N° d'ordre : 2317

# THÈSE

PRÉSENTÉE À

# L'UNIVERSITÉ BORDEAUX I

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

**Par Cedric CHAUVE**

POUR OBTENIR LE GRADE DE

## DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

---

## Structures arborescences : problèmes algorithmiques et combinatoires

---

**Soutenue le :** 11 décembre 2000

**Après avis de :** MM. Dominique Gouyou-Beauchamps Rapporteurs
Mireille Régnier .................
Volker Strehl .....................

**Devant la Commission d'examen formée de :**

| | | | |
|---|---|---|---|
| MM. | Robert Cori | Professeur ........................ | Président |
| | Mireille Bousquet-Mélou | Chargée de recherche au CNRS .. | Rapporteur |
| | Maxime Crochemore | Professeur ........................ | Examinateur |
| | Serge Dulucq | Professeur ........................ | Directeur |
| | Dominique Gouyou-Beauchamps | Professeur ........................ | Examinateur |
| | Mireille Régnier | Directrice de recherche à l'INRIA | Examinatrice |
| | Volker Strehl | Professeur ........................ | Examinateur |

– 2000 –

Au cours de ces trois années de thèse, j'ai croisé de nombreuses personnes que je me suis promis de remercier… que les absents de la (pourtant longue) liste qui suit veuillent bien me pardonner.

Le professeur Volker Strehl a assisté à mes premiers pas balbutiants hors du doux cocon du LaBRI, lors d'un séminaire lotharingien de combinatoire en Allemagne, et sa gentillesse fut d'un grand secours. Je suis donc particulièrement sensible au fait qu'il ait accepté le lourd travail de rapporter cette thèse et d'assister à cette soutenance.

À la sortie de mon DEA, j'ai passé de longues semaines à sécher sur les tableaux de Young de hauteur bornée. Une des motivations de cette tentative résidait dans les beaux articles de Dominique Gouyou-Beauchamps sur ce sujet (qui ont donc accompagné mes premiers échecs). Je suis très heureux de le compter parmi mes rapporteurs et dans mon jury.

Je remercie Mireille Régnier d'avoir apporté le point de vue de l'algorithmicienne sur ce travail.

Le livre de Maxime Crochemore sur l'algorithmique des mots fut mon livre de chevet durant ma première année de thèse. Je suis donc très heureux qu'il ait accepté de participer à ce jury.

Il est à la fois étrange et agréable de retrouver dans son jury ses anciens enseignants (de "quand on était petit"). Ma passion pour l'algorithmique remontant au cours de licence de Robert Cori, je suis particulièrement heureux de sa participation à ce jury. J'ajoute que ses questions et commentaires lors de mes divers exposés en groupe de travail ont sensiblement contribué au développement de cette thèse.

Il y aurait vraiment trop à dire pour exprimer la joie que j'éprouve à compter Mireille Bousquet-Mélou (Mimi) dans mon jury, tant pour des raisons scientifiques (je ne cite que les constellations) qu'extra-scientifiques (vélocipédiques notamment). Je m'arrête donc là avant de noircir une page entière.

Enfin, mon "chef", Serge. Pour exprimer toute ma gratitude à son égard, je dirais qu'en me retournant sur ces trois années écoulées, je m'aperçois qu'il m'a permis de m'épanouir scientifiquement, encadrant de réunions hebdomadaires mes débuts hésitants avant de me laisser progressivement voler de mes propres ailes vers d'autres horizons (à la fois coloriés et québécois), sans toutefois jamais me perdre de vue. J'ajoute que sa relecture attentive (doux euphémisme) de ce mémoire est pour beaucoup dans le fait qu'il soit lisible (je l'espère).

En parlant de chef, je ne peux m'empêcher de dire ici toute l'admiration que je porte à Pierre Leroux, qui m'a gentiment accueilli au LaCIM et m'a guidé dans le monde fascinant de la théorie des espèces. Je m'estime privilégié d'avoir pu travailler avec lui et son compère Gilbert Labelle.

Continuons en mélangeant joyeusement les nombreux combinatoriciens que j'ai régulièrement croisés que ce soit au LaBRI ou au LaCIM : Sacha (dont la gentillesse semble n'avoir d'égal que l'immensité de ses connaissances), Bétréma (Canaris 0 - Girondins 5), Srećko Brlek, François Bergeron, Olivier Guibert, Jean-Guy Penaud, Xavier Viennot, …

L'un des aspects les plus agréables du petit monde de la combinatoire est que chaque conférence est l'occasion de retrouver, en général autour d'une bonne bière, un groupe de jeunes camarades : Gilles (en concaténant les nombreuses questions que je lui ai posées on doit pouvoir couvrir la distance séparant Bordeaux de Nancy), Dominique (compagne d'infortune lors du séminaire lotharingien sus-cité … ça crée des liens), Sylvie (que je n'aurais jamais cru remercier après notre première rencontre), Mike (à qui je dois notamment la découverte du Rocky Horror Picture Show), Étienne, Luigi, Marni (qui m'a presque persuadé que le tofu peut s'apprêter), Michel (dont l'appartement à Montréal fut un nid douillet durant 4 mois) et les nombreux étudiants du LaCIM. Je ne peux oublier de mentionner les amis de Michel : Francis, Michel2 et Nathalie.

Si j'ai beaucoup apprécié mon temps au LaBRI, les gens que j'y ai croisés y sont sans doute pour beaucoup. Je remercie donc en vrac TiNico et GrandNico (qui m'ont initié au culte du dieu Tipunch), Driss (dont la recette de tajine fait encore partie de mes préparations favorites), Akka (dangereux compagnon de voyage), Sami, Laurent, Guillaume, Anne (douce Anne), Olivier (qui finira bien par trouver un vélo), JC, Christophe, Jaco, Isa... Une mention spéciale à Pierre (pour Cortazar, Coltrane, Cassavetes, et bien d'autres choses), Hélène (qui a fini par céder aux sirènes du grand capital), Duduche (dangereux probabiliste) et Stéphane (à qui je dois la recette de ma désormais célébrissime tarte au citron). Pour clore le chapitre LaBRI, je ne peux qu'évoquer les amis qui ont partagé mes divers bureaux (les bars bordelais fréquentés en commun en faisant partie) : Olive (un jour on finira bien par ne plus être dans le même bureau et tu pourras fumer), Jeff (maintenant indissociable de la délicieuse Élisa), Manu (le squatter du jeudi) et Andrew (sur qui j'aurais trop de choses à dire pour le peu de place qui me reste).

Cette thèse fut aussi l'occasion de découvrir le beau métier d'enseignant ... que les collègues et étudiants qui ont accompagné cette découverte soient remerciés à la fois pour leur indulgence face à mes balbutiements et pour le plaisir que j'ai eu à travailler avec eux.

Enfin, pour passer certains caps difficiles au cours de ces dernières années, j'ai toujours su pouvoir compter sur le soutien indéfectible des mes amis au long cours (Henri, Yannis, Lionel, Xavier, Sandrine1, Sandrine2, Christine, Thierry), de mes parents et de mon globetrotter de frère.

# Table des matières

# Introduction

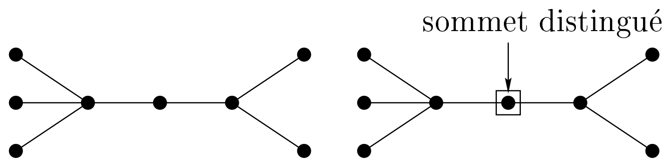
Les travaux que nous présentons dans ce mémoire se situent dans deux domaines connexes de l'informatique, la combinatoire, et plus précisément la combinatoire énumérative, et l'algorithmique. Nous nous sommes intéressés à divers problèmes ayant en commun le fait de manipuler des *structures arborescentes*.

Si la notion de structure arborescente n'a été formalisée, en termes mathématiques, que tardivement (au XIX$^{ème}$ siècle), la présentation arborescente d'un ensemble de données est une pratique bien antérieure. On peut par exemple citer les arbres généalogiques, mais aussi la représentation de vertus par des arbres de la sagesse dans certains livres médiévaux, ou encore les arbres de classification d'espèces animales des naturalistes des XVII$^{ème}$ et XVIII$^{ème}$ siècles.

La première formalisation mathématique du concept d'arbre est généralement attribuée à Kirchhoff et est issue de travaux portant sur l'étude des circuits électriques.

Un *arbre* est un graphe connexe acyclique.
Une *arborescence* est un arbre ayant un sommet distingué[1].

La figure suivante représente un arbre (à gauche) et une arborescence (à droite).




C'est cependant Cayley qui a introduit le terme d'arbre, dans une série d'articles (le premier datant de 1857), portant principalement sur l'énumération d'arbres et d'arborescences en relation avec l'étude de formules algébriques [30]. À la suite de ces premiers travaux de Cayley, l'étude des propriétés des structures arborescentes s'est rapidement imposée comme un domaine classique et fécond de la combinatoire, notamment de la combinatoire énumérative (on peut se reporter aux ouvrages de Moon [120], Harary et Palmer [88, chapitre 3], Stanley [148, chapitre 5], Bergeron, Labelle et Leroux [16]).

Cet intérêt pour les structures arborescentes a ensuite connu un nouvel essor avec le développement de l'informatique. Selon Knuth, les arborescences sont en effet "les plus importantes structures non linéaires apparaissant en algorithmique" [102, page 308]. De nombreux algorithmes manipulant des arborescences (notamment cette structure de donnée

1. On peut noter que cette terminologie, issue des travaux de théorie des graphes (voir par exemple le livre de Berge [14]), est différente de la terminologie utilisée couramment dans les travaux traitant d'algorithmique des structures arborescentes, où on appelle arbre (resp. arbre non enraciné) ce que nous appelons arborescence (resp. arbre).

fondamentale qu'est l'arborescence binaire de recherche), l'analyse des performances en temps de ces algorithmes (en moyenne ou dans le cas le pire) nécessite souvent une connaissance fine de propriétés combinatoires de ces structures (voir les ouvrages de Knuth [102, section 2.3] et Sedgewick et Flajolet [139, chapitre 5]).

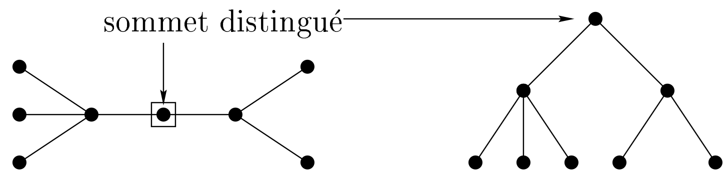
Finalement, on peut noter que de nombreux problèmes (mathématiques, biologiques, etc) pouvant être exprimés (ou modélisés) en termes d'arborescences, la question de la résolution algorithmique de ces problèmes implique la mise au point d'algorithmes efficaces manipulant des structures arborescentes. On peut citer par exemple le problème de la reconstruction phylogénétique source de nombreux travaux algorithmiques récents manipulant des arborescences phylogénétiques (voir l'ouvrage de Gusfield [86, chapitre 17] ou la thèse de Berry [19]).

Notre travail se situe dans ces deux domaines, à savoir l'étude des propriétés combinatoires de structures arborescentes et la mise au point d'algorithmes manipulant de telles structures. Bien que ces deux types problèmes aient comme point commun de manipuler des structures arborescentes, les problématiques, tout comme les méthodes employées, sont de natures différentes, ce qui nous a conduit à séparer ce mémoire en deux parties, une première partie consacrée aux problèmes combinatoires et une seconde partie consacrée aux problèmes algorithmiques.

**Terminologie et conventions.** Nous reprenons la terminologie classique composée de termes provenant essentiellement du vocabulaire de la généalogie et de la botanique.

Soit $A$ une arborescence. On appelle *racine* de $A$ son sommet distingué. Soient $x$ et $y$ deux sommets de $A$. Si le chemin allant de la racine à $y$ passe par $x$, on dit que $x$ est un *ancêtre* de $y$ et que $y$ est un *descendant* de $x$. Si de plus $x$ et $y$ sont reliés par une arête, on dit alors que $x$ est le père de $y$ et que $y$ est un fils de $x$. Si deux sommets ont le même père, on dit qu'ils sont *frères*. On appelle *degré* d'un sommet $x$ de $A$ le nombre de fils de $x$. On dit que $x$ est une *feuille* si son degré est nul, et que $x$ est un *nœud* sinon.

Il existe de nombreuses façons de représenter une structure arborescente, mais dans ce mémoire, nous convenons, à l'instar de ce qui est le plus courant dans la littérature, de dessiner une arborescence en plaçant le sommet distingué en haut, toutes les arêtes s'éloignant de ce sommet (dessin de droite de la figure ci-dessous).




**Énumération de structures arborescentes.** La première partie de ce mémoire est consacrée à l'énumération de diverses familles de structures arborescentes, en général selon le nombre de sommets. La plupart des articles initiaux de Cayley traitant de structures arborescentes étaient consacrés à leur énumération. Dans un article célèbre [29], ce dernier s'est notamment intéressé aux arborescences étiquetées (chacun des $n$ sommets est muni d'un étiquette entière appartenant à $[n]$, deux sommets ne pouvant avoir la même étiquette) libres (seules les relations d'adjacence entre sommets sont prises en compte) et a montré le résultat suivant[2] :

le nombre d'arborescences libres à $n$ sommets et de racine 1 est $n^{n-2}$, le corollaire suivant en découlant immédiatement :

le nombre d'arborescences libres à $n$ sommets est $n^{n-1}$.

Ces arborescences, qui sont parfois appelées arborescences de Cayley, plutôt que libres, sont un objet central apparaissant dans de nombreux problèmes combinatoires, comme par exemple les factorisations minimales de permutations circulaires en produit de transposition [53], l'étude des fonctions de parking ou, plus récemment, le dénombrement des régions de la famille d'arrangements d'hyperplans défini par Shi [147, 124].

Cette partie débute par un chapitre d'introduction, dans lequel nous fixons des notations et définitions que nous utilisons dans les chapitres ultérieurs. Puis, toujours dans ce chapitre introductif, nous présentons deux preuves du résultat de Cayley utilisant des outils que nous reprenons dans les chapitres suivants : le codage de Prüfer, qui illustre l'approche bijective et le codage des arborescences par des mots, et la formule d'inversion de Lagrange, qui illustre l'approche formelle et l'utilisation de séries génératrices.

Le chapitre suivant est consacré à l'étude combinatoire de la famille $\mathcal{A}_{n,k}$ des arborescences de Cayley à $n$ sommets telles que la racine possède exactement $k$ fils qui lui sont inférieurs. Nous montrons notamment de manière bijective le résultat suivant (théorème 16).

$$\text{Le nombre d'arborescences de } \mathcal{A}_{n+1,k} \text{ est } \binom{n}{k} n^{n-k}.$$

Dans le cas $k = 0$, ce résultat implique que le nombre d'arborescences de Cayley à $n+1$ sommets telles que la racine est inférieure à tous ses fils est $n^n$. Nous poursuivons par l'étude de divers paramètres classiques sur les structures arborescentes (l'étiquette de la racine, le degré de la racine ou encore le nombre d'arêtes décroissantes) pour la famille $\mathcal{A}_{n,k}$, qui permet de donner des explications combinatoires pour plusieurs identités faisant intervenir $n^n$, nouvelles pour la plupart à notre connaissance. Nous proposons ensuite une nouvelle description de la bijection due à Moszkowski [122] prouvant que le nombre de factorisations de la permutation cyclique $(1,2,\ldots,n)$ en produit de $n-1$ transpositions est $n^{n-2}$, description qui nous permet de raffiner ce résultat et de relier un sous-ensemble de ces factorisations à $\mathcal{A}_{n,0}$. Finalement, nous concluons ce chapitre en proposant un lemme général pour l'énumération d'arborescences telles que la racine est inférieure à ses fils, que nous appliquons aux arborescences ordonnées, planes et cycliques. Les résultats de ce chapitre ont été obtenus en collaboration avec Serge Dulucq et Olivier Guibert [36].

Dans le chapitre 3, nous poursuivons l'étude combinatoire de la famille d'arborescences $\mathcal{A}_{n,0}$. Nous nous intéressons au polynôme énumérateur des inversions de cette famille (une *inversion* dans une arborescence est un couple $(x,y)$ de sommets tels que $x$ est un ancêtre de $y$ et $x > y$). Dans le cas des arborescences de Cayley de racine 1, ce polynôme a été étudié par Kreweras [106], qui démontre une formule de récurrence qu'il satisfait (permettant par exemple de calculer le nombre d'arborescences ayant un nombre fixé d'inversions) et relie son évaluation en 0, 1, 2 et $-1$ à des objets combinatoires classiques (graphes et nombres d'Euler). Nous proposons une généralisation des résultats de Kreweras pour la famille des arborescences de $\mathcal{A}_{n,0}$ de racine $r$ fixée (en posant $r = 1$ dans nos résultats, nous obtenons ceux de Kreweras). Les résultats de ce chapitre ont été obtenus en collaboration avec Philippe Duchon et Serge Dulucq [34].

Le troisième chapitre est consacré à l'énumération d'arborescences alternantes. Une arborescence est alternante si toute suite $x_1,x_2,x_3\ldots$ de sommets successivement reliés

2. Bien que ce résultat soit en général attribué à Cayley, il avait été auparavant découvert par Sylvester en 1857.

par une arête dans cette arborescence vérifie $x_1 < x_2 > x_3 < \cdots$ ou $x_1 > x_2 < x_3 > \cdots$. Les arborescences de Cayley alternantes ont été introduites par Postnikov [124, 125] en relation avec la famille d'arrangements d'hyperplans de Linial. Dans ce chapitre, nous donnons une preuve bijective de la formule dénombrant le nombre d'arborescences de Cayley alternantes, répondant ainsi à une question de Postnikov [125]. Nous poursuivons en montrant que le nombre d'arborescences ordonnées alternantes à $n+1$ sommets est $2n^n$. Nous donnons deux preuves de ce résultat : une preuve formelle basée sur la composition de séries génératrices et la formule d'inversion de Lagrange, suivie d'une preuve basée sur une bijection avec les arborescences de $\mathcal{A}_{n,0}$. Les résultats de ce chapitre ont été obtenus en collaboration avec Serge Dulucq et Andrew Rechnitzer [37].

Les deux derniers chapitres de cette partie sont centrés autour de l'extension aux séries multivariées de la formule d'inversion de séries formelles de Lagrange, particulièrement adaptée à l'énumération de structures arborescentes coloriées (chaque sommet est muni, en plus de son étiquette, d'une couleur choisie parmi un ensemble de couleurs fixé). La formule de Lagrange est un outil particulièrement utile pour traiter de nombreux problèmes d'énumération de structures arborescentes caractérisées par des séries génératrices à une variable [16, section 3.1]. L'énumération de structures arborescentes coloriées faisant appel à des séries génératrices multivariées, Good [72] a proposé en 1960 une extension de la formule de Lagrange permettant de prendre en compte le cas de séries multivariées. Parmi les différentes formes équivalentes de la formule de Good (voir [16, section 3.2]), il en est une, introduite indépendamment par Goulden et Kulkarni [80] et Bender et Richmond [11], qui est particulièrement intéressante car elle ne fait pas intervenir de calcul de déterminant (contrairement à toutes les autres formes) et s'interprète naturellement en termes d'objets combinatoires classiques (arborescences coloriées et endofonctions coloriées). Goulden et Kulkarni utilisent cette propriété pour donner une preuve bijective de cette formule. Dans les chapitres 5 et 6 nous proposons une approche bijective de cette formule, approche qui nous permet de clarifier et d'expliquer combinatoirement certains résultats obtenus par application des formes classiques de la formule de Good, et de déduire de ces explications des algorithmes de génération aléatoire uniforme pour les structures considérées. Ces deux aspects représentent, à notre avis, un point de vue intéressant sur la formule de Good et l'énumération de structures arborescentes coloriées.

Dans le chapitre 5 nous présentons différentes formes de la formule de Good, dont la forme due à Goulden-Kulkarni et Bender-Richmond, que nous appelons forme arborescente car basée sur la notion de dérivée d'un vecteur de séries formelles multivariées relativement à une arborescence. Nous proposons ensuite une nouvelle preuve bijective de cette formule, plus simple que la seule preuve bijective connue (due à Goulden et Kulkarni [80]), basée sur une variante de la preuve de la formule de Lagrange pour les séries à une variable due à G. Labelle [108]. Comme la preuve de Goulden et Kulkarni, notre preuve repose sur une bijection entre arborescences coloriées et endofonctions coloriées. Nous illustrons ensuite un premier intérêt d'une telle preuve en proposant des explications combinatoires pour plusieurs formules dénombrant des arborescences coloriées dont la partition des arêtes est fixée. Si la plupart de ces formules étaient déjà présentes dans la littérature, les preuves que nous en donnons sont, à notre connaissance, les premières à être purement combinatoires. Nous poursuivons en expliquant deux formules dénombrant des cactus planaires, une famille de graphes coloriés reliés aux factorisations minimales de permutations cycliques en produit de permutations [22, 21] et à la classification topologique des polynômes complexes [87, 58]. Nous déduisons de ces preuves des algorithmes de génération aléatoire uniformes de cactus planaires selon le nombre de sommets de chaque couleur ou selon la partition des degrés. Les résultats de ce chapitre ont été obtenus en collaboration avec Michel Bousquet, Gilbert Labelle, Pierre Leroux et Gilles Schaeffer [23, 24].

Le dernier chapitre de cette partie est consacré à l'énumération et à la génération aléatoire de constellations planaires. Une $m$-constellation planaire est une carte planaire obtenue en collant des polygones à $m$ sommets de sorte que pour chaque face de la carte ainsi obtenue, le nombre d'arêtes la bordant soit un multiple de $m$. Ces objets sont en relation avec les factorisations minimales de permutations en produits de permutations et ont été notamment étudiés par N. Hanusse [87], et M. Bousquet-Mélou et G. Schaeffer [135, 25]. Ces derniers montrent notamment, à l'aide de la conjugaison d'arborescence, technique introduite dans [135], que les $m$-constellations sont en bijection avec une famille particulière d'arborescences : les arborescences eulériennes. Nous montrons alors que l'on peut mettre ces arborescences en bijection avec une famille d'arborescences pour lesquelles on peut appliquer la formule de Good. Nous obtenons ainsi une formule dénombrant le nombre de $m$-constellations ayant $n$ sommets et $f$ faces et en déduisons un algorithme de génération aléatoire pour ces structures. Dans le cas $m = 2$, nous obtenons de plus une nouvelle formule dénombrant les cartes biparties selon le nombre de sommets et de faces.

**Recherche de motifs.** Dans la seconde partie de ce mémoire, nous nous intéressons au problème algorithmique constitué par la recherche de motifs dans une arborescence. Le problème de la recherche de motifs ("pattern matching") est un problème classique en algorithmique (voir par exemple le récent livre édité par Apostolico et Galil [7]) que l'on peut définir de la façon suivante.

> Étant données deux structures de même type, un texte $T$ et un motif $M$, on cherche à repérer toutes les occurrences (cette notion dépendant des structures considérées) de $M$ présentes dans $T$.

Dans le cas des mots, lorsque par exemple, $T = bacacbacbabc$ et $M = acba$, on a deux occurrences de $M$ dans $T$ débutant respectivement en position 4 et 7. Dans le cas des arborescences, on considère habituellement des arborescences ordonnées (pour chaque sommet, on impose un ordre linéaire sur ses fils) telles que tout sommet est muni d'une étiquette appartenant à un alphabet $\Sigma$, deux sommets pouvant ici avoir la même étiquette. Intuitivement, on dit qu'on a une occurrence de $M$ (le *motif*) en un sommet $x$ de $T$ (le *texte*) si en plaçant la racine de $M$ en $x$, on arrive à une situation (que l'on devra préciser) dans laquelle tout sommet (resp. arête) de $M$ recouvre un sommet (resp. arête) de $T$ ayant la même étiquette.

Dans nos travaux, nous nous sommes plus particulièrement intéressés au problème de la recherche de motif dans un texte *statique*. Revenons aux mots et considérons par exemple le cas d'une encyclopédie sur CD-ROM. Une des fonctionnalités indispensables à un tel logiciel est une procédure permettant de repérer toutes les occurrences d'un mot $M$ dans le texte de cette encyclopédie. En remarquant que dans toute instance $(T,M)$ de ce problème le paramètre $T$ est le même (le texte de l'encyclopédie), il est naturel de vouloir traiter ces problèmes de recherche de motifs de la manière suivante : prétraiter le texte $T$ une fois pour toute lors de la création du CD-ROM et utiliser les informations de ce prétraitement pour rechercher rapidement les occurrences d'un motif[3]. La solution généralement employée dans ce cas consiste à construire une structure de données annexe indexant les suffixes

3. Parmi les nombreuses autres applications où l'on retrouve la situation d'un texte statique, on peut citer la recherche de motifs dans des banques de données regroupant des milliers de séquence issues de l'analyse des génomes (une fois qu'une séquence est ajoutée dans la banque, elle n'est plus modifiée et elle peut donc être prétraitée lors de cette insertion).

du texte $T$ : en remarquant que toute occurrence du motif $M$ dans $T$ est préfixe d'un des suffixes de $T$, cet index permet de déterminer rapidement les occurrences de $M$ dans $T$ (en général en temps proportionnel à la seule taille du motif, c'est à dire indépendamment de la taille du texte). Dans le cas des mots, les principales structures de données utilisées pour indexer les suffixes d'un texte sont l'automate des suffixes, le tableau des suffixes et l'arborescence des suffixes [52, chapitres 5 et 6].

Notre but est de traiter le problème de la recherche de motifs dans une arborescence *statique* en nous inspirant essentiellement de l'utilisation de l'arborescence des suffixes dans le cas des mots. Dans un premier chapitre, nous définissons formellement la notion d'occurrence d'un motif $M$ dans une arborescence $T$ et une restriction de cette notion (que nous appelons occurrence compacte). On peut cependant noter que la notion générale d'occurrence est équivalente à la notion d'occurrence compacte si les arborescences considérées sont des termes (encore appelés arborescences d'expression). Nous poursuivons par une rapide présentation du problème de la recherche de motifs dans les mots, en nous intéressant surtout à une structure de données souvent utilisée dans le cas d'un texte statique : *l'arborescence des suffixes d'un mot*, qui est à la base des résultats que nous proposons par la suite. Nous effectuons ensuite une revue des principaux articles traitant de recherche de motifs dans une arborescence. Cette étude nous amène notamment à faire les remarques suivantes :

- le sujet a été relativement peu étudié, notamment d'un point de vue pratique, la plupart des algorithmes proposés étant assez complexes,
- tous les algorithmes proposés recherchent les occurrences compactes du motif dans le texte et ne semblent pas pouvoir être aisément étendus pour rechercher les occurrences générales,
- aucun des algorithmes proposés ne se place dans le cas d'un texte statique.

Nous concluons ce chapitre par une adaptation simple d'un algorithme proposé Hoffmann et O'Donnell [91] pour la recherche des occurrences compactes qui nous permet de rechercher les occurrences générales du motif dans le texte.

Le second chapitre propose une approche du cas statique pour les termes reposant sur le codage d'un terme par un mot et l'utilisation de l'arborescence des suffixes de ce mot. La complexité dans le pire des cas de l'algorithme que nous proposons n'améliore pas les résultats théoriques connus, mais des résultats expérimentaux nous laissent penser qu'il est très intéressant d'un point de vue pratique. Nous proposons aussi une variante de cet algorithme pour la recherche des occurrences compactes du motif dans le texte.

Le dernier chapitre de cette partie propose une approche plus théorique du problème que nous considérons, dans laquelle nous étendons la notion d'arborescence des suffixes des mots aux arborescences. En effet, étant donnée une famille de structures de données et une notion de suffixes de ces structures, il est naturel de vouloir définir une notion d'arborescence des suffixes pour ces structures. Giancarlo et Grossi ont par exemple suivi cette démarche dans le cas des matrices carrées [66, 67]. Récemment, Kosaraju [105] a introduit une notion d'arborescence des suffixes d'une arborescence, dans laquelle un suffixe est un chemin allant d'un sommet vers la racine. Dans ce chapitre, nous introduisons une nouvelle notion de suffixe, inspirée de la notion de suffixe d'un mot : un suffixe d'une arborescence $T$ est une sous-arborescence maximale de $T$. On a donc exactement un suffixe par sommet $x$ de $T$ (la sous-arborescence de $T$ de racine $x$), ce qui permet de définir une notion d'arborescence des suffixes d'une arborescence $T$ occupant un espace linéaire par rapport au nombre de sommets de $T$. Nous proposons ensuite un algorithme permettant de construire une telle arborescence des suffixes, dérivé de l'algorithme de McCreight pour l'arborescence des suffixes d'un mot [118]. Nous concluons en évoquant les applications de de cette structure en termes de recherche de motifs dans une arborescence.