

Utilisation des pyramides pour visualiser la contamination des manuscrits

Marc Le Pouliquen*

*Université de Bretagne Occidentale, Labsticc UMR 3192, BP 832, 29285 Brest Cedex - France
marc.lepouliquen@univ-brest.fr,

1 Introduction

Dans le cadre de l'édition de manuscrits anciens, le rôle de l'éditeur consiste à reconstruire le plus fidèlement possible le manuscrit original à partir des différentes versions du texte disponible. Pour cela, l'éditeur classe les différentes versions du texte afin d'obtenir un arbre généalogique de cette filiation que l'on nomme *stemma codicum* (cf. Fig1).

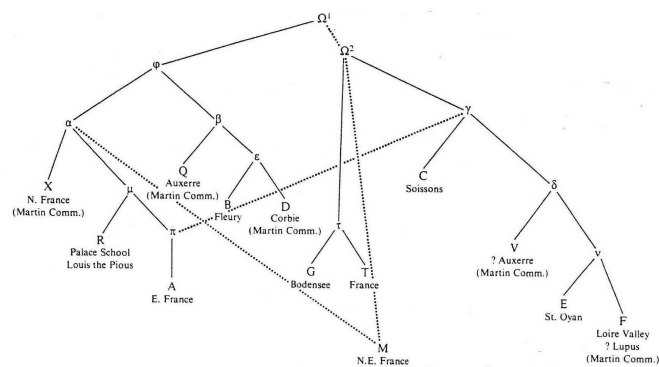


FIG. 1 – *Stemma de De Nuptiis Philologiae et Mercurii établi par Danuta Shanzer.*

La reconstruction généalogique par un arbre suppose que chaque copiste n'a utilisé qu'un seul manuscrit pour réaliser son exemplaire (filiation unique). Malheureusement, il arrive qu'un exemplaire ait été copié, non à partir d'une seule source, mais sur plusieurs manuscrits existants. On parle alors de contamination ou de corruption.

Nous proposons dans cet article, de représenter une tradition contaminée à l'aide une construction pyramidale basée sur la notion de manuscrits intermédiaires.

2 Méthodologie

Une des modélisations philologiques utilisées par les éditeurs pour reconstruire le stemma codicum est du à Don Quentin (1926). Elle permet de retrouver, à partir du corpus, des triplets de manuscrits dont l'un est l'intermédiaire des deux autres et que nous nommons T3M. Nous utilisons alors ces triplets T3M pour bâtir une pyramide généalogique.

Les triplets T3M sont déterminés par un indice qui est nul. L'expérience montre néanmoins que le nombre de triplets T3M obtenus est très faible. Dans des données réelles, la contamination, les erreurs de saisie, etc., empêchent de respecter strictement les conditions de Don Quentin. Nous allons donc être amené à relâcher les conditions strictes de Don Quentin pour créer des triplets T3M-souples où l'on impose à l'indice, non plus d'être nul, mais d'être inférieur à un seuil, seuil déterminé par l'éditeur en fonction du corpus.

Pour construire une pyramide à partir des triplets, définissons quatre propriétés :

- l'**intermédiarité** permet à un ensemble de triplets T3M de respecter la transitivité ;
- La **couverture** impose que tous les manuscrits appartiennent au moins à un triplet T3M afin de pouvoir les situer sur la pyramide finale.
- La **compatibilité** consiste pour un ensemble de triplets T3M à être représentable sur une pyramide (cf. Defays (1979) et Bertrand et Diday (1990))
- La **densité** permet d'avoir suffisamment de triplets T3M pour construire une pyramide. Avec ses propriétés, nous pouvons alors construire l'algorithme suivant :

Algorithme 1

1. Nous vérifions l'intermédiarité et la compatibilité de l'ensemble des triplets T3M et supprimons les triplets contradictoires
2. Nous vérifions la couverture et la densité du corpus par l'ensemble des triplets T3M. Si les propriétés ne sont pas assurées, nous rajoutons un à un les triplets T3M-souples (triés par ordre croissant des indices et qui vérifient 1) jusqu'à obtention des propriétés.
3. Nous reconstruisons ensuite la pyramide de façon itérative en prenant successivement les manuscrits appartenant au maximum de triplets sélectionnés.

Références

- Bertrand, P. et E. Diday (1990). Une généralisation des arbres hiérarchiques : les représentations pyramidales. *Rev. Stat. Appliquée*, 53–78.
- Defays, D. (1979). Tree representation of ternary relations. *Journal of Mathematical Psychology* 19, 208–218.
- Quentin, H. (1926). *Essais de critique textuelle*. Picard.

Summary

In this paper we present a new codicum stemma visualization method. Don Quentin's modeling is used to classify the textual tradition. We supplement the genealogical editor's information of betweenness triplets obtained directly from the corpus. A pyramid depicting the family codicum stemma is then constructed on the basis of information obtained by the triplets