

Matthias GILLE LEVENSON

*Postdoctorant, Université de Versailles – Saint
Quentin en Yvelines*

Né le 22/05/1993

✉ matthias.gille.levenson@ens-lyon.fr

🌐 [matgille](#)

🆔 0000-0001-9488-5986

🐦 [mgillelevenson](#)

Mise à jour : novembre 2025

Table des matières

Parcours professionnel	1
Diplômes obtenus	2
Formation, concours et qualifications	2
Séjours universitaires et de recherche à l'étranger	3
Expérience d'enseignement	4
Production scientifique	10
Projets financés en cours	15
Intégration dans des comités éditoriaux et de lecture	15
Activités d'encadrement et participation à des jurys	15
Compétences	16
Responsabilités organisationnelles	16

Parcours professionnel

- 2025– **Postdoctorant**, *Université de Versailles Saint-Quentin en Yvelines*, DYPAC
- 2024–2025 **Attaché Temporaire d'Enseignement et de Recherche**, *Aix-Marseille Université*, Centre Aixois d'Études Romanes, UE 854.
- 2023–2024 **Post Doctorant**, *École nationale des chartes*, Centre Jean Mabillon, École Normale Supérieure de Lyon, associé au CIHAM, UMR 5648.
- 2022–2023 **Attaché Temporaire d'Enseignement et de Recherche**, *ÉNS de Lyon*.
- 2020–2022 **Membre scientifique**, *Casa de Velázquez – EHEHL*.
- 2018–2020 **Doctorant contractuel**, *École Normale Supérieure de Lyon*.
- 2013–2018 **Élève fonctionnaire**, *ÉNS de Lyon*.

Diplômes obtenus

- 2023 **Doctorat ès Lettres, spécialité Études Hispaniques, École Normale Supérieure de Lyon**
Titre de la thèse de doctorat : *Le Regimiento de los príncipes et sa glose : étude et édition numérique de la partie sur le gouvernement de la cité en temps de guerre (III, 3)*. Thèse réalisée sous la direction de Carlos Heusch (ENS de Lyon) et de Jesús R. Velasco (Yale University).
- 2016 **Master 2 en langues hispanophones et lusophones, École Normale Supérieure de Lyon**
Titre du mémoire de Master 2 : *La version B du "Regimiento de los príncipes glosado"*. Note obtenue : 19.
- 2014 **Licence LLCER espagnol, Université Lumière-Lyon 2.**

Formation, concours et qualifications

- 2024 **Candidature au poste de Maître de Conférence 4683 de l'Université de Tours (section 27)**
■ Auditionné, classé 2ème
- 2024 **Qualifications aux fonctions de Maître de conférences par le CNU**
■ Section 7 : Sciences du langage; no de qualification : 24207362400.
■ Section 14 : Langues et littératures romanes; no de qualification : 24214362400.
- 2018-2023 **Préparation du diplôme de Doctorat, soutenu le 8 septembre 2023**
■ *Visiting Assistant in Research* à l'Université Yale (septembre – décembre 2019).
■ Suivi (sur invitation) de l'atelier DTS (*Distributed Text Services*) « Texts and API » à Hambourg, du 15 au 16 juillet 2019.
■ Suivi de l'école d'été de l'IRHT « Le livre médiéval au regard des méthodes quantitatives » du 25 au 28 juin 2019.
- 2017-2018 **4^e année d'ÉNS; préparation de la thèse et mise à niveau technique et méthodologique.**
■ Participation à l'atelier 2018 du consortium CAHIER (Montpellier, 26-29 juin 2018).
■ Participation à la session 2018 de la formation EDEEN (École d'été en humanités numériques) à Grenoble, organisée par l'Université Grenoble-Alpes et la MSH Alpes (Grenoble, 28 mai-2 juin).
■ Suivi d'un atelier en édition numérique à Cambridge (Royaume-Uni) en avril 2018, dans la continuité du DEMM (*Digital Encoding of Medieval Manuscripts*).
- 2016-2017 **Agrégation externe d'Espagnol. Rang : 3.**
- 2015-2016 **Master II d'études hispanophones et lusophones, ÉNS de Lyon**
■ Suivi des cours du Master *Patrimonio Textual y humanidades digitales* de l'Université de Salamanque dans le cadre du programme Erasmus.
■ Mémoire préparé sous la direction de Carlos HEUSCH  et de Francisco BAUTISTA  : *La version B du Regimiento de los príncipes glosado*. Note obtenue : 19.
■ Suivi de la formation *Digital Encoding of Medieval Manuscripts* (DEMM) financée par l'Union Européenne. Trois semaines de stage à Vienne (Abbaye de Klosterneuburg), Lyon (ÉNS de Lyon), Londres (Queen Mary University), entre janvier 2015 et mai 2016.

2014-2015 **Master I d'études hispanophones et lusophones et hispanophones**, ÉNS de Lyon

- Spécialisation en philologie castillane et en étude de textes du bas-Moyen Âge. Mémoire préparé sous la direction de Carlos Heusch : *Edición crítica de dos textos de Pedro de Chinchilla*. Note obtenue : 18.
- Stage d'ecdotique des Sources chrétiennes (Lyon, février 2015).

2013-2014 **Licence LLCER espagnol**, Université Lumière-Lyon 2

- Séminaire d'ecdotique de Gisèle BESSON, ÉNS Lyon.

2011-2013 **Préparation et obtention du concours de l'ENS de Lyon** (CPGE, Lycée Joffre, Montpellier).

2011 **Baccalauréat général scientifique, option mathématiques**, Mention TB, Lycée Joffre, Montpellier.

Séjours universitaires et de recherche à l'étranger

2020-2022 Membre scientifique de l'École Française de Madrid – Casa de Velázquez (« École des Hautes Études Hispaniques et Ibériques »).

2019 *Visiting Assistant in Research*, Yale University, auprès de Jesús R. Velasco. Septembre-décembre 2019.

2015-2016 Échange universitaire Erasmus (M2) à l'Université de Salamanque. Suivi des cours du master *Patrimonio textual y humanidades digitales*.

Expérience d'enseignement

Thème principal du cours	Nombre d'heures	Unités concernées
Traduction	165 h.	11, 12, 21, 28, 29, 30
Civilisation médiévale hispanique	96 h.	4, 9, 25, 27, 38
Littérature médiévale hispanique	18 h.	3
Langue espagnole	36 h.	7, 10
Initiation à l'HTR	147 h.	1, 5, 8, 14, 15, 16, 17, 24, 32, 34
XML-TEI : modélisation, encodage, transformation	59 h.	2, 33, 35, 40
Méthodologie	102 h.	6, 26
TAL et annotations linguistiques	59 h.	13, 31, 36, 37, 39
Autres cours en Humanités Numériques	48 h.	18, 19, 20, 23
Total	730 h.	

HTR ET ENCODAGE XML-TEI

[1]

Institution: Univ. de Poitiers / ÉNS de Lyon

Niveau: Doctorat

Années: 2023–2025

Nombre total d'heures: 80 h.

Animation d'un groupe d'une demi-douzaine de personnes de la formation EnExDi entre 2023 et 2025 ; séances co-animées avec Ariane Pinche sur le HTR et la transcription de sources au format XML-TEI.

SCRIPTS ET MANUSCRITS – PYTHON, XSLT, XQUERY

[2]

Institution: ENS de Lyon

Niveau: Doctorat

Années: 2025

Nombre total d'heures: 15 h.

Co-animation en mai 2025 avec Ariane Pinche et Jean-Paul Rehr d'une formation au traitement de sources XML à l'aide de trois langages de requête et de programmation : XPath, Python, XSLT, XQuery. Travail sur un corpus de documents inquisitoriaux en latin. Lien sciencesconf.

LITTÉRATURE MÉDIÉVALE HISPANIQUE

[3]

Institution: AMU

Niveau: L3 LLCER espagnol

Années: 2024–2025

Nombre total d'heures: 18 h. (CM)

Le cours se centre sur l'œuvre du *Cantar de Mio Çid*.

COMPÉTENCES GÉNÉRALES DISCIPLINAIRES

[4]

Institution: AMU

Niveau: L3 LLCER espagnol

Années: 2024–2025

Nombre total d'heures: 24 h. (TD)

Ce TD propose des séances d'analyse de dossiers thématiques sur l'Espagne médiévale entre le 13e et le début du 16e siècle.

INITIATION À L'HTR ET À ESCRIPTORIUM

[5]

Institution: UCM

Niveau: Doctorat

Années: 2025

Nombre total d'heures: 3 h.

Animation en février 2025 d'un atelier doctoral à l'Universidad Complutense de Madrid sur l'HTR et eScriptorium.

MÉTHODOLOGIE DU TRAVAIL UNIVERSITAIRE

[6]

Institution: AMU

Niveau: L1 LLCER espagnol

Années: 2024–2025

Nombre total d'heures: 12 h.

COMPRÉHENSION - EXPRESSION LEA

Institution: AMU

Années: 2024-2025

[7]

Niveau: L1 LEA espagnol
Nombre total d'heures: 24 h.

INTRODUCTION À L'HTR

Institution: Université Paris 8

Années: 2025

[8]

Niveau: Doctorat
Nombre total d'heures: 3 h.

Cours sur la reconnaissance automatique de l'écriture dans le cadre d'un séminaire doctoral organisé par Marta López Izquierdo.

CIVILISATION MÉDIÉVALE HISPANIQUE

Institution: AMU

Années: 2024-2025

[9]

Niveau: L2 LLCER espagnol
Nombre total d'heures: 18 h. (CM)

Le cours s'est attaché à proposer un panorama historique général de la péninsule Ibérique entre le Ve siècle et 1516.

COMPRÉHENSION ORALE LEA

Institution: AMU

Années: 2024-2025

[10]

Niveau: L1 LEA espagnol
Nombre total d'heures: 12 h.

VERSION LLCER

Institution: AMU

Années: 2024-2025

[11]

Niveau: L2 LLCER espagnol
Nombre total d'heures: 12 h.

Cours de version moderne, alternance de textes espagnols et latino-américains entre le XIXe et le XXe siècle.

VERSION LEA

Institution: AMU

Années: 2024-2025

[12]

Niveau: L2 LEA espagnol
Nombre total d'heures: 12 h.

Cours de traduction d'articles de presse actuelle.

INITIATION À L'ALIGNEMENT MULTILINGUE DE TEXTES MÉDIÉVAUX

Institution: École nationale des chartes, Biblissima+

Années: 2023-2024

[13]

Niveau: Doctorat
Nombre total d'heures: 3 h.

Co-Animation d'un atelier avec Lucence Ing consacré à l'alignement multilingue de textes romans médiévaux, dans le cadre des journées des clusters 5 et 7 de Biblissima+ en novembre 2024. Travail à l'aide de notebooks python. Lien vers le matériel.

INTRODUCTION À L'HTR

Institution: ENSSIB

Années: 2023-2025

[14]

Niveau: Master
Nombre total d'heures: 9 h.

Participation au séminaire du master HN de l'ENSSIB/Lyon 3/ENS de Marianne Reboul en octobre 2022, 2023 et 2024; séances en collaboration avec Ariane Pinche sur la transcription automatisée.

PRODUCTION COLLABORATIVE D'UN CORPUS D'ATR

Institution: ÉNS de Lyon

Années: 2024-2025

[15]

Niveau: Doctorat
Nombre total d'heures: 15 h.

Animation d'un groupe (cinq personnes) dans le cadre du transcribathon Transcriboquest 2024 (Lyon, 11-13 septembre 2024) organisé par Thibault Clérice et Ariane Pinche. Production d'un corpus d'ATR de manuscrits littéraires.

INITIATION À HTR – SOURCES CASTILLANES MÉDIÉVALES

[16]

Institution: Universidad del Salvador – Buenos Aires

Niveau: Master et Doctorat

Années: 2024

Nombre total d'heures: 3h.

Animation d'une formation à l'HTR à l'Universidad de El Salvador, au sein du Master Humanités Numériques de Gimena del Rio Riande. Travail sur des manuscrits modernes du *Septennaire*, œuvre juridique d'Alphonse X.

INITIATION À HTR – SOURCES CASTILLANES MÉDIÉVALES

[17]

Institution: ÉNS de Lyon

Niveau: Master et Doctorat

Années: 2023–2024

Nombre total d'heures: 5h.

Animation d'une formation HTR dans le cadre du Séminaire Hispanique d'Études Médiévales (Lyon, 5 avril 2024); séances co-animées avec Ariane Pinche, Olivier Brisville et Julie Bordier. Une vingtaine de participants est prévue.

INITIATION À L'ECDOTIQUE NUMÉRIQUE

[18]

Institution: Universités Grenoble-Alpes

Niveau: Master-Doctorat

Années: 2023–2024

Nombre total d'heures: 10 h.

Animation du colloque-école d'hiver « Du calame au clavier » (Grenoble, 17–19 janvier 2024). Communication et ateliers en HTR et sur la collation semi-automatisée à l'aide de Notebooks Python. Lien vers le matériel : github.

SOCIÉTÉ NUMÉRIQUE ET HUMANITÉS

[19]

Institution: UVSQ

Niveau: L2 LLCER espagnol

Années: 2023–2024

Nombre total d'heures: 18 h.

Présentation et utilisation des outils de visualisation de données (StoryMap, StoryLine, TimeLine, Datawrapper), appliqué à des thèmes de civilisation hispanophone. Présentation et application de méthodes de lecture rapide d'articles journalistiques et scientifiques sur les thèmes étudiés en cours.

SOCIÉTÉ NUMÉRIQUE ET HUMANITÉS

[20]

Institution: UVSQ

Niveau: L2 LLCER anglais

Années: 2023–2024

Nombre total d'heures: 18 h.

Présentation et utilisation des outils de visualisation de données (StoryMap, StoryLine, TimeLine, Datawrapper), appliqué à des thèmes de civilisation anglophone. Présentation et application de méthodes de lecture rapide d'articles journalistiques et scientifiques sur les thèmes étudiés en cours.

VERSION

[21]

Institution: ENS de Lyon

Niveau: Agrégation

Années: 2023–2024

Nombre total d'heures: 21 h.

Cours de version classique et moderne dans le cadre de la préparation à l'Agrégation externe d'espagnol (décembre 2023–janvier 2024).

INITIATION À LA COLLATION AUTOMATISÉE

[22]

Institution: École nationale des chartes

Niveau: Doctorat

Années: 2023–2024

Nombre total d'heures: 3 h.

Co-Animation d'un atelier avec Lucence Ing consacré à l'alignement et à la collation automatisée, dans le cadre des journées du cluster 7 de Biblissima+ en décembre 2023. Travail à l'aide de notebooks python. Lien vers le matériel : Alignement, Collation.

INTRODUCTION L'ÉDITION ÉLECTRONIQUE

[23]

Institution: Universidad Complutense de Madrid

Niveau: M2

Années: 2023–2024

Nombre total d'heures: 2 h.

Animation en novembre 2023 d'une séance de séminaire Master (Universidad Complutense de Madrid), consacré à une présentation générale de l'édition électronique, de l'image au texte édité.

LA TRANSCRIPTION AUTOMATISÉE : INTRODUCTION AVANCÉE

[24]

Institution: UNIL

Niveau: Master

Années: 2023–2024

Nombre total d'heures: 3 h.

Animation en octobre 2023 d'un séminaire de Master (Université de Lausanne), consacré à la transcription automatisée.

CIVILISATION MÉDIÉVALE HISPANIQUE

[25]

Institution: ENS de Lyon

Niveau: Agrégation

Années: 2022–2023

Nombre total d'heures: 9 h.

Cours d'introduction historique au siècle d'Alphonse X, dans le cadre de la préparation à l'agrégation externe d'espagnol. 3 séances au début du premier semestre 2022–2023.

OUTILS DE LA RECHERCHE

[26]

Institution: ENS de Lyon

Niveau: M1

Années: 2022–2023; 2019–2020; 2018–2019

Nombre total d'heures: 90 h.

Cours de méthodologie de la recherche. Intitulé du cours : La publication scientifique : enjeux économiques, juridiques, éditoriaux et techniques. Le cours a alterné entre présentation de la publication scientifique sous ses différents prismes (juridique, scientifique, économique) et initiation à la production de documents scientifiques avec le logiciel/langage de composition \LaTeX : structuration de document, mise en page, bibliographie, diapositives.

CIVILISATION MÉDIÉVALE HISPANIQUE

[27]

Institution: ENS de Lyon

Niveau: M1

Années: 2022–2023

Nombre total d'heures: 30 h.

Le sous-titre du cours est : *Droit, pouvoir, légitimité politique, d'Alphonse X aux Trastamares.*

SÉMINAIRE DE TRADUCTION

[28]

Institution: ENS de Lyon

Niveau: M2

Années: 2022–2023

Nombre total d'heures: 30 h.

Le thème du cours est la réécriture. Il s'est agi de travailler à la traduction de textes ayant été repris et réécrit au cours des siècles, du Moyen Âge au 20^e.

VERSION CLASSIQUE

[29]

Institution: ENS de Lyon

Niveau: L3–M1

Années: 2022–2023; 2018–2019

Nombre total d'heures: 60 h.

Cours de version classique.

VERSION MODERNE

[30]

Institution: ENS de Lyon

Niveau: L3–M1

Années: 2022–2023

Nombre total d'heures: 30 h.

Cours de version moderne.

HUMANITÉS NUMÉRIQUES APPLIQUÉES AU MONDE HISPANIQUE

[31]

Institution: ENS de Lyon

Niveau: L3–M1

Années: 2022–2023

Nombre total d'heures: 30 h.

Le sous-titre du cours est *La Brevísima relación de la destrucción de las Indias* : de l'image au texte annoté. Le cours a consisté en une introduction aux différentes méthodes de traitement et d'étude des sources textuelles anciennes, en utilisant une même source littéraire comme fil conducteur, la *Brevísima relación de la destrucción de las Indias* (1552) de Bartolomé de las Casas : courte introduction à la bibliographie matérielle, OCR et travail sur l'apprentissage supervisé, segmentation, annotation lexico-grammaticale du castillan moderne.

INITIATION À L'HTR ET À ESCRIPTORIUM

[32]

Institution: UCM

Niveau: Master et doctorat

Années: 2021-2022

Nombre total d'heures: 4 h.

Animation en mai 2022 d'un atelier doctoral à l'Universidad Complutense de Madrid (dans le cadre du séminaire « *Herramientas digitales* ») sur le HTR et eScriptorium.

DIGITAL SCHOLARLY EDITIONS : MANUSCRIPTS, TEXTS AND TEI

[33]

ENCODING

Institution: Université Lyon II

Niveau: Master et plus

Années: 2021-2022; 2020-2021; 2019-2020

Nombre total d'heures: 18 heures

Animateur secondaire (2019-2021) et principal (2022) du MOOC *Digital Scholarly Editions : Manuscripts, Texts and TEI Encoding* organisé par Marjorie Burghart dans le cadre du Master Mondes Médiévaux de l'Université Lyon II, aux printemps 2020, 2021, 2022.

TRANSCRIPTION AUTOMATISÉE DE SOURCES MANUSCRITS AVEC ESCRIPTORIUM

[34]

Institution: Casa de Velázquez

Niveau: Master et Doctorat

Années: 2021-2022

Nombre total d'heures: 16 h.

Co-organisation et co-animation de la formation à la transcription automatisée de sources manuscrites avec le logiciel eScriptorium, avec María Díez Yáñez et Irene Salvo García, organisée à la Casa de Velázquez en septembre 2021 (vingt participant.es présent.es, trente en distanciel), avec la collaboration et la participation de Peter Stokes, Benjamin Kiessling, Leonor Zozaya-Montes et de Belén Almeida. Le support de la formation était la *General Estoria*. Lien vers la page de l'événement.

DOCUMENTATION D'ÉDITIONS TEI AVEC ODD

[35]

Institution: CIHAM

Niveau:

Années: 2020-2021

Nombre total d'heures: 12 h.

Coorganisation et animation de la formation en ligne à la documentation TEI-ODD, avec Emmanuelle Morlock, Sarah Orsini et Ariane Pinche (mai-juin 2021, trois demi-journées). Lien vers la page de l'événement

LEMMATISER DES TEXTES ET CORRIGER L'ANNOTATION GRÂCE À L'APPRENTISSAGE PROFOND AVEC PYRRHA

[36]

Institution:

Niveau: Mixte

Années: 2020-2021

Nombre total d'heures: 4 h.

Atelier coanimé avec Thibault Clérice, Lucence Ing, Ariane Pinche, Simon Gabay et Jean-Baptiste Camps en mai 2021 lors du congrès Humanistica. Lien vers la page de l'atelier et Abstract.

INTRODUCTION À L'ANNOTATION LEXICO-GRAMMATICALE

[37]

Institution: ENSSIB

Niveau: Master

Années: 2020-2021

Nombre total d'heures: 4 h.

Animation d'une séance d'introduction à l'analyse grammaticale et à l'utilisation de l'outil Pyrrha, dans le cadre du master HN de l'ENSSIB, en octobre 2020.

CIVILISATION MÉDIÉVALE HISPANIQUE

[38]

Institution: ENS de Lyon

Niveau: L3-M1

Années: 2019-2020

Nombre total d'heures: 15 h.

Cours d'introduction à l'histoire culturelle de la Castille sous les Trastamare. Thème : le pouvoir et sa légitimation.

HUMANITÉS NUMÉRIQUES

[39]

Institution: ENS de Lyon

Niveau: L3-M1

Années: 2019-2020

Nombre total d'heures: 15 h.

Cours d'introduction aux humanités numériques : l'annotation lexico-grammaticale du castillan médiéval.

FORMATION À XSLT

[40]

Institution: CIHAM

Niveau: Master et doctorat

Années: 2018–2019

Nombre total d'heures: 16 h.

Organisation et animation de la formation XSLT COSME 2019 en collaboration avec trois autres doctorant.es/docteurs : Simon Gabay, Université de Lausanne, Ariane Pinche, École nationale des Chartes, Jean-Paul Rehr, Université Lyon II. Lien vers le dépôt de la formation.

COLLES DE PRESSE

[41]

Institution: Lycées La Martinière-Duchère et Juliette Récamier

Niveau: Première année de CPGE

Années: 2017–2018

Nombre total d'heures: 40 h (non décomptées)

Colleur dans deux lycées lyonnais. Sujets de presse, environ 40 h.

Production scientifique

Thèse de doctorat

Titre : *Le Regimiento de los príncipes et sa glose : étude et édition numérique de la partie sur le gouvernement de la cité en temps de guerre (III, 3).*

Direction : Carlos HEUSCH (École Normale Supérieure de Lyon) et Jesús R. VELASCO (Yale University).

Date de soutenance : 8 septembre 2023.

Rapportrices : Elena PIERAZZO (Université de Tours) et Susana ALLÉS TORRENT (University of Miami).

Membres du jury : Jean-Baptiste CAMPS (École nationale des chartes), Elena PIERAZZO (Université de Tours), Hélène THIEULIN-PARDO (Sorbonne Université), Marta LÓPEZ IZQUIERDO (Université Paris VIII).

Lien vers la thèse : tel-04337406.

Résumé : Ma thèse de doctorat a consisté en l'étude et en l'édition numérique de la dernière partie du *Regimiento de los príncipes* (III-3), une traduction castillane glosée du *De Regimine Principum* de Gilles de Rome. J'y décris selon les principes de la bibliographie matérielle et de la sociologie des textes tous les témoins du corpus, composé de neuf manuscrits et d'un exemplaire incunable. La thèse développe une chaîne de traitement semi-automatisée qui permet l'édition, depuis l'acquisition individuelle de chacun des témoins du corpus, par HTR, OCR ou transcription manuelle, jusqu'à la production de l'édition au format pdf, en passant par la structuration XML/TEI des données. Une place importante est accordée à la modélisation de la chaîne de traitement développée, à sa description en termes de flux d'information, et à la présentation détaillée des données produites (transcription, segmentation, lemmatisation). Je me suis en particulier intéressé à la question de la pérennité de la production ecdotique en contexte numérique. Lui-même structuré en XML-TEI, le mémoire de thèse a exploré de nouvelles voies de consultation et de mise en relation entre le texte critique et son commentaire (extraction d'information, citations avec appareil critique). La question de la comparaison semi-automatisée de sources livresques préalablement décrites et encodées au format XML-TEI y est essentielle, ouvrant la voie à la collation d'information non textuelle (voir la production 43). Quant au corpus, il a été étudié sous l'angle de la variation, en prenant comme axe le discours politique, afin de déterminer son homogénéité conceptuelle, pour ensuite étudier les relations entre témoins. La thèse a permis de mettre à jour les connaissances sur les différentes versions du texte et d'identifier la première traduction castillane du *De Regimine Principum*, jusqu'alors considérée comme perdue. L'édition a été produite sur les 23 derniers chapitres du texte. Elle aligne, au paragraphe, la traduction originelle, dite ω , avec la version β qui est le texte édité à proprement parler, et avec le texte latin, récupéré à partir du projet PALM, puis structuré en XML-TEI par mes soins. Cette thèse a reçu les félicitations unanimes du jury à l'issue de la soutenance.

Articles

- accepté [1] Alexandre Wauthier, Matthias Gille Levenson, Daphné Mathelier, Marie Flesch, « Programming Historian en français : les défis de la fabrication d'une revue francophone sur les humanités numériques », revue *Données, Expériences, Méthodes, Codes* (DEMC).
- [2] « Towards a general open dataset and models for late medieval Castilian text recognition (HTR/OCR). », *Journal of Data Mining and Digital Humanities (JDMDH), Special Issue : Historical documents and automatic text recognition*, 2023, éd. Ariane Pinche et Peter Stokes, doi :10.5281/zenodo.7387376.
- [3] « Educación del príncipe, nobleza y caballería en el *De regimine principum* castellano (segundo y tercer libro) », *Librosdelacorte* 22, 2021, pp. 285-308. doi :10.15366/ldc2021.13.22

Articles de conférence

- soumis [4] Lucence Ing, Matthias Gille Levenson, Carolina Macedo, « Phrase-Level Segmentation on Medieval Corpora for Aligning Multilingual Texts », soumis à LREC2026
- [5] Matthias Gille Levenson, Lucence Ing, Jean-Baptiste Camps, « Textual Transmission without Borders : Multiple Multilingual Alignment and Stemmatology of the *Lancelot en Prose* (Medieval French, Castilian, Italian) », Proceedings of the Computational Humanities Research Conference 2024 (CHR 2024), 2024, Aarhus, Denmark, éd. Wouter Haverals, Marijn Koolen et Laure Thompson. <https://ceur-ws.org/Vol-3834/paper104.pdf>. **Nominé parmi les trois meilleurs long papers de la conférence**
- [6] Matthias Gille Levenson, Célian Ringwald, Marie Flesch, Jennifer Isasi, Sofia Papastamkou, Riva Quiroga et David Valentine, « Connecter les chapitres linguistiques de *Programming Historian*? Vers une table conceptuelle multilingue constituée semi-automatiquement », Humanistica2024, éd. Simon Gabay. Dépôt hal.
- [7] Clérice, Thibault, Ariane Pinche, Malamatenia Vlachou-Efstathiou, Alix Chagué, Jean-Baptiste Camps, Matthias Gille Levenson, Olivier Brisville-Fertin, et al., « CATMuS Medieval : A Multilingual Large-Scale Cross-Century Dataset in Latin Script for Handwritten Text Recognition and Beyond. », In : Barney Smith, E.H., Liwicki, M., Peng, L. (eds) Document Analysis and Recognition – ICDAR 2024. *Lecture Notes in Computer Science*, vol 14806. Springer, Cham, 2024, doi :10.1007/978-3-031-70543-4_11.
- [8] Simon Gabay, Thibault Clérice, Jean-Baptiste Camps, Jean-Baptiste Tanguy, Matthias Gille-Levenson. *Standardizing Linguistic Data : Method and Tools for Annotating (Pre-Orthographic) French.*, Proceedings of the 2nd International Digital Tools and Uses Congress (DTUC '20), Oct 2020, Hammamet, Tunisia, hal-03018381. doi :10.1145/3423603.3423996

Chapitres d'ouvrages collectifs

- [9] Simon Gabay, Matthias Gille Levenson, Olivier Brisville-Fertin, Malamatenia Vlachou-Efstathiou, « Le projet CATMuS : les langues d'Europe occidentale en alphabet latin » in *Apprendre à lire aux machines*, éd. Thibault Clérice et Alix Chagué, Presses de l'Université de Montréal, 2025 (sous presse).
- [10] « L'évolution du Regimiento de los príncipes (1345-1494), conditionnée par le pouvoir politique? » in *Écritures du Pouvoir*, 2, éd. Véronique Lamazou-Duplan, Ausonius éditions, Scripta Medievalia, 2019, pp. 137-148, hal-02369116.

Communications dans des congrès internationaux avec comité de sélection

- [11] Lucence Ing, Matthias Gille Levenson, Carolina Macedo, Jean-Baptiste Camps, « Premiers jalons de la collation multilingue du *De Regimine Principum* latin et vernaculaire », 31e Congrès International de Linguistique et de Philologie Romane, 30 juin – 5 juillet 2024, Lecce (Italie).
- [12] Jennifer Isasi, Matthias Gille Levenson, Célian Ringwald, Marie Flesch, Sofia Papastamkou, Riva Quiroga and David Valentine, « *Enhancing Programming Historian Multilingual Reach. Creating a Semi-Automated Multilingual DH Vocabulary Table* », ACH2024, USA, 2024.

- [13] Ariane Pinche, Thibault Clérice, Alix Chagué, Jean-Baptiste Camps, Malamatenia Vlachou-Efstathiou, Matthias Gille Levenson, Olivier Brisville-Fertin, Federico Boschetti, Franz Fischer, Michael Gervers, et al., « *CATMuS-Medieval : Consistent Approaches to Transcribing Manuscripts : A Generalized Set of Guidelines and Models for Latin Scripts from Middle Ages (8th-16th century)* ». DH2024, Washington DC, USA, 2024. hal-04346939
- [14] « *TEICollator : a TEI to TEI workflow* », TEI2022 : Text as data, Newcastle (Angleterre) 12-16 septembre 2022.
- [15] « *TEICollator : une chaîne de traitement ecdotique semi-automatisée* », 30e Congrès International de Linguistique et de Philologie Romane, Université de La Lagune, Canaries (Espagne), 4-9 juillet 2022. Dépôt hal.
- [16] Matthias Gille Levenson, Gwenaëlle Patat (ordre alph.). « *Programming Historian en français : Faire communauté pour le partage de ressources éducatives libres sur les méthodes numériques en sciences humaines et sociales francophones* », Colloque Humanistica 2022, Mai 2022, Montréal, Canada. hal-03672420
- [17] Matthias Gille Levenson, Olivier Brisville-Fertin, María Díez Yáñez, Simon Gabay, « *Construcción de un corpus de evaluación de la anotación léxico-gramatical del castellano medieval (siglos XIII-XV)* », HDH2021 : scire vias, Saint Jacques de Compostelle, Espagne, oct. 2021. Lien vers le dépôt git.
- [18] Simon Gabay, Lucie Rondeau Du Noyer, Matthias Gille Levenson, Ljudmila Petkovic et Alexandre Bartz, « *Quantifying the Unknown : How many manuscripts of the marquise de Sévigné still exist ?* », DH2020, Ottawa, Canada. Abstract et diapositives.
- [19] Paul Bertrand, Ariane Pinche, Matthias Gille Levenson et Margot Ferrand, « *COSME - Complexities : 30 Years Of Research Of Medievalists DH Concerning A Thousand Years Of Medieval Sources* » (poster). Congrès « *Digital Humanities 2019* » (DH2019), du 9 au 12 juillet 2019, Utrecht, Pays-Bas. DOI [10.5281/zenodo.3275597](https://doi.org/10.5281/zenodo.3275597)

Communications dans des colloques ou des journées d'étude

- sur invitation [20] Lucence Ing, Matthias Gille Levenson, « *Modelizing textual variation in multilingual context* » (poster). Séminaire « *Seeing the difference* », du 14 au 17 avril 2025, Lorentz Center, Leiden, Pays-Bas, <https://hal.science/hal-05173463>
- [21] Lucence Ing, Alice Leflaëc, Laurence Mellerin, Matthias Gille Levenson, et Sébastien Hamel, (2024). « *Cluster 7 Biblissima+ : projets en cours* » (poster), Journées Biblissima+. Lien hal : hal-04753020.
- sur invitation [22] « *De la donnée avant toute chose. Retour d'expérience de l'utilisation de l'HTR dans des projets d'édition et d'étude des textes médiévaux* », Journées d'étude De la transcription manuelle participative des textes à la reconnaissance automatique de texte (ATR/HTR) : outils, théorie, pratiques , Nancy, 5 avril 2024.
- [23] « *Vers la collation multilingue de textes romans médiévaux. Le Regimiento de los príncipes comme cas d'étude* », Journées du Séminaire d'Études Hispaniques Médiévales : Humanités Numériques et philologie romane, 5 avril 2024.
- sur invitation [24] « *Primeras pistas de estudio computacional de la recepción del Regimiento de los Príncipes : las marcas de lectura del manuscrito K.I.5 (Biblioteca del Escorial)* », Journées d'étude Cartographies du livre ibérique, CEIBA, Toulouse, 15 mars 2024. HAL.

- [25] « Chevaliers et cavaliers dans la dernière partie du *Regimiento de los príncipes* », Séminaire d'Études hispaniques médiévales 2022, co-organisation Julie Marquer et Matthias Gille Levenson, ENS de Lyon, 13 décembre 2022.
- sur invitation [26] *Impressa manent* : en defensa de la producción de ediciones digitales en formato fijo, Congrès Ecdótica y humanidades digitales : historiografía, metodología y praxis, Salamanca (Espagne) 23-24 novembre 2023.
- sur invitation [27] « La transcription automatisée, un changement de paradigme ecdotique ? », colloque Trilobe II, 29-30 septembre 2022, ENS Lyon.
- [28] Matthias Gille Levenson, Sofia Papastamkou et Célian Ringwald (ordre alph.). « *Programming Historian* : un lieu de collaborations et d'interactions multiples », Colloque DH-Nord 2022, juin 2022, en ligne.
- [29] « Réception du roman, fiction et valeur du vrai dans trois traités de littérature politique des XIV^e et XV^e siècles en castillan », Journée d'études « Vérité et mensonge », ENS de Lyon, 6 mai 2021.
- [30] « Describir y prescribir el liderazgo en la Castilla bajomedieval », Séminaire du Madrid Institute of Advanced Studies (MIAS), co-animé avec Edward Holt, 22 mars 2021.
- sur invitation [31] « "E mayormente los reyes". Educación del hijo y educación del príncipe en el segundo libro del *De regimine principum* castellano », I SEMINARIO DE CORTE Y LITERATURA, Instituto Universitario La Corte en España, UAM, 26-27 novembre 2021.
- [32] « *Editing a 15th century political treatise using the computer : a back-and-forth between meaning and information* », *Iberian Connections seminar*, Yale University, 12 novembre 2019. Lien.
- [33] « Du concept au récit : théorie politique et enxiemplo dans le *Regimiento de los príncipes* (III, 3) ». Colloque « Theorica 7 : Politia », ENS de Lyon, Lyon, 6 juin 2019.
- [34] « La variance et ses enjeux dans l'édition numérique du *Regimiento de los príncipes* ». Journée d'étude Humanités Numériques, Agorantic-Ciham, Avignon, 27 mai 2019.
- sur invitation [35] « Construire l'édition numérique d'une tradition textuelle complexe : propositions pour le *Regimiento de los príncipes* ». Atelier IRPALL « La fabrique du texte », Toulouse, 22 février 2019.

Rapports, documents non évalués

- [36] Matthias Gille Levenson, Olivier Brisville et al., « Manuel d'annotation lexicogrammaticale du castillan médiéval », 2025, hal-04917864
- [37] Alix Chagué, Floriane Chiffolleau, Matthias Gille Levenson, Hugo Scheithauer et Ariane Pinche, « Chaînes d'acquisition, de traitement et de publication du texte », Consortium Ariane - Axe 1. 2024. hal-04734959

Jeux de données

[38] Thibault Clérice et al. Jeu de données d'ATR diachronique et multilingue. CATMuS : reconnaissance de l'écriture manuscrite et imprimée. Mars 2025 (version 1.6). Huggingface. *Résumé*: Ce jeu de données de standard *gold*, produit dans le cadre des financements HTRogène (Biblissima+) et HTRomance (Datalab BNF) portés par Thibault Clérice, contient la transcriptions de fragments de plus de 200 manuscrits médiévaux du 7^e au 15^e siècles, dans plus d'une demi-douzaine de langues (ancien et moyen français, italien, vénitien, latin, occitan, anglais, castillan, catalan, navarrais, flamand), représentant environ 200.000 lignes intégralement corrigées. Outre l'apport au corpus du corpus issu de ma thèse (voir item suivant), je suis co-responsable avec Olivier Brisville-Fertin du versant ibérique du projet, de la constitution du corpus et du contrôle de la qualité des données produites en castillan et catalan.

[39] Jeu de données et modèles évalués – reconnaissance de l'écriture manuscrite et imprimée du castillan médiéval. Novembre 2022. 10.5281/zenodo.7386489

Résumé: Ce jeu de données de standard *gold* contient environ 360 pages segmentées (zones et lignes) et transcrites selon des normes allographétique et graphématiques, ainsi qu'un ensemble de modèles de transcription. Les données sont celles de mon corpus de thèse. Ce jeu de données a été évalué dans l'article 2.

[40] Jeu de données non évalué – segmentation du castillan médiéval (14^e–15^e siècle), mai 2022. 10.5281/zenodo.7684879.

Résumé: Ce jeu de données de standard *gold* contient environ 45.000 lignes transcrites selon des normes allographétiques et segmentées (c'est-à-dire, les mots étant identifiés comme des tokens individuels, selon les normes actuelles). Les données sont celles de mon corpus de thèse.

[41] Jeu de données non évalué – annotations du castillan médiéval (14^e–15^e siècle), juin 2023. 10.5281/zenodo.7772058.

Résumé: Ce jeu de données de standard *gold* contient environ 280.000 tokens annotés en lemmes, parties du discours et morphologie. Les données sont celles de mon corpus de thèse.

Logiciels développés

L'ensemble de ma production informatique est disponible en licence ouverte sur Github.

[42] **Aquilign** : <https://github.com/ProMeText/Aquilign>

Aquilign est un outil co-développé avec Lucence Ing et Jean-Baptiste Camps, qui permet l'alignement automatisé de sources textuelles multilingues, adapté aux états de langue médiévaux. Fondé sur des modèles de langue, il permet de dépasser l'alignement formel et permet la réalisation de l'alignement sémantique, ouvrant la voie à la comparaison des différentes traductions d'un même texte. Utilisant l'algorithme d'alignement Bertalign, il s'appuie sur un segmenteur spécifiquement conçu, permettant la création de segments textuels correspondant à des unités sémantiques. Aquilign permet la production de tables d'alignement et inclut une fonctionnalité d'alignement de sources structurées en XML-TEI.

[43] **TeiCollator** : <https://github.com/matgille/tei-collator>

TeiCollator est un outil développé en python et un ensemble de méthodes de collation semi-automatisée TEI to TEI qui s'appuie sur CollateX pour l'alignement des témoins. Il permet en particulier la collation et la comparaison d'éléments non textuels, et offre la possibilité d'obtenir en sortie autant de témoins-base que de témoins disponibles, dans une optique essentiellement bédieriste.

[44] **ClaViTranscr** : https://github.com/matgille/transcription_virtual_keyboard

ClaViTranscr est un clavier virtuel développé en python et qui permet d'aider au travail d'établissement de vérités terrains dans le cadre de l'utilisation de logiciels de transcription automatisée comme eScriptorium. Il autorise en particulier la recherche de caractères et un ajout facilité de nouveaux caractères à la base de données.

[45] **Boudams Fork** : https://github.com/matgille/boudams_like_tokenizer

Boudams Fork est une adaptation du travail et du logiciel Boudams de Thibault Clérice. Il s'agit d'un tokéniseur fondé sur des techniques d'apprentissage supervisé (réseaux neuronaux convolutifs). Cette adaptation permet l'utilisation de l'outil avec des sources XML-TEI (en particulier l'identification des mots coupés à la ligne), ainsi que la conservation de l'information de segmentation originale à l'aide d'entités XML, par ajout de classes de sortie à l'architecture originelle conçue par Thibault Clérice.

[46] **Pyrrha** : <https://github.com/hipster-philology/pyrrha>

Pyrrha est une interface de post-correction d'annotations lexico-grammaticales imaginée et développée principalement par Thibault Clérice, à laquelle je contribue par des corrections de bugs, ajouts et suggestions de nouvelles fonctionnalités.

Développement web

en cours [47] **CATMuS Guidelines** : <https://catmus-guidelines.github.io/>

Le site *CATMuS Guidelines* vise à faciliter l'utilisation et la consultation des normes édictées par le groupe CATMuS. Il consiste en un site web statique hébergé sur Github et généré à partir de fiches au format markdown, qui comprend notamment une fonctionnalité d'indexation et de recherche.

Projets financés en cours

Co-Porteur du projet e-Cam : (2024)

Co-porteur avec Olivier Brisville-Fertin du projet « e-Cam » (Étiquetage lexico-grammatical du Castillan Médiéval), qui vise à la constitution de normes standard d'annotation du castillan médiéval (lemmes, parties du discours, morphologie), ainsi qu'à la création d'un corpus représentatif du castillan médiéval, annoté et intégralement corrigé à la main, dans l'optique de l'entraînement d'un modèle d'annotation plus efficace que les outils existants. Organisme financeur : Agorantic (8000 euros obtenus).

Responsable de la section ibérique des projets HTRomance et HTRogène : (2023–2025)

Membre des projets « HTRomance » (financement DataLab BNF : 8000 euros) et « HTRogène » (financement Biblissima+ : 20.000 euros) dont les porteurs principaux sont Thibault Clérice et Alix Chagué, consacrés à la production de données d'entraînement d'ATR (transcription automatique de l'écriture manuscrite et imprimée). Co-responsable avec Olivier Brisville-Fertin du versant ibérique du projet (constitution de jeux de données en castillan et catalan par supervision de vacations de transcription).

Intégration dans des comités éditoriaux et de lecture

2023– Membre du comité éditorial des *Cahiers d'Études Hispaniques Médiévales*.

2021– Membre du comité de rédaction de *Programming historian en français* (rédacteur en chef adjoint).

Activités d'encadrement et participation à des jurys

- 2023 Jury de soutenance du mémoire de master 1 de Kelian DESNOUES. Titre du mémoire : *Un pasado continuo : estudio lingüístico del idiolecto de Cristóbal Colón*, mémoire réalisé sous la direction d'Olivier BRISVILLE-FERTIN.

Encadrement de stages

- 2024 Co-encadrant du stage de M1 d'Amandine FRÉCHOU à l'ENS de Lyon avec Olivier BRISVILLE-FERTIN. Thèmes de travail : HTR, annotations grammaticales, études lexicales.

Encadrement de vacances

- 2024-2025 Encadrement de trois étudiant.es pour des vacances de production de corpus d'HTR
2024 Encadrement de deux étudiant.es pour des vacances de production de corpus d'annotations linguistiques

Compétences

Langues pratiquées

- **Français, Espagnol** : bilingue.
- **Anglais** : compétence professionnelle et langue de travail, de communication et de publication fréquente.

Compétences informatiques

- **Maîtrise des méthodes et outils d'édition électronique** selon le standard XML TEI/P5 : XML, ODD, XSLT, librairie python-lxml, gestion des espaces de nommage. Compréhension et utilisation du XQuery ; expérience ponctuelle dans la construction de bases de données XML. Utilisation courante des expressions régulières.
- **Compréhension des fondamentaux du web de données et du web sémantique**. Badge FUN du cours « Web Sémantique et Web de données » d'INRIA.
- **Utilisation courante du langage Python** et maîtrise des **fondamentaux de la Programmation Orientée Objet**. Badge FUN du cours « Python : des fondamentaux aux concepts avancés du langage » de l'Université Côte d'Azur et INRIA.
- **Utilisation courante des méthodes de publication** de sources documentaires et/ou littéraires au format fixe (L^AT_EX, reledmac, reledpar, beamer).
- **Maîtrise des outils de programmation lettrée** comme les notebooks jupyter (en local, via Binder ou Colab).
- **Expérience des processus et outils de développement web** (HTML, CSS, Javascript). Expérience ponctuelle avec le framework de développement web flask.
- **Compétence dans l'utilisation des processus de publication statique** (Jekyll, jinja) éventuellement adossés à Github (Github Actions et Github Pages).
- **Maîtrise et utilisation du logiciel de contrôle de version git** et des plateformes github et gitlab.
- **Utilisation courante des techniques et technologies d'apprentissage supervisé** pour le traitement de l'image et du texte (reconnaissance de caractères et d'écriture manuscrite : Ocropy, Kraken, eScriptorium ; analyse grammaticale : freeling, Pie ; segmentation/tokénisation : Boudams ; modèles de langue : Bert). Expérience dans la création d'architectures neuronales (Aquilign et le fork de Boudams, voir plus haut).
- **Maîtrise des systèmes d'exploitation basés sur linux** (Ubuntu, Debian, Raspbian). Utilisation courante de bash.

Responsabilités organisationnelles

Événements scientifiques

- Participation à l'organisation d'une journée d'études du SEHM (Séminaire d'Études Hispaniques Médiévales) du CIHAM, suivie d'une formation à l'HTR : « Humanités Numériques et philologie romane », les 4 et 5 avril 2024.
- Organisation d'une séance du SEHM (Séminaire d'Études Hispaniques Médiévales) du CIHAM (coord. Julie Marquer et Carlos Heusch) : « Actualités de la recherche », le 13 décembre 2022.
- Participation à l'organisation des Journées d'Étude Doctorales 2021 du CIHAM.
- Organisation des Journées des Jeunes Hispanistes 2019 (ENS de Lyon).

Formations académiques

J'ai eu à cœur de participer et d'organiser au moins une formation ponctuelle chaque année depuis le début de ma thèse. Les formations à l'organisation desquelles j'ai contribué sont les suivantes :

- *ScriptMSS*, Lyon, ENS de Lyon, 2025
- *EnExDi*, Lyon, ENS de Lyon, 2024, 2025 et 2026
- *Du Calame au Clavier*, Grenoble, Université Grenoble Alpes, 2024
- *Formation à eScriptorium*, Madrid, CVZ, 2021
- *ODD et documentation des éditions numériques*, CIHAM, 2021
- *Transformation des sources XML avec XSLT*, CIHAM, 2019