Documents de travail du Programme Modélisation des plantes
N° 3-2002

Rapport de recherche Post-doctorale

Développement d'outils conceptuels et algorithmiques pour l'analyse de structures biologiques arborescentes

P. Ferraro

Centre de coopération internationale
en recherche agronomique pour le développement

Documents de travail du Programme Modélisation des plantes
N° 3-2002

Rapport de recherche Post-doctorale

Développement d'outils conceptuels et algorithmiques pour l'analyse de structures biologiques arborescentes

P. Ferraro

Août , 2002

Résumé

La visualisation et la modélisation de l'architecture des plantes est l'un des principaux domaines de recherche du groupe « Biological Modeling and Visualization Group » du laboratoire d'Informatique de l'Université de Calgary. Les activités de recherche abordées par cette équipe, très proches de ceux suivis par l'UMR « Botanique et Bioinformatique de l'Architecture des Plantes » de Montpellier nous a conduit à mettre en place de nombreux contacts entre ces deux institutions, notamment grâce aux collaborations entre C. Godin et P. Prusinkiewicz. Ces collaborations ont donné lieu à un post-doctorat réalisé à l'Université de Calgary, financé par une bourse de l'Institut National de Recherche en Informatique et en Automatique (INRIA), présenté dans ce rapport.

Table des matières

1 Introduction 4

2 Résumé des travaux antérieurs 4

2.1 Le laboratoire d'accueil 4

2.2 Travail de Thèse 5

3 Résumé du travail post-doctoral **5**

3.1 Le laboratoire d'accueil 6

3.2 Présentation du travail 6

4 Couplage d'un moteur de simulation et d'un logiciel d'analyse de plantes **8**

4.1 Introduction 8

4.2 Description des bases de données architecturales dans AMAPmod 8

4.3 L-systèmes 12

4.4 Algorithme de conversion 13

4.5 Applications 14

4.6 Conclusion 15

5 Autosimilarité des structures ramifiées **16**

5.1 Introduction 16

5.2 Motivation Biologique 16

5.3 Évaluation de la ressemblance entre deux systèmes ramifiés 17

5.3.1 Définitions 18

5.3.2 Comparaison d'arborescences 18

5.3.3 Standardisation et normalisation de la distance obtenue 21

5.4 Evaluation de l'autosimilarité dans une arborescence 22

5.4.1 Définitions 22

5.4.2 Étude de la valeur de $\overline{PR}$ 24

5.4.3 Analyse d'inflorescences de lilas 24

5.5 Conclusion 26

6 Perspectives **27**

Références **28**

Remerciements **32**

1 Introduction

La visualisation et la modélisation de l'architecture des plantes est l'un des principaux domaines de recherche du groupe "Biological Modeling and Visualization Group" du laboratoire d'Informatique de l'Université de Calgary. Les activités de recherche abordées par cette équipe, très proches de celles suivies par l'UMR "Botanique et Bioinformatique de l'Architecture des Plantes" de Montpellier nous ont conduit à mettre en place de nombreux contacts entre ces deux institutions, notamment grâce aux collaborations entre C. Godin et P. Prusinkiewicz. Ces collaborations ont donné lieu à un post-doctorat réalisé à l'Université de Calgary, financé par une bourse de l'Institut National de Recherche en Informatique et en Automatique (INRIA), présenté dans ce rapport.

Nous avions préalablement identifié deux axes de recherche majeurs :
- L'extension des méthodes de calcul de distances entre architectures de plantes étudiées durant la thèse.
- l'instantiation d'approximations de plantes pour la représentation et la visualisation de grandes plantes ou d'écosystèmes végétaux.

Dans ce rapport, après avoir rappelé les travaux mis en place durant mon travail de recherche à l'UMR AMAP, je présenterai les résultats mathématiques et informatiques obtenus lors de ce travail post-doctoral et leurs applications à la modélisation de l'architecture des plantes.

2 Résumé des travaux antérieurs

Nous résumons dans cette première partie le travail de thèse et le contexte dans lequel ce travail a été réalisé.

2.1 Le laboratoire d'accueil

Depuis plus d'une dizaine d'années, l'UMR CIRAD-CNRS-INRA-Université Montpellier II "Botanique et Bioinformatique de l'Architecture des Plantes" (AMAP) regroupe des biologistes et des informaticiens qui développent des méthodes mathématiques et informatiques pour la modélisation des plantes. Ce travail a permis de concevoir une méthodologie générique de mesure et d'analyse de l'architecture des plantes (Godin et al., 1997b). De nombreux outils conceptuels et informatiques ont été développés dans ce cadre et sont rassemblés dans le logiciel AMAPmod (Godin et Guédon, 1999).

AMAPmod est aujourd'hui un logiciel libre, adopté par un nombre croissant d'instituts de recherche en France et à l'étranger, qui permet d'établir progressivement un standard de mesure et d'analyse de l'architecture des plantes. L'utilisateur peut créer ses propres bases de données architecturales et en explorer le contenu grâce à des outils statistiques, de modélisation ou de visualisation graphique en 2 ou 3 dimensions. Ce logiciel préfigure les outils agronomiques de demain et suscite un travail de recherche important en mathématiques appliquées et en informatique e.g. (Godin et Caraglio, 1998; Guédon et al., 1999; Ferraro et Godin, 2000).

2.2 Travail de Thèse

Dans ce cadre, mes travaux de thèse ont permis d'introduire de nouvelles méthodes algorithmiques pour la comparaison de l'architecture des plantes. La définition d'une mesure quantitative de la ressemblance entre architectures de plantes a des applications potentielles en sylviculture, en arboriculture, en horticulture et en agronomie. Cette notion permet en effet de prendre en compte de façon très fine des différences de structure entre deux plantes lorsque celles-ci ont des caractéristiques globales équivalentes (même nombre de fruits, même nombre d'axes, etc.). En modélisation, la comparaison structurelle d'individus permet d'aborder sous un angle nouveau le problème de l'évaluation des modèles de simulation de la croissance des plantes (Ferraro et Godin, 2000).

L'organisation structurelle des entités d'une plante, *ie.* son architecture, est souvent représentée par un graphe arborescent. Pour rendre compte des différents niveaux d'organisation dans la description de l'architecture des plantes, l'UMR AMAP a introduit un formalisme basé sur la notion de graphe quotienté : les structures arborescentes multi-échelles (Godin et Caraglio, 1998). Ainsi, comparer la structure de deux plantes se ramène formellement à comparer deux graphes arborescents multi-échelles. Mon travail de thèse a eu pour objectif d'explorer les algorithmes de comparaison entre arborescences, d'en étudier les applications dans les domaines de la biologie végétale et d'envisager différentes extensions, notamment à la définition d'une distance entre arborescences multi-échelles.

La famille d'algorithmes étudiée s'appuie sur une généralisation des méthodes de calcul de distance entre séquences, utilisées par exemple pour la comparaison du génome (Durbin et al., 1998). Leur principe général consiste à reconstruire un graphe cible en appliquant à un graphe initial différentes opérations structurelles élémentaires, appelées opérations d'édition. On peut montrer que la définition d'une telle distance se ramène à la recherche d'une solution optimale dans un espace dont la taille croît exponentiellement avec la taille des objets à comparer. Aussi, pour définir des algorithmes efficaces à la fois en temps et en espace, la solution optimale est construite de façon incrémentale en utilisant des techniques de programmation dynamique.

Après avoir proposé un point de vue unifié de différentes distances entre arborescences (Ferraro et Godin, 2001), j'ai pu donner deux extensions algorithmiques permettant d'envisager le calcul d'une mesure de ressemblance entre arborescences quotientées (Ferraro et Godin, 1999; Ferraro, 2000a; Ferraro et Godin, 2002). J'ai alors pu développer différentes applications de ces méthodes à la comparaison de l'architecture des plantes, en collaboration avec des biologistes (botanistes et agronomes) (Ferraro et Godin, 2000).

3 Résumé du travail post-doctoral

Nous présentons dans cette section un résumé des différents thèmes abordés durant ce post-doctorat. Ces thèmes seront ensuite détaillés dans les sections suivantes.

3.1 Le laboratoire d'accueil

J'ai prolongé le travail de thèse présenté précédemment dans le cadre d'un projet post-doctoral au sein du groupe "visualisation et modélisation biologique" de l'université de Calgary (Prusinkiewicz, 1996) financé par une bourse de l'Institut National de Recherche en Informatique et en Automatique (INRIA). Les axes de recherches de ce laboratoire canadien, similaires à ceux du programme de modélisation des plantes, m'ont permis d'étendre mon champ de compétence en informatique et mathématiques appliquées. Ce laboratoire a développé une approche de la modélisation de la croissance des plantes fondée sur des systèmes de réécriture particuliers appelés L-systèmes (Prusinkiewicz et Lindenmayer, 1990). Ce formalisme a été employé pour modéliser la croissance de nombreuses espèces végétales (Prusinkiewicz, 1994; Prusinkiewicz, 1998), et pour formaliser différentes hypothèses sur la morphogénèse des végétaux et des animaux (Prusinkiewicz, 1994).

3.2 Présentation du travail

Durant cette année post-doctorale, le travail réalisé a été décomposé en trois parties :

- Couplage d'un moteur de simulation et d'un logiciel d'analyse de plantes (Ferraro et al., 2001) :
  Afin de permettre l'évaluation des plantes obtenues par simulation ou de construire des modèles à partir de l'analyse de données observées, j'ai été amené à coupler deux plateformes logicielles très différentes car conçues indépendamment : cpfg (développée à l'université de Calgary) destinée à la simulation de l'architecture des plantes et AMAPmod (développée à l'UMR AMAP de Montpellier) dédiée à l'analyse de l'architecture des plantes. L'implémentation de ce couplage a ensuite été intégrée dans le "virtual laboratory software" (vlab) (Federl et Prusinkiewicz, 1999).
- Algorithmes d'analyse d'arborescences :
  En m'appuyant sur cette implémentation, j'ai pu éprouver la méthode de comparaison entre plantes proposée dans le cadre de ma thèse sur des données simulées par cpfg. Nous avons pu ainsi pour la première fois évaluer les plantes obtenues par simulation, en les comparant structurellement avec des plantes mesurées sur le terrain. Ce travail constitue l'une des toutes premières approches structurelles de l'évaluation de modèles morphogénétiques de plantes. En nous appuyant sur ce premier travail, en collaboration avec l'UMR AMAP et l'équipe du Pr. Prusinkiewicz, nous avons proposé une méthode permettant d'identifier les propriétés autosimilaires des plantes. Cette méthode a été, par exemple, utilisée pour quantifier le degré d'autosimilarité d'inflorescences de lilas (Ferraro et al., 2002b). Deux perspectives majeures ressortent de ce travail, il permet :
  1. d'une part, de définir un nouvel outil d'investigation biologique, permettant d'extraire un déterminisme caché de la structure végétale (son degré de répétition).
  2. d'autre part, d'envisager le développement de méthodes de représentation réaliste d'écosystèmes composés de plantes simulées, en mettant à profit la répétition de certaines structures dans l'architecture des plantes.

  Cette dernière perspective, actuellement à l'étude, a des applications potentielles importantes dans la modélisation et la simulation de grands écosystèmes (Prusinkiewicz et Remphrey, 2000), pour la représentation de couverts forestiers, des paysages ou la synthèse

d'images.

- Infographie :
  Depuis quatre ans, C. Godin et P. Prusinkiewicz travaillent à l'expression des fondements informatiques et mathématiques de la croissance multi-échelles des plantes. Dans cette optique, ils ont engagé un projet de modélisation de cette croissance multi-échelles basée sur les L-systèmes. L'une des difficultés importante de ce travail réside dans la conception de méthodes d'exploration visuelle d'objets multi-échelles. En m'appuyant sur ma connaissance des deux formalismes (graphes multi-échelles et L-systèmes), j'ai développé un outil permettant d'afficher simultanément différents niveaux de détails d'une partie d'un objet multi-échelles tout en conservant le contexte dans lequel elle se trouve. Une publication ainsi qu'une thèse sur ce sujet sont actuellement en cours, dans le cadre de ce projet. Ce travail s'appuie sur une bibliothèque graphique développée par Cathy Jirasek et Sheelag Carpendale (Carpendale et al., 1998) utilisant la distortion des objets en trois dimensions pour simuler un effet de zoom sur ces objets (cf. Figure 1).

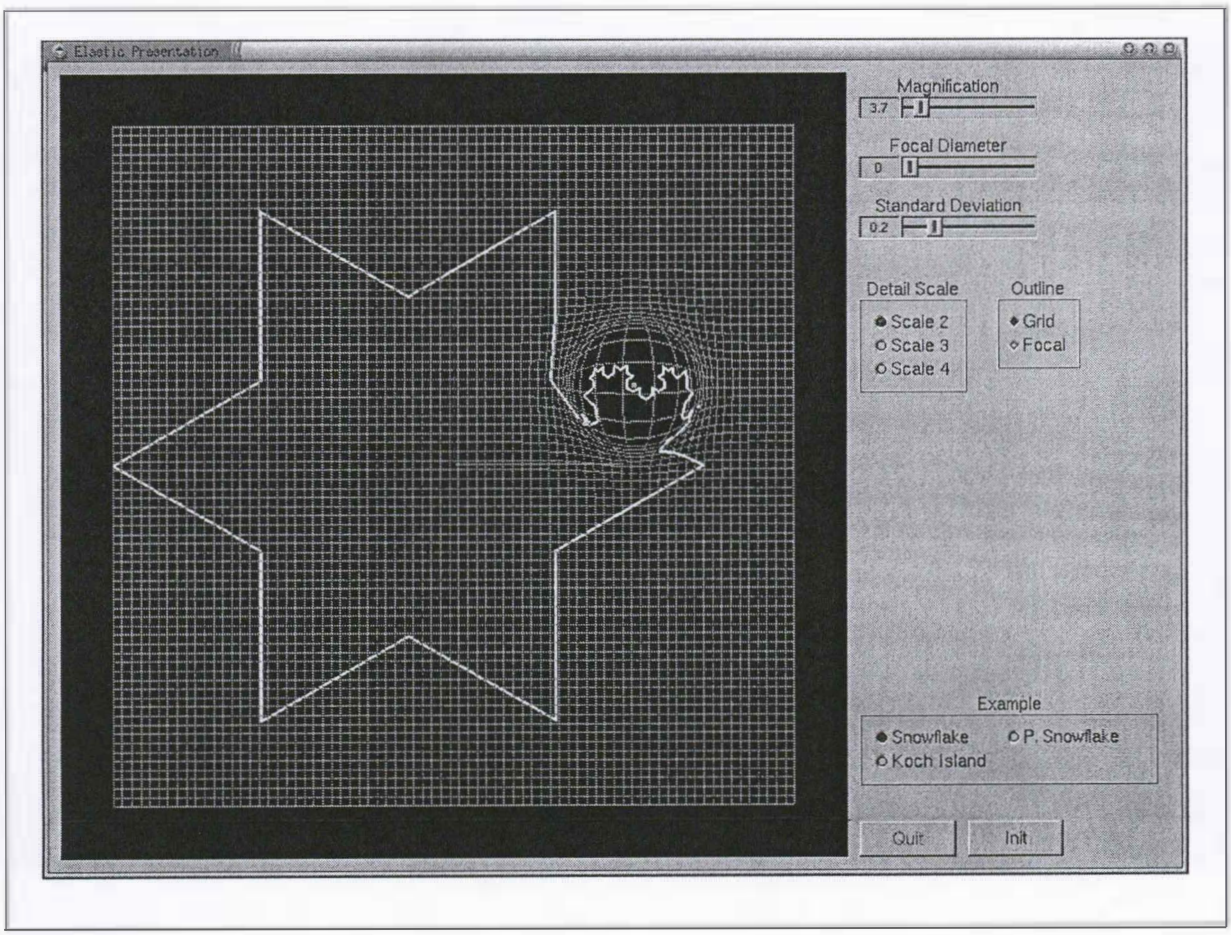

FIG. 1 – Simulation du zoom sur la courbe de Von Koch. Lorsque la loupe se déplace, le développement de la courbe est recalculée à la volée par communication avec cpfg.

La suite de ce rapport décrit de manière détaillée les deux premiers points de ce travail de recherche.

4 Couplage d'un moteur de simulation et d'un logiciel d'analyse de plantes

4.1 Introduction

Les Lindemayer-systèmes (L-systèmes) ont été originellement introduits afin de modéliser le développement d'organismes multicellulaires relativement simples, tels que des algues (Lindenmayer, 1968). Ils sont maintenant utilisés pour la modélisation de plantes de grandes tailles (Prusinkiewicz et al., 1997). Le logiciel cpfg (Mech, 1998), développé à l'Université de Calgary, est un programme de simulation basé sur ce formalisme L-systèmes. Sa particularité est la flexibilité du langage de modélisation qui permet à l'utilisateur de spécifier l'architecture de divers organismes modulaires, des algues aux plantes herbacées, des arbres aux écosystèmes de plantes. Les modèles produits par cpfg peuvent être de nature descriptive ou mécaniste. Dans le second cas, l'utilisateur peut alors étudier l'impact des processus physiologiques ou environnementaux sur le développment des plantes. Les résultats des simulations sont ensuite visualisés sous la forme de diagrammes ou d'images réalistes de plantes.

Le formalisme de graphes arborescents multi-échelles (MTGs) a été introduit par Godin et Caraglio (98) afin de représenter la topologie des plantes à différentes échelles et à différents niveaux d'abstraction, (métamères, unités de croissance ou axes d'une plante, par exemple). Une fois qu'une plante a été décrite sous la forme d'un MTG, les données de la plante peuvent être explorées en utilisant le logiciel AMAPmod (Godin et al., 1997b) développé à l'UMR AMAP. Les outils logiciels contenus dans AMAPmod pemettent d'identifier des régularités au sein de la structure de la plante, de quantifier des différences entre des structures ramifiées, ou d'accomplir d'autres tâches analytiques (*cf.* (Ferraro et Godin, 2000; Guédon et Costes, 1998)...). Les MTGs peuvent représenter des données expérimentales, obtenues par mesure de plantes, ou des structures de plantes simulées par des modèles de croissance.

Dans ce travail, notre objectif était de connecter ces deux outils logiciels afin de permettre l'analyse de plantes obtenues par simulation. Par la suite nous présenterons un algorithme permettant de convertir les structures de plantes générées par des L-systèmes au format MTG et son implémentation dans le logiciel de modélisation cpfg. Le "virtual laboratory sofwtare" (lab) est utilisé pour intégrer cpfg et AMAPmod au sein d'un système cohérent pour la modélisation et l'analyse des plantes (*cf.* Figure 2). Nous présentons de plus comment cette application peut être utilisée pour l'évaluation des plantes virtuelles dans le cas de modèles d'inflorescences d'espèces de Syringa (inflorescences de lilas) en utilisant une comparaison structurelle avec des inflorescences mesurées.

4.2 Description des bases de données architecturales dans AMAPmod

Dans AMAPmod, les plantes sont représentées par des graphes arborescents multi-échelles (MTGs). Un MTG peut être vu comme un ensemble d'arborescences emboîtées les unes dans les autres, représentant la topologie de la plante à différentes échelles (entre-nœuds, unités de croissance, *etc.*). Pour construire un MTG à partir d'une plante, celle-ci est tout d'abord décomposée en ensemble de composants organisés à différentes échelles. Un label spécifiant leur type est

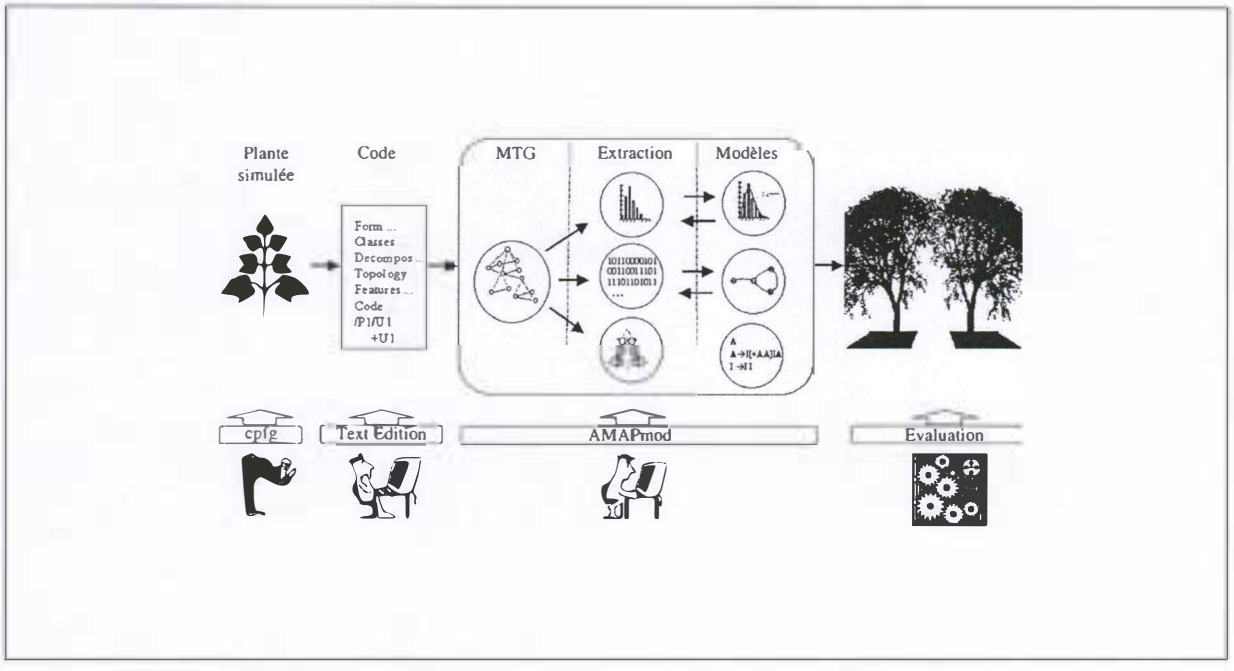

FIG. 2 – Synopsis de la connection entre cpfg et AMAPmod.

affecté à chaque composant (`U` = Unité de croissance, (`F` = site florifère, (`S` = pousse courte, (`I` = Entre-nœud). Ces labels sont alors utilisés pour coder l'architecture de la plante sous une forme textuelle. Le fichier de codage résultant peut ensuite être analysé par AMAPmod et correspond à un MTG. En pratique, dans un MTG, l'organisation des composants de la plante à une échelle de détails donnée est représentée par une arborescence, où un nœud du graphe représente un composant et les arcs représentent les connections entre les composants. A toutes échelles, les composants sont reliés par deux types de relation, correspondant aux deux mécanismes de bases de croissance de la plante : la croissance apicale et le processus de ramification. La croissance apicale est responsable de la création des axes, par production de nouveaux composants (correspondant à de nouvelles portions de la tige et des feuilles) au sommet du composant précédent. La connexion entre deux composants résultant de la croissance apicale est une relation dite de "succession" et on la note par un '<'. D'un autre côté, le processus de ramification est responsable de la création axillaire (ces bourgeons peuvent alors créer des axes axillaires par leur propre croissance apicale). La connexion entre deux composants résultant du processus de ramification est du type "porteur-porté" et on la note '+'. Un MTG regroupe au sein d'un modèle unique les différentes représentations qui correspondent aux différentes échelles décrivant la plante.

A toute échelle, différents types d'attributs peuvent être affectés aux composants d'une plante représentée par un MTG. Ces attributs peuvent être géométriques (*e.g.* diamètre des tiges, surface d'une feuille ou coordonnées 3D des composants de la plante) ou morphologiques (*e.g.* nombre de fleurs, type de production axillaire, - bourgeon latent, pousse courte ou longue -, nature des feuilles). Différentes stratégies ont été proposés pour coder la structure des plantes, *e.g.* (Hanan et Room, 1997), pour des plantes représentées à une échelle de décomposition donnée et, (Godin et al., 1997a; Godin et al., 1997b), pour des représentations multi-échelles. La structure topologique d'une plante représentée sous la forme d'une arborescence multi-échelles dans AMAPmod est spécifiée en utilisant un langage de codage. Ce langage consiste en une stratégie de nommage des nœuds et des arcs du graphe multi-échelles. Le nom d'un nœud est défini de telle manière

qu'il indique clairement la position topologique du nœud dans tout le graphe multi-échelles. Les nœuds et leurs attributs sont décrits en utilisant ce langage formel dans un fichier de codage. Nous illustrons par la suite le principe général de ce langage de codage sur la plante représentée Figure 3.

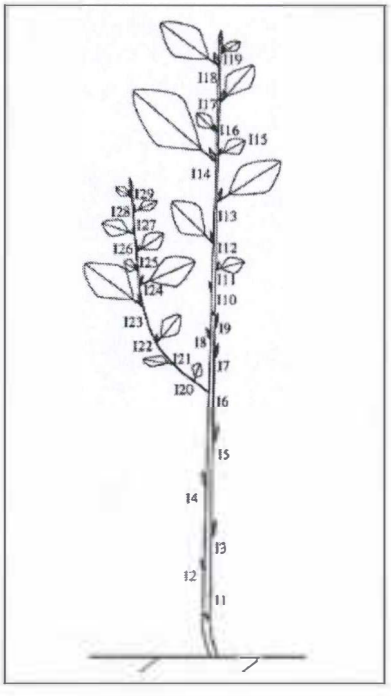

FIG. 3 – Codage de la structure topologique d'un peuplier de deux ans.

A chaque nœud est affecté une *étiquette* constituée d'une lettre, appelée sa *classe*, et d'un entier, appelé son *index*. La classe d'un nœud référence souvent la nature de l'entité botanique correspondante, *e.g.* `I` pour un entre-nœud, `U` pour une unité de croissance, `B` pour un système ramifié, *etc.* . L'index d'un nœud est un entier qui permet à l'utilisateur d'identifier localement un nœud parmi ses voisins immédiats.

A une échelle donnée, la plante est décrite depuis la base du tronc et en remontant vers le sommet. La chaîne de codage commence par le symbole '/'. La série des entre-nœuds du tronc de la plante représentée Figure 3 correspond à la chaîne suivante :

```
/I1<I2<I3<I4<I5<I6<I7<I8<I9<I10<I11<I12<I13<I14<I15<I16<I17<I18<I19
```

Le codage d'une structure arborescente par une séquence linéaire de symboles conduit à l'introduction d'une notation spéciale, fréquemment utilisée en informatique (*e.g.* (Prusinkiewicz et Lindenmayer, 1990)). L'arborescence est décrite de la racine vers les feuilles. Un crochet est ouvert chaque fois qu'un point de bifurcation est rencontré durant la visite (*ie.* pour les nœuds ayant plus d'un fils). Un crochet est fermé chaque fois qu'un nœud terminal est visité (*ie.* pour les nœuds n'ayant pas de fils). Dans l'exemple ci-dessus, l'entité `I6` est un point de bifurcation car le processus de description peut être poursuivi soit en visitant l'entité `I7` soit en visitant `I20`. Dans ce cas, la bifurcation `I6` est tout d'abord stockée dans une pile (qui est initialement vide).

Puis, un crochet ouvert est inséré dans la chaîne et enfin, le processus de visite renseigne les deux directions possibles, par exemple I20, conduit au code :

```
/I1<I2<I3<I4<I5<I6[+I20
```

La branche entière I20 à I28 est alors codée comme les entité 1 à 6. L'entité I29 n'a pas de fils, c'est donc une entité terminale. Ceci implique qu'un crochet fermant est inséré dans la chaîne.

```
/I1<I2<I3<I4<I5<I6[+I20<I21<I22<I23<I124<I25<I26<I27<I28<I29]
```

Le dernier point de bifurcation peut alors être dépilé et le processus de visite peut être poursuivi dans l'une des bifurcations possible de I6 non encore visitée,*ie.* I7. On obtient alors à la chaîne de codage finale :

```
/I1<I2<I3<I4<I5<I6[+I20<I21<I22<I23<I124<I25<I26<I27<I28<I29]<I7<I8<I9<I10<I11<
I12<I13<I14<I15<I16<I17<I18<I19
```

La stratégie de croissance multi-échelles n'étant pas prise en compte par le formalisme L-système, nous ne présenterons pas l'extension de cette stratégie à la description de structures multi-échelles.

Dans des applications pratiques, un certain nombre d'attributs sont souvent mesurés sur différentes entités. Les valeurs mesurées peuvent alors être attachées à l'entité correspondante en utilisant une notation parenthésée : '(...)'. Supposons, par exemple, qu'on veuille noter les longueurs et diamètres observés des unités de croissance. Une paire de valeurs numériques définissant respectivement la longueur et le diamètre de l'unité de croissance est alors affecté au dernier entre-nœud de celle-ci. La chaîne précédente devient alors :

```
/I1(10,5.9)<I2<I3<I4<I5<I6[+I20<I21<I22<I23<I24(4,2.1)<I25<I26<I27<I28<I29]<I7<
I8<I9(8,4.3)<I10<I11<I12<I13<I14<I15(7.5,3.9)<I16<I17<I18<I19
```

Dans cette chaîne, le premier entre-nœud du tronc, I1 a une longueur de 10 mm et un diamètre de 5.9 mm (les unités sont supposés connues et fixées).

Cependant, en pratique, le codage des plantes sous la forme d'une séquence de symboles devient rapidement illisible. Afin de donner à l'utilisateur un meilleur feedback de la topologie de la plante dans le code, le codage précédent peut être modifié pour une meilleure lisibilité. Chaque crochet est alors remplacé par une nouvelle ligne et un niveau d'indentation correspondant au degré d'emboîtement des crochets. De la même manière, une nouvelle ligne est créee après chaque ensemble d'attributs et les valeurs correspondantes sont écrites dans des colonnes spécifiques. Le tableau suivant donne le code final de l'exemple de la Figure 3 :

```
                                        Length  Diameter
/I1                                       10       5.9
<I2<I3<I4<I5<I6
            +I20<I21<I22<I23<I24          4        2.1

            <I25<I26<I27<I28<I29
<I7<I8<I9                                 8        4.3
<I10<I11<I12<I13<I14<I15                 7.5       3.9
<I16<I17<I18<I19
```

4.3 L-systèmes

En 1968, Lindemayer introduit un formalisme, appelé L-système (Lindenmayer, 1968), pour la modélisation d'organismes multicellulaires simples. Un L-système peut être défini en s'appuyant sur la théorie des langages comme un alphabet de symboles, une chaîne initiale et un ensemble de règles de réécriture. Dans le contexte biologique, les symboles de l'alphabet représentent des cellules ou des modules de plantes (Prusinkiewicz et al., 1997), comme des entre-nœuds, des apex, des feuilles, ou des fleurs, l'axiome spécifie l'état initial de la plante, et les productions représente les règles de développement du module durant des intervalles de temps fixés.

Un L-système est composé d'un axiome ω et de deux productions p_1 et p_2, comme dans l'exemple ci-dessous :

ω : A
p_1 : A->AB
p_2 : B->BB

La quatrième chaîne du développement de la séquence S générée par ce L-système est alors : ABBBBBBB.

Afin de permettre la génération de structures ramifiées, Lindenmayer et Prusinkiewicz (90) introduisent la notion de L-systèmes parenthésés, dans lesquels les branches sont délimitées par des crochets. La définition d'un tel L-système est généralement étendu par les règles suivantes :

1. les crochets sont remplacés par eux-mêmes ;
2. à tout crochet ouvrant '[' correspond un unique crochet fermant ']'.

La représentation graphique des structures ramifiées a été développée par Lindemayer puis clarifiée par Prusinkiewicz (*cf.* (Mech, 1998) pour une revue). Cette représentation basée sur *l'interprétation de la tortue* est définie comme suit. La tortue est tout d'abord positionnée sur l'origine du repère. La chaîne produite par le L-système est alors interprétée comme un ensemble de commande permettant le déplacement de la tortue dans l'espace. Par exemple, le module F génère un segment dans la direction du vecteur directeur de la tortue et positionne la tortue à la fin de la ligne. Les modules '+', −', '&', ∧', '/' et '\' permettent d'effectuer différentes rotations de la position de la tortue. Ces modules sont, par définition, toujours remplacés par eux-même. Afin de visualiser les structures ramifiées générées par des L-systèmes, les crochets sont interprétés de la manière suivante : '[' place l'état actuel de la tortue dans une pile, ']' remplace l'état courant de la tortue par l'état en haut de la pile. Les L-systèmes paramétriques ont par la suite été introduits, afin de prendre en compte les différents angles d'insertion ou les

longueurs.

Dans un L-système paramétrique, chaque module est composé d'une lettre issue de l'alphabet du L-système et associé à des paramètres réels. Dans les règles de productions, la valeur des paramètres est référencée en utilisant des paramètres formels. Les paramètres formels peuvent être utilisés dans des expressions arithmétiques ou logiques. Par exemple, la production :

```
A(t)->B(t+1)CD(t∧0.5,t-2)
```

peut être appliquée au module `A(9)` donnant le mot `B(10)CD(3,7)`.

4.4 Algorithme de conversion

Cette section décrit la transformation d'une chaîne de symboles obtenue par un L-système en un MTG. Considérons le L-système paramétrique suivant :

ω : `A(1)`
p_1 : `A(t)->B(t)[+(45)A(t*1.3)][-(45)A(t*1.3)]A(t)`
p_2 : `B(t)->B(t+1)`

Les deux premières chaînes générées par le L-système sont :

```
A(1)->B(1)[+(45)A(1.3)][-(45)A(1.3)]A(1)->B1(2)[+(45)B2(1.3)[+(45)A3(1.69)]
[-(45)A4(1.69)]A5(1.3)][-(45)B6(1.3)[+(45)A7(1.69)][-(45)A8(1.69)]A9(1.3)]
B10(1)[+(45)A11(1.3)][-(45)A12(1.3)]A13(1)]
```

Cette chaîne peut être interprétée comme un MTG de la façon suivante. Tout d'abord un identifiant numérique est affecté à chaque module `A` ou `B` :

```
B1(2)[+(45)B2(1.3)[+(45)A3(1.69)][-(45)A4(1.69)]A5(1.3)][-(45)B6(1.3)[+(45)A7(1
.69)][-(45)A8(1.69)]A9(1.3)]B10(1)[+(45)A11(1.3)][-(45)A12(1.3)]A13(1)]
```

A ce point, chaque paramètre attaché à un module est interprété comme une longueur. De plus, les symboles '+' et '−' sont des symboles nécessaires à l'interprétation de la tortue et peuvent donc être analysés comme des attributs géométriques :

```
B1(2)[B2(1.3,45)[A3(1.69,45)][A4(1.69,-45)]A5(1.3)][B6(1.3,-45)[A7(1.69,45)][A8
(1.69,-45)]A9(1.3)]B10(1)[A11(1.3,45)][A12(1.3,-45)]A13(1)]
```

Le deuxième paramètre de chaque module sera alors interprété comme un angle d'orientation (relatif) du module dans l'espace. Finalement, le tableau suivant indique le code final du MTG correspondant à cet exemple :

```
                                        Length  Diameter
/B1                                       2
            +B2                          1.3       45
                        +A3             1.69       45
                        +A4             1.69      -45
            ^<A5                         1.3
            +B6                           1.3      -45
                        +A7             1.69       45
                         +A8            1.69      -45
            ^<A9                         1.3
<B10                                      1
            +A11                         1.3       45
            +A12                         1.3      -45
<A13                                      1
```

L'algorithme de conversion a été développé de manière à filtrer la géométrie produite par la chaine issue du L-système d'origine. L'algorithme générique permettant de transformer la chaine L-système en un graphe multi-échelle est de ce fait non trivial.

4.5 Applications

Cet algorithme a été implémenté dans cpfg et a été utilisé pour l'évaluation des plantes virtuelles dans le cas d'inflorescences d'espèces de Syringa en utilisant une comparaison structurelle.

Durant ma thèse, j'ai pu proposer une distance entre plantes (Ferraro et Godin, 2000) qui permet l'évaluation des plantes simulées en les comparant avec des plantes réelles. Cette distance est définie comme le coût minimum de la séquence d'opérations d'édition transformant une plante en une autre. La distance est, par définition, normalisée et toujours comprise en 0. et 100. Outre le calcul de la distance entre plantes, l'algorithme met également en évidence *alignement* entre les plantes qui est intuitivement une description de la transformation de la plante initiale en la plante finale. Ce type d'analyse permet de montrer des similarités locales entre les deux plantes. Ces outils ont été implémentés dans AMAPmod et utilisés pour évaluer des inflorescences de lilas.

En 2001, Prusinkiewicz *et al.* (Prusinkiewicz et al., 2001) ont proposé une méthode de modélisation réaliste de plantes basées sur l'utilisation des propriétés d'auto-similarité des plantes (*cf.* section suivante). Ainsi, des inflorescences de lilas ont été modélisées en utilisant uniquement les mesures effectuées sur le sommet de la branche la plus longue. Une telle hypothèse a été par la suite vérifiée de manière quantitative selon deux méthodes distinctes (*cf.* section suivante et (Ferraro et al., 2002a; Münderman, 2002))

Considérons les cinq inflorescences représentées à la Figure 4. Ces modèles ont été simulés en utilisant cpfg, convertis selon l'algorithme précédent et comparés dans AMAPmod avec les inflorescences correspondantes mesurées sur le terrain. Le tableau suivant indique la distance et le pourcentage de correspondances locales parfaites durant la comparaison.

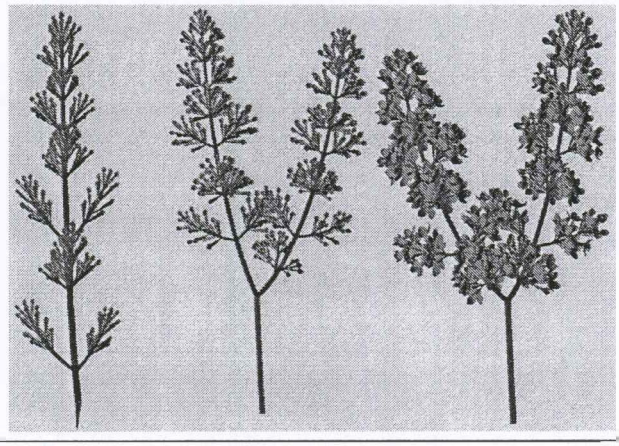

FIG. 4 – Cinq inflorescences de lilas simulées en utilisant cpfg (d'après L. Münderman).

Inflorescences	Distance	% de correspondances parfaites entre modèles et branches mesurées
A1	1.64	70.83
A2	1.96	78.26
A3	2.12	78.18
A4	2.17	79.16
A5	1.92	72.72

4.6 Conclusion

La conversion des plantes générées par des L-systèmes présentées dans cette section a été implémentée dans le logiciel de simulation CPFG. Le virtual laboratory (vlab) a été par la suite utilisé afin d'intégrer cpfg et AMAPmod dans un système cohérent et unique pour la modélisation et l'analyse des plantes.

5 Autosimilarité des structures ramifiées

5.1 Introduction

Le rôle joué par l'architecture dans la modélisation des plantes (Frijters et Lindenmayer, 1976; Hallé et al., 1978; Honda, 1971; Ross, 1981) a généré le besoin de nouveaux outils d'investigation. Des outils génériques ont été développés pour visualiser l'architecture des plantes en 3-Dimensions (de Reffye et al., 1989; Prusinkiewicz et Lindenmayer, 1990), pour modéliser le développement de la structure de la plante e.g. (Kurt, 1994; Mech et Prusinkiewicz, 1996; deReffye et al., 1997), pour mesurer son architecture (Godin et al., 1997a; Hanan et Room, 1997; Sinoquet et al., 1997), et pour explorer et analyser une plante (Godin et al., 1997b). Cet article introduit un nouvel outil pour l'identification des propriétés d'autosimilarité des plantes.

En 2001, Prusinkiewicz *et al.* (2001) ont proposé une méthode permettant de modéliser de manière réaliste un ensemble de plantes, basée sur l'utilisation de leurs propriétés (supposées) d'autosimilarité. Ainsi, pour des besoins de visualisation, une inflorescence de lilas a été modélisée en utilisant uniquement les mesures effectuées sur le sommet d'une branche puis étendues à toute la structure (*cf.* figure 17a (Prusinkiewicz et al., 2001)). L'utilisation d'une telle modélisation nécessite d'évaluer au préalable la ressemblance des branches deux à deux, ou plus précisément d'évaluer un certain degré d'autosimilarité des inflorescences. Ferraro et Godin (2000) ont proposé une distance entre architecture de plantes qui offre un outil efficace pour évaluer la ressemblance structurelle entre systèmes ramifiés. Dans cet article, nous développons une procédure, qui, partant des observations d'un système ramifié et utilisant cette distance, permet de quantifier le degré d'autosimilarité de la structure d'une plante et d'identifier les motifs identiques d'une structure arborescente.

Bien que la notion d'autosimilarité dans les plantes ne soit pas une notion botanique en tant que telle, ce concept peut être mis en relation avec les notions de "branch mapping", de relations paracladiales, d'axes de référence ou de réitérations développées dans la littérature. En nous appuyant sur une revue de ces concepts, (Section 5.2), nous développons les fondations mathématiques d'une distance entre architecture de systèmes ramifiés (Section 5.3) que nous utilisons pour formaliser le problème de la quantification de l'autosimilarité des plantes (Section 5.4). Nous illustrons enfin cette méthodologie sur des inflorescences de lilas comme un cas d'école. Nous terminons cette section par une discussion ouvrant les perspectives sur les travaux ultérieurs (Section 5.5).

5.2 Motivation Biologique

Pour de nombreux processus de croissance d'organismes vivants, et en particulier les plantes, la répétition régulière de certaines structures est aisément identifiable. Dans les cas les plus simples, les mêmes structures sont répétées périodiquement le long d'un axe, comme, par exemple les feuilles le long de la tige. Dans des cas plus complexes, la structure entière d'une étape précédente est répétée dans une partie de l'organisme à l'état suivant (M. Herman et al., 1975).

L'idée que les branches ont un comportement similaires aux branches qui les portent a fait l'objet de nombreuses études (Frijters et Lindenmayer, 1976; Honda, 1971; Edelin, 1984; White,

1979; Barthélémy, 1991; Prusinkiewicz et al., 2001). Troll définit par exemple une inflorescence comme un système constitué d'une inflorescence principale et de *paracladia* (Frijters et Lindenmayer, 1976), où un *paracladium* est une branche qui porte le même type de structure que celle portée dans la portion terminale de la branche mère (Frijters et Lindenmayer, 1976; Lindenmayer, 1977). En s'appuyant sur cette définition, Fritjers et Lindenmayer ont introduit la notion de relation paracladiale entre deux systèmes ramifiés (Frijters et Lindenmayer, 1976) : il y a relation paracladiale si "le comportement d'une branche est exactement le même que celui de la partie de la branche porteuse qui s'est développée depuis la formation de la branche fille". Fritjers et Lindenmayer (Frijters et Lindenmayer, 1976) puis Robinson (Robinson, 1986) ont ainsi proposé une procédure permettant l'identification de formules de récurrence des relations paracladiales sur les formes ramifiées. Si ces concepts biologiques peuvent être observés visuellement dans des cas évidents, pour certaines plantes l'auto-similarité de la structure peut être cachée par la complexité de la structure. De tels développements peuvent donc être observés dans la nature mais rarement de manière exacte et dans tous les cas, ils n'ont été que très rarement évalués quantitativement.

De Reffye *et al.* (de Reffye et al., 1990) ont proposé une approche de modélisation de la croissance des plantes s'appuyant sur un principe similaire. Pour cela, ils ont introduit la notion *d'axe de référence* basée sur le regroupement et le classement de toutes les étapes de différentiation d'un arbre le long d'un axe théorique. L'étude de *Zelkova Serrata* (Orme du Japon) à différents stades de son développement a montré que la croissance et les processus de ramification étaient similaires dans toutes les parties de l'arbre. Ils ont ainsi statistiquement identifié un axe théorique qui, de la graine à la mort de l'arbre, donne tous les états possible de la différentiation morphologique d'une unité de croissance[1]. Dans ce sens, "l'axe de référence" interprète le comportement des bourgeons dans une plante en s'appuyant sur une analyse quantitative des gradients des tailles d'organes sur les différents axes.

Nous proposons dans ce qui suit un outil permettant d'évaluer automatiquement la ressemblance entre différentes structures ramifiées dans une plante et ainsi d'identifier les relations paracladiales ou les structures homologues à la base du modèle d'axe de référence. Cet outil est basé sur une représentation particulière des sytèmes ramifiés à l'aide de graphes et s'appuie sur une comparaison structurelle de ces représentations.

5.3 Évaluation de la ressemblance entre deux systèmes ramifiés

Pour comparer deux systèmes ramifiés, une première approche consiste à résumer chaque individu par un petit nombre de variables synthétiques et globales (*e.g.* production de fruits, taille de la couronne, *etc.*). L'évaluation de la ressemblance de deux systèmes ramifiés est alors réduite à la comparaison des variables synthétiques. Ceci définit les méthodes de comparaison globales pour lesquelles l'organisation topologique des entités dans la structure ramifiée n'est pas prise en compte. Il existe cependant des domaines pour lesquels la structure d'un système ramifié joue un rôle important. Dans de tels cas, la notion de distance entre deux systèmes ramifiés devrait naturellement prendre en compte l'organisation spatiale des entités. Ceci définit les méthodes de comparaison analytique qui sont basées sur une comparaison élément par élément (Miclet, 1984).

[1]portion de la tige mise en place durant une période d'allongement continue (Caraglio et Barthélémy, 1997)

5.3.1 Définitions

Une plante est considérée comme un ensemble d'entités botaniques (*e.g.* entre-noeuds et noeuds, unités de croissance, pousses annuelles). L'organisation topologique peut être représentée par un graphe (Godin et Caraglio, 1998). Un graphe $G = (V, E)$ est constitué d'un ensemble V de vertex et d'un ensemble E d'arcs, chaque arc étant représenté par une paire de vertex (Preparata et Yeh, 1973). Les vertex représentent les entités botaniques et les arcs correspondent aux connections physiques entre celles-ci. Si (v_1, v_2) est un arc de E, le vertex v_1 est appelé père de v_2 et le vertex v_2 est appelé le fils de v_1 (Preparata et Yeh, 1973). L'ensemble des fils de v_1 est noté $son[v_1]$. Dans un graphe, un *chemin* de x_1 à x_n est une séquence de vertex $(x_1, x_2.., x_n)$ telle que pour deux vertex consécutifs $\{x_i, x_{i+1}\}$ de la séquence, (x_i, x_{i+1}) est un arc. Dans un graphe, un vertex v_1 est appelé l'*ancêtre* d'un autre vertex s'il existe un chemin de v_1 à v_2.

On affecte à chaque vertex un ou plusieurs attributs qui représentent des caractéristiques biologiques des entités. Chaque attribut est une valeur numérique réelle (*e.g.* le diamètre ou la longueur d'une entité) ou symbolique (*e.g.* le type d'une entité). Ces attributs seront donc considérés comme choisis dans un ensemble de label. Soit $\Sigma = a, b, c, \ldots$ un ensemble fini ou infini de labels, et soit λ un unique symbole dans Σ. Soit α une fonction d'étiquetage qui affecte un label à tout vertex d'un graphe. Nous suposons qu'une distance est d, appelée distance élémentaire, est définie sur l'ensemble des labels Σ. Une distance entre deux vertex v_1 et v_2 d'un graphe est alors définie en utilisant la distance élémentaire sur Σ : $d(v_1, v_2) = d(\alpha(v_1), \alpha(v_2))$. Nous supposons de plus, qu'un coût $\gamma(x)$ est affecté à tout label x tel que :

$$\forall (x, y) \in \Sigma : d(x, y) \leq \gamma(x) + \gamma(y)$$

Un coût sera de plus affecté à chaque vertex v : $\gamma(v) = \gamma(\alpha(v))$.

Dans une plante $\mathcal{P}$, toute entité étant physiquement attachée à, au plus, une entité parente, la structure topologique est représentée par une arborescence (Godin et Caraglio, 1998) $T = (V, E)$. Une arborescence est un graphe dans lequel tout vertex excepté un seul, r, appelé *racine*, a un et un seul parent. La racine n'a pas de parent. Dans la suite une arborescente enracinée en r sera représentée par $T[r]$. Afin d'identifier les différents axes (*cf.* section 5.4) d'une plante représentée par une arborescence $T[r]$, les arcs de l'arborescence sont étiquetés : une entité peut soit précéder (symbôle '<') soit porter (symbôle '+') une autre entité. Nous supposerons par la suite qu'une entité ne peut être connectée qu'à, au plus, une entité par un arc '<'. Une forêt est un graphe dont les composantes connexes sont des arbres. Pour tout vertex x d'une arborescence $T[v]$, $F[x]$ représente la *forêt enracinée en* x, *ie.* la forêt obtenue en enlevant à l'arborescence $T[x]$ sa racine x et tout les arcs incident de x. Par la suite, nous utiliserons le terme *forêt* pour désigner une forêt enracinée. L'ensemble de tous les sous-arbres et forêts seront respectivement notés $\mathcal{T}$ et $\mathcal{F}$.

5.3.2 Comparaison d'arborescences

La représentation des données sous forme d'arborescences totalement étiquetées permet d'exprimer une information structurelle, soit de type concret (géométrique, dans le cas d'images, par exemple), soit de type abstrait (relations hiérarchiques entre les éléments primitifs). La manipulation de ces structures nécessite la définition d'algorithmes efficaces. En reconnaissance de formes, la nécessité de comparer deux objets représentés par des arborescences (*e.g.* reconnaissance de

caractères (Lu et FU, 1978), corrélation de signaux (Cheng et Lu, 1985), sciences du comportement (Boorman et Oliver, 1973), classification hiérarchique (Lu, 1979), chimie (Takajashi et al., 1987), biologie moléculaire (Shapiro et Zhang, 1990), ...), pose fréquemment le problème de la comparaison de deux arborescences.

On peut distinguer deux approches spécifiques permettant de définir une mesure de dissimilarité entre deux arborescences. Une première approche, développée par des mathématiciens algébristes, définit des métriques sur des ensembles partiellement ordonnés (dont les arborescences sont un cas particulier). Ces travaux concernent généralement les arborescences dont seules les feuilles sont étiquetées, et sont appliqués en particulier à la comparaison des résultats d'algorithmes de classification hiérarchique sur un ensemble de données (Boorman et Oliver, 1973; Monjardet, 1981; Karonski et Palska, 1977). Une deuxième approche, communément appelée *problème de l'édition d'arborescences* (*tree editing problem* (Selkow, 1977), en anglais), consiste à effectuer une suite de transformations permettant de construire une arborescence cible à partir d'une arborescence initiale. La classification hiérarchique (Lu, 1979), la reconnaissance de formes (Miclet, 1984; Sanfeliu, 1990; Banegas et al., 1999), l'analyse de scène (Wernecke, 1995), la chimie (Takajashi et al., 1987) ou la biologie moléculaire (Shapiro et Zhang, 1990; Shasha et al., 1994) sont autant de domaines d'applications du calcul d'une distance d'édition. Ce calcul est basé sur la définition de trois opérations : l'insertion, l'élision et la substitution, d'autres opérations pouvant être envisagées, comme la fusion ou l'éclatement de vertex (Sanfeliu, 1990). Ces méthodes font apparaître plusieurs stratégies spécifiques qui s'appuient sur des définitions différentes des opérations d'édition. Selkow (Selkow, 1977) a proposé dans un premier temps d'appliquer les opérations d'insertion et d'élision uniquement aux feuilles des arborescences. Taï (Tai, 1979) définit une métrique pour laquelle l'insertion et l'élision de vertex s'appliquent à tout vertex des arborescences concernées mais qui conserve un ordre partiel prédéfini sur les vertex des arborescences (*arborescences ordonnées*). Entre ces deux métriques, de nombreuses autres distances ont été définies (Lu, 1979; Nakabayashi et Kamata, 1983; Tanaka, 1984; Tanaka, 1985) dérivant toutes de la distance de Taï (Sanfeliu, 1990). Celle de Whilhelm (Wilhelm, 1981), par exemple, considère une extension des distances de Taï qui s'affranchit de la conservation de la numérotation des vertex pour des arborescences ordonnées dont les étiquettes sont toutes deux à deux distinctes. Enfin, dans le cas d'arborescences non-ordonnées, Zhang (Zhang, 1993; Zhang, 1996) propose une modification des contraintes de Taï permettant de définir une nouvelle distance entre arborescences. Les trois opérations d'édition considérées sur toute arborescence non ordonnée sont les suivantes :

- la substitution d'un vertex v_1 par un vertex v_2 (cette opération consiste à modifier les labels des deux vertex v_1 et v_2) ;
- l'élision d'un vertex v , *ie.* l'ensemble des fils de v deviennent les fils du père v et v est effacé de l'arborescence T (ainsi que les arcs incidents de v) ;
- l'insertion d'un vertex v comme fils d'un vertex x : *ie.* l'opération symétrique de l'élision, l'ensemble des fils de x devient l'ensemble des fils de v et v devient l'unique fils de x.

Les différentes opérations d'insertion et d'élision sont représentées sur la Figure 5. Notons qu'une opération de substitution est une fonction de l'ensemble des arborescences sur lui même alors qu'une fonction d'élision est une fonction de l'ensemble des arborescences $\mathcal{T}$ vers l'ensemble des forêts $\mathcal{F}$ (avec la relation $\mathcal{T} \subset \mathcal{F}$). L'ensemble des opérations d'édition possible est noté $\mathcal{E}$.

Pour toute opération d'édition s, un coût $\gamma(s)$ est défini comme suit :

- $\gamma(s) = \gamma(C(x,y)) = d(x,y)$ si s est une opération $C(x,y)$ qui substitue les vertex x et y ;
- $\gamma(s) = d(D(x)) = \gamma(x,\lambda)$ si s est une opération $D(x)$ qui efface le vertex x et,

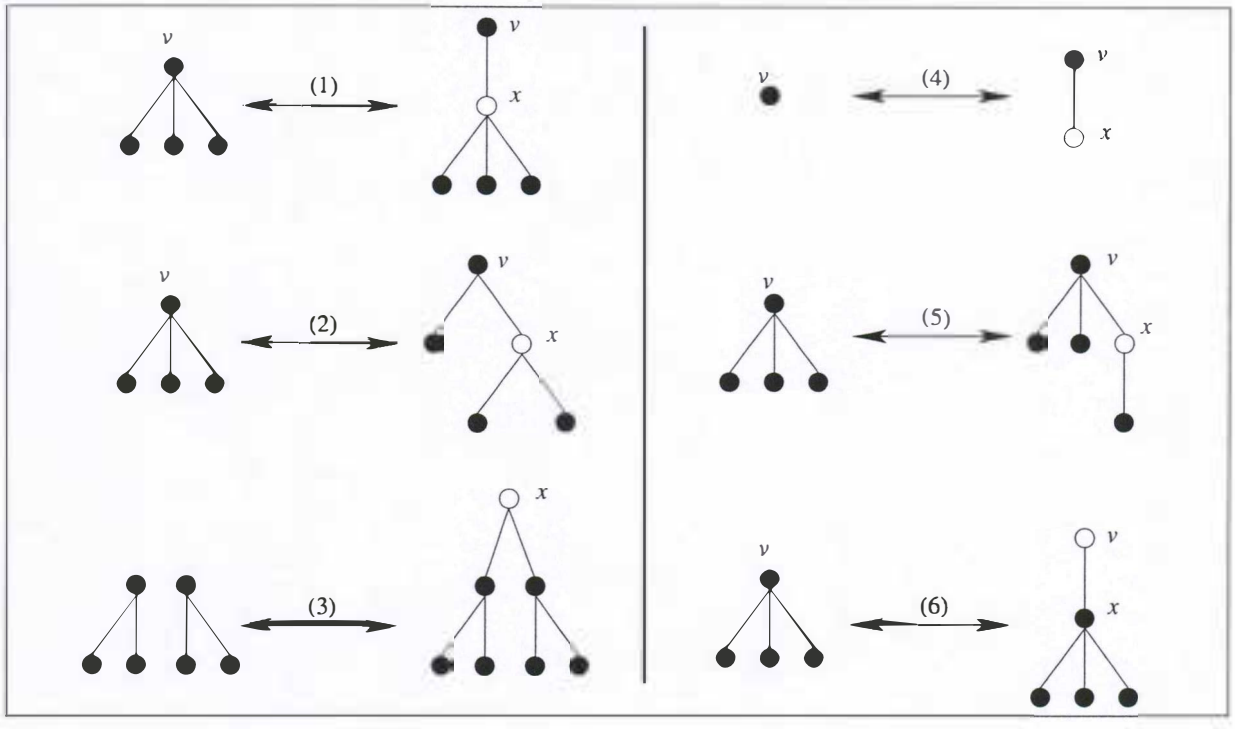

FIG. 5 – Les différentes opérations d'insertion et d'élision possible sur une arborescence. Les flèches de gauche à droite représentent une opération d'insertion, et les flèches de droite à gauche représentent les opérations d'élision. (1) L'ensemble des fils de v devient l'ensemble des fils de x, (2) un sous-ensemble strict des fils de v devient l'ensemble des fils de x, (3) x est inséré comme racine d'un ensemble d'arbres. Les opérations (4) à (6) sont des restrictions des opérations (1) et (3) mais ce sont des opérations rencontrées fréquemment. Dans l'algorithme, les opérations de type (2) ne sont pas autorisées.

– $\gamma(s) = d(I(y)) = \gamma(\lambda, y)$ si s est $I(y)$ une opération qui insère le vertex y.

Nous rappelons que d et γ vérifie la propriété suivante :

$$d(x, y) \leq \gamma(x) + \gamma(y) \tag{1}$$

Nous considérons deux arborescences $T_1[v]$ et $T_2[w]$, nous noterons $\mathcal{S}(v, w)$ l'ensemble de toutes les séquences d'opérations d'édition qui transforme $T_1[v]$ en $T_2[w]$. Soit S une séquence de n opérations d'édition $(s_1, s_2, \ldots, s_n)$ de $\mathcal{S}(v, w)$, le coût $\gamma(S)$ de la séquence est défini en sommant le coût des différentes opérations qui composent la séquence : $\gamma(S) = \sum_{s_i \in S} \gamma(s_i)$. La mesure de ressemblance $D(T_1[v], T_2[w])$ entre les deux arborescences est définie comme la séquence de $\mathcal{S}(v, w)$ de coût minimum :

$$D(T_1, T_2) = \min_{S \in \mathcal{S}} \{\gamma(S)\} \tag{2}$$

Il est possible de montrer (Ferraro et Godin, 2001) que la mesure de ressemblance ainsi définie est égale à la distance entre deux arborescences définie par Zhang (Zhang, 1996). Cet auteur a proposé un algorithme de programmation dynamique (Zhang, 1993; Zhang, 1996) calculant la distance entre deux arborescences T_1 et T_2 en un temps :

$$O\left(|T_1| \times |T_2| \times \max\left(\deg(T_1) + \deg(T_2)\right) \times \log_2\left(\deg(T_1) + \deg(T_2)\right)\right)$$

Où $\deg(T_1)$ (resp.$\deg(T_2)$ représente le nombre maximal de fils de l'arborescence T_1 (resp. T_2). Cet algorithme (Zhang, 1993; Zhang, 1996) s'appuie sur un calcul récursif des distances entre les sous-arborescences de $T_1[v]$ et $T_2[w]$ (détaillé dans (Ferraro, 2000b)) :

$$D(T_1[v], T_2[w]) = \min \begin{cases} D(\theta, T_2[w]) + \min_{w_k \in son[w]} \{D(T_1[v], T_2[w_k]) - D(\theta, T_2[w_k])\} \\ D(T_1[v], \theta) + \min_{v_k \in son[v]} \{D(T_1[v_k], T_2[w]) - D(T_1[v_k], \theta)\} \\ D(F_1[v], F_2[w]) + d(v, w) \end{cases}$$

Ce calcul fait apparaître le calcul de la distance entre les deux forêts $F_1[v]$ et $F_2[w]$ issues de v et w. On montre de la même façon que ce calcul s'effectue récursivement et s'appuie lui même récursivement sur le calcul de la distance entre arborescences (Zhang, 1993; Zhang, 1996).

Algorithmiquement, le calcul de la distance $D(T_1[v], T_2[w])$ nécessite celui d'une matrice M de dimension $n \times p$, où n et p sont respectivement les nombres de vertex de $T_1[v]$ et $T_2[w]$. Si on donne un ordre arbitraire (préfixé par exemple) sur les vertex des deux arborescences, le coefficient $m_{i,j}$ de M représente alors la distance $D(T_1[x], T_2[y])$ entre les sous-arborescences $T_1[x]$ et $T_2[y]$, où x et y sont respectivement les vertex numéro i et j des arborescences $T_1[v]$ et $T_2[w]$, la distance entre les deux arborescences $T_1[v]$ et $T_2[w]$ étant obtenue au coefficient $m_{0,0}$ de M (avec l'ordre préfixé).

5.3.3 Standardisation et normalisation de la distance obtenue

La distance entre arborescences basée sur une transformation d'édition, s'appuie sur la définition d'une distance élémentaire entre les vertex des arborescences. Cette distance affecte un réel positif $d(x_1, x_2)$ à toute paire de vertex (x_1, x_2) des deux arborescences $T_1[v]$ et $T_2[w]$ et un coût, réel positif $\gamma(x)$ à tout vertex x, vérifiant :

$$\forall (x, y) \in T_1[v] \times T_2[w] : d(x, y) \leq \gamma(x) + \gamma(y)$$

Une première distance élémentaire est la *distance de Levenstein* (Levenstein, 1966). C'est une fonction de coût local élémentaire basée sur une distance binaire. Dans ce cas, un coût nul est affecté à toute opération de substitution, et un coût de un pour toute opération d'insertion ou d'élision. Le coût local ainsi défini ne tient pas compte de la nature des vertex. La distance d'édition qui en découle est donc indépendante des vertex concernés par l'opération. Une distance basée sur cette fonction de coût local tient compte uniquement de la topologie des structures comparées, nous l'appellerons *distance topologique*. Le coût binaire ainsi défini peut être raffiné en utilisant les attributs liés aux entités tels que la longueur, le diamètre de la moëlle, le type, etc. Nous étudions dans ce qui suit ce coût variable en séparant les coûts de substitution de ceux de l'élision et de l'insertion. Dans le cas où l'on considère plusieurs attributs numériques ou symboliques (Ferraro et Godin, 2000), il est nécessaire d'homogénéiser la dynamique des différents attributs afin qu'ils aient une importance comparable dans la définition de la distance. Une procédure de standardisation est alors nécessaire.

Nous avons montré dans (Ferraro et Godin, 2000) que, dans le cas d'une distance locale topologique, la distance maximale entre deux arborescences est égale à la somme des tailles (du nombre de vertex) des deux arborescences comparées. La distance obtenue par l'algorithme de

Zhang sera donc normalisée de la manière suivante :

$$D'(T_1[v], T_2[w]) = \frac{D(T_1[v], T_2[w])}{(|T_1[v]| + |T_2[w]|)} \tag{3}$$

La distance ainsi obtenue est nécessairement comprise entre 0. et 1. et s'affranchit au mieux de la différence des tailles des deux arborescences.

5.4 Evaluation de l'autosimilarité dans une arborescence

Afin d'évaluer l'autosimilarité au sein d'une arborescence, nous proposons d'appliquer la méthode de comparaison structurelle, présentée précédemment à une même arborescence. En effet, le calcul de la distance entre une arborescence $T[v]$ et elle même selon l'algorithme de Zhang nécessite le calcul de la distance de l'ensemble des sous-arborescences de $T[v]$ deux à deux. En comparant $T[v]$ avec elle-même, nous faisons apparaître le calcul d'une matrice de distance M, de diagonale nulle et dont les coefficient $m_{i,j}$ sont les valeurs de la distance entre les sous-arborescences $T[v_i]$ et $T[v_j]$, où v_i et v_j sont deux vertex de $T[v]$. L'étude de cette matrice nous permet de mettre alors en évidence les sous-arborescences de $T[v]$ homologues.

De manière générale, pour l'évaluation de l'autosimilarité d'une arborescence, nous appliquerons des méthodes de classification sur cette matrice de distance. Dans le cas des plantes (représentées par une arborescence), nous émettons l'hypothèse qu'une plante reproduit sur ces ramifications toute (ou partiellement) la structure de la partie sommitale de la plante. Dans la suite de cette section, nous introduisons les définitions mathématiques permettant de vérifier une telle hypothèse.

5.4.1 Définitions

Dans ce qui suit, nous considérons une plante $\mathcal{P}$ représentée par un graphe $T[r]$. L'ensemble de toutes les sous-arborescences de $T[r]$ est noté $\mathcal{T}(r)$.

Soit v un vertex de $T[r]$ et soit $T[v]$ une sous-arborescence enracinée en v, nous définirons alors *l'axe* $A[v] = (V, E)$ de $T[v]$ comme le sous-graphe de $T[v]$ tel que $V = \{v_1, v_2, \ldots, v_n\}$ où $v = v_1$ et, pour tout $i \in \{1 \ldots n-1\}$, (v_i, v_{i+1}) est un arc de type '<' (figure 6). La hauteur de l'arborescence $T[v]$ est définie comme le nombre d'éléments de l'axe $A[v]$.

L'ensemble des sous-arborescences de $T[r]$ enracinées en un vertex de $A[v]$ est noté $\mathcal{A}(v)$:

$$\mathcal{A}(v) = \{T[v_1], T[v_2], T[v_3], \ldots, T[v_n]\}$$

Remarquons que dans cette définition v_{i+1} est nécessairement un fils de v_i (pour tout i dans $\{v_1, \ldots, v_{n-1}\}$). De plus, $\mathcal{A}(v)$ est un ensemble d'arborescences emboîtées les unes dans les autres (*cf.* figure 6). L'axe $A[r]$ est appelé *axe principal* de $\mathcal{P}$.

En nous appuyant sur la définition de la notion de *relation paracladiale* (Frijters et Lindenmayer, 1976; Lindenmayer, 1977), nous donnons une évaluation de l'autosimilarité en identifiant la ressemblance des différentes structures arborescentes avec la tige principale :

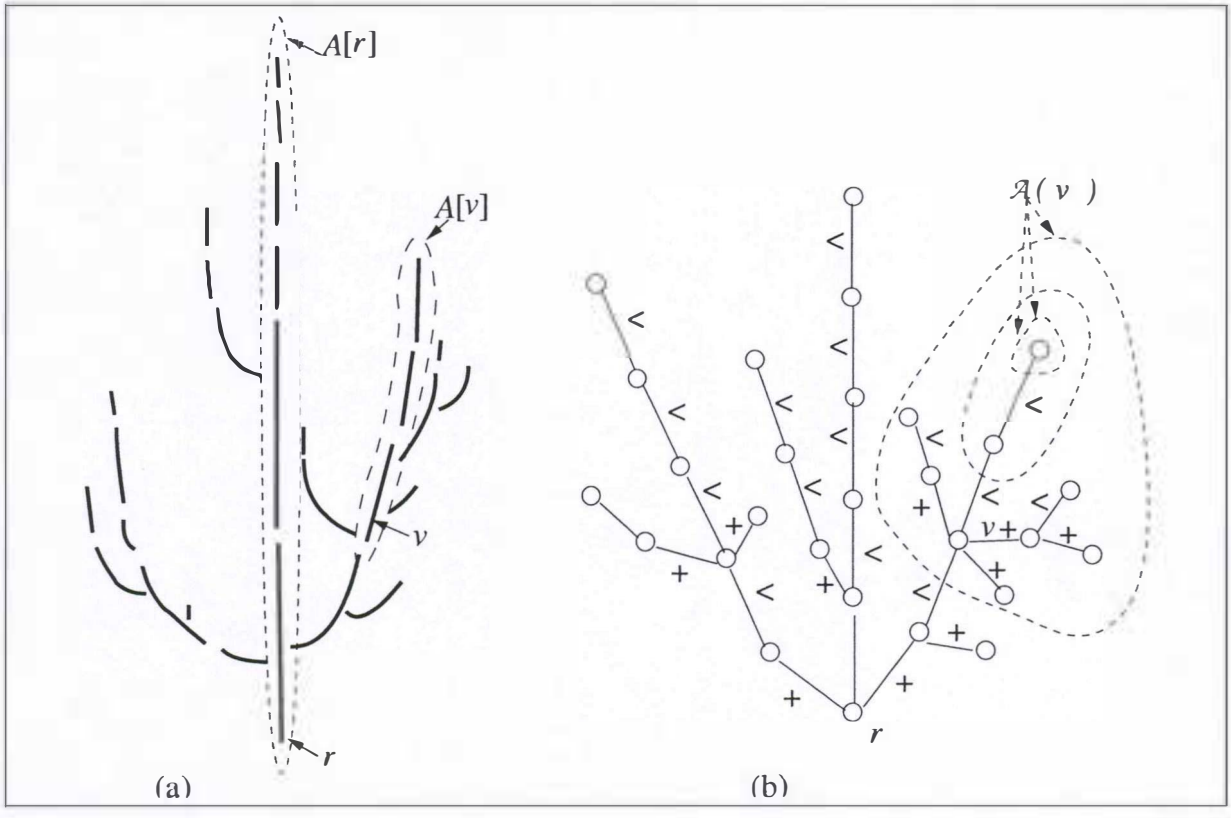

FIG. 6 – (a) un système ramifié et (b) sa représentation sous forme d'arborescence.

Definition 1 *Pour toute sous-arborescence $T[v]$ de $\mathcal{T}(r)$, le* coefficient paracladial *de $T[v]$ est la quantité :*

$$pr(v) = \min_{r_i \in A[r]} \{D(T[v], T[r_i]))\}$$

Pour une ramification quelconque $T[v]$, ce coefficient correspond à l'évaluation de la ressemblance entre la branche et la portion sommitale de l'axe principal qui est la plus ressemblante avec la branche $T[v]$. Nous dirons qu'une plante, représentée par une arborescence $T[r]$ est "*autosimilaire*" si toute sous-arborescence $T[v]$ de $\mathcal{T}(r)$ vérifie :

$$\forall T[v] \in \mathcal{T}(r) \qquad pr(v) = 0$$

En général, dans la nature on ne rencontre pas de système ramifié vérifiant parfaitement des propriétés d'autosimilarité. Dans des applications pratiques, pour évaluer de manière quantitative ces notions, nous considérons la valeur moyenne sur toutes les sous-arborescences $T[v]$ de $\mathcal{T}(r)$ du coefficient paracladial :

$$\overline{PR} = \frac{1}{|\mathcal{T}(r)|} \times \sum_{T[u] \in \mathcal{T}(r)} D(T[u], T[r_u]) \tag{4}$$

où $|\mathcal{T}(r)|$ représente le nombre d'arborescences dans $\mathcal{T}(r)$ (*ie.* le nombre de vertex de $T[r]$)

5.4.2 Étude de la valeur de $\overline{PR}$

Les valeurs de $\overline{PR}$ sont entre 0.0 et 1.0, la valeur maximale (1.0) étant atteinte lorsqu'on compare une arborescence avec un arbre nul (sans vertex, ni arc). Lorsque $\overline{PR}$ a une valeur proche de 0.0 on peut alors estimer que la structure est "quasi-autosimilaire".

D'un point de vue pratique, pour évaluer l'auto-similarité , nous devons tout d'abord définir une distance locale entre les vertex représentants les entités élémentaires de la plante. Dans nos exemples, nous n'utiliserons qu'une distance topologique, *ie.* nous ne tiendrons pas compte des éventuels attributs numériques ou qualitatifs affectés aux vertex. Dans ce cas là, la valeur de $\overline{PR}$ correspond au pourcentage moyen de vertex qui ont été insérés ou supprimés durant la comparaison. Un faible pourcentage indique alors que les ramifications sont très ressemblantes à la tige principale, et donc qu'il y a une forte autosimilarité de la structure.

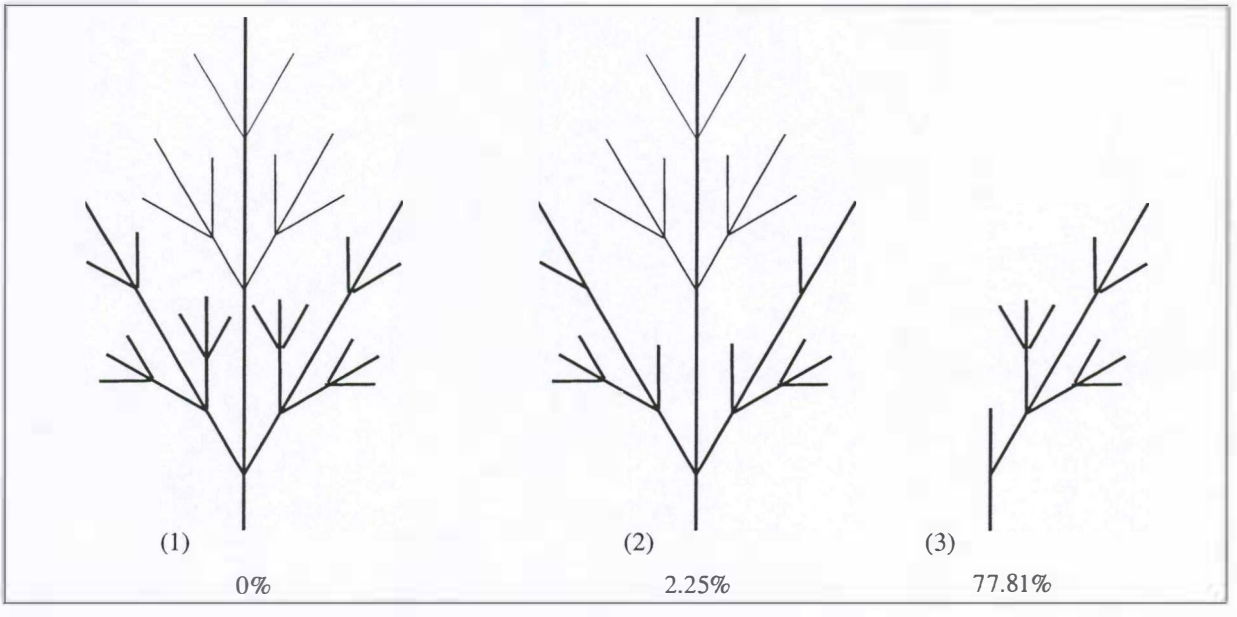

FIG. 7 – Trois structures ramifiées (1) parfaitement autosimilaire, (2) partiellement autosimilaire et (3) ne présentant pas un caractère autosimilaire. Le pourcentage indiqué est la valeur de $\overline{PR}$ dans chacun des cas.

Considérons par exemple les deux structures ramifiées (1), (2) et (3) de la figure 7. Pour la structure ramifiée (1), la structure parfaitement autosimilaire nous permet d'obtenir un coefficent paracladial nul. On peut constater que ce coefficient croit légérèrement lorsqu'on dégrade cette structure d'origine de quelques vertex, mais augmente très sensiblement lorsque la structure est fortement déséquilibrée (grande ramification portée par un tronc réduit à son minimum).

5.4.3 Analyse d'inflorescences de lilas

Nous présentons dans ce qui suit une application dans un cas réel de l'approche présentée ci-dessus. Dans cet exemple, nous considérons 5 inflorescences de lilas (noté A_1 à A_5), choisies pour leurs propriétés d'auto-similarité. Nous cherchons à évaluer quantitativement ces propriétés qui peuvent être observées visuellement. Les systèmes ramifiés représentant 5 inflorescences de lilas

sont représentés sous la forme d'une arborescence, chaque vertex de l'arborescence représente un "entre-nœud" (portion d'axe entre deux ramifications ou fleurs). Les longueur et diamètre de chaque entre-nœud ont été mesurés.

D'un point de vue pratique, pour évaluer l'auto-similarité , nous devons tout d'abord définir une distance locale entre entités élémentaires (ici les entre-nœuds) des inflorescences. Dans notre exemple, nous n'utiliserons qu'une distance topologique, *ie.* nous ne tiendrons pas compte des mesures réalisées sur les entre-nœuds. Dans ce cas là, la valeur de $\overline{PR}$ correspond au pourcentage moyen de nœuds qui ont été insérés ou supprimés durant la comparaison. Un faible pourcentage indique alors que les ramifications sont très ressemblantes à la tige principale, et donc il y a une forte autosimilarité de la structure.

Le tableau 1 indique les valeurs de $\overline{PR}$ pour chaque inflorescence, indiquant pour chacune d'elle un certain degré d'autosimilarité.

	(a) $\overline{PR}$	(b)
A_1	1.64%	70.83%
A_2	1.96%	78.26%
A_3	2.12%	68.18%
A_4	2.27%	79.16%
A_5	1.92%	72.72%

TAB. 1 – (a) Valeur de $\overline{PR}$ pour chaque inflorescence, et (b) pourcentage de branches d'ordre 2 ayant un axe de même taille que la structure homologue identifiée par l'algorithme sur la tige principale.

Nous pouvons remarquer que pour chaque inflorescence, les valeurs de $\overline{PR}$ sont proches de 0. Pour mettre en évidence cette remarque, la distance entre la branche et la portion la plus ressemblante sur la tige principale peut être observée graphiquement en utilisant une reconstruction 3-D des inflorescences (Mech, 1998). La figure 8 indique la valeur du cœfficient paracladial pour chaque branche. On peut remarquer que les plus fortes valeurs de $pr(v)$ sont généralement obtenues pour les branches du sommet. Les branches à la base de l'inflorescence qui ont eu une croissance plus importante et donc plus de chance de "subir" des accidents extérieur. On peut également observer une certaine variablilité du cœfficient paracladial au sein de la population, l'inflorescence A_1 possédant le meilleur taux de répétition dans sa structure.

La méthode présentée précédemment permet de plus d'identifier quelles sont les sous-parties homologues à l'intérieur de l'inflorescence. Si nous supposons qu'il existe une relation entre la taille de l'axe principal et une branche, cette supposition peut être vérifiée en utilisant une telle analyse. Le tableau 1 donne le nombre de branche d'ordre 2 qui ont exactement la même taille que la structure homologue identifiée par l'algorithme sur la tige principale et indique une "bonne" corrélation entre le nombre d'entités de l'axe de référence et une branche d'ordre 2.

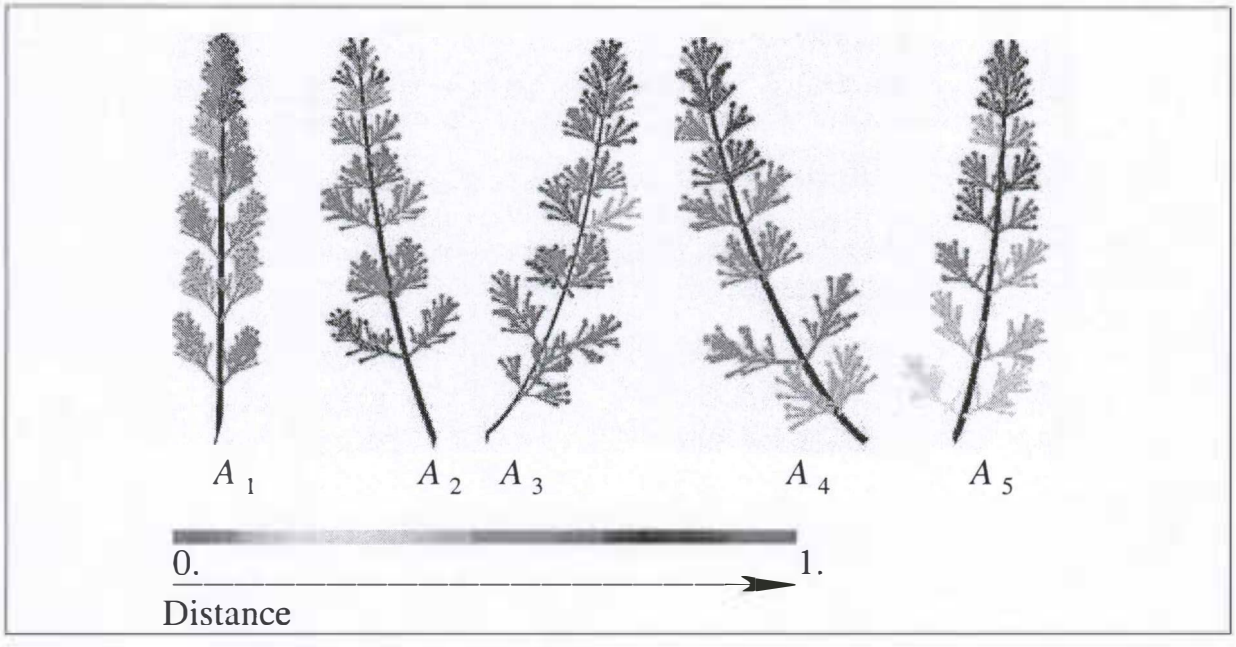

FIG. 8 – Représentation du coefficient paracladial pour chaque branche de chaque inflorescence.

5.5 Conclusion

Dans cet article, nous avons proposé une méthode analytique d'évaluation de l'autosimilarité des plantes. La quantification de l'autosimilarité dans une plante s'effectue à partir d'une comparaison structurelle des différents sous-systèmes ramifiés de la plante avec la tige principale. La distance entre deux structures ramifiées, fondée sur une méthode d'édition, nécessite le calcul d'une matrice de distances des sous-systèmes ramifiés deux à deux de la plante. La méthode propose une analyse d'une partie de la matrice, plus précisément, celle correspondant aux cœfficients de la tige principale. L'utilisation partielle de cette matrice nous permet tout de même l'identification de propriétés d'autosimilarité chez certaines plantes autorisant ainsi la réduction du nombre de mesures nécessaires à la construction de modèles (*cf.* (Prusinkiewicz et al., 2001)). Néanmoins, pour une évaluation plus complète de l'autosimilarité, les développements futurs de cette méthode prendront en compte l'ensemble de la matrice, c'est à dire la comparaison de toutes les sous-parties de plantes deux à deux.

L'appproche que nous proposons ici est une étape dans la l'évaluation quantitative de la production de structures équivalentes et emboitées chez les plantes. Nous envisageons par exemple d'utiliser cette méthode et ses extensions pour automatiser (au moins partiellement) l'inférence de modèles de croissance de plantes.

6 Perspectives

L'essor des bases de données contenant des descriptions de l'architecture de plantes crée un important besoin de nouveaux outils de traitement automatique et d'analyse (Godin et al., 1997a). Mon travail de thèse a jeté les bases de nouvelles techniques algorithmiques exploratoires prenant en compte la nature structurelle des données (Ferraro et Godin, 2000).

Le projet de recherche proposé consiste à étendre cette démarche selon deux axes :

- Axe théorique. Extension des outils conceptuels et algorithmiques ; l'idée est ici de concevoir les structures de données et les algorithmes associés qui permettent d'explorer des bases de données biologiques dans lesquelles les objets représentés ont une forte composante structurelle (e.g. arborescences, treillis, graphes multi-échelles, grammaires, réseaux de Pétri). Les outils développés seront des outils de caractérisation combinatoire et statistique ou structurelle (détermination d'un degré d'autosimilarité), des méthodes de comparaison (généralisation des methodes de comparaison d'arborescences aux graphes multi-échelles), de visualisation.
- Axe applicatif. Application de ces techniques à différents domaines biologiques. La modélisation des plantes constitue un domaine d'application privilégié. Toutefois, la généricité des méthodes développées dans l'axe théorique nous permet d'envisager des applications dans d'autres domaines biologiques et notamment en génomique et en phylogénie. Les méthodes développées s'apparentent en effet de très près aux méthodes utilisées dans l'analyse du génome (Durbin et al., 1998). Les structures manipulées dans le contexte de l'architecture des plantes sont toutefois structurellement plus complexes.

Pratiquement, le projet commencera par étudier les extensions des algorithmes de comparaison d'arborescences développés pendant la thèse et le post-doctorat et permettront de concevoir de nouveaux et puissants outils d'investigation. Par exemple :

- Étiquetage automatique. Ce problème consiste à retrouver dans une base de donnée architecturale toutes les occurrences d'un motif particulier de système ramifié afin de constituer, par exemple, un échantillon homogène. Actuellement, seul un repérage manuel de ces entités, occurrence après occurrence est possible. Un outil logiciel sera développé afin de propager l'étiquetage, effectué manuellement sur un ensemble limité de plantes, à l'ensemble de la base de données.
- Identification de motifs. C'est un problème voisin, mais techniquement très différent, qui consiste à identifier les sous-parties homologues au sein d'une ou plusieurs plantes. Cette fois, l'étiquetage n'est pas fait à l'avance, mais résulte au contraire de l'identification de structures homologues (étiquetées par un même label). Une telle technique pourrait être utilisée pour établir des schémas "d'auto ou d'inter-corrélation" sur l'architecture des plantes : à quels endroits une plante répète-t-elle les mêmes structures dans son architecture ? Seul un premier travail permettant de quantifier un certain degré d'autosimilarité des plantes a été accompli durant mon année de post-doctorat. Celui-ci ne répond que partiellement à une telle question considérée comme fondamentale dans la compréhension biologique de la plante.

L'enjeu est de constituer un ensemble de méthodes et d'outils de très haut niveau qui permettront aux biologistes d'analyser quantitativement la structure des végétaux sur des bases de données de grande taille. Ces outils seront mis en œuvre dans la plateforme logicielle AMAPmod et seront évalués sur des bases de données architecturales au sein de l'UMR AMAP, du Cirad,

de l'Inra et du réseau international PAIS (Plant Architecture Information System).

Enfin, nous chercherons comment ces méthodes peuvent être adaptées à d'autres champs d'application où les structures arborescentes et la notion de multi-échelles peuvent être également rencontrées : en informatique théorique, en génomique (Robinson et Foulds, 1981; Sankoff et al., 1983; Shapiro et Zhang, 1990) ou plus généralement en modélisation biologique.

Références

Banegas, F., Michelucci, D., Roelens, M., et Jaeger, M. (1999). « Significant features extraction from 3D points cloudsby ellipsoidal skeletons. Application in vision and geometric characterization. ». Dans *Internationnal Conference in Visual Computing.*

Barthélémy, D. (1991). « Levels of organization and repetition phenomena in seed plants ». *Acta Biotheoretica*, 39 :309–323.

Boorman, S. et Oliver, D. (1973). « Metrics on spaces of finite trees ». *J. Math. Psychol*, 10(26-59).

Caraglio, Y. et Barthélémy, D. (1997). Revue critique des termes relatifs à la croissance et à la ramification des tiges des végétaux vasculaires. Dans édition, I., éditeur, *Modélisation et simulation de l'architecture des végétaux*, Chapitre Part I, pages 11–87. Science Update.

Carpendale, M., Tigges, M., Cowperthwaite, D. J., et Fracchia, F. D. (1998). « Bringing the Advantages of 3D Distortion Viewing into Focus ». Dans *INFO-VIS'98 : Hot Topics, IEEE Conference on Information Visualization*, pages 17–20. IEEE Computer Society Press.

Cheng, Y. et Lu, S. (1985). « Waveform correlation by tree matching ». *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7 :299–395.

de Reffye, P., Dinouard, P., et Barthélémy, D. (1990). « Architecture et modélisation de l'Orme du Japon Zelkova serrata (Thunb.) Makino (Ulmaceae) : la notion d'axe de référence ». Dans *De la forêt cultivée à l'industrie de demain : 3e colloque Sciences et Industries du Bois, Arbora*, pages 351–352, Bordeaux, France.

de Reffye, P., Elguero, E., et Costes, E. (1989). « Growth units construction in trees : a stochastic approach ». *Acta Biotheoretica*, 39 :325–342.

deReffye, P., Fourcaud, T., Blaise, F., Barthélémy, D., et Houllier, F. (1997). « A Functional Model of Tree Growth and Tree Architecture ». *Silva Fennica*, 31(3) :297–311.

Durbin, R., Eddy, S. R., et Krogh, A. (1998). *Biological sequence analysis. Probabilistic models of proteins and nucleic acids.* Cambridge University Press, Cambridge, UK.

Edelin, C. (1984). « *L'architecture monopodiale : l'exemple de quelques arbres d'asie tropicale* ». PhD thesis, USTL, Montpellier II.

Federl, P. et Prusinkiewicz, P. (1999). « Virtual Laboratory : An interactive software environment for computer graphics ». Dans *Proceedings of Computer Graphics International 1999*, pages 93–100 and 242. IEEE CS.

Ferraro, P. (2000a). « Définitions d'une métrique entre arborescences quotientées ». Rapport Technique, Cirad.

Ferraro, P. (2000b). « *Méthodes algorithmiques de comparaison d'arborescences. Applications à la comparaison de l'architecture des plantes.* ». Informatique, Institut National Polytechnique de Toulouse (France), Toulouse (France).

Ferraro, P. et Godin, C. (1999). « An algorithm for comparing unordered tree graphs based on a minimum cost mapping with a minimal connectivity ». Dans *Third International Conference on Orders, Algorithms and Applications*, pages 179–195, Montpellier (France).

Ferraro, P. et Godin, C. (2000). « A distance measure between plant architectures ». *Annals of Forest Science*, 57 :445–461.

Ferraro, P. et Godin, C. (2002). « An Edit Distance Between Quotiented Trees ». Rapport Technique 1-2002, Cirad, Montpellier (France).

Ferraro, P. et Godin, C. (Soumis Dec. 2001). « An unified point of view on tree graph comparison ». *Pattern Recognition.*

Ferraro, P., Godin, C., et Prusinkiewicz, P. (2002a). « L-systems and MTGs : Integrating simulation and formal analysis of architectural plant models. ». Dans *16th European Simulation Conference*, Darmstadt, Germany.

Ferraro, P., Prusinkiewicz, P., Münderman, L., et Godin, C. (2001). « L-systems and MTGs : Integrating simulation and formal analysis of architectural plant models. ». Dans *Third International Workshop on Functional-Structural Trees Models*, pages 24–26, Montréal, QB, Canada.

Ferraro, P., Prusinkiewicz, P., Münderman, L., et Godin, C. (Soumis Jan. 2002b). « A structural method for assessing self-similarity in plants ». *Journal of Theoretical Biology.*

Frijters, D. et Lindenmayer, A. (1976). « Developmental descriptions of branching patterns with paracladial relationships ». *Automata, Langauges, Development*, pages 57–73.

Godin, C. et Caraglio, Y. (1998). « A multiscale model of plant topological structures ». *Journal of theoretical biology*, 191 :1–46.

Godin, C., Costes, E., et Caraglio, Y. (1997a). « Exploring plant topological structure with the AMAPmod Software : an outline ». *Silva Fennica*, 31 :355–366.

Godin, C. et Guédon, Y. (1999). *AMAPmod v1.2 Introduction and reference manual.* Cirad, Montpellier, France.

Godin, C., Guédon, Y., Costes, E., et Caraglio, Y. (1997b). Measuring and analyzing plants with the AMAPmod software. Dans Michalewicz, M., éditeur, *Advances in computational life sciences, Vol I : Plants to ecosystems*, volume January, pages 63–94. Csiro, Australia. chapitre 4.

Guédon, Y., Barthélémy, D., et Caraglio, Y. (1999). « Analyzing spatial structures in forests tree architectures ». Dans Amaro, A. et Tmoé, M., éditeurs, *Empirical and process-based models for forest tree and stand growth simulation*, pages 23–42, Oeiras (Portugal). IUFRO.

Guédon, Y. et Costes, E. (1998). « A statistical approach for analyzing sequences in fruit tree architecture ». Dans *Proceedings of Fifth international symposium on computer modelling in fruit research and orchard management.*

Hallé, F., Oldeman, R., et Tomlinson, P. (1978). *Tropical trees and forests. An architectural analysis.* Springer Verlag, New York.

Hanan, J. S. et Room, P. (1997). Practical aspects of plant research. Dans Michalewicz, M., éditeur, *Advances in computational life sciences, Vol I : Plants to Ecosystems*, volume January, pages 28–43. CSIRO, Australia. Chapitre 2.

Honda, H. (1971). « Description of the form of trees by the parameters of tree-like body ». *Journal of theoretical biology*, 31 :331–338.

Karonski, M. et Palska, Z. (1977). « On MS type distances between hypergraphs ». *Zatosowania Matermatyki*, XVI.1.

Kurt, W. (1994). « Growth grammar interpreter GROGRA 2.4 : A software for the 3-dimentional interpretation of stochastic, sensitive growth grammar in the context of plant modelling ». Introduction and reference manual, Forschungszentrum Waldokosysteme der Universitat Gottingen.

Levenstein, A. (1966). « Binary codes capable of correcting deletions, insertions and reversals ». *Sov. Phy. Dohl.*, 10 :707–710.

Lindenmayer, A. (1968). « Mathemathical models for cellular interaction in development ». *Journal of theoretical biology*, 18 :280–315.

Lindenmayer, A. (1977). « Paracladial relationships in leaves ». *Ber. Deutsch. Bot. Ges. Bd*, 90 :287–301.

Lu, S. et FU, K. (1978). « Error-correcting tree automata for syntactic pattern recognition ». *IEEE Trans. Computers*, 27 :1040–1053.

Lu, S.-Y. (1979). « A tree-to-tree distance and its application to cluster Analysis ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1 :219–224.

M. Herman, G., Lindenmayer, A., et Rozenberg, G. (1975). « Description of developmental languages using recurence system ». *Mathematical systems theory*, 8(4) :316–341.

Mech, R. (1998). « CPFG version 3.4 user's manual ». Rapport Technique, Department of Cmputer Science, University of Calgary, Alberta, Canada.

Mech, R. et Prusinkiewicz, P. (1996). « Visual models of plants interacting with their environment. ». Dans Rushmeier, H., éditeur, *SIGGRAPH 96 Conference Proceedings*, pages 397–410, New Orleans (Louisiana, USA). Addison-Wesley.

Miclet, L. (1984). *Méthodes Structurelles Pour la Reconnaissance Des Formes.* Eyrolles, 61, bd Saint-Germain - 75005 Paris.

Monjardet, B. (1981). « Metrics on partially ordered sets, a review ». *Discr. Math.*, 35 :173–184.

Münderman, L. (2002). « *? ? ? ? ?* ». PhD thesis, University of Calgary.

Nakabayashi, K. et Kamata, K. (1983). « A distance between two binary trees and its computing algorithm ». *Tans. IECE*, J66-D(4) :445–451.

Preparata, F. P. et Yeh, R. T. (1973). *Introduction to Discrete Structures for Computer Science and Engineering.* Addison-Wesley Publishing Company.

Prusinkiewicz, P. (1994). « Visual Models of morphogenesis ». *Artificial Life*, 1 :61–74.

Prusinkiewicz, P. (1998). « In search of the right abstraction : the synergy between art, science, and information technology in the modeling of natural phenomena ». *Scienta Horticulturae*, 74 :113–149.

Prusinkiewicz, P., Hammel, M., Hanan, J., et Mech, R. (1997). « *Plants to Ecosystems. Advances N Computanional Life Sciences,* », Chapitre L-System : From the Theory to Visual Models of Plants, pages 1–27. CSIRO publishing, michalewicz, m.t. ed. edition.

Prusinkiewicz, P. et Lindenmayer, A. (1990). *The algorithmic beauty of plants.* Springer Verlag, New York.

Prusinkiewicz, P., Mündermann, L., Karwowski, R., et Lane, B. (2001). « The use of positional information in the modeling of plants ». Dans SIGGRAPH, A., éditeur, *Computer Graphics Proceedings, Annual Conference Series, 2001*, pages 289 – 300, Los Angeles, CA.

Prusinkiewicz, P. et Remphrey, W. R. (2000). « Characterization of architectural tree models using L-systems and Petri nets ». Dans *4th International Symposium on the tree*, Montréal, QB, Canada.

Prusinkiewicz, P. p. l. (1996). « Virtual Plant Laboratory. A hypertext document and software distribution ». Rapport Technique, Departement of Computer Science. University of Calgary, Calgary, AB, Canada.

Robinson, D. F. (1986). « A notation for the growth of inflorescences ». *The New Phytologist*, 53 :587–596.

Robinson, D. F. et Foulds, L. (1981). « Comparison of phylogenetics trees ». *Mathematical biosciences*, 53 :131–147.

Ross, J. (1981). *The radiation regim and the architecture of plant stands.* Junk W. Pubs., The Hague, The Netherlands.

Sanfeliu, A. (1990). Matching tree structures. Dans Bunke, H. et Sanfeliu, A., éditeurs, *Syntactic and Structural pattern recognition. Theory and applications*, volume 7 de *Series in Computer Science*, pages 144–178. Word Scientific.

Sankoff, D., Kruskal, J. B., Mainville, S., et Cedergen, R. J. (1983). Fast algorithms to determine RNA secondary structures containing Multiple Loops 93. Dans Sankoff, D. et Kruskal, J. B., éditeurs, *Time Wraps, Strings Edits, and Macromolecules : the theory and practice of sequence comparison*, Chapitre 3, pages 93–120. Addison-Wesley Publishing Company Inc, University of Montreal, Montreal, Quebec, Canada.

Selkow, S. M. (1977). « The tree-to-tree editing problem ». *Information processing letters*, pages 184–186.

Shapiro, B. A. et Zhang, K. (1990). « Comparing multiple RNA secondary structures using trees comparisons ». *Cabios*, 6 :309–318.

Shasha, D., Wang, J. T.-L., Zhang, K., et Shih, F. Y. (1994). « Exact and approximate algorithms for unordered tree matching ». *IEEE transactions on systems, man, and cybernetics*, 24(4) :668–678.

Sinoquet, H., Rivet, P., et Godin, C. (1997). « Assessment of the three-dimensional architecture of walnut trees using digitizing ». *Silva Fennica*, 3 :265–273.

Tai, K.-C. (1979). « The tree-to-tree correction problem ». *Journal of the Association for Computing Machinery*, pages 422–433.

Takajashi, Y., Satoh, Y., Suzuki, H., et Saski, S. (1987). « Recognition of largest common structural fragment among a variety of chemical structures ». *Anal. Sci*, 3 :23–28.

Tanaka, E. (1984). « The metric between trees based on the strongly structure preserving mapping and its computing method ». *Trans. IECE*, J67-D(6) :722–723.

Tanaka, E. (1985). « A computing algorithm for the tree metric based on the structure preserving mapping ». *Trans. IECE*, E68(5) :317–324.

Wernecke, J. (1995). *The inventor Mentor. Programming Object-Oriented 3D Graphics with Open Inventor, Release 2.* Addison Wesley.

White, J. (1979). « The plant as a metapopulation ». *Ann. Rev. Ecol. Syst.*, 10 :109–145.

Wilhelm, R. (1981). « A modified tree-to-tree correction problem ». *Information processing letters*, 12(3) :127–132.

Zhang, K. (1993). « A new editing-based distance between unordered trees ». Dans *Combinatorial Pattern Matching, 4th Ann. Symp.*, pages 254–265, Padala (Italy). CPM'93.

Zhang, K. (1996). « A constrained edit distance between unordered labeled trees ». *Algorithmica*, 15 :205–222.

Remerciements

Le travail présenté dans ce mémoire a été réalisé au sein du groupe "Biological Modeling and Visualization Group" du laboratoire d'Informatique de l'Université de Calgary. Je tiens à remercier, ici celles et ceux qui y ont apporté leur contribution.

A Przemek Prusinkiewicz, directeur du groupe "Biological Modeling and Visualization Group", et professeur à l'Université de Calgary, je tiens à exprimer ici toute ma gratitude pour m'avoir accueilli au sein de son laboratoire pendant cette année de post-doctorat. Sa disponibilité, la passion et l'enthousiasme qui l'animent dans ses activités de recherche, son sens aigu du perfectionnisme m'ont servi de référence tout au long de la réalisation de ce travail. Son accueil chaleureux ont de plus rendu notre séjour des plus agréable.

Je souhaite également remercier Christophe Godin qui a supervisé depuis la France une partie de ce travail.

Je remercie Radek Karwoski et Lars Münderman qui m'ont initié à l'utilisation de cpfg et L-Studio, les riches discussions que nous avons pu avoir m'ont épaulé dans la réalisation de mon travail.

J'adresse un vif remerciement à Cathy Jirasek et Sheelagh Carpendale avec qui j'ai développé la loupe "multi-échelles" (qui n'est hélas pas présenté dans ce mémoire).

Comment ne pas remercier Camille ? Sa gentillesse et son efficacité dans toutes les démarches administratives qui ont émaillé mon séjour canadien m'ont été d'une aide précieuse.

Je tiens également à remercier chaleureusement, Calum, Maï, Robson, Pavel et Wayne (qui m'ont bien souvent apportés leur aide technique dans mes problèmes systèmes), Brian ainsi que Lyne et toutes les autres personnes qui ont rendu notre séjour agréable que ce soit sur un plan personnel ou professionnel.