

Projet doctoral : Les Registres de la Comédie-Française / Questions d'intelligence artificielle

Sous la direction de Florence Naugrette à Sorbonne Université, avec le concours scientifique de Sara Harvey (University of Victoria, Canada), des équipes RCF du Massachusetts Institute of Technology (Boston, USA), de Glenn Roe et Motasem Al Rahabi (Sorbonne Université), et l'expertise extérieure de Jean-Philippe Moreux (BnF).

Le programme de registres de la Comédie-Française (RCF) est une initiative de recherche internationale (France, É-U, Canada) en humanités numériques et histoire du théâtre. Centré sur l'exploration des registres journaliers de la première institution théâtrale française depuis sa fondation en 1680 jusqu'au début du 20^e siècle, RCF entend approfondir et renouveler connaissances et méthodes en histoire du théâtre ancien à partir des nouvelles technologies. Les archives de la Comédie-Française permettent non seulement d'explorer la programmation quotidienne du théâtre, jusqu'à ses recettes, ses dépenses, ses distributions, elles ouvrent à une meilleure compréhension de l'administration d'une entreprise culturelle qui oscille entre les intérêts professionnels et économiques de la troupe et sa dépendance à l'État. Les archives permettent en outre une visibilité nouvelle en regard des professionnels qui circulent autour de la CF - auteurs et artisans, souffleurs et comptables, mais aussi en regard de l'extrême diversité du répertoire. Parce que notre recherche est fondée sur le récit et l'administration au jour le jour du théâtre, elle décentre l'histoire institutionnelle et monumentale du Français par une approche multipliant les contextes de lecture du passé. Les premières phases du programme, axées sur la création et l'exploitation de bases de données relationnelles s'inscrivent dans une approche pleinement computationnelle et quantitative en raison des premières sources exploitées : la programmation, la distribution et les finances. Or un autre corpus de registres de toute première importance reste encore à explorer : les Assemblées qui traversent le 17^e, 18^e et 19^e siècle et qui constituent une collection d'archives qualitatives. L'écart entre la source et les possibilités d'une représentation numérique dans le cas des Assemblées exige une approche critique mettant l'accent sur la nuance et les contrastes entre la construction de savoirs humanistes et scientifiques. L'Intelligence artificielle apparaît ici comme la piste la plus prometteuse et féconde pour aborder ce corpus.

Les registres de l'Assemblée constituent une source qualitative extrêmement précieuse parce qu'elle représente un exemple unique de la constitution progressive du modèle administratif de la Comédie-Française tout en mettant au jour les mécanismes, les conflits et la politique interne à la troupe. Archives de la vie administrative et témoignages des relations interpersonnelles qui régissent l'institution, les Assemblées ouvrent de nombreuses pistes pour une compréhension originale de la vie des acteurs et actrices à cette période en regard de leurs rapports au pouvoir politique, mais aussi dans la perspective d'un corps professionnel et humain. La source qui s'échelonne sur

plus de deux siècles est également digne d'intérêt sur le plan numérique en tant que manuscrit riche et nuancé qui compte une grande variété de formats et de types de contenu. Il semble que l'inconsistance et la diversité représentent un défi important et utile à l'avancée des connaissances.

Les pistes potentiels du projet doctoral :

- L'analyse de région et segmentation de documents à des fins d'extraction de données et l'étude de l'évolution de la structure et de la fonction des documents ainsi que de possibles catégories et hiérarchies ;
- La création ou l'enrichissement d'un modèle d'océrisation, la transcription et transformation automatisée en TEI/XML et la création d'outils une d' interface d'exploitation adaptée à la recherche humaniste, c'est-à-dire des interfaces ancrées fortement contextualisées et critiques ;
- L'extraction des entités nommées et la modélisation des rapports entre documents - y compris les autres ressources du programme - en vision de l'étude des réseaux de pouvoir et d'influence (par ex. : l'érosion du pouvoir des femmes au XVIIe-XIXe siècle, la nature des rapports entre la société et les auteurs, etc.).

Tout système de traitement et d'analyse des données devrait surtout viser à intégrer les logiques et épistémologies des deux domaines constitutifs des humanités numériques engagés dans ce projet (histoire du théâtre et intelligence artificielle). Si le projet doctoral doit faire avancer le domaine de l'application des techniques numériques pour l'étude du patrimoine culturel, il doit également exposer et explorer ses limites et la manière dont ils influent et façonnent - parfois de manière implicite - nos visions et interprétations de l'histoire.

Le projet doctoral vise donc l'exploitation des Assemblées dans une perspective technologique et critique. Il s'agit d'une exploration résolument transdisciplinaire et la recherche devra autant être au service des avancées en histoire du théâtre qu'en intelligence artificielle; les deux domaines doivent être mis en valeur, à travers les tensions naissant de leur unification. Il s'agit donc d'une démarche critique des humanités numériques.

Le profil du candidat recherché est orienté du côté de la technologie (profil en humanités numériques ou sciences de l'informatique). Une connaissance en histoire du théâtre, en archivistique est un atout considérable.

Bibliographie indicative

Ben Salah Mohammed, Moreux Jean-Philippe, Ragot Nicolas, Paquet Thierry, « OCR performance prediction using cross-OCR alignment », IEEE 2015.

D'ignazio, Catherine, & Lauren F. Klein. Data feminism. MIT press, 2020.

Drucker, Johanna. « Humanities approaches to graphical display. » Digital Humanities Quarterly 5.1 (2011): 1-21.

Filgueira, Rosa, et al. « Extending defoe for the Efficient Analysis of Historical Texts at Scale. » 2021 IEEE 17th International Conference on eScience (eScience). IEEE, 2021.

Harvey, Sara & Agathe Sanjuan. « Les Humanités numériques, dialogue entre le monde de la recherche et de la documentation : le projet des registres journaliers de la Comédie-Française », Bulletin des Bibliothèques de France, 2016, p. 102-109.

Moreux, Jean-Philippe, « Recherche d'image dans les bibliothèques numériques patrimoniales et expérimentation de techniques d'apprentissage profond », ASTED, vol. 65, nbr 2, avril-june 2019.

Moreux, Jean-Philippe, "Innovative Approaches of historical newspapers: Data Mining, Data visualization, Semantic Enrichment. Facilitating access for various profiles users", IFLA 2016 Satellite Meeting.

Shen, Zejiang, et al. « LayoutParser:A unified toolkit for deep learning based document image analysis. » International Conference on Document Analysis and Recognition. Springer, Cham, 2021.

Terras, Melissa. « Towards a Reading of the Vindolanda Stylus Tablets: Engineering Science and the Papyrologist. » Human IT: Journal for Information Technology Studies as a Human Science 4.2-3 (2000).

Verhoeven, Deb, et al. « Controlling for openness in the male-dominated collaborative networks of the global film industry. » PloS one 15.6 (2020): e0234460.