



HAL
open science

Texte et image : les livres d'heures manuscrits vus par l'intelligence artificielle

Dominique Stutzmann, Louis Chevalier

► To cite this version:

Dominique Stutzmann, Louis Chevalier. Texte et image : les livres d'heures manuscrits vus par l'intelligence artificielle. *Humanistica*, May 2021, Rennes, France. hal-03234821

HAL Id: hal-03234821

<https://hal.science/hal-03234821>

Submitted on 25 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Texte et image : les livres d'heures manuscrits vus par l'intelligence artificielle

Colloque Humanistica (Rennes, 10-12 mai 2021)

Dr. Dominique Stutzmann

Dr. Louis Chevalier

Les livres d'heures médiévaux, livres de prière à l'usage des laïcs, sont le « best-seller » du Moyen Âge. Produits en grand nombre à partir du XIII^e siècle dans toute l'Europe occidentale et plus nombreux que les manuscrits bibliques, ils sont une source de premier ordre pour l'histoire du sentiment religieux, des pratiques sociales, mais aussi de la circulation des textes et des transmissions entre régions différentes, ou encore des processus d'industrialisation de la production livresque, car ils sont largement standardisés¹. Avant l'invention de l'imprimerie en Occident, ils sont les premiers livres dits « d'étal », réalisés par les libraires en dehors de toute commande, faisant passer le livre d'une économie de la demande à une économie de l'offre. Livres de luxe, ils sont souvent richement ornés et ont donné lieu à la formalisation de cycles iconographiques stables, comme celui de l'Enfance du Christ.

Ces livres sont nombreux et volumineux, compilés à partir de centaines de pièces élémentaires et, contrairement aux enluminures, parfois peintes par les plus grands artistes du temps, leur contenu textuel a découragé les études systématiques. Or il faut examiner un très grand nombre de ces documents pour saisir la relation qui lie leurs différentes parties, le degré de fréquence ou de rareté des chants, des lectures, oraisons et prières, et comprendre la circulation de ces textes et les adaptations individuelles d'un produit de masse.

Les résultats présentés aujourd'hui ont été élaborés dans le cadre du projet ANR *HORAE Hours : Recognition, Analysis, Edition*, dont nous avons déjà exposé les principes au colloque *Digital Humanities* d'Utrecht en 2019². Le programme de recherche réunit trois partenaires complémentaires, spécialisés en humanités, en Traitement automatique du Langage Naturel (TALN) et en intelligence artificielle. Les défis technologiques, maintenant relevés, portent sur l'analyse d'image, en particulier la lecture des écritures médiévales par ordinateur ou l'analyse de la mise en page³, l'identification de textes par « alignement » ou *text reuse identification*⁴

¹ Dominique Stutzmann, « Résistance au changement ? Les écritures des livres d'heures dans l'espace français (1200-1600) », in « *Change* » in *medieval and Renaissance scripts and manuscripts. Proceedings of the 19th Colloquium of the Comité international de paléographie latine (Berlin, 16-18 September, 2015)*, éd. par Eef Overgaauw et Martin J. Schubert, *Bibliologia* 50 (Turnhout: Brepols, 2019), 97-116.

² Dominique Stutzmann et al., « Integrated DH. Rationale of the HORAE Research Project » (*Digital Humanities 2019, Utrecht, 2019*), <https://dev.clariah.nl/files/dh2019/boa/0192.html>.

³ Mélodie Boillet et al., « HORAE: an annotated dataset of books of hours », in *the 5th International Workshop on Historical Document Imaging and Processing (International Workshop on Historical Document Imaging and Processing, Sydney: ACM Press, 2019)*, 7-12, <https://doi.org/10.1145/3352631.3352633>; Emanuela Boros et al., « Automatic Page Classification in a Large Collection of Manuscripts Based on the International Image Interoperability Framework », in *2019 International Conference on Document Analysis and Recognition (ICDAR) (International Conference on Document Analysis and Recognition, Sydney: IEEE, 2019)*, 756-62, <https://doi.org/10.1109/ICDAR.2019.00126>.

⁴ Amir Hazem et al., « Réutilisation de textes dans les manuscrits anciens », in *TALN-RECITAL Conférence sur le Traitement Automatique des Langues Naturelles PFIA 2019 (TALN-RECITAL 2019, Toulouse, 2019)*, 417-25, https://www.irit.fr/pfia2019/wp-content/uploads/2019/07/actes_TALN-RECITAL-courts_CH_PFIA2019-2.pdf.

et l'établissement automatique de tables de matières hiérarchiques ou « segmentation »⁵. Cette communication a maintenant pour objectif de montrer la possibilité et l'intérêt d'une analyse des livres d'heures, du point de vue du texte comme de l'image, soutenue par l'intelligence artificielle.

Pour cela, nous décrivons, dans une première partie, le cadre scientifique, institutionnel et technologique de l'étude, en mettant l'accent sur les aspects historiques. Dans une seconde partie, nous présenterons les résultats déjà obtenus par les traitements automatiques, en particulier pour ce qui est de l'analyse graphique afin de souligner les enjeux de volumétrie et de mesure de la précision. Enfin, dans une troisième partie, nous exploiterons les résultats de l'analyse textuelle sur un sous-corpus validé. Nous mettrons en évidence des caractéristiques textuelles encore inconnues et établirons des corrélations entre texte et image dans les manuscrits médiévaux.

1 Cadre et enjeux

1.1 ANR HORAE : partenaires et tâches

Le projet HORAE vise à l'analyse des livres d'heures à grande échelle, en étudiant les textes transmis et en les mettant en relation avec les caractéristiques matérielles, codicologiques ou décoratives.

Le financement par l'Agence Nationale de la Recherche, d'un montant de quatre cent mille euros, couvre une durée de quatre ans. Le consortium interdisciplinaire rassemble des partenaires en humanités, en Traitement automatique du Langage Naturel (TALN) et en intelligence artificielle, particulièrement en apprentissage machine. Il s'agit de l'Institut de Recherche et d'Histoire des textes (CNRS) pour les Humanités, du LS2N Laboratoire des Sciences du numérique de Nantes pour le TAL, et de la société privée Teklia pour l'intelligence artificielle.

Ce consortium a donc des liens forts, institutionnels et personnels, avec des projets de recherche connexes qui ont déjà été présentés dans les conférences d'humanités numériques, dont Oriflamms, Himanis et Home History of Medieval Europe.

Les grandes tâches du projet ont été définies comme suit :

- Recensement des livres d'heures conservés dans le monde et des textes principaux qu'ils transmettent ;
- Préparation de la vérité terrain pour l'apprentissage machine ;
- Analyse graphique par ordinateur pour produire des annotations automatiques de type de page, d'éléments sur la page tels que miniatures, décoration, rubriques et initiales, mais aussi lignes de texte ;
- Analyse graphique des lignes de texte pour la production automatique d'un texte transcrit, dit aussi « HTR » ;
- Analyse par ordinateur du texte, d'une part « l'alignement » produisant une identification des textes, et, d'autre part, la « segmentation hiérarchique », créant des tables de matières à plusieurs niveaux.

Tout au long de ces étapes, les humanistes interviennent pour annoter, valider et corriger les annotations automatiques, puis, *in fine*, pour procéder à l'analyse des données, en particulier

⁵ Amir Hazem et al., « Hierarchical Text Segmentation for Medieval Manuscripts », in *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020, Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020)*, 6240-51, <https://www.aclweb.org/anthology/2020.coling-main.549>.

par mécanismes de fouilles de données ou d'analyse de réseaux. Cette architecture des tâches, à plusieurs étages, est requise en raison de la matière et des enjeux SHS du projet.

1.2 Circulation des textes : histoire textuelle et liturgique

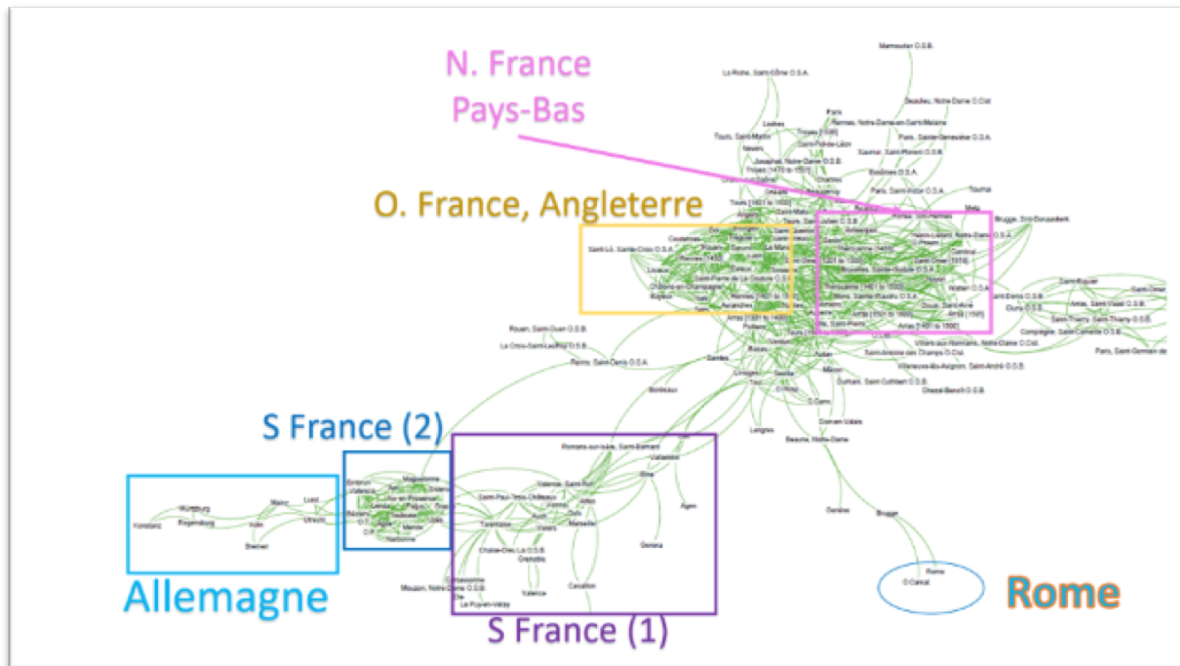
Tout d'abord, l'étude des textes et de leur circulation. Le texte des livres d'heures est complexe, marqué par une très forte structuration granulaire. Chaque manuscrit est une « compilation de compilations », composé, au niveau supérieur, de différentes sections, notamment le calendrier, les péripécies évangéliques, les heures de la Vierge, l'office des morts, les heures abrégées de la Croix et de l'Esprit Saint, les psaumes pénitentiels et les litanies, et les suffrages, auxquelles peuvent s'ajouter diverses prières.

Les heures et les offices sont constitués, à un niveau inférieur, de chants, de lectures et d'oraisons répartis selon les heures de la journée. Leur structure est identique en tous lieux, ainsi que le choix de certaines pièces comme celui des cantiques évangéliques des laudes, des vêpres et des complies. Cependant, le choix des autres pièces, réalisé dans un large répertoire, comprenant dans le cas des heures de la Vierge près de quatre cents pièces, varie selon les établissements religieux, et ces choix caractéristiques sont appelés « usage ». L'étude des livres d'heures nous permet d'éclairer les particularités et la diffusion de ces *usages* à travers l'Europe, ainsi que des réseaux d'échanges entre des groupes d'usages.

La Figure 1, réalisée à l'occasion du colloque *Networks of Manuscripts, Networks of Texts* (octobre 2020, Amsterdam), présente un réseau des usages liturgiques des heures de la Vierge⁶. Elle donne à voir une cartographie intellectuelle, montrant la proximité des usages selon le choix et l'ordre de leurs pièces liturgiques. Elle met en évidence l'existence de deux groupes d'usages provenant du Sud de la France, d'un groupe d'usages anglais et de l'Ouest de la France, et d'un groupe du Nord de la France et des Pays-Bas méridionaux, ainsi que d'un groupe d'usages allemands, dépendant d'un modèle dominicain et teutonique comme en témoigne leur parenté avec des usages du Sud de la France.

⁶ Dominique Stutzmann et Louis Chevalier, « Books of hours as text compilations in the Low countries », in *Networks of Manuscripts, Networks of Texts* (Amsterdam, Netherlands: Evina Steinová (Huygens ING), 2020), <https://hal.archives-ouvertes.fr/hal-03112110>.

Figure 1 : Heures de la Vierge - Réseau des proximités textuelles



Selon les intentions des commanditaires et le lieu de production, on peut observer des phénomènes d'hybridation des usages liturgiques ou des variations dans le choix des prières, des saints honorés dans les litanies et les suffrages, ou la copie d'offices rares. Ainsi, les fréquences relatives d'apparition, les variantes textuelles des prières, ou les relations des textes dits accessoires avec les autres parties du livre recèlent des indications sur l'identité, la dévotion et le jeu de représentation sociale des commanditaires du livre d'heures.

1.3 Codicologie quantitative

Un autre domaine concerné par notre projet est la « codicologie quantitative ». Ce terme désigne les études entreprises depuis les années 1980 sur les aspects matériels des manuscrits médiévaux⁷. Cela couvre les problématiques économiques, les rythmes et coûts de production et les taux de conservation, mais aussi les questions de format et de mise en page et de mise en texte, référant non seulement aux techniques artisanales, mais aussi aux problématiques de lisibilité, vues par le prisme des longueurs de ligne, nombre de colonnes, de mots, de caractères et d'abréviations, et encore à des questions esthétiques et de luxe.

Parmi les travaux récents, plusieurs proposent une évaluation de la richesse du décor pour compléter leurs approches sociologiques et classificatoires et ainsi de mieux comprendre la réception d'œuvres médiévales et les milieux et conditions de commande et de copie⁸.

⁷ Ezio Ornato et Carla Bozzolo, *Pour une histoire du livre manuscrit au moyen âge : trois essais de codicologie quantitative* (Paris: Ed. du C.N.R.S., 1980); Ezio Ornato, *La face cachée du livre médiéval : l'histoire du livre vue par Ezio Ornato, ses amis et ses collègues*, I libri di Viella 10 (Roma: Viella, 1997).

⁸ Emilie Cottureau, « La copie et les copistes français de manuscrits aux XIVe et XVe siècles. Etude sociologique et codicologique » (Paris, Université Paris 1 - Panthéon-Sorbonne, 2005); Emilie Cottureau-Gabillet, « Manuscrits de luxe et distinction sociale à la fin du Moyen Âge », in *Marquer la prééminence sociale*, éd. par Jean-Philippe Genet et E. Igor Mineo, Histoire ancienne et médiévale (Paris: Éditions de la Sorbonne, 2015), 283-301, <http://books.openedition.org/psorbonne/3352>; Emilie Cottureau-Gabillet, « Revealing Some Structures and Rules of Book Production (France, Fourteenth and Fifteenth Centuries) », in *Ruling the Script in the Middle Ages. Formal Aspects of Written Communication (Books, Charters, and Inscriptions)*, éd. par Sébastien Barret, Dominique Stutzmann, et Georg Vogeler, Utrecht Studies in Medieval Literacy 35 (Turnhout: Brepols, 2016),

Les livres d'heures sont certes, pris dans leur ensemble, des livres de luxe, où l'écriture occupe en moyenne moins de 40 % de la page, mais leur diversité presque infinie va nous permettre d'observer aujourd'hui des niveaux variés et de mieux catégoriser et comprendre leur production.

2 Résultats de l'analyse automatique

Pour ce faire, avant d'entrer dans le détail d'une enquête nouvelle, passons à notre seconde partie et présentons l'état actuel de nos travaux.

2.1 Base Heurist

Les opérations de recensement des manuscrits et d'établissement des textes les plus fréquents ont permis d'alimenter une base Heurist avec près de 7500 manuscrits, 400 organisations, plus de 200 usages liturgiques décrits finement et plus de 15 000 œuvres. Une liste de plus de mille manifests IIF, aussi sur Github, permet d'accéder aux images de façon normalisée.

2.2 Arkindex

Cette liste de manifests alimente la plateforme Arkindex où est réalisée l'analyse graphique des images. À ce jour, nous avons traité 993 volumes, plus de 260 000 images, et généré automatiquement des annotations pour environ 38 000 miniatures, un million d'initiales, six millions de lignes de texte.

Cette analyse graphique, ainsi que la reconnaissance du texte manuscrit, est obtenue avec des réseaux neuronaux profonds, entraînés à partir d'images annotées. Les modèles peuvent être entraînés avec des corpus de taille modeste.

La reconnaissance est de bonne qualité sans être parfaite. Elle est suffisante pour passer à une étape ultérieure : l'exploitation du texte tout d'abord pour identifier les textes unitaires à l'intérieur du volume, comme ici le début du psaume 6, marqué par une initiale historiée, mais aussi ici, le début du psaume 142 qui, exceptionnellement, n'est indiqué par aucune marque graphique spéciale.

Elle permet ensuite, avec des modèles entraînés de la même façon, de reconnaître les structures des livres d'heures et les différentes sections et sous-sections de façon hiérarchisée.

3 Psaumes et usages liturgiques

Pour montrer dans la présente communication l'intérêt de ces opérations, nous choisissons un sous-corpus de 50 livres d'heures complets, pour lesquels nous avons validés les résultats de l'analyse automatique, et d'exploiter spécifiquement le texte des psaumes. Pourquoi les psaumes ? Textes bibliques de longueur variable, au nombre de cent-cinquante, les psaumes sont au centre de la prière chez les Chrétiens. Ils apparaissent, au sein des livres d'heures, dans les heures de la Vierge, dans l'office des morts et dans la section des psaumes pénitentiels, ainsi que dans certains offices votifs plus rares comme les heures de la Passion.

129-63; Octave Julien, « Déliver, lire et relier », *Hypothèses* 19, n° 1 (13 octobre 2016): 211-24; Octave Julien, « Lire, écrire, relier : la composition des recueils vernaculaires français et anglais de la fin du Moyen âge », <http://www.theses.fr> (thesis, Paris 1, 2016), <http://www.theses.fr/2016PA01H014>; Chiara Ruzzier, « Continuité et rupture dans la production des bibles au XIII^e siècle », in *Comment le Livre s'est fait livre. La fabrication des manuscrits bibliques (IVe-XVe siècle)*, éd. par Chiara Ruzzier et Xavier Hermand, vol. 40 (Turnhout: Brepols Publishers, 2015), 155-68, <https://doi.org/10.1484/M.BIB-EB.5.105426>.

Les heures de la Vierge comptent généralement une trentaine de psaumes, répartis par groupes de trois ou cinq entre huit offices, et l'office des morts, une vingtaine.

Très courants, ces textes ne sont pas catalogués au sein des livres liturgiques ou des livres d'heures. Or nous allons montrer que, même en ne se fondant que sur ces textes très connus, il est possible de faire des découvertes.

3.1 Repérage des psaumes

L'analyse graphique, tout d'abord, sur les 50 volumes, soit près de douze mille pages, a donné 800 miniatures, 50 000 initiales, 250 000 lignes de texte.

L'analyse automatique du texte produit par HTR a donné une liste de 2240 psaumes présumés. Parmi ceux-ci, seuls 17 étaient des erreurs, dont quinze confusions entre les psaumes 39 et 69 : le psaume 39 a été identifié quinze fois comme psaume 69 et les psaumes 94 et 116 ont été identifiés chacun une fois comme psaumes 8 et 145. La précision est excellente, dépassant les 99 %. Par ailleurs, des prédictions en doublon sont apparues en raison de la méthode d'optimisation du temps de calcul, à savoir un découpage du texte reconnu par HTR en sections de la taille de chaque psaume recherché avec glissement d'un pas de 0.75. Ainsi, deux sections successives peuvent être suffisamment proches du texte recherché et se voir affecter la même identification. Dans notre corpus de test, 214 prédictions étaient des doublons, soit une précision de 90 % si on pénalise les doublons.

Après validation, nous avons une liste de 2029 psaumes, avec 36 psaumes ajoutés par correction. Le rappel est donc de 98 %. Les erreurs et omissions se concentrent sur des volumes où la reconnaissance de texte a été moins performante.

Le nombre de psaumes, allant de 1 à 75 par volume, n'est pas corrélé au nombre de pages. Tel manuscrit (Karlsruhe Lichtenthal 45) ne copie qu'un psaume en entier sur 171 pages ; les autres sont indiqués par leurs premiers mots et récités de mémoire. Inversement, un même psaume peut être copié jusqu'à quatre fois dans le même volume, en l'occurrence le psaume 129 dans le manuscrit Beaune 53. Ce psaume 129 connaît en tout 88 copies dans les 50 volumes.

Figure 2 : Détection des psaumes au sein des livres d'heures du corpus validé

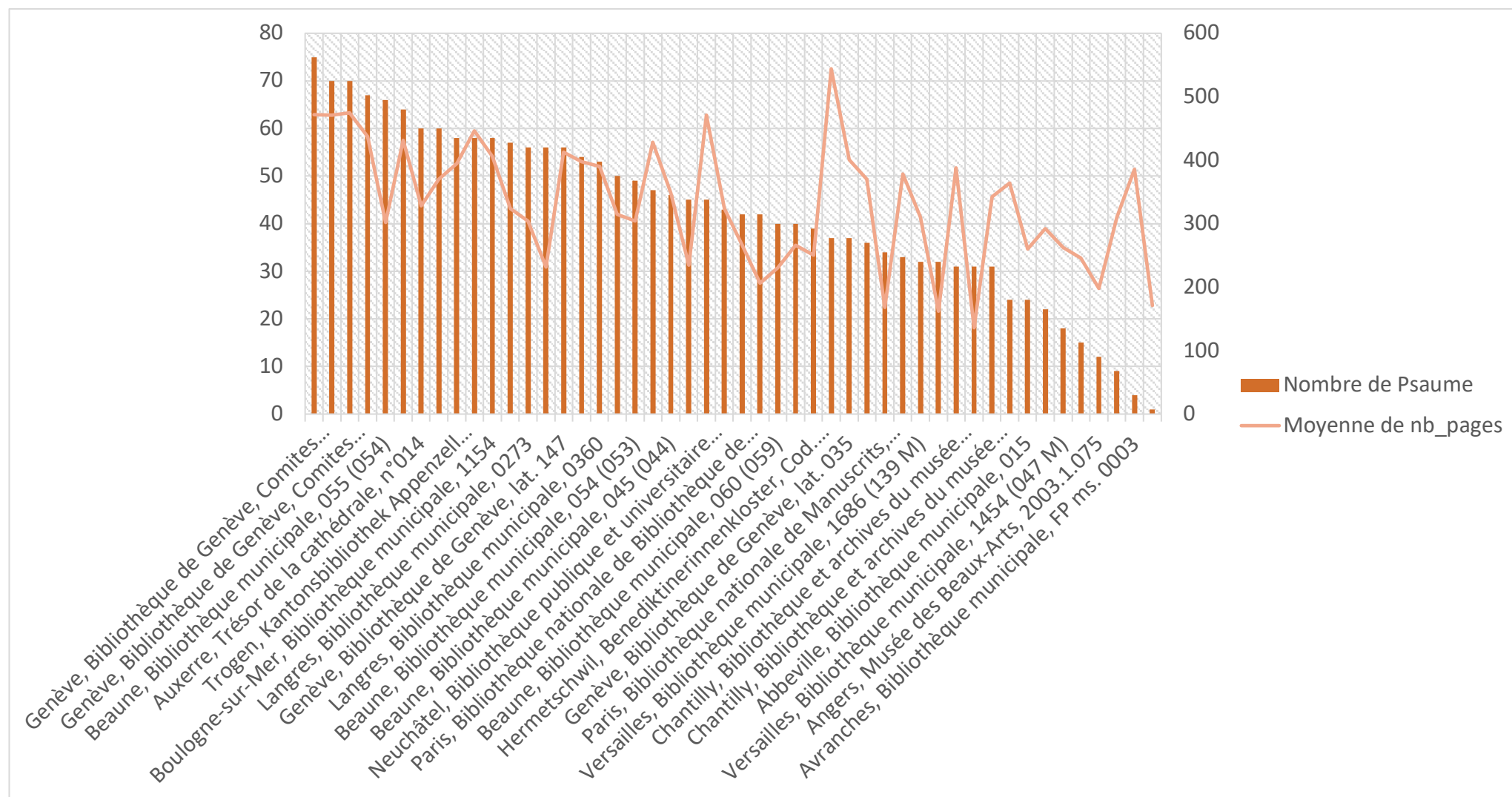
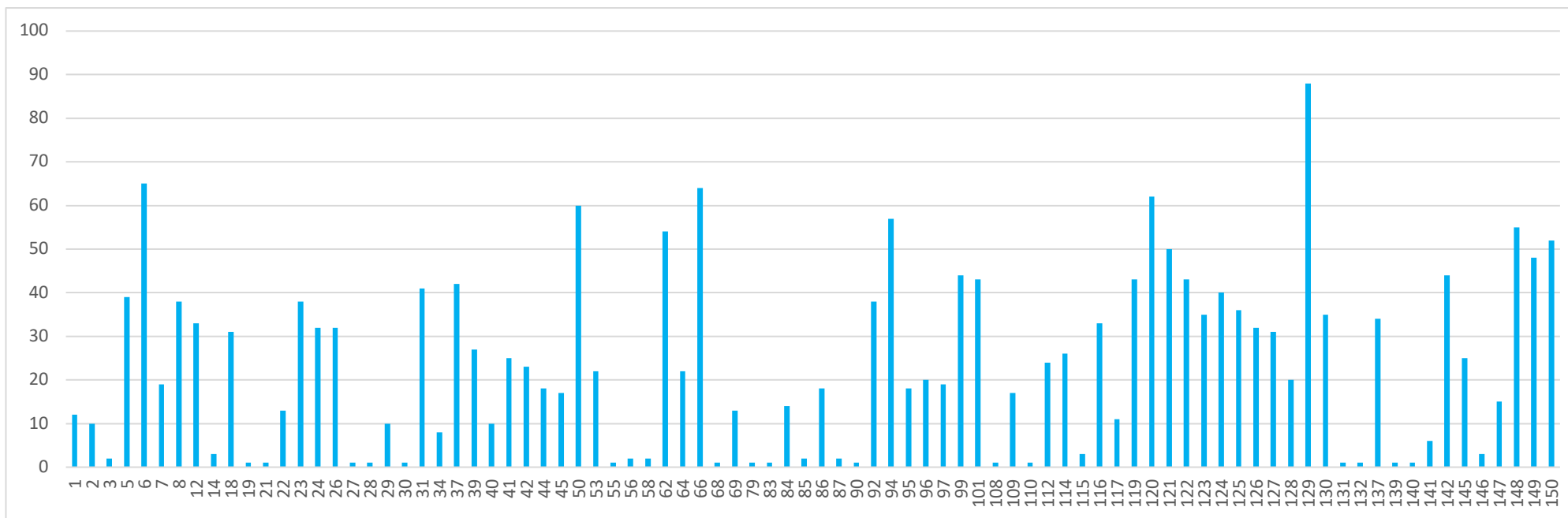


Figure 2 : Représentation des psaumes au sein des livres d'heures du corpus validé



3.2 Psaumes et usage liturgique

Si l'étude des psaumes ne suffit pas à déterminer l'origine d'un office, elle permet d'observer des phénomènes de mutation des usages dans certains livres d'heures, et de former des familles d'établissements géographiquement ou institutionnellement proches, de restituer une relation liturgique entre des textes présents dans des sections distinctes, d'observer des variations des séries de psaumes et de déceler des pratiques régionales.

3.2.1 Usage du psaume 69 au sein des litanies

Par exemple, dans les « litanies ». Celle-ci sont composées d'une suite d'invocations aux saints, du *Pater noster*, et de versicules et d'oraisons. Dans neuf livres d'heures de notre corpus, un psaume, le psaume 69, est copié en entier entre le *Pater noster* et la suite de versicules⁹. Aucun autre psaume n'apparaît dans les litanies de ces manuscrits, ni dans les litanies des autres manuscrits du corpus. Or, les neuf manuscrits concernés possèdent des heures de la Vierge à l'usage de Rome et nous montrons ici une corrélation liturgique qui était inconnue.

3.2.2 Heures de la Vierge à l'usage de Metz

L'analyse des psaumes permet de révéler des incohérences au sein d'un même usage liturgique. Au sein de l'office des vêpres de la Vierge, l'usage de Metz tel qu'attesté par les mss. Paris, BnF, lat. 1403 et Karlsruhe, Badische Landesbibliothek, Lichtenthal 55, présente une série de psaumes qui semble rare : 112-116-145-146-147. Nous la trouvons dans deux manuscrits de notre corpus et elle est identique à celle relevée par Victor Leroquais dans des heures de Metz imprimées en 1498¹⁰. Mais les heures de Metz imprimées en 1507¹¹ présentent la série : 109-112-121-126-147 que l'on retrouve dans de très nombreux usages de toute l'Europe¹². Cette observation nouvelle peut donner lieu à des interprétations variées : coexistence de deux séries de psaumes des heures de la Vierge à l'usage de Metz, ou réforme liturgique à l'orée du xvi^e siècle, ou hybridation par l'imprimeur-libraire Simon Vostre officiant à Paris.

⁹ Liste des manuscrits concernés : Genève, Bibliothèque de Genève, Comites Latentes, ms 38 ; Genève, Bibliothèque de Genève, ms lat. 147 ; Genève, Bibliothèque de Genève, ms Comites Latentes 124 ; Chantilly, Bibliothèque et archives du musée Condé, ms 69 (1456) ; Le Mans, Bibliothèque municipale, ms 127 ; Genève, Bibliothèque de Genève, ms Comites Latentes 54 ; Neuchâtel, Bibliothèque publique et universitaire de Neuchâtel, ms AF A28 ; Caen, Musée des Beaux-Arts, ms FMM 241 ; Angers, Musée des Beaux-Arts, ms 2003.1.076.

¹⁰ Cf. Paris, BnF, ms nouv. acq. lat. 3162, n° 62.

¹¹ *Ces presentes heures a lusaige de Metz au long sans requerir*, Paris, Simon Vostre, c.1507 (exemplaire Paris BnF, Vélins-1610). Dépouillement publié par Erik Drigsdahl, « Hore Beate Marie Virginis - Use of Metz », *Late Medieval and Renaissance Illuminated Manuscripts - Books of Hours 1300-1530*, 1999, http://manuscripts.org.uk/chd.dk/use/hv_metz.html.

¹² Usages considérés, selon les relevés d'Erik Drigsdahl : Amiens, Anvers, Arras, Autun, Beauvais, Besançon, Brème, Saint-Donatien de Bruges, Sainte-Gudule de Bruxelles, Camaldules, Carmes, Chartreux, Cisterciens, Clermont, Cologne, Dominicains, Saint-Amé de Douai, Grasse, Hénin-Liétard, Laon, Le Mans, Liège, Lille, Limoges, Lund, Mâcon, Mayence, Meaux, Metz, Nantes, Noyon, Poitiers, Ratisbonne, Saint-Hermès de Renaix, Rome, Saint-Bertin de Saint-Omer, Saint-Denis, Saintes, Saint-Omer, Senlis, Sion, Strasbourg, Saint-Vaast d'Arras, Teutoniques, Théroouanne, Toul, Toulouse, Tournai, Utrecht, Worms, Wurtzburg.

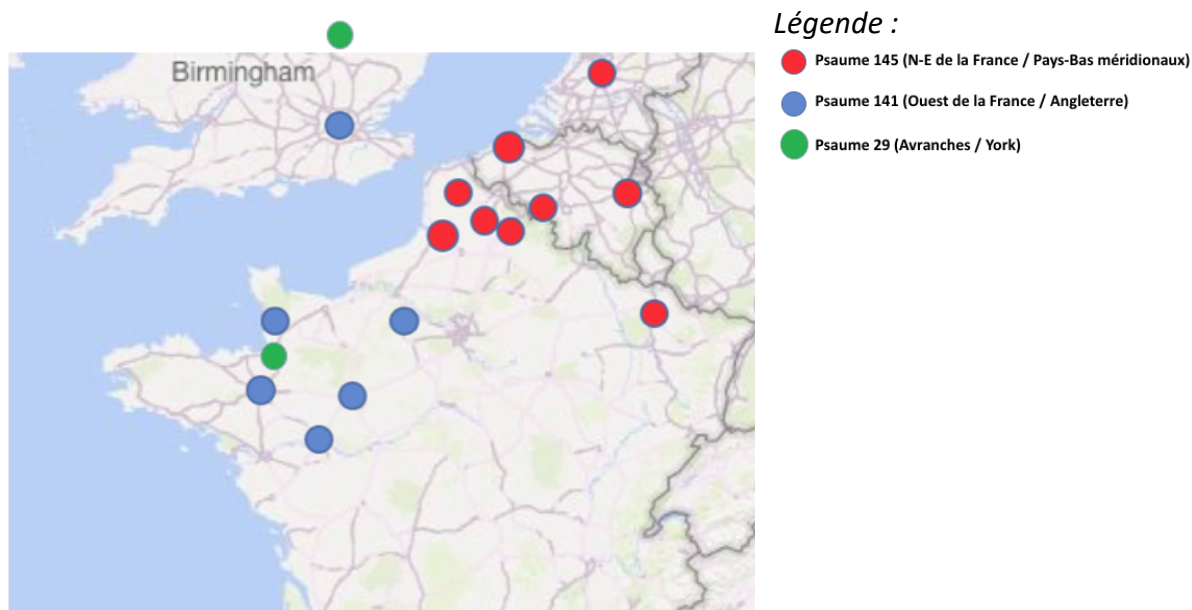
Tableau 1 : Usage liturgique de Metz : psaumes des vêpres des Heures de la Vierge

	Paris, BnF, lat. 1403	Karlsruhe, Badische Landesbibliothek, ms Lichtenthal 55	Heures imprimées de Metz (1498)	Heures imprimées de Metz (1507)
position 1	Psaume 112	112	112	109
position 2	Psaume 116	116	116	112
position 3	Psaume 145	145	145	121
position 4	Psaume 146	146	146	126
position 5	Psaume 147	147	147	147

3.2.3 *Preces* de l'office des morts

Élargissons la perspective, avec l'office des morts. Les vêpres et les laudes y possèdent une structure semblable : à une série de cinq à huit psaumes, identique dans tous les usages, succèdent un cantique du Nouveau Testament, puis une série de pièces formant les « *preces* », et enfin des oraisons. Les *preces* comprennent en général entre un et quatre psaumes. À vêpres, le psaume 145 est le plus courant, et, à laudes, c'est le psaume 129. Notre étude révèle des choix plus rares, locaux ou régionaux : le manuscrit Beaune 55 présente dans les *preces* des vêpres le psaume 142, ce qui est caractéristique de l'usage d'Autun. Par ailleurs, le manuscrit BnF latin 1403, dont l'office des morts est à l'usage de Metz, prescrit dans les *preces* des laudes le psaume 145. Ce choix de psaume ne se retrouve que dans des usages des Pays-Bas méridionaux et du Nord-Est de la France, parfois suivi par le psaume 129 (Arras, Saint-Donatien de Bruges, Thérouanne, Saint-Riquier, Liège, Metz, Utrecht, Saint-Waudru de Mons, Cambrai). Deux autres manuscrits du corpus (Versailles, BM, 845, usage du Mans, et Caen, MBA, Mancel 273, usage de Coutances) ont pour premier psaume des *preces* des laudes le psaume 141. Ce choix ne se retrouve que dans des usages anglais et de l'Ouest de la France (Westminster, Rennes, Coutances, Angers, Le Mans, Évreux) - mais pas dans les usages d'Avranches ni de York qui ont le psaume 29.

Figure 3 : Diffusion des psaumes 145 et 141 dans les *preces* des laudes en position 1



Il apparaît ainsi que, dans le cas de certains usages liturgiques, les psaumes des *preces* des vêpres et des laudes de l'office des morts révèlent par leurs variantes des caractéristiques locales ou des parentés textuelles entre plusieurs usages.

4 Image et texte : perspectives et conclusion

4.1 Psaumes des heures de la Passion et des heures longues de l'Esprit Saint

Au terme de l'analyse automatique du corpus, introduisons une nouvelle notion : celle d'offices rares, tels que les heures de la Passion et les heures longues de l'Esprit Saint, copiés dans un faible nombre de livres d'heures. Ils nous sont révélés par certaines séries de psaumes, inconnues dans les heures de la Vierge et l'office des morts. Dans notre corpus, les heures de la Passion se trouvent dans deux manuscrits, et les heures longues de l'Esprit-Saint dans trois volumes ; le ms. Genève, C.L. 124 contenant ces deux offices rares.

Pour les heures de la Passion, la série de psaumes est la même, alors que les volumes concernés sont à l'usage l'un de Paris, l'autre de Rome.

En revanche, pour les heures longues de l'Esprit-Saint, deux manuscrits dont les heures de la Vierge sont à l'usage de Metz possèdent une série de psaumes qui les distingue d'un troisième volume, dont les heures de la Vierge sont à l'usage de Rome. Les psaumes 127 et 128 sont prescrits aux vêpres et aux complies pour Rome, alors que, pour Metz, l'on prescrit les psaumes 92 et 127. Comme pour les litanies, nous découvrons ainsi des différences liturgiques selon les usages grâce à l'étude des psaumes à des positions où les liturgistes n'attendaient pas de telles différences.

Tableau 2 : Psaumes des Heures longues de l'Esprit-Saint (P : Paris, BnF, ms lat. 1403 ; K : Karlsruhe, Badische Landesbibliothek, ms Lichtenthal 55 ; G : Genève, Bibliothèque de Genève, ms Comites latentes 124)

Manuscrits	Usage (Heures de la Vierge)	Matines / Laudes	Prime	Tierce	Sexte	None	Vêpres	Complies
P	Metz	1	14	53	69	86	92	127
K	Metz	1	14	53	69	86	92	127
G	Rome	1	14	53	69	86	127	128

4.2 Texte et image : le statut des textes et des volumes

De cela découle de nouvelles interrogations. Quel est le statut des textes et des volumes qui les transmettent ? La décoration peut-elle informer notre analyse ?

Les miniatures permettent une première approche du luxe des manuscrits. Dans le corpus de cinquante manuscrits, une miniature apparaît en moyenne toutes les vingt-et-une pages. Parmi les quatre manuscrits comportant des offices rares, trois ont davantage de miniatures que la moyenne, tandis que le quatrième n'en présente pas du tout.

Tableau 3 : Images, pages, miniatures dans le corpus validé et dans les manuscrits à offices rares

	Genève, Comites Latentes 124	Karlsruhe, Lichtenthal 55	Paris, BnF, lat. 1403	Versailles 845	Total CORPUS 50 MSS
Offices rares	Passion Esprit-Saint	Esprit-Saint	Esprit-Saint	Passion	
Images	475	324	370	185	11800
~pages	475	324	370	370	16570
Miniatures	37	0	31	19	767
Pages/miniature	12,8	324	11,9	19,5	21

Une autre approche est donnée par la proportion des pages couvertes par les initiales. Cette mesure, à la fois plus fine car portant sur tout le volume, mais aussi très approximative car dépendant des modalités de numérisation, place clairement les trois premiers manuscrits parmi ceux les plus richement exécutés. Le quatrième est placé en position basse, mais de façon plus intelligible que ne le faisait l'absence complète de miniatures.

Tableau 4 : Surface des initiales reconnues sur la page (manuscrits à office rare et corpus validé)

	Genève, Comites Latentes 124	Karlsruhe, Lichtenthal 55	Paris, BnF, lat. 1403	Versailles 845	Corpus 50 MSS
Offices rares	Passion Esprit-Saint	Esprit-Saint	Esprit-Saint	Passion	
Pages/miniature	12,8	324	11,9	19,5	
Surface des initiales reconnues (px ²) / surface totale des images (px ²)	2,59	1,84	5,75	3,24	2,03 - 1 ^{er} quartile : 1,41 - médiane : 1,87 - 3 ^e quartile : 2,40 - max : 5,75

Un regard sur la numérisation permet d'ailleurs de révéler que des miniatures étaient prévues (cf. Figure 4), mais n'ont pas été réalisées, ce que ne signale pas le catalogue publié¹³. Nous pouvons poser ainsi deux hypothèses dans un rapport nuancé. Les textes rares sont généralement un surcroît de luxe, sortant le manuscrit concerné de la catégorie générique des livres d'heures produits en masse. Mais ils peuvent aussi correspondre à un besoin pragmatique d'un commanditaire ayant à réciter ce texte pour des raisons qui lui sont propres. Le manuscrit se révèle alors certes de moins haut niveau formel que la moyenne, mais, dans le cas qui nous occupe, il n'en est pas pour autant négligé ou un pur manuscrit d'usage.

¹³ Felix Heinzer et Gerhard Stamm, *Die Handschriften von Lichtenthal, mit einem Anhang Die heute noch im Kloster Lichtenthal befindlichen Handschriften des 12. bis 16. Jahrhunderts*, Die Handschriften der Badischen Landesbibliothek in Karlsruhe 11 (Wiesbaden: Otto Harrassowitz, 1987), 155.

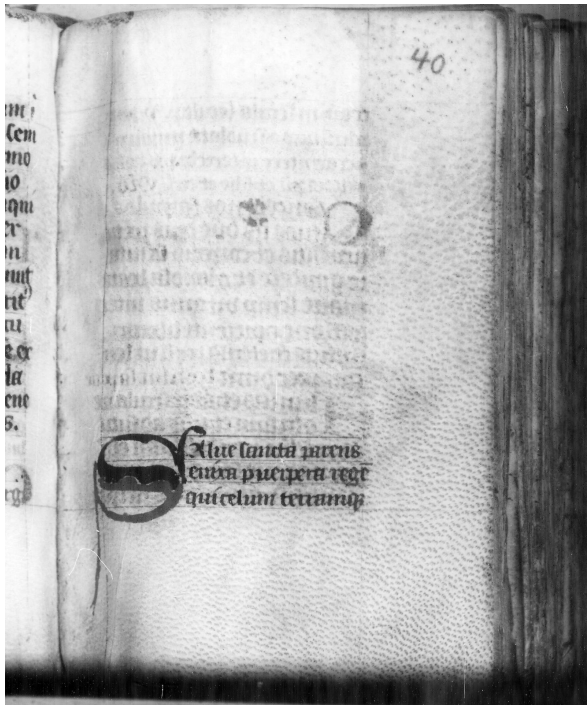


Figure 4: Miniature non réalisée dans le ms. Karlsruhe, Badische Landesbibliothek, Lichtenthal 55, fol. 40r

4.3 Conclusion

Cette conclusion historique nous invite à d'autres réflexions. Tout d'abord sur l'organisation de la recherche et la responsabilité des intervenants. Dans notre projet, la recherche se fonde sur les numérisations disponibles via le protocole IIF et nous ne gardons pas copie des fichiers images, systématiquement affichés sur notre site via IIF ; la base de données Heurist est hébergée par Huma-Num ; il s'agit donc de structures expertes, mais légères, où chaque acteur se concentre sur son domaine, patrimonial, recherche en TAL, en IA ou en SHS, tout en permettant des valorisations vers le grand public : repérage automatique des miniatures, transcription et traduction de documents.

Ensuite, nous nous insérons évidemment dans une approche holistique des sources médiévales, à la fois texte et image, contenu et objet, avec des niveaux multiples de lecture : différents types de décor et différents types de structurations textuelles.

Enfin, il faut y revenir, nos travaux font entrer le Moyen Âge dans le domaine des données massives et changent les modalités d'interrogation et les volumétries dépassent la capacité d'absorption de l'esprit. En lien avec le développement d'intelligences artificielles dédiées, nous avons formalisé des « caractéristiques de niveau moyen » ou *mid-level features*, un « moyen terme » au sens logique qui évite d'affronter directement une masse trop grande en transigeant à un niveau intermédiaire qui permette la vérification et la validation. Ici, par exemple, nous passons du « plein texte », à la fois brut et bruité, à la notion de « texte unitaire », le psaume, qui permet d'explorer des motifs récurrents et des exceptions dans la transmission. Même si l'on revient à une notion heuristique traditionnelle, le passage par l'IA et le traitement massif permettent de réinterroger les données, en croisant des critères sans corrélation préalablement connue, comme la présence de certains textes et la décoration ou la coprésence de textes.

5 Références

- Boillet, Mélodie, Marie-Laurence Bonhomme, Dominique Stutzmann, et Christopher Kermorvant. « HORAE: an annotated dataset of books of hours ». In *the 5th International Workshop on Historical Document Imaging and Processing*, 7-12. Sydney: ACM Press, 2019. <https://doi.org/10.1145/3352631.3352633>.
- Boros, Emanuela, Alexis Toumi, Erwan Rouchet, Bastien Abadie, Dominique Stutzmann, et Christopher Kermorvant. « Automatic Page Classification in a Large Collection of Manuscripts Based on the International Image Interoperability Framework ». In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 756-62. Sydney: IEEE, 2019. <https://doi.org/10.1109/ICDAR.2019.00126>.
- Cottureau, Emilie. « La copie et les copistes français de manuscrits aux XIVe et XVe siècles. Etude sociologique et codicologique ». Université Paris 1 - Panthéon-Sorbonne, 2005.
- Cottureau-Gabillet, Émilie. « Manuscrits de luxe et distinction sociale à la fin du Moyen Âge ». In *Marquer la prééminence sociale*, édité par Jean-Philippe Genet et E. Igor Mineo, 283-301. Histoire ancienne et médiévale. Paris: Éditions de la Sorbonne, 2015. <http://books.openedition.org/psorbonne/3352>.
- Cottureau-Gabillet, Emilie. « Revealing Some Structures and Rules of Book Production (France, Fourteenth and Fifteenth Centuries) ». In *Ruling the Script in the Middle Ages. Formal Aspects of Written Communication (Books, Charters, and Inscriptions)*, édité par Sébastien Barret, Dominique Stutzmann, et Georg Vogeler, 129-63. Utrecht Studies in Medieval Literacy 35. Turnhout: Brepols, 2016.
- Drigsdahl, Erik. « Hore Beate Marie Virginis - Use of Metz ». Late Medieval and Renaissance Illuminated Manuscripts - Books of Hours 1300-1530, 1999. http://manuscripts.org.uk/chd.dk/use/hv_metz.html.
- Hazem, Amir, Béatrice Daille, Dominique Stutzmann, Jacob Currie, et Christine Jacquin. « Réutilisation de textes dans les manuscrits anciens ». In *TALN-RECITAL Conférence sur le Traitement Automatique des Langues Naturelles PFIA 2019*, 417-25. Toulouse, 2019. https://www.irit.fr/pfia2019/wp-content/uploads/2019/07/actes_TALN-RECITAL-courts_CH_PFIA2019-2.pdf.
- Hazem, Amir, Beatrice Daille, Dominique Stutzmann, Christopher Kermorvant, et Louis Chevalier. « Hierarchical Text Segmentation for Medieval Manuscripts ». In *Proceedings of the 28th International Conference on Computational Linguistics*, 6240-51. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020. <https://www.aclweb.org/anthology/2020.coling-main.549>.
- Heinzer, Felix, et Gerhard Stamm. *Die Handschriften von Lichtenthal, mit einem Anhang Die heute noch im Kloster Lichtenthal befindlichen Handschriften des 12. bis 16. Jahrhunderts*. Die Handschriften der Badischen Landesbibliothek in Karlsruhe 11. Wiesbaden: Otto Harrassowitz, 1987.
- Julien, Octave. « Délier, lire et relier ». *Hypotheses* 19, n° 1 (13 octobre 2016): 211-24.
- . « Lire, écrire, relier : la composition des recueils vernaculaires français et anglais de la fin du Moyen âge ». <http://www.theses.fr>. Thesis, Paris 1, 2016. <http://www.theses.fr/2016PA01H014>.
- Ornato, Ezio. *La face cachée du livre médiéval : l'histoire du livre vue par Ezio Ornato, ses amis et ses collègues*. I libri di Viella 10. Roma: Viella, 1997.
- Ornato, Ezio, et Carla Bozzolo. *Pour une histoire du livre manuscrit au moyen âge : trois essais de codicologie quantitative*. Paris: Ed. du C.N.R.S., 1980.
- Ruzzier, Chiara. « Continuité et rupture dans la production des bibles au XIII^e siècle ». In *Comment le Livre s'est fait livre. La fabrication des manuscrits bibliques (IVe-XVe*

siècle), édité par Chiara Ruzzier et Xavier Hermand, 40:155-68. Turnhout: Brepols Publishers, 2015. <https://doi.org/10.1484/M.BIB-EB.5.105426>.

Stutzmann, Dominique. « Résistance au changement ? Les écritures des livres d'heures dans l'espace français (1200-1600) ». In « *Change* » in *medieval and Renaissance scripts and manuscripts. Proceedings of the 19th Colloquium of the Comité international de paléographie latine (Berlin, 16-18 September, 2015)*, édité par Eef Overgaauw et Martin J. Schubert, 97-116. *Bibliologia* 50. Turnhout: Brepols, 2019.

Stutzmann, Dominique, et Louis Chevalier. « Books of hours as text compilations in the Low countries ». In *Networks of Manuscripts, Networks of Texts*. Amsterdam, Netherlands: Evina Steinová (Huygens ING), 2020. <https://hal.archives-ouvertes.fr/hal-03112110>.

Stutzmann, Dominique, Jacob Currie, Béatrice Daille, Amir Hazem, et Christopher Kermorvant. « Integrated DH. Rationale of the HORAE Research Project ». Utrecht, 2019. <https://dev.clariah.nl/files/dh2019/boa/0192.html>.