



HAL
open science

Projet Fabliaux

Jules Nuguet, Corinne Pierreville, Ariane Pinche, Alexei Lavrentiev

► **To cite this version:**

Jules Nuguet, Corinne Pierreville, Ariane Pinche, Alexei Lavrentiev. Projet Fabliaux : De la reconnaissance d'écriture à la textométrie : la (re-)définition du genre. *Humanistica* 2023, Association francophone des humanités numériques, Jun 2023, Genève, Suisse. hal-04099492

HAL Id: hal-04099492

<https://hal.science/hal-04099492>

Submitted on 30 Jun 2023

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

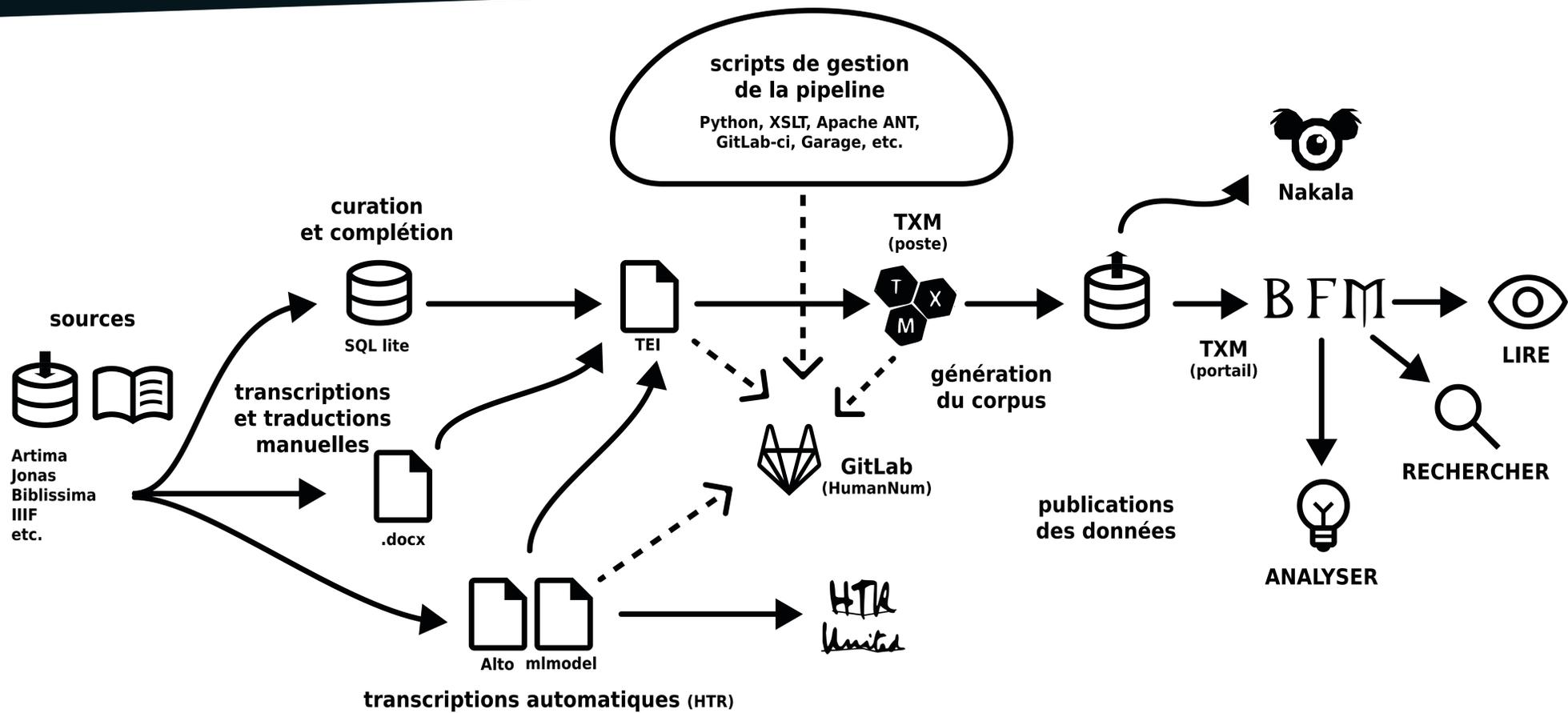
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Projet Fabliaux

De la reconnaissance d'écriture à la textométrie



■ Fabliaux ?

Un fabliau est un genre littéraire médiéval caractérisé par une courte histoire humoristique, souvent grivoise, dont l'intrigue s'inscrit dans la vie quotidienne. Le plus souvent, les fabliaux mettent en scène des personnages issus du peuple ou de la petite bourgeoisie au sein d'intrigues jouant sur les ressorts de la ruse, l'adultère, la cupidité et la tromperie.

■ HTR ?

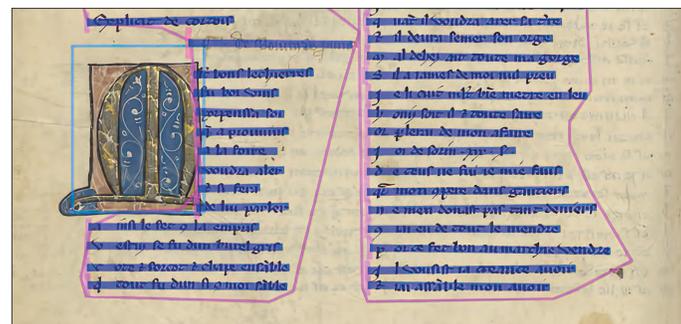
Une partie de la transcription des manuscrits est accélérée à l'aide de la reconnaissance automatique d'écriture manuscrite (HTR, Handwritten Text Recognition). La transcription d'un fabliau prend alors quelques secondes contre plusieurs heures sans cette technologie. Des procédés de conversion semi-automatique permettront de passer des sorties XML ALTO de l'HTR vers XML-TEI, puis de les intégrer au logiciel TXM. L'ensemble des prédictions et des modèles HTR sont déposés sur le GitLab du projet. Toutes les données d'entraînement produites sont référencées dans le catalogue HTR-United.

■ BFM ?

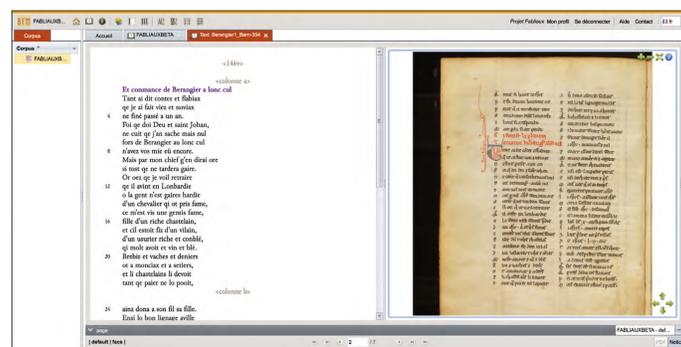
La Base de français médiéval (BFM) est une ressource en ligne destinée aux études linguistiques, littéraires et historiques. Grâce au logiciel Portail TXM, elle permet de lire et d'analyser un ensemble de corpus, dont le principal (BFM2022) contient près de 6,5 millions de mots.

■ Vers l'analyse...

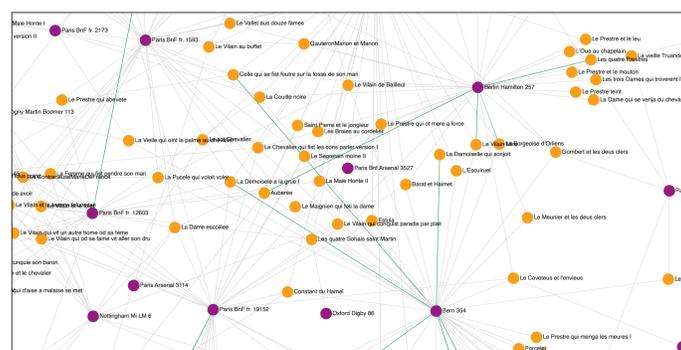
Après avoir rassemblé l'ensemble des données, le portail BFM donnera accès à un large corpus de fabliaux (23 fabliaux mi-2023, plus de 150 à terme). Il permettra de mieux comprendre ce qui fait leurs spécificités : thèmes, personnages, vocabulaire, etc. Les textes pourront être lus de manière linéaire ou interrogés par un moteur de recherche (CQP). Des outils de visualisation, comme des graphes de réseaux, permettront d'explorer le corpus à partir des informations intégrées dans les métadonnées (titre des fabliaux, date de production, manuscrits, etc.), complétant ainsi les outils de la BFM, comme les concordanciers ou les analyses factorielles des correspondances.



eScriptorium, BnF 837, fol. 66v



Interface de consultation BFM



Détail d'un graphe de réseau mettant en évidence les relations entre les manuscrits et les fabliaux

Ce travail a été réalisé grâce à une aide de Biblissima+, qui bénéficie d'une aide de l'Etat gérée par l'ANR au titre du Programme d'investissements d'avenir intégré à France 2030, portant la référence ANR-21-ESRE-0005