

DU DESSIN DES ARBRES RESULTANT DE CLASSIFICATIONS HIERARCHIQUES

Gérard Thauront

Institut national de recherche sur les transports et leur sécurité

2, avenue du Général Malleret-Joinville

F-94114 ARCUEIL CEDEX

tél. 33 (1) 47 40 71 09 fax 33 (1) 45 47 56 06

thauront@inrets.fr

1. Introduction

Les programmes de classification arborescente, par exemple de classifications ascendantes hiérarchiques (CAH), fournissent comme résultat une distance ultramétrique sur les objets à classer. Cette distance ultramétrique est représentable sous forme d'un arbre hiérarchique.

Les programmes de CAH offrent des représentations graphiques prédéfinies, pas toujours satisfaisantes, au point que la rédaction d'une étude "soignée" nécessite de redessiner les arbres de résultat.

Ce travail présente :

- des considérations générales sur le dessin des arbres de classification ;
- des représentations alternatives ;
- des représentations adaptées à des cas particuliers.

Certaines de ces représentations sont programmées et accessibles par FTP anonyme.

L'étude a fait l'objet d'un exposé aux XXVIII journées de statistiques de l'ASU à Québec en mai 1996.

2. Trois visions du même modèle

La structure fournie par les programmes de classification hiérarchique est très riche. Elle peut être regardée selon différents points de vue selon que l'on veut mettre en évidence :

- la forme : **structure** d'ordre sur un arbre ;
- l'aspect classificatoire : système de **classes** emboîtées ;
- la notion de ressemblance : **distance** ultramétrique ;

Nous allons d'expliquer ces trois points de vues à l'aide de la plus ancienne et la plus fameuse des classifications hiérarchiques: celle des êtres vivants.

2.1 La vision en structure arborescente

C'est, par exemple, la vision de l'évolutionniste qui s'intéresse à la façon dont une population primitive s'est diversifiée en phylums distincts. C'est bien sûr la structure d'arbre qui va présenter le mieux cette vision : Un nœud de l'arbre pourra être identifié à "l'ancêtre commun" à toutes les feuilles qui en sont issues. Les branches issues de ce nœud représentant des lignées distinctes.