

éphéméride

Le magazine de l'École Pratique des Hautes Études – PSL

DOSSIER SPÉCIAL

Humanités: quand l'IA transcrit les textes anciens.



sommaire

Éditorial du Président	p.3
Les actualités de l'EPHE – PSL	p.4
Les actualités des instituts	p.6
Le Fonds de dotation de l'EPHE : de l'ambition à l'impact concret	p.10



Dossier spécial : Humanités : quand l'IA transcrit les textes anciens. p.16

Introduction : eScriptorium	p.18
eScriptorium : une aventure collective au croisement des humanités et de l'intelligence artificielle.	p.20
Portrait : Robin Tissot, le concepteur d'eScriptorium	p.22
Intelligence artificielle et Sciences des données pour les Humanités dans la stratégie de l'Université PSL.	p.25
Portrait : Benjamin Kiessling, l'architecte IA de l'EPHE – PSL.	p.26
Biblissima : observatoire des cultures écrites anciennes	p.28
Dunhuang : déjouer la loi de puissance	p.30
MiDRASH : reconstruire la culture juive médiévale écrite.	p.32
CATMuS : lire les langues de l'Europe au Moyen-Âge	p.34
Syriac Transcribathon : un exercice collaboratif pour décrypter des textes syriaques	p.35
L'initiative Open Islamicate Texts (OpenITI) : étudier les textes arabes.....	p.36
ERC PRIMA : les manuscrits italiens du XVI ^e au XVIII ^e siècle ..	p.37
DeLiCaTe : déchiffrer le patrimoine écrit du Caucase ancien ..	p.38
Conclusion : « Aussi ouvert que possible, aussi fermé que nécessaire »	p.39
Lexique	p.40
 Nos publications	p.42
Sur le terrain	p.44

Éphéméride

Le magazine de l'École Pratique
des Hautes Études – PSL

Numéro 17

hiver 2025

ISSN 2679 - 7631

Directeur
de la publication
Michel Hochmann

Rédacteur en chef
Fabrice Auffret van der Kemp

Création graphique
Jean-Baptiste Lasserre

Remerciements :

Colin Brisson
Marc Bui
Thibault Clerice
Sébastien Couette
Serena Crespi
Jost Gippert
Benjamin Keissling
George Kiraz
Mathew Miller
Catherine Paul
Elena Pierazzo
Stéphanie Plenquette
Christine Roughan
Sarah Savant
Peter Stokes
Daniel Stökl Ben Ezra
Robin Tissot
Anne-Marie Turcan-Verkerk
Stéphane Verger

Couverture

Weißenauer Passionale,
Coligny,
Fondation Martin Bodmer,
Cod. Bodmer 127

Impression
Paragon

École Pratique des
Hautes Études – PSL
4 rue Ferrus 75014 Paris
www.ephe.psl.eu

L'École Pratique des Hautes Études est, depuis ses origines, un centre majeur pour ce qu'on appelle les sciences de l'érudition. On a pu considérer, à un certain moment, que ces disciplines avaient été largement remises en cause par les nouvelles approches des sciences sociales. Or, le développement spectaculaire des humanités numériques et, en particulier, de l'intelligence artificielle ont, ces dernières années, complètement bouleversé ce paysage et démontré le dynamisme et la validité des spécialités qui ont toujours été les nôtres, le déchiffrement des langues et des écritures rares ou la constitution de corpus, pour prendre quelques exemples. L'École a su développer des programmes originaux dans ce domaine et a joué un rôle pionnier. Comme l'indique Marc Bui, il faut rappeler que Denis Pelletier et Hubert Bost, deux de mes prédécesseurs, ont su fédérer les forces et lancer des initiatives fondamentales pour le développement de ces recherches. Je pense, en particulier, à la création de l'Initiative de recherches interdisciplinaires et stratégiques Scripta, développée au sein de notre Université PSL, qui a permis la naissance de la plateforme eScriptorium, dont le succès international se trouve illustré par les multiples programmes de recherche auxquels elle a contribué dans les domaines les plus divers, des manuscrits hébreux aux manuscrits médiévaux, à Princeton, à l'Université du Maryland, et, évidemment, à l'EPHE. Grâce en partie à eScriptorium, nous avons pu aussi obtenir de grands contrats de recherche européens, comme l'ERC Synergy MiDRASH.

Au cours des mêmes années et grâce à l'action d'Anne-Marie Turcan, c'est aussi au sein de l'École qu'a été créé l'Equipex (équipement d'excellence) Biblissima, qui structure des millions de données relatives au textes anciens et qui va sans doute connaître bientôt une ultime métamorphose en devenant une infrastructure de recherche, l'une des très rares dans le domaine des sciences humaines et sociales. La prise en compte de l'importance du développement des humanités numériques a abouti à la création d'une direction d'études spécifique dans ce domaine, qu'occupe Peter Stokes depuis sa création, et, lui-même et ses collègues Daniel Stökl Ben Ezra et Marc Bui ont joué un rôle capital dans le développement d'eScriptorium. Ils ont pu s'appuyer sur des collaborateurs d'exception, Benjamin Kiessling et Robin Tissot, dont vous découvrirez les portraits dans ce numéro d'Éphéméride. La recherche, à l'École, est toujours étroitement liée à la formation, et nous avons participé à la mise en place du master Humanités numériques porté par l'École nationale des Chartes au sein de PSL. Ce croisement entre des disciplines rares et des technologies d'avant-garde démontre la fécondité de nos approches et la capacité d'innovation de nos directeurs d'études et maîtres de conférences, et je m'en réjouis. Nous travaillons à présent pour que ces avancées trouvent pleinement leur place dans PSL et, notamment, au sein de l'institut Prairie qui regroupe les forces de notre université dans le domaine de l'intelligence artificielle.



Michel Hochmann,
Président
de l'EPHE – PSL

Je suis en tout cas très reconnaissant à tous ceux qui ont contribué à l'élaboration de ce dossier thématique qui témoigne, encore une fois, du dynamisme de notre École. Ce numéro d'Éphéméride donne d'ailleurs bien d'autres témoignages de ce dynamisme, avec le développement du mécénat, les actualités de la recherche (la découverte opérée dans le cadre de la mission archéologique de Tanis, les succès remportés à l'ERC et à l'ANR), les publications. Je regarde donc l'avenir avec confiance (malgré les difficultés budgétaires que nous traversons comme la plupart des établissements d'enseignement supérieur), et j'adresse mes vœux les plus chaleureux à notre École et à tous ceux qui la font vivre, enseignants-chercheurs, personnels administratifs, étudiants et auditeurs.

Michel Hochmann



École Pratique
des Hautes Études

PSL

édito

Alessandro Buccheri lauréat de l'ERC Starting Grants 2025

Le Conseil européen de la recherche (ERC) a sélectionné le projet POLYBOTA dans le cadre des Appels ERC-WP-2025 en septembre dernier.

Alessandro Buccheri, chercheur rattaché à l'UMR ANHIMA, en est un des lauréats, après avoir bénéficié du programme ANR Access ERC, depuis 2023, auprès de notre institution. Le projet qu'il porte s'intitule « POLYBOTA. Polytheism and Botany: An Experimental Approach to the Environmental Dimensions of Ancient Greek Religion ».

Il a pour objectif de mettre en lumière la manière dont les textes de l'Antiquité grecque représentent les relations entre les dieux et l'environnement naturel. C'est le quatrième projet ERC distinguant l'EPHE – PSL depuis 2024.



actualités

Colloque « Regards sur une mémoire savante »

Les 19 et 20 novembre 2025, l'EPHE – PSL a organisé un colloque exceptionnel au Collège de France et à l'IEA de Paris intitulé « Les archives et les collections de l'École pratique des hautes études : regards sur une mémoire savante ».

Ces deux journées, qui ont rassemblé des professionnels des archives comme des bibliothèques et des chercheurs issus de nombreuses disciplines, en sciences humaines comme en Sciences de la Vie et de la Terre, visaient à l'étude des fonds relatant l'histoire institutionnelle et intellectuelle de l'École. Elles ont rencontré un vif succès.



© Jean-Baptiste Lasserre

Importante découverte dans la nécropole royale de Tanis (Égypte)

La Mission Française des Fouilles de Tanis (École Pratique des Hautes Études – PSL), dirigée par le professeur Frédéric Payraudeau (Sorbonne Université) et le Dr Ahmed Nakshara (Ain Shams University), a découvert un ensemble de statuettes funéraires dans la nécropole royale de Tanis (San el-Hagar). Cette capitale de l'Égypte ancienne est déjà connue pour les trésors qui y ont été trouvés en 1939, joyaux du Musée égyptien de la place Tahrir.

Des statuettes conservées dans la boue des millénaires

Dans une salle de la tombe du roi Osorkon II (22^e dynastie, vers 850 avant J.-C.), l'équipe a trouvé un ensemble de 225 statuettes funéraires

royales (ouchebtis) à leur emplacement d'origine, près d'un sarcophage anonyme en granit rose. C'est en nettoyant le sol de la boue qui s'y était accumulée depuis des millénaires que cette découverte inattendue a été rendue possible. Ces statuettes de faïence appartiennent au pharaon Chéchong III (830-791 avant J.-C.), bien connu pour ses constructions à Tanis. Leur identification confirme donc que ce pharaon est bien le propriétaire de ce sarcophage et des vestiges du mobilier funéraire trouvé à côté (poterie, perles). La mission a également découvert de nouvelles inscriptions dans la chambre funéraire qui confirment cette identification.



© MFFT - Simone Nannucci

Une première en 80 ans

C'est la première fois depuis 1946 qu'une découverte aussi importante est faite dans les tombes royales de Tanis. La mission française y poursuit son programme d'étude et de conservation depuis 1929, en étroite collaboration avec le Ministère égyptien du Tourisme et des Antiquités et l'équipe égyptienne locale.



Retrouvez toutes les actualités de l'EPHE – PSL sur le site web

Ouverture de l'entrepôt Recherche Data EPHE – PSL

L'École Pratique des Hautes Études poursuit son engagement en science ouverte et complète son offre de services en ouvrant l'entrepôt institutionnel Recherche Data EPHE – PSL. Ouvert au sein de Recherche Data PSL, il propose une solution souveraine de partage et de diffusion des données et corpus de recherche produits par ou en collaboration avec la communauté scientifique de l'École. Cet entrepôt s'adresse principalement aux chercheurs qui ne disposent pas d'un entrepôt disciplinaire ou thématique reconnu.

Vietnamica – bilan du projet ERC

Lancé en novembre 2019, le projet ERC Advanced Grant, Historical Research and Data Processing on Ancient Vietnamese Inscriptions (Vietnamica) s'est achevé en octobre 2025. Porté par Philippe Papin, il s'est consacré à l'étude historique et philologique des stèles de donation érigées dans les villages au Vietnam entre le XVI^e et le XX^e siècle.

Ces stèles, rédigées en caractères chinois et vernaculaires vietnamiens fourmillent d'informations qui permettent d'écrire une histoire locale et populaire du Vietnam rural, ce que le manque de documentation rendait impossible jusqu'à présent. Une partie des résultats scientifiques du projet sont accessibles depuis le site Internet du projet.

Le projet, coordonné par l'EPHE – PSL à Paris, avait pour opérateur, au Vietnam, l'Agence Universitaire de la Francophonie et, pour partenaires scientifiques, l'Université vietnamienne des Sciences Sociales et Humaines ainsi que l'Institut Han-Nôm (sino-vietnamien) relevant de l'Académie des Sciences Sociales du Vietnam.



© Philippe Papin

Un supergène, liant sédentarité et coopération, découvert chez une petite fourmi européenne

Une équipe internationale de chercheurs menée par Stefano Mona et Clémence Doums (EPHE – PSL) a pu mettre en évidence un supergène – section de chromosome transmis à la descendance – déterminant le comportement de coopération et de sédentarité des reines chez une petite fourmi européenne : *Myrmecina graminicola*.

Cette découverte a fait l'objet d'une publication dans la revue *Current Biology*, ce

20 novembre. C'est la première fois que sont liés de façon inéluctable coopération et sédentarité. L'exploration des génomes d'espèces proches de *M. graminicola* permettra de comprendre les mécanismes génétiques à l'œuvre dans cette association. (associations étudiantes) sont également renforcées par la création d'un poste dédié.



© Antoine Taupenot

L'EPHESE a renouvelé son équipe pour l'année 2025-2026 !

Cette année marque un tournant historique pour l'EPHESE, le bureau des étudiants de l'EPHE – PSL. Pour la première fois, les 3 sections de l'école sont représentées au sein du bureau (Sciences de la Vie et de la Terre, Sciences historiques et philologiques et Sciences des religions) créé en 2008 pour rassembler et représenter les étudiants de l'EPHE – PSL. Les relations de l'EPHESE avec les partenaires de PSL (associations étudiantes) sont également renforcées par la création d'un poste dédié.



De la découverte des langues rares (ILARA) à la présence des faits religieux dans une société sécularisée (IREL), l'accompagnement des personnes âgées (ITEV), les trois instituts de l'EPHE – PSL abordent un grand nombre de thématiques au travers de formations accessibles à tous.

actualités des instituts

IREL

Les Mercredis de l'IREL

Environ une fois par mois, l'IREL organise une rencontre autour d'un livre paru récemment avec son autrice ou son auteur en débat avec d'autres spécialistes.

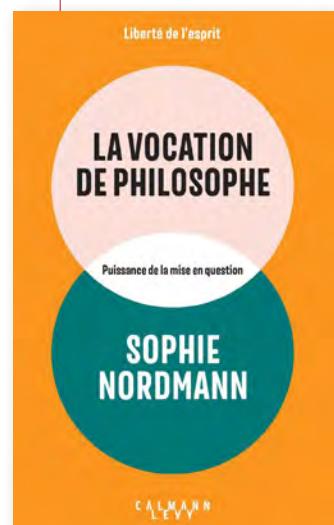
Un moment d'échanges enrichissant dont voici la programmation hivernale :

Le 11 février 2026 avec **Sophie Nordmann** (EPHE – PSL, GSRL) autour de son livre *La vocation de philosophe, Puissance de la mise en question* (Calmann-Lévy, 2025).

Le 11 mars 2026 avec **Sébastien Fath** (CNRS, GSRL) autour de son livre *Croire et prospérer. Théologie(s) de la prospérité en perspective postcoloniale. Redéploiements francophones depuis les années 1950*, à l'IPT Paris en débat avec Gilles Vidal.

Le 15 avril 2026 avec **Guillaume Cuchet** (Université Paris 1 Panthéon Sorbonne) autour de son livre à paraître sur *La religion des morts*, à la bibliothèque du Saulchoir.

Le 20 mai 2026 avec **Xavier Boniface** (Université de Picardie) autour de son livre à paraître sur la religion en Ukraine (Hermann, 2026).





Libre de droit

ILARA

Cycles de conférences de langues rares

Depuis 2020, l'ILARA dispense des formations innovantes et variées à des langues peu décrites, peu enseignées ou peu parlées, parfois en danger d'extinction, ou déjà éteintes.

Cette offre n'a cessé de se diversifier d'année en année, reflétant l'actualité des recherches et l'intérêt croissant du public, issu ou non du monde universitaire, pour ce domaine. En 2025-2026, pas moins de vingt-six langues et familles linguistiques sont enseignées à l'ILARA. Le choix affirmé d'un enseignement hybride favorise le suivi de participants dans toute la France, comme dans le reste du monde.

Ouverts à un large public, en présentiel et à distance, les cycles de conférences de l'ILARA sont filmés et retransmis en ligne. En 2024, le soutien de la Fondation Polonsky a permis l'organisation du cycle « L'araméen dans tous ses états », coordonné par Muriel Debié (EPHE – PSL). En 2025, le cycle coordonné par Andréas Stauder (EPHE – PSL) « Déchiffrements en cours : les écritures anciennes » a bénéficié du soutien de la Fondation Singer-Polignac et de la Fondation Jan Michalski. Celles et ceux qui souhaitent réécouter les conférences peuvent y accéder depuis le site et la chaîne YouTube de l'ILARA. Plusieurs conférences seront prochainement mises en ligne, dont « Quelle herméneutique pour l'étrusque ? », par Jean Hadas-Lebel et « The 'Minoan' Linear A Script of Bronze Age Crete: Is decipherment the only goal? », par Ester Salgarella.

Libre de droit

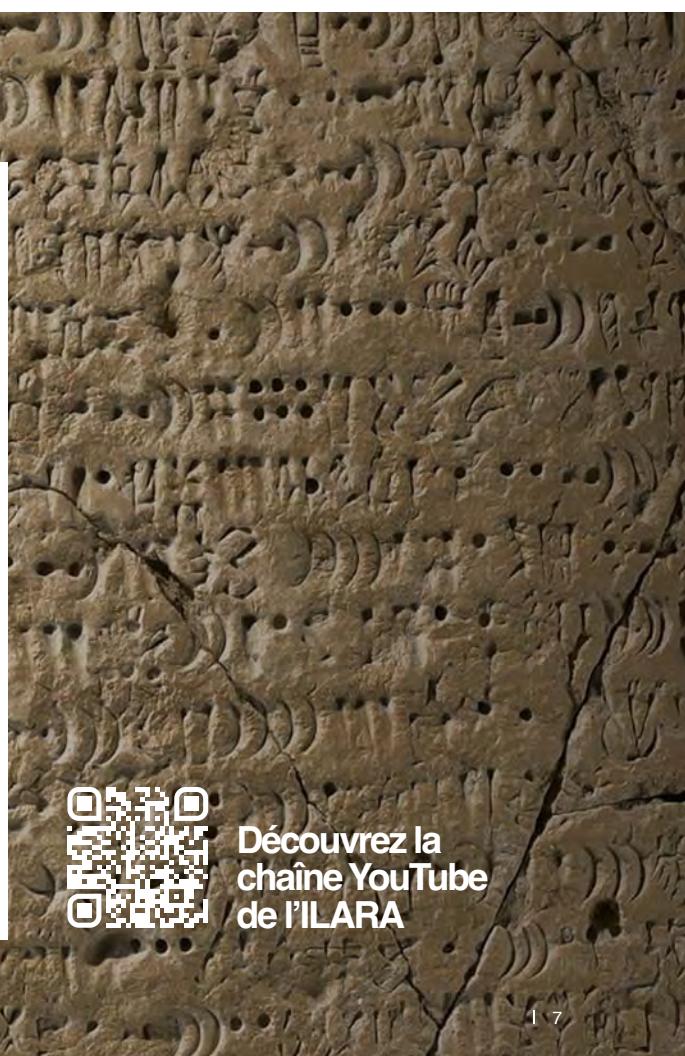
ILARA

ILARA en ligne

Depuis sa création par Amina Mettouchi en 2020, « l'ILARA en ligne » s'est attaché à communiquer à grande échelle sous la forme de vidéos, de posts, ou de billets de blogs. En parallèle, avec les formations et événements mis en place par l'ILARA, ce format digital bilingue s'attache à diffuser la connaissance des langues rares auprès du plus grand nombre, contribuant ainsi au rayonnement de l'EPHE – PSL.

La chaîne YouTube de l'ILARA compte à ce jour 5 300 abonnés et 100 vidéos en ligne. Elle s'articule autour de cinq playlists : les *Ouvertures* (de courtes vidéos d'introduction aux formations dispensées par l'ILARA), les *Cycles* (des séries de conférences sur des thèmes ou des groupes de langues), les *Cultures* (des entretiens en direct précédés par la diffusion de documentaires anthropologiques réalisés par les invitées), les *Locuteurs* (des entretiens en différé mettant en valeur le rapport de locuteurs à leur langue menacée), et les *Chercheurs* (des entretiens ou reportages présentant la recherche sur les langues rares à la société civile).

Cette chaîne, au programme diversifié, s'accompagne d'un ensemble de réseaux sociaux servant à relayer les contenus de médiation et les informations sur nos activités et formations. Nos comptes Facebook, Instagram et LinkedIn témoignent d'une véritable stratégie de transmission, de médiation et de communication, sur laquelle reposent le dynamisme et la créativité de l'ILARA.



Découvrez la chaîne YouTube de l'ILARA



Conférence de restitution des résultats de la démarche de «Design citoyen», Entre-Vignes, 2024

© ITEV

ITEV

Le programme « Design Citoyen » : de la démarche participative au débat académique

L’Institut Transdisciplinaire d’Étude du Vieillissement (ITEV) participe depuis 2024 au programme de recherche « Design Citoyen » consistant au déploiement d’une méthode de diagnostic participatif afin d’accompagner la formulation de solutions et l’amélioration des conditions du vieillissement dans les territoires.

Ce programme repose sur l’implication active des citoyens âgés, accompagnés et outillés par les chercheurs en charge de la mise en œuvre de la méthode, dans un processus de dialogue et d’élaboration de solutions qui répondent à la diversité de leurs besoins et expériences du vieillissement. Les recherches menées en 2024 sur 6 territoires différents en France a conduit à l’organisation d’un symposium sur ce sujet et à la présentation de nos résultats (« *Entre-Vignes : Comment inclure les aînés en milieu rural ?* ») lors d’une communication orale au 7^{ème} Colloque international du REACTIS (Réseau d’Études Internationales sur l’Âge, la CitoyenneTé et l’Intégration Socio-économique) en juin 2024 à Montréal, ainsi qu’à la participation aux journées d’étude Cap’Lab organisées par la Chaire internationale Sociétés Inclusives et Avancée en Age (SIAGE), la communauté épistémique Cap Droits et le programme PPR AUVI sur le thème « *Regards croisés sur les méthodes participatives dans le champ de l’avancée en âge* » (27-28 mars 2025, Metz). Ces travaux de recherche feront l’objet de la rédaction d’un ouvrage collectif (publication prévue en 2026).

ITEV

École d’Automne de l’ILVV : croiser les approches et disciplines scientifiques pour transmettre

L’Institut Transdisciplinaire d’Étude du Vieillissement (ITEV), aux côtés du Dr Aline Chamahian (MCF en sociologie, Aix-Marseille Université et membre du conseil de direction de l’ITEV), a co-organisé, du 19 au 21 novembre 2025 à Aix-en-Provence, la 7^{ème} École d’Automne de l’Institut de la longévité, des vieillesses et du vieillissement (ILVV) avec pour objectif de rassembler les doctorants et chercheurs en début de carrière autour d’une thématique structurante ayant un enjeu majeur dans le contexte du vieillissement.

Cette année le thème que nous avons porté « La galaxie des savoirs autour du vieillir » avait pour but de sensibiliser les étudiants à la nécessité de croiser les regards, qu’ils soient experts ou profanes, afin de diffuser, partager et transmettre les savoirs sur le vieillissement ou acquis par le vieillissement.

Lors de ces journées, nous avons également organisé et animé la table ronde intitulée « Croiser les savoirs autour du vieillir entre expérience, expertise et compétences » au cours de laquelle nous

avons confronté les savoirs académiques et les savoirs profanes avec les expériences portées par les personnes elles-mêmes, leurs proches ou les professionnels qui les accompagnent. Ces travaux seront diffusés prochainement par l’ILVV et en accès libre via les captations de vidéos et la publication des comptes-rendus des séances.

L’EPHE – PSL est un des membres fondateurs de l’ILVV (Groupement d’Intérêt Scientifique pluri-organisme).



© ITEV

Affiche de l’École d’Automne d’Aix-en-Provence. Session introductory, ILVV.



Enseigner la
recherche par
la pratique de
la recherche



Le Fonds de dotation de l'EPHE : de l'ambition à l'impact concret

« Il y a quatre ans, l'École Pratique des Hautes Études – PSL donnait naissance à son Fonds de dotation avec une mission claire : apporter un soutien financier aux missions d'intérêt général de l'EPHE – PSL et amplifier son action au bénéfice de la recherche et de l'enseignement. »



édito

Le mot

de la

directrice

Marie-Pierre Lamotte,
directrice du Mécénat et du Fonds de dotation
de l'École Pratique des Hautes Études – PSL

Le Fonds de dotation de l'EPHE : 4 ans d'engagement envers la science et la connaissance

« Chères lectrices, chers lecteurs,

Il y a quatre ans, l'École Pratique des Hautes Études – PSL donnait naissance à son Fonds de dotation avec une mission claire : apporter un soutien financier aux missions d'intérêt général de l'EPHE – PSL et amplifier son action au bénéfice de la recherche et de l'enseignement.

Grâce à l'engagement de généreux donateurs, à la contribution active du Conseil d'Administration et à l'appui de toute la communauté EPHE – PSL, le Fonds de dotation a pris son envol et peut s'enorgueillir de réussites que d'aucuns lui envient.

Création de deux fonds individualisés, organisation de conférences, attribution de bourses, édition d'ouvrages, soutien à l'Institut des Langues Rares, ces projets s'inscrivent dans la droite ligne de la mission de l'EPHE – PSL : préserver des

savoirs rares et fragiles, étudiés, transmis et protégés grâce aux travaux de ses enseignants-chercheurs.

À travers ses actions, le Fonds réaffirme chaque jour sa vocation : agir aux côtés de l'EPHE – PSL pour soutenir une recherche d'excellence et garantir que les savoirs produits bénéficient au plus grand nombre.

Le Fonds de dotation se doit aujourd'hui de tirer parti du chemin parcouru pour développer de nouvelles ambitions et se tourner vers l'avenir. C'est donc encore et toujours animés du désir de contribuer au rayonnement d'un établissement d'exception que nous abordons les années à venir, heureux et fiers de voir toute une communauté mobilisée à nos côtés.

Nous lui en sommes sincèrement reconnaissants. »

Le Centre Sarikhani d'études élamites (CSEE)

Un ouverture vers une culture et une période souvent négligée de l'histoire millénaire de l'Iran

Le Centre a pour mission de soutenir l'**étude de l'Élam**, ainsi que de la langue et de la culture élamites, sous toutes leurs formes géographiques et temporelles. Son champ d'action s'étend du quatrième millénaire avant notre ère jusqu'à la période parthe, couvrant l'ensemble du sud-ouest de l'Iran, de l'Illam au Fars.

Il se consacre principalement au soutien de la recherche en histoire, philologie, archéologie et histoire de l'art.

Au-delà du milieu académique, le CSEE vise également à promouvoir la culture élamite auprès d'un public plus large, notamment par des collaborations avec des artistes contemporains.

Il s'attachera par ailleurs à établir de vastes coopérations avec des chercheurs individuels spécialisés dans les études élamites.

En complément de ces projets de recherche, des ateliers thématiques, réunissant des groupes de spécialistes, pourront être soutenus ou organisés. Plus globalement, le Centre s'efforce de sensibiliser les milieux universitaires et le grand public à une période et une culture souvent négligée au sein de l'**histoire millénaire de l'Iran**.

Le CSEE a pu voir le jour grâce à la Fondation Sarikhani (« The Sarikhani Foundation ») qui encourage la recherche, les expositions, les conférences et événements autour des arts de l'Iran de l'Antiquité à nos jours. Basé à l'EPHE – PSL, il a été inauguré le mercredi 6 novembre 2024 au prestigieux Grand Salon de la Sorbonne en présence de chercheurs de renom et de **Mme Ina Sarikhani**, directrice de la Fondation.

Le Centre a démarré ses activités en 2024 et son mandat initial est fixé à cinq ans. La direction en a été confiée à Monsieur Wouter F.M. Henkelman, maître de conférences à l'EPHE – PSL.



Le Centre Elahé Omidyar Mir-Djalali d'études persanes (CEOMEP)

Un centre de recherche international d'excellence dédié à la culture persane

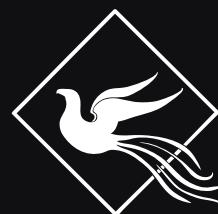
Le Centre a pu voir le jour grâce à un don exceptionnel de 2 millions de dollars de **Madame Elahé Mir-Djalali Omidyar**, Fondatrice et Présidente du Roshan Cultural Heritage Institute (RCHI) et ancienne élève de l'École.

Le CEOMEP a pour mission de servir de centre international d'excellence pour la recherche, la formation et la diffusion des connaissances sur **tous les aspects de la civilisation et de la culture persane**, s'adressant aux universitaires comme au grand public. Cette initiative s'inscrit dans

la longue tradition d'excellence de l'EPHE – PSL en études iraniennes, englobant des disciplines allant de la linguistique à l'archéologie.

Inauguré solennellement le 6 novembre 2024 au Grand Salon de la Sorbonne, le Centre, dont le mandat initial est fixé à cinq ans, est placé sous la direction de Monsieur Mohammad Ali Amir-Moezzi, directeur d'études émérite à l'EPHE – PSL, avec un comité scientifique dédié.

Ses activités se concentrent sur un cycle annuel de conférences d'iranologues de notoriété internationale (Iran préislamique, médiéval, moderne), des aides à la publication d'ouvrages, et l'octroi de bourses aux meilleurs étudiants en iranologie (Master et Doctorat).



« [...] en ce qui concerne les études iraniennes, les conférences les plus impressionnantes, les recherches les plus rigoureuses et les personnalités les plus fascinantes que j'ai rencontrées pendant toutes mes études, je les ai connues à l'École Pratique des Hautes Études [...] »

Madame Elahé Omidyar Mir-Djalali,
Fondatrice et Présidente du Roshan Cultural Heritage Institute



Libre de droit

Campagne Giving Tuesday



Une participation active au mouvement international Giving Tuesday

Dans un contexte où la solidarité et le soutien académique sont primordiaux, l'École Pratique des Hautes Études – PSL s'est engagée avec conviction dans l'initiative Giving Tuesday, participant ainsi pour la première fois à cette campagne philanthropique mondiale, en novembre 2024.

Initié en 2012 par la 92nd Street Y de New York et la Fondation des Nations Unies, le Giving Tuesday se présente comme une réponse altruiste à la surconsommation du Black Friday et promeut la générosité sous toutes ses formes au bénéfice de l'intérêt général. Cette initiative a depuis lors acquis une portée internationale. En France, elle a été introduite en 2018.

L'EPHE – PSL a saisi cette opportunité pour initier une campagne de collecte de fonds destinée à soutenir les étudiants. La somme collectée a alimenté le Fonds de dotation pour être ensuite redistribuée par la commission du Fonds de solidarité et de développement des initiatives étudiantes (FSDIE).

Les fonds récoltés ont permis de :

- > compléter les bourses de mobilité, facilitant l'accès aux terrains de recherche et aux ressources nécessaires pour les étudiants limités par les coûts de scolarité ;
- > offrir des aides d'urgence pour les étudiants en situation précaire, leur permettant de subvenir à leurs besoins fondamentaux.

Pour marquer le Giving Tuesday, l'EPHE – PSL a organisé le mercredi 9 avril 2025 et le mardi 2 décembre 2025 au Centre Mazet de la Vie Étudiante de PSL, deux tables rondes axées sur l'égalité des chances et le mécénat.

Nous tenons à remercier tous les donateurs particuliers qui ont permis de soutenir les étudiants en difficulté de l'EPHE – PSL.

Quelques repères chiffrés

Créé en
2021

+15
événements

Dynamique Alumni

Contribuer à la dynamisation du réseau des anciens : un projet dans la droite ligne de l'opération Giving Tuesday, en lien étroit avec la Direction de la vie étudiante

Dans une optique transverse et collaborative avec l'ensemble de la communauté EPHE – PSL, le Fonds de Dotation et la Direction de la vie étudiante travaillent en étroite collaboration, notamment pour bâtir un réseau dynamique et vivant qui crée des ponts entre les diplômés et favorise le dialogue tant intergénérationnel qu'interdisciplinaire.

C'est dans ce contexte qu'ont vu le jour les tables-rondes annuelles « L'EPHE – PSL et après ? » qui font interagir anciens élèves et étudiants. Dans le même esprit, un cycle de conférences portées par d'anciens élèves de l'école au parcours exemplaire émaillera l'année 2026 et au-delà.

2 fonds individualisés

+110 donateurs

7 bourses octroyées

1 structure abritante aux Etats-Unis



Soutenir l'EPHE – PSL et ses projets

En participant aux projets de l'EPHE – PSL vous contribuez à :

- › Soutenir nos ambitions d'excellence et d'innovation
- › Renforcer notre recherche dans des disciplines rares
- › Accompagner nos étudiants vers la réussite
- › Valoriser notre patrimoine
- › Favoriser nos liens avec le monde socio-économique

Comment nous soutenir et de quelles façons

- › Faire un don sur le site ephe.psl.eu rubrique « nous soutenir » ou par chèque à l'ordre de « Fonds de Dotation de l'EPHE » à retourner à EPHE Mécénat, 4-14 rue Ferrus, 75014 Paris
- › Devenir ambassadeur de l'école auprès de votre entourage

En pratique

- › Votre don vous permet de bénéficier des déductions fiscales en vigueur en France, dans certains pays d'Europe et aux Etats-Unis
- › Notre fonds de dotation est habilité à recevoir dons, legs, assurances-vie et donations

Exemple de projets à financer :

- › **Investir dans une recherche de pointe :** identifier et protéger les éponges dans les forêts animales marines ; comprendre leurs capacités d'acclimatation aux changements environnementaux

500 000 € sur 3 ans

- › **Soutenir nos étudiants :** octroi de bourses de mobilité internationale

50 000 € par an

- › **Sauvegarder et transmettre l'histoire et la mémoire des civilisations :** protéger la nécropole de Tanis, site patrimonial prestigieux et fer de lance de l'égyptologie française

150 000 € sur 4 ans

- › **Relier le monde socio-économique et la recherche :** financement de formations d'accompagnants au grand âge

30 000 € par an



DOSSIER SPÉCIAL

Humanités: quand l'IA transcrit les textes anciens



Jean Miélot copiant un livre dans son atelier,
Jean Le Tavernier, 1456

Saurez-vous retrouver l'anomalie dans l'image ?

L'Intelligence Artificielle (IA). Rien n'échappe à cette révolution déclenchant à des systèmes automatisés, des activités jusque-là vues comme relevant de la production intellectuelle. L'agenda de l'IA s'est accéléré en 2022 avec les premiers modèles de langues génériques, les fameux « LLM ». Dans les même temps, l'Union européenne (UE) publiait son « règlement sur l'intelligence artificielle (AI Act)¹ » afin de rendre ces technologies plus fiables, plus sûrs et dignes de confiance.

Dans ce cadre, les modèles d'IA en « *Open source* », c'est-à-dire dont le code est ouvert, jouent un rôle primordial. Aussi, nous vous proposons de découvrir la plate-forme e-Scriptorium et son module d'IA Kraken, portés par l'EPHE – PSL. Cet écosystème, implanté en Europe mais aussi au Moyen-Orient et en Amérique du Nord, contribue pleinement à la maîtrise de systèmes d'IA souverain.

L'Inria, par le biais de son équipe ALMAnaCH, s'est très tôt associé à l'EPHE – PSL pour développer et déployer cette plate-forme qui a vocation à devenir un service de référence pour les sciences humaines.

La partage de données ouvertes au sein de ces communautés scientifiques offre en particulier une opportunité de dynamiser les recherches en Humanités numériques (HN) partout dans le monde. Puisse ce dossier exceptionnel contribuer à cette ouverture et apporter à cette révolution en marche un éclairage éthique et scientifique.

Laurent ROMARY, Inria

¹ <https://www.europarl.europa.eu/topics/en/article/20230601ST093804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-regulation-in-europe-the-first-comprehensive-framework-4>

Introduction eScriptorium

Peter STOKES (EPHE – PSL)

eScriptorium est un logiciel qui utilise Kraken, un moteur d'Intelligence Artificielle (IA), pour transcrire automatiquement des documents manuscrits et imprimés. Chacun d'eux ont été développés principalement par une équipe en humanités numériques (HN) à l'EPHE – PSL dans le laboratoire AOROC (CNRS, ENS – PSL, EPHE – PSL).

Des millions de documents historiques sont conservés dans des bibliothèques, des archives et des collections à travers le monde, rédigés dans de nombreuses langues et écritures différentes. Jusqu'à tout récemment, l'accès à ces documents restait lent et laborieux ; il fallait se rendre dans différentes archives, obtenir l'autorisation de consulter les documents, disposer de l'expertise nécessaire pour les lire, et naturellement, d'avoir le temps et l'argent requis. Des millions de documents restent encore à peine étudiés voire tout simplement méconnus des chercheurs, malgré le travail considérable accompli pour les numériser depuis la fin des années 90. Par exemple, le portail du projet Biblissima+ comprend à lui seul près de 300 000 manuscrits et livres imprimés, dont environ 125 000 sont disponibles en ligne sous forme d'images numériques. Ceux-ci sont rédigés dans des langues telles que le latin, le grec, l'hébreu, l'arabe, le copte, l'arménien, le persan et le syriaque, pour n'en citer que quelques-unes.

Les chercheurs souhaitent lire et étudier tous ces textes. La chaîne numérique qui a été développée les aides pour analyser chacun de ces textes à grande échelle en répondant à de nombreuses questions telles que : « quels autres textes ont cité des passages de ce texte ? », « où, quand et par qui a-t-il été écrit ? », « comment l'utilisation d'un mot particulier a-t-elle évolué au fil du temps ? » et même « quels sont les textes contenus dans ce manuscrit ? ». Mais répondre à chacune de ces questions comme à bien d'autres, nécessite de disposer de caractères lisibles par la machine. C'est

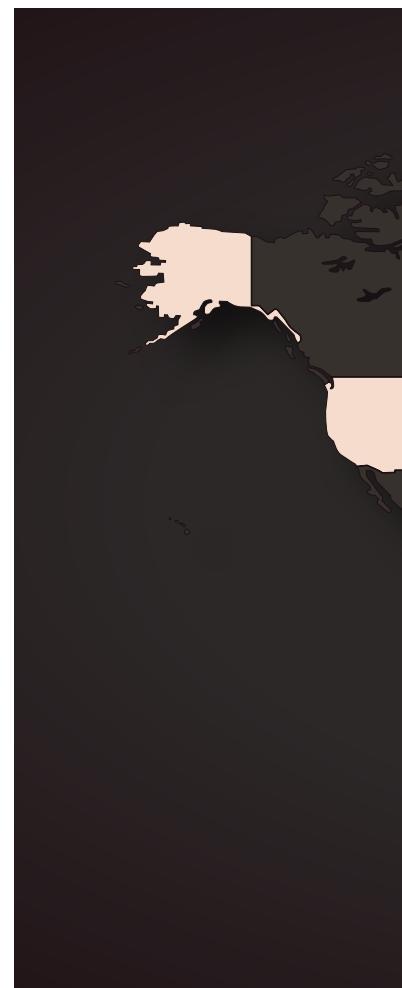
là qu'intervient la reconnaissance optique de caractères (OCR). Elle convertit l'image en texte brut, et c'est exactement à cela que sert eScriptorium.

Le logiciel eScriptorium a été conçu avec deux objectifs principaux : premièrement, que l'ensemble du processus de transcription automatique soit aussi ouvert que possible, et deuxièmement, que le logiciel soit conçu pour prendre en charge autant de systèmes d'écriture différents que possible. La première exigence est de plus en plus importante, et nous y reviendrons dans ce dossier à travers plusieurs exemples d'instances eScriptorium.

eScriptorium est le fruit d'une vingtaine de co-financements, dont les plus importants viennent de France 2030, de la Mellon Foundation et des grands projets européens. Le logiciel a vu le jour en 2019 dans le cadre du projet Scripta financé par l'université PSL. Celui-ci aborde l'histoire et les pratiques de l'écriture. Ben Kiessling travaillait déjà depuis un certain temps sur le logiciel Kraken (voir p.26), mais dont l'usage nécessite des compétences techniques pointues. L'équipe de l'EPHE – PSL a donc décidé de développer une plateforme web intégrée avec une interface ergonomique (orientée utilisateur).

eScriptorium et Kraken sont des logiciels que tout le monde peut télécharger et utiliser, à la condition de disposer d'ordinateurs relativement puissants pour entraîner et utiliser les modèles d'IA. Une partie importante du projet eScriptorium consiste à assurer l'installation et le fonctionnement du logiciel sur des serveurs et à en garantir l'accès depuis un navigateur web classique.

Distribution des instances eScriptorium dans le monde



Nous avons identifié 39 instances dans le monde, dont les plus importantes en nombre de pages sont celles de Paris, de Jérusalem, de Mannheim, de Leipzig et du Maryland. À notre connaissance, plus d'une vingtaine d'institutions et de consortiums différents à travers le monde utilisent ce logiciel, y compris deux gérés par les équipes de PSL. Une

des instances d'eScriptorium s'appelle MSIA ; elle est actuellement gérée par l'EPHE – PSL dans le cadre de Biblissima+ (voir p. 28) avec des

ordinateurs hébergés à l'Observatoire de Paris. L'« équipement d'excellence » MésoPSL y fournit un centre de calcul haute performance. Une autre instance, CREMMA, est hébergée par l'Inria en collaboration avec l'ENC – PSL et l'EPHE – PSL.

La reconnaissance optique de caractères convertit l'image en texte brut, et c'est exactement à cela que sert eScriptorium.

Télécharger des images

Depuis les bibliothèques, archives, le portail de Biblissima+, ...

Segmenter les images

- > Identifier automatiquement les lignes de texte, la pagination, ... avec un modèle adapté
- > Eventuellement corriger les résultats à la main et réentraîner le modèle

Reconnaissance du texte

- > Transcrire automatiquement chaque ligne de texte avec un modèle adapté
- > Eventuellement corriger les résultats à la main et réentraîner le modèle

Export des résultats

- > Export de la transcription, par exemple pour préparer une édition
- > Export des modèles et d'autres données, pour que d'autres puissent en bénéficier



eScriptorium, une aventure collective au croisement des humanités et de l'intelligence artificielle

Marc BUI (EPHE – PSL)

L'avènement de la plateforme eScriptorium est le fruit d'un long processus de maturation scientifique et humaine, nourri par les contributions de nombreuses personnes au sein de l'École Pratique des Hautes Études – PSL¹.

Dès l'origine, le projet s'inscrit dans le paysage singulier de l'EPHE – PSL, marqué par une culture de l'érudition, une ouverture interdisciplinaire et une méthode de recherche fondée sur la pratique en séminaire. C'est dans ce contexte qu'a émergé, à l'initiative de Daniel Stökl Ben Ezra, directeur d'études, titulaire

de la *Chaire de philologie et linguistique de l'hébreu et de l'araméen anciens*, l'idée de créer un outil informatique à la fois sophistiqué mais simple d'usage, destiné aux chercheurs travaillant sur les sources écrites.

Avec l'essor spectaculaire de l'IA et du deep learning, cette

intuition – permettre la lecture automatique et l'analyse à grande échelle de documents historiques – a trouvé un écho favorable. Le président de l'EPHE de l'époque, Hubert Bost, puis son successeur, Denis Pelletier², ont soutenu avec conviction cette initiative, inscrite dans la volonté de promouvoir les humanités numériques au sein de l'Université PSL alors en pleine construction.

Des synergies institutionnelles fondatrices

C'est également dans le cadre de l'*Initiative de Recherches Interdisciplinaires et Stratégiques Scripta-PSL*, consacrée à l'histoire et aux pratiques de l'écrit, et pilotée par Andreas Stauder avec le soutien de Philippe Huyse, que le projet a pris forme et cette initiative a servi d'incubateur scientifique pour eScriptorium, tout en étant soutenue par la direction de l'école et deses trois sections.

Le projet a également bénéficié de l'appui déterminant du programme Biblissima+ (voir p.28), dirigé par Anne-Marie Turcan-Verkerk³. Grâce à ce soutien, à la fois financier et technique, eScriptorium a pu se développer dans un environnement favorable, en étant étroitement lié au laboratoire AOrOc (voir lexique).

¹. Université PSL, dans le cadre du projet Scripta-PSL (2018–2021) et le Grand programme de recherche Les Fabriques de l'Antique (2025–).

². Programme de recherche et d'innovation Horizon 2020 de l'UE, dans le cadre des conventions de subvention n° 871127 (RESILIENCE, 2019–21) et n° 101071829 (MIDRASH, projet 2023–29, soutien financier eScriptorium 2023–25)

³. L'État français via l'Agence Nationale de la Recherche, à travers le Programme d'Investissements d'Avenir, réf. ANR-21-ESRE-0005 (Biblissima+, projet 2021–29, soutien financier eScriptorium 2021–24 et 2026)

Une aventure humaine et technologique

Au-delà de ces soutiens institutionnels, eScriptorium est avant tout une aventure humaine. La plateforme a permis de rassembler philologues, spécialistes des écritures rares et informaticiens dans un réseau scientifique d'une grande dynamique.

Deux figures majeures du développement, Benjamin Kiessling (voir portrait p.26) et Robin Tissot (voir portrait p.22), illustrent parfaitement cette rencontre entre les humanités et le monde de l'ingénierie. Le développement constant de la plateforme repose sur une interaction dynamique entre les chercheurs, les développeurs et la communauté d'utilisateurs, notamment autour de son moteur d'OCR Kraken.

L'arrivée de Peter Stokes à l'EPHE – PSL comme directeur d'études en humanités numériques, a par ailleurs renforcé le pilotage scientifique et donné toute la dimension pédagogique d'eScriptorium, dans le paysage des formations de niveau Master en humanités.

Des collaborations notables

De nouvelles collaborations se sont nouées avec des partenaires du monde de la recherche et de la culture, notamment l'équipe Inria Almanach dirigée par Benoît Sagot, avec notamment Alix Chagué ou Laurent Romary, pour répondre à divers défis (développements, documentation) mais aussi de pérennité technologique de l'instance principale (infrastructures de calcul, stockage, en fin 2025)

Le réseau inclut naturellement l'École nationale des chartes – PSL, en particulier via le Master Humanités numériques de PSL. L'article consacré au projet CATMuS illustre la vitalité de ces échanges et la place qu'occupe eScriptorium dans la formation et la recherche sur les écritures anciennes.

Une plateforme devenue référence internationale

Entrée dans sa maturité, eScriptorium est désormais au cœur de grands projets de recherche européens, notamment MiDRASH (hébreu), DeLiCaTe (arménien-géorgien) et PRIMA (italien). Ces projets démontrent la polyvalence de la plateforme, capable de traiter des écritures variées – jusqu'aux sinogrammes, qui posent de nouveaux défis scientifiques (voir l'article sur la loi de puissance et le chinois classique).

Les initiatives OpenITI ou Syriac Transcribathon témoignent également du caractère vivant et collectif : la recherche y prend la forme d'événements collaboratifs et participatifs, où chercheurs et étudiants travaillent ensemble.

4. Outre les financements principaux précités, et grâce au soutien continu du laboratoire AOrOc (CNRS, ENS – PSL, EPHE – PSL) et de l'EPHE – PSL, eScriptorium a bénéficié d'autres contributions, par exemple celles des projets Vietnamica (porté par l'EPHE – PSL), LectAuRep (EPHE – PSL, Inria et Archives nationales), ManuscriptologIA (Domaine de recherche et d'Innovation Majeur Sciences du Texte et Connaissances Nouvelles [DIM STCN]), EPHE – PSL et l'Institut de Recherche et d'Histoire des Textes CNRS), Schmidt DataX Fund (Université de Princeton), FonDUE (Formes Numérisées et Détection Unifiée des Écritures, Université de Genève), ReadChinese (EPHE – PSL et BnF DataLab) et BACK IN TIME (Inria).

5. L'équipe eScriptorium est reconnaissante à différentes formes de contributions, telles que la fourniture de jeux de données ouvertes (en particulier ceux sur HTR-United), les modèles ouverts (en particulier ceux déposés sur Zenodo), les rapports des bogues et des améliorations, la documentation, l'organisation de formations, la production de « plugins » et d'autres extensions pour Kraken et eScriptorium

Un écosystème pour la science ouverte

Si Biblissima+ et les programmes de recherche de PSL ont servi d'incubateur, ce sont les grands projets ERC et la démarche de science ouverte⁴ qui ont assuré la visibilité internationale d'eScriptorium⁵.

Aujourd'hui, cette plateforme incarne un écosystème unique pour toutes celles et ceux qui s'intéressent à l'écrit, aux écritures et aux langues – qu'il s'agisse de philologie, de linguistique ou d'IA.

Ce numéro, placé sous le haut patronage du président Michel Hochmann, offre un panorama de ces réalisations et de l'actualité de l'IA appliquée aux humanités. Il marque avant tout une réussite collective, où l'innovation technologique se met au service de la connaissance et du patrimoine écrit.

Portrait Robin Tissot, le concepteur d'eScriotorium

Propos recueillis par FAVDK

On doit à Robin Tissot, développeur principal à l'EPHE – PSL, l'interface du logiciel eScriotorium associé au module d'IA Kraken. Rencontre avec un ingénieur informatique changé par les humanités, qui a passé ces six dernières années à créer eScriotorium.



D'où venez-vous et pourquoi avoir embarqué dans ce projet ?

À l'origine je ne viens pas des humanités numériques (HN). Ma culture s'est faite dans l'industrie, la communication et les sociétés de services en ingénierie informatique (SSII). Un jour, j'ai découvert une annonce publiée par l'Université PSL. J'ai passé l'entretien et cela s'est bien passé. J'ai donc rejoint le projet à ses débuts sans avoir le « background académique ». J'ai pris un peu de temps pour m'acclimater, mais je dois dire que les HN ont changé ma façon de voir les choses. Cela m'a permis de donner plus de sens à mon travail.

Quel est votre rôle dans le projet ?

eScriotorium est principalement une interface entre les utilisateurs et Kraken. Je me suis donc occupé d'abord de l'analyse des besoins utilisateurs et les ai traduits dans un schéma de bases de données, des fonctionnalités, des pages et des interfaces. Cela touche, par exemple, à comment afficher le plus ergonomicement possible les images ou comment les stocker le plus économiquement. L'essentiel de mon activité porte, aujourd'hui, sur la maintenance du code et l'amélioration du programme. Je développe notamment quelques nouvelles fonctionnalités tout en gérant les contributions de la communauté.

A quoi ressemble cette communauté e-Scriotorium ?

La plus grande partie - peut-être 3 000 personnes - est constituée d'étudiants et d'enseignants-chercheurs (philologues ou linguistes) qui n'ont pas nécessairement une connaissance du code. J'interagis plutôt avec ceux dont les

eScriotorium est principalement une interface entre les utilisateurs et Kraken.

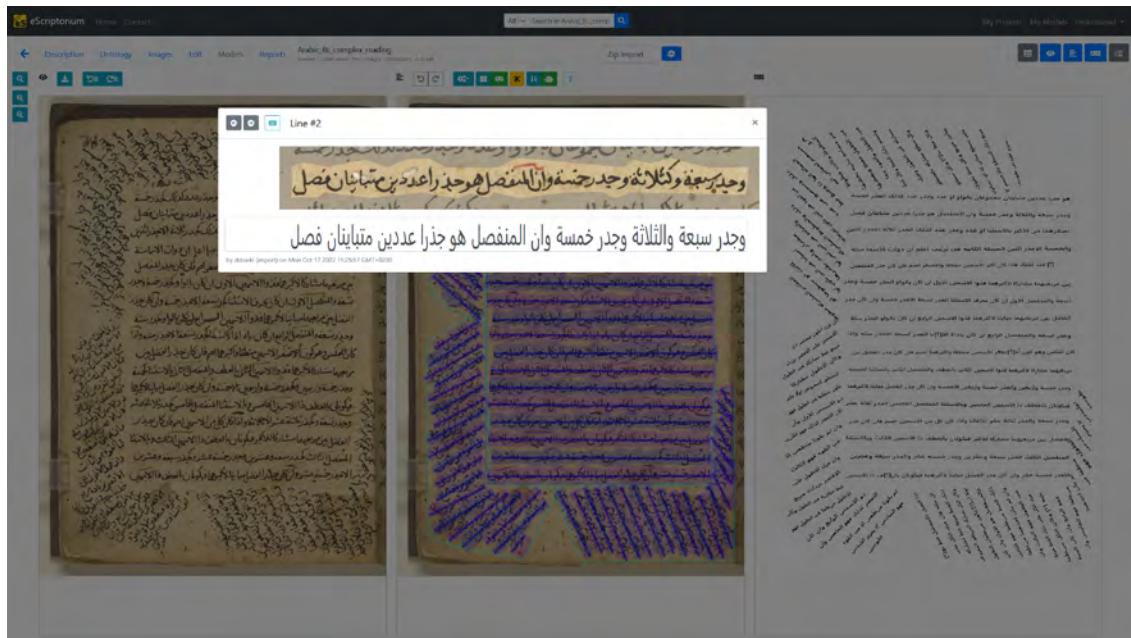
connaissances techniques sont suffisantes et plus rarement avec les utilisateurs ordinaires qui font des transcriptions. Ce qu'on a essayé de faire est de se rapprocher d'un développement classique pour un logiciel *Open Source*. On compte aujourd'hui quatre principaux contributeurs au code et une vingtaine de réguliers. On apprend l'existence des problèmes via le Gitlab (Kraken passe Github).

Finalement, il y a les gestionnaires des instances, c'est-à-dire des installations d'eScriotorium, qu'il faut épauler de temps en temps. C'est une question de bonne pratique – les méthodes fonctionnent – pour un projet très ouvert, utilisable par tous.

Comment décrire le développement d'eScriotorium ?

Il y a eu une première étape d'apprentissage, notamment s'agissant de la terminologie comme de l'adaptation aux besoins des chercheurs. Ensuite, il a fallu définir une architecture et des solutions techniques inédites (il n'y avait que le logiciel Transkribus sur le même créneau) en se posant les questions essentielles : comment, pourquoi ou encore qu'est-ce qu'on pourrait faire ou ne pas faire ?

Concrètement, cela consistait à se demander si on allait stocker par « ligne » ou bien par « lettre ».



Interface de transcription d'eScriptorium

Cela a eu un impact profond sur la structuration du code comme l'interface et in fine, cela a garanti le succès d'eScriptorium. Il s'avère que l'analyse qu'on avait à l'époque tient encore la route malgré une interface complètement remodelée par les designers. Pour parvenir à ce succès, il y a eu naturellement beaucoup d'essais, souvent infructueux, quelquefois avec les utilisateurs. Mais je me suis toujours fixé comme objectif de conserver les choses les plus simples dans ce processus, complété à partir de 2023 avec des focus group. Et il y a aujourd'hui près de 120 fonctionnalités !

Quelle méthode de management a-t-elle été utilisée ?

Ce processus itératif ne s'inscrit pas nécessairement dans les méthodes dites Agile qui ont eu le vent en poupe dans les années 2010. Comme beaucoup de développeurs que je connais, nous avons préféré mobiliser notre temps sur le développement plutôt que sur le management. Et je dois dire que je peux compter sur des gens très compétents, qui savent ce qu'ils ont à faire, et des contributeurs qui savent ce qu'ils veulent et l'exprimer dans un environnement de travail principalement anglophone.

Quel est le plus d'eScriptorium aujourd'hui ?

Il fait gagner un temps fou à des gens qui n'ont pas beaucoup de temps et de moyens. Au-delà, eScriptorium essaie de ne pas prendre parti dans la façon dont les gens vont mettre en œuvre le programme scientifique. C'est un écosystème *Open Source* où les gens donnent volontiers leur retour alors que le logiciel demande beaucoup de ressources. De nombreuses universités le font tourner et c'est une grande fierté.

De quoi êtes-vous le plus fier ?

Sans aucun doute la fonctionnalité « fenêtre de transcription » pour laquelle nous avons eu énormément de retours positifs, ou le jumelage entre l'affichage du manuscrit et sa transcription.

eScriptorium continue-t-il d'évoluer ?

Un écosystème sain est un écosystème fonctionnant avec des outils indépendants mais partageant un même standard. Dans l'immédiat nous visons la mise en place de la nouvelle interface ; ensuite, l'évolution est potentiellement sans fin mais le logiciel doit se stabiliser, c'est-à-dire collaborer avec d'autres outils dédiés spécifiquement pour l'annotation, la visualisation...

Master Humanités Numériques

Porté par PSL avec l'École nationale des chartes et l'EPHE, le master Humanités numériques forme à l'exploitation et à l'analyse des données en sciences humaines et sociales. Il s'adresse à des étudiants de ces domaines souhaitant développer des compétences en méthodes numériques, IA et data science.





Intelligence artificielle et Sciences des données pour les Humanités dans la stratégie de l'Université PSL

Stéphane VERGER

Vice-président délégué Arts & Humanités de l'Université PSL
directeur d'études à l'EPHE – PSL

L'Université PSL compte parmi les leaders européens et internationaux en intelligence artificielle et en Sciences des données. Elle porte le plus important des projets lauréats du programme IA Cluster, une initiative phare du Secrétariat Général pour l'Investissement (SGPI) déployée dans le cadre de la stratégie France 2030. Financée à hauteur de 75 M€, la Paris School of AI (PR[AI]RIE-PSAI), qui est portée par l'Université PSL en partenariat avec l'université Paris Cité, le CNRS, l'Inria et l'Institut Pasteur, a été inaugurée en 2025.

Forte d'un potentiel scientifique de premier ordre, la Paris School of AI propose des parcours de formation ambitieux et innovants, de la licence au doctorat. Ces cursus sont conçus pour répondre aux besoins des étudiants, du monde académique et des professionnels des secteurs privés et publics en quête d'expertise en intelligence artificielle. Fidèle à l'approche interdisciplinaire de l'Université PSL, elle forme aussi les étudiants aux usages de l'intelligence artificielle dans de nombreux domaines d'application, en prenant en compte les questions éthiques, de régulation et de soutenabilité.

Dans le domaine des Humanités numériques et computationnelles ou des Sciences des données pour les Humanités, l'Université PSL et ses établissements composantes ont développé un dispositif unique, qu'il conviendra dans les prochaines années de renforcer et de structurer au sein d'un grand pôle de formation et de recherche en Humanités.

Pour ce qui concerne la formation de premier cycle, les Humanités computationnelles sont au cœur du parcours du Cycle Pluridisciplinaire d'Études Supérieures (CPES) en « Sciences des données, arts et cultures » organisé en partenariat avec le Lycée Louis-le-Grand. L'offre de formation se poursuit en master avec les parcours « Humanités numériques », fruit de la collaboration de l'École

nationale des chartes – PSL, de l'École normale supérieure – PSL et de l'École Pratique des Hautes Études – PSL, et « Technologies numériques appliquées à l'Histoire » de l'École Nationale des Chartes – PSL, qui sont communs aux trois programmes gradués en Humanités (Sciences Historiques, Sciences des religions et Philosophie et Littératures)

et peuvent donner accès à un contrat doctoral spécifique réservé aux Humanités numériques.

La recherche de l'Université PSL dans ce domaine est quant à elle structurée au sein du programme CultureLab

– Computational Science of Culture, l'un des quatorze Grands Programmes de Recherche sélectionnés pour une durée de cinq ans, qui réunit des équipes de recherche de cinq établissements

composantes (ENC – PSL, ENS – PSL, EPHE – PSL, Dauphine Paris – PSL et Observatoire de Paris – PSL). Les programmes phares de l'EPHE – PSL, comme Biblissima+ et eScriptorium, assurent à l'établissement une position clé dans la stratégie générale de l'Université PSL dans le domaine des Sciences des données pour les Humanités.

L'Université PSL compte parmi les leaders européens et internationaux en intelligence artificielle et en Sciences des données.

Portrait Benjamin Kiessling, l'architecte IA de l'EPHE – PSL

Propos recueillis par FAVDK

On doit à Benjamin Kiessling, aujourd'hui ingénieur de recherche à l'Inria, le module d'IA du logiciel eScriptorium : Kraken. Rencontre avec un ingénieur informatique (développeur principal) passionné par les humanités.

Où vous est venue l'idée de Kraken ?

J'étais assistant de recherche (HiWi) en Allemagne, il y a tout juste dix ans. L'outil « Ocropus » était un logiciel déjà flexible, mais faible en termes d'architecture logiciel. Il faut dire que ce type d'outils de recherche sont souvent très fragiles. Il était donc nécessaire de développer des modules facilement intégrables dans un chaîne de traitement, tel que eScriptorium. Kraken reste un logiciel de recherche mais davantage proche des standards utilisés par le monde de l'archivistique.

Pourquoi avoir rejoint notre école ?

En 2018, Daniel Stökl Ben Ezra m'a approché pour rejoindre l'EPHE – PSL et le projet Scripta-PSL (Université PSL). C'était une belle opportunité pour développer Kraken ; je rêvais de changement à l'époque, même si le modèle de l'EPHE (l'enseignement de la recherche par la recherche), dérivé d'un modèle allemand, m'était familier. Je savais en tous les cas que l'EPHE était un haut lieu de recherche dans des langues et écritures rares et que Scripta-PSL s'inscrivait parfaitement dans cette tradition. De surcroît, l'EPHE est une école dont la taille permet une certaine agilité et où un problème trouve toujours une solution. De fait, Kraken a énormément évolué

et n'a plus grand-chose à voir avec ce qu'il était avant.

Et qu'est-ce que Kraken aujourd'hui ?

L'objectif de Kraken est de fournir un outil de numérisation, dit de transcription automatique pour les « matériaux » (textes) qui ne sont pas bien traités par les solutions OCR commerciales. Kraken est le seul logiciel qui est conçu comme étant agnostique (pas de biais culturel) par rapport à ces matériaux. Durant ces 7 années, notre équipe a pu implémenter une méthode entraînable automatique (méthode supervisée) pour toutes les tâches liées à l'OCR : trouver les lignes, leur ordre, leur hiérarchie et les transcrire ; les défis de la recherche et du développement étaient immenses à tel enseigne que les modules de la chaîne de reconnaissance de Kraken sont désormais entraînables.

Cette flexibilité en a fait un outil particulièrement utile et largement adopté dans les humanités numériques. Cette communauté se parle et tire, aussi, profit des proximités scripturales. J'ai conçu Kraken pour la transcription de l'arabe et cela a créé des opportunités pour la transcription de l'hébreu. Et les synergies entre projets peuvent être encore plus nombreuses et parfois



invisibles. Aujourd'hui, je suis donc devenu le développeur principal du logiciel avec une communauté de plusieurs milliers d'utilisateurs. Il est probable que plusieurs centaines d'instances de Kraken et des dizaines d'instances d'eScriptorium soient installées sans que nous n'ayons connaissance. Généralement c'est quand quelque chose ne fonctionne pas que nous interagissons via Github. Mais je préfère largement l'interaction dans la vie réelle lors des nombreuses formations ou transcribathons organisées en Europe et ailleurs.

Kraken continue-t-il d'évoluer et de quoi rêvez-vous ?

En 2017, je visais l'arabe, le grec et le latin. Aujourd'hui, je rêve d'une transcription universelle. Plus la communauté s'élargit plus ce rêve devient atteignable. Il y a quelques semaines, lors d'un colloque en Chine, j'ai par exemple appris que Kraken avait été utilisé pour des textes en

écriture arabe et en langues sub-sahariennes utilisées au Congo, au Mozambique, au Ghana, au Sénégal et bien d'autres pays.

L'objectif de Kraken est de fournir un outil de transcription automatique pour les textes qui ne sont pas bien traités par les solutions commerciales.

Ces langues partagent des traces d'arabe liées à leurs échanges commerciaux avec l'Inde, Oman... Il s'agit d'un corpus de centaines de milliers de manuscrits pour lequel le projet OpenITI a apporté sa contribution. C'est très enthousiasmant et c'est aussi cela eScriptorium : un écosystème décentralisé et ouvert.

Kraken reste toutefois limité s'agissant de la transcription des textes cunéiformes et hiéroglyphiques (Égypte et Amérique centrale) pour lesquels il n'y a souvent même pas une liste de signes.

Outre les développements incrémentaux (écritures, logiciels, fonctionnalités...), Kraken continue d'évoluer notamment sous l'influence des LLM (modèles de langage). Aujourd'hui, le système reste très dépendant d'une détection précise des lignes. J'ai donc réutilisé LLAMA (le LLM de Meta, en adaptant et allégeant son architecture) pour déployer une reconnaissance plus robuste pour les ajouts marginaux et interlinéaires des textes.

Kraken pourrait-il être utilisé en art, en astronomie ou en SVT ?

Bien sûr, Kraken pourrait servir à bien d'autres choses. On pourrait imaginer qu'il identifie des peintres, des traces d'empreintes ou des galaxies, mais je ne suis pas certain que cela soit l'outil le plus adapté pour cela bien qu'il fasse de la classification.

Cela n'empêche qu'il sert déjà à la numérisation d'archives (dossiers de patients) dans le domaine médical notamment. Mais je suis certain qu'il y aura toujours des gens pour trouver une application bizarre pour kraken que je n'aurais jamais imaginée.

Qu'est-ce que ce projet vous a apporté ?

Comme informaticien, je n'avais pas la formation en écriture historique. Avec le temps, mes ambitions sont devenues plus grandes, notamment pour être utile aux chercheurs en sciences humaines. J'ai trouvé extrêmement satisfaisant de travailler avec les sciences humaines car il y a une pratique de la nuance et la vision n'est pas binaire comme en informatique. Dans ce domaine, j'ai souvent constaté que « They don't know what they don't know ». À l'inverse, en sciences humaines, le questionnement est permanent, riche et extrêmement intéressant. Le monde historique y est beaucoup plus vaste que ce

qu'on apprend à l'école. Je n'ai qu'un seul regret, ne pas avoir appris l'arabe, mais j'espère le pouvoir dans le cadre de l'ILARA par exemple.

PSL Week AljamiaTech : manuscrits aljamiados et humanités numériques

La chaire d'Histoire et codicologie du livre manuscrit arabe, dirigée par Nuria de Castilla (EPHE – PSL), organise du 2 au 6 mars 2026 une semaine consacrée aux écritures en caractères arabes et à la translittération automatique dans le cadre de la PSL-Week.

Ce cours intensif place les humanités numériques (HN) au cœur de l'exploration d'un corpus encore largement méconnu : les manuscrits aljamiados. Ces textes, rédigés en alphabet arabe pour préserver en espagnol les pratiques et croyances des communautés musulmanes de l'Espagne médiévale et moderne (XVe-XVII^e siècle), offrent un terrain privilégié pour expérimenter de nouvelles méthodes numériques de translittération automatique vers l'alphabet latin. Le programme mobilise la plateforme eScriptorium et des approches fondées sur l'IA.

Prolongé par un hackathon collaboratif, ce cours permet aux participants d'acquérir des compétences en HN, de réfléchir aux enjeux méthodologiques et épistémologiques liés à l'application des technologies aux écritures anciennes, et de contribuer à la création collective d'un jeu de données inédit.

Apprenez une langue rare avec l'ILARA



Biblissima !

Observatoire des cultures écrites
de l'argile à l'imprimé



Le portail Biblissima fédère des données numériques sur l'histoire de la transmission des textes et leurs images produits de l'Antiquité à l'avènement de l'imprimé, en Orient comme en Occident, dans une grande variété de supports, de langues ou d'écritures.

© Biblissima

Biblissima : observatoire des cultures écrites anciennes

Anne-Marie TURCAN-VERKERK (EPHE – PSL)

eScriptorium fait partie d'un vaste écosystème de programmes consacrés au patrimoine écrit, dont l'un des plus importants est Biblissima+. Cet équipement d'excellence porté par le Campus Condorcet fédère et structure des millions de données relatives aux textes anciens, produites par ses partenaires français et internationaux.

Dépasser l'hétérogénéité des bases de données, des bibliothèques numériques, des catalogues, des éditions numériques pour donner à l'internaute un accès unique aux données sur les textes anciens et leurs images : c'est ce que réalise progressivement Biblissima pour les quelque 150 projets qu'il soutient depuis 2012.

Un premier ÉquipEx, Biblissima (2012-2021), a permis de favoriser la production de données sur la transmission des textes anciens, surtout en Occident,

du VIII^e au XVIII^e siècle, de créer l'infrastructure numérique et de mettre en place le processus de réconciliation de données qui consiste à comparer et vérifier des ensembles de données provenant de sources différentes pour s'assurer qu'elles sont cohérentes, exactes et complètes, permettant ainsi d'atteindre l'interopérabilité profonde utile aux chercheurs.

Un second programme, Biblissima+ (2021-2029), a permis de consolider ce socle technique en soutenant le

développement de ressources numériques complémentaires (logiciels, référentiels, nouveaux jeux de données) et l'activité de groupes thématiques autours de 7 domaines d'innovation pour les disciplines associées (utilisation du standard IIIF pour la diffusion des images sur le web, encodage de textes, reconnaissance des écritures avec l'IA...). Il met l'accent sur l'élargissement et l'automatisation des processus d'agrégation de données et développe les liens avec d'autres infrastructures, comme Persée. Il favorise la création d'outils, mais aussi la production de nouveaux types de données : désormais, Biblissima concerne potentiellement toutes les langues anciennes, sur tous supports, de toutes époques. L'écosystème s'ouvre donc à de nouvelles aires

géographiques, à de nouveaux objets et à de nouvelles communautés de recherche et d'utilisateurs, en particulier grâce à son appel à manifestation d'intérêt annuel, qui permet aux chercheurs et aux projets de rejoindre l'infrastructure quand ils se sentent prêts à le faire : c'est ainsi que Biblissima accueille de nouvelles langues, comme le chinois ou le catalan médiéval, de nouveaux supports de textes, comme les rouleaux, les dalles funéraires et même la tapisserie de Bayeux, et que l'infrastructure encourage le développement de nouvelles technologies, en particulier dans le domaine de l'IA, par exemple pour l'analyse des enluminures.

Enseignants-chercheurs et chercheurs, ingénieurs, conservateurs, étudiants, qu'ils soient historiens, philologues, archéologues, numismates, sigillographes, paléographes, épigraphistes, musicologues, archivistes etc., qu'ils appartiennent aux équipes partenaires de l'EquipEx+ ou qu'ils collaborent avec elles, sont réunis au sein de sept clusters transversaux organisés selon le processus de découverte des sources : depuis le cluster 1, consacré à la constitution de corpus numériques interopérables grâce au protocole IIIF, jusqu'au cluster 7 pour l'analyse profonde des textes, entre autres par l'IA, en passant par l'analyse physico-chimique des documents (cluster 2), l'automatisation de la lecture des documents par l'IA (cluster 3, au sein duquel eScriptorium occupe une place de choix), l'analyse structurelle des écritures et des formes (cluster 4), l'édition numérique (TEI, standard ÉpiDoc : cluster 5), la musicologie

numérique (cluster 6). Les clusters sont des lieux de dialogue, de partage et de rencontre de problématiques, de création d'outils communs, de formation et d'accompagnement.

En 2026, Biblissima sera pérennisé sous forme de fédération de recherche et développera de nouveaux axes, en particulier avec la création d'un portail Biblissima-Textes, grâce à Inria-Paris, et celle d'un site de partage des savoirs sur les langues et cultures anciennes, Odysseum-Biblissima, avec l'ENS – PSL et le ministère en charge de l'Éducation nationale. eScriptorium deviendra l'une des briques de l'infrastructure, qui lui permettra de poursuivre son développement. Un accord de partenariat entre le Campus Condorcet et UCL, signé le 21 novembre 2025, implique l'infrastructure Biblissima dans une réflexion commune sur les SHS à l'ère de l'IA. Biblissima doit faciliter l'intégration de l'IA dans des processus métiers et des chaînes de traitement des données, contribuer à la formation des communautés, mais aussi au développement de la recherche en IA sur des corpus très spécialisés. Écosystème où s'exerce avant tout la critique scientifique, l'infrastructure sera aussi un lieu de concertation et de débat sur l'opportunité et la traçabilité des résultats obtenus par l'IA.

Biblissima a été construit grâce aux Investissements d'avenir intégrés dans France 2030 (EquipEx+ réf. ANR-21-ESRE-0005), et est coordonné depuis 2012 par Anne-Marie Turcan-Verkerk, directrice d'études à l'EPHE – PSL, section des Sciences historiques et philologiques.

Le programme associe aujourd'hui 15 établissements d'enseignement supérieur et de recherche, le Service interministériel des archives de France et une entreprise, Teklia. Il est porté par le Campus Condorcet.

The screenshot shows the Biblissima portal interface. At the top, there is a search bar with the placeholder "Rechercher une œuvre, une personne, un lieu, une cote, une enluminure...". Below the search bar, a banner reads "Accès unifié aux données sur les textes anciens et leurs images". The main content area is titled "Bibliothèque de Saint-Benoit-sur-Loire". It displays a list of digitized manuscripts under the heading "Numérisations". The list includes:

- France, Orléans, fonds principal, 0277 (233) (décor) - 9 images
- Reg. lat. 215 - 300 images
- Fragment (Latin) - 6 images
- BnF: Département des Manuscrits. Latin 7926 - 434 images
- Vita, miracula, transitus Martialis Lemoviensis / Pseudo-Aurélianus of Limoges, VLF 12 F - 19 images
- BnF: Département des Manuscrits. Latin 7903 - 178 images

On the left side of the manuscript list, there is a sidebar with icons for search, filters, and other functions. At the bottom of the page, there is a footer with the Biblissima logo and a copyright notice: "© Biblissima".

Consulter dans une même interface les numérisations IIIF des manuscrits ayant appartenu à l'abbaye de Fleury à Saint-Benoit-sur-Loire, qu'ils soient conservés aujourd'hui à Leyde, au Vatican, à la BnF ou dans des bibliothèques publiques françaises.

Dunhuang: déjouer la loi de puissance

Colin BRISSON & Marc BUI (EPHE – PSL)

Avec plus de 200 000 titres d'ouvrages uniques, des dizaines de millions de documents d'archives, et des centaines de milliers de sources épigraphiques et manuscrites, le patrimoine littéraire chinois est si vaste qu'il pourrait « faire transpirer un bœuf et remplir une pièce jusqu'au plafond ».

Malgré le rôle pionnier de la République populaire de Chine dans l'utilisation de l'IA pour faciliter l'accès à ces textes¹, on estime qu'aujourd'hui, seuls 5 % du patrimoine littéraire chinois est disponible sous forme de texte numérique. Sur ces 5 %, une grande partie est inaccessible aux établissements d'enseignement supérieur européens en raison du coût prohibitif des licences commerciales.

Composée de chercheurs des laboratoires AOrOc (UMR 8546) et du CRCAO (UMR 8155), sous la tutelle de l'EPHE – PSL, l'équipe Numerica Sinologica contribue à créer de vastes corpus gratuits et libres d'accès. Pour y parvenir, nous avons résolu des difficultés techniques inhérentes à la nature même de l'écriture chinoise. Le succès de notre approche sur le chinois classique ouvre de nouvelles perspectives pour l'étude des documents historiques en sinogrammes au-delà des frontières de la Chine.

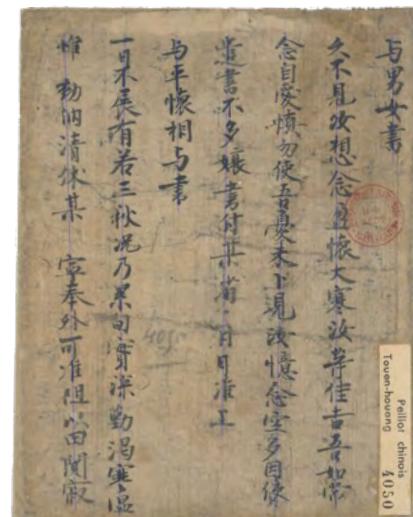
Les défis de la transcription automatique du chinois classique

On estime qu'environ 14 000 sinogrammes sont nécessaires pour transcrire la majorité des textes en chinois classique. Cela représente un ensemble de classes de signes cent fois supérieur à celui requis pour les langues alphabétiques.

Le deuxième défi réside dans la similarité visuelle de nombreux sinogrammes, qui est source de nombreuses confusions. Un exemple bien connu des sinologues est la triade : *jǐ 己*, « soi », *yǐ 已*, « révolu », et *sǐ 已*, « le sixième ». Dans les imprimés anciens ou les manuscrits, de légères variations graphiques – deux traits qui ne se rejoignent pas parfaitement, une courbure altérée – peuvent rendre ces caractères indiscernables hors d'un contexte linguistique précis.

Troisièmement, la distribution de la fréquence des caractères chinois correspond à une loi de puissance : une minorité d'éléments apparaît fréquemment, tandis que la majorité est extrêmement rare. Il en résulte que moins de 20 % des sinogrammes représentent plus de 95 % des occurrences, tandis que l'immense majorité est extrêmement rare.

Ce phénomène pose un double problème. D'abord les algorithmes d'apprentissage ont une tendance naturelle à privilégier l'apprentissage des caractères les plus fréquents au détriment de la longue traîne des signes rares. Or, ces caractères rares sont souvent cruciaux car ils portent un poids sémantique fort. Un modèle peut donc atteindre un taux de précision statistique très élevé tout en produisant un texte sémantiquement erroné.



Manuel de correspondance, manuscrit sur papier, Pelliot chinois 4050, département des Manuscrits, Bibliothèque nationale de France.

De plus, cette distribution statistique pose une difficulté majeure lors de la collecte des données. Pour garantir que l'algorithme soit exposé, même de manière minimale, aux milliers de caractères de la longue traîne, il est nécessaire de réunir un volume beaucoup plus important de données que pour les écritures alphabétiques.

1. Il s'agit du projet de transcription du *Sikuquanshu* qui a débuté en 1996 et qui a permis de numériser en à peine trois ans 3 461 ouvrages, représentant un total de 800 millions de sinogrammes.

Déjouer la loi de puissance des sinogrammes

Afin de contrer les effets délétères de la loi de puissance, plutôt que de classifier des caractères isolés, l'équipe Numerica Sinologica a réalisé la transcription au niveau des lignes de texte. Initialement développée pour les écritures ligaturées où les caractères sont difficiles à séparer, cette approche permet, pour le chinois, d'obtenir une distribution plus équilibrée de l'information.

Cette approche se heurte cependant à une contrainte qui a freiné son adoption en Chine. Pour modéliser efficacement la probabilité conditionnelle au niveau de la ligne d'écriture, l'entraînement requiert des paires image-texte parfaitement alignées.

La distribution de la fréquence des caractères chinois correspond à une loi de puissance : une minorité d'éléments apparaît fréquemment, tandis que la majorité est extrêmement rare.

Afin de constituer un jeu de données couvrant au moins 10 occurrences de 50 % des 14 000 caractères, il faut en moyenne réunir 800 000 paires image-texte. La création d'un tel jeu de données,

qui resterait par ailleurs largement insuffisant, en utilisant l'annotation manuelle nécessiterait l'équivalent d'environ cinq années de travail à temps plein.

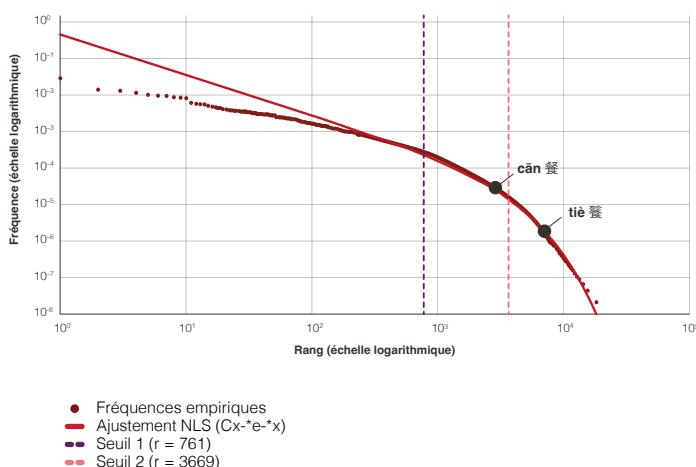
Pour surmonter cet obstacle, nous avons exploité des transcriptions existantes, notamment celles issues de projets

collaboratifs Kanripo et CBETA. Nous avons développé pour cela un algorithme capable d'identifier si une image contient un texte que nous possédons, puis d'aligner avec une très haute précision chaque ligne de l'image avec sa transcription. Ce procédé nous a permis de générer les données nécessaires à l'apprentissage sans avoir recours à l'annotation manuelle.

Nous avons dans un premier temps appliqué cet algorithme à plusieurs corpus de facsimilé². Cela nous a permis de constituer un premier jeu de données de plus de deux millions de lignes. Les modèles entraînés à l'aide du moteur d'OCR Kraken ont atteint un taux d'erreur de 1 %, confirmant ainsi l'efficacité de notre stratégie.

2. https://github.com/colibrissen/CHAT_models

Distribution rang-fréquence des caractères chinois



La densité informationnelle des sinogrammes

Celle-ci peut être mesurée par l'entropie de Shannon. Tandis qu'une lettre de l'alphabet latin possède une entropie d'environ 1,5 bit, celle d'un sinogramme atteint en moyenne 10,20 bits. Cette densité est comparable à celle d'un mot complet en anglais (estimée par Shannon à 11,8 bits). Un erreur de transcription occasionne donc la perte d'un mot complet en chinois. Sachant un taux d'erreur de 4 % en latin, cela reviendrait à perdre 1 mot sur 25 en chinois.

Sur le plan statistique, cela revient à utiliser la probabilité conditionnelle d'un caractère au lieu de sa probabilité fréquentiste. Le sinogramme *tiè 餐* illustre bien les bénéfices de cette approche. A l'échelle d'un corpus, ce caractère est extrêmement rare ($P < 0,0002\%$). Il est en effet presque exclusivement utilisé dans l'expression *tāotíè 餐餐* qui désigne une créature mythique. Sa rareté le place dans la longue traîne de la distribution et les systèmes de transcription automatique ont tendance à le confondre avec le caractère *cān 餐*, « manger » ou « repas », qui est dix fois plus fréquent. Toutefois, au niveau de la ligne, quand il est précédé du caractère *tāo 餐*, la probabilité de *tiè 餐* est proche de 1 ($P(\text{餐}|\text{餐}) \approx 1$). Cette forte probabilité conditionnelle devient un signal puissant qui facilite l'apprentissage et la reconnaissance des caractères rares.

MiDRASH : reconstruire la culture juive médiévale écrite

Daniel STÖKL BEN EZRA (EPHE – PSL)

Le projet ERC synergy MiDRASH, porté par l'EPHE – PSL en collaboration avec l'Université de Tel Aviv, l'Université de Bar Ilan ainsi que la Bibliothèque nationale d'Israël et l'Université de Haïfa, vise à reconstruire et analyser la culture juive écrite jusqu'au XVI^e siècle en hébreu, en araméen et en judéo-arabe principalement.

Plus de 10 millions d'images venant de 80 000 manuscrits et 300 000 fragments (sans compter les 10 millions d'images de livres imprimés) sont réunies sous format électronique à la Bibliothèque nationale d'Israël. Ce fonds rassemble pratiquement tous les manuscrits qui ont survécu aux guerres, aux parasites et aux persécutions au fil des siècles.

Ces manuscrits et fragments ont profité des techniques avancées de transcription automatique de l'équipe de Daniel Stökl Ben Ezra à l'EPHE – PSL pour ensuite être soumis au traitement automatique de l'équipe de Avi Shmidman de l'Université de Bar Ilan, spécialisée sur d'autres dimensions (morphologie et syntaxe, lexique...).

La transcription automatique, fournie par eScriptorium et son moteur IA Kraken, fonctionne déjà au-delà de nos espérances. Le taux de précision de cette transcription atteint 96-99 % pour les manuscrits en écriture livresque. Il reste toutefois des difficultés à surmonter pour transcrire les écritures cursives comme les mises en page complexes.

L'équipe de Daniel Stökl Ben Ezra vient d'achever une transcription complète de tous les fragments de la Guéniza du Caire (plus de 400 000 documents), le graal pour reconstituer tous les codices dispersés dans des collections diverses aux quatre coins du monde.

Au volet de la transcription automatique, s'ajoutent ceux de l'analyse paléographique - à la fois manuelle (par l'équipe de Judith Olszowy-Schlanger de l'EPHE – PSL) et automatique (par l'équipe de Nachum Dershowitz à l'Université de Tel Aviv) - et de la mise en page (par les équipes EPHE – PSL).

Le cœur du projet consiste en sept études de cas, dont le résultat sera une base de données liant plusieurs dizaines de milliards de mots et de lettres, formes et contenus, de la granularité la plus fine jusqu'à la plus grande échelle. Le service devrait être supérieur à Google Books pour les manuscrits juifs, en rendant possible le traçage de chaque unité textuelle ou manuscrite dans le temps et dans l'espace.

Cette combinaison permettra de suivre la transmission et la migration des idées écrites, l'ADN de la littérature intellectuelle juive, au sein d'un vaste espace allant de l'Iran jusqu'à la France, en passant par l'Irak, la Palestine, le Yémen, l'Egypte et l'Italie. Une occasion unique de percevoir les influences des cultures musulmanes et chrétiennes. Il serait par exemple possible d'analyser la paléographie des corrections interlinéaires et marginales et de retracer les détenteurs des livres pour écrire

une histoire des lecteurs - ou des bibliothèques - à grande échelle. Nous pourrions travailler comme les astronomes qui, grâce à une quantité énorme des données et des ordinateurs de très haute puissance, peuvent comparer les amas de galaxies tout en ayant accès aux étoiles ou même planètes ou lunes. Impensable sans eScriptorium.

Transcription automatique de Kraken superposée au manuscrit.
Cambridge University Library, T-S Ar. 50.3 1v



CATMuS : lire les langues de l'Europe au Moyen Âge

Thibault CLÉRICE (Inria)

Dans le monde des manuscrits médiévaux, chaque projet de transcription avait jusqu'ici ses propres règles : certains développant les abréviations quand d'autres les conservaient ; les graphies anciennes étaient parfois uniformisées, parfois respectées jusqu'à dans ses variations les plus précises.

Ces différences, potentiellement anodines pour un lecteur humain, sont un casse-tête pour l'IA, qui apprend à partir des données qu'on lui fournit. Résultat : les modèles et données de reconnaissance d'écriture restaient difficilement comparables, et souvent peu réutilisables.

De ce constat est né CATMuS Medieval. Ce projet international, né en 2021 dans un séminaire d'Ariane Pinche (CNRS) à l'École nationale des Chartes – PSL et soutenu par l'EPHE – PSL, veut créer un terrain commun pour la transcription des manuscrits et imprimés anciens.

Aujourd'hui, son développement est mené de concert entre le CNRS et l'Inria. Plus de 300

documents, 10 langues et près de 200 000 lignes de texte, du VIII^e au XVI^e siècle sont ainsi réunis. Sa force : proposer des règles partagées qui concilient fidélité aux sources et lisibilité pour la machine.

Quand on transcrit un document ancien, un certain nombre de phénomènes viennent bousculer notre rapport au texte : les copistes abrègent avec de nombreux signes, parfois rares, peuvent faire des « fautes »... Ces pratiques changent en fonction des langues (on n'abrége pas autant en français qu'en latin par exemple), des époques, des genres (on abrège plus les documents médicaux que les poèmes).

CATMuS a réussi à fédérer des équipes de plusieurs pays, mutualisé des données dispersées et proposé une réponse commune à ces phénomènes. En les uniformisant, CATMuS permet de bâtir des modèles plus robustes, capables de lire au-delà d'un seul manuscrit, d'un seul genre, d'un siècle ou d'une langue, preuves à l'appui. Stabiliser ces pratiques, c'est réfléchir à la manière dont nous construisons, avec l'aide des machines, notre accès au passé.

L'écosystème eScriptorium et son infrastructure ont permis à CATMuS de voir le jour en annotant quelque 300 manuscrits. CATMuS est devenu l'un des jeux de données et des modèles de base les plus utilisés pour les écritures latines. Aujourd'hui, CATMuS et eScriptorium avancent de concert, formant un socle commun pour l'étude et la transmission des écritures médiévales.

<i>Caroline</i> 8 th -13 th	<i>Prægæthica</i> 12 th -13 th	<i>Gothica Textualis</i> 13 th -16 th	<i>Semientualis</i> 13 th -16 th	<i>Cursiva</i> 14 th -15 th	<i>Hybrida</i> 15 th -16 th	<i>Humanistic</i> 16 th	<i>Incunabulum</i> 15 th
eute ^x t euteric ^x qui iner ^x pre ^x for eutac ^x eutex t euteric ^x qui iner ^x pre ^x fortunat ^x iota qb; ifort ^x ymago d ^x ppe sic picta ^x e qre paunon ^x qre anser ^x gallina refug ^x	euteric ^x qui iner ^x pre ^x for eutac ^x eutex t euteric ^x qui iner ^x pre ^x fortunat ^x iota qb; ifort ^x ymago d ^x ppe sic picta ^x e qre paunon ^x qre anser ^x gallina refug ^x	eute ^x t euteric ^x qui iner ^x pre ^x for eutac ^x eutex t euteric ^x qui iner ^x pre ^x fortunat ^x iota qb; ifort ^x ymago d ^x ppe sic picta ^x e qre paunon ^x qre anser ^x gallina refug ^x	omo que creo no fuesen ^x C omo que creo no fuessen	potūt hēr i bona q̄titate f̄ q̄ tibi videb̄t potūt hēr i bona q̄titate f̄ q̄ tibi videb̄t domos .s. uii milia ho ³ quos 7 ipē inandria sic aut inopia & cognatorum negligē	potūt hēr i bona q̄titate f̄ q̄ tibi videb̄t potūt hēr i bona q̄titate f̄ q̄ tibi videb̄t domos .s. uii milia ho ³ quos 7 ipē inandria sic aut inopia & cognatorum negligē	De Sperchio fluui. xxiiii. oceāi filio De Sperchio fluui. xxiiii. oceāi filio	Typologie d'écritures médiévale

Syriac Transcribathon : un exercice collaboratif pour décrypter des textes syriaques

Christine ROUGHAN (Université de Princeton) et George KIRAZ (Beth Mardutho, Institut d'études avancées, Princeton)

L'utilisation stratégique d'algorithmes associée aux contributions participatives de chercheurs spécialisés dans le patrimoine culturel pourrait rapidement produire des modèles d'IA pour les manuscrits médiévaux.

C'est qui a été démontré en mars 2024 à l'Université de Princeton, dans le cadre du "Syriac Transcribathon", un événement hybride, visant à lancer un modèle d'IA capable de transcrire automatiquement le texte de manuscrits rédigés dans l'une des trois principales écritures syriaques. Un modèle aussi général doit néanmoins s'appuyer sur une grande quantité de données pour apprendre. Le Transcribathon s'est donc concentré sur les manuscrits contenant des textes bibliques. En fournissant un algorithme d'alignement de texte performant, eScriptorium permet de combiner ces manuscrits avec les textes numériques existants de la Bible syriaque tout en générant rapidement des correspondances plausibles entre les lignes des manuscrits et le texte numérique.

Le transcribathon a été plus spécifiquement utile pour mener à bien cet alignement des données grâce à l'expertise (humaine celle-là) de participants : 4 jours durant des locuteurs natifs et des spécialistes du syriaque, venus de la région, d'Europe et du Moyen-Orient, ont

été réunis pour apprendre à utiliser eScriptorium avec des manuscrits prémodernes. Ils ont examiné les données alignées et les ont modifiées afin de s'assurer qu'elles reflétaient fidèlement ces manuscrits. Le logiciel eScriptorium a finalement été utilisé pour entraîner une IA sur 38 manuscrits datant du VI^e jusqu'au XX^e siècle. Finalement, cet effort collaboratif a abouti à un modèle capable de lire des textes syriaques vieux de plusieurs siècles avec une précision supérieure à 97 %.

"Syriac Transcribathon" est un événement hybride, visant à lancer un modèle d'IA capable de transcrire automatiquement le texte de manuscrits rédigés dans l'une des trois principales écritures syriaques.

Feuillet du XI^e siècle provenant d'un manuscrit liturgique syriaque. Collection privée de John Ingels (1925-2016).



L'initiative Open Islamicate Texts (OpenITI) : étudier les textes arabes

Matthew MILLER (University of Maryland),
Sarah SAVANT (Aga Khan University),
David SMITH (Northeastern University)

Tandis qu'il y a depuis longtemps des grands corpus de textes électroniques en latin ou en grec, et même certains en arabe, on a plus de mal à s'imaginer une seule collection des textes islamiques, avec un seul format, qui pourrait être accédé et analysé à grande échelle. Depuis sa création en 2016, l'initiative OpenITI a pour but de répondre à ce manque.

Ce projet de recherche implique plusieurs partenaires et vise à développer l'infrastructure numérique nécessaire à l'étude des cultures islamiques. Pendant des décennies, différents acteurs se sont livrés sans concertation à une gigantesque entreprise de transcription manuelle, parfois en double emploi, de plusieurs milliers de livres. Il est arrivé à plusieurs reprises, que le même livre ait été transcrit manuellement par plusieurs équipes. OpenITI s'est donc concentré sur le développement de solutions numériques pour l'amélioration de la transcription automatique (OCR) de textes manuscrits et imprimés en écriture arabe, leur analyse computationnelle et l'élaboration de standards d'encodage.

eScriptorium étant *Open Source*, nous avons décidé d'utiliser les ressources du projet "OCR Catalyst Project" (AOCP), financé par la fondation Mellon, pour améliorer l'interface utilisateur d'eScriptorium et pour ajouter de

nouvelles fonctionnalités adaptées aux besoins spécifiques de notre projet. Par exemple, nous avons soutenu l'intégration de l'alignement de la sortie OCR d'un manuscrit avec des livres imprimés déjà numérisés. Cela réduit considérablement le temps de préparation pour entraîner l'IA dans ce cas comme pour bien d'autres utilisateurs d'eScriptorium. Désormais cette interface et ces fonctionnalités peuvent servir à tous les autres utilisateurs d'eScriptorium. Le code source ouvert de Kraken nous a également servi d'inspiration pour créer une chaîne de traitement OCR de masse sur mesure. Et comme il est gratuit, nous pouvons numériser un grand nombre

d'œuvres à moindre coût, via le serveur de l'institution hôte du projet, l'Université du Maryland (USA). Grâce à cela, s'ajouteront bientôt 2 000 œuvres récemment "OCÉRisés". Dans une prochaine phase nous nous pencherons sur des manuscrits, leur transcription et leur catalogage automatiques.

Ben Kiessling est en effet parvenu à ce que le système apprenne par lui-même à reconnaître les points diacritiques de l'écriture arabe autour d'une ligne de base, et ce quel que soit leur éloignement.

Le lien que nous avons avec l'équipe porteuse d'eScriptorium (EPHE – PSL) nous a permis d'avoir une influence décisive sur l'architecture du module d'IA Kraken. Ben Kiessling

est en effet parvenu à ce que le système apprenne par lui-même à reconnaître les points diacritiques de l'écriture arabe autour d'une ligne de base, et ce quel que soit leur éloignement. Un modèle de langage assure de son côté une transcription lisible malgré le caractère erratique de ces points dans les manuscrits.

ERC PRIMA : les manuscrits italiens du XVI^e au XVIII^e siècle

Serena CRESPI et Elena PIERAZZO (U. Tours)

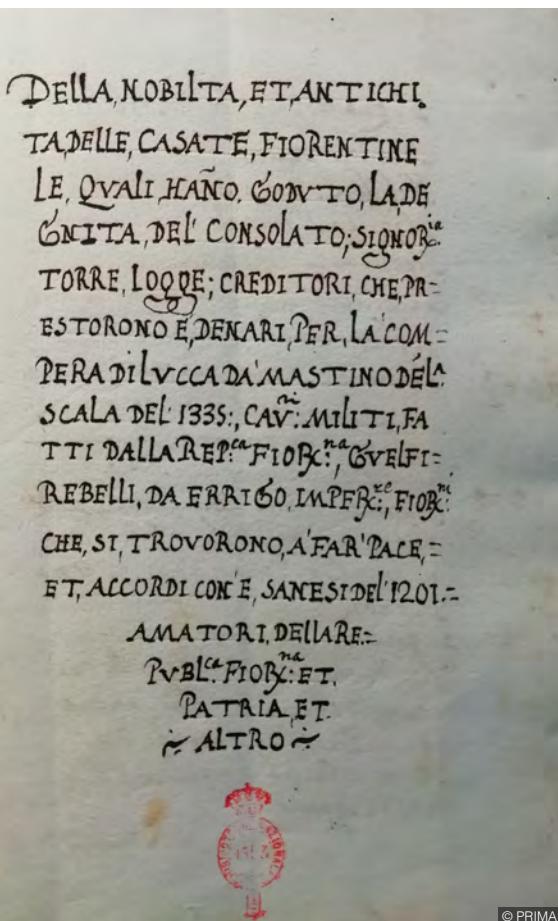
Depuis novembre 2024, le projet européen PRIMA (Manuscripts in the Age of Print), financé par le Conseil européen de la recherche (ERC) et hébergé au Centre d'Études Supérieures de la Renaissance (CESR) de l'Université de Tours, s'intègre pleinement à l'écosystème eScriptorium en tant que terrain d'expérimentation pour l'étude de corpus manuscrits modernes (XVI^e–XVIII^e siècle).

Le projet vise à explorer les potentialités offertes par la reconnaissance automatique d'écriture manuscrite (OCR) appliquée à des traditions textuelles peu stabilisées, tant du point de vue du graphisme que des pratiques d'écriture. L'utilisation d'eScriptorium a permis aux chercheurs de PRIMA de se former et de développer des modèles d'OCR qui n'auraient pas été possibles si eScriptorium ou Kraken n'étaient pas *Open Source*. En effet, les

écritures modernes ont des spécificités graphiques qui posent des difficultés pour la constitution de la vérité de terrain (jeu de données vérifiables et vérifiées) nécessaire à l'entraînement de la machine ; pour y remédier, des données artificielles - images manuscrites générées automatiquement pour simuler la diversité des graphies - ont été générées afin d'enrichir le corpus et d'accroître la robustesse des modèles de 80 à 98 %. De plus, beaucoup de photos de nos objets ont été prises par des téléphones portables selon différents angles et expositions. Là aussi, la possibilité d'ajouter des images créées artificiellement était la solution.

S'appuyant sur ces fondations, le projet PRIMA est en train d'explorer le potentiel de l'IA appliquée à la reconnaissance des écritures manuscrites, en développant une chaîne de traitement fondée sur les données synthétiques et des approches de partitionnement des graphies et des caractères.

Néanmoins, la transcription n'est pas une fin en soi. Celles obtenues servent de base à des analyses plus poussées : repérage de caractéristiques graphiques, identification des éléments de la page (texte principal, ajouts dans les marges des pages,...), classification automatique de différences entre les versions et regroupement des copies similaires. Cette première phase a également ouvert la voie à des réflexions méthodologiques plus larges, notamment pour améliorer la compréhension de la façon dont les textes ont été copiés, modifiés ou transmis au fil du temps.



Extrait du manuscrit Magliabechiano XXVI.14, fol. 1r, conservé à la Bibliothèque nationale centrale de Florence.

DeLiCaTe : déchiffrer le patrimoine écrit du Caucase ancien

Jost GIPPERT (Université de Hambourg)

Les littératures arménienne et géorgienne ont connu un développement continu depuis l'apparition de leurs alphabets au V^e siècle ap. J-C. Pendant longtemps, ces littératures ont été transmises par des manuscrits, dont beaucoup sont conservés.

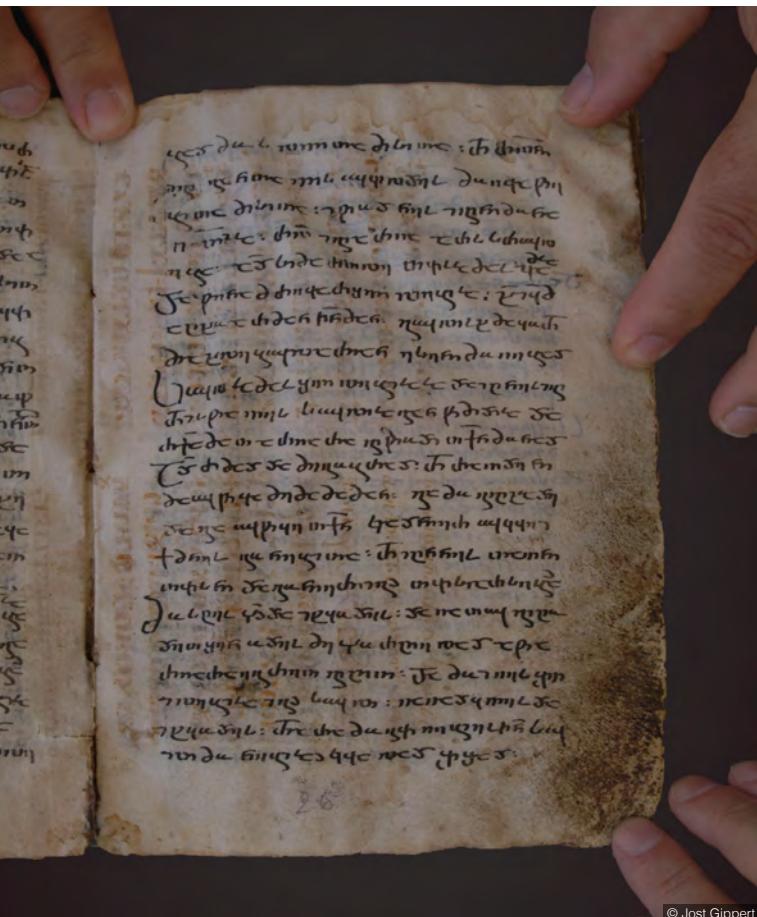
Cet immense patrimoine culturel n'a été exploré que de manière partielle et beaucoup d'œuvres sont encore inédites. Afin d'accélérer ce nécessaire travail d'édition, le projet DeLiCaTe (à Hambourg) a commencé en février 2024 à développer

des modèles eScriptorium pour la lecture automatique des manuscrits. Sur la base des modèles préexistants pour le latin et l'hébreu, trois adaptations différentes ont été réalisées : en géorgien, pour les majuscules (taux de précision actuel : 93,8 %) et pour les manuscrits mixtes (88,7%) et en arménien (90,2 %). Ces développements ont été grandement facilités par le caractère intuitif du processus et de l'interface d'eScriptorium et par le soutien continu de Daniel Stökl Ben Ezra.

La base de développement comprenait les images de 296 pages de manuscrits qui avaient été transcrits manuellement et 118 transcriptions automatiques alignées sur des éditions existantes pour le géorgien, et 36 images transcrits manuellement pour l'arménien.

Malgré des problèmes particuliers, notamment des défauts de lisibilité dus à des dommages sur les manuscrits originaux et la qualité inégale des images disponibles, 18 554 pages géorgiennes et 6 189 pages arméniennes ont été lues avec succès au cours des quinze derniers mois.

Depuis avril 2025, nous avons installé une instance locale au CHAI de Hambourg, ce qui attire de plus en plus d'utilisateurs.



Ms. Sin. georg.
NF 13, fol. 4v:
Géorgien avec
sous-titres
albanais

© Jost Gippert

Conclusion

« Aussi ouvert que possible, aussi fermé que nécessaire »

Peter STOKES (EPHE – PSL)

Comme nous l'avons évoqué dans les contributions précédentes, eScriptorium et la recherche contemporaine reposent notamment sur le principe de la *Science Ouverte*.

Cela concerne les logiciels mais aussi les modèles d'IA et leurs données d'entraînement. Cependant, dans certains cas, cette ouverture n'est ni possible ni souhaitable, ce qui nous amène à un objectif plus nuancé : « aussi ouvert que possible, aussi fermé que nécessaire ». Le cas le plus évident concerne les données personnelles, parfois sensibles, de personnes vivantes, qu'il est possible de retrouver des documents contemporains. Dans la pratique, les textes traités habituellement dans eScriptorium ne posent aucun problème.

Le cadre éthique et juridique posé par le RGPD (règlement général sur la protection des données), fixant l'usage approprié de ces données, est bien entendu à respecter. Il s'applique à la publication de transcriptions de la vérité de terrain, mais cela peut également s'appliquer aux modèles entraînés, dont certains ont été présentés dans ce dossier.

Des études ont montré qu'il peut être possible de récupérer les données utilisées pour l'entraînement à partir de ces modèles par le biais de ce qu'on appelle un « attaque par reconstruction ». En outre, les images comme les textes sont protégées par le droit d'auteur. A cet égard, les chercheurs ne sont pas autorisés à les publier dans le cadre d'un jeu de données comme l'exige les principes

de la *Science Ouverte*. Le plus facile reste d'ignorer ces documents et à en choisir d'autres pour l'entraînement des modèles d'IA, mais cela n'est pas toujours possible ! On peut débattre du bien-fondé ou non de l'utilisation de ces données, mais leur exclusion *a priori* est très contraignante.

En outre, le téléchargement de données personnelles vers des services cloud suscite (à juste titre) de plus en plus d'inquiétudes. Il est le plus souvent impossible de savoir comment ces données sont utilisées par les plateformes opérant en dehors de l'UE et de la législation qui s'y applique. C'est une autre raison pour laquelle nous laissons aux utilisateurs la possibilité d'installer

Kraken et eScriptorium sur leurs propres machines s'ils souhaitent conserver la maîtrise de leurs données.

Enfin, il existe des considérations financières concrètes pour les universités. L'IA est désormais un secteur d'activité important, et certains soutiennent qu'elles devraient protéger leur propriété intellectuelle et la vendre afin d'assurer leur viabilité à long terme. Nous avons choisi de ne pas le faire, en rendant Kraken et eScriptorium gratuits et ouverts, parce que nous estimons que la recherche financée par des fonds publics doit rester publique.

Toutefois, le débat reste ouvert. Et si nous encourageons les utilisateurs à ouvrir leurs données et modèles, l'usage d'eScriptorium n'oblige en rien à les partager selon les principes de la *Science Ouverte*.

Il est le plus souvent impossible de savoir comment ces données sont utilisées par les plateformes opérant en dehors de l'UE et de la législation qui s'y applique. C'est une autre raison pour laquelle nous laissons aux utilisateurs la possibilité d'installer Kraken et eScriptorium sur leurs propres machines s'ils souhaitent conserver la maîtrise de leurs données.

Lexique

ALMAnaCH

Automatic Language Modelling and Analysis & Computational Humanities, une équipe de recherche de l'Inria.

ANR

Agence Nationale de la Recherche, l'agence française de financement de la recherche sur projets.

AOrOc

Archéologie et Philologie d'Orient et d'Occident (CNRS/ENS – PSL/EPHE – PSL, UMR 8546), le laboratoire de l'équipe eScriptorium.

CATMuS

Consistent Approaches to Transcribing Manuscripts, un projet pour établir les standards de la transcription et pour publier des données pour l'OCR.

CBETA

Chinese Buddhist Electronic Text Association, une plateforme qui donne accès à des textes bouddhistes en chinois sous forme numérique.

CNRS

Centre National de la Recherche Scientifique

CREMMA

Consortium Reconnaissance d'Écriture Manuscrite des Matériaux Anciens, un projet de l'Inria, l'ENC – PSL, l'IRHT, l'EFEO et l'EPHE – PSL pour la création d'un service de mise à disposition de ressources pour favoriser l'accès à l'OCR pour les manuscrits.

DeLiCaTe

The Development of Literacy in the Caucasian Territories, un projet de recherche sur le développement de l'alphabétisation dans le Caucase du Sud, en particulier du V^e au X^e siècle.

DIM STCN

Domaine d'intérêt majeur Science du Texte et Connaissances nouvelles, un des réseaux scientifiques pluridisciplinaires de la région Île-de-France.

FonDUE

Formes Numérisées et Détection Unifiée des Écritures, un projet de recherche à l'Université de Genève pour créer une infrastructure pour l'OCR des documents manuscrits.

Inria

Institut National de Recherche en Informatique et en automatique, un établissement public pour la recherches et innovations dans le domaine du numérique.

LectAuRep

LECTure Automatique de REpertoires, un projet de recherche conjoint du Minutier central des notaires de Paris des Archives Nationales, de l'équipe ALMAnaCH et de l'EPHE – PSL en partenariat avec le Ministère de la Culture, sur les registres des actes notariés.

LLM

"Large Language Model", un grand modèle de langage entraîné avec une très grande quantité de texte (milliards et billions de mots).

MiDRASH

Migrations of Textual and Scribal Traditions via Large-Scale Computational Analysis of Medieval Manuscripts in Hebrew Script, un projet sur l'analyse computationnel à grande échelle des manuscrits médiévaux hébreux.

MSIA

ManuscriptologIA, une installation d'eScriptorium gérée par l'EPHE – PSL et soutenue actuellement par Biblissima+.

OCR

“Optical Character Recognition”, la reconnaissance automatique des caractères ou des textes, c'est à dire l'extraction de texte d'une image. Équivalent de l'HTR (reconnaissance automatique de l'écriture manuscrite).

OpenITI

“Open Islamicate Texts Initiative”, un projet de recherche des universités Aga Kahn (Londres), du Maryland (USA) et de Hambourg (Allemagne). Financé par la Fondation Mellon, il apporte les capacités nécessaires aux études islamiques.

PRIMA

“Manuscripts in the Age of Print (1575-1800)” est un projet de recherche sur les manuscrits italiens du XVI^e au XVIII^e siècle et sur leur relation avec les livres imprimés.

Transcribathon

Un événement d'une ou plusieurs journées où les personnes intéressées peuvent se réunir pour produire des transcriptions à la main, normalement avec l'objectif de les utiliser pour entraîner eScriptorium ou un autre système OCR.

nos publications

Bibliothèque de l'École des hautes études, Sciences religieuses, 209 (série Etudes persanes)

Sous la direction de
Orkhan Mir-Kasimov
et Mathieu Terrier

De la lettre à l'esprit *Travaux en hommage à Mohammad Ali Amir-Moezzi*

La parution du *Guide divin dans le shī'isme originel* en 1992 a marqué le début de l'itinéraire scientifique de Mohammad Ali Amir-Moezzi. Suivant une ligne directrice cohérente aux ramifications nombreuses, cet islamologue a transformé en profondeur les études shī'ites d'abord, notamment en soulignant l'importance de la « tradition ésotérique originelle » dans l'histoire de ce grand courant de l'islam, et les études coraniques ensuite, par la prise en compte des sources shī'ites anciennes et la critique du récit traditionnel « orthodoxe ». Enseignant depuis 1984 à la section des Sciences religieuses de l'école pratique des hautes études où il a détenu, de 1996 à 2024, la chaire « Exégèse et théologie de l'islam shī'ite », Mohammad Ali Amir-Moezzi est également un professeur hors du commun dont l'enseignement a formé et inspiré de nombreux chercheurs actuels. Cet ouvrage, en deux volumes, réunit quarante-quatre contributions de ses collègues et amis, dont nombre d'anciens étudiants, français et étrangers, portant sur les domaines d'études chers au dédicataire : études shī'ites, études coraniques, antiquité tardive et débuts de l'islam, traditions mystiques de l'islam, aspects de la vie religieuse et intellectuelle contemporaine. Témoignages d'amitié et de reconnaissance pour une œuvre scientifique majeure, les contributions savantes réunies dans ces deux volumes en font un ouvrage de référence.



Bibliothèque de l'École des hautes études, Sciences religieuses, 210

Sous la direction de
Anna Van den Kerchove,
Mohammad Ali Amir-Moezzi,
Pierre-Olivier Léchot

Protestantisme et islam. regards croisés (XVI^e-XXI^e siècle) *Actes du colloque organisé par l'Institut protestant de théologie et l'École pratique des hautes études, 19-21 janvier 2022*

Comment une confession chrétienne considère-t-elle une religion non chrétienne ? Peut-on déceler certaines spécificités dans son approche de cette religion en regard des autres traditions confessionnelles ? Et réciproquement, comment les représentants de cette religion non chrétienne perçoivent-ils les différences entre confessions au sein du christianisme ? Ce sont à ces questions somme toute assez générales que les contributions à ce volume collectif ont cherché à répondre en concentrant leurs investigations sur les regards croisés entre islam et protestantisme. L'orientation du volume se veut clairement historique et sociologique – pour le dire en d'autres termes : notre souci n'était pas prioritairement celui du dialogue interreligieux, même si les éditeurs du présent volume sont bien conscients de l'importance des travaux réunis ici pour un tel dialogue.



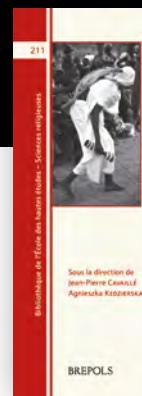
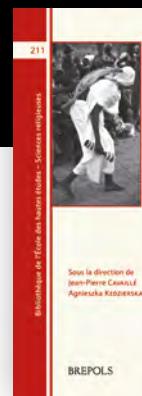
Bibliothèque de l'École des hautes études, Sciences religieuses, 211

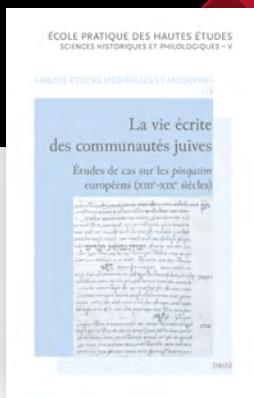
Sous la direction de Jean-Pierre Cavaillé et Agnieszka Kedzierska Manzon

Possession *Représentaions, pratiques, interprétations*

Issue d'un long processus de maturation, d'abord dans un cadre religieux chrétien puis, à partir du XIX^e siècle, dans un cadre médical où la psychiatrie substitue à l'action diabolique l'explication pathologique, la notion de possession renvoie toujours à une forme d'altérité – qu'elle soit religieuse, biologique, sociale, géographique, ethnique ou raciale. Elle est souvent associée à un ailleurs, qu'il soit spatial ou temporel. Pourtant, à la lecture des contributions rassemblées dans ce volume, la possession apparaît comme un phénomène universel, présent dans une multitude de sociétés anciennes et contemporaines. Elle se présente comme l'une des expériences ou pratiques religieuses les plus répandues et partagées dans le monde. On pourrait la qualifier de pratique « ordinaire », mais ce serait oublier qu'elle soulève partout des questions, résiste aux analyses trop simplificatrices ou généralisantes et demeure irréductible dans la diversité infinie de ses manifestations.

Ce livre rassemble des spécialistes en histoire et en anthropologie qui explorent la possession à travers ses pratiques culturelles, les représentations qu'en donnent les acteurs comme ceux qui s'en distancient, les discours qui l'entourent, ainsi que les interprétations qui évoluent au gré des transformations des paradigmes philosophiques et scientifiques. L'ambition est double : dresser un état des lieux des connaissances actuelles et ouvrir de nouvelles pistes de réflexion. L'ouvrage s'organise autour d'une série d'études de cas, précédées d'une introduction à trois voix qui propose un réexamen de la notion, en tenant compte à la fois de son histoire et de ses usages les plus contemporains.





Hautes Études médiévales et modernes, 118

Édité par Elena N. Barile & Miruna-Ştefana Belea

La vie écrite des communautés juives *Études de cas sur les pinqasim européens (XII^e-XIX^e siècle)*

L'étude des manuscrits en caractères hébreuques a une longue histoire à l'École pratique des hautes études. En 2009, Judith Olszowy-Schlanger et ses doctorants se sont engagés dans l'organisation d'une journée visant à réunir élèves et chercheurs, venant de l'EPHE – PSL et de l'extérieur, tous spécialistes des manuscrits hébreux. Cette première rencontre a inauguré les Journées d'études doctorales et postdoctorales en paléographie et codicologie hébreuques qui, depuis, se sont tenues chaque année sans interruption. Ce volume recueille huit études liées à la XIII^e journée d'étude intitulée Pinqasim. *Écrire la communauté juive en Europe médiévale et moderne*. À côté d'une introduction essentielle aux registres juifs au Moyen Âge et à l'époque moderne, l'on trouvera des synthèses méthodologiques à partir d'études de cas particuliers : le rôle socio-historique des registres communautaires en Italie et le changement de perspective apporté par la lecture dialogique entre les documents hébreuques et latins, ainsi que les caractéristiques identitaires des communautés juives nouvellement formées en Europe ashkénaze. Les contributeurs intègrent également le rôle de la codicologie dans l'étude des registres, comme les caractéristiques spécifiques de la mise en page italienne et la circulation du papier parmi les utilisateurs de *pinqasim* en Pologne.

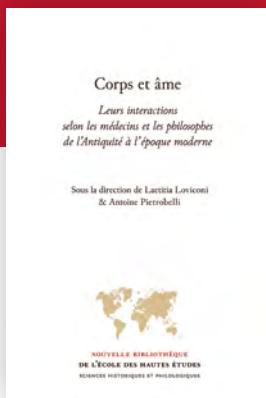


Nouvelle Bibliothèque de l'Ecole des Hautes Études, Sciences historiques et philologiques, 1

Jérôme Petit

La mission de Charles d'Ochoa *Un jalon dans l'histoire des collections indiennes de la Bibliothèque nationale de France*

En s'appuyant sur des sources primaires inédites, cet ouvrage retrace l'histoire des collections indiennes de la Bibliothèque nationale de France, depuis les premiers manuscrits envoyés de l'Inde par les jésuites dans les années 1730 jusqu'aux acquisitions les plus récentes. Des aventuriers rapportèrent des manuscrits importants dans la seconde moitié du XVIII^e siècle, avant que des philologues comme Eugène Burnouf organisent des collectes de corpus, tout au long du XIX^e siècle, pour nourrir les différents champs des études indiennes naissantes. Le XX^e siècle, dominé par la figure de Jean Filliozat, vit cet effort se poursuivre. L'ouvrage s'attarde ensuite sur la figure de Charles d'Ochoa (1816-1846), un jeune indianiste qui mena une mission en Inde centrale, financée par le ministère de l'Instruction publique, afin de collecter les matériaux nécessaires à la rédaction d'une histoire des littératures indiennes. Journaux de voyages, lettres et rapports offrent une meilleure compréhension des enjeux et des implications des missions scientifiques en ce milieu du XIX^e siècle. L'étude matérielle et philologique de la collection d'Ochoa permet d'en dégager les lignes de force. En parallèle des documents d'archives, la lecture des colophons et des paratextes des manuscrits permet de suivre le collectionneur dans sa démarche et de mettre au jour son réseau savant, dans le cadre général de l'histoire des études indiennes. Face aux antiquisants à la recherche de la langue des origines, Charles d'Ochoa fait figure de moderniste et offre l'image d'une Inde où se mêle une multiplicité de langues et de cultures.



Nouvelle Bibliothèque de l'Ecole des Hautes Études, Sciences historiques et philologiques, 2

Sous la direction de Laetitia Loviconi & Antoine Pietrobelli

Corps et âme *Leurs interactions selon les médecins et les philosophes de l'Antiquité à l'époque moderne*

Comment l'âme interagit-elle avec le corps ? Ces deux entités sont-elles unies ou séparées ? Quel est leur point d'intersection ou quels sont les médiateurs entre le corps et l'âme ?

Le présent volume, qui s'inscrit dans la perspective d'une histoire intellectuelle et conceptuelle au long cours, regroupe onze contributions qui explorent comment différents auteurs, depuis l'Antiquité jusqu'au XVIII^e siècle, ont questionné et analysé les modalités d'interaction et d'union du corps et de l'âme.

Les contributions sont regroupées en fonction de leur période (Antiquité, Moyen Âge, période moderne) et abordent les sujets suivants : les relations physiologiques qu'entretiennent le corps et l'âme considérés dans leur globalité, la question des facultés végétales et végétatives qui traversent les différentes espèces vivantes, la nature de certains médiateurs comme le *pneuma* ou *spiritus* et les *species* ou encore ce que révèlent les contextes pathologiques sur ces liens entre le corps et l'âme.

Sur le terrain

L'EPHE – PSL au centre de la recherche à Dijon

Les 5 et 6 novembre 2025, la Direction de la communication de l'EPHE – PSL s'est rendue à Dijon à l'invitation de nos équipes rattachées au laboratoire Biogéosciences (CNRS/ Université de Bourgogne Europe/EPHE – PSL) et au Centre de recherche Translationnelle en Médecine moléculaire (INSERM UMR1231/ Université de Bourgogne Europe/EPHE – PSL).

Nous vous proposons de découvrir en images l'excellence de leurs travaux conduits au cœur d'un territoire marqué par son histoire géologique et figurant parmi les principaux pôles de recherche contre le cancer en France.



L'imagerie, un des points forts au sein de l'UMR Biogéosciences

L'équipe BIOME de Biogéosciences

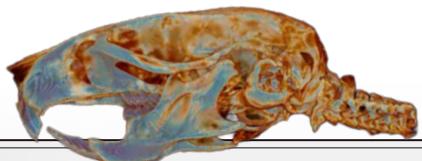
Les personnels de l'EPHE – PSL opèrent à la fois au sein de l'équipe BIOME ainsi que dans le cadre du GRET « Écologie-Évolution » qui s'intéresse aux dynamiques éco-évolutives de la biodiversité à différents niveaux d'organisation depuis les gènes jusqu'aux écosystèmes et différentes échelles temporelles.

Nos chercheurs présents à Dijon étudient plus spécifiquement l'évolution de la morphologie des organismes grâce à l'imagerie 3D et à la morphométrie géométrique. Les objectifs sont multiples, par exemple, de mieux comprendre les réponses morphologiques adaptatives ou non des organismes au changement climatique ou encore d'identifier les gènes contribuant aux variations morphologiques communes ou pathologiques.

Prochainement, l'équipe enverra en Afrique du Sud Laetitia Confuron (doctorante EPHE – PSL) pour travailler sur la communication vocale d'une espèce de primate, afin de comprendre l'importance des variations individuelles de leurs interactions à l'environnement sur cette communication, mais aussi du rôle des différences morphologiques sur ces variations. Restez connectés avec nos réseaux sociaux pour avoir prochainement de leurs nouvelles au Printemps.

Acquisition
2D et 3D de
la forme des
organismes

Rendu volumétrique
d'un CT scan de
souris



La fiche du labo

Ecole
Pratique
Hauts
Etudes

Nom BGS - Biogéosciences

UMR UMR 6282

Ville Dijon

Objectifs

Étudier la morphologie pour mieux comprendre la biodiversité

Comprendre l'histoire évolutive épigénétique (forme des individus)

Équipe EPHE – PSL

Sébastien Couette, maître de conférences, directeur adjoint de l'UMR

Sophie Montuire directrice d'études

Nicolas Navarro, maître de conférences, coordinateur de l'équipe BIOME

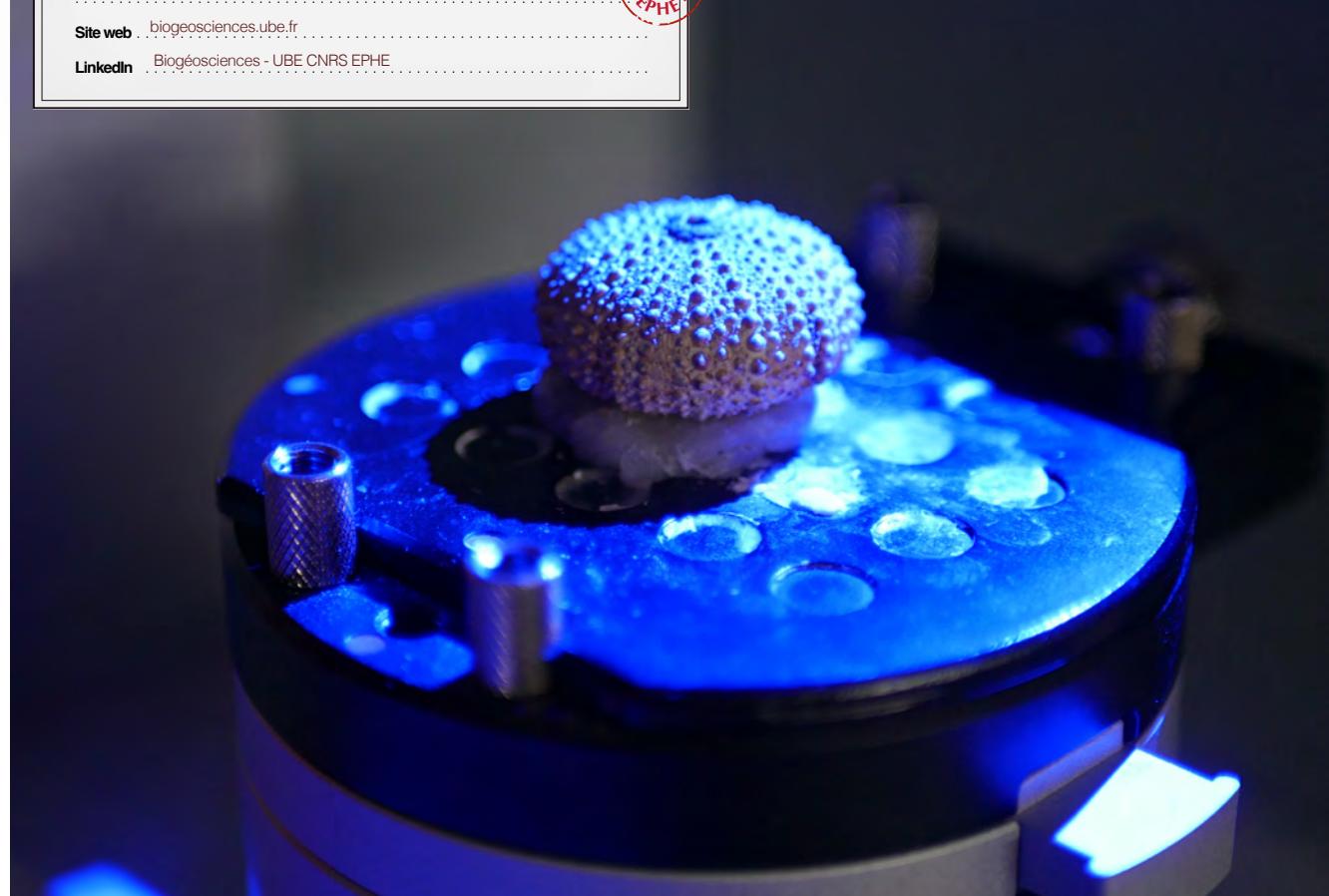
Lauriane Poloni, technicienne de laboratoire

Laëtitia Confuron, doctorante

Myriam Marsot, doctorante

Site web biogeosciences.ube.fr

LinkedIn Biogéosciences - UBE CNRS EPHE



Analyse de la forme par
Morphométrie Géométrique 3D

L'équipe du CTM (Center for Translational and Molecular Medicine)

Nos chercheurs travaillent dans le cadre du GRET « Biologie des cancers » de l'EPHE – PSL. Celui-ci développe des projets de recherche axés sur l'étude des mécanismes associés au développement tumoral et aux thérapies innovantes.

Nos équipes présentes à Dijon étudient plus spécifiquement les mécanismes associés au cancer du sein ainsi qu'aux cancers digestifs et du poumon. Un enjeu important est de comprendre le fonctionnement interne des tumeurs pour susciter des réponses immunitaires adaptées et ciblées. Deux voies sont explorées avec l'effet antitumoral du monoxyde d'azote (NO) et l'utilisation de composés issus de la bactérie pour activer la production de cytokines ou activer les cellules immunitaires. La fluorescence (dont l'étude remonte aux origines de l'EPHE) est aussi utilisée pour mieux visualiser et éliminer les tumeurs par chirurgie.

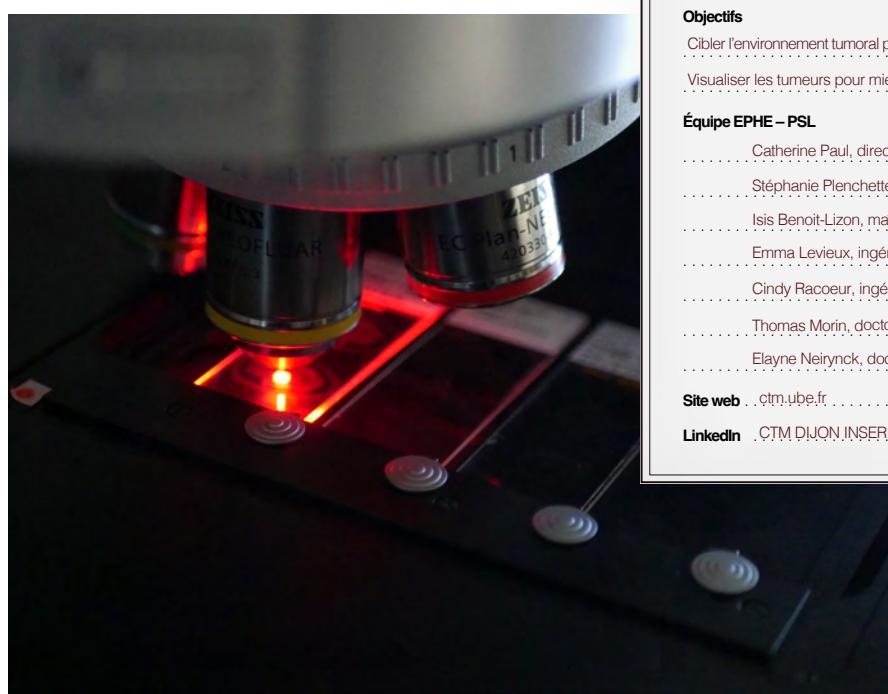
L'équipe a enfin la chance de pouvoir compter sur un réseau doctoral Marie Curie. Ce dernier finance Elayne Neirynck (réseau « NO CANCER NET ») pour développer de nouvelles stratégies contre un type de cancer du sein très agressif grâce à un médicament qui libère du NO.



Membres de l'EPHE – PSL du CTM, Dijon



Immunomarquage
sur coupe de tissus



Microscopie à
fluorescence

La fiche du labo

École
Pratique
Hauts
Études

Nom CTM - Center for Translational and Molecular Medicine

UMR UMR 1231

Ville Dijon

Objectifs

Cibler l'environnement tumoral pour augmenter l'efficacité des traitements anticancéreux

Visualiser les tumeurs pour mieux les éliminer avec la fluorescence

Équipe EPHE – PSL

Catherine Paul, directrice d'études

Stéphanie Plenquette, directrice d'études

Isis Benoit-Lizon, maître de conférences

Emma Léveillé, ingénieur d'études

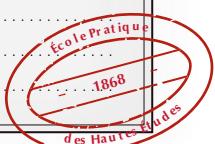
Cindy Racour, ingénieur d'études

Thomas Morin, doctorant

Elayne Neirynck, doctorante

Site web ctm.ube.fr

LinkedIn CTM DIJON INSERM



Enseigner la recherche par la pratique de la recherche

