The directed tree obtained by connecting A with all paths $\sigma(s)$ for $s \in S$ is called the ramified directed tree of A with respect to σ and is denoted $ramif(A,\sigma)$ (see Figures 1.*d*, 1.*e*, 3.*a*, 3.*b* and 3.*c*).

Remark 2.3. If A_k is an araucaria of type $(p_1,\dots,p_{k-1})$ and A'_k the ramified tree $ramif(A_k, A(p_k))$, Definitions 2.2 and 2.3 are applicable to A'_k. Moreover, by Definition 2.4, A'_k is isomorphic to a subtree of the araucaria $A(p_1,\dots,p_k)$.

The next lemma is based on this idea and shows the importance of the terminal truncations.

Lemma 2.2. Let $(p_1,\dots,p_k)$ be a sequence of positive integers, i and j such that $0 < i \leq k$, $0 < j \leq k$ and $i \neq j$, A_i and A_j araucarias of types respectively

$$(p_{i+1},\dots,p_k,p_1,\dots,p_{i-1}) \text{ and } (p_{j+1},\dots,p_k,p_1,\dots,p_{j-1}),$$

$A'_i = ramif(A_i, A(p_i))$ and $A'_j = ramif(A_j, A(p_j))$.
For each maximal path σ of A'_i, there exists a maximal path σ' of A'_j such that σ and σ' admit isomorphic canonical decompositions up to equivalence if and only if the terminal truncation τ of σ contains a maximal section which is attributable to p_j.

Proof:

(i) By Corollary 2.2, if there exist in A'_j a maximal path σ' which admits the same canonical decomposition as σ up to equivalence, then σ and σ' admit the same terminal truncation τ. Since the last section of the terminal truncation of σ' is attributable to p_j, τ contains indeed a maximal section attributable to p_j.

(ii) Let σ be a maximal path of A'_i and $\sigma_1 \dots \sigma_h$ its canonical decomposition with σ_h maximal and attributable to p_i. By Corollary 2.2, the terminal truncation of σ is of the form $\sigma_g ... \sigma_h$ with $g \in \{1, \dots, h-1\}$. For each section $\sigma_l \in \{\sigma_g, \dots, \sigma_{h-1}\}$, there exists $j \in \{1, \dots, k\} \setminus \{i\}$ such that σ_l is attributable to p_j. By Definition 2.4, A_j contains a maximal path $\lambda = (s_0, \dots, e)$ whose canonical decomposition is $\sigma_1 ... \sigma_{l-1}\sigma_{l+1} ... \sigma_h$. Extending λ with a path isomorphic to $A(p_j)$ and σ_l, we obtain a maximal path λ' of $A'_j = ramif(A_j, A(p_j))$ which admits the canonical decomposition $\sigma_1 ... \sigma_{l-1}\sigma_{l+1} ... \sigma_h\sigma_l$ which is equal to $\sigma_1 ... \sigma_h$ up to equivalence by Corollary 2.2. □

Definition 2.9. Let $\sigma = (s_0, \dots, s_h)$ and $\sigma' = (s'_0, \dots, s'_h)$ be two paths of equal length in a graph G. Merging σ and σ' consits in merging, for all i of $\{0, \dots, h\}$, the vertices s_i and s'_i into a unique vertex, and, for all i in $\{0, \dots, h-1\}$, in merging the edges (s_i, s_{i+1}) and (s'_i, s'_{i+1}) into a unique edge.

Definition 2.10. Let k be an integer such that $1 < k$ and $(p_1, \dots, p_k)$ a sequence of positive integers. $\forall i \in \{1, \dots, k\}$, let A_i be an araucaria of type $(p_{i+1}, \dots, p_k, p_1, \dots, p_{i-1})$ and $A'_i = ramif(A_i, A(p_i))$ and let B be the disjoint union of the directed trees $A'_1, \dots, A'_k$.
The directed graph A obtained by merging, for each couple of maximal paths (σ, σ') of B having isomorphic canonical decompositions up to equivalence, the terminal truncations of σ and σ', is called the shuffle product of the elementary araucarias $A(p_1), \dots, A(p_k)$.

Theorem 2.2. Let k be an integer such that $k > 1$ and $(p_1, \dots, p_k)$ any sequence of positive integers. The shuffle product of the elementary araucarias $A(p_1), \dots, A(p_k)$ is an araucaria of type $(p_1, \dots, p_k)$.

Proof:

(i) The directed graph A does not depend on the order in which the merging of the terminal truncations are realized. For every $i \in \{1, \dots, k\}$, A'_i admits a maximal path σ_i of length $p_1 + \dots + p_k$ and if we merge such paths σ_i and σ_j, we merge the roots of A'_i and A'_j. Hence, if we merge first all the maximal paths of B of length $p_1 + \dots + p_k$, we obtain a directed tree B'. Moreover, the merging of a maximal path σ with another maximal path σ' transforms a directed tree into another since σ and σ' have the same first vertex. The same thing happens when merging two terminal truncations issued from the same vertex. But, in each decomposition $\tau_1 \dots . \tau_t$ into truncations of a maximal path σ of rank $t > 1$, τ_{t-1} is a section of a terminal trunk τ'_{t-1} by Lemma 2.1 and $\sigma' = \tau_1 \dots . \tau_{t-2} . \tau'_{t-1}$ is then a maximal path whose rank $t-1$ is strictly less than the rank of σ. It follows that, if we realize the mergings of the terminal truncations for all couples of maximal paths (σ, σ') in increasing rank, we realize only mergings of this type. This proves that A is a directed tree.

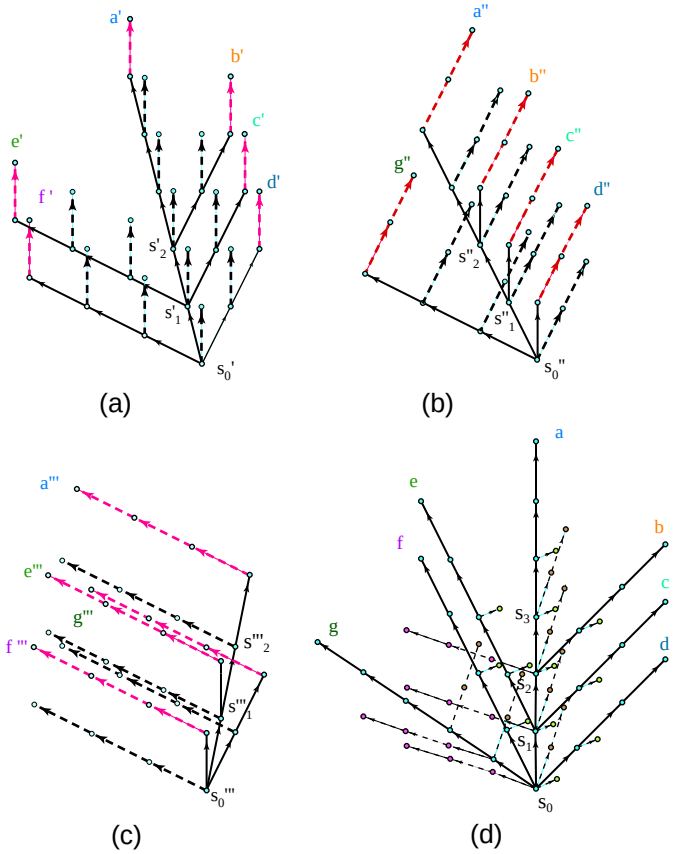

Figure 3. The trees $A'_1 = ramif(A(3,2), A(1))$, $A'_2 = ramif(A(3,1), A(2))$ and $A'_3 = ramif(A(2,1), A(3))$ are isomorphic to subtrees of $A(3,2,1)$. Moreover if, in the union B of A'_1, A'_2 and A'_3, we merge the maximal paths whose extremities are respectively a', a'', a''', b', b'', c', c'', d', d'', g', g''', e'', e''' and f'', f'''' we obtain $A(3,2,1)$.

(ii) For each $i \in \{1, \ldots, k\}$, since the araucaria A_i is of type $(p_{i+1}, \ldots, p_k, p_1, \ldots, p_{i-1})$, each maximal path σ of A_i admits a canonical decomposition into sections $\sigma_1, \ldots, \sigma_h$ attributable to two by two distinct integers of the set $\{p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_k\}$ by Definition 2.2 and Theorem 2.1 and this decomposition is unique up to equivalence. The extremity e of σ is a leaf and, if σ_{h+1} is the path issued from e isomorphic to $A(p_i)$, the path $\sigma' = \sigma.\sigma_{h+1}$ in the ramified directed tree $A'_i = ramif(A_i, A(p_i))$ is maximal and admits $\sigma_1 ... \sigma_h.\sigma_{h+1}$ as canonical decomposition and this decomposition is unique up to equivalence in A'_i. By the proof of Theorem 2.1, A'_i is isomorphic to a directed subtree of the araucaria of type $(p_1, \ldots, p_k)$.

(iii) Since, as in the construction of A, we join together all directed trees A'_1,...,A'_k and then, by Lemma 2.2, we merge all pairs of maximal paths which admit isomorphic canonical decompositions up to equivalence, each maximal path of A admits a canonical decomposition and this decomposition is unique up to equivalence.

Since, for each $i \in \{1, \ldots, k\}$ the araucaria A_i is complete for the canonical decomposition, A'_i contains all maximal paths whose terminal section is attributable to p_i. It follows that A is complete for the

canonical decomposition. By Definition 2.4, A is therefore an araucaria of type $(p_1, \ldots, p_k)$. □

2.3. The size of an araucaria

Now we introduce a family of remarkable polynomials which will be helpful for the calculation of the size of the araucarias.

Definition 2.11. Let $\{X_1, \ldots, X_k\}$ be a set of k variables.
For each $m \in \{1, \ldots, k\}$, let $\Psi_m(X_1, \ldots, X_k)$ be the elementary symmetric polynomial of degree m on variables $X_1, \ldots, X_k$ and let $\Psi_0(X_1, \ldots, X_k) = 1$.
The polynomial

$$\Upsilon_k(X_1, \ldots, X_k) = \sum_{m=0}^{m=k} m! \times \Psi_m(X_1, \ldots, X_k)$$

is called the araucaria polynomial in k variables.
The first araucaria polynomials are:

$\Upsilon_1(X_1) = X_1 + 1,$
$\Upsilon_2(X_1, X_2) = 2X_1X_2 + X_1 + X_2 + 1,$
$\Upsilon_3(X_1, X_2, X_3) = 6X_1X_2X_3 + 2(X_1X_2 + X_2X_3 + X_3X_1) + X_1 + X_2 + X_3 + 1.$

Lemma 2.3. If χ is a cyclic permutation of the set $\{1, \ldots, k\}$, for each $m \in \{1, \ldots, k-1\}$,

$$\sum_{i=1}^{i=k} \Psi_m(X_{\chi^i(1)}, \ldots, X_{\chi^i(k-1)}) \times X_{\chi^i(k)} = (m+1) \times \Psi_{m+1}(X_1, \ldots, X_k).$$

Proof:

Since χ permutes circularly the integers $1, \ldots, k$, $\sum_{i=0}^{i=k-1} \Psi_m(X_{\chi^i(1)}, \ldots, X_{\chi^i(k-1)}) \times X_{\chi^i(k)}$ is a symmetric function of $X_1, \ldots, X_k$.
The product $X_1X_2 \ldots X_mX_{m+1}$ appears in $\Psi_m(X_{\chi^i(1)}, \ldots, X_{\chi^i(k-1)}) \times X_{\chi^i(k)}$ if and only if the variable $X_{\chi^i(k)}$ belongs to $\{X_1, X_2, \ldots, X_{m+1}\}$ i.e. if and only if $\chi^i(k) \in \{1, \ldots, m+1\}$. This product appears therefore $m+1$ times in the sum. By symmetry, the same thing happens for the other products and this proves the relation. □

Theorem 2.3. For each cyclic permutation χ of the set $\{1, \ldots, k\}$,

$$\sum_{i=1}^{i=k} \Upsilon_{k-1}(X_{\chi^i(1)}, \ldots, X_{\chi^i(k-1)}) \times X_{\chi^i(k)} + 1 = \Upsilon_k(X_1, \ldots, X_k).$$

Proof:

By Lemma 2.3,

$$\sum_{i=1}^{i=k} \Upsilon_{k-1}(X_{\chi^i(1)}, \ldots, X_{\chi^i(k-1)}) \times X_{\chi^i(k)} + 1$$

$$= \sum_{i=1}^{i=k} \left(\sum_{m=0}^{m=k-1} m! \times \Psi_m(X_{\chi^i(1)}, \ldots, X_{\chi^i(k-1)}) \right) \times X_{\chi^i(k)} + 1$$

$$= \sum_{m=0}^{m=k-1} m! \times \left(\sum_{i=1}^{i=k} \Psi_m(X_{\chi^i(1)}, \ldots, X_{\chi^i(k-1)}) \times X_{\chi^i(k)} \right) + 1$$

$$= \sum_{m=0}^{m=k-1} (m+1)! \times \Psi_{m+1}(X_1, \ldots, X_k) + 1$$

$$= \Upsilon_k(X_1, \ldots, X_k).$$

□

Theorem 2.4. An araucaria of arity k and of type $(p_1, \ldots, p_k)$ has a size equal to $\Upsilon_k(p_1, \ldots, p_k)$ and the number of internal vertices is equal to $k! \times p_1 \times \ldots \times p_k$.

Proof:

(i) If $k = 1$, for each positive integer p_1, the araucaria $A(p_1)$ is reduced to a path of length p_1, its size is therefore $p_1 + 1 = \Upsilon_1(p_1)$.

Assume that the size of each araucaria of arity $k-1$ is given by the araucaria polynomial in $k-1$ variables and let A be an araucaria of arity k and of type $(p_1, \ldots, p_k)$. By Theorem 2.2, A is isomorphic to the shuffle product of the elementary araucarias $A(p_1), \ldots, A(p_k)$.

For all $i \in \{1, \ldots, k\}$, let A_i be an araucaria of type $(p_{i+1}, \ldots, p_k, p_1, \ldots, p_{i-1})$ and considere the ramified directed tree $A'_i = ramif(A_i, A(p_i))$. By Definition 2.10, if B is the disjoint union of the directed trees $A'_1, \ldots, A'_k$, then A is obtained by merging, in B, the terminal truncations of each couple of maximal paths (σ, σ') which admit isomorphic canonical decompositions up to equivalence.

For each i of $\{1, \ldots, k\}$, let V_i be the set of vertices of A'_i which do not belong to the directed subtree A_i. By Lemma 2.2 and the proof of Theorem 2.2, we can realize the merging of the terminal truncations for all couples of maximal paths (σ, σ') with respect to increasing rank.

Let σ be a maximal path of A and let τ be its terminal truncation. Since τ is a product of maximal attributable sections, there exist sections $\sigma_1, \ldots, \sigma_h$ respectively attributable to $p_{i_1}, \ldots, p_{i_h}$ where $i_1 < \ldots < i_h$. $\forall i \in \{1, \ldots, h\}$, let σ'_i be the set of vertices of σ_i distinct from the first vertex.

Since σ is a path of A'_i if and only if τ contains a section which is attributable to p_i, none of the sets $\sigma \cap V_{i_1}, \ldots, \sigma \cap V_{i_h}$ is empty. These sets are not disjoint but if we replace, for each $g \in \{2, \ldots, h\}$, the part $\sigma \cap V_{i_g}$ of V_{i_g} by the set σ'_g then they become two by two disjoint.

Let $V'_1, \ldots, V'_k$ be the residual sets obtained respectively from $V_1, \ldots, V_k$ by these substitutions for all maximal paths of A whose terminal truncations are not reduced to a unique attributable section. Then, for each maximal path of A, the sets $V'_1 \cap \sigma, \ldots, V'_k \cap \sigma$ which are not empty, are two by two disjoint and $(V'_1 \cap \sigma) \cup \ldots \cup (V'_k \cap \sigma)$ contains all vertices of σ distinct from the first vertex. It follows that the residual sets $V'_1, \ldots, V'_k$ are two by two disjoint and $V'_1 \cup \ldots \cup V'_k$ contains all vertices of A out of the root. Hence, $card(A) = 1 + \sum_{i=1}^{i=k} card(V'_i)$.

Since, for each $i \in \{1, \ldots, k\}$, each replacement of a part of V_i does not modify its cardinality, $card(V'_i) = card(V_i)$.

Moreover each vertex t of V_i is the extremity of an edge which belongs to a section which is attribuable to p_i and the set of these edges is the union of all sections attributable to p_i. It follows, by the induction

hypothesis, that $card(V_i) = p_i \times \Upsilon_{k-1}(p_{i+1}, \ldots, p_k, p_1, \ldots, p_{i-1})$.
Therefore,

$$card(A) = 1 + \sum_{i=1}^{i=k} card(V_i) = 1 + \sum_{i=1}^{i=k} p_i \times \Upsilon_{k-1}(p_{i+1}, \ldots, p_k, p_1, \ldots, p_{i-1})$$

and, by Theorem 2.3,

$$card(A) = \Upsilon_k(p_1, \ldots, p_k).$$

(ii) We prove now, thanks to a double induction on k and p_k, that an araucaria of arity k and of type $(p_1, \ldots, p_k)$ has $\sum_{m=0}^{m=k-1} m! \times \Psi_m(p_1, \ldots, p_k)$ leaves.
If $k = 1$, for each positive integer p_1, the araucaria $A(p_1)$ is reduced to a path of length p_1 and has a single leaf.
Definition 2.4 can be generalized for the value $p_k = 0$. Then an araucaria of arity k and of type $(p_1, \ldots, p_{k-1}, 0)$ can be identified with an araucaria of arity $k-1$ and of type $(p_1, \ldots, p_{k-1})$. If we assume that the property is true for each araucaria of arity $k-1$, it follows that an araucaria of arity k and of type $(p_1, \ldots, p_{k-1}, 0)$ admits $\sum_{m=0}^{m=k-2} m! \times \Psi_m(p_1, \ldots, p_{k-1})$ leaves .
Assume now that the result is true for an araucaria of type $(p_1, \ldots, p_k)$ and let B be an araucaria of type $(p_1, \ldots, p_{k-1}, p_k + 1)$.
By Theorem 2.1, we can identify A with a directed subtree of B and, furthermore:

- A admits a subaraucaria A_{k,p_k} of type $(p_1, \ldots, p_{k-1})$ whose root is the vertex s_{p_k} of height p_k on the trunk τ of A and B admits, in the place of the subaraucaria A_{k,p_k}, two subaraucarias B_{k,p_k} and B_{k,p_k+1} such that, to each subaraucaria of arity $k-2$ contained in $B_{k,p_k+1} \setminus A_{k,p_k}$ corresponds an isomorphic subaraucaria of root s_{p_k} which is common to B_{k,p_k} and to A_{k,p_k} and conversely. The number of leaves of $(B_{k,p_k} \cup B_{k,p_k+1}) \setminus A_{k,p_k}$ is therefore, by the induction hypothesis,

$$f_0 = \sum_{m=0}^{m=k-2} m! \times \Psi_m(p_1, \ldots, p_{k-1}).$$

- For each $i \in \{1, \ldots, k\}$, B admits p_i subaraucarias $B_{i,0}, \ldots, B_{i,p_i-1}$ of type

$$(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_k + 1)$$

in the place of the subaraucarias $A_{i,0}, \ldots, A_{i,p_i-1}$ of type $(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_k)$ of A.
Since

$$\Psi_m(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_k + 1) - \Psi_m(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_k)$$
$$= \Psi_{m-1}(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_{k-1})$$

the number f_i of leaves of this type is, by the induction hypothesis,

$$f_i = \sum_{m=1}^{m=k-2} m! \times p_i \times \Psi_{m-1}(p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_{k-1}).$$

- For each $i \in \{1, \ldots, k\}$, B admits the subaraucaria B_{i,p_i} in the place of A_{i,p_i} but each leaf of $B_{i,p_i} \setminus A_{i,p_i}$ belongs to a $B_{k,p_k+1} \setminus A_{k,p_k}$ and is therefore counted.

Since all these sets of leaves are two by two disjoint, it follows, by Lemma 2.3, that the number of leaves in $B \setminus A$ is:

$$\sum_{i=0}^{i=k-1} f_i = \sum_{m=0}^{m=k-2} (m+1)! \times \Psi_m(p_1, \ldots, p_{k-1}).$$

The number of leaves of B is, as expected:

$$\sum_{m=1}^{m=k-1} m! \times \big(\Psi_m(p_1, \ldots, p_k) + \Psi_{m-1}(p_1, \ldots, p_{k-1})\big) + 1 = \sum_{m=0}^{m=k-1} m! \times \Psi_m(p_1, \ldots, p_k + 1).$$

This proves the result by induction and shows that the number of internal vertices is equal to

$$k! \times p_1 \times \ldots \times p_k.$$

□

3. On some subautomatas of the minimal automaton of the shuffle of a set of words $u_1, u_2, ..., u_k$

In this section we prove that, for every sequence of positive integers $(p_1, \ldots, p_k)$, there exist words $u_1, u_2, ..., u_k$ such that the minimal automaton of $L = u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$ contains an araucaria of type $(p_1, \ldots, p_k)$.

Definition 3.1. (i) Let u and v be two words of the free monoid A^*.
The language whose words are of the form $u_1 v_1 u_2 v_2 \ldots u_m v_m$ where $u_1 u_2 \ldots u_m$ is a factorisation of u, $v_1 v_2 \ldots v_m$ a factorisation of v and the factors u_1 and v_m are possibly empty, is called the shuffle of the words u and v and is denoted $u \sqcup\!\sqcup v$.
(ii) If I and J are two languages of A^*, the union of the sets $u \sqcup\!\sqcup v$ for $u \in I$ and $v \in J$ is called the shuffle of the languages I and J and is denoted $I \sqcup\!\sqcup J$.
(iii) Let $u_1, \ldots, u_k$ be k words of A^*. If we assume that the shuffle K of the words $u_1, \ldots, u_{k-1}$ is defined, the language $L = K \sqcup\!\sqcup u_k$ is called the shuffle of the words $u_1, \ldots, u_k$ and is denoted

$$L = u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k.$$

Remark 3.1. If all words of a langage L have the same length, then the minimal automaton of L has a unique terminal state and an absorbing state z ($\forall a \in A, z.a = z$). In particular, this is the case for the langage $L = u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$.

Definition 3.2. (i) Let $u = a_1 \ldots a_n$ be a word of length n of A^*. The partial automaton $PA(u)$ whose set of states is $\{s_0, \ldots, s_n\}$, whose transitions are (s_i, a_i, s_{i+1}) with $i \in \{0, \ldots, n-1\}$, s_0 the initial state and s_n the terminal state, is called the partial minimal automaton of u.
(ii) Let $u_1, \ldots, u_k$ be k words of A^*, $L = u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$, $A(L)$ the minimal automaton of L, d its

initial state and f its terminal state.
The partial automaton $PA(L)$ obtained from $A(L)$ by deleting his absorbing state z and all transitions towards z or issued from z, is called the partial minimal automaton of L.
(iii) The directed graph whose vertices are the states of $PA(L)$ and whose edges are the pairs (s,t) such that there exists a letter $a \in A$ such that (s,a,t) is a transition of $PA(L)$ is called the graph of $PA(L)$.

Definition 3.3. (i) Let S_0 be the set of states of the partial minimal automaton $PA(K)$, d_0 the initial state of $PA(K)$, f_0 its unique terminal state and $u_k = a_1 \dots a_n$.
For each $i \in \{0,\dots,n\}$, let $\theta_i : s \to s^{(i)}$ be an isomorphism from $PA(K)$ on an automaton $PA(K)^{(i)}$ such that the sets $S^{(i)} = \theta_i(S_0)$ are two by two disjoint.
The non-deterministic automaton $M_1(L)$ which is the disjoint union of the partial automata $PA(K)^{(0)}$, $\dots$, $PA(K)^{(n)}$ and which admits, for each $s \in S_0$ and each $i \in \{0,\dots,n-1\}$, $(s^{(i)}, a_{i+1}, s^{(i+1)})$ as transition, is called the shuffle product of the partial automata $PA(K)$ and $PA(u_k)$ and is denoted $PA(K) \sqcup\!\sqcup PA(u_k)$.
It admits $S_1 = S_0^{(0)} \cup \dots \cup S_0^{(n)}$ as set of states, $d_1 = d_0^{(0)}$ as initial state and $f_1 = f_0^{(n)}$ as terminal state (see [5] for a more general definition and for complements).
Each transition of the form $(s^{(i)}, a_{i+1}, s^{(i+1)})$ is called vertical and each transition of one of the partial automata $PA(K)^{(i)}$ is called horizontal.
(ii) Let $M_2'(L)$ be the subautomaton of the automaton of subsets of $PA(K) \sqcup\!\sqcup PA(u_k)$ generated by $d_2 = \{d_1\}$.
The partial subautomaton $M_2(L)$ of the automaton $M_2'(L)$ obtained by deleting the empty set of S_1, is called the determinization of $M_1(L) = PA(K) \sqcup\!\sqcup PA(u_k)$.

Lemma 3.1. (i) The non-deterministic automaton $PA(K) \sqcup\!\sqcup PA(u_k)$ recognizes language L.
(ii) There exists a morphism from $M_2(L)$ onto the partial automaton $PA(L)$.
(iii) If the alphabet of the word u_k is disjoint from that of the language K, then the automata $PA(L)$, $M_2(L)$ and $PA(K) \sqcup\!\sqcup PA(u_k)$ are isomorphic.

Proof:
The proof of this lemma is straightforward. □

Definition 3.4. (i) Each factorisation (x, a^p, y) of a word u where a is a letter of A which is not a right factor of x nor a left factor of y, is called an a-factorisation of u.
Such an a-factorisation is called degenerate if $p = 0$.
If, for each $i \in \{1,...,k\}$, (x_i, a^{p_i}, y_i) is an a-factorisation of word u_i and if, at least one of these a-factorisations is not degenerate, $\psi = ((x_i, a^{p_i}, y_i)_{i\in\{1,\dots,k\}})$ is called an a-factorisation of $(u_1,\dots,u_k)$.
(ii) Let d be the initial state of the partial minimal automaton $PA(L)$ and let ψ be an a-factorisation of $(u_1,\dots,u_k)$.
Let T be the set of states t such that there exist
- left factors v_1,..., v_k respectively of u_1,..., u_k such that, for all i of $\{1,...k\}$, $|x_i| \le |v_i| \le |x_i| + p_i$
- and a word v of $v_1 \sqcup\!\sqcup \dots \sqcup\!\sqcup v_k$ such that $t = d.v$.
The partial automaton N of $PA(L)$ which admits T as set of states and the a-transitions of the form $(t, a, t.a)$ such that $t.a \in T$ as only transitions, is called the nest of a-transitions associated to the a-factorisation ψ (see Figure 4.b).

The number of positive integers p_i is called the dimension of the nest N.
(iii) For each $i \in \{1, ..., k\}$, such that $h_i \neq 0$ and $x_i \neq 1$, the last letter b_i of x_i is called the entry letter of N of index i and, for each $i \in \{1, ..., k\}$, such that $y_i \neq 1$, the last letter c_i of y_i is called the exit letter from N of index i.
Each state t of T with no predecessor in T, is called entry state of N.
Each state t such that there exists a transition relatively to an exit letter c_i of N issued from t, is called an exit state.

Definition 3.5. (i) Let N_0 be a nest of a-transitions of $PA(K)$ associated to the a-factorisation $\psi_0 = ((x_i, a^{p_i}, y_i)_{i\in\{1,...,k-1\}})$ of $(u_1, \ldots, u_{k-1})$, (x_k, a^{p_k}, y_k) an a-factorisation of u_k and $r_k = |x_k|$.
If T_0 is the set of states of N_0 and, if for each $i \in \{r_k, \ldots, r_k + p_k\}$, $T^{(i)} = \theta_i(T_0)$, the partial subautomaton N_1 of $M_1(L) = PA(K) \sqcup\!\sqcup PA(u_k)$ which admits $T_1 = T^{(r_k)} \cup \ldots \cup T^{(r_k+p_k)}$ as set of states, admits only transitions relatively to a and is isomorphic to $N_0 \sqcup\!\sqcup PA(a^{p_k})$.
N_1 is called the non-deterministic nest of a-transitions of $M_1(L)$ associated to the a-factorisation $\psi = ((x_i, a^{p_i}, y_i)_{i\in\{1,...,k\}})$ (see Figure $4.a$).
(ii) If S_2 is the set of states of $M_2(L)$, let T_2 be the set of elements of S_2 which contain at least one state of the nest N_1.
The partial subautomaton N_2 of $M_2(L)$ which admits T_2 as set of states and only transitions relatively to the letter a, is called the determinization of N_1.

In the sequel of this section we will study the following particular case:
$\forall i \in \{1, \ldots, k\}$, the word u_i is of the form $u_i = b_i a^{p_i} c_i$ with positive integers $p_1, \ldots, p_k$ and the letters of $\{b_1, \ldots, b_k\} \cup \{c_1, \ldots, c_k\}$ are two by two distinct up to the equalities $b_1 = c_1, \ldots, b_k = c_k$.
Such a set $(u_1, \ldots, u_k)$ is called special.

Lemma 3.2. Let $\psi = ((b_i, a^{p_i}, c_i);\ 1 \leq i \leq k)$ an a-factorisation of the special set $(u_1, \ldots, u_k)$. Then the graph of the nest N of a-transitions associated to ψ in the partial minimal automaton $PA(L)$ is the opposite of a directed tree.

Proof:

We prove this property by induction on k.
For $k = 1$, the graph of the nest of a-transitions of $PA(u_1)$ is a path of length p_1. The property is therefore verified.
Assume that the property is verified for each special set of $k-1$ words and let $(u_1, \ldots, u_k)$ be a special set of k words. We use the notations of Definition 3.5.

(i) First we study the part of the graph obtained from the entries corresponding to letter b_k.
For all $i \in \{1, \ldots, k-1\}$, $h_i \in \{0, \ldots, p_i\}$ and $x \in b_1 a^{h_1} \sqcup\!\sqcup \ldots \sqcup\!\sqcup b_{k-1} a^{h_{k-1}}$, $d_0^{(0)}.x$ is reduced to a unique state $s^{(0)}$ of $S^{(0)}$ since $b_k \notin \{b_1, \ldots, b_{k-1}\}$ and, therefore, $d_0^{(0)}.x b_k = \{s^{(1)}\}$. Each entry relative to b_k is therefore reduced to a unique state of $S^{(1)}$.
Let $\psi_0 = ((x_i, a^{p_i}, y_i)_{i\in\{1,...,k-1\}})$ be the a-factorisation of $(u_1, \ldots, u_{k-1})$. The opposite graph G_0 of the graph of the nest N_0 associated to ψ_0 is a directed tree by induction. Let $(s_q, s_{q-1}, \ldots, s_0)$ be the unique path from the root $s_q = r_0$ of the directed tree G_0 to the vertex $s_0 = s$.
For each h of $\{0, \ldots, min(p_k, q)\}$,

$$\{s^{(1)}\}.a^h = \{s_0^{(1+h)}, s_1^{(h)} \ldots, s_h^{(1)}\}.$$

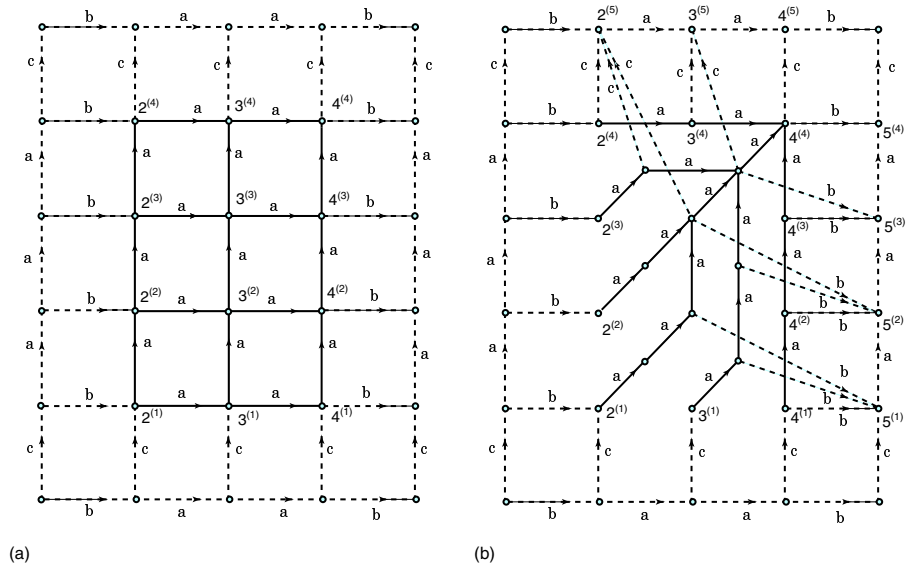

Figure 4. Let $u_1 = baab$, N_0 the nest of $PA(u_1)$ associated to the a-factorisation (b, a^2, b) of u_1, $u_2 = caaac$ and $L = u_1 \sqcup\!\sqcup u_2$. (a) The non-deterministic nest $N_1 = N_0 \sqcup\!\sqcup PA(a^3)$ in $PA(u_1) \sqcup\!\sqcup PA(u_2)$; (b) The corresponding nest N in the partial minimal automaton $PA(L)$ and its subgraph which is the opposite of an araucaria of type $(2, 3)$.

For each h of $\{min(p_k, q) + 1, \ldots, max(p_k, q)\}$,

$$\{s^{(1)}\}.a^h = \{s^{(1+p_k)}_{h-p_k}, \ldots, s^{(1)}_h\} \text{ when } p_k < q \text{ and}$$
$$\{s^{(1)}\}.a^h = \{s^{(1+h)}_0, \ldots, s^{(h+1-q)}_q\} \text{ when } q < p_k.$$

For each h of $\{max(p_k, q) + 1, \ldots, p_k + q\}$,

$$\{s^{(1)}\}.a^h = \{s^{(1+p_k)}_{h-p_k}, \ldots, s^{(h+1-q)}_q\}$$

In particular $\{s^{(1)}\}.a^{p_k+q} = \{s^{(1+p_k)}_q\}$ and this proves that there exists, in the opposite graph G_2 of the graph of the nest N_2, a path from the vertex $\{r^{(1+p_k)}_0\}$ to the vertex $\{s^{(1)}\}$.

(ii) We study now the part of the graph obtained from entries corresponding to a letter b_j distinct from b_k.

As $\forall i \in \{1, \ldots, k\} \setminus \{j\}$ and $\forall x \in b_1 a^{h_1} \sqcup\!\sqcup \ldots \sqcup\!\sqcup b_{j-1} a^{h_{j-1}} \sqcup\!\sqcup b_{j+1} a^{h_{j+1}} \sqcup\!\sqcup \ldots \sqcup\!\sqcup b_k a^{h_k}$, where $h_i \in \{0, \ldots, p_i\}$ there exists a left factor y of x and a right factor z of x such that $x = y b_k z$ and the set $d^{(0)}_0.y b_k$ is reduced to a unique state $s^{(1)}$ of $S^{(1)}$. Each occurrence of a letter b_i in z induces a horizontal transition in $M_1(L)$ although each occurrence of the letter a in z induces simultaneously a horizontal transition and a vertical transition and, when this vertical transition is used, the corresponding horizontal transition is occulted.

Let $\Lambda(x)$ be the set of pairs (z_λ, r_λ) such that z_λ is obtained by deleting r_λ occurrences of the letter a in z with $0 \leq r_\lambda \leq p_k$.

$$d^{(0)}_0.x b_j = \{(s.z_\lambda b_j)^{(1+r_\lambda)}; (z_\lambda, r_\lambda) \in \Lambda(x)\}$$

is then an entry relative to letter b_j in the nest N_2 and, for each integer $r \geq 0$,

$$d_0^{(0)}.xb_ja^r = \bigcup_{(z_\lambda, r_\lambda)\in\Lambda(x)} (s.z_\lambda b_j)^{(1+r_\lambda)}.a^r.$$

Since G_0 is a directed tree, for each state t of G_0 there exists a path from the root r_0 of G_0 to t and therefore, if q is the length of this path, then $t.a^q = r_0$ and $\{t^{(1)}\}.a^{q+p_k} = \{r_0^{(1+p_k)}\}$. Since all states of $X = d_0^{(0)}.xb_j$ have the same rank, there exists therefore also an integer q such that $X.a^{q+p_k} = \{r_0^{(1+p_k)}\}$. There exists therefore also a path from $\{r_0^{(1+p_k)}\}$ to X in the opposite graph G_2 of the graph of the nest N_2.

(iii) By (i) and (ii), for each state s of G_2, there exists a path $(s_0, \ldots, s_q)$ from $s_0 = r_0^{(1+p_k)}$ to $s_q = s$ and, since $\forall i \in \{1, \ldots, q\}$, $s_i.a = s_{i-1}$ in the automaton $M_2(L)$, this path is unique. This proves that G_2 is a directed tree.
Since, by Lemma 3.1, there exists a morphism κ from $M_2(L)$ on the partial automaton $PA(L)$, the opposite graph $G = \kappa(G_2)$ of the nest associated to the a-factorisation ψ is also a directed tree. □

Definition 3.6. The word u is called left factor of the language K if there exists a word v such that $uv \in K$.
Let $d' = d.u_k$, Q the set of states s of S such that there exists a left factor u of K such that $s = d'.u$ in $PA(L)$ and, for such an s, $U(s)$ the set of transitions $(s, a, s.a)$ with $a \in A$ such that ua is a left factor of K.
The partial subautomaton $Ult(L)$ of $PA(L)$ which admits Q as set of states and, for each $s \in Q$, $U(s)$ as set of transitions issued from s, is called the ultimate face of $PA(L)$.

Lemma 3.3. For each special set $(u_1, \ldots, u_k)$, the ultimate face of $PA(L)$ with d' as initial state and f as terminal state, is isomorphic to $PA(K)$.

Proof:

Since c_k does not belong to the alphabet of K, $\{d_0^{(0)}\}.u_k = \{d_0^{(p_k+2)}\}$ in $M_2(L)$.
Assume that there exists a left factor u of K such that $\{d_0^{(p_k+2)}\}.u$ contains a state $t^{(h)}$ of $S_1 \setminus S^{(p_k+2)}$. Then there exists a right factor v_1 of u_k distinct from the empty word and a right factor v_2 of K such that $\{t^{(p_k+2)}\}.v_1v_2 = \{f_0^{(p_k+2)}\}$. Then $\{d_0^{(0)}\}.u_kuv_1v_2 = \{f_0^{(p_k+2)}\}$ and, by Lemma 3.1, $w = u_kuv_1v_2 \in L$ which is not possible because w contains an excess occurrence of c_k. It follows that, for each left factor u of K, $\{d_0^{(p_k+2)}\}.u \subset S^{(p_k+2)}$ and, since θ_{p_k+2} is a morphism from $PA(K)$ on $PA(K)^{(p_k+2)}$, $\{d_0^{(p_k+2)}\}.u$ is reduced to the unique element $\theta_{(p_k+2)}(d_0.u)$.
If κ is the morphism from $M_2(L)$ on $PA(L)$ of Lemma 3.1, $\kappa(\theta_{(p_k+2)}(d_0)) = \kappa(\{d_0^{(0)}\}.u_k) = d.u_k = d'$ and, for each left factor u of K, $\kappa(\theta_{(p_k+2)}(d_0.u)) = \kappa(\{d_0^{(p_k+2)}\}.u) = d'.u$. $\kappa o\theta_{(p_k+2)}$ is therefore a morphism from $PA(K)$ on $PA(L)$ whose image is the ultimate face $Ult(L)$ of $PA(L)$ and, since this morphism is injective, $Ult(L)$ is isomorphic to $PA(K)$. □

Lemma 3.4. Let N be the nest of a-transitions associated to the a-factorisation $\psi = ((x_i, a^{p_i}, y_i)_{1\leq i\leq k})$ of the special set $(u_1, \ldots, u_k)$ in the automaton $PA(L)$ and let N_0 be the nest of a-transitions associated to the a-factorisation $\psi_0 = ((x_i, a^{p_i}, y_i)_{i\in\{1,\ldots,k-1\}})$ of $(u_1, \ldots, u_{k-1})$ in the automaton $PA(K)$.
If G_0 is the opposite graph of the graph of the nest N_0, then the opposite graph G of the graph of the nest N admits a subgraph which is isomorphic to the ramified directed tree $ramif(G_0, A(p_k))$.

Proof:

(i) By the proof of Lemma 3.2, each entry of $N_1 = N_0 \sqcup\!\sqcup PA(u_k)$ relative to letter b_k is of the form $\{s^{(1)}\}$ with $s \in T_0$ and, the sets $\{s^{(1)}\}.a^h$ for $h \in \{0, \ldots, p_k\}$, are two by two disjoint. There exists therefore a path $c(s)$ from $\{s^{(1)}\}.a^{p_k}$ to $\{s^{(1)}\}$ in the graph G_2 which is the opposite of the graph of the nest N_2 in the determinization $M_2(L)$ of $M_1(L)$ and, since the length of $c(s)$ is equal to p_k, $c(s)$ is a path isomorphic to the elementary araucaria $A(p_k)$.
Consider the mapping $\theta : s \to \{s\}.a^{p_k}$ from N_0 to N_2. For each edge (t, s) of G_0, $(s, a, t = s.a)$ is a transition and, therefore

$$\theta(t) = \{t^{(1)}\}.a^{p_k} = \{(s.a)^{(1)}\}.a^{p_k} = (\{s^{(1)}\}.a^{p_k}).a = (\theta(s)).a.$$

Hence $(\theta(t), \theta(s))$ in an edge of the graph G_2 and θ is a morphism from the graph G_0 to the graph G_2. Since θ is injective, the subgraph $\theta(G_0)$ of G_2 is isomorphic to G_0.
Moreover, for all states s and s' of G_0, the paths $c(s)$ and $c(s')$ are disjoint if $s \neq s'$.
It results from them that the graph G_2 admits a subgraph H_2 which is isomorphic to the ramified directed tree $ramif(G_0, A(p_k))$.

(ii) By Lemma 3.1, there exists a morphism κ from $M_2(L)$ to $PA(L)$. For all states s and s' of the nest N_0 such that $s \neq s'$, $\theta(s).c_k = \{s^{(2+p_k)}\}$ and $\theta(s').c_k = \{s'^{(2+p_k)}\}$. But by the proof of Lemma 3.3, the states $s^{(2+p_k)}$ and $s'^{(2+p_k)}$ of $PA(K)^{(2+p_k)}$ are not equivalent. $\theta(s)$ and $\theta(s')$ are not equivalent in $M_2(L)$ and, therefore, $\kappa(\theta(s)) \neq \kappa(\theta(s'))$.
Moreover, for each state s of the nest N_0, the states of the path $c(s)$ form a sequence of a-transitions from $\{s^{(1)}\}$ to $\{s^{(1)}\}.a^{p_k}$ in $M_2(L)$. The states of the path $c(s)$ are therefore not equivalent, nor equivalent to a state of another path $c(s')$.
The restriction of κ to the subgraph H_2 of the graph G_2 and to the graph G which is the opposite of the graph of the nest N of $PA(L)$ is therefore injective and this proves that G admits also a subgraph which is isomorphic to a ramified directed tree $ramif(G_0, A(p_k))$. □

Theorem 3.1. If the set $(u_1, \ldots, u_k)$ is special, the graph of the nest N of a-transitions associated to the a-factorisation $\psi = ((b_i, a^{p_i}, c_i)_{1 \leq i \leq k})$ of $(u_1, \ldots, u_k)$ in the automaton $PA(L)$ is the opposite of an araucaria of type $(p_1, \ldots, p_k)$

Proof:

(i) The property is trivial for $k = 1$.
Assume that, for each sequence $(p_1, \ldots, p_{k-1})$ of positive integers, the graph of the nest associated to the a-factorisation $\psi_0 = ((b_i, a^{p_i}, c_i)_{i \in \{1,\ldots,k-1\}})$ of $(u_1, \ldots, u_{k-1})$ in the partial minimal automaton $PA(K)$ is the opposite of an araucaria of type $(p_1, \ldots, p_{k-1})$.
By Lemma 3.2, the opposite graph G of the graph of the nest N contains a directed subtree A'_k which is isomorphic to the ramified directed tree $ramif(A_k, A(p_k))$ where $A_k = A(p_1, \ldots, p_{k-1})$.
Since, for each permutation χ of $\{1, \ldots, k\}$, the languages $u_{\chi(1)} \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_{\chi(k)}$ and $u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$ are the same, the graph G contains also, for each $i \in \{1, ..., k-1\}$, a directed subtree A'_i which is isomorphic to $ramif(A_i, A(p_i))$ where A_i is an araucaria of type $(p_{i+1}, \ldots, p_k, p_1, \ldots, p_{i-1})$.
Since A'_k contains all entries relative to letter b_k, the same thing happens for the other entry letters $b_1, \ldots, b_{k-1}$. Therefore G is covered by the directed trees $A'_1, \ldots, A'_k$.

(ii) For each entry e relative to a letter b_i in the nest N_0 associated to the a-factorisation ψ_0 in the

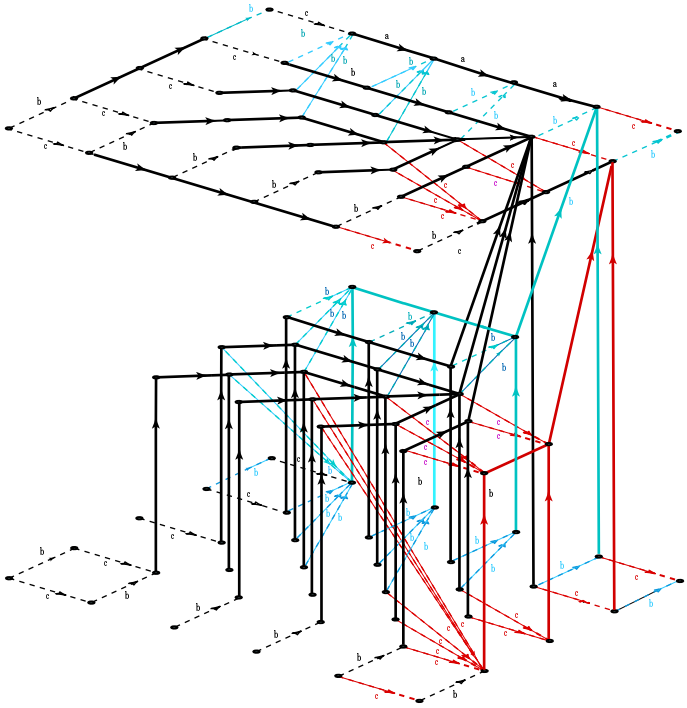

Figure 5. Let $u_1 = baab$, $u_2 = caaac$, $u_3 = dad$, $L = u_1 \sqcup\!\sqcup u_2 \sqcup\!\sqcup u_3$, N the nest of $PA(L)$ associated to the a-factorisation $((b, a^2, b), (c, a^3, c), (d, a, d))$ of (u_1, u_2, u_3). The directed subtree of the graph G opposite of the graph of the nest N which is isomorphic to $A'_1 = ramif(A(2, 3), A(1))$ of an araucaria of type $(2, 3)$ and the exit transitions of the states supported by the trunk of this araucaria. Only the a-transitions have been drawn with unbroken lines and some b-transitions and c-transitions with dotted lines.

automaton $PA(K)$, $e^{(1)}$ is simultaneously an entry for b_i and for b_k (see the proof of Lemma 3.2). If κ is a morphism from $M_2(L)$ on $PA(L)$, there exists therefore, in G, a maximal path σ from the root r of G to vertex $\kappa(e^{(1)})$ and this vertex is common to subtrees A'_i and A'_k. The directed trees A'_i and A'_k are therefore merged with respect to the maximal path σ. The same argument applies for each common entry letter of any number of entry letters of $\{b_1, \ldots, b_{k-1}\}$ and, permuting the words $u_1, \ldots, u_k$, to each part of $\{b_1, \ldots, b_k\}$.

First we merge the paths issued from the root which end with the unique entry common to all entry letters $b_1, \ldots, b_k$, then we do the same thing for the entries common to $k-1$ letters and so forth until the common entries of two letters. The operation consists then in merging the terminal truncations.

G is therefore isomorphic to the shuffle product of the elementary araucarias $A(p_1), \ldots, A(p_k)$ and is, by Theorem 2.2, isomorphic to an araucaria of type $(p_1, \ldots, p_k)$. □

4. Shuffle products of words and partial commutations

Definition 4.1. Given an alphabet A and a set Θ of pairs (a, b) of letters of A such that $a \neq b$, let w and w' be words of A^* such that $w = uabv$ and $w' = ubav$ with $(a, b) \in \Theta$ or $(b, a) \in \Theta$: the transformation of w in w' is called a partial commutation.

If w is a word of A^*, the set of words $w' \in A^*$ such that there exists a sequence of partial commutations which transforms w in w' is called the commutation class of w.

The present authors have designed an efficient sequential algorithm for the generation of commutation classes [14] and two optimal parallel algorithms on the commutation class of a given word [15]. During our investigation, it became clear to us that some of our results extend to the shuffle product of words. In particular, we have been motivated by the fact that all words in a commutation class have the same length as in shuffle products of words and by the following result.

Proposition 4.1. If the respective alphabets $A_1, \ldots, A_k$ of the words $u_1, \ldots, u_k$ are two by two disjoint, then the shuffle product $L = u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$ is equal to the commutation class of the word $u_1 ... u_k$ relative to $\Theta = \bigcup_{1 \leq i < j \leq k} A_i \times Aj$.

Proof:
The proof of this proposition is straightforward. □

Remark 4.1. This result does not remain true if the words $u_1, \ldots, u_k$ share a common letter. In the case where the words $u_1, \ldots, u_k$ are special (see Section 3), we are far away from the assumptions of Proposition 4.1 and it appears that the study of the shuffle of words is much more intricate than the study of commutation classes.
Partial commutation theory assumes that a letter never commutes with itself but, in the shuffle product, the occurrences of the same letter of two distinct words still commute. This fundamental difference implies that the minimal automaton of the shuffle of words contains graphs such as araucarias.

5. Conclusion

In this paper we have investigated directed trees which appear in the construction of the minimal automaton of the shuffle of words. We hope to be able to prove that, if the set of words is special, then the size of the partial minimal automaton of the shuffle of words is maximal. The design of an optimal algorithm for the construction of this automaton is under investigation by the present authors.

Acknowledgments: The authors are grateful to anonymous referees for pertinent comments and suggestions.

References

[1] Allauzen C., Calcul efficace du shuffle de k mots, Prépublication de l'Institut Gaspard Monge, 36, 2000.

[2] Berstel J. and Boasson L., Shuffle factorization is unique, *Theoretical Computer Science*, **273**, 2002, 47-67.

[3] Cori R. and Perrin D., Automates et commutations partielles, *RAIRO Inf. Theor.*, **19**, 1985, 21-32.

[4] Diekert V., Combinatorics of traces, *Lecture Notes in Computer Science*, **454**, Springer Verlag, 1990.

[5] Eilenberg S., Automata, Languages and Machines, Academic Press, 1974.

[6] Françon J., Une approche quantitative de l'exclusion mutuelle, *Theoretical Informatics and Applications*, **20**, 3, 1986, 275-289.

[7] Gómez A. C. and Pin J.-E., On a conjecture of Schnoebelen, *Lecture Notes in Computer Science*, **2710**, 2003, 35-54.

[8] Lothaire M., Combinatorics on words, Addison-Wesley, Reading, MA, 1982.

[9] Mateescu A., Shuffle of ω-words: algebraic aspects, Proceedings of STACS 98, *Lecture Notes in Computer Science*, **1373**, 150-160, Springer Verlag 1998.

[10] Nivat M., Ramkumar G.D.S., Pandu Rangan C., Saoudi A., and Sundaram R., Efficient parallel shuffle recognition, *Parallel Processing Letters*, **4**, 1994, 455-463.

[11] Pradeep B. and Siva Ram Murthy C., A constant time string shuffle algorithm on reconfigurable meshes, *Internat. J. Comput. Math.*, **68**, 1998, 251-259.

[12] Radford D. E., A natural ring basis for the shuffle algebra and an application to group schemes, *J. Algebra*, **58**, 1979, 432-454.

[13] Schnoebelen P., Decomposable regular languages and the shuffle operator, *EATCS Bull.*, **67**, 1999, 283-289.

[14] Schott R. and Spehner J.-C., Efficient generation of commutation classes, *Journal of Computing and Information*, **2**, 1, 1996, 1028-1045.
Special issue: Proceedings of Eighth International Conference of Computing and Information (ICCI'96), Waterloo, Canada, June 19-22, 1996.

[15] Schott R. and Spehner J.-C., Two optimal parallel algorithms on the commutation class of a word, *Theoretical Computer Science*, **324**, 2004, 107-131.

[16] Schott R. and Spehner J.-C., On the minimal automaton of the shuffle of words and araucarias (Extended abstract), Proceedings of MCU-2004 (International Conference on Machines, Computing and Universality, St Petersburg, September 2004), *Lecture Notes in Computer Science*, **3354**, 316-327, Springer Verlag 2005.

[17] Spehner J.-C., Le calcul rapide des mélanges de deux mots, *Theoretical Computer Science*, **47**, 1986, 181-203.

[18] Zielonka W., Notes on finite asynchronous automata and trace languages, *RAIRO Inf. Theor.*, **21**, 1987, 99-135.