

Aide à la construction du stemma codicum, Application à un corpus de « Chain Letters »

Marc Le Pouliquen¹

¹Telecom Bretagne, Labsticc UMR 3192, BP 832, 29285 Brest Cedex - France

marc.lepouliquen@telecom-bretagne.eu

Abstract

Automatic methods inspired mainly from the phylogenetic to classify a corpus of texts are often not used by critical editor because their results seem often artificial and few interpretable. We therefore propose a new method based on philological arguments that the editor will be able to understand and validate

This paper presents a stemma codicum's classification of a corpus of chain letters collected and studied by Bennett and his team. For this we propose two philological indices one of Lachmann and another of Quentin allowing to visualize globally and locally the corpus.

Résumé

Les méthodes automatiques pour classer un corpus de textes qui sont inspirées principalement de la phylogénétique sont peu utilisées par les éditeurs critiques car leur résultats semblent parfois artificiels et peu interprétables. Il faut donc revenir à des outils plus philologiques pour que l'éditeur puisse comprendre, interpréter et valider les résultats obtenus.

Ce papier présente une classification sous forme de stemma codicum d'un corpus de chaîne de lettres collecté et étudié par Bennett et son équipe. Pour cela nous proposons deux indices philologiques celui de Lachmann et celui de Quentin permettant de visualiser globalement et localement le corpus.

Mots-clés : Chain letters, édition critique, stemma codicum.

1. Introduction

Qui n'a pas le souvenir, avant l'arrivée des nouvelles technologies, d'avoir participé à une chaîne de lettres ? Pour ma part, j'avais reçu un courrier dont le message contenait une injonction de prendre part à une chaîne de lettres. Il suffisait d'envoyer ce message à dix de mes amis et de renvoyer de l'argent (une petite valeur) à trois personnes en amont dans la chaînes et dont le nom et l'adresse se situait dans le courrier. Si personne ne rompait la chaîne, je recevrais en retour une importante somme d'argent provenant de toutes les personnes situées en aval et liée à leur croissance exponentielle.

Ces chaînes de lettres ont des motivations diverses : argent, chance, religion, amour, et elles ont évoluées vers les courriels avec l'arrivée des nouvelles technologies. Dans cet article, nous nous intéresserons à un corpus de chaînes de lettres traditionnelles¹ récupérée par Bennett entre les années 1980 et 1995 et motivée par la chance. Dans son article Bennet (Bennett, Li, et Ma, 2003) propose de classifier ces lettres par un arbre censé représenter la filiation « au sens de la copie » entre les différentes lettres (cf. Figure 2). C'est un travail similaire à celui de l'éditeur critique qui trie les différentes versions d'un même manuscrits

¹ Le corpus de Bennett est disponible à l'adresse : <https://cs.uwaterloo.ca/~mli/chain.html>

(appelées témoins) en recherchant les différences (appelées variantes) afin d'établir une filiation entre eux. Par cette méthode due à Lachmann (Lachmann, Müllenhoff, et Vahlen, 1876), l'éditeur peut dresser un arbre généalogique de cette filiation que l'on nomme stemma codicum (cf. Figure 3). 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 ».

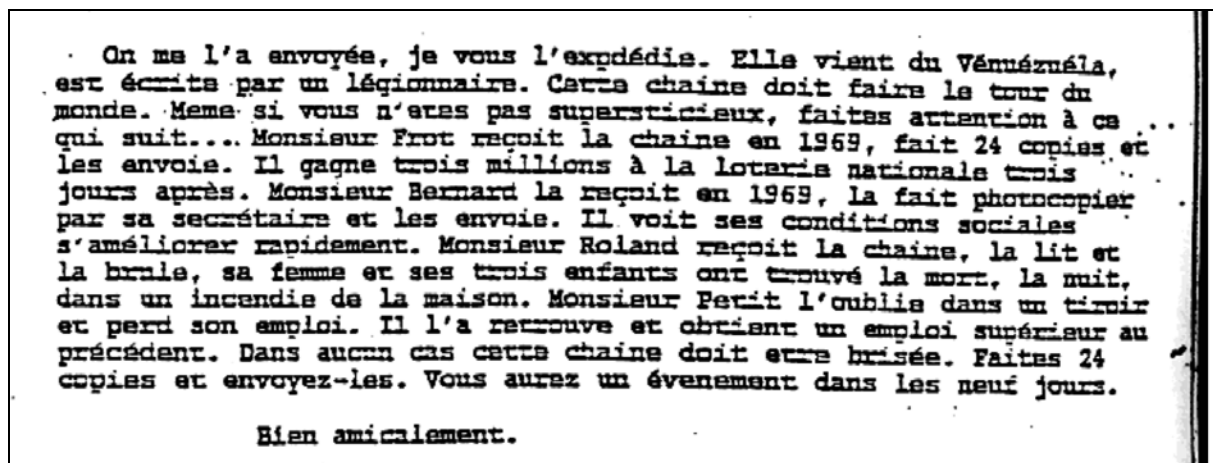


Figure 1 : Exemple de chaîne de lettres tirée de l'Hôtel de S. Cale²

Depuis vingt ans, de nombreuses méthodes informatiques sont apparues pour résoudre plus efficacement ce problème, inspirées principalement des méthodes phylogénétiques comme celle qu'a utilisée Bennett. Malheureusement, ces méthodes restent peu employées par les éditeurs critiques généralement littéraires non scientifiques et dont la méthode de construction du stemma codicum, si elle est basée sur des calculs un tant soit peu compliqués, est généralement artificielle et peu interprétable.

Dans cet article, nous proposons à l'éditeur de bâtir sa filiation de façon traditionnelle. Ce stemma codicum est alors confronté au corpus des textes (manuscrits ou ici les lettres) afin d'en tirer des incohérences philologiques issues des méthodes de Don Quentin (Quentin, 1926) et de Lachmann. Nous proposerons alors à l'éditeur d'observer les branches de sa filiation qui semblent les plus contradictoires, et ceci par une information philologique qu'il pourra alors utiliser.

2. De la règle philologique à la modélisation mathématique

Pour chacune des méthodes classiques d'établissement du stemma codicum, nous utilisons un exemple fictif de filiation simple.

2.1. Exemple fictif de filiation

Soit la tradition textuelle composée de 5 textes et représentée par le stemma de la Figure 3. Les cinq textes sont représentés par les lettres A, B, C, D et E. Pour simplifier, nous avons résumé nos textes à une seule phrase. « Voici une phrase inventée pour l'exemple »

² Lettres disponible à l'adresse <http://www.desordre.net/textes/bibliotheque/calle.html>

Elle propose 4 lieux variants (*Voici/Voilà/C'est*), (*phrase/phase*), (*inventée/créée/crée*) et (*exemple/modèle*) pour 10 variantes. Lors de la collation, l'éditeur obtient le tableau suivant :

N° du lieu variant	Variantes	Phrases associées	Variantes	Phrases associées	Variantes	Phrases associées
1	<i>Voici</i>	<i>A, C</i>	<i>Voilà</i>	<i>B</i>	<i>C'est</i>	<i>D, E</i>
2	<i>phrase</i>	<i>A, B, C, D</i>	<i>phase</i>	<i>E</i>		
3	<i>inventée</i>	<i>A, B</i>	<i>créée</i>	<i>C, E</i>	<i>créé</i>	<i>D</i>
4	<i>exemple</i>	<i>C, D, E</i>	<i>modèle</i>	<i>A, B</i>		

Tableau 1 : Table de collation des trois phrases

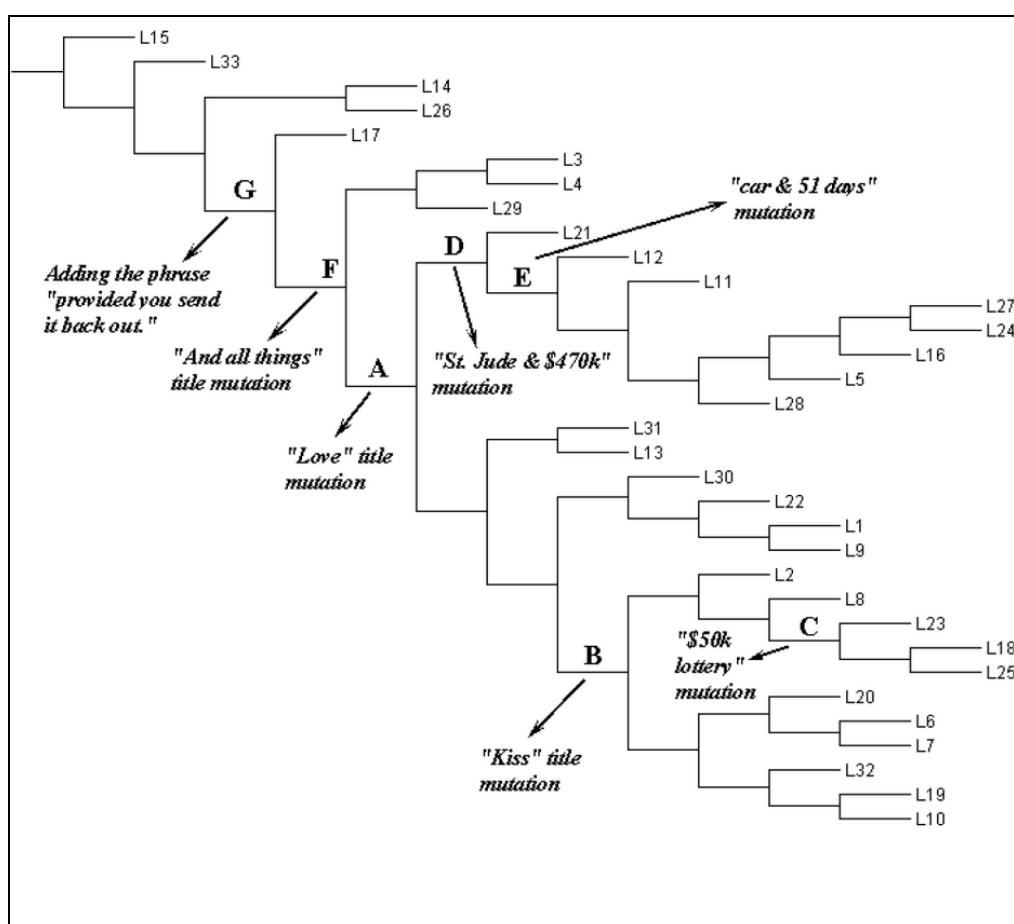


Figure 2 : Arbre ou stemma codicum établi par Bennett

2.2. Règle de Lachmann

2.2.1. Présentation

La méthode dite de Lachmann est simple et scientifiquement solide. Pour construire le stemma codicum à partir des textes, on part du constat suivant : 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.

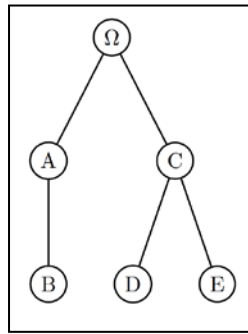


Figure 3 : Stemma codicum de l'exemple fictif

Pour classer les témoins, on recourt à la règle suivante : « la communauté des fautes implique la communauté d'origine » (cf. Froger, 1968)

Le problème reste pour l'éditeur à distinguer pour un lieu variant, la variante originale de la variante qui est une faute.

Pour notre modélisation, nous allons calculer pour chaque paire de textes un indice I_{Lach} indiquant le ratio du nombre de variantes différentes sur le nombre de lieux variants.

Soit $I_{Lach} = \frac{n_1}{n}$ avec $I_{Lach} \in [0,1]$ où :

- n le nombre de lieux variants du corpus ;
- n_1 le nombre de variantes qui diffèrent entre les deux textes.

En classant les paires de textes dans l'ordre croissant de l'indice, nous devrions légitimement retrouver les arêtes de notre stemma codicum dans les ratios les plus faibles et donc pourvoir alerter l'éditeur sur une branche suspecte.

2.2.2. Exemple :

En reprenant l'exemple fictif, le calcul des indices (en pourcentage) indique que les arêtes AB, AC, CD, CE qui sont les arêtes du stemma de la Figure 3 sont effectivement celles qui ont les plus faibles indices dans le Tableau 2. Le stemma codicum proposé semble donc cohérent au sens de Lachmann

Arêtes	AB	AC	CD	CF	DE	AD	BC	BD	AE	BE
Indices I_{Lach}	10%	20%	20%	20%	20%	30%	30%	30%	40%	40%

Tableau 2 : Indice inspiré par Lachmann

2.2.3. Remarque :

Cet indice constitue une indication, mais ne peut en aucun cas servir de preuve car beaucoup de biais sont à noter :

- lors de la copie, on peut tomber sur un scribe peu scrupuleux qui va enchaîner les fautes et donc augmenter l'indice I_{Lach} de façon anormale ;

- deux copistes peuvent réaliser la même erreur indépendamment et donc entraîner une impression de proximité entre deux textes ;

2.3. Règle de Don Quentin

2.3.1. Présentation

Une autre méthode historique est celle de Don Quentin, 1926. Elle s'attache à reconstituer l'enchaînement des textes au moyen de comparaison trois par trois. La modélisation philologique utilisée par Don Quentin pour la recherche des textes 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 .

Nous voulons construire ici deux indices directement à partir de ces conditions. Ils sont nuls en cas d'intermédiarité parfaite et plus ils sont grands, 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.

2.3.2. Notation

- 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 ;
- où $n_1 = \text{Card}(A \cap \overline{B} \cap C)$ le nombre de variantes communes à A et C qui n'appartiennent pas à B et $n_2 = \text{Card}(\overline{A} \cap B \cap \overline{C})$ le nombre de variantes de B qui n'appartiennent ni à A , ni à C ;
- pour exprimer le fait que B est intermédiaire entre A et C , on écrira le triplet sous la forme (A, B, C) . On pourrait éventuellement l'écrire sous la forme (C, B, A) car la relation d'intermédiarité est symétrique, mais on conservera l'ordre lexicographique classique.

2.3.3. Exemple

En reprenant l'exemple fictif précédent, le Tableau 3 résume les résultats obtenus en calculant les indices de Quentin (en pourcentage). Sur les trois triplets (A, B, C) , (B, A, C) et (A, C, B) , on ne conserve que celui qui minimise l'indice I_{q1} , ici (B, A, C) . On constate que seul le triplet (B, A, C) revendique une intermédiarité parfaite au sens de Quentin. L'expérience montre néanmoins que le nombre de triplets obtenus avec des indices nuls 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 de Quentin souples où l'on impose à l'indice, non plus d'être nul, mais d'être inférieur à un seuil (ici 25%). Ce seuil est déterminé par l'éditeur en fonction du corpus.

La méthode permet alors de retrouver les lettres intermédiaires, c'est-à-dire les lettres A et C . La lettre D semble intermédiaire entre A et E ce qui est en effet impossible à savoir si l'on

s'en tient aux données, c'est-à-dire aux variantes. On peut donc légitimement interroger l'éditeur sur son choix ?

Triplets	BAC	BAD	BAE	ACD	ACE	ADE	BDC	BCE	BDE	DCE
I_{q1}	0%	0%	0%	0%	0%	0%	0%	0%	0%	25%
I_{q2}	0%	25%	25%	25%	0%	25%	50%	25%	25%	25%

Tableau 3 Indices inspirés de Don Quentin

	Titre	La lettre est en	Quel officier reçoit des \$?	Combien reçoit-il ?	Qui perd sa vie ou sa femme ?	Auteur de la lettre	Qui perd son travail ?	Femme californienne ?	St Jude ?	Date
L1	T1	New England	RAF	\$70k	Gene Walsh, wife	Decroup	Carle Daddit	No	No	
L2	T2'	New England	officer	\$70k	Gene Welch, wife	de Croole	Carlo Dodds	No	No	9/23/85
L3	T3	Netherland	ex-officer	\$70k	General Welch, life	De Vold	Carlo Daddit	No	No	
L4	T3'	Netherlands	U.S.A.F.	\$70k	General Wales, life	De Caiod	Carlo Depot	No	No	1981
L5	T1	New England	R.A.F.	\$470k	Gene Welch, wife	de Group	Carlo Daddit	Yes	Yes	Summer 1989
L6	T2	New England	A.R.P. office	\$70k	Gene Welch, wife	DECROOF	Carla Daddit	No	No	1986
L7	T2	New England	A.R.F. office	\$70k	Gene Welch, wife	deCroof	Arla Daddit	No	No	
L8	T2	New England	R.A.F.	\$70k	Gene Welch, wife	deCroof	Arla Daddit	No	No	
L9	T1	New England	RAF	\$70k	Gene Welch, wife	De Croup	Carlo Daddit	No	No	1988
L10	T2'	New England	R.A.F.	\$70k	Gene Welch, wife	DeCroof	Anala Daddit	No	No	
L11		New England	R.A.F.	\$470	Gene Welch, wife	DeGroot	Carlo Daddit	Yes	Yes	1988
L12	T1'				Gene welch		Carlo Dadiot	Yes	Yes	Fall 1990
L13	T1	New England	RAF	\$70k	he, wife	de Group	Carlo Dabbit	No	No	
L14	Thank	Netherlands	U.S.	\$7k	Gen Walsh, life	de Cader	Charles Brer	No	No	
L15	T4'	Netherlands	RAF	\$70k	General Welch, life	de Cadif	Carlo Cradu	No	No	early 1980's
L16	T1'	New England	RAF	\$470k	George Welsch, wife	DeGrou	Carl Daddit	Yes	Yes	7/92
L17	T4	Netherlands	RAF(?)	\$70k	General Welch, life	Da Dadif	Carlo Raditt	No	No	
L18	T2	New England	R.A.F.	\$70k	Gene Welch, wife	de Croof	Aria Daddit	No	No	Jan. 1986
L19	T2	New England	R.A.F.	\$70k	Gene Welsh, wife	DeCroof	Aria Daddit	No	No	1986
L20	T2	New England	R.A.F.	\$70k	Gene Welch, wife	deCroof	Arla Daddit	No	No	
L21	T1	New England	R.A.F.	\$470k	Gene Welch, wife	DE GROUP	Carlo Daddit	No	Yes	
L22	T1'	New England	RAF	\$70k	Gene Walsh, wife	De Croup	Carlo Daddit	No	No	
L23	T2'	New England	Air Force	\$70k	Gene Welch, wife	Decroe	Andy Babbit	No	No	févr-92
L24	T1	New England	R.A.F.	\$470k	George Welch, wife	DeGrou	Carlo Daddit	Yes	Yes	10/91
L25	T2	New England	R.A.F.	\$70k	Gene Welch, wife	de Creff	Aria Daddit	No	No	4/1986
L26	T4'	Netherlands	R.A.	\$70k		De Codi	Carol Brand	No	No	
L27	T1	New England	R.A.F.	\$470k	George Welch, wife	DeGrou	Cario Daddit	Yes	Yes	
L28	T1	New England	R.A.P.	\$70k	Gen. Welch, wife	DECROUP	Carlo Daddit	Yes	Yes	
L29		Netherlands	R.A.F.	\$70k	General Welch, wife	Dowell	Carlo Davitt	No	No	
L30	T1	New England	R.A.F.	\$70k	Gene Welch, wife	DE GROUP	Carlo Daddit	No	No	
L31	T1	New England	an officer	\$40k	Gene Welch, wife	De Croup	Carlo Daddit	No	No	
L32	T2	New England	R.A.F.	\$70k	Gene Welch, wife	DECROOF	Arla Daddit	No	No	
L33	T4'	Netherlands	R.A.F.	\$70k	General Welsh, life	de Cadif	Corco Cradu	No	No	
T1=	With love all things are possible									
T2=	Kiss someone you love when you get this letter and make magic									
T2=	And all things whatever ye shall ask in prayer, believing, ye shall receive									
T4=	Trust in the Lord with all your heart and he will light the way									

Tableau 4 : Résumé des chaînes de lettres

2.4. Bilan

Nous avons donc construit deux indices I_{q1} et I_{q2} qui nous permettent de savoir si un triplet

de lettres (A, B, C) peut être considéré comme un triplet de Quentin, 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.

Nous disposons également de l'indice de Lachmann qui permet d'alerter l'éditeur sur une branche suspecte dans son stemma codicum. Si l'indice obtenu semble trop important, l'éditeur pourra alors confronter son choix aux données philologiques.

3 Application au corpus de « Chain Letters »

3.1. Présentation du corpus

Nous allons utiliser 33 versions de lettres d'une taille moyenne de 2000 caractères, recueillies en 1980 et 1995. Cette chaîne de lettres sur le thème de la chance s'est transmise par courrier ; les lettres sont arbitrairement étiquetées de L1 à L33. Il n'y a pas de deux lettres identiques, en fait, la plupart d'entre elles diffèrent de façon significative. Il y a 15 titres différents, 25 noms différents pour l'auteur. Faute de frappe, phrase ajoutée, manquante, ou échangée sont monnaie courante. Le Tableau 4 dresse la liste de certaines caractéristiques distinctives de celles-ci et la Figure 2 présente une modélisation graphique du stemma codicum par l'intermédiaire d'un dendrogramme réalisés par Bennett. Après un alignement automatique des lettres, nous avons obtenu une table de collation composée de 342 variantes.

3.2. Application de l'indice de Lachmann

Pour chaque paire de lettres, nous avons calculé l'indice de Lachmann puis ordonné cet indice de façon à obtenir une liste des paires de lettres plus ou moins « Lachmann-compatibles ». Les premières valeurs de ces calculs sont rassemblées dans le Tableau 5.

Lettres	Lettres	Distance de Bennett	Indice de Lachmann	Classement de la paire chez Bennett
L20	L32	0,165343	0.014	2
L18	L25	0,145988	0.020	1
L9	L22	0,259401	0.026	18
L8	L32	0,205799	0.026	4
L7	L32	0,249053	0.026	12
L9	L21	0,309512	0.026	33
L21	L30	0,363316	0.026	63
L10	L32	0,207373	0.029	5
L19	L32	0,20748	0.029	6
⋮	⋮	⋮	⋮	⋮
L9	L20	0,28648	0.076	13

Tableau 5 : Indice de Lachmann et distance de Bennett

Nous constatons une véritable corrélation (0.95) entre la distance utilisée par Bennett et l'indice de Lachmann. L'avantage de l'indice de Lachmann est qu'il est facilement compréhensible par l'éditeur qui peut éventuellement revoir sa conjoncture.

En effet, la ligne en jaune alerte l'éditeur sur un possible lien entre les lettres L21 et L30 qui n'était pas détecté par la distance utilisée par Bennett. Avec des arguments philologiques (similarité des lieux variants), l'éditeur sera en mesure de choisir d'agréger cette paire et éventuellement de détecter une contamination (Le Pouliquen et Barthelemy, 2009)

La ligne verte révèle à l'inverse une paire L9 L20 de lettres qui semblent proches (classée 13^{ème}) par la distance utilisée par Bennett et qui comporte beaucoup de variantes

incompatibles ? On constate d'ailleurs que sur l'arbre phylogénétique de la Figure 2, l'algorithme utilisé (Vincent Berry, 2000) n'a pas véritablement rapproché les deux lettres et donc probablement réévalué la distance.

Dans les 52 premières paires (12% d'erreur sur les variantes au maximum) on retrouve 41 paires classés par la distance utilisée par Bennett. 9 lettres ne sont pas représentées dans les 52 premières paires. Pour obtenir une représentation de toutes les lettres, il faut monter à 24% d'erreur sur les variantes et donc prendre en compte 186 paires.

3.3. Application des indices de Quentin

Pour chaque triplet de lettres, nous avons calculé les indices de Quentin puis ordonné ces triplets par rapport à $I_q = \frac{1}{2}(I_{q_1} + I_{q_2})$. Le corpus comporte 33 lettres, c'est-à-dire 5456 triplets ($33 * 32 * 31 / 6$) Nous obtenons alors une liste de triplets de lettres plus ou moins « Quentin-compatibles » dont les premières valeurs sont rassemblés dans le Tableau 6.

Triplets	Indice Iq
L7 L8 L29	0,0%
L7 L32 L16	0,0%
L7 L32 L17	0,0%
L7 L32 L29	0,0%
L7 L32 L30	0,0%
L8 L32 L11	0,0%
L8 L32 L16	0,0%
L8 L32 L28	0,0%
L8 L32 L30	0,0%
L1 L32 L7	0,5%
L1 L32 L8	0,5%
L4 L32 L7	0,5%
L5 L32 L7	0,5%
L5 L22 L9	0,5%
L6 L27 L8	0,5%
L6 L27 L20	0,5%
L6 L32 L8	0,5%
L6 L32 L20	0,5%
L6 L27 L24	0,5%

Tableau 6 : Indice de Quentin

Seule 2 lettres semblent être des intermédiaires stricts, L8 et L32. En examinant les relations obtenues pour les indices les plus faibles, les lettres L32, L22, L20 et L27 se retrouvent dans 222 des 233 triplets à moins de 1%. Ce ne sont pas uniquement des intermédiaires dans différents triplets, ils occupent peut être une position centrale dans le corpus. Nous pouvons alors nous interroger sur le rôle d'intermédiaire de ces lettres entre différentes sous-traditions textuelles du corpus. Cette hypothèse ne semble pas relevée par le stemma proposé par Bennett (cf Figure 2), nous allons donc bâtir un graphe à partir de nos indices pour tenter de vérifier cette hypothèse.

3.4. Représentation des données

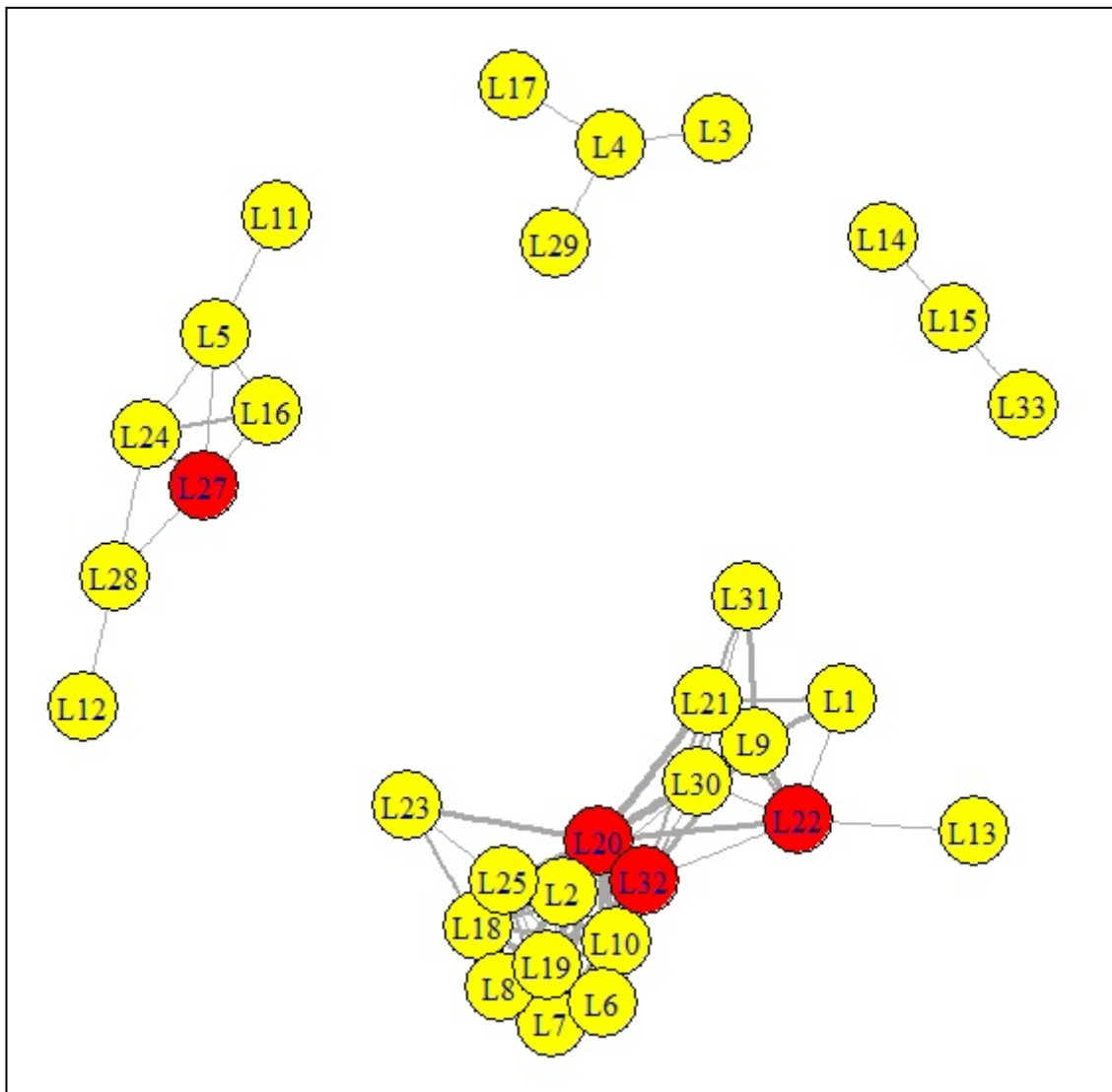


Figure 4 : Graphe ou stemma codicum établi par Le Pouliquen

Nous avons réalisé un graphe sous R (cf. Figure 4) à partir des arêtes obtenues avec les 52 premières paires (12% d'erreur sur les variantes au maximum) auxquelles on a rajouté les premières relations obtenues pour les 9 lettres qui n'étaient pas représentées. L'épaisseur des arêtes est inversement proportionnelle à l'indice obtenu, c'est-à-dire qu'elle indique la cohérence philologique de l'arête (pourcentage de variantes communes entre les deux lettres). Nous avons par ailleurs coloré en rouge les lettres recensées comme étant les plus intermédiaires et il semble clair que les lettres L20 et L32 sont de véritables intermédiaires (des ancêtres commun ?) entre deux sous-traditions déjà visibles sur l'arbre de Bennett et séparées par la lettre B. Pour les lettres L27 et L22, leur intermédianité est moins apparente.

Ce graphe distingue assez clairement 5 sous-traditions dans le corpus, dont les deux du haut (celle avec les lettres L17, L4, L3 et L29 et celle avec les lettres L14, L15 et L33) sont constituées des lettres en néerlandais ce qui semble prouver leur communauté d'origine.

4. Conclusion

Les méthodes automatiques inspirées principalement de la phylogénétique pour classer un corpus de textes sont souvent peu utilisées par les éditeurs critiques car leur résultats semblent souvent artificiels et peu interprétables. Il faut donc revenir à des outils plus philologiques pour que l'éditeur puisse comprendre, interpréter et valider les résultats obtenus.

Ce travail amorcé sur un corpus de chaîne de lettres complète celui réalisé par Bennett et son équipe, et présente deux indices philologiques celui de Lachmann et celui de Quentin permettant de visualiser globalement et localement le corpus.

Il existe d'autres corpus de chaîne de lettres sur lesquels il doit être intéressant de tester les indices proposés et l'on peut facilement obtenir une visualisation animée du graphe en augmentant ou en diminuant le nombre de paires ou de triplet pris en compte.

Enfin, il serait intéressant de collaborer avec un éditeur critique pour améliorer les outils et recueillir son ressenti.

Références

Bennett C., Li M., Ma B. « Linking Chain Letters ». *Sci. Am.* 2003.

Lachmann K., Müllenhoff K., Vahlen J. *Kleinere Schriften von Karl Lachmann*. [s.l.] : Berlin G. Reimer, 1876. 294 p.

Le Pouliquen M., Barthelemy J.-P. « Mathématiques et sciences humaines - Mathematics and social sciences ». *Mathématiques Sci. Hum. Math. Soc. Sci.*. 30 décembre 2009. n°187, p. 93- 105.

Quentin H. *Essais de critique textuelle (ecdotique)* [En ligne]. Paris : A. Picard, 1926. 3 p. 1., [9]-177, [2] p. p. Disponible sur : < <http://catalog.hathitrust.org/Record/001933116> > (consulté le 20 novembre 2013)

Vincent Berry D. B. « A Practical Algorithm for Recovering the Best Supported Edges of an Evolutionary Tree ». 2000. p. 287- 296.