



Mathématiques et sciences humaines

Mathematics and social sciences

187 | Automne 2009

Journée 2007 de la Société Francophone de
Classification

Construction d'arbres à partir de relations d'intermédiarité, application au stemma codicum

*Tree construction starting from betweenness relations, application to the
stemma codicum*

Marc Le Pouliquen et Jean-Pierre Barthelemy



Édition électronique

URL : <http://journals.openedition.org/msh/11144>

DOI : 10.4000/msh.11144

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée

Date de publication : 30 décembre 2009

Pagination : 93-105

ISSN : 0987-6936

Référence électronique

Marc Le Pouliquen et Jean-Pierre Barthelemy, « Construction d'arbres à partir de relations d'intermédiarité, application au stemma codicum », *Mathématiques et sciences humaines* [En ligne], 187 | Automne 2009, mis en ligne le 15 décembre 2009, consulté le 23 juillet 2020. URL : <http://journals.openedition.org/msh/11144> ; DOI : <https://doi.org/10.4000/msh.11144>

CONSTRUCTION D'ARBRES À PARTIR DE RELATIONS D'INTERMÉDIARITÉ, APPLICATION AU *STEMMA CODICUM*

Marc LE POULIQUEN^{1,2}, Jean-Pierre BARTHÉLEMY^{1,3}

RÉSUMÉ – *Dans cet article, nous allons modéliser la relation ternaire d'intermédiarité dans le cadre de l'édition critique de texte. L'éditeur doit essayer de reconstituer au mieux, à partir des manuscrits préservés, le manuscrit original tel que l'auteur l'a écrit. Le corpus est constitué de manuscrits copiés les uns sur les autres. Une des méthodes utilisées consiste à élaborer un arbre de filiation des manuscrits restants, appelé le *stemma codicum*. Nous proposons de construire cet arbre à partir des relations d'intermédiarité entre les manuscrits : Un manuscrit B est entre les manuscrits A et C si le manuscrit C a été copié à partir du manuscrit B et que lui-même a été copié sur le manuscrit A.*

MOTS CLÉS – Arbre, Filiation de manuscrits, Relation d'intermédiarité, *Stemma codicum*

SUMMARY – Tree construction starting from betweenness relations, application to the *stemma codicum*

*In this paper, we model the ternary betweenness relation within the framework of the critical edition of texts. The editor must try to reconstruct as well as possible, starting from the various preserved manuscripts, the original manuscript as the author wrote it. The corpus is made up of many manuscripts which are copied from one another. To do so, it appears interesting to draw up a family tree of these manuscripts called *stemma codicum*. A manuscript B is between the manuscripts A and C, i.e. the manuscript C was copied starting from the manuscript B which itself was copied from A. This is this concept of betweenness by copy act which one wishes to model.*

KEYWORDS – Betweenness, Filiation of manuscripts, *Stemma codicum*, Tree

1. INTRODUCTION

L'édition critique de manuscrits anciens reconstitue, au mieux, à partir des différents manuscrits conservés, l'œuvre telle que l'auteur l'a voulue. Pour retrouver le texte original, l'éditeur confronte les différentes versions du texte qui ont été transmises par les copistes. En effet, l'auteur a pu écrire différentes versions de son texte, les copistes ont fait des erreurs (oubli de mot, saut de ligne, amélioration...), et les évolutions du temps et de

¹ Département LUSI TAMCIC, UMR CNRS 2872, ENST Bretagne, BP832, 29285 Brest Cedex {marc.lepouliquen, jp.barthelemy}@enst-bretagne.fr

² IUP Génie Mécanique et Productique, Université de Bretagne Occidentale, 6 avenue Le Gorgeu, CS93837, 29238 Brest Cedex 3

³ CAMS, UMR CNRS 8557, École des hautes études en sciences sociales, 54 boulevard Raspail, 75270, Paris cedex 06

l'espace (trous dans le papier, paragraphe illisible, évolution de la langue...) multiplient exponentiellement les dissemblances entre versions conservées.

La réalisation de l'édition critique classique se fait en plusieurs étapes dont l'une consiste à établir un arbre de filiation des manuscrits pour savoir lequel a été copié sur l'autre et de détecter les chaînons manquants : c'est l'établissement du *stemma codicum*. Le *stemma codicum* (cf. Figure 1.) se présente donc sous la forme d'un arbre généalogique des manuscrits représentant la filiation entre les différentes versions du même texte.

Cet article s'intéresse à la modélisation mathématique de l'intermédialité entre manuscrits permettant de construire ou de préciser le *stemma codicum*. Les méthodes phylogénétiques ou les treillis précédemment utilisés (cf. [Le Pouliquen *et al.*, 2006 ; Le Pouliquen, 2007]) ne permettent pas de détecter de manuscrits ancêtres communs qui apparaîtraient comme des nœuds intérieurs dans les graphes ou les arbres servant de modèles au *stemma codicum*.

En premier, nous allons tenter de modéliser cette relation dans le cadre de l'intermédialité de textes. Sur le *stemma* de la Figure 1., nous constatons que le manuscrit F est intermédiaire entre les manuscrits G et I, c'est-à-dire que le manuscrit G a été copié à partir du manuscrit F qui lui-même a été copié sur le I. C'est cette notion d'intermédialité par la copie que nous souhaitons modéliser.

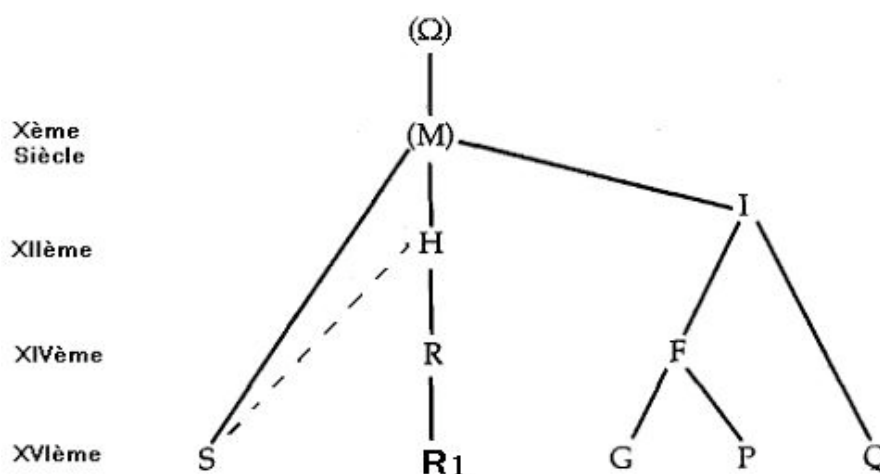


FIGURE 1. *Stemma codicum* chaque lettre correspond à un manuscrit, les parenthèses indiquent les manuscrits perdus ou supposés, les pointillés représentent les *contaminations*⁴

Dans un deuxième temps, nous nous intéressons aux possibilités que nous avons de reconstruire un arbre à partir d'un certain nombre de relations d'intermédialité entre les sommets. Nous pouvons nous demander si un ensemble de relations d'intermédialité est modélisable par un arbre et, si oui, comment peut-on le reconstruire ?

Pour finir, nous détaillerons les algorithmes que nous avons testés sur un corpus réel. Nous regarderons dans quelles mesures ils permettent d'améliorer les techniques de classification de manuscrits avant de conclure.

⁴Contamination : copie d'un manuscrit sur plusieurs modèles (syn : corruption, hybridation).

2. DIFFÉRENTES MODÉLISATIONS DE L'INTERMÉDIARITÉ AU NIVEAU DES MANUSCRITS

C'est à Quentin [1926] que l'on doit l'idée d'utiliser la notion d'intermédialité afin de dresser le stemma. En effet, il se propose de reconstituer des petites chaînes de trois manuscrits dont l'un est l'intermédiaire des deux autres puis, d'assembler ces petites chaînes afin d'inférer l'arbre complet. En s'inspirant de cette démarche, nous allons essayer de caractériser l'intermédialité.

Pour définir l'intermédialité, nous utilisons quatre caractérisations différentes, celle de la géométrie, celles liées à une structure définie par une relation binaire ou par un score et celle de la théorie des ensembles.

2.1. INTERMÉDIARITÉ EN GÉOMÉTRIE

Des nombreuses caractérisations géométriques, celle qui nous intéresse est celle introduite par Menger [1928] sous le nom de relation métrique d'intermédialité de la façon suivante :

DÉFINITION 1. *Une relation ternaire (a, b, c) sur une ensemble E est dite relation métrique d'intermédialité s'il existe une métrique d sur E telle que :*

$$(a, b, c) \Leftrightarrow d(a, b) + d(b, c) = d(a, c)$$

Il faut désormais définir une métrique qui permet de comparer les manuscrits de telle sorte que l'intermédialité au sens de la copie corresponde à celle de la métrique. Pour cela, prenons un exemple. Soient les 3 phrases suivantes correspondant aux trois mêmes phrases de différents manuscrits copiés les uns sur les autres.

A=Voici une phrase courte inventée pour l'exemple

B=Voici une phrase inventée pour cet exemple

C=Voici une phrase créée pour cet exemple

La logique veut que la phrase B soit intermédiaire au sens de la copie entre A et C. En effet, le copiste de B a omis *courte* et a modifié *l'* en *cet* et le copiste de C a remplacé *inventée* par *créée*. Il est en revanche peu probable que C soit l'intermédiaire, car le copiste de C a supprimé *inventée* qui est réintroduit par le copiste suivant.

Soit la distance d définie par le nombre de mots insérés, supprimés ou substitués entre 2 textes alignés (une sorte de distance d'édition [Levenshtein, 1966] au niveau des mots). Si nous réalisons les calculs sur l'exemple précédent, nous obtenons alors la Table 1.

A	Voici une phrase	courte	inventée	pour	l'	exemple
B	Voici une phrase		inventée	pour	cet	exemple
C	Voici une phrase		créée	pour	cet	exemple
$d(A, B) = 2$		$d(B, C) = 1$		$d(A, C) = 3$		

TABLEAU 1. Alignement des trois phrases pour compter le nombre de variantes

Nous constatons en effet que (A, B, C) puisque $d(A, C) = d(A, B) + d(B, C)$ et nous n'avons pas (A, C, B) .

2.2. INTERMÉDIARITÉ DÉFINIE PAR UNE RELATION D'ORDRE

Nous pouvons aussi présenter une relation ternaire avec une relation d'ordre.

DÉFINITION 2. *Une relation ternaire B que nous appelons intermédialité d'ordre est définie de la façon suivante : (a, b, c) si et seulement si $a \leq b \leq c$ ou $c \leq b \leq a$. Il faut donc définir une relation d'ordre entre les manuscrits au sens de :*

$$M1 \leq M2 \Leftrightarrow M1 \text{ a été copié sur } M2 \text{ ou est égal à } M2.$$

Malheureusement, étant donnés deux manuscrits, nous n'avons aucune façon de savoir si l'un a été copié sur l'autre ou vice versa à l'examen du texte. Nous pouvons nous douter que deux manuscrits trop dissemblables ne sont pas des copies l'un de l'autre. Mais les copistes étant plus ou moins précis, les manuscrits copiés sont plus ou moins proches et leurs proximités ne traduisent pas forcément une filiation directe. Il n'y a donc pas de modélisation envisageable de la relation binaire et donc de l'intermédialité qu'elle induirait.

2.3. INTERMÉDIARITÉ DÉFINIE PAR UN SCORE

Le but est de construire un indice qui permette de détecter si un manuscrit est intermédiaire entre deux autres au sens de la copie. Nous observons en premier lieu que trois stemmas différents peuvent être construits conformément à *B est intermédiaire entre A et C* (cf. Figure 2).

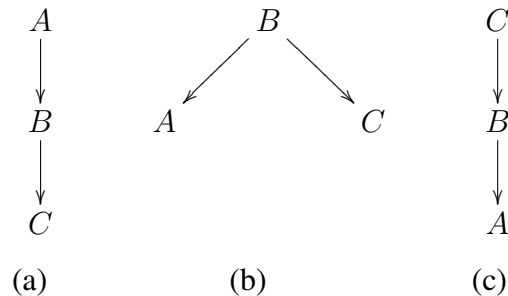


FIGURE 2. Stemmas possibles si « *B est intermédiaire entre A et C* »

Pour construire le stemma à partir des différents manuscrits de la tradition textuelle, nous partons du constat suivant, à savoir que toutes les copies qui contiennent, aux mêmes endroits, les mêmes fautes, ont été faites les unes sur les autres et donc dérivent toutes d'une copie où ces fautes existaient. Ces fautes sont appelées variantes ou leçons.

Dans les trois cas de la Figure 2, il y a un moyen de déterminer l'enchaînement des manuscrits à partir des variantes mais pas d'orienter le sens de la copie. En effet, pour que B soit intermédiaire entre A et C, il suffit que A et C s'accordent tour à tour avec B au niveau des variantes mais surtout qu'ils ne s'accordent jamais contre lui. Reprenons notre exemple précédent :

A=Voici une phrase courte inventée pour l'exemple

B=Voici une phrase inventée pour cet exemple

C=Voici une phrase créée pour cet exemple

Nous avons ici 3 lieux variants : courtes, inventée/créée et l' /cet correspondant à 6 variantes et résumés dans le tableau suivant :

n° du lieu variant	Variante	Phrases associées	Variante	Phrases associées
1	courte	A	{∅}	B,C
2	inventée	A,B	créée	C
3	l'	A	cet	B,C

Pour les lieux variants 1 et 3, nous voyons que B s'accorde avec C. Pour le lieu variant 2, B s'accorde avec A et surtout A et C ne s'accordent jamais contre B. B est donc bien intermédiaire entre A et C. Nous pouvons construire un indice en divisant le nombre de lieux variants où ils ne s'accordent pas par celui où B s'accorde avec l'un des deux. Nous déterminons ainsi si un manuscrit est « plus ou moins intermédiaire » (s'il est nul, il est intermédiaire). Cela correspond bien à l'indice recherché.

$$\text{Ici } \text{Indice}_B = \frac{0}{3} = 0, \text{Indice}_A = \frac{2}{1} = 2 \text{ et } \text{Indice}_C = \frac{1}{2} = 0,5$$

2.4. INTERMÉDIARITÉ EN THÉORIE DES ENSEMBLES

Nous allons ici nous intéresser à la relation d'intermédiarité introduite par Restle [1959]. Cette relation définit la notion d'intermédiarité au niveau des ensembles.

DÉFINITION 3. Soient trois ensembles A,B et C. On considère que B est intermédiaire entre A et C ssi :

- (i) $A \cap \overline{B} \cap C = \emptyset$
- (ii) $\overline{A} \cap B \cap \overline{C} = \emptyset$

Cela correspond à la Figure 3

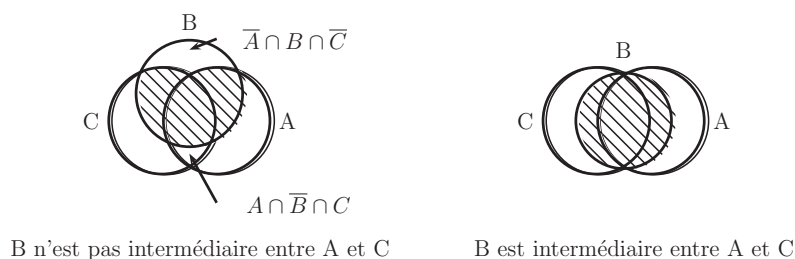


FIGURE 3. Intermédiarité au sens de Restle

Nous pouvons associer nos manuscrits à l'ensemble de leurs variantes. C'est en effet ce qui les caractérise les uns par rapport aux autres. Reprenons à nouveau notre exemple précédent :

A=Voici une phrase courte inventée pour l'exemple

B=Voici une phrase inventée pour cet exemple

C=Voici une phrase créée pour cet exemple

L'ensemble A est constitué des variantes {courte, inventée, l'}, B de {inventée, cet} et C de {créée, cet}. Nous pouvons alors le modéliser par la Figure 4.

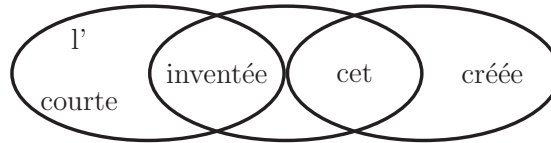


FIGURE 4. Visualisation de l'ensemble intermédiaire

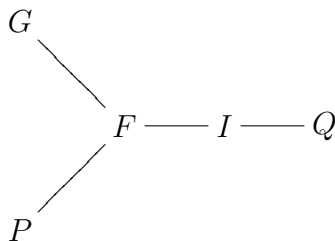
Cela correspond à nouveau au fait que B est entre A et C. Si nous retrouvons la même variante plusieurs fois dans le texte, nous devons les indiquer afin de les identifier et ne pas les mélanger.

3. UTILISATION DE L'INTERMÉDIARITÉ POUR LA REPRÉSENTATION DU STEMMA PAR UN ARBRE

3.1. INTRODUCTION

Dans le paragraphe précédent, nous avons modélisé la notion d'intermédiarité sur un corpus de manuscrits. Le résultat que nous obtenons est un ensemble de manuscrits et un certain nombre de relations ternaires entre ces manuscrits.

Supposons que le résultat soit le suivant : {F,G,I,P,Q} avec les relations (G, F, I) , (G, F, P) , (G, F, Q) , (F, I, Q) , (G, I, Q) , (P, F, Q) , (P, I, Q) et toutes les symétriques, alors l'arbre correspondant est :



(Arbre non orienté associé aux données)

Nous pouvons constater qu'il n'y a ici qu'un seul arbre non orienté répondant aux données. Nous pouvons nous demander si l'on peut toujours représenter des relations ternaires par un arbre ?

3.2. UN PEU DE THÉORIE

Pour la représentation du stemma, nous utilisons la notion d'arbre ainsi que toutes les notations qui lui sont liées définies par Barthélemy et Guénoche [1988]. Les manuscrits sont représentés par les sommets et les arêtes sont les relations de filiation. Dans un article,

Defays [1979] établit les conditions nécessaires et suffisantes pour pouvoir représenter, par un arbre, un ensemble de relations d'intermédiarité.

DÉFINITION 4. Soit Co un ensemble fini et $()$ une relation ternaire sur Co . $()$ est une relation ternaire d'intermédiarité ssi $\forall a, b, c \in Co$, les axiomes (i) et (ii) sont satisfaits.

- (i) (a, b, c) et $(b, a, c) \Leftrightarrow a = b$
- (ii) si (a, b, c) alors (c, b, a)

THÉORÈME 1. Soit $()$ une relation ternaire d'intermédiarité sur Co et les axiomes suivants.

- (i) $\forall a, b, c, d \in Co (a, b, c) \Rightarrow (a, b, d) \text{ ou } (d, b, c)$
- (ii) $\forall a, b \in Co$ Soit $\exists c$ tel que (c, b, a) sinon $\forall d (a, b, d) \text{ ou } (b, a, d)$
- (iii) $\forall a, b, c \in Co$, on a (a, b, c) ou (b, a, c) ou (a, c, b)

Si (i), (ii) et (iii) sont satisfaits, une représentation par un arbre linéaire (aussi appelé chaîne) est possible. Si (i) et (ii) sont satisfaits, une représentation par un arbre est possible avec comme correspondance un sommet de l'arbre pour un élément de Co . Si (i) est uniquement satisfait, une représentation par un arbre est possible avec plus de sommets que d'éléments de Co .

Dans le cadre de nos manuscrits, il suffit de vérifier le (i) pour savoir si notre corpus Co est représentable par un arbre. Cela nous impose un algorithme avec une complexité en n^4 pour vérifier cette faisabilité. Ensuite, il faut réaliser un algorithme de reconstruction d'arbre à partir de nos relations.

3.3. ALGORITHME ET EXEMPLES

ALGORITHME 1. Soit Co l'ensemble des manuscrits et B l'ensemble des relations d'intermédiarités

1. On vérifie la compatibilité de B avec un arbre, il faut que le (i) du Théorème 1 soit satisfait.
2. On détermine ensuite l'ensemble F des manuscrits qui ne sont pas des intermédiaires.
3. Pour chaque sommet i de F , on recherche le nœud k tel que l'on n'a pas de j intermédiaire entre i et k . On conserve alors les arêtes ik .
4. On enlève alors les manuscrits F de Co et les relations contenant des manuscrits de F de B
5. On réitère alors les étapes précédentes à partir de l'étape 2 jusqu'à ce que B soit vide.
6. Il nous reste alors N sommets dans Co .
 - Si $N=1$, on a un arbre que l'on peut reconstruire avec l'ensemble des arêtes conservées.
 - Si $N=2$ et qu'une relation d'intermédiarité contient les 2 sommets, on les relie et on peut reconstruire l'arbre.
 - Sinon, il faut rajouter des sommets et des arêtes ou des relations si l'on veut reconstruire un arbre.

Afin de mieux comprendre l'algorithme, voici 3 exemples qui prennent en compte les 3 sous-cas précédents. Le premier exemple correspond au cas $N=1$, le deuxième au cas $N=2$ et le dernier au cas où il faut rajouter des arêtes.

EXEMPLE N° 1 :

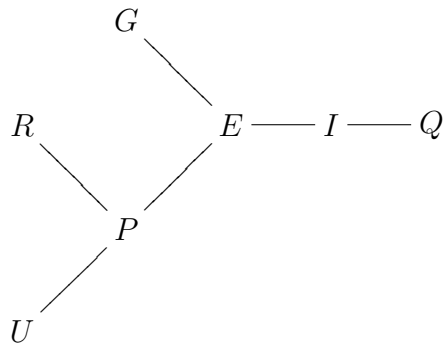
Soient l'ensemble des manuscrits $Co=\{E,G,I,P,Q,R,U\}$ avec les relations suivantes :

(R, P, U) , (R, P, E) , (R, P, G) , (R, P, I) , (R, P, Q) , (U, P, E) , (U, P, G) , (U, P, I) , (U, P, Q) , (R, E, G) , (R, E, I) , (R, E, Q) , (U, E, G) , (U, E, I) , (U, E, Q) , (P, E, G) , (P, E, I) , (P, E, Q) , (G, E, I) , (G, E, Q) , (R, I, Q) , (U, I, Q) , (P, I, Q) , (G, I, Q) , (E, I, Q) et toutes les symétriques. Nous avons alors $F=\{G,Q,R,U\}$ car ils ne sont pas intermédiaires et nous conservons les arêtes suivantes $G - E$, $Q - I$, $R - P$ et $U - P$.

Après avoir enlevé les manuscrits de F , nous obtenons au 2^e tour $Co=\{E,I,P\}$ une seule relation (P, E, I) et sa symétrique. Donc $F=\{I,P\}$ tandis que les arêtes sont $I - E$ et $P - E$.

Enfin, au 3^e et dernier tour, nous avons $Co=\{E\}$ sans relation.

En reconstituant l'arbre nous obtenons :

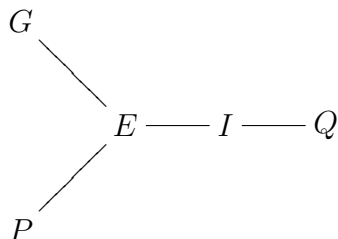


EXEMPLE N° 2 :

Soit l'ensemble des manuscrits $Co=\{E,G,I,P,Q\}$, les relations suivantes : (G, E, I) , (G, E, P) , (G, E, Q) , (E, I, Q) , (G, I, Q) , (P, I, Q) et toutes les symétriques. Nous avons ici l'ensemble des intermédiaires $F=\{G,P,Q\}$ et les arêtes associées $G - E$, $P - E$, $Q - I$.

Après avoir enlevé les manuscrits de F , nous obtenons au 2^e tour $Co=\{E,I\}$ sans relation.

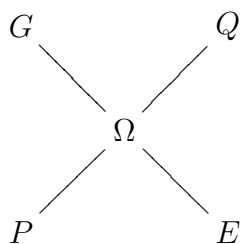
Comme E et I se retrouvent dans la relation (E, I, Q) , nous les relient et nous pouvons alors construire l'arbre suivant :



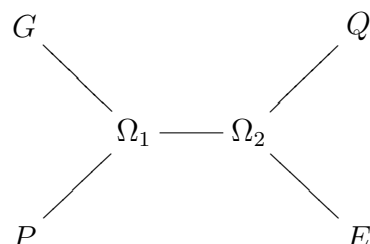
EXEMPLE N° 3 :

Soit l'ensemble des manuscrits $Co=\{E,G,P,Q\}$ sans relation d'intermédiarité. Nous pouvons alors envisager les arbres suivants :

Nous avons choisi 2 arbres quelconques, mais ce ne sont pas les seules possibilités. Nous constatons aussi que seul, dans ce dernier exemple le (ii) du Théorème 1, n'est pas satisfait. Si l'on veut malgré tout inférer notre stemma, deux alternatives s'offrent à nous :



Arbre avec un nœud intermédiaire



Arbre avec deux nœuds intermédiaires

1. soit nous complétons l'ensemble des manuscrits par des sommets supplémentaires correspondants aux manuscrits disparus ;
2. soit nous affaiblissons le critère d'intermédiarité afin d'obtenir plus de relations entre les manuscrits pour permettre de construire le stemma.

Finalement, trois méthodes sont envisagées selon la modélisation de l'intermédiarité choisie :

- Pour la relation métrique d'intermédiarité, nous choisissons de rajouter des sommets supplémentaires lors de la construction de l'arbre. Nous proposons un algorithme similaire à l'algorithme Neighbor-Joining (NJ) de Saitou et Nei [1987]. À chaque étape, NJ sélectionne deux sommets pour lequel il infère un ancêtre commun. Nous proposons donc, quand il n'y a pas assez de relations d'intermédiarité, de construire un nœud supplémentaire selon les critères de NJ, puis de reprendre l'Algorithme 1 avec les nouvelles relations obtenues grâce à ce nouveau sommet.
- Dans le cas où l'intermédiarité est définie par un score, c'est la deuxième alternative qui est retenue. La définition du score étant trop contrainte pour un corpus réel, souvent, nous n'obtenons pas beaucoup de solutions (relations d'intermédiarité). Il apparaît naturel de permettre à certaines contraintes d'être relaxées en augmentant l'indice d'intermédiarité minimum afin de faire en sorte que les manuscrits *presque intermédiaires* deviennent intermédiaires.
- Enfin, pour l'intermédiarité en théorie des ensembles, nous pouvons rajouter des sommets supplémentaires à l'ensemble des manuscrits. Nous pouvons regrouper les manuscrits qui ont la plus grande intersection (le plus grand nombre de variantes communes) et considérer qu'ils sont issus d'un même manuscrit perdu que nous rajoutons au corpus. Il suffit alors de relancer l'algorithme précédent avec les nouveaux manuscrits et les nouvelles relations obtenues.

4. APPLICATIONS

4.1. MÉTHODES

Précédemment, nous avons envisagé trois méthodes de reconstruction d'arbres à partir de corpus. Le schéma (cf. Figure 5) nous résume les différentes possibilités. Trois types d'arbres sont proposés dans le schéma et décrits par Barthélemy et Guénoche [1988]. Nous utilisons surtout les arbres hiérarchiques pour leurs ressemblances avec un *stemma codicum* classique.

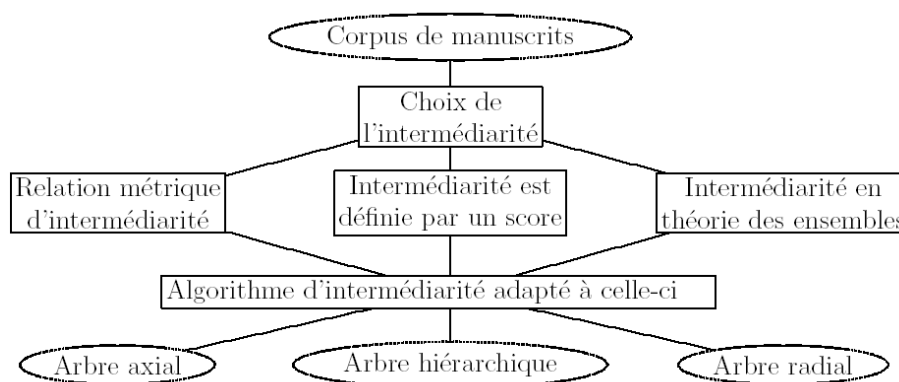


FIGURE 5. Méthodes de reconstruction d'arbres par l'intermédiarité

4.2. APPLICATION AU CORPUS DE TEXTES : TERTULLIANUS

Afin de réaliser les premières expérimentations, nous avons utilisé le corpus Tertullianus, De Pallio, de Marie Turcan [2007]. En plus de la collation⁴ des textes sous forme d'apparat critique⁵ (cf. Figure 6), Marie Turcan nous propose un stemma (cf. Figure 7). Nous avons utilisé les différentes méthodes proposées précédemment pour reconstruire le stemma que nous comparons avec celui proposé. Afin de permettre une comparaison plus aisée, nous nous sommes limités aux manuscrits suivants : $Co = \{F, G, L, N, R1, R2, R3, S, V, X\}$

Ligne 1 *post uiri add. affrice [-ca M] uiri NFX ||*
 Ligne 3 *annonae LatPC Rig : annona et mssR*
Salm ||
 Ligne 5 *uobis NR : nobis FX || olim : enim N ||*
 Ligne 6 *subteminis : -tegninis NR || concilio*
Iun Salm : consilio mssR || mensurae R³ Salm :
mensura e X mensura et M^{pc} FR¹ ||
 Ligne 7 *crura : cura M^{pc} || nec intra : ne intra X ||*
inuerecundae R³ Salm mss :
uerec- R¹ ||
 Ligne 8 *parae testes : parce V^{pc} edd. praet. Bu ||*

FIGURE 6. Exemple d'apparat critique de Marie Turcan

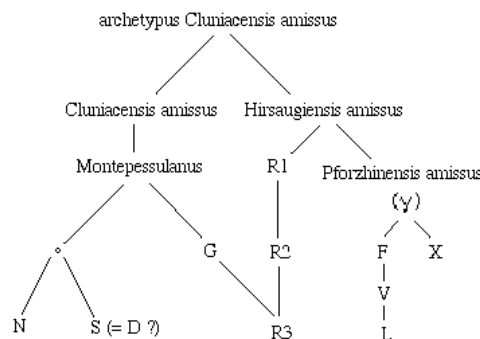


FIGURE 7. Stemma codicum adapté du stemma établi par J.-C. Fredouille

⁴Dans le travail philologique, la collation consiste à relever les variantes d'un texte afin de permettre l'établissement de celui-ci. Ces variantes proviennent généralement de l'auteur lui-même, de copistes ou d'éditeurs.

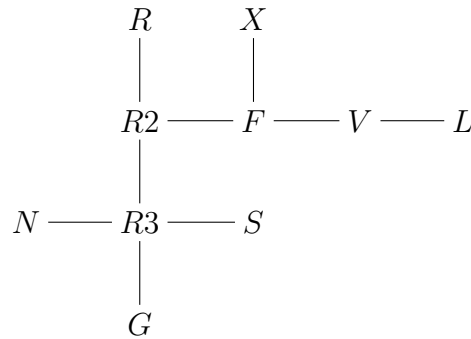
⁵L'ensemble constitué par les indications de variantes et les éventuelles notes associées constitue ce que l'on appelle « l'apparat critique » du texte.

Après une première itération de l'algorithme avec les scores, il s'avère qu'aucun manuscrit n'est un intermédiaire. Nous décidons alors d'augmenter l'indice d'intermédiarité minimum afin d'obtenir des relations de *presque intermédiarité* convenables. Nous obtenons alors les relations suivantes :

(L, F, X) , (V, F, X) , (F, V, L) , (L, V, X) , $(G, R2, R1)$, $(R3, R2, R1)$, $(G, R3, R1)$, $(G, R3, R2)$ ce qui correspond aux 2 chaînes suivantes :

$$G \text{ — } R3 \text{ — } R2 \text{ — } R1 \qquad L \text{ — } V \text{ — } F \text{ — } X$$

Et après trois itérations, le (ii) du Théorème 1 est satisfait ce qui correspond à l'arbre suivant :



Le résultat obtenu n'est pas convaincant. Si l'augmentation de l'indice d'intermédiarité est intéressant au début, rapidement l'algorithme trouve des intermédiarités partout, même où il n'y en a pas (Ex. $(R2, F, X)$). C'est donc à l'éditeur de choisir les limites de ce qu'il estime raisonnable lors de la construction du stemma.

Dans le cas de l'utilisation de l'algorithme associé à la relation métrique d'intermédiarité, une seule relation apparaît, c'est (F, V, L) . Nous construisons alors un nouveau sommet à partir de NJ, ce qui nous donne finalement de nouvelles relations d'intermédiarité. Après plusieurs itérations, le (ii) du Théorème 1 est satisfait et nous obtenons alors l'arbre suivant :

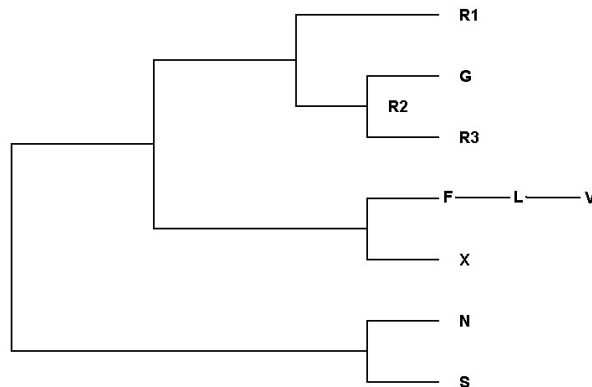


FIGURE 8. Arbre obtenu par la relation métrique d'intermédiarité et inspiré par NJ

L'arbre obtenu est relativement proche du stemma, ce qui laisse présager l'intérêt de la méthode même si peu de relations intermédiarité apparaissent à la première itération.

La troisième méthode n'ayant pas encore été testée sur ce corpus, les résultats sont en attente.

5. CONCLUSION

Les méthodes d'intermédiarité constituent un outil indispensable dans la classification de tous les documents. Il est toujours intéressant de savoir qu'un objet s'intercale entre deux autres. Dans le cadre de documents textuels, le classement peut s'effectuer selon plusieurs axes : les thèmes, les auteurs, la grammaire, le style, les dates... Dans tous ces cas, il faut définir l'intermédiarité pour positionner notre document parmi les autres.

Dans l'édition critique, et plus précisément dans la reconstitution de l'histoire d'un texte et sa schématisation (*stemma codicum*), nous avons tenté de modéliser la notion de *copie intermédiaire*. Nous constatons que cette modélisation stricte ne permet pas de retrouver les intermédiaires dans le cas d'un corpus réel. Soit nous trouvons peu d'intermédiaires *purs*, soit nous recherchons des manuscrits presque intermédiaires que seul un expert, l'éditeur, est capable d'ajuster pour ne pas dénaturer la relation ternaire obtenue. Ce n'est pas étonnant puisque c'est le reproche qui était fait à Don Quentin et à sa méthode par de nombreux philologues. Si la reconstruction du *stemma* est difficile, la méthode permet en revanche de détecter si un manuscrit est intermédiaire entre deux branches de la tradition textuelle ou s'il a un rôle de manuscrit central dans une sous-tradition.

De plus, si sur un corpus, le nombre de relations d'intermédiarité était trop important, la reconstruction du *stemma* peut devenir très lente. Il faut alors diminuer le nombre de relations ternaires obtenues. La suppression de relations pour obtenir une sorte de base minimum (« re-déployable » par transitivité et symétrie (cf. Pitcher et Smiley [1942]) permettrait d'améliorer l'algorithme et de simplifier le problème.

Un autre point d'intérêt se situe dans l'orientation des relations. Non seulement elle permettrait d'orienter le *stemma* et donc de déterminer la racine de notre arbre (ici le manuscrit original encore appelé « archétype »), mais en plus, elle simplifierait énormément la reconstruction de l'arbre.

Enfin, la modélisation permet aussi de visualiser la contamination d'un corpus. Si l'on examine la Figure 1., on constate un trait en pointillés entre les manuscrits H et S. Cela signifie que le manuscrit S a été copié à partir de (M) mais aussi à partir de H. Notre *stemma* n'est plus un arbre (il y a un cycle M,S,H) que l'on observe par les intermédiarités (M, S, H) , (S, H, M) et (H, M, S) .

BIBLIOGRAPHIE

- BARTHÉLEMY J.-P., GUÉNOCHE A., *Les arbres et les représentations des proximités*, Paris, Masson, 1988.
- DEFAYS D., "Tree representation of ternary relation", *Journal of Mathematical Psychology* 19, 1979, p. 208-218.
- LE POULIQUEN M., "Using lattices for reconstructing *stemma*", *Actes de la fifth international conference on Concept Lattices and their Applications, CLA, CEUR workshop proceedings, CEUR-WS org*, vol. 331, 2007, [<http://www.lirmm.fr/cla07/>].

LE POULIQUEN M., BARTHÉLEMY J.-P., BERTRAND P., « Filiation de manuscrits sanskrits et arbres phylogénétiques », RNTI-C-2 classification : points de vue croisés, 2008, *Actes de la XIII^e rencontre de la Société Francophone de Classification*, 2008.

[<http://lita.sciences.univ-metz.fr/~sfc06/>]

LEVENSHTAIN V.I., “Binary codes capable of correcting deletions, insertions and reversals”, *Soviet physics - Doklady* 10(8), February 1966, p. 707-710.

MENGER K., “Untersuchungen über allgemeine metrik”, *Mathematische Annalen* 100, 1928, p. 75-163.

PITCHER E., SMILEY M.F., “Transitivities of betweenness”, *Transactions of the American Mathematical Society* 52, 1942, p. 95-114.

QUENTIN H., *Essais de critique textuelle*, Paris, Picard, 1926.

RESTLE F., “A metric and an ordering on sets”, *Psychometrika* 24, 1959, p. 207-220.

SAITOU N., NEI M., “The neighbor-joining method : a new method for reconstructing phylogenetic trees”, *Molecular Biology and Evolution* 4, 1987, p. 406-425.

TURCAN M., « Le manteau », *Tertullien, De pallio. Introduction, édition critique*, traduction française, commentaire et index, Paris, éditions du Cerf, Coll. Sources chrétiennes, 2007.

[URL <http://www.tertullian.org/articles>]