

Lire, reconnaître, éditer : L'intelligence artificielle face aux archives romanes

Colloque international (30-31 mars 2027)

Montpellier Salle cinéma et rencontres de la Médiathèque Emile Zola

Comité d'organisation :

Archives Montpellier ville et métropole : Christine Feuillas, Claire Garcia, Pierre-Joan Bernard

Université Montpellier Paul-Valéry : Gilda Caiti-Russo, Hervé Lieutard, Camilla Talfani, Silvan Chabaud, Yan Lespoux

Comité scientifique :

Geneviève Dumas (Université de Sherbrooke)

Daniel Le Blévec (Université Paul Valéry)

Philippe Martel (Université Paul Valéry)

Florence Clavaud (Archives Nationales)

Thibault Clérice (INRIA-Institut national de recherches en sciences et technologies du numérique)

Ariane Pinche (CNRS – CIHAM UMR 5648, Histoire, Archéologie, Littératures des mondes chrétiens et musulmans médiévaux, Université Jean Moulin Lyon 3)

Federico Boschetti (Istituto di linguistica computazionale Zampolli Pise, CNR italien)

Francesca Frontini (Istituto di linguistica computazionale Zampolli Pise, CNR italien)

*

Dans quelques mois le registre occitan des comptes de la claverie du consulat de Montpellier de 1357¹ sera édité à l'aide de l'intelligence artificielle : cet ambitieux projet², soutenu par MIRANDA³, sera l'occasion de faire le point sur l'apport que l'IA peut d'ores et déjà fournir

¹ Conservé dans le fonds ancien des Archives de la ville sous la cote [CC138](#).

² [Projet DOP : Digital Occitan paleography](#).

³ [Montpellier Institute for research-creation on art, culture and heritage in a new digital age](#).



pour la connaissance des archives vernaculaires du Moyen âge roman.

Longtemps demeurées dans l'ombre des manuscrits latins, les archives rédigées en langues romanes constituent aujourd'hui l'un des principaux lieux de la recherche sur les sociétés et les langues médiévales. Registres urbains, comptabilités, actes notariés, cartulaires, livres de raison, délibérations consulaires, procédures judiciaires, enquêtes administratives ou correspondances offrent un accès privilégié à l'histoire des langues, des pratiques de l'écrit et des institutions entre le XII^e et le XV^e siècle.

Si les vastes campagnes de numérisation menées depuis une vingtaine d'années ont considérablement facilité l'accès à ces sources, les progrès récents de la paléographie assistée par l'intelligence artificielle — notamment grâce à la reconnaissance automatique de l'écriture manuscrite (*Handwritten Text Recognition*, HTR) — ouvrent désormais un nouvel horizon pour leur édition, entendue au sens de « mise en lumière », ainsi que pour l'étude de ces textes relevant de l'écriture pragmatique.

Ces développements ne constituent pas seulement une évolution technologique ; ils invitent à repenser les méthodes mêmes de la philologie documentaire. Si les outils d'intelligence artificielle permettent aujourd'hui d'identifier automatiquement des formes graphiques et de traiter des masses documentaires jusque-là difficilement exploitables, ils interrogent également la place de l'expertise philologique face aux modèles statistiques, la qualité des données d'entraînement et les modalités de validation scientifique.

Les archives romanes médiévales constituent un terrain d'expérimentation particulièrement exigeant pour l'intelligence artificielle : variation graphique, absence de normalisation orthographique, alternance entre latin et langues vernaculaires, abondance des abréviations et diversité des traditions scripturaires imposent le développement de modèles adaptés à ces corpus à faibles ressources. L'enjeu n'est pas seulement de mieux lire les documents, mais aussi de renouveler notre connaissance de l'histoire des langues romanes à partir des archives.

Le colloque se propose de réunir archivistes, historiens, philologues et spécialistes des humanités numériques afin d'examiner les nouvelles approches offertes par ces technologies à l'étude des archives romanes médiévales. Les communications s'articuleront autour de trois axes et pourront porter sur les archives rédigées dans les principales langues romanes médiévales : occitan, ancien français, catalan, galicien-portugais, italien ancien, sarde, francoprovençal, ainsi que sur d'autres variétés romanes documentées.

1. Lire

- Quelle place pour l'expertise paléographique traditionnelle à l'ère de l'intelligence artificielle ?



- Retours d'expérience sur l'utilisation de la HTR (*Handwritten Text Recognition*).
- Segmentation des documents, détection des abréviations, identification des mains et des scribes.

2. Reconnaître

- Contrôle, évaluation et validation des transcriptions produites par l'IA.
- Traçabilité et reproductibilité des traitements.
- Constitution et entraînement de modèles de reconnaissance adaptés aux écritures romanes médiévales.

3. Éditer

- Enrichissement des éditions numériques en TEI : annotation diplomatique, philologique et linguistique.
- Gestion, stockage, hébergement et partage des données sources.
- Application des principes FAIR (Findable, Accessible, Interoperable, Reusable) aux corpus d'archives médiévales.

Une attention particulière sera accordée aux corpus documentaires vernaculaires encore peu étudiés, récemment numérisés ou faisant l'objet de projets innovants de transcription, d'édition ou d'analyse assistées par l'intelligence artificielle.

Nous vous prions de bien vouloir envoyer vos propositions de communication pour le **30 novembre 2026** à :

PierreJoan.BERNARD@montpellier.fr

gilda.russo@umpv.fr

camilla.tani@gmail.com

Retour sur les propositions prévu pour le 15 janvier 2027.

