

Le livre médiéval au regard des méthodes quantitatives

Paris, 19-23 juin 2023

Programme de l'école d'été

Direction : Christine Bénévent, Émilie Cottereau-Gabillet,
Catherine Rideau-Kikuchi et Dominique Stutzmann

Coordination scientifique et pédagogique :
Émilie Cottereau-Gabillet et Octave Julien

Organisé par le Laboratoire de Médiévisstique
occidentale de Paris, l'Institut de Recherche et
d'Histoire des Textes, l'École nationale des chartes et
le laboratoire Dynamique patrimoniales et culturelles

Avec le soutien du LabEx Histoire et anthropologie
des savoirs, des techniques et des croyances



Lundi 19 juin Sorbonne

L'état de l'art

9h30 – Accueil

10h – Le livre médiéval au regard des méthodes quantitatives : un état de l'art

par Émilie Cottereau-Gabillet et Octave Julien

11h15 – La philologie computationnelle pour le livre médiéval : entre données et modèles

par Jean-Baptiste Camps

Ateliers d'application

14h – Observer les livres pour quantifier : les manuscrits

atelier animé par Émilie Cottereau-Gabillet

Cet atelier a pour objectif de montrer, pour les différents éléments constitutifs d'un livre manuscrit (support, structure matérielle, réglure et mise en page, aspects périgraphiques, décoration) ainsi que pour les données relatives au colophon (forme matérielle, structure syntaxique) ou issues de ce dernier (lieu et date de production, commanditaire, copiste, etc.), la nature des paramètres susceptibles de faire l'objet d'une quantification au regard de leur exploitation. Dans cette perspective on reviendra, pour chacun de ces éléments, sur : la nature des observations qui peuvent être menées et les protocoles de relevés qui y sont associés ; la manière dont les données peuvent être formalisées ; les usages statistiques qui peuvent être envisagés ; et enfin les perspectives d'analyses historiques liées à cette utilisation des méthodes statistiques. L'enjeu est ici de faire le lien entre l'observation des volumes et leur interprétation, dans une démarche qui doit placer la questionnement (des données comme des résultats) au cœur du propos.

17h – Discussion avec les participantes et participants autour de leurs projets

9h – Observer les livres pour quantifier : les incunables

atelier animé par Christine Bénévent

Les incunables, livres imprimés avant 1501, constituent un corpus quantitativement bien défini grâce aux nombreux catalogues et entreprises bibliographiques qui ont permis de les identifier et de les décrire. Il s'agit toutefois d'un corpus hétérogène, qui ne peut s'appréhender qu'en tenant compte des enjeux techniques propres à l'imprimerie mais aussi d'une dialectique d'imitation / opposition par rapport aux manuscrits.

Dans la continuité de la séance consacrée au livre manuscrit, il s'agira de se pencher sur les éléments matériels constitutifs des incunables : support et structure matérielle (cahiers, formats) certes, mais aussi et surtout choix typographiques, mise en pages et inachèvement du produit, à l'origine conçu pour être complété par des éléments manuscrits. Il s'agira donc d'étudier ce faisant l'évolution de la structure et de la physionomie du nouveau produit au cours de la cinquantaine d'années que couvre la période incunable, mais aussi de constater des pratiques différentielles selon la nature des textes et le rapport avec l'existant manuscrit. Le panorama proposé, centré sur les caractéristiques matérielles des incunables, permettra d'insister sur les ruptures et continuités par rapport aux manuscrits, en essayant d'en expliquer les tenants et les aboutissants.

13h30 – Papier et filigranes au regard des méthodes quantitatives : l'exemple des manuscrits sur papier de Saint-Omer (XIV^e-XV^e siècle)

atelier animé par Marlène Helias-Baron

L'atelier propose une initiation à l'étude des manuscrits sur papier et de leurs filigranes en s'appuyant sur un corpus de 153 manuscrits conservés à la Bibliothèque d'agglomération du Pays de Saint-Omer, produits entre le XIV^e et le début du XVI^e siècle, tous établissements religieux confondus. La BAPSO dispose de plus de 400 manuscrits sur papier des XIV^e-XVIII^e siècles.

Après une courte présentation des usages du papier en Occident depuis la fin du XIII^e siècle et un l'état de l'art sur les filigranes, seront exploitées les données sur les filigranes des manuscrits audomarois disponibles sur la Bibliothèque virtuelle des manuscrits médiévaux et sur le portail WZIS-Wasserzeichen-Informationssystem (Stuttgart, Landesarchiv Baden-Württemberg). À partir d'un corpus test de filigranes issus de manuscrits datés, les stagiaires seront conduits à dater des manuscrits non datés en comparant leurs filigranes avec ceux des manuscrits datés et avec les données issues des bases de données en ligne (Briquet, WZIS).

16h – Présentation d'une sélection de manuscrits et d'incunables de l'École nationale des chartes et de l'IRHT

par Christine Bénévent et Dominique Stutzmann

9h – De la modélisation à l’exploitation : les manuscrits

atelier animé par Dominique Stutzmann

Nous étudierons au cours de cet atelier les problématiques inhérentes à la formalisation des caractéristiques matérielles des manuscrits et à la formation d’un corpus concernant le livre manuscrit au Moyen Âge. À cette fin, nous utiliserons les données collectées au sein du projet HORAE (Hours - Recognition, Analysis, Editions) et portant sur un ensemble de 9000 livres d’heures manuscrits. Les livres d’heures, copiés à partir du XIII^e siècle pour l’usage religieux des fidèles et conservés en grand nombre dans les institutions patrimoniales et les collections privées, constituent un précieux témoignage des traditions liturgiques médiévales et des pratiques de dévotion et des représentations sociales de leurs commanditaires.

Les participant·e·s seront mené·e·s pas à pas pour découvrir l’utilisation de fonctions d’analyse statistiques pour exploiter les données (Excel et tableau croisé dynamique), pour régulariser le corpus afin de réaliser de meilleurs analyses, notamment grâce à l’outil OpenRefine, et enfin dans la création d’une base de données en ligne pour visualiser, analyser, publier et pérenniser les données de la recherche.

13h30 – De la modélisation à l’exploitation : les incunables

atelier animé par Catherine Rideau-Kikuchi

L’objectif de cet atelier est de proposer des éléments méthodologiques relatifs au traitement des données concernant le livre incunable. En guise d’introduction, on réfléchira d’abord à la manière dont certaines bases de données ont été construites, en s’appuyant notamment sur l’*Incunabula Short Title Catalogue*, *Material Evidence in Incunabula* et le *Gesamtkatalog der Wiegendrucke*. Il s’agira de réfléchir aux apports et limites de telles bases dans le cadre d’une approche quantitative systématique.

Dans un second temps, cet atelier visera à aborder la manière dont l’analyse de réseau peut être utilisée pour l’étude des imprimés anciens. Nous verrons d’un point de vue théorique ce que cette approche peut (ou ne peut pas) apporter à l’étude des imprimés anciens et les problèmes méthodologiques qu’elle pose. Si les réseaux peuvent être utilisés à des fins purement visuelles, l’analyse de réseau repose sur l’analyse des formes des objets, que l’on peut examiner à partir de certains outils statistiques simples, mais qui ne peuvent pas non plus être utilisés dans toutes les circonstances. Nous aborderons en particulier les notions de connexité, de cœur et de centralité.

Nous réaliserons enfin une première prise en main du logiciel Gephi à partir d’une base de données préétablie, afin de réaliser quelques manipulations simples.

17h – Discussion avec les participantes et participants autour de leurs projets

9h – Introduction à l'analyse statistique : l'exploration d'un corpus

atelier animé par Octave Julien

Cet atelier consiste en une introduction aux concepts et outils statistiques utiles à la codicologie quantitative. On travaillera pour cela sur un corpus d'environ 200 recueils copiés en France et en Angleterre entre le XIII^e et le XV^e siècle, à la forme et au contenu variés. L'hétérogénéité de ce corpus suppose une première phase d'exploration des données pour en mesurer la diversité et les possibles biais. On verra quels indicateurs statistiques (moyenne, médiane, quartiles, écart type) et quelles représentations graphiques (diagrammes en barres, histogrammes, boxplots, courbes de densité) permettent d'appréhender un tel corpus. Dans le prolongement des ateliers précédents, on reviendra aussi sur les outils permettant d'analyser une corrélation (tris croisés, comparaison d'indicateurs par sous-populations). Les méthodes d'analyses factorielles seront enfin abordées afin de construire une typologie de manuscrits.

Cet atelier se fera sur R, un logiciel d'analyses statistiques puissant, libre et gratuit. On verra comment l'utiliser pour importer ses données, les explorer, les transformer grâce à quelques méthodes simples, et produire des tableaux et graphiques.

13h30 – La construction d'un protocole d'analyse : étude et interprétation de corrélations statistiques

atelier animé par Octave Julien

En partant des analyses de Carla Bozzolo et Ezio Ornato concernant le lien entre format des manuscrits et disposition du texte, on cherchera à comprendre ce qui a déterminé les choix en matière de mise en page et de décoration dans le corpus de manuscrits utilisé le matin. La nature diversifiée de ce corpus conduit à étudier d'autres corrélations et à proposer d'autres interprétations pour rendre compte de ces choix fonctionnels et esthétiques. On réfléchira à la manière de quantifier les éléments décoratifs des livres manuscrits. Ce sera également l'occasion de mobiliser les méthodes et concepts vus le matin, et d'aller plus loin dans la démarche quantitative en utilisant des tests statistiques destinés à confirmer ou infirmer ces hypothèses (tests du χ^2 , coefficient R^2 de corrélation linéaire).

17h – Discussion avec les participantes et participants autour de leurs projets

9h – La paléographie à l’heure de l’intelligence artificielle : Analyse d’images

atelier animé par Dominique Stutzmann

L’objectif de cet atelier est de mettre en œuvre des technologies d’analyse d’image par ordinateur fondées sur l’intelligence artificielle et de réfléchir à l’utilisation et à la signification historique des résultats générés suite à l’apprentissage machine. L’atelier s’ouvrira sur une présentation des enjeux de la paléographie comme discipline historique étudiant l’écriture dans ses formes et ses emplois sociospatiaux.

Plus particulièrement, le cas de l’écriture des livres d’heures permettra de faire le lien avec l’atelier du mercredi matin. Dans la partie pratique, les participant·e·s manipuleront des images disponibles selon le protocole IIIF (visualisation et import). Nous nous intéresserons ensuite à l’analyse d’image d’écritures médiévales et à la forme et à la signification des résultats des traitements automatiques, notamment ceux produits dans le cadre des compétitions internationales. Sur la base des manipulations et observations des participant·e·s, nous discuteront les enjeux de la visualisation de données et de classification dans les différentes tâches de la paléographie (paléographie d’expertise), en particulier la reconnaissance des mains.

Enfin les participant·e·s verront différentes plateformes et logiciels de traitement des documents manuscrits et d’analyse d’image, en particulier pour l’analyse du décor et de la mise en page.

13h30 – La paléographie à l’heure de l’intelligence artificielle : Transcription automatique

atelier animé par Jean-Baptiste Camps

Cet atelier présentera, de manière théorique et surtout pratique, les tenants et aboutissants de la transcription des manuscrits assistée par l’intelligence artificielle. Y seront abordées les questions liées à l’analyse de mise en page, aux choix de transcription, à la préparation de données d’entraînement, et à l’utilisation d’algorithmes de reconnaissance des écritures (*handwritten text recognition*, ou HTR). Les sujets de l’interopérabilité et du partage de données seront également évoqués.

L’atelier consacrera une large part à la pratique, depuis la récupération d’images (via le protocole IIIF), jusqu’à l’utilisation d’interfaces pour la production ou la correction de données et l’entraînement ou l’application de modèles. En particulier, les travaux pratiques seront réalisés avec les logiciels eScriptorium et Kraken, et des modèles déjà produits pour le latin et le français médiévaux (notamment les modèles CREMMA).

17h – Discussion avec les participantes et participants autour de leurs projets, bilan de l’école d’été et remise des certificats de participation

Inscription

L'accueil à l'école d'été est limité à 20 inscrits.

Formulaire d'inscription en ligne : <https://forms.gle/yQEZoHy1RokWTwJc7>

Clôture des inscriptions : 7 avril 2023

Sélection des dossiers : 17 avril 2023

Cette école d'été s'adresse en priorité aux chercheurs en formation (masterants, doctorants et post-doctorants). Des bourses destinées à rembourser des frais de transport et de logement pourront être distribuées, à hauteur d'un plafond qui sera déterminé lors de l'examen des dossiers, en fonction du nombre et de l'origine géographique des candidats sélectionnés.