

Tous des *hapax* ? Quantifier la variabilité dans une tradition manuscrite à l'aide de l'outil statistique ; retour épistémologique autour d'une expérimentation en humanités numériques sur la *Lex Saracenorum*

Antoine Tiendrebeogo

Référence de l'article

TIENDREBEOGO Antoine, « Tous des *hapax* ? Quantifier la variabilité dans une tradition manuscrite à l'aide de l'outil statistique ; retour épistémologique autour d'une expérimentation en humanités numériques sur la *Lex Saracenorum* », *Revue de l'Association des Jeunes Chercheurs de l'Ouest* [en ligne], n°1, 2021. URL : <https://ajco49.fr/2021/07/21/tous-des-hapax-quantifier-la-variabilite-dans-une-tradition-manuscrite-a-laide-de-loutil-statistique-retour-epistemologique-autour-dune-experimentation-en-humanites-nume>

Résumé

Le récent champ des humanités numériques offre aux chercheurs — et en particulier aux historiens — de nouveaux outils afin d'éprouver un regard analytique original sur leurs objets d'étude. Ce fut notre cas lors de notre travail de mémoire : une démarche éditoriale sur la *Lex Saracenorum*. Lors de ce travail, notre parti pris fut de mobiliser les statistiques afin de créer une grille de détermination des filiations. Cette innovation empirique fut concluante – autant que la mesure de notre échantillonnage puisse être sujette à induction. Cependant elle appelle des réflexions épistémologiques à la croisée de celles déjà présentes en humanités numériques mais également dans les sciences mobilisées (statistiques, biologie, histoire, philologie, etc.). C'est ce que nous nous proposons de mettre en lumière dans cet article. D'abord nous reviendrons sur le cadre particulier de la discipline philologique au sein de la famille des sciences auxiliaires de l'histoire. Puis nous expliciterons plus avant le cadre expérimental de nos travaux. Enfin nous prendrons de la hauteur en interrogeant le lieu des porosités épistémologiques qui apparaissent et méritent d'être interrogées ainsi que les critères de scientificité et les points de subjectivité qui ont marqué l'expérience.

Mots-clés : Stemmologie, Philologie médiévale, *Lex Saracenorum*, Statistiques appliquées, Epistémologie, Humanités numériques.

Abstract

All *hapax* bar none ? Quantifying variability in a medieval manuscript tradition using the statistical tool ; an epistemologic look-back about an experiment on the *Lex saracenorum* in digital humanities

Digital humanities as a youthful field of research, provides — to historians especially — new tools in order to have a brand new look on their research. This was our case when approaching our master thesis ; an editorial attempt at the first two surats of the *lex saracenorum*. During this workshop our approach of this work consisted of using statistical tools to bring out a brand new framework to determine filiation. This empirical innovation was a success – as much as we could induce based on our experimental sample. Nonetheless it raises several epistemological afterthought at the crossroads of some preexisting in digital humanities but some others coming from the aforementioned fields (statistics, biology, history, philology, etc.). This is what we aim to shed light on in this article. Firstly coming back on the particularity of the science of philology within the field of History's "auxiliary sciences", we will then explain further the experimental process of our master thesis experiment. Finally we will circle back to question potential transmissions or epistemological porosities that may arise and should be put forth or better answered along with the scientific standards and the subjectivity points that inevitably affected the experiment. Keywords: Stemmology, Medieval philology, *Lex Saracenorum*, Applied Statistical methods, Epistemology, Data-driven Humanities

Tous des *hapax* ? Quantifier la variabilité dans une tradition manuscrite à l'aide de l'outil statistique ; retour épistémologique autour d'une expérimentation en humanités numériques sur la *Lex Saracenorum*

Antoine Tiendrebeogo

Commençons en avant-propos par évoquer nos inspirations méthodologiques. La première vient des analyses granulométriques dans l'étude des mortiers dont nous avons été témoin au sein du Pôle archéologie de Maine-et-Loire – spécialisé dans l'archéologie du bâti médiéval. Leur méthode de reconstruction du processus d'un mur maçonné par l'étude des mortiers repose sur une analyse granulométrique des roches inspirée des pratiques géologiques¹. Elle se divise en étapes qui sont prioritairement le prélèvement, l'échantillonnage par réduction, le tamisage des mortiers et enfin une mesure présentant la quantification sous forme de graphique. Dans ce graphique, la taille des grains est disposée en abscisse et la proportion relative en ordonnée.

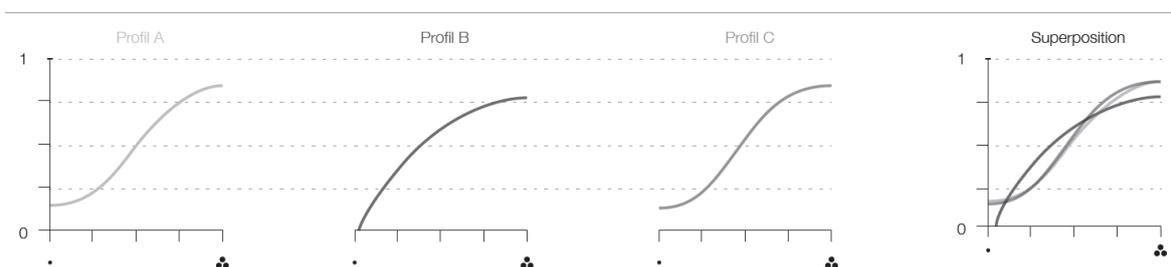


Fig. 1. Les profils granulométriques et la méthode de superposition. Ces courbes de profils A, B et C caractérisent la nature granulométrique de prélèvements de mortier. En les superposant, on se rend compte d'une analogie entre les mortiers issus des prélèvements A et C.

L'hypothèse qui soutient cette analyse est celle de la simultanéité de conception des mortiers dont les courbes possèdent des profils analogues car il est hautement improbable de produire – à échelle microscopique – le même mortier plusieurs fois. Cette mécanique d'association par la superposition de profils fut le premier jalon de notre réflexion philologique. La paternité de notre méthode reste nôtre. Cependant, une part du mérite en revient au Pôle archéologie de Maine-et-Loire – nous lui rendons donc ici hommage.

Cette hypothèse de la concordance des profils est-elle réinscriptible dans l'étude d'une tradition manuscrite ? Est-ce possible de construire des profils de témoins ? Si oui, sur

¹ Cette méthode est introduite dans la décennie 1980 au Pôle archéologie de Maine-et-Loire par Daniel Prigent, alors responsable de Pôle. Un exemple est présenté dans LITOUX Emmanuel et al., *Château d'Angers – Étude archéologique du logis royal et de ses abords*, Service archéologique de Maine-et-Loire, septembre 2013. URL : <https://rihvage.univ-tours.fr/wp-content/uploads/2014/09/Litoux-chateau-angers-logis-royal.pdf> (site Internet consulté le 05/05/2021).

quelles caractéristiques faut-il baser la construction de ces profils ? Comment dessiner cette représentation des différents témoins et comment en tirer quelque chose de pertinent d'un point de vue stemmatologique ?

La seconde inspiration vient du génome ou davantage des représentations vulgaires et scolaires de l'étude du génome. Les cours de secondaire présentent le phénomène des mutations comme des expressions multiples pour un même gène donné. L'analogie semble efficace et pouvoir se combiner à la précédente pour évoquer les variations textuelles d'une œuvre au cours de sa tradition, chaque texte étant une expression phénotypique d'une mutation spécifique de l'œuvre comme code génétique.

À partir de ces deux analogies instinctives qui semblent « découler d'elles-mêmes », il faut toutefois comprendre comment concilier des grilles d'analyse issues de disciplines et de champs si différents (biologie, génétique et géologie). Il nous faut donc démontrer comment s'aborde l'association – d'un point de vue épistémologique et méthodologique – de champs parfois en dissonance, qui n'embrassent pas les mêmes concepts et qui ne répondent donc pas aux mêmes logiques.

Au sein de la philologie, l'éditeur scientifique est tenu de revenir sur le parcours de l'œuvre afin de construire son édition critique sur une œuvre historicisée. Cette historicisation est nécessaire car il est rare de posséder le texte original d'une œuvre, celui inscrit par le primo-auteur. Les œuvres médiévales aujourd'hui éditées scientifiquement nous sont parvenues – presque intégralement – par des mécanismes de copie².

Dans ce processus toujours fautif se pose le questionnement de la qualité de l'héritage ou de l'inspiration. Chaque copie pose – par son existence – la question de la conservation du propos et du sens de son modèle. D'aucuns peuvent par leur « sens commun » donner une appréciation instinctive de la qualité d'une copie pour peu qu'on leur présente la source et la copie générée. C'est dans l'espace de cette appréciation que se pose notre questionnement.

Comment peut s'apprécier et se quantifier la conservation du propos et du sens dans une tradition philologique ? Par la « patte » et l'« œil » du philologue répond la discipline³. Face à cette prémisse, nous proposons une mathématisation, une rationalisation de cette appréciation et souhaitons même la développer en système de filiation. Ce système est à mettre en concordance avec les grilles analytiques déjà existantes en philologie. Pour cela, nous nous basons sur l'empirisme de notre expérimentation de mémoire universitaire et notre maîtrise récente de la discipline philologique.

Si les mathématiques peuvent ouvrir une dérivation à une partie limitée du protocole éditorial, quels avantages et écueils nouveaux créent-elles ? Est-ce une addition aux humanités numériques ? Un développement ou une mutation d'outils préexistants dans ce champ ? Enfin, quelles interrogations épistémologiques ces possibilités éveillent-elles ?

² Pour une synthèse, consulter la somme de l'École Nationale des Chartes en trois tomes : VIELLIARD Françoise, GUYOTJEANNIN Olivier (dir.), *Conseils pour l'édition des textes médiévaux*. Fasc. I : *Conseils généraux*, Paris, CTHS, 2001 ; GUYOTJEANNIN Olivier (dir.), *Conseils pour l'édition des textes médiévaux*. Fasc. II : *Actes et documents d'archives*, Paris, CTHS, 2001 ; BOURGAIN Pascale, VIELLIARD Françoise (dir.), *Conseils pour l'édition des textes médiévaux*. Fasc. 3 : *Textes littéraires*, Paris, CTHS, 2002.

³ Ce n'est plus tant le cas depuis la controverse bédieriste sur l'héritage disciplinaire de Karl Lachmann. Malgré tout, cette aptitude de l'éditeur est toujours présentée comme un prérequis. Cf. VIELLIARD Françoise, GUYOTJEANNIN Olivier (dir.), *op. cit.*, p. 9.

LA PHILOGIE, PARENT DE LA DISCIPLINE HISTORIQUE

L'œuvre dans le temps, un objet poreux

L'œuvre incarnée en texte est irrémédiablement marquée au fer du contexte de sa copie. Lorsque les copistes médiévaux entendent reprendre et recopier les textes grecs anciens à partir du XII^e siècle, le contexte social, linguistique et même dogmatique a varié⁴. À partir de l'entendement de cette diachronie, il faut considérer ce qui est conservé du texte d'origine. La volonté de préservation des copistes se porte-t-elle sur la conservation de formes syntaxiques ataviques ou les mettent-elles à jour ? Si le dernier cas se présente ; à partir de quel moment une adaptation d'un propos ancien à une *épistémè* contemporaine devient-elle une logique éditoriale ? Dans le cas où ce serait le fond du propos qui emporte la priorité dans la logique des copistes, à partir de quelle « quantité » de distorsion une adaptation du propos à un nouveau contexte devient-elle une réécriture voire une œuvre nouvelle ?

Constante d'intention et constante de formulation

Nous avons choisi arbitrairement de construire notre analyse sur la préservation de deux constantes⁵. Ces constantes, nous les avons nommées « constante d'intention » et « constante de formulation »⁶.

La première regroupe les mécaniques de copie qui se positionnent dans un angle traditionnel au sens où le copiste – non sans biais – cherche à préserver le caractère sémantique avant la nature syntaxique de l'œuvre qui est recopiée. Au sein d'une tradition manuscrite, les témoins qui font œuvre de « constance d'intention » sont souvent marqués par de grandes variations textuelles qui donnent à percevoir le contexte idéologique et historique dans lequel ils ont été produits⁷. Du point de vue de l'analyse statistique quantitative, ces témoins sont parmi ceux qui possèdent une plus grande variabilité entre eux sans pour autant être raisonnablement qualifiés d'*hapax*⁸.

Dans les manuscrits qui s'inscrivent dans la seconde mécanique de copie – soit celle de « constance de formulation » – la prétention à la tradition n'est pas absente mais elle s'exerce d'une toute autre manière. On trouve beaucoup de témoins de ce type lors des copies proches chronologiquement de la production primordiale de l'œuvre mais aussi – plus étonnement – lors d'époques plus distantes. On constate alors dans ces manuscrits une présence élevée d'atavismes syntaxiques et iconographiques. Ces atavismes témoignent d'une volonté du copiste – ou du centre de copie – de s'inscrire dans une fidélité mimétique

⁴ BOURGAIN Pascale, VIELLIARD Françoise (dir.), *op. cit.*

⁵ Une approche similaire de cette division de processus de copie peut être entendue dans les traditions de copie dites *tradizione queiscente* et *tradizione attiva* de la philologie italienne.

⁶ Il existe un concept similaire en statistique nommé ρ et qui est utilisé pour définir le coefficient de corrélation théorique (ou empirique) observé entre deux séries de données : cf. RAKOTOMALALA Ricco, « Analyse de corrélation. Étude des dépendances - Variables quantitatives », *Université Lumière Lyon 2*, 2017. URL : https://eric.univ-lyon2.fr/~ricco/cours/cours/Analyse_de_Correlation.pdf (site Internet consulté le 05/04/2021).

⁷ Considérons que toute évolution linguistique n'est pas forcément un produit du contexte historique et idéologique mais peut également relever d'un choix conscient et marqué du copiste.

⁸ *Hapax* est ici employé au sens où ce témoin produit une nouvelle branche de la tradition stemmatologique.

au témoin de référence, perçu comme étant de meilleure qualité en raison de son antériorité⁹. Les témoins de cette famille sont ceux dans lesquels on observe une grande constance du point de vue de l'analyse statistique quantitative. Cette constance se poursuit même généralement dans l'ordonnement des textes au sein des corpus.

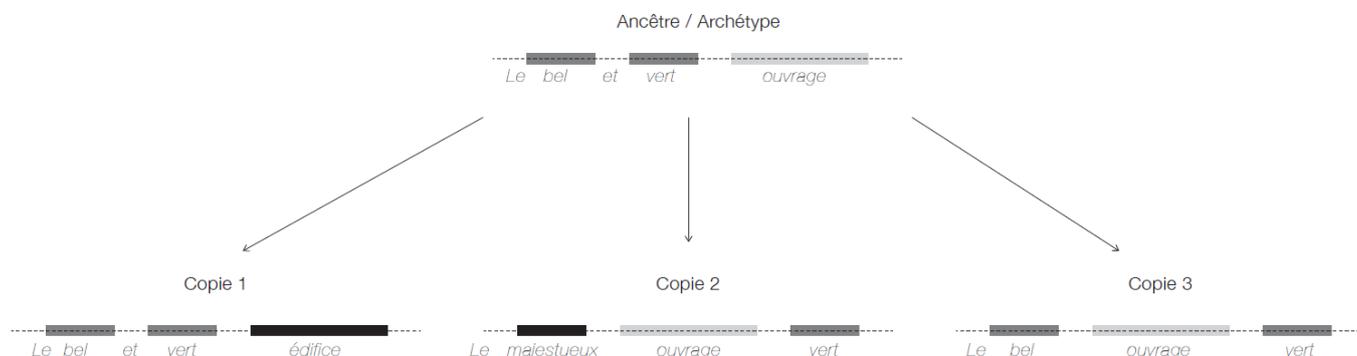


Fig. 2. Présentation des modalités de copie, constances d'intention et de formulation. Un archétype produit trois nouveaux témoins dont aucun n'est identique. La copie 1 est une copie d'intention car la forme globale est respectée mais pas nécessairement le détail de sa composition. La copie 3 a préservé le détail en perdant l'intégralité du sens général ; c'est une manifestation de constante de formulation. La copie 2 est un hybride.

Du croisement des grilles d'analyse

Comme suggéré plus haut, l'étude des œuvres ou de l'œuvre dans ses variations textuelles et des manuscrits qui les recueillent ne suffit pas. La tradition philologique le reconnaît d'ailleurs et cela nous impose donc de passer par l'étude matérielle des éléments de la tradition¹⁰. Ces grilles d'analyse complémentaires sont essentielles afin d'appréhender toutes les facettes d'une tradition manuscrite mais les analyses codicologique et historique le sont tout particulièrement.

L'analyse codicologique nous permet de détecter des intentions et les logistiques de copie qui peuvent expliquer une partie des variations syntaxiques ou sémantiques. Elle nous permet de comprendre non seulement le contexte et parfois les motivations de la recopie mais aussi l'intention qui a présidé à la recopie, cela même si cette dernière n'est que peu présente dans la forme nouvelle que prend l'œuvre au sein du texte produit.

Nous pensons ici au *codex* Cusain produit durant le concile de Bâle (1431–1437) lors de la querelle eugéniste. L'analyse codicologique de ce manuscrit nous révèle qu'il a vraisemblablement été produit en urgence¹¹.

⁹ Nous pensons ici en particulier à la tradition philologique italienne du *Quattrocento*. Cf. : TIENDREBEOGO Antoine, *Édition critique de la Summa et de l'Epistola de Translatione Sua, Œuvres de Petrus Venerabilis présentées dans le Corpus Toledanum*, Mémoire de master 1 [non publié] sous la direction de DESWARTE Thomas, Université d'Angers, 2018, p. 58.

¹⁰ VIELLIARD Françoise, GUYOTJEANNIN Olivier (dir.), *op. cit.*, p. 9.

¹¹ Analyse codicologique détaillée présentée dans TIENDREBEOGO Antoine, *Édition critique de la Lex Saracenorum, Première traduction du Coran en latin présentée dans le Corpus Toledanum et commandée par Petrus Venerabilis : Cas d'espèce ecdotique sur les sourates d'ouverture de l'Œuvre*, Mémoire de master 2 [non publié] sous la direction de DESWARTE Thomas, Université d'Angers, 2019, p. 94-98.

L'analyse codicologique nous indique que ce manuscrit a très probablement été produit par une méthode proto-industrielle au sein d'un *scriptorium* éphémère où la précipitation a prévalu à toute autre considération. Or il est crédible que ce témoin ait été celui de référence pour d'autres copies vaticanes. L'analyse

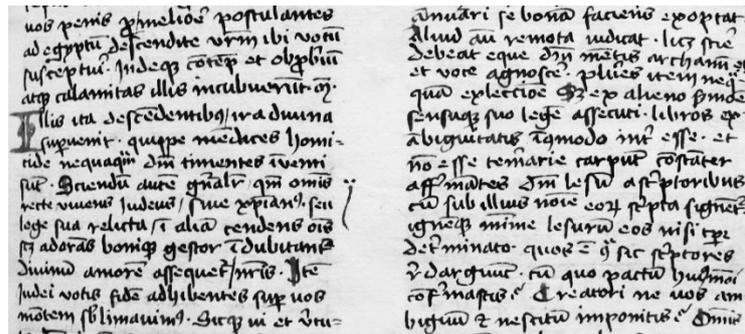


Fig. 3. L'écriture en gothique cursive du Liber Alcr Alchuman Lex Perdi Machumet

codicologique nous permet ici de contextualiser une distance relative (visible par l'analyse statistique mathématique) et de créer – via une mise en relation des indices – une sous-branche stemmatologique au sein de la tradition.

L'analyse historique est la seconde grille d'analyse à considérer par l'établissement de contextes idéologiques. Sans l'étude historique au sens large, il nous est possible de quantifier les variations mais impossible de les qualifier¹². Nous pensons ici en particulier aux éditions produites à la fin de l'époque médiévale dans les cercles protestants. Les copies et éditions produites dans ce contexte particulier et original sont marquées d'une évolution sémantique qui les distancie radicalement du reste de la tradition. Pour autant, sans la connaissance historique du contexte protestant et de l'influence éditoriale de théologiens de la foi protestante, il serait impossible d'appréhender qualitativement ces témoins.

CAS D'ECOLE : LA *LEX SARACENORUM*

La subjectivité linguistique

Comme suggéré ci-avant, le philologue doit appréhender la langue d'une œuvre au travers des différents textes où elle s'exprime afin de connaître une tradition et de pouvoir l'éditer scientifiquement. Pour cela, l'étude seule ne peut pas suffire et il lui faut l'aide des analyses historique et matérielle afin de réduire le plus possible les « angles morts » de son étude. Or dans la pratique philologique, l'accent est mis sur la maîtrise de la langue de l'œuvre étudiée, ainsi que sur les évolutions de l'utilisation de cette langue au sein de différents contextes culturels et historiques¹³.

Ce n'était pas notre cas, nous avons donc pris le parti de forger un outil de façon à pallier une lacune conséquente dans notre maîtrise de la langue – ici latine. L'historien comme scientifique de l'interprétation doit se prémunir d'un tropisme que nous pourrions qualifier comme étant celui de la « stratification des subjectivités ». Pour cela nous avons souhaité créer un outil mathématique de quantification des variabilités de l'œuvre au sein des différents textes.

¹² Sur les variations liées à des évolutions idéologiques, la seconde édition de la *Lex Saracenorum* de 1550 par Théodor Bibliander est un cas d'école particulièrement saisissant. Cf. l'imprimé conservé à la BNF, URL : <https://catalogue.bnf.fr/ark:/12148/cb33327703c> (site Internet consulté le 05/05/2021), ainsi que l'édition scientifique de VIGLIANO Tristan, LAMARQUE Henri, *Le Coran en latin et autres textes sur l'islam. La traduction latine du Coran, tel qu'éditée par Théodore Bibliander en 1550*, 2010, 258 p.

¹³ VIELLIARD Françoise, GUYOTJEANNIN Olivier (dir.), *op. cit.*, p. 9.

Des inspirations

Dans ce but, nous nous sommes inspirés d'autres méthodes de quantification au premier rang desquelles figure la lexicométrie. Cette première inspiration nous a porté à considérer la quantification comme palliatif à un ressenti philologique construit sur le seul instinct du locuteur. Ainsi, en voulant quantifier « combien » un texte mutait, nous avons remobilisé sans le savoir une vieille notion philologique, celle des profils scribaux¹⁴.

Dans notre production académique, nous avons nommé cela des profils de distorsion. Ces profils sont conçus par une étude statistique globale des différents textes entre eux en relevant chaque variante et en leur attribuant une pondération et une typologie¹⁵. À partir de ce relevé, nous avons obtenu une distance relative témoin à témoin.

En considérant chaque témoin comme un clone de son ou de ses modèles, et en filant la métaphore génomique, il est alors possible de décrire comment et à quel point l'œuvre mute à chaque génération de témoin.

Acquis expérimentaux sur la Lex Saracenorum

Au-delà de la réflexion de principe sur la pertinence de l'utilisation de la statistique pour observer une tradition manuscrite, il nous a fallu l'appliquer. Notre objet d'étude fut la *Lex Saracenorum* – première traduction du Coran en latin – produite par les réseaux clunisiens au milieu du XII^e siècle. Cette traduction est conservée avec d'autres œuvres au sein du codex primordial nommé *Corpus Toledanum*. Lors de notre étude, nous avons relevé plus d'une vingtaine de témoins entre 1142 et 1550, date de son édition par Théodor Bibliander¹⁶.

La bibliographie sur cette tradition est ténue et repose principalement sur les travaux de M.-T. d'Alverny et ses quelques successeurs¹⁷. Pour autant ces travaux ont été d'une grande aide pour identifier les différentes autorités commanditaires et responsables des copies produites durant les quatre siècles que recouvre la tradition.

Les études codicologiques antérieures ainsi que les nôtres nous ont permis de confirmer des hypothèses de filiations énoncées par l'historiographie. À partir de ces filiations manifestes – et confirmées par la grille d'analyse historique et les études codicologiques doublées – nous avons pu étalonner notre coefficient de distance.

¹⁴ MARICHAL Robert cité par CAMPS Jean-Baptiste, « Louis Havet, Cesare Segre, critique verbale et diasystème », *Sacré Gr@@l*, 2012, URL : https://graal.hypotheses.org/550#identifiant_8_550 (site Internet consulté le 02/03/2021) : « Plus profondément, par cet examen [des fautes et des aspects codicologiques] on peut espérer tracer, comme on dit aujourd'hui, le profil, plus ou moins caractérisé, bien sûr, d'un copiste, découvrir son propre code, son système, différent de celui de l'auteur et, par conséquent, décrypter ces diasystèmes dont M. Segre nous a montré l'importance et dont l'étude peut dépasser l'ecdotique pour aborder les secrets de la création ».

¹⁵ Ces variances doivent être ventilées sur un axe morpho-syntaxo-sémantique. Un détail de cette ambition est à trouver dans TIENDREBEOGO Antoine, *Édition critique de la Summa...op. cit.*, p. 283.

¹⁶ Ces témoins sont consignés dans notre mémoire universitaire : TIENDREBEOGO Antoine, *Édition critique de la Lex Saracenorum...op. cit.*, chap. II.

¹⁷ ALVERNY (d') Marie-Thérèse, « Deux traductions latines du Coran au Moyen Âge », *Archives d'histoire doctrinale et littéraire du Moyen Âge*, n° 22-23, 1947-1948, p. 69-131 ; BURMAN E. Thomas., *Reading the Qur'an in Latin Christendom 1140-1160*, University of Pennsylvania Press, 2007 (rééd.) ; IOGNA-PRAT Dominique, *Ordonner et exclure. Cluny et la société chrétienne face à l'hérésie, au judaïsme et à l'islam, 1000-1150*, Aubier, 1998 ; KRITZECK John, *Peter the Venerable and Islam*, Princeton University Press, 1964 ; VIGLIANO Tristan, *Parler aux musulmans : quatre intellectuels face à l'Islam à l'orée de la Renaissance, Les seuils de la modernité*, vol. 21., Genève, Droz, 2017.

Notre méthode statistique ne produit jamais d'indice de distance absolu, ce caractère de distanciation doit donc être apprécié subjectivement. De même, cette méthode d'observation, parce qu'elle détermine tout changement comme significatif et toute corrélation de variantes comme une préservation, doit s'accompagner d'une prudence de son utilisateur. Ce dernier doit se garder de produire un indice qui serait uniquement et totalement appuyé sur les corrélations, un indice qui ne tiendrait donc – par exemple – pas compte des phénomènes de polygénèse¹⁸. En effet, toute variance pérenne n'est pas forcément significative et une corrélation n'est pas nécessairement une concordance¹⁹. De ce point de vue purement quantitatif, la statistique offre des solutions par l'utilisation du Kappa de Fleiss et du Kappa de Cohen²⁰.

À l'aide de ces outils, nous avons pu déterminer la taille minimum d'un échantillon ainsi que la part indiscernable des corrélations au sein d'une étude des concordances. Face aux résultats obtenus et biais constatés nous avons considéré les résultats expérimentaux comme concluants. Cette méthode de quantification statistique et l'indice de distance relatif témoin à témoin fonctionnent et produisent de façon assez attendue un réseau maillé où la distance relative est proportionnelle à la corrélation et/ou concordance des variances témoin à témoin. Avec l'utilisation croisée des grilles d'analyse – historiques, iconographiques et codicologiques – et l'élimination des liens faibles, cette structure réticulaire en mailles régulières est disposée en structure hybride de façon à reconstruire la chronologie des

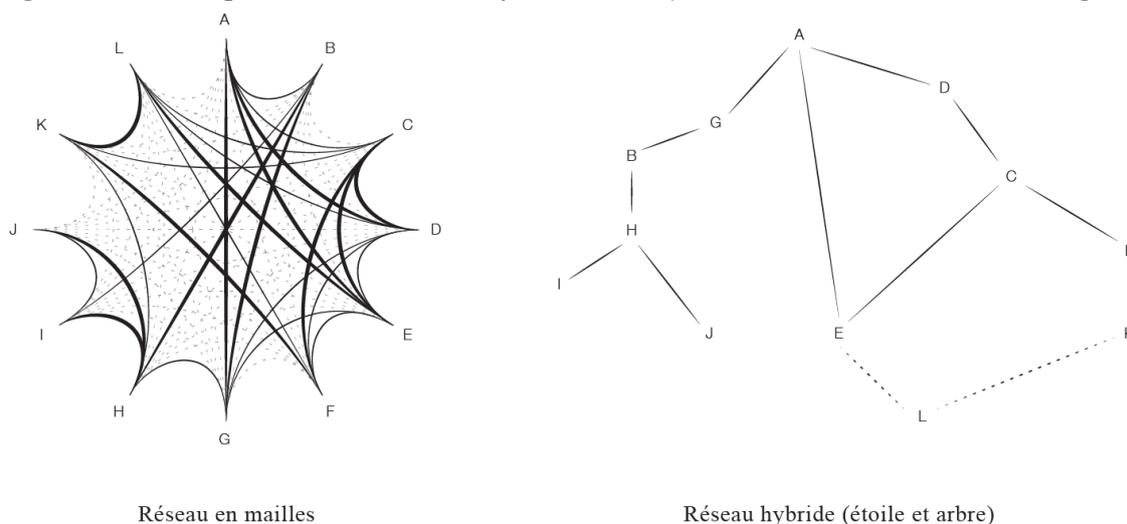


Fig. 4. Réseaux maillés et stemma.

¹⁸ Précisons ici que chaque variante doit être observée dans sa nature de façon à déterminer son éventuel caractère innové. Ainsi une branche stemmatologique ne naît pas tant par son caractère variant mais par l'innovation de cette variance détectée et sa perennité dans ladite branche.

¹⁹ Deux variables indépendantes sont non corrélées mais la réciproque n'est pas toujours vraie. Autrement dit, lorsque deux variables ou mesures sont corrélées alors elles sont dépendantes. La notion de concordance se relie à cette relation dans le sens où deux variables ou mesures concordantes sont corrélées et sont par conséquent dépendantes. Les concepts de concordance, de dépendance et de corrélation sont distincts même s'ils restent liés entre eux, leur relation et leur intrication peuvent être mises en équation ou figurées comme ci-après : [Concordance \in Corrélation \in Dépendance]. Cf. TIENDREBEOGO Antoine, *Édition critique de la Lex Saracenorum...op. cit.*, p. 159.

²⁰ Le Kappa de Fleiss et le Kappa de Cohen permettent via leur utilisation de déterminer la part des concordances qui ne peut se distinguer d'une corrélation hasardeuse dans la comparaison de séries de données. Le détail de leur intégration dans notre protocole est présent dans TIENDREBEOGO Antoine, *Édition critique de la Lex Saracenorum...op. cit.*, p. 151-166.

filiations. Le réseau s'ordonne alors et se recompose en un outil philogénétique connu : le stemma²¹.

Notre méthode – en mettant la variance au centre – produit à première vue un éclatement complet de la tradition manuscrite en ne créant plus que – à l'échelle du texte complet – des *hapax*, chaque témoin n'étant jamais un « clone parfait » de son géniteur. Prise seule, cette méthode écartèle la tradition. Pour autant la distance relative de cet éclatement – une fois associée avec d'autres grilles d'analyse – permet de pondérer la distance, non plus par l'écart chronologique mais par un indice nouveau de distance relative et quasi-objectivé.

Un produit inattendu mais bienvenu de cette méthode est de pouvoir détecter, par l'anomalie d'un indice sensiblement hors moyenne, des mailles fautives dans l'arbre philogénétique et donc de pouvoir estimer le nombre et l'emplacement de témoins manquants.

Les biais sont à considérer et prioritairement celui de l'unité lexicale. Une lecture quantitative du nombre de variances relevées texte à texte serait lourdement modulée par une considération différente de l'unité des lemmes pertinents. Pour autant, il n'est pas certain que la répercussion se perçoive sur les relations de proportionnalité entre les indices de distance produits.

L'appréciation subjective des pondérations à appliquer aux différents types de variances est également un biais qui n'est pas troublant pour peu qu'il soit protocolairement explicite et ouvert à modification²². Nous avons choisi de considérer les variances à coefficient égal mais une pondération précise et justifiée ne pourra être produite que par une construction empirique et par la multiplication des applications de notre méthode à différentes traditions littéraires (la jurisprudence, les collections canoniques, les *vitae*, les décrétales, les textes alchimiques et astronomiques, etc.).

Le point que nous n'avons pas pu apprécier est de savoir combien notre méthode perd de sa validité en ce qui concerne les fragments. Comment légitimement étendre des résultats produits sur un échantillon à un texte complet qui ne nous est pas parvenu²³? Cette question ne limite que peu le potentiel des expérimentations possibles pour éprouver l'usage de la statistique dans les processus philologiques. Cependant elle met en lumière le fait qu'il faut penser l'utilisation des statistiques – et plus largement de notre méthode – à hauteur épistémologique.

²¹ Sur la naissance et l'importance du stemma dans la philologie lachmannienne, voir TIMPANARO Sebastiano, « Critica testuale e linguistica, e crisi di entrambe nell'ultimo Ottocento e nel Novecento », *La Genesi del metodo del Lachmann*, Torino, UTET Libreria, 2003 (4e éd. – 1ère éd. : 1963), p. 99-110, via CAMPS Jean-Baptiste, « Copie, authenticité, originalité dans la philologie et son histoire », *Questes*, 29, 2015. URL : <http://journals.openedition.org/questes/3535> (site Internet consulté le 05/04/2021).

²² Un exemple de ce à quoi cela pourrait ressembler est présent dans CAMPS Jean-Baptiste, ING Lucence, SPADINI Elena, *Collating Medieval Vernacular Texts. Aligning Witnesses, Classifying Variants*, DH2019 Digital Humanities Conference 2019, juillet 2019, Utrecht, Netherlands. URL : <https://hal.archives-ouvertes.fr/hal-02268348> (site Internet consulté le 25/02/2021).

²³ Une possible inspiration serait à tirer de la loi de probabilité dite « Loi de Poisson » qui permet de décrire le nombre d'événements attendus dans un intervalle spatial à partir d'une mesure échantillonnée. Cf. POISSON Siméon-Denis, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile ; précédées des Règles générales du calcul des probabilités*, Paris, Bachelier, 1837, passage 81, p. 205.

DE L'ÉTAT DU DIALOGUE EPISTEMOLOGIQUE ENTRE LES STATISTIQUES ET L'HISTOIRE

Il nous paraît maintenant nécessaire de prendre de la hauteur par rapport à notre étude de cas et d'examiner les natures épistémologiques qui asseyent la légitimité des disciplines ici invoquées : la littérature, l'histoire et les mathématiques. S'il serait trop long de décliner les débats qui existent au sein de chaque discipline, on peut du moins dessiner une ligne de fracture entre les sciences dont l'exercice est prioritairement conceptuel et celles dont le caractère expérimental prévaut à leur constitution.

Entre itération spécifique et reproductibilité

Ne renforçons pas le tropisme qui veut discriminer les sciences « dures » des sciences « molles » et qui tient très peu le feu de l'analyse. Préférons-lui celui qui discrimine les sciences par le matériau sur lequel elles s'exercent. Les premières comme la biologie et les sciences naturelles – dans leur majorité – et l'informatique reposent sur la déduction expérimentale pour construire leurs axiomes. La mécanique inhérente à la construction de leur savoir articule l'hypothèse puis l'expérience et enfin la déduction et l'induction pour la construction de théorisations ou « lois ». La controverse dans le domaine, pour peu que l'expérience ait été faite « de bonne foi », ne peut qu'invalider la loi et pas les résultats de ladite expérience. Le caractère expérimental est donc fondamental.

À contrario sont les sciences que nous comprenons comme celles de l'itération spécifique. Dans ces familles se trouvent toutes les sciences humaines mais aussi les mathématiques. En effet le postulat de discrimination refuse la possibilité qu'il puisse exister deux événements strictement identiques. C'est évidemment le cas en histoire grâce à la conception chrétienne d'une temporalité linéaire, finie et non-cyclique. C'est aussi le cas en mathématiques où les nombres sont des contenants d'événements et non pas des événements eux-mêmes²⁴.

Les statistiques et la place de l'aléatoire

L'histoire refuse catégoriquement l'existence d'événements identiques, simplement par l'observation que l'un se situant nécessairement après l'autre, l'un posséderait *a minima* un caractère inédit non partagé par l'autre. Il existe des permanences (rares et relatives) mais pas de lois. C'est également le cas en mathématiques. Pour autant, elles ont conceptualisé une notion qui nous sera utile plus avant qui est celle de l'Aléatoire²⁵. Jamais réellement atteinte, cette notion a néanmoins été conceptuellement cernée.

En histoire, l'Aléatoire est considéré mais jamais démontrable car l'histoire est un raisonnement déductif limité aux sources auxquelles elle peut accéder et la nature même d'une source est d'avoir été le produit d'une intention. Si l'historien fait parfois parler

²⁴ Le nombre n'est pas – en mathématiques – une valeur intrinsèque mais davantage un concept qui s'applique à une quantité abstraite, concrète ou un rapport de grandeur. (ENCYCLOPAEDIA UNIVERSALIS, art. « Nombres »).

²⁵ Pour une description mathématique du « Hasard » et donc de l'« Aléatoire », voir ALLAIS Maurice, « Fréquence, probabilité et hasard », *Journal de la société française de statistique*, t. 124, n° 2, 1983, p. 70-102.

l'absence de sources dans tel ou tel domaine, la présence d'une absence ne peut se résoudre à l'évocation d'une action due à l'Aléatoire que sous le régime d'une hypothèse.

Il est certain que l'inconnu, la part de choix impossible à retracer ayant donné naissance à tel état de fait, peut être qualifié d'« aléatoire » au sens où il est insaisissable et que lui donner du sens relève de la téléologie. Pour autant cet « inconnu » historique ne recouvre que très partiellement les contours de ce qu'est l'Aléatoire en mathématiques. Il ne nous est pas possible en l'état d'aller beaucoup plus loin que la constatation de cette parenté conceptuelle. L'application computationnelle de la statistique à la philologie entraîne nécessairement un transfert de certains concepts d'un champ épistémologique à un autre. S'il est impossible de considérer notre méthode sans constater que les concepts de l'Aléatoire et de la Corrélation traversent la frontière entre les disciplines par porosité, la question demeure : comment est-il possible de penser scientifiquement le hasard en histoire ?

L'histoire comme discipline de l'interprétation

La discipline historique est donc résumable de façon abrupte à une science de l'interprétation de phénomènes inédits. Les répétitions strictes n'y ont aucune place, au mieux des analogies ou des similarités peuvent y exister. L'historien étudie donc les phénomènes en amont de son objet d'étude tout en se gardant d'étendre une adjonction de sens à des événements en aval ou en amont de son objet.

Le cas de la philologie est original à cet enseigne car – lors de l'étude d'une tradition – le présumé est celui de la conservation d'une œuvre au travers de multiples occurrences que nous nommons textes. En effet, lors de la copie d'une œuvre ou de sa recopie, l'intention présumée est de conserver l'œuvre, de lui permettre de continuer à exister alors même que son *medium* primordial arrive – le plus souvent – en fin de vie. Les scribes recopient pour diffuser et conserver une œuvre. Pour autant, quel que soit l'écart qui sépare une primorédaction d'une œuvre de sa copie, plusieurs éléments sont toujours variés : le premier est l'auteur ; le second l'*épistémé* de l'époque.

Là réside le rôle du philologue dans sa démarche éditoriale, celui de pouvoir considérer toute la filiation d'une œuvre afin d'en offrir à l'utilisateur une version qui puisse lui laisser entrevoir toute la variété des textes qui l'ont portée. En ce sens, l'étude des mécaniques de copie et de recopie est un présumé scientifique indispensable à la diffusion et à l'édition d'une œuvre. Celui qui emploierait notre méthode devrait considérer comment son utilisation viendrait affecter son positionnement et la qualité de sa critique éditoriale. Il devrait également savoir comment l'insertion des concepts nouveaux comme la Corrélation ou la part aléatoire des phénotypes textuels similaires doivent être présentés en appareil éditorial²⁶.

CONCLUSION

En joignant les mathématiques à notre étude historique, nous avons pu apprécier que son emploi ne protège pas d'un biais de subjectivité, notre expérimentation le prouve. Elles nous ont également permis de nous confronter aux concepts de Corrélation et donc de Hasard qui n'existent pas dans la discipline historique, ou plus précisément qui sont des impensés

²⁶ Un addendum doit donc dans ce cas être ajouté à l'apparat qui est un élément indispensable de l'édition critique comme explicité dans les ouvrages suivants : VIELLIARD Françoise, GUYOTJEANNIN Olivier (dir.), *op. cit.* ; GUYOTJEANNIN Olivier (dir.), *op. cit.* ; BOURGAIN Pascale, VIELLIARD Françoise (dir.), *op. cit.*

épistémologiques. En amenant les statistiques au sein des méthodes d'analyse philogénétique des manuscrits médiévaux, de nouvelles voies s'ouvrent.

La première est celle d'une quantification et d'une qualification mathématiques des variabilités de la copie ; la seconde est une possibilité de détection voire de quantification des concordances scripturales et, plus encore, une possibilité d'une mise en algorithmes des profils scribaux, du « bruit auctoral²⁷ ». Nous entendons ces profils comme la translation du concept mathématique de delta de variation²⁸. Il faudra toutefois se prémunir de considérer ces deltas comme définitifs. En effet, les copistes évoluent dans leur pratique, leur érudition et leur *scriptorium* d'implantation²⁹.

Cette question nouvelle des profils de distorsion, ou davantage des profils scribaux, pourrait permettre d'interroger les permanences de distorsion de tel ou tel scriptorium dans un contexte historique donné et même in fine de proposer des avatars de témoins perdus. Ces avatars pourraient être reconstitués à partir d'un témoin parent (en aval ou en amont) à la condition que les sources nous informent du *scriptorium* de production.

Enfin, le dernier avantage de la méthode que nous proposons est de permettre à l'historien novice d'étudier une traduction manuscrite ou toute autre mécanisme historique de copie quelle que soit la langue de production. Étonnement, c'est parce que notre approche statistique se détache de l'exigence de maîtrise linguistique qu'elle permet d'ouvrir le champ de la philologie aux novices dans la pratique d'une langue, voire dans la discipline.

Somme toute, nos travaux s'inscrivent dans le champ récent des humanités numériques³⁰. En ceci, ils défrichent de nouveaux territoires épistémologiques. Dans le cas de la philologie, la « patte » du linguiste a longtemps été une compétence surdéterminante dans les entreprises philologiques – en restreignant donc l'accès aux non-érudits. Nous avons – selon nous – commencé à prouver ici que des dérivations au protocole stemmatologique et éditorial étaient vraisemblables par l'utilisation de la statistique.

En parcourant succinctement cette voie, la silhouette d'un continent nouveau pour les disciplines historiques et philologiques s'est dessinée. Il appartient maintenant aux découvreurs d'en parcourir les rives et d'en bâtir une cartographie.

²⁷ Sur le sujet, saluons Jean-Baptiste Camps qui a produit une lumineuse synthèse à la fin de cet article : CAMPS Jean-Baptiste, « Copie, authenticité, originalité... », *op. cit.*

²⁸ Le delta (Δ) est ici à considérer comme la variation globale sur la superficie globale des loci, i.e. la concaténation des apports (δ) dits de quantité élémentaire. Le delta global est donc – dans notre cadre expérimental – la somme des variations partielles et/ou globales locales.

²⁹ Par exemple, Pierre de Poitiers, secrétaire personnel de Pierre le Vénérable qui, dans la première moitié du XII^e siècle, travaille de façon intermittente comme secrétaire à l'abbaye de Cluny. Après 1156, on le retrouve probablement à l'abbaye d'Anchin dans le nord de la France. Cf. TIENDREBEOGO Antoine, *Édition critique de la Lex Saracenorum...op. cit.*, p. 83.

³⁰ À comprendre ici comme le champ récent des « Data-driven Humanities » comme énoncé par Jean Baptiste Camps dans son article CAMPS Jean-Baptiste, « Où va la philologie numérique... », *op. cit.*

Bibliographie

BERRA Aurélien, « Connaître aujourd'hui. L'épistémologie problématique des humanités numériques », *Conférence du colloque DHnord2014. Humanités numériques : des outils, des méthodes, une culture*, 26-28 mai 2014, MESHS, URL : <http://publi.meshs.fr/ressources/connaître-aujourd'hui-lepistemologie-problematique-des-humanites-numeriques> (site Internet consulté le 17/04/2021).

BOURDEAU Michel, « L'idée de mathématiques appliquées chez Comte », *Mathématiques et Sciences Humaines*, Centre de Mathématique Sociale et de statistique, EPHE, 2011, p. 35-44.

CAMPS Jean-Baptiste, CAFIERO Florian, « Stemmatology: an R package for the computer-assisted analysis of textual traditions », *Corpus-Based Research in the Humanities CRH-2*, Vienne, janvier 2018, p. 65-74.

CAMPS Jean-Baptiste, « Où va la philologie numérique ? », *Fabula-LhT*, n° 20, « Le Moyen Âge pour laboratoire », janvier 2018, URL : <https://www.fabula.org/lht/20/camps.html> (site Internet consulté le 10/03/2021).

———, « Copie, authenticité, originalité dans la philologie et son histoire », *Questes* [En ligne], n° 29, 2015, mis en ligne le 20 juin 2015, URL : <http://journals.openedition.org/questes/3535> (site Internet consulté le 05/04/2021).

HORVATH Robert A., « Épistémologie et méthodologie de la statistique en relation avec la sociologie », *Journal de la société française de statistique*, t. 130, n° 4, 1989, p. 209-218. URL : http://www.numdam.org/item/JSFS_1989__130_4_209_0/ (site Internet consulté le 10/03/2021).

NOEL Michel, *Épistémologie, histoire et historiens : considérations conceptuelles, méthodologiques et empiriques autour du discours que les historiens tiennent sur leur savoir*, thèse de doctorat de l'Université de Laval, Québec, Canada, 2014. URL : <https://corpus.ulaval.ca/jspui/bitstream/20.500.11794/25426/1/30906.pdf> (site Internet consulté le 15/05/2021).

ROELLI Philipp (éd.) et al, *Handbook of Stemmatology History, Methodology, Digital Approaches*, De Gruyter, Berlin/Boston, 2020.

SCHØSLER Lene, « Scribal Variations: When Are They Genealogically Relevant and When Are They to Be Considered as Instances of 'Mouvance'? », *Studies in Stemmatology*, vol. 2, 2004, p. 207-226.

SEGRE Cesare, « Critique textuelle, théorie des ensembles et diasystème », *Bulletin de la classe des lettres et des sciences morales et politiques de l'Académie royale de Belgique*, n° 62, 1976, p. 279–292.

———, « Les transcriptions en tant que diasystèmes », *La pratique des ordinateurs dans la critique des textes*, Paris, 1979, p. 45-49.

TIENDREBEOGO Antoine, *Édition critique de la Lex Saracenorum, Première traduction du Coran en latin présentée dans le Corpus Toledanum et commandée par Petrus Venerabilis : Cas d'espèce ecdotique sur les sourates d'ouverture de l'Œuvre*, Mémoire de master 2 [non publié] sous la direction de DESWARTE Thomas, Université d'Angers, 2019.