

# Utilisation d'une méthode de super-arbre pour établir le stemma codicum

Marc Le Pouliquen<sup>1</sup>, Marc Csernel<sup>1</sup>

<sup>1</sup> Telecom Bretagne, Labsticc UMR 3192, BP 832, 29285 Brest Cedex - France  
marc.lepouliquen@telecom-bretagne.eu

## Abstract

The goal of this paper is to propose a new stemma codicum reconstruction method. Despite the development of many methods of automatic reconstruction (in particular in phylogenetic), editors are still reluctant to approve of stemmata obtained automatically. We therefore propose a new method based on philological arguments that the editor will be able to understand and validate. In the field of the phylogenetics, the consensus methods or super-tree methods allow a global tree from various local trees to be reconstructed. Here we present a new method for super-tree based on the use of betweenness triplets (philologically defined by Don Quentin) and inspired by the Build method of (Aho et al., 1981) which is based on rooted triplets.

## Résumé

Dans cet article, nous allons proposer une nouvelle méthode pour reconstruire le stemma codicum. Malgré le développement de nombreuses méthodes de construction automatique (en particulier phylogénétique), les éditeurs restent toujours réticents envers les stemmata obtenus automatiquement. Nous proposons donc une nouvelle méthode basée sur des arguments philologiques que l'éditeur comprendra et pourra valider. Dans le domaine de la phylogénétique, ce sont les méthodes de consensus ou de super-arbre qui permettent de reconstruire un arbre global à partir de différents arbres locaux. Nous présentons ici une nouvelle méthode de super-arbre basée sur l'utilisation de triplets d'intermédiarité (définis philologiquement par Don Quentin) et inspirée sur la méthode Build de (Aho et al., 1981) qui, elle, travaille sur les triplets enracinés.

**Mots-clés :** Super-arbre, consensus, édition critique, stemma codicum

**Dédicace :** Marc Csernel nous a quittés le 21 Août 2013. Nous avons alors en commun sept ans de collaboration scientifique et d'amitié dont la rédaction de cet article qui était déjà bien avancée et la participation à plusieurs JADT. Je dédie cet article à cet homme de culture qui était aussi un grand voyageur.

## 1. Introduction

Une édition critique classique propose l'établissement d'un texte qui soit le plus fiable possible à partir des différents manuscrits de l'œuvre dont on dispose. L'éditeur doit alors trier les différentes versions du texte et faire des choix entre ces versions afin d'essayer de reconstituer le manuscrit original avec fidélité. En effet, l'auteur a pu écrire différentes versions du texte, les copistes ont fait des erreurs (oubli de mot, saut de ligne, amélioration...), et les évolutions du temps (trous dans le papier, évolution de la langue...) multiplient les dissemblances entre les versions conservées. Une des méthodes classiques pour découvrir le texte original à partir des différentes versions est due à (Lachmann et al., 1876). Dans cette méthode, l'éditeur utilise les différences (appelées variantes) entre les versions du texte (appelées témoins) pour découvrir les filiations entre ces versions. Il dresse alors un arbre généalogique de cette filiation que l'on nomme stemma codicum (*cf.* Figure 1). Ce stemma

codicum permet ensuite à l'éditeur de reconstruire le texte original en examinant les variantes des différentes branches de la filiation et en choisissant les « meilleures ».

Il est évident que dans le cas de tradition textuelle importante, la reconstruction du stemma codicum constitue un défi. Le nombre des témoins, la taille du texte, les différentes langues utilisées ainsi que la contamination<sup>1</sup> sont autant d'obstacles à la création de l'arbre ou du graphe de la filiation.

Depuis vingt ans, de nombreuses méthodes informatiques sont apparues pour résoudre plus efficacement ce problème, inspirées principalement des méthodes phylogénétiques :

- le projet Collate de Robinson et O'Hara (Robinson, 2003) avec l'édition Canterbury Tales ;
- le corpus fictif sur le poème allemand Parzival de Wolfram von Eschenbach (cf. Figure 1). Il a été créé par Spencer et Howe (Spencer et al., 2004) ;
- l'édition de la Glose de Bénarès ou Kāshikāvṛitti (en Sanskrit) par Le Pouliquen (Le Pouliquen, 2011).

La stémmatologie dont le but est de reconstruire l'arbre généalogique des témoins a utilisé des modélisations différentes comme les arbres (souvent phylogéniques), les graphes, les treillis et même les cardiogrammes. Nous allons proposer une nouvelle voie basée sur l'utilisation des arbres et qui nous a été inspirée par les constats suivants :

- L'éditeur est généralement un littéraire et non un scientifique. La méthode de construction du stemma codicum, si elle est basée sur des calculs un tant soit peu compliqués, est généralement vide de sens pour lui, et lui paraît par conséquent artificielle et peu interprétable.
- Il dispose en revanche de connaissances pas toujours facilement modélisables comme les études codicologiques<sup>2</sup> et paléographiques<sup>3</sup> qui lui apportent des informations supplémentaires lui permettant de valider ou d'infirmer certaines paternités qui doivent être conservées.

Dans la réalisation manuelle du stemma, l'éditeur se concentre en général sur un ensemble réduit de manuscrits ou de variantes et réalise alors une sous-branche du stemma global que nous nommerons stemma codicum local qui consiste en un petit arbre. Notre méthode permet alors de rassembler ces arbres locaux, pour obtenir un arbre global représentant le stemma codicum en utilisant une modélisation philologique des manuscrits intermédiaires proposée par Don Quentin (Quentin, 1926).

---

<sup>1</sup> La contamination est la copie d'un manuscrit sur plusieurs modèles. La reconstruction généalogique suppose que chaque copiste n'a utilisé qu'un modèle pour réaliser son exemplaire (filiation unique). La contamination admet la possibilité qu'un copiste ait employé plus d'un exemplaire pour obtenir son manuscrit. (Syn. : corruption, hybridation, transmission horizontale).

<sup>2</sup> La codicologie étudie le manuscrit comme objet matériel afin de mieux comprendre l'histoire du texte (ou des textes) qui est parvenu jusqu'à nous. C'est ainsi qu'elle étudie les techniques de fabrication et les divers accidents qui ont pu affecter les ouvrages comme l'interversion de cahiers, la numérotation des folios, les accidents destructifs, les types de support...

<sup>3</sup> La paléographie a pour objet les écritures anciennes, leur déchiffrement, leur datation...

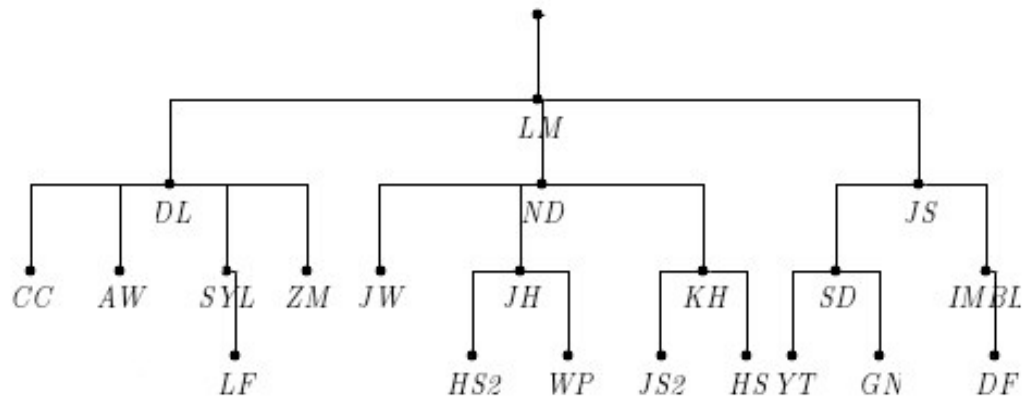


Figure 1. Exemple du stemma codicum de Parzival. Dans ce corpus artificiel, les manuscrits sont repérés par les initiales de leurs scribes. Le premier texte LM est copié directement de l'édition imprimée

## 2. De Don Quentin à une modélisation mathématique

Nous devons à Don Quentin l'idée d'utiliser la notion d'intermédiarité afin de dresser la généalogie. En effet, il suggère 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 comme nous l'avons déjà abordé (Le Pouliquen et al., 2010).

Ici, la problématique est différente puisque nous disposons d'arbres locaux que nous voulons rassembler en conservant le maximum de Triplets de trois Manuscrits parmi lesquels l'un est intermédiaire des deux autres et que nous nommerons T3M.

La modélisation philologique utilisée par Don Quentin pour la recherche des intermédiaires impose deux conditions pour que  $B$  soit intermédiaire entre  $A$  et  $C$  :

- (i)  $A$  et  $C$  s'accordent tour à tour avec  $B$  sur les variantes ;
- (ii)  $A$  et  $C$  ne s'accordent jamais contre  $B$ .

**Exemple 1 :** Soient les 3 phrases suivantes correspondant aux trois mêmes phrases de différents manuscrits copiés les uns sur les autres.

$A$ =« Maître Renard tenait dans son bec un fromage »

$B$ =« Renard tenait dans son bec le fromage »

$C$ =« Renard portait dans son bec le fromage »

Nous avons ici 3 lieux variants : **Maître**, **tenait/portait** et **un/le**, résumés dans le tableau suivant :

N° du lieu variant	Variantes	Phrases associées	Variantes	Phrases associées
1	Maître	$A$	$\{\emptyset\}$	$B, C$
2	tenait	$A, B$	portait	$C$
3	Un	$A$	le	$B, C$

Tableau 1. Table de collation des trois phrases

Pour les lieux variants 1 et 3, nous voyons que  $B$  s'accorde avec  $C$  ; pour le lieu variant 2,  $B$  s'accorde avec  $A$  donc la condition (i) est remplie. La condition (ii) est elle aussi validée car  $A$  et  $C$  ne s'accordent jamais contre  $B$ .  $B$  est donc bien intermédiaire entre  $A$  et  $C$ . Ces deux conditions correspondent bien à la notion d'intermédiaire au sens de la copie qui, ici, peut être vu comme : le copiste de  $B$  a omis `Maître` et a modifié `un` en `le` et le copiste de  $C$  a remplacé `tenait` par `portait`.

Il est en revanche peu probable que  $C$  soit l'intermédiaire, car le copiste de  $C$  a supprimé `tenait` qui serait réintroduit par le copiste suivant. Dans ce cas,  $A$  et  $B$  s'accordent sur la variante `tenait` contre  $C$  qui contient lui, la variante `portait`, donc le (ii) n'est pas valide et  $C$  n'est pas considéré comme intermédiaire entre  $A$  et  $B$ .

**Notation 1** : Pour exprimer le fait que  $B$  est intermédiaire entre  $A$  et  $C$ , on écrira le triplet T3M sous la forme  $(A, B, C)$ .

Pour définir l'intermédiarité au niveau des textes, nous pouvons utiliser quatre caractérisations mathématiques différentes :

- Celle de la géométrie élaborée à partir d'une distance par Menger (Menger, 1928) sous le nom de relation métrique d'intermédiarité ;
- Celle construite naturellement à partir d'une relation d'ordre ;
- Celle introduite par Restle (Restle, 1959) qui définit la notion d'intermédiarité au niveau des ensembles ;
- Celle obtenue en construisant un indice d'intermédiarité avec les conditions de Don Quentin.

Des quatre caractérisations, nous utiliserons seulement les deux dernières car elles semblent plus interprétables par l'éditeur de part leur proximité avec les conditions de Don Quentin.

### 2.1. Intermédiarité en géométrie

Nous allons nous intéresser à la relation d'intermédiarité introduite par Restle. Cette relation définit la notion d'intermédiarité au niveau des ensembles. Cette modélisation peut parfaitement convenir, car en assimilant chaque manuscrit à un ensemble de variantes, la relation est assez proche de celle définie par les conditions de Don Quentin, donc d'une modélisation philologique

**Définition 1** : Soient trois ensembles  $A$ ,  $B$  et  $C$ . On considère que  $B$  est intermédiaire entre  $A$  et  $C$  si et seulement si :

$$A \cap \bar{B} \cap C = \emptyset$$

$$\bar{A} \cap B \cap \bar{C} = \emptyset$$

Finalement, la modélisation peut se schématiser par la Figure 2.

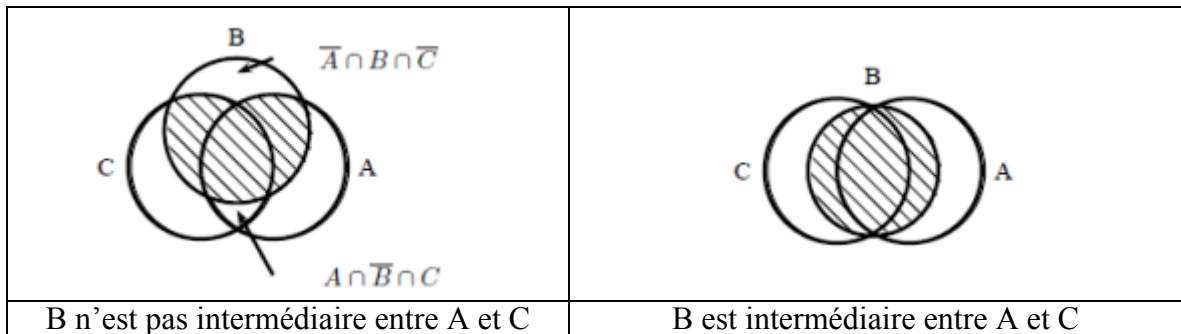


Figure 2. Intermédierité au sens de Restle

Reprenons notre exemple précédent. L'ensemble  $A$  est constitué des variantes {Maître, tenait, un},  $B$  de {tenait, le} et  $C$  de {portait, le}. Nous pouvons alors le modéliser par la Figure 3.

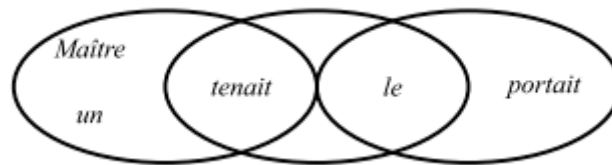


Figure 3. Visualisation sur l'exemple

**Notation 2 :**

- $n$  le nombre de lieux variants du corpus.
- $n_1 = \text{Card}(A \cap \bar{B} \cap C)$  le nombre de variantes communes à  $A$  et  $C$  qui n'appartiennent pas à  $B$ .
- $n_2 = \text{Card}(\bar{A} \cap B \cap \bar{C})$  le nombre de variantes de  $B$  qui n'appartiennent ni à  $A$ , ni à  $C$ .

Nous construisons alors l'indice suivant :  $I_E = \frac{n_1 + n_2}{\text{Card}(B)}$  qui est nul si  $(A, B, C)$ . Nous avons normalisé l'indice par le cardinal de  $B$  pour éviter d'être influencé par la taille des manuscrits

## 2.2. Indice d'intermédierité suivant Don Quentin

L'indice créé précédemment semble cohérent avec la notion d'intermédierité mais, il ne correspond pas exactement aux conditions (i) et (ii) établies par Don Quentin. Nous voulons construire ici un indice directement à partir de ces conditions. Il est nul en cas d'intermédierité parfaite et plus il est grand, plus le manuscrit examiné est éloigné d'un manuscrit parfaitement intermédiaire. Enfin, nous voulons pouvoir comparer les valeurs de l'indice sans être influencé par la taille des manuscrits.

**Notation 3 :**

- Soit  $I_{Q1} = \frac{n_1}{n}$   $I_{Q1} \in [0,1]$  et si  $I_{Q1} = 0$ , la condition (i) de Don Quentin est satisfaite.
- Soit  $I_{Q2} = \frac{n_2}{n}$   $I_{Q2} \in [0,1]$  et si  $I_{Q2} = 0$ , la condition (ii) de Don Quentin est satisfaite.

- En prenant,  $I_Q = \frac{1}{2}I_{Q1} + \frac{1}{2}I_{Q2}$  l'indice est nul si les deux conditions de Don Quentin sont vérifiées.

### Exemple 1 :

En reprenant l'exemple précédent et si l'on étudie l'intermédiarité de  $B$  entre  $A$  et  $C$ , nous obtenons  $n = 3$ ,  $n_1 = 0$ ,  $n_2 = 0$ ,  $I_{Q1} = I_{Q2} = I_Q = 0$  Donc  $B$  est intermédiaire entre  $A$  et  $C$ . En revanche, si l'on examine l'intermédiarité de  $C$  entre  $A$  et  $B$ , alors  $n = 3$ ,  $n_1 = 1$ ,  $n_2 = 1$ ,  $I_{Q1} = I_{Q2} = I_Q = \frac{1}{2}$  et donc  $C$  n'est pas intermédiaire entre  $A$  et  $B$ .

### 2.3. Bilan

Nous avons donc construit deux indices  $I_E$  et  $I_Q$  qui nous permettent de savoir si un triplet de manuscrits  $(A, B, C)$  peut être considéré comme un triplet T3M, c'est-à-dire que  $B$  est intermédiaire entre  $A$  et  $C$ . Ces deux indices ne sont pas identiques puisqu'ils utilisent des modélisations différentes, cependant, elles sont compréhensibles par l'éditeur qui peut facilement exploiter les informations philologiques qu'ils transmettent. Il appartient donc à l'éditeur de choisir l'indice qui lui semble le plus pertinent.

Nous disposons maintenant de deux sources de triplets T3M : les triplets obtenus à partir des arbres locaux proposés par l'éditeur nommés T3M1 et ceux qui annulent l'un des indices construits précédemment appelés T3M2.

Si les ensembles de triplets T3M1 et T3M2 ne coïncident pas, nous pouvons alors alerter l'éditeur sur d'éventuelles erreurs. Les indices étant construits sur des notions philologiques, l'éditeur peut facilement vérifier les lieux variants qui semblent discordants avec sa construction.

L'expérience montre néanmoins que le nombre de triplets T3M2 obtenus par un indice nul est très faible. Dans des données réelles, la contamination, les erreurs de saisie, etc., empêchent de respecter strictement les conditions de Don Quentin pour chaque lieu variant. Nous allons donc être amenés à relâcher ces conditions pour créer des triplets T3M2-souples où l'on impose à l'indice, non plus d'être nul, mais d'être inférieur à un seuil. Ce seuil est déterminé par l'éditeur en fonction du corpus.

Nous allons maintenant examiner comment rassembler les arbres locaux, pour obtenir un arbre global représentant le stemma codicum en utilisant les triplets T3M1, T3M2 et T3M2-souples.

## 3. Du consensus à l'algorithme

Pour la représentation du stemma, nous utilisons la notion d'arbre ainsi que toutes les notations définies par Barthélémy et Guénoche (Barthélémy et al., 1988).

### 3.1. Consensus

En phylogénétique, lorsque l'on dispose en entrée d'un ensemble d'arbres obtenus à partir de gènes différents (des variantes différentes dans le cas de manuscrits), l'arbre qui résume le mieux cette collection est obtenu par des méthodes de consensus ou de super-arbre.

Avec les méthodes de consensus comme celle de (Adams, 1972), chaque arbre de la collection contient le même ensemble d'espèces ou de manuscrits ce qui n'est pas notre cas.

Les méthodes de super-arbre comme la méthode *MRP* pour « *Matrix Representation using Parsimony* » de (Baum, 1992) et (Ragan, 1992) sont applicables sur nos arbres locaux mais elles sont, pour un éditeur, une construction artificielle qui ne se base sur aucun argument philologique.

De plus, si les arbres locaux sont compatibles, les méthodes de super-arbres produisent des résultats comparables. En revanche, quand les arbres locaux sont contradictoires, les solutions sont souvent divergentes et liées à la méthode choisie. Certaines de ces méthodes suppriment les conflits en choisissant d'enlever les arrêtes ou les feuilles impliquées. D'autres méthodes suppriment les conflits par une reconstruction de l'arbre global qui optimise des critères comme la parcimonie ou le maximum de vraisemblance. À nouveau, le comportement de ces méthodes reste peu compréhensible et peu interprétable pour un éditeur.

C'est donc pour permettre à l'éditeur de comprendre la construction du stemma global, que nous proposons d'utiliser une nouvelle méthode de super-arbre basée sur la notion d'intermédiarité de Don Quentin. Parmi les méthodes classiques, la méthode *Build* de (Aho et al., 1981) propose la reconstruction à partir de triplets enracinés. Les triplets enracinés n'étant pas facilement détectables en utilisant des manuscrits, nous allons utiliser les triplets T3M.

### 3.2. Propriétés

Pour que les différentes méthodes de reconstruction du superarbre s'effectuent de manière optimale, (Goloboff et al., 2002) ont établi un certain nombre de propriétés sur les triplets enracinés. Nous avons utilisé des propriétés similaires adaptées aux triplets T3M. Commençons par définir les propriétés simples de nos triplets T3M :

**Proposition 1** : Les triplets T3M vérifient les axiomes d'intermédiarité suivants :

- (i)  $(A, B, C)$  et  $(B, A, C) \Leftrightarrow A = B$  ;
- (ii) si  $(A, B, C)$  alors  $(C, B, A)$  (propriété de symétrie) ;
- (iii)  $(A, B, C)$  et  $(A, D, C) \Rightarrow (D, B, C)$  (propriété de transitivité) ;
- (iv)  $(A, B, C)$  et  $(A, D, B) \Rightarrow (A, D, C)$  (propriété de transitivité) ;
- (v)  $(A, B, C)$  et  $(B, C, D)$  et  $B \neq C \Rightarrow (A, B, D)$  (propriété de transitivité).

**Remarque** : les propriétés de transitivité sont définies pour quatre manuscrits. Dans le cas d'un nombre supérieur de manuscrits, il devient difficile d'identifier l'ensemble des propriétés comme le montrent (Pitcher et al., 1942).

La propriété de **couverture** impose que tous les manuscrits appartiennent au moins à un triplet T3M afin de pouvoir les situer sur l'arbre global.

**Proposition 2** : Soit  $E$  ensemble de triplets T3M sur  $C_0$ , un corpus de manuscrits, On dit que la propriété de **couverture** s'applique à  $C_0$  par rapport à  $E$  si (i) est satisfaite :

- (i)  $\forall A \in C_0 \exists B, C \in C_0 \Rightarrow (A, B, C)$  ou  $(B, C, A)$  ou  $(B, A, C)$

La propriété de **compatibilité** consiste pour un ensemble de triplets T3M à être représentable sur un arbre (Defays, 1979).

**Proposition 3** : Soit E ensemble de triplets T3M sur  $C_0$ . On dit que la propriété de **compatibilité** s'applique à E si (ii) est satisfaite :

$$(ii) \quad \forall A, B, C, D \in C_0 \quad (A, B, C) \Rightarrow (A, B, D) \text{ ou } (D, B, C)$$

La propriété de **densité** permet d'avoir suffisamment de triplets T3M pour construire l'arbre. Les critères nécessaires et suffisants pour vérifier cette propriété n'étant pas établis, nous nous contenterons de critères suffisants.

**Proposition 4** : Soit E ensemble de triplets T3M sur  $C_0$ . On dit que la propriété de **densité** s'applique à E si (iii) est satisfaite :

$$(iii) \quad \forall A, B \in C_0, \exists C \in C_0 \Rightarrow (A, B, C) \text{ ou } (B, C, A) \text{ ou } (B, A, C)$$

### 3.3. Algorithme et exemple

Nous décrivons maintenant notre algorithme de superarbre qui se sert des triplets M3P :

#### Algorithme 1 :

1. À partir des arbres locaux définis par l'éditeur, nous construisons l'ensemble des triplets T3M1.
2. Nous vérifions la **compatibilité** de l'ensemble des triplets T3M1. Nous établissons la liste des triplets incompatibles. Nous proposons alors à l'éditeur de choisir ceux qu'il préfère supprimer, sinon nous supprimons automatiquement ceux qui maximisent l'indice d'intermédiarité.
3. Nous vérifions la **couverture** du corpus par l'ensemble des triplets T3M1. Si tous les manuscrits ne sont pas couverts, nous pouvons alors rajouter des triplets T3M2 (qui annulent l'indice) qui sont compatibles avec les triplets précédents.
4. Nous rajoutons un à un les triplets T3M-souples (triés par ordre croissant des indices), qui sont compatibles avec les triplets déjà sélectionnés, jusqu'à assurer **couverture** et **densité**.
5. Nous reconstruisons ensuite l'arbre (ou l'arbre partiel<sup>4</sup>) de façon itérative en prenant successivement les manuscrits appartenant au maximum de triplets sélectionnés.

#### Remarques :

1. L'augmentation, par l'éditeur, du seuil d'acceptation des triplets T3M-souples permet de grossir le nombre de ces triplets et donc d'obtenir un arbre plutôt qu'un arbre partiel.
2. L'éditeur dispose alors d'un indice (la moyenne des indices de tous les triplets T3M de l'arbre global) permettant de quantifier l'adéquation de l'arbre avec les relations de Don Quentin. L'éditeur peut, s'il le souhaite, connaître les lieux variants en contradiction avec le stemma global proposé.

---

<sup>4</sup> L'arbre partiel n'est pas forcément un arbre, c'est un ensemble d'arêtes et de sommets provenant d'un arbre.



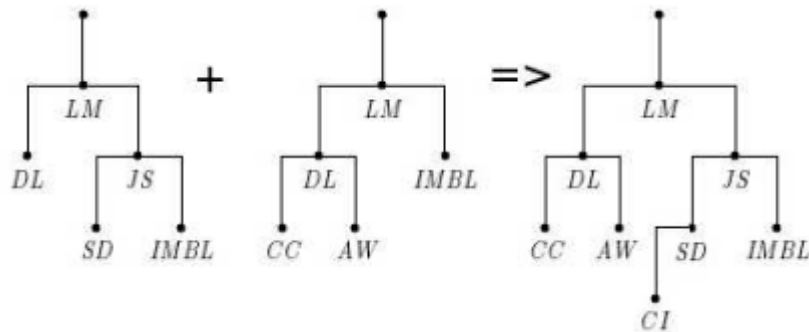


Figure 4. Des deux premiers stemmata ou arbres locaux, on en déduit le troisième

#### Exemple 4 :

Du premier arbre local de la figure 3, on en déduit les triplets T3M1 suivants :  $(DL, LM, JS)$ ,  $(DL, LM, SD)$ ,  $(DL, LM, IMBL)$ ,  $(LM, JS, SD)$ ,  $(LM, JS, IMBL)$ ,  $(SD, JS, IMBL)$  et tous les symétriques.

Du second arbre local, on en déduit les triplets T3M1 suivants :  $(CC, DL, LM)$ ,  $(CC, DL, AW)$ ,  $(CC, LM, IMBL)$ ,  $(AW, DL, LM)$ ,  $(AW, LM, IMBL)$ ,  $(DL, LM, IMBL)$  et tous les symétriques.

Il n'existe pas ici de triplets T3M1 incompatibles entre les triplets des deux arbres locaux.

Il existe en revanche un triplet commun :  $(DL, LM, IMBL)$ .

Le manuscrit *CI* n'est pas couvert par les deux arbres locaux. Cependant, à partir des calculs réalisés sur les manuscrits, on trouve un triplet T3M2-souple pour lequel l'indice  $I_Q$  est suffisamment faible. En effet, la première relation obtenue avec le manuscrit *CI* en les classant par indice croissant est  $(CI, SD, IMBL)$  pour laquelle on a :  $I_Q = 0,05$ . Elle permet alors de placer *CI* dans l'arbre global.

On peut signaler que l'exemple proposé est un cas idéal : aucun conflit entre les deux arbres locaux n'apparaît et la couverture et la densité sont pratiquement assurées avec les deux arbres locaux proposés par l'éditeur.

## 4. Conclusion

Les premiers tests de la méthode sont en cours sur le corpus Parzival (cf. Figure 1). C'est un corpus fictif mais qui a deux avantages, nous connaissons le stemma codicum et tous les manuscrits sont conservés. Nous pourrions alors analyser les résultats obtenus et regarder dans quelle mesure, l'algorithme peut être amélioré.

Il reste aussi un travail important dans la complexité des algorithmes afin d'accélérer les calculs, même si la taille des données dont on dispose en philologie n'est pas comparable avec celle de la phylogénétique.

Nous nous intéresserons enfin au problème de l'orientation des arbres locaux par les éditeurs, ainsi que les conséquences sur le stemma global.

## Références

- Adams E. N. (1972). Consensus Techniques and the Comparison of Taxonomic Trees. *Systematic Zoology*, 21(4), 390.
- Aho A. V. et al. (1981). Inferring a Tree from Lowest Common Ancestors with an Application to the Optimization of Relational Expressions. *SIAM Journal on Computing*, 10(3), 405-421.
- Barthélemy J.-P. et Guénoche A. (1988). *Les arbres et les représentations des proximités*. Méthode + programmes. Paris Milan Barcelone : Masson.
- Baum B. R. (1992). Combining Trees as a Way of Combining Data Sets for Phylogenetic Inference, and the Desirability of Combining Gene Trees. *Taxon*, 41(1), 3.
- Defays D. (1979). Tree representations of ternary relations. *Journal of Mathematical Psychology*, 19(2), 208-218.
- Goloboff P. A. et Pol D. (2002). Semi-strict supertrees. *Cladistics*, 18(5), 514-525.
- Lachmann K., Müllenhoff K. (Karl) et Vahlen J. (1876). *Kleinere Schriften von Karl Lachmann*. Berlin G. Reimer.
- Le Pouliquen M. (2011). Mathématiques et sciences humaines - Mathematics and social sciences. *Mathématiques et sciences humaines. Mathematics and social sciences*, (192), 57-91.
- Le Pouliquen M. et Csernel M. (2010). Stemma codicum et relation d'intermédiarité, utilisation de la méthode de Don Quentin. *Statistical Analysis of Textual Data Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles Lexicométrica*, 1, 309-320.
- Menger K. (1928). Untersuchungen über allgemeine Metrik. *Mathematische Annalen*, 100(1), 75-163.
- Pitcher E. et Smiley M. F. (1942). Transitivity of Betweenness. *Transactions of the American Mathematical Society*, 52(1), 95-114.
- Quentin H. (1926). *Essais de critique textuelle (ecdotique)*. Paris : A. Picard. Consulté à l'adresse <http://catalog.hathitrust.org/Record/001933116>.
- Ragan M. A. (1992). Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution*, 1(1), 53-58.
- Restle F. (1959). A metric and an ordering on sets. *Psychometrika*, 24(3), 207-220.
- Robinson P. (2003). The History, Discoveries, and Aims of the Canterbury Tales Project. *The Chaucer Review*, 38(2), 126-139.
- Spencer M. et al. (2004). Phylogenetics of artificial manuscripts. *Journal of Theoretical Biology*, 227(4), 503-511.