

Traditions archivistiques et analyse computationnelle : modéliser les légendiers français à partir des données de Jonas

Shannon Bruderer¹ , and Ariane Pinche² 

¹ Maison des Sciences sociales et des Humanités - Lyon St-Étienne, Lyon, France

² CIHAM-UMR 5648, CNRS, Lyon, France

Abstract

Nowadays, Digital Humanities are challenged by the growing prominence of computational and AI-based approaches, sometimes at the expense of data structuration. This article argues that rigorous data modelling remains a central epistemological component of Digital Humanities. Through the construction of a dynamic relational database dedicated to French medieval hagiographic manuscripts (legendiers), based on the metadata of the Jonas database (IRHT), we demonstrate how modelling choices in the database are intertwined with philological issues. Focusing on the composition, of legendiers, the article shows how standards-based modelling enables the analysis of textual variance, compilation logics, and manuscript relationships at multiple levels. By combining philological concepts with interoperable data structures, this work shows the importance of data modelling as an essential practice at the core of Digital Humanities research.

Mots-clés: humanités numériques, sciences humaines, base de données, philologie médiévale

Keywords: digital humanities, humanities, database, medieval philology

Introduction

Une base de données peut être envisagée comme une représentation numérique d'un objet ou d'un concept. Par « représentation », nous entendons ici le travail de transposition qui consiste à rendre quelque chose présent sous la forme d'un substitut, en restituant les composants, les traits et les caractéristiques, puis en explicitant la manière dont ceux-ci entrent en relation les uns avec les autres.

Il y a ainsi quelque chose de l'ordre de l'« imitation » qui fonctionne comme une forme de traduction : la base de données transpose un objet inscrit dans le monde matériel vers un autre régime d'existence, celui d'une structure conceptuelle et formelle. Cela étant dit, et dans la mesure où le numérique lui-même repose sur des infrastructures physiques (serveurs, stockage, technologies), l'idée d'une supposée « dématérialisation » peut être relativisée.

Toute la difficulté de ce travail réside dans la capacité à rendre compte d'une réalité complexe à travers une architecture qui impose ses propres traditions et modalités techniques. Toute modélisation suppose dès lors une série de choix : déterminer ce qui sera représenté, ce qui sera laissé en marge, et la manière dont les éléments retenus seront articulés entre eux et, par alignement, au-delà du système lui-même.

Shannon Bruderer, and Ariane Pinche. "Traditions archivistiques et analyse computationnelle : modéliser les légendiers français à partir des données de Jonas." *Actes de la Conférence Humanistica*, éd. par Serena Crespi, Simon Gabay, Martin Grandjean, Ariane Pinche, Marie Puren et Léa Saint-Raymond. Vol. 4. Anthology of Computers et the Humanities. 2026, 166–174. <https://doi.org/10.63744/87xcm3uNiTMz>.

© 2026 par les auteurs. Sous licence Creative Commons Attribution 4.0 International (CC BY 4.0).

Les bases de données, en tant que structures organisées d'information, procèdent ainsi en deux temps. Premièrement, par décomposition : un objet est dissocié en un ensemble de propriétés distinctes. Deuxièmement, on en propose une nouvelle réécriture, où l'objet d'étude n'est pas simplement reproduit, mais décrit dans une forme qui n'est plus linéaire ou narrative, mais fragmentée, normalisée et interconnectée.

Au-delà de leur fonction de centralisation de l'information, les bases de données permettent de modéliser un objet à différentes échelles d'existence. Contrairement à l'idée selon laquelle le numérique induirait une forme d'absolutisme descriptif, souvent ressenti face aux choix de catégories qui semblent figer l'information dans sa qualification, la structuration des données introduit au contraire une finesse et une nuance accrues : un lieu, par exemple, peut être saisi simultanément à plusieurs niveaux, local comme global, et devient un label, une URI, des coordonnées ou encore un code multilingue et interopérable. Enfin, le caractère interrogeable de la base de données rend possible d'être simultanément précis et granulaire, tout en permettant une lecture généralisée des tendances lorsque l'on adopte une vue d'ensemble.

À travers la présentation des étapes de constitution d'une base de données relationnelle consacrée aux manuscrits hagiographiques français dits « légendiers », nous souhaitons montrer que ce travail « numérique » est étroitement lié à son objet d'étude, tant dans son existence matérielle que conceptuelle¹. La base de données s'appuie sur les métadonnées de la base Jonas², qui ont été récupérées, restructurées et mises en relation de manière à pouvoir être analysées afin de mieux comprendre les légendiers français, leurs logiques de composition et leurs évolutions entre le XIII^e et le XV^e siècle. Dans un premier temps, nous reviendrons sur les compilations hagiographiques et sur leurs enjeux philologiques ; nous présenterons ensuite les choix de modélisation opérés pour répondre au mieux aux questions de recherche ; enfin, nous proposerons une étude de cas afin d'illustrer la manière dont ces données peuvent être mobilisées.

1 Les compilations hagiographiques françaises

Les Vies de saints ont connu une large diffusion au cours du Moyen Âge [8], et ce, sous de multiples formes : textes unitaires, petits recueils thématiques ou d'auteur ou légendiers contenant des compilations de Vies de saints [7]. Dans le cas des légendiers français, cette circulation s'accompagne d'une transformation de l'organisation de la compilation. Alors que les légendiers latins sont le plus souvent organisés selon le calendrier liturgique, les légendiers français du XIII^e siècle présentent majoritairement une organisation thématique (apôtres, martyrs, confesseurs et saintes femmes). Les premiers ensembles hagiographiques ont été classés en familles par P. Meyer [6]. Elles fonctionnent de manière cumulative : les légendiers de type A regroupent les Vies des apôtres, les légendiers B associent apôtres et martyrs, les légendiers C apôtres, martyrs, confesseurs et les saintes femmes, tandis que les autres familles sont constituées sur le modèle de C avec des ajouts de nouveaux textes. Au cours du XIV^e s., les légendiers français tendent à revenir à une organisation liturgique. Ils proposent alors des versions abrégées des Vies de saints sur le modèle de la *Legenda Aurea* de Jacques de Voragine, qui connaît un vif succès et fait l'objet de nombreuses traductions [4].

Les légendiers contiennent des textes qui présentent des variations qu'il convient de prendre en compte afin de modéliser au plus près la mouvance propre aux textes médiévaux [9]. En effet, ceux-ci sont constitués de textes qui narrent la vie d'un saint, comme la *Vie de saint Lambert de Liège*. Cette trame narrative donne alors lieu à des mises en récit distinctes, produites par des auteurs différents qui, elles-mêmes, peuvent varier au cours des copies (variations graphiques,

1. Les données utilisées dans cette étude ainsi que les notebooks permettant leur traitement et leur analyse sont accessibles dans le dépôt suivant : https://gitlab.huma-num.fr/sbruderer/legendier_manuscrit.

2. Institut de Recherche et d'Histoire des Textes Section romane, « JONAS - Section Romane - IRHT », <https://jonas.irht.cnrs.fr>.

modernisation linguistique, erreurs, etc.). Ainsi, deux compilations peuvent partager une même trame narrative tout en intégrant des récits distincts, ce qui impose un niveau de granularité fin dans la modélisation.

Si la philologie traditionnelle s’est principalement appuyée sur l’unité textuelle afin d’analyser les relations entre les manuscrits, l’usage des méthodes numériques, et en particulier la modélisation au sein d’une base de données des traditions manuscrites, ouvre la possibilité d’aborder plus finement la transmission des textes au niveau des compilations et de prolonger les travaux fondateurs de Meyer, tout en offrant à la communauté scientifique un ensemble de données structuré, susceptible d’être réinvesti. Une telle démarche requiert une modélisation rigoureuse, ancrée à la fois dans les standards existants et dans les concepts philologiques, afin de constituer un outil scientifique pérenne et réutilisable.

2 Du manuscrit à sa représentation numérique : standards et interopérabilité

Dans le cadre de la modélisation de notre base de données, nous avons d’abord constitué un corpus initial de taille maîtrisable, tant du point de vue humain que computationnel, centré sur la présence de la *Vie de Lambert de Liège* dans les compilations. En partant des identifiants Jonas des manuscrits pour aller vers leurs contenus et métadonnées, un travail de moissonnage des données du site Jonas a été réalisé. Celui-ci a mis en évidence l’importance des identifiants (ARK), des permaliens et d’un schéma de métadonnées stable, lesquels ont permis de récupérer un nombre considérable de données et de faire le pont entre plusieurs dépôts³.

Après la collecte et le nettoyage des données, nous avons élaboré un premier modèle conceptuel. Celui-ci est centré sur les « textes » au sein des manuscrits et mobilise des ontologies permettant un traitement normalisé des métadonnées associées aux documents. L’usage de vocabulaires normalisés assurent la désambiguïsation et préparent un requêtage apte à répondre à des questions en apparence simples portant sur les périodes de composition, les lieux, et les participants (auteur, commanditaire, copiste, etc.), qui sont essentielles dans l’étude des corpus [5]. Conformément aux recommandations de FRBRoo, issues des pratiques bibliothéconomiques et du CIDOC-CRM, ancrées dans le champ de la conservation patrimoniale [1], ainsi qu’à d’autres ontologies, le modèle s’organise autour de quatre grandes entités :

- l’*Œuvre* (F1 Work)⁴, définie comme une entité abstraite et immatérielle, correspondant ici à une trame narrative centrée sur une figure de saint.
- l’*Expression* (F2 Expression)⁵, qui correspond à une réalisation intellectuelle de cette œuvre, c’est-à-dire une mise en récit particulière de la trame narrative. La relation entre l’œuvre et ses expressions peut être qualifiée par la propriété *R19 created a realisation of (was realised through)*⁶, qui permet de relier une œuvre abstraite aux différentes réalisations textuelles qui en dérivent.

3. Par exemple, le manuscrit Paris, Bibliothèque nationale de France, Manuscrits, fr. 00412 dispose d’une notice dans la base Jonas (https://jonas.irht.cnrs.fr/consulter/manuscrit/detail_manuscrit.php?projet=45533), d’un identifiant ARK dans la base ARCA (<https://arca.irht.cnrs.fr/ark:/63955/md182j62st0s>), d’un fac-similé numérique accessible sur Gallica (<https://gallica.bnf.fr/ark:/12148/btv1b84259980/f5.item>), d’une notice dans le catalogue Archives et Manuscrits de la BnF (<https://archivesetmanuscrits.bnf.fr/ark:/12148/cc50522d>), d’un manifeste IIF (<https://iiif.biblissima.fr/collections/manifest/9024c5d42e7e08f849652ff9c8dac5bec28dec0c>), ainsi que d’une notice dans le portail Data Biblissima recensant les différents identifiants attribués au manuscrit (<https://data.biblissima.fr/w/Item:Q51134>). L’ensemble de ces dépôts témoigne de l’existence d’un environnement documentaire numérique vaste, dont le degré d’interconnexion reflète les efforts menés pour produire des données interopérables et conformes aux principes FAIR.

4. <https://cidoc-crm.org/f1-work>.

5. <https://cidoc-crm.org/f2-expression>.

6. <http://iflastandards.info/ns/fr/frbr/frbroo/R19>.

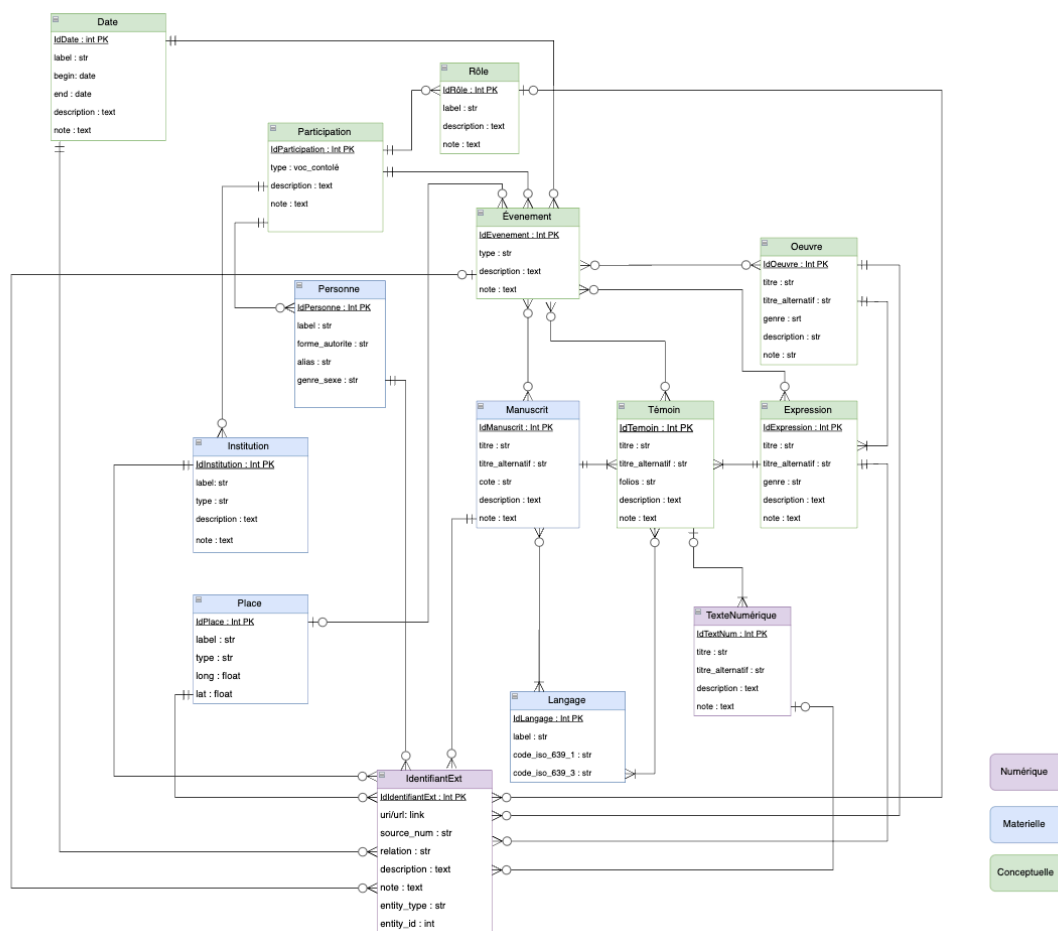


FIGURE 1 – Modèle logique de la base de données en notation *Crow's Foot*.

- le *Témoin*, entendu comme l'attestation d'une expression donnée dans un support matériel précis (par exemple dans un manuscrit à une localisation donnée dans celui-ci). Cette entité permet de relier une expression abstraite à sa présence effective dans un document historique, tout en distinguant plusieurs attestations possibles d'une même expression. L'alignement ontologique du témoin pose toutefois question en raison de sa nature à la fois conceptuelle et matérielle. Afin de distinguer le contenu textuel du support qui le transmet, nous avons fait le choix d'aligner le témoin sur la classe *E33 Linguistic Object*⁷ du CIDOC-CRM, qui permet de modéliser un contenu linguistique identifiable indépendamment de son support matériel. Dans ce cadre, le témoin correspond au texte attesté dans un manuscrit donné. Sa relation avec l'expression peut être qualifiée par la propriété *R13 is realised in (realises)*⁸, qui relie une expression à sa réalisation textuelle concrète.
- la *Manifestation Singleton* (F4 Manifestation Singleton)⁹, entendue comme l'objet matériel unique dans lequel le texte est conservé, par exemple un manuscrit particulier. La relation entre le manuscrit et le témoin est exprimée par la propriété *P128 carries (is carried by)*¹⁰, le manuscrit étant ici compris comme le support matériel qui porte le contenu textuel attesté par le témoin.

7. http://www.cidoc-crm.org/cidoc-crm/E33_Linguistic_Object.
 8. <http://iflastandards.info/ns/fr/frbr/frbroo/R13>.
 9. <https://cidoc-crm.org/f4-manifestation-singleton>.
 10. http://www.cidoc-crm.org/cidoc-crm/P128_carries.

Manuscrit : frbroo:F4_Manifestation_Singleton					
Champ	Type	Relation	Description	Exemple	Alignement sémantique
id_manuscrit	INT PK	—	Identifiant interne et unique du manuscrit dans la base de données (identifiant interne dérivé de la base Jonas).	45533	dcterms:identifier
titre	STRING	—	Titre normalisé ou label descriptif du manuscrit permettant d'identifier le manuscrit.	Paris, Bibliothèque nationale de France, Manuscrits, fr. 00412	dc:title rdfs:label
titre_alternatif	STRING	—	Autres titres utilisés pour désigner le manuscrit (titres historiques, variantes ou appellations secondaires).	—	dcterms:alternative
cote	STRING	—	Cote institutionnelle permettant d'identifier le manuscrit dans une collection.	fr. 00412	dcterms:identifier
description	TEXT	—	Description textuelle du manuscrit	NaN	dc:description
note	TEXT	—	Remarques complémentaires sur le manuscrit (commentaires, analyses ou informations diverses).	ms de grand luxe ; contient aux ff. 1-4 (dans la moitié supérieure de la page) un calendrier en latin, et une table de comput au f. 4	skos:note rdfs:comment
langue_id	FK	1 à n	Référence vers la langue principale du manuscrit dans la table Langue.	oil-français	dcterms:language
témoin	FK	1 à n	La relation relie le manuscrit aux entrées de la table Témoin, chacune correspondant à l'attestation d'une expression textuelle dans ce manuscrit.	manuscrit_id → témoin_id → expression_id	frbroo:F2_Expression
identifiantExt	FK	1 à n	Permet d'associer le manuscrit à des identifiants externes provenant d'autres bases (Wikidata, BnF, Jonas, Arca, etc.).	001	schema:sameAs skos:exactMatch rdfs:seeAlso
événement	FK	1 à n	Relation reliant le manuscrit aux événements qui lui sont associés (production, copie, enluminure, conservation, etc.), permettant également de relier personnes et institutions impliquées.	event_enluminure_456	crm:E7_Activity

FIGURE 2 – Structure de la table *Manuscrit* et alignement sémantique.

Pour donner un exemple concret, la *Vie de saint Lambert de Liège* constitue une œuvre comportant treize expressions recensées dans la base Jonas. Parmi celles-ci, on peut citer l'expression attribuée à Jean de Vignay¹¹ ou encore celle de Jeanne de Malone¹². Cette dernière est attestée par un témoin conservé dans le manuscrit Leiden, Bibliotheek der Rijksuniversiteit, BPL 0046 A, aux folios 81vb–95vb. Dans la base de données, une table *Texte Numérique*, alignée sur la classe *Item*¹³, permet de distinguer la matérialisation textuelle unique attestée dans le manuscrit des différentes représentations numériques du texte. Les numérisations, les transcriptions ou encore les éditions encodées en TEI peuvent ainsi être référencées séparément et associées chacune à un identifiant unique, permettant de tracer les différentes formes de circulation et de transformation du texte.

3 Analyse de réseau et logique de compilation des légendiers

Nous avons réalisé de premières analyses quantitatives en exploitant les différentes tables de données et modélisé des familles de légendiers à l'aide de visualisations en graphe. Notre corpus se compose au total de 71 manuscrits. Parmi les manuscrits les plus fournis, on compte le British Library, Stowe MS 0050, le BnF, fr. 00416 et Bibliothèque de l'Arsenal, 3682 qui sont tous les trois

11. Anne-Françoise Leurquin et Marie-Laure Savoye, notice *Vie de saint Lambert de Liège, Jean de Vignay*, base Jonas-IRHT/CNRS, <http://jonas.irht.cnrs.fr/oeuvre/3058>, consultation du 12/03/2026.

12. Section romane, notice *Vie de saint Lambert de Liège, Jeanne de Malone*, base Jonas-IRHT/CNRS, <http://jonas.irht.cnrs.fr/oeuvre/8764>, consultation du 12/03/2026.

13. <https://cidoc-crm.org/f5-item>.

des *Légendes dorées* du 15^e siècle traduites par Jean de Vignay. Le plus petit légendier du corpus est celui de la Bibliothèque royale de Belgique, 10326 (Légendier de la famille B) (voir figure 3).

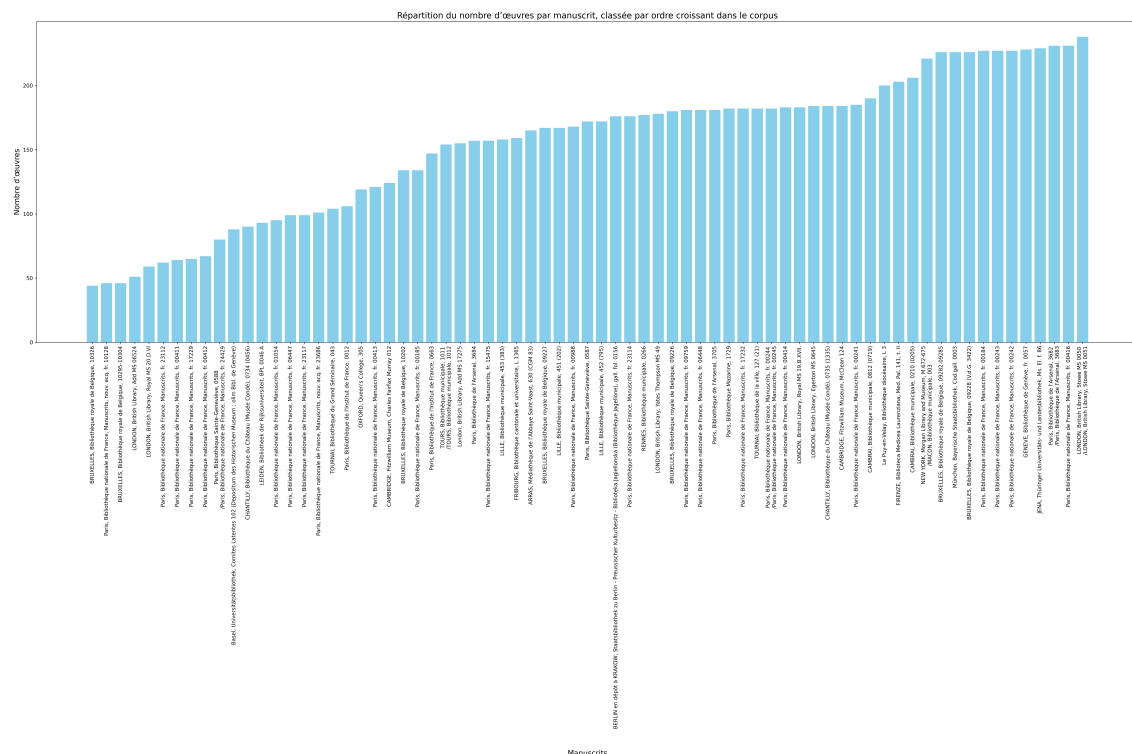


FIGURE 3 – Répartition du nombre d’œuvres par manuscrit, classées par ordre croissant dans le corpus.

Dans notre corpus, 687 trames narratives distinctes ont été identifiées, correspondant au niveau de l’*Œuvre*. Celles-ci donnent lieu à 2 653 expressions textuelles distinctes, chacune correspondant à une mise en récit particulière d’une même trame narrative. Enfin, ces expressions sont attestées par 11 038 occurrences textuelles dans les manuscrits du corpus.

Comme nous l’avons mentionné précédemment (2), une même trame narrative peut se réaliser en plusieurs récits différents possédant chacun leur identifiant propre (voir section 1). En effet, la *Vie de saint Lambert de Liège* connaît treize mises en récit distinctes, parfois sensiblement différentes [3]. La version ancienne porte l’identifiant 1958 dans la base Jonas, tandis qu’une autre version, correspondant à la traduction par Jean de Vignay de la version latine de la *Legenda aurea*, porte l’identifiant 3058.

Si la première question adressée aux données est relativement simple – à savoir comment les œuvres se répartissent à travers les manuscrits –, sa réponse met en évidence un écart important dans les résultats selon que l’on considère l’identifiant attribué aux récits

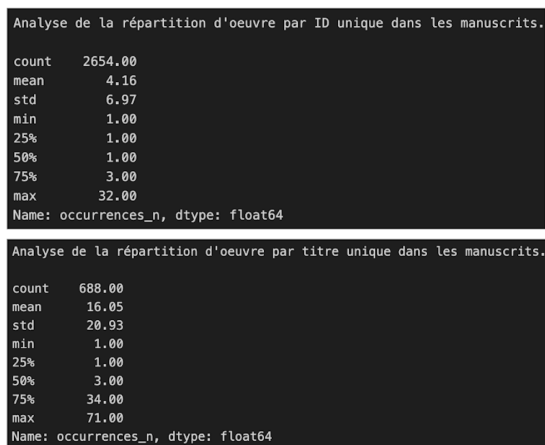


FIGURE 4 – Comparaison de la répartition du contenu des manuscrits selon un comptage par titre normalisé (trame narrative) ou par identifiant d’expression.

(c'est-à-dire aux expressions) ou que l'on regroupe les données selon leur titre comme référence, analysant ainsi la trame narrative, soit l'œuvre¹⁴. En effet, en prenant appui sur le titre, un total de 687 œuvres se répartissent entre les manuscrits avec en moyenne d'apparition dans le corpus de 16%. Dans cette perspective le troisième quartile est à 34, ce qui signifie que 25% des œuvres sont attestées plus de 34 fois (voir figure 4). Ici, la *Vie de saint Matthieu* et la *Vie de saint Lambert de Liège* sont les plus répétées¹⁵.

Si nous examinons les expressions à partir de leur identifiant, les récits de la *Vie de saint Clément* ou la *Commémoration des défunts* dominent le corpus (chacune présente dans trente-deux manuscrits), leur trame narrative s'étant réalisée dans un nombre relativement limité de mises en récit différentes.

Grâce au jeu de données, une première visualisation des logiques de compilation des légendiers a été réalisée. Nous avons élaboré un graphe traduisant les relations qu'entretiennent les manuscrits, en fonction des contenus qu'ils partagent. Plusieurs étapes de préparation ont été indispensables, dont l'ajout d'une colonne titre_id. Elle nous a permis d'attribuer à l'ensemble des œuvres un identifiant fondé sur le titre, afin de rendre compte des logiques de compilation, non pas uniquement des versions textuelles, mais aussi au niveau des trames narratives (F1 Work). En effet, les œuvres (trames narratives) correspondant à un même titre ont ainsi été regroupées sous un identifiant commun, afin de permettre un second niveau d'analyse de compilation des légendiers¹⁶. Nous avons ainsi construit un graphe bipartite dans lequel les nœuds sont répartis en deux ensembles distincts : les manuscrits d'un côté et les trames narratives de l'autre. À ce stade, les relations manuscrit-manuscrit sont indirectes et transitent par les nœuds représentant les trames narratives qui servent de nœuds-ponts. La proximité entre deux manuscrits est ensuite calculée à partir du nombre de trames narratives partagées ($weight(M1, M2) = |O(M1) \cap O(M2)|$), ce qui définit le poids du lien dans le graphe projeté et permet l'identification de clusters de manuscrits.

Le graphe bipartite initial comporte 758 nœuds (trames narratives et manuscrits), ce qui correspond à l'ensemble des 687 trames et des 71 manuscrits du corpus. Une fois le graphe projeté, les résultats montrent une forte standardisation des légendiers : le graphe manuscrit-manuscrit compte 71 nœuds et 2 485 arêtes, indiquant que la quasi-totalité des manuscrits partage au moins une trame avec les autres. Les statistiques montrent également que la moitié des paires partagent plus ou moins 75 titres entre eux, le maximum est de 226, le minimum de 6. Certains manuscrits comme Thüringer Universitäts – und Landesbibliothek, Ms. El. f. 86 et BnF, fr. 00416 possèdent des compilations presque identiques avec 226 œuvres partagées. L'analyse permet d'identifier deux clusters, respectivement composés de 39 et 32 manuscrits, ce qui souligne l'existence d'une logique de compilation largement partagée entre les manuscrits (voir figure 5). En raison de la forte densité du réseau, causée par la présence de trames narratives communes largement diffu-

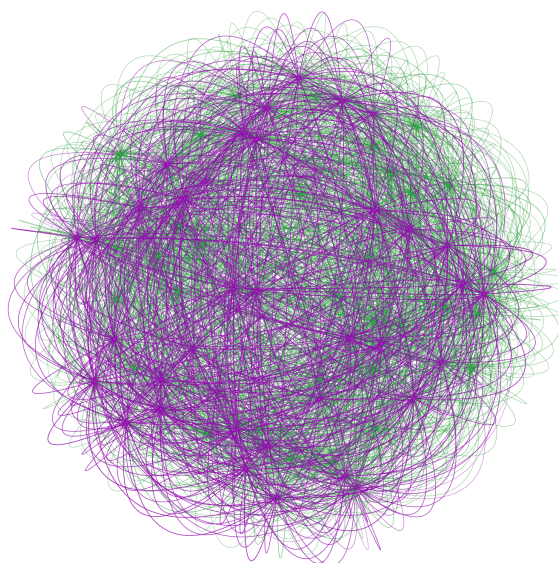


FIGURE 5 – Graphe des relations entre manuscrits fondé sur le partage des trames narratives.

14. La question se complique encore si plusieurs saints portent le même nom, comme c'est le cas pour saint Grégoire, le pape, l'auteur et le bon pécheur.

15. Ce qui est relativement normale car la *Vie de saint Lambert de Liège* a servi à la constitution du corpus.

16. Si jamais un copiste conservait la même compilation, mais pour une raison ou une autre décide d'insérer une autre version d'un texte.

sées, les deux principales familles ne sont distinguables que par leur couleur.

En comparaison, le graphe basé sur les différentes expressions d'une même trame narrative propose une autre visualisation des relations entre les manuscrits. Comportant 2 724 nœuds, il se compose de cinq clusters, regroupant respectivement 31, 28, 9, 2 et 1 manuscrits (voir figure 6). Cette visualisation souligne combien le niveau d'analyse retenu influe sur la représentation des relations entre les manuscrits. Les deux niveaux coexistent en effet simultanément dans les sources, ce qui montre l'intérêt d'une modélisation capable de rendre compte de ces différentes échelles d'existence du texte.

L'une des perspectives de travail consisterait à filtrer les trames les plus fréquentes afin d'atténuer le poids du standard imposé par ces textes communs et de faire émerger des liens plus spécifiques. Cette approche repose sur l'hypothèse que, dans des pratiques de compilation fortement standardisées, ce sont les entrées rares qui témoignent d'une proximité plus significative entre les manuscrits¹⁷.

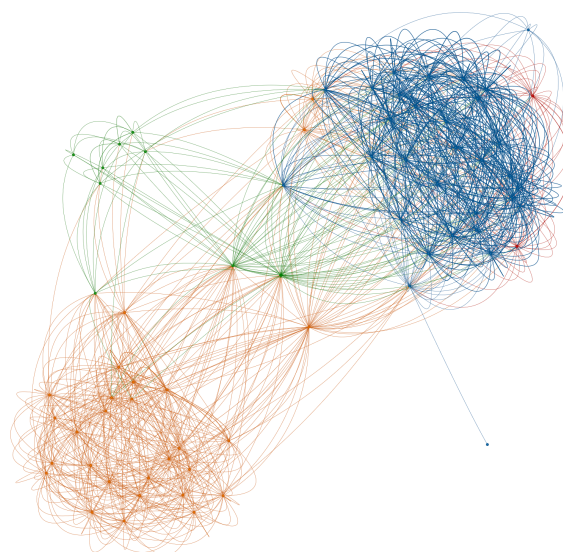


FIGURE 6 – Graphe des relations entre manuscrits fondée sur le partage des expressions.

Conclusion

Si plusieurs pistes quant à l'analyse des logiques de compilation ont déjà été évoquées, ce travail appelle également des analyses textuelles plus fines. Notamment avec la mise en ligne récente de CoMMA, qui donne accès à la transcription de plus de 30 000 manuscrits [2], il serait possible d'élargir l'étude et de dépasser des analyses uniquement basées sur les métadonnées pour s'orienter vers des méthodes computationnelles appliquées au texte lui-même : analyse du nombre de tokens, mesure du *token overlap*, ou encore calcul de similarités cosinus. Ces approches permettraient de proposer des analyses à plusieurs niveaux, de la trame narrative aux témoins manuscrits.

Plus largement, ce travail met en évidence la zone d'intersection dans laquelle se situent les humanités numériques. La juxtaposition des termes « humanités » et « numériques » engage, depuis l'émergence de la discipline, de longues discussions sur la compatibilité de ces deux « mondes ». Cependant, nous avançons que, dans un environnement archivistique qui tend vers l'interconnexion de ses données, les frontières deviennent plus floues. Pour la constitution de notre base de données, il a en effet été indispensable de croiser trois champs de réflexion majeurs : la matérialité de l'objet manuscrit, ses spécificités conceptuelles autour de la notion de « texte », et son contexte numérique. Chacune de ces dimensions s'est trouvée intégrée au modèle conceptuel, non comme un simple ajout, mais comme un élément structurant.

Il convient toutefois de souligner que ce processus est loin de suivre un déroulement linéaire allant de la collecte des données au traitement, puis à l'analyse. La pratique a montré que le travail sur les données repose au contraire sur de constants allers-retours entre ces différentes phases. Bien souvent, ce sont les premiers résultats quantitatifs ou les visualisations issues des méthodes computationnelles qui révèlent des incohérences, des défauts de normalisation ou des indexations

17. Nous avons mené la même expérience à partir des différents récits des Vies (expressions), en initiant les regroupements à un seuil de dix Vies partagées. Cette approche a permis de dégager trois clusters. Après vérification à partir de la *Vie de saint Lambert*, chaque cluster semble associé à des mises en récit spécifiques, montrant que la transmission de ces textes ne s'opère pas principalement à l'échelle du texte isolé, mais bien à celle de la compilation.

inadéquates. La réalité professionnelle des ingénieur·es en humanités numériques, amené·es à intervenir à différents niveaux de complexité, montre que la maîtrise des traditions archivistiques et une connaissance du document restent des conditions essentielles à la production d'analyses pertinentes.

Enfin, ce travail rappelle que les opérations dites « manuelles » ou de base sur les données, comme le nettoyage, l'harmonisation ou la documentation, ne constituent pas une étape secondaire, mais forment le socle même à partir duquel des analyses solides, d'un point de vue computationnel comme disciplinaire, peuvent émerger.

Références

- [1] BUARD, Pierre-Yves. *Modélisation des sources anciennes et édition numérique*. Thèse de doct. Université de Caen, 2015. URL : <https://hal.science/tel-01279385>.
- [2] CLÉRICE, Thibault, GABAY, Simon, VLACHOU-EFSTATHIOU, Malamatenia, PINCHE, Ariane et SAGOT, Benoît. « CoMMA, a Large-scale Corpus of Multilingual Medieval Archives ». 2025. URL : <https://inria.hal.science/hal-05299220>.
- [3] CLÉRICE, Thibault et PINCHE, Ariane. « Wauchier, Is That You? A Multi-Manuscript Authorship Analysis of Saint Lambert's Life ». In : *Computational Humanities Research 2025*, sous la dir. de Taylor ARNOLD, Margherita FANTOLI et Ruben Ros. Anthology of Computers and the Humanities, 2025, p. 149-165. DOI : 10.63744/QsBV0XYj8wRC.
- [4] Barbara FLEITH et Franco MORENZONI, éd. par. *De la sainteté à l'hagiographie : genèse et usage de la Légende dorée*. Publications romanes et françaises 229. Genève : Droz, 2001.
- [5] HASSNER, Tal, REHBEIN, Malte, STOKES, Peter A. et WOLF, Lior. « Computation and Palaeography : Potentials and Limits (Dagstuhl Perspectives Workshop 12382) ». Rapp. tech. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013. DOI : 10.4230/DAGMAN.2.1.14.
- [6] MEYER, Paul. « Légendes hagiographiques en français ». In : *Histoire littéraire de la France*. Imprimerie nationale. T. 33. Paris : Paris : Imprimerie nationale, 1906, p. 328-458. URL : <http://archive.org/details/histoirelitttra33riveuoft>.
- [7] PHILIPPART, Guy. *Les Légendiers latins et autres manuscrits hagiographiques*. Turnhout : Brépols, 1977.
- [8] TRIGALET, Michel. « Compter les livres hagiographiques. Aspects quantitatifs de la création et de la diffusion de la littérature hagiographique latine (II-XVe siècle) ». In : *Gazette du livre médiéval* 38, no. 1 (2001), p. 1-13.
- [9] ZUMTHOR, Paul. « Intertextualité et mouvance ». In : *Littérature* 41, no. 1 (1981), p. 8-16. DOI : 10.3406/litt.1981.1331.