



Colloque **Humanistica 2021**

Recueil des résumés



Rennes, 10-12 mai 2021

Comité scientifique du colloque :

Martin Grandjean (co-président), Nicolas Thély (co-président),
Aurélien Berra, Emmanuel Château-Dutier, Aurélie Hess,
Fatiha Idmhand, Morgane Mignon et Florence Thiault

Illustration: CC-BY-SA Édouard Hue (Wikimedia Commons)

Argumentaire du colloque

Humanistica, l'association francophone des humanités numériques créée en 2014, a organisé en 2020 son premier colloque (www.humanisti.ca/colloque2020). La seconde édition de cette manifestation se déroule en ligne depuis Rennes à l'invitation de la MSHB du 10 au 12 mai 2021.

Cette année, nous suggérons de porter une attention plus particulière à trois grands thèmes importants pour les humanités numériques en raison de leurs implications sociales et de leur actualité : les pratiques d'enseignement à l'ère du numérique ; l'effet des méthodes computationnelles sur la recherche en SHS ; la conservation et la transmission des patrimoines culturels, artistiques et scientifiques.

Que changent les moyens numériques aux formations, et que change l'existence de formations au numérique ou en humanités numériques dans l'enseignement secondaire et supérieur ? Quelle est leur influence sur les pratiques pédagogiques, la transformation matérielle des supports et des ressources, les compétences acquises par les élèves, leurs choix d'orientation ? Qu'en est-il du côté des formations universitaires qui initient les étudiants à des techniques d'étude de corpus et leur offrent de nouvelles perspectives professionnelles ?

Quel est le niveau de complexité et quels sont les résultats des méthodes computationnelles mises en œuvre dans les projets d'humanités numériques ? Assiste-t-on à une poussée des expérimentations impliquant l'apprentissage automatique, l'intelligence artificielle et les approches *data driven* dans le domaine des sciences humaines et sociales ? Quelles en sont les effets épistémologiques dans la production et la diffusion des savoirs ?

Quelles sont les stratégies de conservation dans les domaines culturels et artistiques (danse, théâtre, musique, cinéma, arts plastiques, littérature) ou scientifiques (collections d'instruments, anciens et contemporains, archives de chercheurs) ? Comment les principes d'ouverture des données patrimoniales et des collections (par exemple selon les principes des données FAIR) sont-ils appliqués ? Quels nouveaux usages de recherche ou de médiation cette ouverture fait-elle apparaître ?

Le colloque est porté par les institutions suivantes : Maison des sciences de l'homme en Bretagne, université Rennes 2, SCD Rennes 2, URFIST Bretagne Pays de la Loire, Champs libres/musée de Bretagne et INRIA. Partenaires : Huma-Num et RnMSH.



Les champs libres



Humanistica, l'association francophone des humanités numériques

L'association Humanistica cherche à réunir autour d'actions communes toutes les personnes intéressées par le mouvement des *digital humanities* telles qu'elles peuvent s'exercer et se penser en langue française. Voici ses missions :

RASSEMBLER. Humanistica rassemble la communauté des acteurs et actrices de la recherche et de l'enseignement supérieur (chercheurs, ingénieurs, enseignants, étudiants, métiers de l'information scientifique et technique, etc.) qui font usage des outils numériques dans leurs travaux en sciences humaines et sociales ou prennent les transformations numériques du savoir pour objet d'étude.

REPRÉSENTER. Humanistica s'implique, d'une part, vis-à-vis des institutions académiques francophones, pour lesquelles les enjeux et pratiques ne sont pas toujours évidents, d'autre part, vis-à-vis des autres organisations internationales œuvrant pour le développement des humanités numériques, auprès desquelles la spécificité francophone peut et doit être valorisée. Humanistica est une organisation constituante de l'Alliance of Digital Humanities Organizations (ADHO). À ce titre, elle est impliquée dans les différentes activités d'ADHO, en particulier l'organisation du colloque international annuel *Digital Humanities*.

ANIMER. S'associant à des partenaires ou des événements ponctuels, Humanistica suscite des débats et des échanges sur des questions propres à ce champ en pleine institutionnalisation. Elle maintient et anime la liste de diffusion francophone *Digital Humanities*. Sur demande, ou lorsque l'occasion s'y prête, l'association apporte également son soutien à des événements déjà existants ou facilite leur accès à des étudiantes et étudiants.

RÉALISER. Ne se limitant pas à être un lieu de débat pour une communauté constituée, l'association est porteuse de projets concrets : la publication de la revue *Humanités numériques* (deux numéros annuels depuis 2020) et un colloque annuel (à partir de 2020). Humanistica héberge plusieurs groupes de travail, notamment les suivants : arts, design et humanités numériques ; enjeux professionnels du numérique en SHS ; francophonie.

Site Web : www.humanisti.ca



À propos de ce recueil

Les textes présentés dans ce recueil sont les résumés soumis par les autrices et auteurs et sélectionnés par le comité scientifique du colloque Humanistica 2021. Leur contenu ainsi que leurs métadonnées sont une version harmonisée des textes extraits de la plateforme humanistica2021.sciencesconf.org.

Le comité scientifique du colloque remercie les expertes et experts ayant accepté de participer à l'évaluation des 61 communications reçues (entre 5 et 8 évaluations par communication) :

René Audet, Anne Baillot, Clarisse Bardiot, Francesco Beretta, Aurélien Berra, Michael Bourgatte, Carmen Brando, Jean-Baptiste Camps, Emmanuel Château-Dutier, Frédéric Clavert, Thibault Clérice, Claire Clivaz, Francesca Frontini, Ioana Galleron, Martin Grandjean, Aurélie Hess, Fatiha Idmhand, Vincent Jolivet, Olivier Le Deuff, Sabine Loudcher, Jean-Philippe Magué, Benoît Majerus, Sophie Marcotte, Morgane Mignon, Emmanuelle Morlock, Enrico Natale, Marie Puren, Vincent Razanajao, Marianne Reboul, Marine Riguet, Yannick Rochat, Glenn Roe, Sara Tandar, Laurent Tessier, Nicolas Thély et Florence Thiault.

Table des matières

PLÉNIÈRE D'OUVERTURE	7
L'intelligence artificielle pour décrire, indexer et explorer des collections multimédias : comment faire sans Google ?	8
COMMUNICATIONS ORALES	9
Automatisation des réseaux de personnages : proposition de cadre formel pour démystifier les algorithmes	10
Comment exploiter un corpus à l'aide des technologies du Web sémantique ? Le cas de la correspondance d'Henri Poincaré	12
Contextualisation de la critique francophone d'art contemporain	14
Contribution à l'histoire des méthodes de l'enregistrement archéologique de terrain. État d'avancement de travaux en cours sur l'étude d'archives de fouille par leur transcription numérique	16
Des prononciations et des cartes : une fenêtre ouverte sur la longue durée. Un exemple à partir de la langue bretonne	18
Dialogue interdisciplinaire en humanités numériques : le cas de l'évolution des formes discursives en critique d'art	21
Du « Chantre tout divin » au « Ménétrier déconfit » : exploiter les sources anciennes d'un mythe pour mieux détecter les correspondances entre ses réécritures	24
Exploration du théâtre alsacien à travers ses listes de personnages pendant la période 1870-1940	27
FAIRiser des données : état des lieux, barrières et choix. Une réflexion à partir des données des corpus d'auteurs	30
IA et approches participatives dans les humanités numériques : de la conjonction à la coopération	32
Image et texte : les livres d'heures manuscrits vus par l'intelligence artificielle	34
Interactions avec la sirène de Cagniard-Latour en réalité virtuelle	36
Katabase : à la recherche des manuscrits vendus	39
L'observation (numériquement équipée) des problèmes publics	42
L'utilisation de l'outil Stylo dans le cadre de la revue <i>Sens public</i>	44
Le recours aux humanités numériques pour (re)constituer le dossier génétique d'une création collective	46
Le rôle des modalités linguistiques dans l'analyse automatique du style critique	48
Le style polémique dans les correspondances de Voltaire	51
Les humanités numériques pour valoriser le patrimoine minier	53
Les matériaux de recherche ethnographiques à l'épreuve de l' <i>open data</i>	56

Modéliser le feuilleton. Créer et analyser une nouvelle archive patrimoniale avec les méthodes computationnelles	58
Moyen Âge et performances (MAP) : (modéliser un) répertoire de performances médiévales	61
POPP. Projet de reconnaissance optique des caractères des recensements de la population parisienne	64
Préserver et diffuser le patrimoine textile européen grâce aux technologies du Web sémantique. La réutilisation des données patrimoniales par le projet SILKNOW	66
Saisir le buzz politique sur Internet : le cas Juan Branco autour de son livre <i>Crépuscule</i>	68
Seconde traversée (<i>Défteros ploús</i>) : nouvelles et développements du projet d'édition collaborative numérique de l' <i>Anthologie palatine</i>	70
Théâtre classique et difficultés numériques : quelques réflexions sur l'analyse dramatique assistée par ordinateur	73
Traduire sans trahir, la vocation des logiciels ou celle des chercheurs ? Programmes de recherche et programme informatique au carrefour des cultures et des méthodologies	75
Un outil de dépouillement de sources archivistiques basé sur des technologies XML	78
Utilisation d'approches automatiques pour la reconnaissance des expériences de lecture	81
« Je pense que ça traite d'expérience de lecture, à voir... » : retour sur une expérience d'annotation collaborative	84
PRÉSENTATIONS ÉCLAIR	86
Adressbuch 2021 : la troisième vie de l'Annuaire des Allemands à Paris de 1854	87
Éditer et transmettre le <i>Parallèle des Anciens et des Modernes</i> de Charles Perrault	89
Essai de cartographie de médias imaginaires du XIX ^e siècle	91
Géo-visualisation des contenus de la Perséide Athar : le cas du Caire moderne. Un exemple d'intégration de données documentaires, sémantiques et spatiales au service des historiens	94
La médiation du patrimoine culturel et l'éducation : le rôle de la plate-forme numérique <i>e-learning</i> dans la conservation et la transmission de patrimoine	95
Langues celtiques, typographie et hybridation	97
Le projet ITHAC entre méthodes numériques, recherche multidisciplinaire et propositions pédagogiques	98
Le projet « Démêler le cordel » : une bibliothèque numérique pour l'étude de la littérature éphémère espagnole du XIX ^e siècle	100
Les bibliothécaires et leur rapport au livre numérique au Québec : vers une vision intégrative du patrimoine écrit	102
Paul Otlet en son réseau. Présentation de l'Otletosphère	104
Penser la phase post-pandémique au supérieur : quels processus d'appropriation des technologies éducatives dans les pratiques enseignantes ?	106
Reconnaissance et certification dynamique dans les dispositifs contributifs du secondaire : opportunité et levier de développement des habiletés numériques des élèves et des enseignants	108
SITT : Streaming, interactivité, téléperformance, transmédia. Un projet de sauvegarde du patrimoine chorégraphique par le développement d'outils technologiques spécifiques	109

ATELIERS DE GROUPE DE TRAVAIL	111
Humanités numériques, éducation et formation	112
Les éditions critiques numériques : langues, textes et fragments	114
Littérature et numérique : les sessions d'atelier du groupe CCLA	115
Réflexions sur l'évolution des formes et des structures argumentatives	116
ATELIERS DE FORMATION	118
Atelier de formation TraduXio : une plateforme de traduction collaborative en ligne	119
Écriture et édition scientifiques avec l'éditeur de texte Stylo	121
Lemmatiser des textes et corriger l'annotation grâce à l'apprentissage profond avec Pyrrha	122
ATELIER PARTENAIRE	124
L'open content au musée : diffusion, contribution et innovation	125
PROGRAMME DU COLLOQUE	126

Plénière d'ouverture

L'intelligence artificielle pour décrire, indexer et explorer des collections multimédias : comment faire sans Google ?

Guillaume Gravier¹

1 : IRISA
CNRS : UMR6074

Décrire, indexer, explorer, analyser des collections de textes, d'images, de vidéos constitue une part importante du travail en humanités numériques, pour lequel les techniques d'intelligence artificielle ouvrent en théorie de nombreuses possibilités. En pratique, cependant, les limites des technologies grand public disponibles et immédiatement applicables à la description, l'indexation et la fouille de tels contenus sur une problématique donnée – les « technologies Google » par abus de langage – les rendent la plupart du temps inexploitable sans un considérable effort d'ingénierie pour les adapter. Nous argumenterons que faire « sans Google » requiert une co-construction de projets de recherche en humanités numériques et en intelligence artificielle pour, d'une part, faire évoluer la capacité des outils d'intelligence artificielle à appréhender de nouveaux problèmes rapidement et sans expertise particulière et, d'autre part, faire progresser la recherche en sciences humaines et sociales en l'outillant plus efficacement.

Après une brève présentation des principes sous-jacents de l'intelligence artificielle pour en comprendre l'intérêt et les limites, nous analyserons des projets de recherche à l'interface entre intelligence artificielle et humanités pour mettre en évidence les difficultés rencontrées et les questions de recherche qui en résultent en informatique. Nous prendrons notamment appui sur plusieurs projets : iCODA, qui concerne l'extraction d'information dans les archives du quotidien Ouest-France ; READ-IT, qui concerne la création d'une base de données outillée pour l'étude du patrimoine de la lecture ; ou encore ARCHIVAL, qui concerne l'exploration de fonds documentaires sur l'autogestion. Tous ces projets allient « concepteurs » et « consommateurs » de technologie pour répondre aux besoins des deux catégories.

Guillaume Gravier est directeur de recherche au CNRS et directeur de l'Institut de recherche en informatique et systèmes aléatoires (UMR 6074). Il est spécialiste de l'apprentissage statistique pour l'analyse de contenus multimédias. De 2014 à 2018, il a été responsable de l'équipe de recherche Linkmedia dédiée aux technologies pour l'exploration de collections multimédias. Il a été impliqué dans de nombreux projets en lien avec les humanités numériques, notamment aux côtés de l'Institut national de l'audiovisuel, de la Bibliothèque nationale de France ou encore de la Fondation des maisons des sciences de l'homme. Récemment, il a coordonné le projet iCODA sur l'exploration des archives du journal Ouest-France et, en binôme avec Brigitte Ouvry-Vial, le projet européen READ-IT sur le patrimoine de la lecture en Europe.

Communications orales

Automatisation des réseaux de personnages : proposition de cadre formel pour démystifier les algorithmes

Coline Métrailler¹

1 : Section des sciences du langage et de l'information, université de Lausanne

Les réseaux de personnages offrent un regard particulier sur les œuvres de fiction, permettant de prendre de la distance avec la narration, d'en proposer une vision globale et de dégager des tendances difficiles à capturer avec une approche de type *close reading* [7]. Mais tandis que la génération de tels réseaux peut s'avérer fastidieuse si l'on doit relever manuellement les occurrences des personnages, son automatisation [2, 3, 5] soulève des questions sur la qualité des résultats obtenus [8]. Pour contrer l'aspect « boîte noire » des algorithmes informatiques et inviter une communauté littéraire à s'en emparer, il semble ainsi primordial de mettre en perspective les deux approches, et d'aborder la notion de qualité des résultats avec des outils formels permettant d'évaluer la perte d'information résultant de l'automatisation et son incidence concrète sur le réseau engendré.

L'objectif de cette présentation est d'appliquer ce cadre formel à la question de la génération automatique d'un réseau de personnages. Ce processus étant en réalité construit comme une succession d'outils d'automatisation, il ne s'agit pas seulement d'évaluer le résultat final, mais également la performance de chacun de ces outils, afin de mesurer l'impact des différentes étapes sur le réseau engendré (qui est par définition une agrégation des informations obtenues, et dont on peut espérer qu'il soit robuste vis-à-vis de petites variations dans les étapes de la construction des données). La démarche est illustrée ici à l'aide d'un corpus constitué de romans de fiction anglophones et francophones, mais une telle approche méthodologique peut évidemment se décliner pour d'autres types de médias et de formats.

Dans le cas des romans de fiction, le premier module à considérer est celui de la reconnaissance d'entités nommées (*natural language processing*, NER), qui consiste en la récupération automatisée des noms propres contenus dans le texte (personnages, lieux, organisations, etc.). Cette étape est cruciale dans la création du réseau, puisqu'elle permet d'identifier les éléments à mettre en lien : un personnage qui n'aurait pas été détecté à ce stade ne figurera pas dans le réseau final. Il s'agit également d'une étape très chronophage si elle est effectuée manuellement, ce qui rend l'emploi de méthodes automatisées particulièrement tentant.

Une première évaluation systématique formelle a été mise en œuvre afin de choisir l'algorithme de NER le plus approprié pour les données choisies. La méthodologie utilisée pour départager les deux candidats les plus prometteurs [10] (*spaCy* [6] et *Flair* [1]) repose sur les éléments suivants : un référentiel commun (par exemple un texte annoté manuellement sur lequel seront testés les deux algorithmes), un protocole d'uniformisation des résultats obtenus, et enfin des scores de performance applicables aux deux algorithmes. Après comparaison desdits scores, *Flair* apparaît comme plus précis et plus exhaustif dans sa reconnaissance d'entités nommées. La même méthodologie peut être appliquée pour mesurer la performance d'un modèle automatisé face à une annotation entièrement manuelle, ou encore l'impact d'ajustements apportés à l'algorithme (en comparant les scores de l'ancienne et la nouvelle version du modèle).

Une approche similaire permet d'évaluer la qualité du réseau final. Si toute tentative d'établir un gold standard pour un réseau de personnages dépend nécessairement des choix de l'annotateur (type d'interactions que l'on cherche à illustrer, mais également nombre d'informations à afficher, entre autres choix tout aussi importants), il est tout de même possible de construire un réseau manuel, selon les mêmes choix et définitions que la version automatisée, et de s'en servir comme exemple à suivre, comme objectif à atteindre pour notre modèle algorithmique. Des mesures de comparaison permettent ensuite de quantifier non seulement les différences entre le réseau automatique et ce réseau de référence, mais également les

différences entre deux réseaux générés automatiquement (toujours en regard du réseau de référence) et donc l'impact des changements faits en début, milieu ou fin du processus de génération sur le résultat final.

Selon la complexité des données et des informations que l'on souhaite extraire et mettre en évidence, ce type de cadre formel peut donner une bonne intuition de la quantité de personnages et d'interactions à détecter pour engendrer un réseau très proche du gold standard, ou du moins identifier les étapes les plus cruciales pour chercher à les améliorer. Le développement de ce type de méthodologie, basée sur des fondements précis et clairement énoncés, semble en tous les cas une approche intéressante en vue de démystifier les algorithmes [4], d'évaluer la faisabilité (ou non) d'une automatisation satisfaisante de la création de réseaux de personnages, et de fournir une base stable de discussion entre sciences humaines et sciences computationnelles autour de ces objets.

Bibliographie

1. Akbik, Alan, Duncan Blythe et Roland Vollgraf. « Contextual String Embeddings for Sequence Labeling ». Dans *COLING 2018, 27th International Conference on Computational Linguistics*. 2018, p. 1638-1649.
2. Amancio, Diego R. « Network analysis of named entity co-occurrences in written texts ». *EPL (Europhysics Letters)* 114, no 5 (1 juin 2016). <https://doi.org/10.1209/0295-5075/114/58005>.
3. Celikyilmaz, Asli, Dilek Hakkani-Tur, Hua He, Greg Kondrak, et Denilson Barbosa. « The Actor-Topic Model for Extracting Social Networks in Literary Narrative ». Dans *NIPS Workshop : Machine Learning for Social Computing*, 7.
4. Da, Nan Z. « The Computational Case against Computational Literary Studies ». *Critical Inquiry* 45, no 3 (1 mars 2019) : 601-639. <https://doi.org/10.1086/702594>.
5. Elson, David, Nicholas Dames, et Kathleen McKeown. « Extracting Social Networks from Literary Fiction ». Dans *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 138-147. Uppsala, Sweden: Association for Computational Linguistics, 2010. <https://www.aclweb.org/anthology/P10-1015>.
6. Honnibal, Matthew et Ines Montani. « spaCy 2 : Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing ». *A paraître*. 2017.
7. Moretti, Franco, « Network Theory, Plot Analysis ». Dans *New Left Review*, 2011.
8. Labatut, Vincent, et Xavier Bost. « Extraction and Analysis of Fictional Character Networks: A Survey ». *ACM Computing Surveys* 52, no 5 (13 septembre 2019) : 1-40. <https://doi.org/10.1145/3344548>.
9. Soundarajan, Sucheta, Tina Eliassi-Rad, et Brian Gallagher. « A Guide to Selecting a Network Similarity Method ». Dans *Proceedings of the 2014 SIAM International Conference on Data Mining*, 1037-45. Society for Industrial and Applied Mathematics, 2014. <https://doi.org/10.1137/1.9781611973440.118>.
10. Stanislawek, Tomasz, Anna Wróblewska, Alicja Wójcicka, Daniel Ziemnicki, et Przemyslaw Biecek. « Named Entity Recognition – Is There a Glass Ceiling ? » Dans *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 624-633. Hong Kong, China : Association for Computational Linguistics, 2019. <https://doi.org/10.18653/v1/K19-1058>.

Comment exploiter un corpus à l'aide des technologies du Web sémantique ? Le cas de la correspondance d'Henri Poincaré

Nicolas Lasolle^{1,2}, Pierre Willaime¹

1 : UMR 7117 Laboratoire d'histoire des sciences et de philosophie – Archives Henri-Poincaré, université de Strasbourg, université de Lorraine, CNRS

2 : UMR 7503 Laboratoire lorrain de recherche en informatique et ses applications, Institut national de recherche en informatique et en automatique, université de Lorraine, CNRS

Le Web sémantique peut être vu comme particulièrement adapté à la description d'un corpus dans le cadre d'un projet d'humanités numériques. En effet, cette extension du Web permet de modéliser les entités, relations, concepts, ou autres métadonnées pouvant décrire les documents d'archives, les acteurs historiques en question et leur contexte. Cette structuration s'appuie sur des technologies standardisées (W3C) constituant le Web sémantique (Berners-Lee et al, 2001) telles que RDF (Resource Description Framework), RDFS (RDF Schema) et OWL (Web Ontology Language). La « sémantisation » d'un corpus ouvre la voie à son traitement automatique par des méthodes computationnelles.

Dans cette communication, nous nous proposons de décrire notre usage des technologies du Web sémantique dans le cadre du projet d'édition de la correspondance d'Henri Poincaré (1854-1912). Ce corpus est composé d'environ 2100 lettres pour lesquelles une numérisation du document original, une transcription, un appareil critique un ensemble de métadonnées descriptives sont disponibles. Ce projet est porté collectivement par les Archives Henri Poincaré (UMR 7117 – CNRS/université de Lorraine/université de Strasbourg). Plus précisément, nous souhaitons développer les méthodes et réflexions entamées pour répondre à deux problématiques rencontrées lors de ces travaux. La première problématique est liée à la représentation des connaissances de ce corpus. La seconde concerne l'exploration des données du corpus avec des outils computationnels.

Le premier enjeu de représentation des connaissances permet de mesurer l'expressivité permise par les ontologies (ici à comprendre comme des schémas de métadonnées structurés sur plusieurs niveaux). Les métadonnées décrivant finement le corpus ont pu être ainsi piochées dans différentes ontologies (FOAF, BIO, BIBO, Relationship, Dublin Core). Les manques d'expressivité sur des points précis, problématiques pour les enjeux scientifiques, ont été comblés par la création d'une nouvelle ontologie, nommée *ahpo*. Celle-ci décrit la structuration du modèle de connaissances liant les différents schémas de métadonnées préexistants et précise l'introduction de métadonnées propres. L'alignement partiel permet de conserver une bonne interopérabilité. Nous nous intéresserons tout particulièrement à deux difficultés : la prise en compte d'une validité temporelle dans le renseignement des métadonnées (Gutierrez et al, 2005) et la gestion de l'incertitude sur leur contenu (Stoilos et al, 2006). La question de la temporalité peut être exemplifiée par la ville de résidence d'une personne. Il est souhaitable de pouvoir conditionner la validité des contenus de cette métadonnée à des contraintes temporelles. L'incertitude quant à elle peut concerner la datation des lettres. Il peut être possible de spécifier une date butoir d'envoi, Poincaré faisant référence à la lettre en question dans une autre, sans pour autant être en mesure de préciser. Ces deux questions, la représentation temporalisée des connaissances et la gestion de l'incertitude et de l'approximation, sont tout particulièrement intéressantes pour l'exploitation scientifique du corpus.

Ce second enjeu d'exploitation se propose d'utiliser, et d'étendre, les capacités d'interrogation permises par le Web sémantique. Dans le domaine, le langage SPARQL est le standard pour interroger des graphes de données. C'est un langage expressif qui permet de formuler des requêtes complexes pour exploiter les liens entre les éléments d'un graphe de données. Par exemple, un historien pourrait formuler une requête afin de récupérer les lettres échangées entre David Hilbert et Henri Poincaré en 1890 et qui traitent de géométrie

non-euclidienne. Cependant, les résultats de l'exécution de requêtes ne sont parfois pas suffisants pour apporter des réponses à une problématique de recherche et cela oblige ainsi les chercheurs à devoir formuler de nouvelles requêtes. L'idée du mécanisme de recherche approchée est de créer des règles de transformation de requêtes afin d'assouplir ou modifier certaines contraintes de façon automatique. Par exemple, dans le cas de la requête présentée ci-dessus, il est possible de relâcher des contraintes en étendant les bornes temporelles liées à l'expédition de la lettre, en recherchant les lettres avec des thèmes scientifiques proches, ou en s'intéressant aux échanges avec les correspondants de Poincaré qui ont collaboré avec David Hilbert. Une autre idée pour exploiter les connaissances du corpus est de mettre en place des règles d'inférences de manière à compléter automatiquement des bases de connaissances. L'objectif est de compléter l'édition des données du corpus et de dégager de nouvelles connaissances pour s'assurer que les recherches sur le corpus retournent l'ensemble des résultats attendus.

Ces travaux de recherche et méthodes ne sont pas spécifiques au corpus d'Henri Poincaré et ont pour objectif d'être réutilisés pour exploiter d'autres corpus en SHS et de proposer des extensions aux outils traditionnels du Web sémantique (Bruneau et al, 2021).

Références

- Bruneau, Olivier, Nicolas Lasolle, Jean Lieber, Emmanuel Nauer, Siyana Pavlova, et Laurent Rollet 2021. « Applying and Developing Semantic web Technologies for Exploiting a Corpus in History of Science : The Case Study of the Henri Poincaré Correspondence ». *Semantic Web*, 12(2), IOS Press
- Berners-Lee, Tim, James Hendler, et Ora Lassila. 2001. « The Semantic Web ». *Scientific American*, 284(5), 34-43.
- Ghorbel, Fatma, Faycal Hamdi, Elisabeth Métails, Nebrasse Ellouze, et Faiez Gargouri. 2018 « A Fuzzy-Based Approach for Representing and Reasoning on Imprecise Time Intervals in Fuzzy-OWL 2 Ontology ». *International Conference on Applications of Natural Language to Information Systems*. 167-178.
- Gutierrez, Claudio, Carlos Hurtado, et Alejandro Vaisman. 2005. « Temporal RDF ». *European Semantic Web Conference*. 93-107.
- Stoilos, Giorgos, Nikos Simou, Giorgos Stamou, et Stefanos Kollias. 2006. « Uncertainty and the semantic Web ». *IEEE Intelligent Systems*, 21(5), 84-87.

Contextualisation de la critique francophone d'art contemporain

William Diakité¹

1 : ED Arts, lettres, langues (ALL), EA 7472 Pratiques et théories de l'art contemporain (PTAC), université Rennes 2, EA Linguistique, ingénierie et didactique des langues (LIDILE), université Rennes 2

Cette proposition de communication a pour objectif de présenter une expérimentation conduite dans le cadre d'une recherche doctorale, initiée en 2017, et qui s'effectue dans le domaine de l'esthétique et des sciences de l'art.

À l'initiative des équipes d'accueil Pratiques et Théories de l'Art Contemporain (PTAC) et Linguistique, Ingénierie et Didactique des Langues (LIDILE) de l'université Rennes 2 et soutenue par la Maison des sciences de l'homme en Bretagne, cette recherche vise à concevoir un appareillage permettant aux chercheurs en art d'étudier d'une part les discours de l'art relevant de la critique professionnelle, d'autre part de mettre en relation les corpus étudiés avec d'autres corpus comprenant des informations liées aux événements sociaux et politiques.

La période concerne les années 1990 à nos jours et relève de la communauté linguistique francophone. Cette période est caractérisée par de profondes transformations des arts plastiques et de la critique. En France, ces transformations prennent leurs racines au moment de ce que l'on a nommé la « crise de l'art contemporain » (Michaud 2011). Ces transformations ont invité de nombreux philosophes de l'art et esthéticiens à revoir leur position et leur rôle face à ces différents renouvellements. Le philosophe Jean-Pierre Cometti, par exemple, tend à radicalement distinguer la critique d'art de la philosophie en insistant sur le fait que « ni l'interprétation (au sens herméneutique) ni l'évaluation [des œuvres d'art] ne sont à proprement parler des tâches philosophiques » (Cometti 2009). Ce dernier constat l'amène à proposer une *esthétique minimale* définie comme « une esthétique qui reconnaît à la critique une priorité et qui peut se proposer d'examiner les conditions d'effectuation (y compris effectives) de la reconnaissance dont l'art est solidaire » (*ibid.*).

À l'occasion de cette communication, il s'agira de présenter les premiers résultats d'une expérimentation qui vise à fournir une assistance à la contextualisation des productions de critiques d'art. Si ces travaux portent effectivement sur un genre de discours particulier, soulignons que l'expérience pourrait être reproduite pour l'étude d'autres genres discursifs.

Cette expérimentation repose sur une interface web qui permet de charger un corpus encodé en TEI afin d'en extraire des informations. Cependant, celles-ci ne sont pas destinées à être interprétées directement par l'analyste comme c'est le cas dans les disciplines de l'analyse du discours. En effet, les informations produites par les algorithmes d'extractions sont utilisées pour construire une requête http permettant de communiquer avec des API (Application Programming Interface) de moteurs de recherche telles que celle de Google ou alors de Bing. Ces APIs permettent de questionner le web et mettent à disposition un ensemble de paramètres permettant d'affiner les requêtes (date de mise en ligne, opérateurs logiques, filtrage des sites web ou des types de ressources, etc.) Ainsi, l'hypothèse que nous discuterons durant la présentation est que, pour un document du corpus, il est possible de reconstruire son contexte d'énonciation ou autrement dit, de retrouver les ressources du web (articles de presse, de blog, de Wikipédia, etc.) en lien avec ledit document. La visée de l'outil s'inscrit donc dans une démarche d'objectivation du regard porté par l'analyste lors de l'examen de productions textuelles.

Notre corpus de travail est constitué de textes produits par le critique d'art Frédéric Bonnet. Ceux-ci ont été collectés à l'aide d'une méthode de scraping[1] depuis le site du Journal des Arts (www.lejournaldesarts.fr), un bimensuel fondé en 1994 consacré aux actualités sur l'art et son marché. Le corpus a été encodé en TEI

(« édition de lecture ») en suivant les recommandations du consortium CAHIER en matière de corpus d'auteurs (Galleron *et al.* 2018). Il est composé de 946 documents, soit un total de 653308 mots, étalés sur une période allant de 2001 à 2017 (sans production en 2004).

Dans ce cadre, nous souhaitons présenter les étapes de constitution de notre corpus ainsi que l'application. Après avoir présenté les méthodes et les résultats des différentes évaluations de notre système, nous souhaitons discuter des apports d'une telle approche dans le cadre d'une étude de cas.

Cette expérimentation entre en résonance avec plusieurs recherches conduites ces dernières années : dans le domaine de la critique d'art, Nicolas Thély a proposé quatre opérations de contextualisation des productions critiques (Thély 2016). Ses expérimentations ont pour enjeu de révéler les conditions de production des textes critiques en les confrontant à différents types d'informations comme des quantités mesurées, des données GPS décrivant le parcours des critiques d'art, des informations extraites par des outils du TAL (Traitement Automatique des Langues) ainsi que des graphes.

Dans le domaine des sciences politiques, Hamzaoui *et al.* présentent une expérimentation portée sur la contextualisation du discours politique français (2019). Leur approche consiste à extraire des termes multimots pour constituer une requête de recherche. Une fois exécutée, cette requête fait remonter les articles Wikipédia les plus pertinents. Enfin, ces articles sont agrégés en un court résumé qui constitue le résultat de la procédure. Notre approche diverge de la leur en ce sens qu'il ne nous incombe aucunement de constituer un résumé à partir d'un corpus de référence mais bien de faire remonter des objets que l'on considérera se constituer en contexte d'énonciation.

L'approche que nous souhaitons présenter dispose de nombreux avantages. D'abord, elle repose sur un corpus extrait automatiquement du web. Se faisant, nous nous affranchissons des coûts liés à la numérisation-description-océrisation des documents. D'autre part, contrairement à l'approche proposée par Hamzaoui *et al.*, elle ne nécessite pas la constitution d'un corpus de référence. Enfin, et bien que les conséquences ne soient pas neutres, l'utilisation d'API permet là encore de s'affranchir de nombreuses difficultés lors de la mise en place technique. En ce sens notre approche tend à autonomiser l'analyste dans un cadre de recherche en environnement numérique.

Bibliographie

- Cometti, Jean-Pierre. 2009. *La Force d'un malentendu. Essais sur l'art et la philosophie de l'art*. Questions Théoriques. Vrin.
- Galleron, Ioana, Marie-Luce Demonet, Cécile Meynard, Fatih Idmhand, Elena Pierazzo, Geoffrey Williams, Yves Buard, et Julia Roger. 2018. « Les publications numériques de corpus d'auteurs », 18. <https://cahier.hypotheses.org/guides-juridiques/les-publications-numeriques-de-corpus-dauteurs>
- Hamzaoui, Ouassim, Tania Jiménez, Christèle Lagier, et Eric SanJuan-Ibekwe. 2019. « Contextualisation du discours politique ». *Document numérique* Vol. 22 (1) : 63-84.
- Michaud, Yves. 2011. *La crise de l'art contemporain*. Presses Universitaires de France. <https://doi.org/10.3917/puf.micha.2011.01>.
- Thély, Nicolas. 2016. « Quatre opérations pour étudier la critique d'art à l'époque des humanités numériques ».

[1] Rappelons que la régulation française sur le scraping est particulièrement ardue à interpréter et se faisant, nous nous sommes reposés sur les mentions légales du site, celles-ci ne faisant pas état de contre-indications concernant l'usage de méthodes d'extraction automatisée (<https://www.lejournaldesarts.fr/mentions-legales>)

Contribution à l'histoire des méthodes de l'enregistrement archéologique de terrain. État d'avancement de travaux en cours sur l'étude d'archives de fouille par leur transcription numérique

Christophe Tufféry^{1,2}, Claudia Marinica³, Maximilien Rioult³, Yulin Xie³

1 : Institut national de recherches archéologiques préventives (INRAP), ministère de la Culture et de la Communication, ministère de l'Enseignement supérieur et de la Recherche scientifique

2 : UMR 7324 Cités, territoires, environnement et sociétés, université de Tours, CNRS

3 : Polytech Nantes, université de Nantes

Pour une thèse en cours à Cergy Paris Université (EUR Humanités, Création, Patrimoine) en partenariat avec l'Institut National du Patrimoine, une étude est en cours des contenus de plusieurs carnets de terrain concernant des fouilles archéologiques programmées et préventives réalisées entre les années 1970 et aujourd'hui. Les contenus des carnets sont d'une grande richesse pour reconstituer l'histoire d'un chantier de fouille et pour contribuer à l'historiographie de l'archéologie française qui, entre les années 1970 et les années 1990, connaît de profondes mutations.

D'une part pendant cette période l'archéologie de sauvetage qui devient préventive connaît le début de son essor, liée à la multiplication de grands travaux d'aménagement du territoire. Ce développement s'appuie notamment sur la création en 1973 de l'Association pour les fouilles archéologiques nationales (AFAN). En nombre de chantiers, l'archéologie préventive devient plus importante que l'archéologie programmée et prend une dimension internationale par la mise en œuvre, en 1992, de la Convention européenne pour la protection du patrimoine archéologique (Demoule, 2007).

D'autre part, l'apparition de la micro-informatique à partir du milieu des années 1970 et sa diffusion large à partir du milieu des années 1980 se traduit par le développement des premiers systèmes d'enregistrement numérique de terrain, comme par exemple le système Syslat (Py, 1991), souvent sur des matériels de la marque Apple dont l'ergonomie et la facilité d'utilisation rencontre un vif succès dans les communautés archéologiques.

Une partie des carnets de terrain d'un chantier de fouille archéologique, conduit entre 1972 et 1990, est pris comme cas d'étude. Renseignés par plusieurs des fouilleurs, les contenus de ces carnets consistent en des annotations manuscrites concernant la vie du chantier, des observations scientifiques, des découvertes de mobilier archéologiques, des choix dans la conduite de la fouille, des hypothèses d'interprétation. D'autres types de contenus d'apparence plus anecdotiques ont aussi leur importance : annotations sur la météorologie, sur les comportements des fouilleurs, des passages humoristiques, des notes et des croquis réalisés sur des morceaux de papier puis scotchés, des tirages photographiques collés, des cartes postales insérées.

La numérisation de ces carnets de terrain a été réalisée pour en entreprendre la transcription numérique. Des premiers essais ont été conduits avec le logiciel Transkribus qui permet notamment la reconnaissance automatique d'écriture. Après des débuts décevants avec cette solution, une évaluation d'autres solutions a été entreprise dans le cadre d'un projet exploratoire en collaboration avec le Laboratoire des Sciences du Numérique de Nantes (équipe DUKe) et Polytech Nantes (université de Nantes). Le choix s'est porté sur le développement d'une application basée sur la plateforme Scribe[1]. Avec le prototype développé, le travail de transcription manuelle a pu être commencé.

Une proposition de typologie a été nécessaire pour guider la distinction entre les divers types de contenus : zones de texte autonome, zones de commentaires associés, objets graphiques, objets importés, etc. Cette

typologie permet d'obtenir des contenus numériques qui servent à observer de la tenue de ce type d'archives comme de celle des pratiques de chantier. Ainsi, si les carnets de terrain sont des œuvres collectives pendant les premières campagnes de fouilles, ils ne sont plus renseignés que par un seul auteur lors des dernières campagnes. Par ailleurs, l'analyse des contenus permettrait de faire ressortir les choix effectués des zones fouillées, des structures dégagées, des relevés en plan et en coupe, des blocs de sédiments prélevés pour des analyses en laboratoires, des interprétations avancées, individuelles ou collectives, etc. Selon les fouilleurs, les proportions dans les divers types de contenus ne sont pas identiques.

Les premiers résultats seront présentés et mis en perspective par rapport à d'autres travaux en partie similaires, conduits en France sur ce même type d'archives comme le projet Bulliot, Bibracte et moi (Depalle *et al.*, 2019)[2].

Des relations étroites apparaissent entre la nature des informations et les modalités de leur inscription selon que celle-ci est faite sur des carnets de terrain et des fiches d'enregistrement papier ou dans des fichiers d'enregistrement numérique. Les effets des formalismes de l'enregistrement numérique sur les savoirs archéologiques se vérifient directement. Ainsi, l'usage progressif de fiches normées d'enregistrement prend le pas sur la tenue des carnets de terrain qui tendent à contenir de moins en moins d'indications au fur et à mesure des campagnes de fouilles.

Les conséquences de l'usage des dispositifs numériques sur la conservation et la diffusion des archives de fouille sont aussi évaluées. Elles relèvent de choix en matière de pérennité et d'accessibilité à des archives de fouille anciennes et de leur ouverture, dans la perspective des principes FAIR, permettant de nouvelles formes de valorisation et de médiation de données de la recherche archéologique considérées comme des données patrimoniales.

Enfin, une mise en perspective de ces travaux est proposée. Elle s'appuie sur l'évaluation de l'importance que revêtent l'inscription spatiale et temporelle des savoirs « ce qui les ancre dans un lieu et un moment particuliers ou ce qui leur permet de circuler et de se diffuser » (Jacob, 2011).

Références

Demoule Jean-Paul. 2007. « La naissance de l'archéologie préventive », Pour la Science, novembre 2007 : 146-151.

Depalle Claire, Durost Sébastien, Girard Jean-Pierre et Perrin-Touche Emmanuelle, 2019. « "Bulliot, Bibracte et moi" une expérience de sciences participative en archéologie », Culture et Recherche, « Recherche culturelle et sciences participatives », n° 140 : 78.

Jacob Christian. 2011. Lieux de savoirs. Tome 2 La main de l'intellect, Paris, Albin Michel.

Py Michel. 1991. « Principes, outils et implantation de "Syslat", Système d'enregistrement, de gestion et d'exploitation de la documentation issue des fouilles de Lattes », Lattara 4 : 117-122.

[1] <https://scribeproject.github.io/>

[2] <https://bbm.hypotheses.org/>

Des prononciations et des cartes : une fenêtre ouverte sur la longue durée. Un exemple à partir de la langue bretonne

Tanguy Sollicec¹

1 : Centre de recherche bretonne et celtique (CRBC), université de Bretagne occidentale (UBO)

Contextualisation

La réflexion sur l'histoire des langues et des populations qui les parlent se base avant tout sur l'analyse des formes anciennes des idiomes étudiés et de leur généalogie selon les principes, établis de longue date, de la méthode comparative en linguistique. La constitution des atlas linguistiques et de la géographie linguistique au tournant des XIX^e et XX^e siècles a permis d'ouvrir de nouvelles perspectives sur ces questions (Le Dù 2006, 2008, 2013). Cette approche offre un changement de point de vue en se proposant de réfléchir au temps long à partir de l'étude de formes linguistiques vernaculaires recueillies à une époque plus ou moins récente.

L'exploitation récente des données numérisées d'un atlas linguistique de la langue bretonne, le *Nouvel atlas linguistique de la Basse-Bretagne* (Le Dù 2001) par le biais d'un traitement quantitatif ou dialectométrie a ainsi permis d'apporter un nouvel éclairage sur les débuts de l'histoire de la Bretagne continentale, à la suite du déplacement de populations entre l'île de Bretagne et la péninsule armoricaine, à une période située entre les V^e et VII^e siècles (Sollicec 2021). Cet épisode historique, s'il a pu être considéré comme fondateur du point de vue de l'historiographie bretonne, n'en demeure par moins mal documenté en raison de la pauvreté des sources contemporaines (Fleuriot 1980; Coumert 2010; Brett 2011, 2019).

Au-delà de l'objet particulier de notre recherche, l'exploitation systématique des données d'un type de corpus particulier, notre communication envisage l'usage de certaines méthodes computationnelles comme une procédure d'objectivisation des données analysées. Elles permettent en effet de dévoiler des structures invisibles à l'œil nu. Le statut épistémologique de ces résultats demeure toutefois à préciser, de même que leur fonction au sein d'une démarche scientifique. Malgré ce rôle incertain, c'est le rendu visuel de ces résultats qui autorise un dialogue avec d'autres disciplines et permet ainsi de prolonger la réflexion.

Les atlas linguistiques, un type de corpus particulier

Un atlas linguistique correspond à une collection de cartes présentant pour un ensemble de points d'enquête les formes linguistiques correspondant à une notion précise (Dalbera 2002, 2007). Cette méthodologie d'analyse de la variation linguistique, à présent plus que centenaire, a été inaugurée dans le domaine francophone par l'*Atlas linguistique de la France* de Edmont & Gilliéron (1902-1910).

Les atlas linguistiques forment donc des corpus de données bien particuliers à plusieurs titres. Les données qu'ils contiennent sont avant tout linguistiques, reflétant principalement des usages vernaculaires. Elles possèdent également une très forte dimension spatiale, ce qui en fait des données géolocalisées. Néanmoins, un atlas linguistique se caractérise en tant que corpus de par ses limites. Il est une prolongation particulière de la liste de mots, dans la mesure où celle-ci est répétée autant de fois qu'il comporte de points d'enquêtes. Au delà de cette dernière idéalisation, l'analyse des données contenues dans un atlas linguistique renvoie aux spécificités des petits corpus, en opposition aux grands corpus nationaux. Nombre de langues humaines ne sont et ne seront documentées que par ce type d'outil (Adamou 2016). De ce fait, la question de leur statut méthodologique et épistémologique se pose, et notamment de quelle(s) façon(s) ces petits corpus contribuent à la description du fonctionnement des langues humaines malgré leur taille restreinte.

L'intérêt des atlas linguistiques pour décrire les faits linguistiques a ainsi été discuté (Canobbio 2004) et la disparition du groupe de recherche (GdR.) des atlas linguistiques au sein du CNRS (Le Dù 1997) a pu paraître comme un échec épistémologique à un moment où la dialectologie s'interroge sur les enjeux de sa discipline (Thun 2000; Dalbera 2002).

L'exploitation numérique des atlas linguistiques

L'application à ce type de données de techniques et de méthodes de type computationnel permet d'une part à la dialectologie de dépasser de l'ornière méthodologique dans laquelle elle se trouvait alors et d'autre part, de dévoiler la complexité de la stratigraphie historique sur laquelle repose les faits langagiers décrits dans les atlas linguistiques.

Notre démarche illustre la possibilité méthodologique de l'exploitation numérique de ces corpus particuliers par le biais d'une approche computationnelle ou dialectométrique (Séguy 1973; Goebel 1984, 2010; Nerbonne 2010; Nerbonne & Heeringa 2010; Wieling & Nerbonne 2015). La représentation cartographique des résultats obtenus grâce à un outil SIG permet d'accéder à la longue durée et ainsi, aux modalités selon lesquelles les dynamiques historiques modèlent un territoire (Braudel 1958). Ces visualisations créent les conditions d'un dialogue avec d'autres disciplines, notamment la génétique des populations (Pellen & Sollic accepté).

Références

- Adamou, Evangelia. 2016. *A Corpus-Driven Approach to Language Contact. Endangered Languages in a Comparative Perspective* (Language Contact and Bilingualism 12). Berlin, Boston : Walter de Gruyter.
- Braudel, Fernand. 1958. Histoire et Sciences sociales : La longue durée. *Annales* 13(4). 725-753. <https://doi.org/10.3406/ahess.1958.2781>.
- Brett, Caroline. 2011. Soldiers, Saints, and States ? The Breton Migrations Revisited. *Cambrian Medieval Celtic Studies* 61. 1-56.
- Brett, Caroline. 2019. An Invisible Migration ? The Origins of Brittany. In *Migrations et territoires celtiques : mouvement spatial et mutations culturelles/ Divroañ hag enbroañ ar Gelted : dilec'biañ ha cheñch sevenadur*, 11-26. Longueil/Rennes : Tir/L'instant même.
- Canobbio, Sabina. 2004. L'Atlas linguistique comme outil de recherche ? A propos de quelques expériences italiennes. *La Bretagne linguistique* 13 numéro spécial Dialectologie et Géolinguistique. 281-312.
- Coumert, Magali. 2010. Le peuplement de l'Armorique : Cornouaille et Domnonée de part et d'autre de la Manche aux premiers siècles du Moyen Age. In Magali Coumert & Hélène Tétrel (eds.), *Histoires des Breagnes – 1. Les mythes fondateurs*, 15-22. Brest : CRBC.
- Dalbera, Jean-Philippe. 2002. Géolinguistique : Un Nouveau Souffle ? *Revue Belge de Philologie Et D'Histoire* 80(3). 831-849.
- Dalbera, Jean-Philippe. 2007. Linguistic Atlases – Objectives, Methods, Results, Prospects –. In *Corpus-Based Perspectives in Linguistics*, 39-54. Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Edmont, Edmond & Jules Gilliéron. 1902. *Atlas linguistique de la France*. Paris : Champion.
- Fleuriot, Léon. 1980. *Les origines de la Bretagne. L'émigration* (Bibliothèque Historique Payot). Paris : Payot.
- Goebel, Hans. 1984. *Dialektometrische Studien : Anband italaromanischer, rutoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Vol. 3. Tübingen : Max Niemeyer Verlag.
- Goebel, Hans. 2010. Dialectometry and quantitative mapping. In Alfred Lameli, Roland Kehrein & Stefan Rababus (eds.), *Language and Space. An International Handbook of Linguistic Variation*, 433-457. Berlin, Boston : De Gruyter Mouton. <https://doi.org/10.1515/9783110219166.1.433>.
- Le Dù, Jean. 1997. La disparition du groupe des atlas et l'avenir de la géographie linguistique. *Le français moderne* LXV(1). 6-12.

- Le Dù, Jean. 2001. *Nouvel Atlas Linguistique de la Basse-Bretagne*. Brest : CRBC.
- Le Dù, Jean. 2006. A quoi servent les atlas linguistiques ? In *A l'arpenteur inspiré : mélanges offerts à Jean Barnabé*, 79-87. Matoury (Guyane) : Ibis Rouge.
- Le Dù, Jean. 2008. Les atlas linguistiques : une fenêtre sur le passé des langues. In Gianmario Raimondi & Luisa Revelli (eds.), *La dialectologie aujourd'hui. Atti del convegno internazionale "Dove va la dialettologia ?" Saint-Vincent, Aosta, Cogne, 21-24 settembre 2006*. Alessandria : Edizioni dell'Orso.
- Le Dù, Jean. 2013. Après la collecte, l'interprétation. Que faire des atlas linguistiques ? In *Mémoires de terrain. Enquêtes, matériaux, traitements des données. Actes du colloque de Lyon. Centre d'Etudes Linguistiques, université de Lyon 12 et 13 mars 2009*. (Publications du Centre d'Etudes Linguistiques [Linguistique, dialectologie] Nouvelle série), 7-25. Lyon : université Lyon 3 Jean Moulin.
- Nerbonne, John. 2010. Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B : Biological Sciences* 365(1559). 3821-3828. <https://doi.org/10.1098/rstb.2010.0048>.
- Nerbonne, John & Wilbert Heeringa. 2010. Measuring dialect differences. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and Space. An International Handbook of Language Variation* (Handbücher Zur Sprach- Und Kommunikationswissenschaft 30), vol. 1. Theories and Methods, 550-567. Berlin : De Gruyter Mouton.
- Pellen, Nadine & Tanguy Sollic. accepté. Dialectometry and population genetics – when results converge : the case of Western Brittany.
- Séguy, Jean. 1973. La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane* 37(145-146). 1-24.
- Sollic, Tanguy. 2021. *Distance linguistique et dialectométrie, une application à la langue bretonne. Enjeux, méthodologie et interprétations*. Brest : université de Bretagne Occidentale Doctorat.
- Thun, Harald. 2000. Introduction à la table ronde. In Annick Englebert, Michel Pierrard, Laurence Rosier & Dan van Raemdonck (eds.), *Actes du XXIIe Congrès International de Linguistique et de Philologie Romanes – Bruxelles, 23-29 juillet 1998*, vol. III Vivacité et diversité de la variation linguistique, 404-410. Berlin, New York : Walter de Gruyter.
- Wieling, Martijn & John Nerbonne. 2015. Advances in Dialectometry. *Annual Review of Linguistics* 1(1). 243-264. <https://doi.org/10.1146/annurev-linguist-030514-124930>.

Dialogue interdisciplinaire en humanités numériques : le cas de l'évolution des formes discursives en critique d'art

Orélie Desfriches Doria¹, Josquin Debaz², Gérald Kembellec^{3, 4}, Marie Gispert⁵

1 : EA 349 Laboratoire Paragraphe, université Paris 8 Vincennes – Saint-Denis

2 : Groupe de sociologie pragmatique et réflexive (GSPR), École des hautes études en sciences sociales (EHESS)

3 : EA 7339 Laboratoire Dispositifs d'information et de communication à l'ère numérique (DICEN), Conservatoire national des arts et métiers, université Gustave-Eiffel, université Paris-Nanterre

4 : Institut historique allemand de Paris, MaxWeber Stiftung

5 : Histoire culturelle et sociale de l'art, université Panthéon-Sorbonne

Le mouvement des humanités numérique porte des implications bien plus profondes que ne l'entend une acception restreinte (Cormerais *et al.* 2016), comme une pratique centrée sur la numérisation de corpus, l'encodage, la fouille-extraction de données. Il ne s'arrête pas à une « technologie » dans laquelle les humanités travaillent des corpora déjà constitués avec des outils numériques, afin d'en renouveler la compréhension. Dans une première partie, nous proposerons un cadre contextuel déployant ces implications, en lien avec le travail interprétatif, les interactions intra- et interdisciplinaires, et même une dimension humaniste, dans la constitution de sujets « numériquement » émancipés. La deuxième partie de notre communication visera la mise en pratique de ce cadre avec la présentation d'un projet de collaboration interdisciplinaire, en cours de construction entre des chercheuses et chercheurs en SIC, Histoire de l'art, Épistémologie et Sociologie.

La « technologisation » des humanités trame et draine des représentations liées aux technologies, des peurs, des énergies, des espoirs liés aux algorithmes, à l'Intelligence Artificielle, aux datas... et donc des dimensions affectives, dimension qui rappelle les propositions du manifeste des Humanités Numérique 2.0 (Schnapp, 2008). Selon ce texte, les H.N. se reconnaissent à leur caractère qualitatif, au-delà de traitement quantitatifs de données numériques, elles impliquent un travail interprétatif, qui pourra se traduire dans les pratiques, relevant par exemple du close reading (Moretti, 2013 ; Jänicke et al, 2015). Close et distant reading permettent un questionnement fécond : parmi ces pratiques, lesquelles sont à même d'assurer une pratique herméneutique ? Quelle implication de l'expérientiel, soit selon notre proposition, l'engagement des chercheurs et chercheuses comme « sujets socio-politiques » dont les appartenances disciplinaires composent, à travers leurs dimensions épistémologiques et méthodologiques, des formes d'existence socio-politiques, qui embarquent des formes réifiées des questionnements et des modes de raisonnements ?

Cette dimension des interactions disciplinaires et sociales rejoint l'approche proposée par Citton (2015), sur la strate 2.0 des humanités. S'y trament de nouvelles formes de connexions, d'écoutes, d'échanges et de collaborations entre des mondes qui jusqu'ici fonctionnaient en silos (scientifiques, artistes et activistes, selon Citton) : « Les propriétés connectives du numérique [sont à même de] pluraliser et redynamiser les interprétations créatives » dans l'optique de reconfigurer les lignes de démarcation sédimentées dans le partage de pouvoirs et de savoirs.

Dans une acception la plus large, les humanités numériques aboutissent à poser la question de l'humanisme numérique. Ainsi, comment les humanités numériques peuvent-elles accompagner la construction de la subjectivation à l'ère numérique, autrement dit, la subjectivation computationnelle, dans les termes d'Yves Citton ? Questionner la pratique computationnelle (collecte de données, algorithmes, prédictibilité des comportements) et des formes de littérature numérique (Millerand, 2003 ; Simonnot, 2009 ; Le Deuff, 2008) dans une perspective de recherche d'émancipation intellectuelle des individus revient à questionner la gouvernamentalité algorithmique (Rouvroy et Berns, 2013). Dans la mesure où l'émancipation vise à échapper aux déterminismes sociaux, qui traversent tous les individus, et considérant par ailleurs, que la reproduction de ces déterminismes est soutenue par le computationnel, renforcée par les algorithmes, basés sur l'exploitation des données numériques, le projet humaniste d'encourager et soutenir la subjectivation et

l'éducation à l'autonomie de pensée, entre en tension avec le déploiement massif des technologies numériques.

C'est au prisme de ces considérations, sur les risques des déterminismes computationnels sur nos subjectivités de chercheurs, que nous entendons positionner notre projet interdisciplinaire, en vue de permettre une (ré) appropriation des connaissances qui sont et seront produites et dans le but d'éviter de contribuer à la génération de nouvelles « victimes computationnelles » (une certaine forme de « digitaux naïfs »).

Dans le cadre d'approches d'analyses qualitatives de « grands » corpus de textes, existantes et mises en œuvre à travers le logiciel Prospéro (lui-même initialement développé dans les années 90 par des sociologues et des informaticiens, pour des analyses à « long » terme, a posteriori ou au fur et à mesure), un rapprochement interdisciplinaire s'opère actuellement entre des chercheuses et chercheurs en SIC ou histoire de l'art et ces travaux de socio-informatique. L'objet du travail qui sera présenté consistera à questionner de manière réflexive ces rapprochements disciplinaires, afin de positionner ces futures recherches vis-à-vis des travaux existants en lien avec les questions d'ordre épistémologiques dans l'usage des technologies textuelles (Carbou, 2017a et b) appliquées au champ de l'histoire de l'art.

Le projet impliquera un usage du logiciel Prospéro dans une approche historique de l'art à propos de l'évolution de la rubrique de critique artistique de la revue *Mercure de France*. À travers l'application des méthodes computationnelles et qualitatives mises en œuvre en interaction avec cet outil, et selon une démarche itérative d'exploration du corpus, permettant des allers-retours constants du texte à son interprétation, de l'analyse à l'enrichissement des grilles conceptuelles de traitement, de la lecture aux opérations de calcul, nous espérons révéler, dans cette étude interdisciplinaire, à travers l'évolution des formes discursives présentes dans cette revue, des pistes de recherche sur les évolutions sociologiques du monde de la critique d'art entre 1890 et 1913.

Dans cette perspective de collaboration interdisciplinaire nous viserons à clarifier d'une part les particularités de notre approche en regard des autres technologies textuelles mobilisables, et d'autres part, les positionnements épistémologiques et leurs éventuelles conflictualités. Enfin, les questionnements méthodologiques émergents au fil de la construction de ce travail interdisciplinaire seront présentés.

Bibliographie

Carbou, G. 2017a. Quelques questions à l'attention des utilisateurs de statistique textuelle pour l'analyse des discours. *Texte ! Textes et cultures*, 22(4).

Carbou, G. 2017b. Analyser les textes à l'ère des humanités numériques : Quelques questions pour l'analyse statistique des données textuelles. *Les Cahiers du numérique*, vol. 13(3-4), 91-114.

Citton, Y. 2015. « Humanités numériques. Une médiapolitique des savoirs encore à inventer », *Multitudes* 2015/2 (n° 59), p. 169-180

Cormerais F., Le Deuff O, Lakel A., Pucheu D. 2016. « Les SIC à l'épreuve du digital et des Humanités : des origines, des concepts, des méthodes et des outils », *Revue française des sciences de l'information et de la communication* [Online], 8 | 2016, Online since 23 March 2016, URL : <http://journals.openedition.org/rfsic/1820>

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. 2015. On Close and Distant Reading in Digital Humanities : A Survey and Future Challenges. In *EuroVis (STARs)* (p. 83-103).

Le Deuff, O. 2008. La culture de l'information : Quelles « littératies » pour quelles conceptions de l'information ? *VI.ème Colloque international du chapitre français de l'ISKO*, 7 et 8 juin 2007, à Toulouse, IUT de l'université Paul Sabatier., Juin 2008, France. p. 97-116, 2008.

Millerand, F. 2003. *L'appropriation du courrier électronique en tant que technologie cognitive chez les enseignants chercheurs universitaires : vers l'émergence d'une culture numérique ?*. Thèse de Doctorat, université de Montréal.

Moretti, F. 2013. *Distant reading*. Verso Books.

Rouvroy, A. & Berns, T. 2013. Gouvernamentalité algorithmique et perspectives d'émancipation : Le disparate comme condition d'individuation par la relation ? *Réseaux*, 177(1), 163-196.
<https://doi.org/10.3917/res.177.0163>

Schnapp J. 2008, *Digital Humanities Manifesto 2.0*. www.humanitiesblast.com/manifesto/Manifesto_V2.pdf

Du « Chantre tout divin » au « Ménétrier déconfit » : exploiter les sources anciennes d'un mythe pour mieux détecter les correspondances entre ses réécritures

Karolina Suchecka¹, Nathalie Gasiglia²

1 : EA 1061 Analyses littéraires et histoire de la langue (ALITHILA), université Lille III

2 : UMR 8163 Savoirs, textes, langage (STL), CNRS, université Lille 3

Dans l'édition numérique comparative de réécritures du mythe d'Orphée et Eurydice que nous concevons, nous visons à étudier et à représenter l'intertextualité par le biais d'un réseau des correspondances détectées entre les œuvres. Pour ce faire, nous cherchons à prouver la pertinence qu'il y a à exploiter deux logiciels de détection des réutilisations, Tracer[1] et TextPAIR[2], afin de comparer leurs produits pour des œuvres, du XV^e au XXI^e siècle, de contenu semblable, mais qui sont très éloignées structurellement et lexicalement.

L'idée de pouvoir détecter automatiquement des relations intertextuelles est passionnante, tant du point de vue de la recherche littéraire et linguistique, que de l'édition numérique. Cependant, si les logiciels employés sont très performants pour les correspondances proches, grâce à des techniques informatiques de l'intelligence artificielle, comme les réseaux neuronaux ou la vectorisation, même les technologies les plus récentes ne permettent pas la détection optimale des réutilisations et évocations plus complexes. Or, notre corpus regroupe plus de 70 réécritures françaises de genres et d'époques différents, et dont la proximité au mythe est très variable : certaines sont assez proches dans des transpositions pour le théâtre ou l'opéra, alors que d'autres sont très implicites, notamment dans les modernisations. Le rapport au mythe n'est pas non plus le même. Les parodies, par exemple, jouent sur la ridiculisation de la musique d'Orphée[3] ou de son amour pour Eurydice[4]. Les modernisations, quant à elles, présentent des amalgames, des interventions de nouveaux personnages ou des décors modernes qui participent à l'orientation du mythe vers un sens nouveau (Brunel 1997). Face à cette diversité, il serait optimiste d'espérer que les outils de détection automatique des réutilisations soient capables de repérer les relations directes pertinentes entre des réécritures aussi différentes que des modernisations contemporaines et des travestissements de la Renaissance.

Fort heureusement, un socle commun relie ces différentes œuvres. Parmi une trentaine de mentions du mythe recensées dans les littératures grecque et romaine antiques (Béague *et al.* 2019), deux œuvres, *Les Géorgiques* de Virgile (37-30 av. J.-C.) et *Les Métamorphoses* d'Ovide (1^{er} s. après J.-C.), sont généralement considérées comme reprises le plus fréquemment dans les réécritures postérieures. Durant notre communication, nous nous attacherons à démontrer qu'en intégrant à notre corpus les séries traductives françaises de ces deux sources anciennes, chacune comptant une vingtaine de traductions publiées depuis le XV^e s., en vers ou en prose, certaines en ancien français, d'autres modernisées et adaptées au jeune public, nous améliorons nos résultats de manière importante.

Tout d'abord, ces traductions nous permettent d'opérer des traitements linguistiques fins comme l'annotation morpho-syntaxique qui lisse les différences entre états de langue, la détection de mots-clés saillants pour chaque épisode du mythe, ou encore l'exploration des périphrases des entités nommées. En effet, nous exploitons les séries traductives pour optimiser quantitativement et qualitativement les résultats par l'élaboration et l'enrichissement de lexiques adaptés à la spécificité du corpus que nous couplons à des dictionnaires de langue générale (notamment de synonymes) et à des ressources polyvalentes (utilisées pour le repérage de chaînes de références). Nous pouvons alors nous focaliser sur les facteurs communs balisés et sur les unités linguistiquement pertinentes, comme les entités nommées et leurs périphrases (« Hyménée »/« Hymen »/« dieu de noçoiement »), la synonymie (« elle fut replongé de nouveau au même lieu »/« elle roule au lieu où elle était avant »), l'hyperonymie (« il fut présent à la vérité »/« il est là »), ou

encore la métaphorisation (« elle ne put s'animer bien qu'il l'agite »/« le dieu qui l'agite ne peut ranimer ses mourantes clartés »). Enfin, grâce aux relations détectées entre des traductions et des réécritures, nous sommes en mesure de construire un réseau de correspondances croisées (où un extrait du corpus correspond à plusieurs extraits d'autres textes) et, ainsi, nous lions indirectement des réécritures entre elles (grâce à un réemploi détecté au sein de chacune d'un même extrait de la traduction). Pour ce type de traitement, nous montrons enfin l'intérêt d'une visualisation des données à l'aide des graphes, qui s'avèrent particulièrement utiles pour révéler les relations intertextuelles intéressantes à analyser[5].

[1] Cf. <https://www.etrapp.eu/research/tracer/> (eTRAP, Georg-August-Universität de Göttingen) et (Büchler *et al.* 2012, 2013 ; Franzini *et al.* 2014).

[2] Cf. <https://artfl-project.uchicago.edu/text-pair> (ARTFL Project, université de Chicago) et (Horton, Olsen, et Roe 2010 ; Allen et Cooney 2010 ; Abdul-Rahman *et al.* 2016).

[3] « et vous, mon tendre ami, à qui jetez-vous ce chant passionné de votre... crin-crin ? » (Offenbach et Crémieux 1936).

[4] « pendant sept mois, il vécut dans les bois, dans les antres, jetant aux échos ses gémissements, ses grincements de dents, oubliant probablement de peigner sa barbe et ses cheveux » (Eimann 1854).

[5] Différentes types de graphes ont déjà été exploités pour présenter les relations intertextuelles, cf. notamment : (Jänicke *et al.* 2014 ; Ganascia 2019 ; Suchecka, Gasiglia, et Zieger 2019 ; Del Lungo et Suchecka [À paraître]).

Abdul-Rahman, Alfie, Glenn Roe, Mark Olsen, Clovis Gladstone, Robert Morrissey, Nicholas Cronk et Min Chen. 2016. « Constructive Visual Analytics for Text Similarity Detection ». *Computer Graphics Forum* 36 (1), 237-248. <https://doi.org/doi:10.1111/cgf.12798>.

Allen, Timothy et Charles Cooney. 2010. « Plundering Philosophers : Identifying Sources of the Encyclopédie ». *Journal of the Association for History and Computing*. <http://hdl.handle.net/2027/spo.3310410.0013.107>.

Béague, Annick, Jacques Boulogne, Alain Deremetz et Françoise Toulze. 2019. *Les Visages d'Orphée. Les Visages d'Orphée. Savoirs Mieux*. Villeneuve d'Ascq : Presses universitaires du Septentrion. <http://books.openedition.org/septentrion/50854>.

Brunel, Pierre. 1997. « Orphée moderne ». Dans *Apollinaire entre deux mondes. Le contrepoint mythique dans Alcools. Mythocritique II*, 63-82. Écriture. Paris : Presses Universitaires de France.

Büchler, Marco, Gregory Crane, Maria Moritz et Alison Babeu. 2012. « Increasing recall for text re-use in historical documents to support research in humanities ». Dans *Theory and Practice of Digital Libraries*, édité par P. Zaphiris, G. Buchanan, E. Rasmussen et F. Loizides, 95-100. Berlin : Springer Berlin Heidelberg. 10.1007/978-3-642-33290-6_11.

Büchler, Marco. 2013. *Informationstechnische Aspekte des Historical Text Re-use*. Thèse de doctorat dirigée par G. Heyer et K. Schulz. Université de Leipzig. <https://ul.qucosa.de/api/qucosa%3A11877/attachment/ATT-0/>.

Del Lungo, Andrea et Karolina Suchecka. [À paraître]. « Projet eBalzac : construire une bibliothèque hypertextuelle des sources intertextuelles ». Dans *Actes du colloque DHNord 2019 « Corpus et archives numériques »*. Villeneuve d'Ascq : Presses universitaires du Septentrion.

Eimann, D.-L. 1854. « Une Histoire renouvelée des Grecs ». Dans *Histoires cavalières*, 23-35. Paris : Charpentier.

Franzini, Greta, Emily Franzini, Marco Büchler, Martin Mueller et Philip Burns. 2014. « Towards a Historical Text Re-use Detection ». Dans *Text Mining. From Ontology Learning to Automated Text Precession Applications*, édité par C. Biemann et A. Mehler, 221-238. Suisse : Springer International Publishing. https://doi.org/10.1007/978-3-319-12655-5_11.

- Ganascia, Jean-Gabriel. 2019. « Graphes et intertextualité ». Dans *Humanités numériques*. Centre Universitaire Méditerranéen, Nice.
- Horton, Russell, Mark Olsen et Glenn Roe. 2010. « Something Borrowed : Sequence Alignment and the Identification of Similar Passages in Large Text Collections ». *Digital Studies/Le Champ numérique* 2 (1). <https://doi.org/10.16995/DSCN.258>.
- Jänicke, Stefan, Annette Geßner, Marco Büchler et Gerik Scheuermann. 2014. « Visualizations for Text Re-Use ». Dans *Proceedings of the 5th International Conference on Information Visualization Theory and Applications*, 59-70.
- Offenbach, Jacques et Hector Crémieux. 1936. *Orphée aux Enfers*. Paris : Calmann-Lévy.
- Suchecka, Karolina, Nathalie Gasiglia et Karl Zieger. 2019. « Édition comparative intermédiaire de séries traductives : exploiter les homologies pour créer des visualisations modulables ». *Traitement Automatique des Langues, TAL et humanités numériques*, 60 (3), 37-62.
https://www.atala.org/sites/default/files/TAL-60-3_SucheckaEtAl_EditionComparativeIntermedialeDeSeriesTraductives-final.pdf.

Exploration du théâtre alsacien à travers ses listes de personnages pendant la période 1870-1940

Pablo Ruiz Fabo¹, Carole Werner¹

1 : UR 1339 LiL.Pa Linguistique, langues, parole, université de Strasbourg

Le projet MeThAL [1] vise à créer un corpus large de théâtre dialectal en alsacien, encodé en TEI dans le respect de principes FAIR (Wilkinson *et al.* 2016), pour faciliter des recherches quantitatives en sociolinguistique historique de l'Alsace et en analyse dramatique. Le corpus s'étend de 1870 à 1940. Nous partons de 150 pièces numérisées comme fichiers d'image par la Bnu (portail Numistral) [2].

Un défi : Comment arriver à un aperçu de la collection pendant que l'océrisation et l'encodage TEI sont en cours [3] ? Notre approche a consisté à transcrire les listes de personnages de toutes les pièces et de les enrichir avec des métadonnées décrivant la situation sociale des personnages : leur sexe, profession et âge si disponibles, et une estimation de leur classe sociale. La base de données ainsi créée inclut des annotations pour 109 pièces et 1091 personnages, dont le *dramatis personæ* identifie la profession pour 611 personnages (240 activités professionnelles uniques) [4]. Les métadonnées bibliographiques (date et lieu de publication etc.) sont elles aussi disponibles.

Quel intérêt présentent ces données pour l'exploration du corpus ? La pertinence des listes de personnages comme indice des groupes socio-politiques d'une pièce a été récemment rappelée par Wiedmer *et al.* (2020). Les caractéristiques des personnages conditionnent les conflits et intrigues possibles et le *dramatis personæ* fournit une trame tant sociobiographique que dramatique (cf. Asmuth 2016, 97-98 ; Pfister 2001, 220). La variation lexicale et phonologique représentée dans le texte attribué aux personnages peut de son côté être reliée à des variables sociales ou relatives à l'intrigue.

Plusieurs travaux, dont Huck (2005) et von Hülsen (2003) analysent les intrigues et caractéristiques linguistiques des pièces représentées au Théâtre Alsacien de Strasbourg (TAS) entre 1898 et 1914. Selon von Hülsen (2003, 100-104) l'opposition entre personnages alsaciens et représentants de l'autorité allemande, qui remplace ou s'ajoute à l'opposition entre sexes ou générations typique de certains sous-genres comédiques, est l'un des moteurs du conflit dramatique. S'ajoute à ces schémas récurrents l'opposition parents-enfants ; les parents cherchant l'ascension sociale par le mariage de leurs enfants, ces derniers préférant le mariage d'amour. Huck (2005, 206-210) analyse dans quelle mesure les stratégies de (re)présentation du parler dialectal dans les pièces pourraient être réalistes par rapport à la situation sociale des personnages. Il souligne par exemple que Stoskopf avait probablement connaissance des pratiques langagières de la petite-bourgeoisie dont sont souvent issus ses personnages. L'importance de la petite et moyenne bourgeoisie dans les pièces est aussi abordée par von Hülsen (2003, 106, 140) [5].

Les travaux cités analysent autour de 20 pièces, parmi les plus importantes du TAS. Notre hypothèse est que nos métadonnées concernant les personnages sont utiles pour examiner jusqu'à quel point les tendances repérées dans ces pièces se reproduisent dans un corpus plus large, varié en époque et origine : ces métadonnées permettent de cibler des pièces qui fournissent une combinaison de personnages avec des caractéristiques sociales pertinentes, p. ex. des pièces qui mettent en scène divers membres de la (petite et moyenne) bourgeoisie avec plus ou moins de pouvoir économique (artisans, commerçants, professions libérales et manufacturiers), en interaction entre eux ou avec d'autres groupes (ouvriers, domestiques) [6].

Un autre intérêt des métadonnées élaborées est qu'elles aident à décrire la composition sociale des pièces et son évolution. Un exemple concerne l'évolution des rôles des personnages féminins. Le *dramatis personæ* les présente généralement en relation avec les personnages masculins (*épouse de, fille de*) sans mentionner d'activité professionnelle. Quand celle-ci est présente, elle correspond dans la plupart des cas au domaine de la domesticité. Or, des différences selon la décennie sont constatées : les rôles féminins avec une activité

autre que « aide ménagère » sont absents dans le corpus avant 1900, mais couvrent 18 % des cas (5 sur 28) dans la décennie de 1900, avec des rôles comme « sage-femme », « actrice » ou « étudiante de philosophie ». Dans la décennie de 1930, nous trouvons des rôles féminins avec des activités comme « dactylographe » ou « commerçante de tissus ». Il reste à déterminer si ces chiffres sont dus à des biais dans la sélection du corpus ou s'ils reflètent une vraie évolution dans les contenus typiques représentés dans chaque décennie.

Quels sont les points à améliorer dans nos métadonnées de personnages ? D'abord, une classification de professions selon des groupes pertinents pour l'analyse dramatique est en cours, ce qui présente des défis. L'application d'une classification existante pour le théâtre classique français (Fièvre 2017) ne serait pas pertinente ; les catégories sont obsolètes pour la période du corpus. L'adaptabilité au corpus des classifications professionnelles standard actuelles comme la CITEP-08 ou ESCO [7] est limitée. Il existe des classifications d'activités professionnelles historiques, comme HISCO (van Leuwen *et al.* 2002) ou l'ontologie de Moeller et Nasarek (2018) pour l'espace germanophone, mais leur objectif n'étant pas l'analyse de personnages littéraires, leur application au corpus n'est pas immédiate. Hormis l'ajout d'une classification des professions, il serait pertinent d'annoter l'origine (alsacienne, allemande, française) des personnages, en suivant l'idée de von Hülsen (2003, 100) que les différences entre les alsaciens et l'autorité allemande sont source de conflit dramatique. Enfin, une autre amélioration serait l'intégration des métadonnées au corpus en format TEI. Une approche sur la base de *feature structures* (cf. Galleron, 2017) pourrait être utilisée, comme nous l'avons décrit dans Ruiz Fabo *et al.* (2020).

En résumé, les métadonnées de personnages peuvent fournir un aperçu utile d'une collection dramatique et aider à trouver des textes pertinents pour aborder une question de recherche. La création de ces métadonnées sur la base du *dramatis personae* demande peu de ressources et peut constituer une première étape utile dans un effort d'encodage de corpus. Nous travaillons à l'amélioration de nos annotations par la création d'une classification de professions pertinente pour nos analyses et prévoyons la formalisation de celle-ci en TEI.

Notes

[1] MeThAL (« Vers une Macroanalyse du THéâtre en ALSacien ») : <https://methal.pages.unistra.fr>

[2] Bnu : Bibliothèque nationale et universitaire de Strasbourg. Numistral : <https://numistral.fr/fr>

[3] Les pièces encodées sont disponibles sur <https://git.unistra.fr/methal/methal-sources> ; le dépôt est mis à jour à mesure que l'encodage progresse.

[4] Notre transcription couvre les genres comédiques et les drames de la collection Numistral ; les contes de Noël ne sont pas couverts pour le moment.

[5] Gall (1974) analyse un corpus antérieur aux pièces examinées par les travaux suscités, qui comporte les pièces les plus importantes du théâtre dialectal en alsacien au cours du XIX^e siècle. Gall observe déjà (1974, 65) l'importance des différents groupes de la bourgeoisie urbaine dans ces pièces.

[6] Pour les pièces déjà encodées en TEI, il est possible de cibler des répliques dont les personnages ont les caractéristiques voulues.

[7] La CITEP est la *Classification internationale type des professions* de l'Organisation internationale du travail : <https://www.ilo.org/public/french/bureau/stat/isco/isco08/index.htm>. ESCO (*Classification européenne des aptitudes/compétences, certifications et professions*) a été créée par la Commission européenne : <https://ec.europa.eu/esco/portal/occupation>

Références

Asmuth, Bernhard. 2016. *Einführung in die Dramenanalyse*. 8., aktualisierte und erweiterte Auflage. Sammlung Metzler, Band 188. Stuttgart : J.B. Metzler Verlag.

Fièvre, Paul. 2017. « Rubrique concernant les personnages ». Théâtre Classique. <http://www.theatre-classique.fr/pages/PagePersonnages.html>.

Gall, Jean-Marie. 1974. *Le théâtre populaire alsacien au XIX^e siècle*. Strasbourg : Istra.

- Galleron, Ioana. 2017. « Conceptualisation of Theatrical Characters in the Digital Paradigm : Needs, Problems and Foreseen Solutions ». *Human and Social Studies* 6 (1) : 88-108. <https://doi.org/10.1515/hssr-2017-0007>.
- Huck, Dominique. 2005. « Le “Théâtre Alsacien de Strasbourg” et la production dramaturgique de ses fondateurs (1898-1914) ». Dans *Culture et histoire des spectacles en Alsace et en Lorraine : De l'annexion à la décentralisation (1871-1946)*, par Jean-Marc Leveratto, Jeanne Benay, Olivier Thomas et Séverine Wuttke, 197-222. Peter Lang.
- Hülßen, Bernhard von. 2003. *Szenenwechsel im Elsass : Theater und Gesellschaft in Strassburg zwischen Deutschland und Frankreich 1890-1944*. Leipziger Universitätsverlag.
- Leeuwen, Marco H. D. van, Ineke Maas et Andrew Miles. 2002. *HISCO Historical International Standard Classification of Occupations*. Leuven, Belgium : Leuven University Press.
- Moeller, Katrin et Robert Nasarek. 2018. « Die Ontologie historischer deutschsprachiger Berufs- und Amtsbezeichnungen. Interoperationalität und Berufsklassifizierung durch semantisches Topic Modeling ». Dans *DHd 2018. Kritik der Digitalen Vernunft. Konferenzabstracts*, 173-177. Köln : Universität zu Köln.
- Pfister, Manfred. 2001. *Das Drama : Theorie und Analyse*. 11^e éd. Information und Synthese ; Bd. 3. München : W. Fink.
- Ruiz Fabo, Pablo, Delphine Bernhard et Carole Werner. 2020. « Création d'un corpus FAIR de théâtre en alsacien et normalisation de variétés non-contemporaines ». Dans *2èmes journées scientifiques du Groupement de Recherche Linguistique Informatique Formelle et de Terrain (LIFT)*, édité par Thierry Poibeau, Yannick Parmentier et Emmanuel Schang, 34-43. Montrouge, France : CNRS. <https://doi.org/10.5281/zenodo.4323301>.
- Wiedmer, Nathalie, Janis Pagel et Nils Reiter. 2020. « Romeo, Freund des Merkurio : Semi-Automatische Extraktion von Beziehungen zwischen dramatischen Figuren ». Dans *DHd2020 Spielräume. Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts.*, 194-200. Paderborn : DHd. Digital Humanities im deutschsprachigen Raum e.V. <https://doi.org/10.5281/zenodo.3666689>
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg *et al.* 2016. « The FAIR Guiding Principles for Scientific Data Management and Stewardship ». *Scientific Data* 3 (1) : 160018. <https://doi.org/10.1038/sdata.2016.18>.

FAIRiser des données : état des lieux, barrières et choix. Une réflexion à partir des données des corpus d'auteurs

Ioana Galleron¹, Fatiha Idmhand²

1 : UMR 8094 LATTICE, université Sorbonne nouvelle – Paris 3, CNRS

2 : CRLA-Archivos, UMR 8132 ITEM

Depuis leur publication en 2016 dans *Scientific Data* (1), les principes FAIR ont été présentés comme des « lignes directrices » destinées à faciliter l'accessibilité et la réutilisation des données et ressources scientifiques numériques. En tant que « principes », ils orientent, suggèrent et encouragent leur mise en œuvre en vue d'une recherche plus ouverte, plus partagée et transparente. Toutefois, depuis que certains États, organismes (2) et la Commission Européenne exigent des projets financés de les suivre de façon impérative, leur application est devenue un enjeu financier et technique qui élude trop souvent la nécessaire réflexion scientifique qui doit accompagner la circulation des objets d'études en milieu numérique.

Du point de vue technique, respecter l'ensemble des principes requiert des efforts conséquents et onéreux que seules de très grandes infrastructures (comme Huma-Num en France) peuvent assumer. Ainsi, l'existence d'un écosystème numérique est essentielle pour stocker les données et leurs métadonnées, pour leur délivrer un identifiant pérenne unique (ex. DOI ou URL) de sorte qu'elles soient faciles à trouver (F), pour accéder à long terme aux données et les télécharger (A), pour pouvoir les utiliser et les croiser avec d'autres jeux de données (I) et, enfin, pour les réutiliser dans de nouvelles recherches scientifiques (R).

Si, à première vue, les principes FAIR semblent dessiner un horizon technique à destination des systèmes informatiques d'abord, et des humains ensuite, créer de nouvelles connaissances à partir de ces données nécessite de réfléchir à la façon dont les pratiques actuelles des humanistes numériques encouragent ou, à l'inverse, bloquent, la recherche à partir de leurs données et métadonnées. Dans un travail antérieur, nous avons engagé la réflexion en interrogeant la notion de « Réutilisabilité » (R) dans l'édition électronique (3), au terme de nos observations, nous avons suggéré que pour un même projet, différents états du fichier numérique soient proposés de façon à faciliter la réutilisation de celui-ci. Cette contribution propose de prolonger la réflexion initiée sur la « Réutilisabilité » et de questionner les autres dimensions des principes FAIR : pourquoi « facile à trouver » et interface (un site web le plus souvent) peuvent entrer en contradiction ? quels sont, dans le contexte des Humanités numériques, les freins à l'accessibilité ? enfin, on se demandera pourquoi l'interopérabilité n'est pas qu'une question de langage et de technique, mais un problème de positionnement de la recherche scientifique sur les textes dans les nouvelles frontières des disciplines ?

Pour exemplifier notre propos, nous nous appuyons sur le dépouillement récemment réalisé des éditions et données produites par les membres du consortium CAHIER. Avec une soixantaine de projets, plus de 327000 fichiers texte et près de 500.000 images, CAHIER représente un terrain d'étude éclairant au sujet des obstacles et difficultés qui jalonnent le chemin de la FAIRisation dans une communauté ayant développé antérieurement des pratiques et des standards autour de l'édition papier (4). L'acculturation nécessaire pour passer d'un régime de publication à un autre, ainsi que l'accompagnement technique mis en œuvre dans le cadre de CAHIER seront abordés tour à tour dans notre communication.

Bibliographie

(1) <https://www.nature.com/articles/sdata201618>

(2) <https://www.science-ouverte.cnrs.fr/>

(3) Ioana Galleron et Fatiha Idmhand, « De l'interopérabilité à la réutilisabilité des éditions électroniques », *Humanités numériques* [Online], 1 | 2020. URL : <http://journals.openedition.org/revuehn/350> ; DOI : <https://doi.org/10.4000/revuehn.350>

(4) <https://www.go-fair.org/fair-principles/>

IA et approches participatives dans les humanités numériques : de la conjonction à la coopération

Olivier Aubert, Guillaume Raschia¹, Benjamin Hervy²

1 : UMR 6241 Laboratoire d'informatique de Nantes-Atlantique (LINA), CNRS, université de Nantes, École nationale supérieure des mines – Nantes

2 : Maison des sciences de l'homme Ange-Guépin (MSH Nantes)

Le projet de recherche CIRESEFI (*Contrainte et Intégration : pour une réévaluation des spectacles forains et italiens sous l'Ancien Régime*) vise à apporter de nouveaux éclairages sur l'histoire des spectacles. Parmi les angles d'attaque utilisés, le projet RECITAL (*Registres Comptables de la Comédie Italienne*) (Rubellin et Raschi, 2020) développé au sein de CIRESEFI vise à exploiter la matière présente dans les registres comptables du théâtre de la Comédie Italienne, en se focalisant particulièrement sur le XVIII^e siècle.

Ainsi, les numérisations sous forme image de 64 registres comptables, couvrant (avec des manques) la période de 1717 à 1794, ont été acquises par le projet. L'enjeu de RECITAL est de procéder à l'extraction des informations présentes dans les 26000 pages des registres, afin de pouvoir les analyser et voir ce que nous enseigne la comptabilité du théâtre sur la structure sociale du public, la popularité du répertoire et des auteurs, les décors et la mise en scène (accessoires), etc.

Le corpus de 26000 pages présente des caractéristiques non homogènes : il est manuscrit, avec plus de 7 rédacteurs différents, donc autant de styles d'écriture. Par ailleurs, la langue utilisée a varié au cours des ans, entre le français et différents types de dialectes italiens. Enfin, la forme et les règles des bilans comptables ont également évolué au cours des années.

Deux approches parallèles ont été testées dès le début du projet : d'une part, une approche purement automatique, mobilisant des algorithmes d'Intelligence Artificielle, visant à extraire et transcrire automatiquement les informations. D'autre part, une approche plus manuelle de *crowdsourcing*, basée sur la mobilisation d'un large public au travers d'une interface participative.

L'approche automatique (A. Granet et al, 2018) s'est retrouvée confrontée aux caractéristiques difficiles du corpus : une forte hétérogénéité des données (scripteurs, langues, formats, etc) ainsi qu'à un manque de données d'entraînement pour l'apprentissage. Elle a permis de faire progresser l'état de l'art, mais n'a pas produit de données utilisables. L'approche de *crowdsourcing* a pris la forme de la plateforme <https://recital.univ-nantes.fr/> utilisant et adaptant le logiciel ScribeAPI. Cette plate-forme offre différentes fonctionnalités précieuses, dont la possibilité d'enchaîner les tâches de marquage, transcription et vérification, ainsi que sa capacité à adapter les formulaires de saisie des informations aux contenus.

La plate-forme permet ainsi de découper le travail en trois étapes : marquage (détermination et indexation des zones présentant des informations à transcrire), transcription (du contenu des zones identifiées dans l'étape de marquage) et enfin vérification (par présentation des contenus saisis).

L'approche participative a permis de générer plus de 115000 transcriptions, qu'il faut ensuite traiter pour homogénéiser les informations ainsi qu'en évaluer la qualité (B. Hervy et al, 2019). En effet, un écueil commun dans les approches participatives est l'évaluation de la qualité des résultats, dont les producteurs ont des compétences et des niveaux de motivation variés. Les données produites font donc l'objet d'une vérification participative au sein de la plate-forme, permettant d'aboutir à une première forme de consensus parmi les transcriptions. Leur évaluation peut ensuite prendre plusieurs formes. On trouve tout d'abord une première évaluation de la cohérence syntaxique et sémantique des données individuelles, tirant parti de la connaissance de leur nature (par exemple, la recette de la journée doit nécessairement être sous une forme numérique), via des scripts python dédiés. On peut également exploiter dans cette phase les informations issues de la phase participative de vérification, qui fournissent une indication de la qualité des données,

venant pondérer l'indice de confiance associé à la donnée. Étant donné la nature comptable des informations présentées, on peut également procéder, toujours via des scripts dédiés, à une phase de détection d'incohérences sur un ensemble de données entre elles (recettes mensuelles ne correspondant pas à la somme des recettes quotidiennes par exemple – ce qui peut provenir d'erreurs à la source ou lors de la transcription), mais cette étape nécessite une couverture complète d'un ensemble de données. Enfin, une validation manuelle par les experts du domaine est nécessaire pour évaluer la qualité des informations. Cette phase passe par la mise à disposition d'interfaces permettant une consultation rapide et synthétique de l'ensemble des informations produites. Chacun des traitements mentionnés ci-dessus fait l'objet d'une traçabilité dans le système d'information, de manière à pouvoir retracer la provenance des données nettoyées qui seront finalement présentées aux utilisateurs du système.

Vers une coopération IA/approche participative

Les approches automatiques et participatives ont été dans un premier temps considérées indépendamment. En effet, l'approche participative nécessite un certain temps pour mettre en place l'infrastructure puis récolter un nombre suffisant de données. Nous sommes à présent dans une situation différente, où l'approche participative a déjà permis de couvrir près de la moitié des pages du corpus. Nous nous attachons donc maintenant à étudier la manière dont on peut combiner les deux approches, en utilisant l'une pour valider les résultats de l'autre, ou de manière plus poussée en cherchant à obtenir une coopération plus fine. Nous avons donc repris l'ensemble de la chaîne de traitement de la plate-forme RECITAL pour en identifier les différentes phases et étudier pour chacune d'elles dans quelle mesure elle pourrait se prêter à une approche de coopération humain/algorithmique.

Nous avons identifié trois niveaux d'intégration pour cette coopération. Le premier niveau donne à l'IA le rôle d'*assistant à l'annotateur*, qui propose à l'utilisateur (humain) des informations pertinentes afin de l'assister dans sa tâche – mais c'est bien l'humain qui prend les décisions et saisit les informations. Le second niveau consiste en une *assistance à l'expert* chargé de qualifier et valider les informations, offrant des signaux provenant tant des données saisies que des traces d'activité des utilisateurs sur la plate-forme visant à qualifier la qualité des informations qu'ils produisent. Ces signaux permettent à l'expert d'évaluer les informations, voire d'en proposer des corrections. Enfin, un troisième niveau envisagé consiste à intégrer l'IA sous forme d'*agent autonome* dans le système, en tant qu'annotateur ou qu'expert, qui participerait alors aux côtés des acteurs humains à la production et la validation collaborative des informations, interagissant en autonomie comme un utilisateur standard avec la plate-forme RECITAL pour y saisir des données.

Nous avons commencé à mettre en oeuvre les deux premiers niveaux au sein de la plate-forme RECITAL, exploitant les données déjà saisies. Le troisième niveau requiert des investigations plus poussées que nous poursuivons, tant sur ses principes que sa faisabilité technique dans son intégration à la plate-forme.

Références

Granet, Adeline, Benjamin Hervy, Geoffrey Roman-Jimenez, Marouane Hachicha, Emmanuel Morin, Harold Mouchère, Solen Quiniou, Guillaume Raschia, Françoise Rubellin et Christian Viard-Gaudin, 2018. Crowdsourcing-based Annotation of the Accounting Registers of the Italian Comedy. LREC.

Hervy, Benjamin, Pierre PÉTILLON, Hugo PIGEON et Guillaume RASCHIA, 2019. Correction des données : retour d'expérience sur la plate-forme RECITAL de transcription participative. In About Variety in Humanities Big Data, Recherche d'information, document et web sémantique. Vol. 19, No. 1, ISTE OpenScience.

Rubellin, Françoise et Guillaume Raschia, 2020. Redécouvrir les Théâtres de la Foire et la Comédie-Italienne avec les bases THEAVILLE et RECITAL. Revue d'Historiographie du Théâtre. Numéro 5. Trimestre 1.

« Site RECITAL ». 2021. Université de Nantes. 4 janvier. <https://recital.univ-nantes.fr/>

« Projet CIRESEFI ». 2021. CETHEFI. 4 janvier. <http://cethefi.org/ciresfi/>

Image et texte : les livres d'heures manuscrits vus par l'intelligence artificielle

Dominique Stutzmann¹, Louis Chevalier¹

1 : UPR841 Institut de recherche et d'histoire des textes (IRHT), CNRS

Le projet ANR *HORAE – Hours Recognition, Analysis, Editions* étudie les livres d'heures médiévaux, livres de prière à l'usage des laïcs et le « best-seller » du Moyen Âge. Produits en grand nombre à partir du XIII^e siècle dans toute l'Europe occidentale et plus nombreux que les manuscrits bibliques, ils documentent le sentiment religieux, les pratiques sociales, la circulation des textes et les connexions liturgiques entre des régions différentes, mais aussi les processus d'industrialisation de la production livresque car ils sont largement standardisés (Stutzmann 2019).

L'étude des livres d'heures est complexe. En effet, ces livres sont nombreux et volumineux, compilés à partir de centaines de pièces élémentaires. Ils ont découragé les études systématiques. Or il importe de pouvoir examiner un très grand nombre de ces documents pour saisir la relation qui lie leurs différentes parties, le degré de fréquence ou de rareté des chants, des lectures, oraisons et prières, et comprendre la circulation de ces textes et l'individualisation d'un produit de masse.

Le programme de recherche réunit trois partenaires complémentaires, spécialisés en humanités, en TALN et en IA (particulièrement apprentissage machine) : l'Institut de Recherche et d'Histoire des textes (CNRS), le Laboratoire des Sciences du numérique de Nantes, et Teklia. Les défis technologiques, maintenant relevés, portent sur analyse d'image, tels que la lecture des écritures médiévales par ordinateur ou l'analyse de la mise en page (Boillet *et al.* 2019 ; Boros *et al.* 2019), l'identification de textes par « alignement » ou *text reuse identification* (Hazem *et al.* 2019) et établissement automatique de tables de matières hiérarchiques ou « segmentation » (Hazem *et al.* 2020 ; Daille *et al.* 2019). L'analyse historique affronte particulièrement les problèmes de masse de données, de granularité et de complexité des réseaux (Stutzmann *et al.* 2019).

S'inscrivant dans l'axe 2 de l'appel (effet des méthodes computationnelles en SHS), cette communication présentera les premiers résultats à grande échelle de l'analyse automatique. Pour la démonstration, en utilisant un sous-corpus de 50 manuscrits (11800 images, 248932 lignes de texte, 674 miniatures), nous traiterons d'abord de la méthodologie historique. Les questions et typologies usuelles de la codicologie quantitative, d'une part, et les méthodes plus récentes d'analyse des réseaux textuels sont renouvelées (Ornato 1997 ; É. Cottureau-Gabillet 2015 ; E. Cottureau-Gabillet 2016 ; Julien 2016 ; Riva 2019). L'élaboration des critères d'exploitation et d'interrogation historique peut prendre appui sur des données autrefois soit inaccessibles, soit formalisées à un autre niveau de granularité : la comparaison du luxe des manuscrits peut tenir compte du nombre de miniatures, initiales, éléments de décoration autant que de leur superficie réelle ; la comparaison des contenus textuels peut se fonder sur les structures et les pièces unitaires dans des collections de citations ; l'on peut aussi corréler ces domaines généralement séparés.

La communication discutera aussi la mise en œuvre concrète de l'apprentissage automatique et de l'intelligence artificielle. Parfois inutile face aux méthodes dites « traditionnelles » (temps de calcul, création de données directement utiles aux historiens), elle permet de croiser des critères sans corrélation préalablement connue. Elle pose aussi des problèmes heuristiques. Les *mid-level features* définis dans ce projet (sur ce concept, voir (Hassner *et al.* 2013)) aboutissent à une production massive de données largement brutes et « à plat », posant des difficultés d'interprétation (discrétisation d'un continuum de données et clustering, poids excessif des critères d'analyse sur les savoir créés).

L'axe 3 (transmission des patrimoines culturels) sera concerné secondairement, car la recherche se fonde sur les numérisations disponibles via le protocole IIIF, la base de données Heurist produite est hébergée

par Huma-Num et permet des valorisations vers le grand public : repérage automatique des miniatures, transcription et traduction de documents autrement hermétiques.

Boillet, Mélodie, Marie-Laurence Bonhomme, Dominique Stutzmann et Christopher Kermorvant. 2019. « HORAE : an annotated dataset of books of hours ». Dans the 5th International Workshop on Historical Document Imaging and Processing, 7-12. Sydney : ACM Press. <https://doi.org/10.1145/3352631.3352633>.

Boros, Emanuela, Alexis Toumi, Erwan Rouchet, Bastien Abadie, Dominique Stutzmann et Christopher Kermorvant. 2019. « Automatic Page Classification in a Large Collection of Manuscripts Based on the International Image Interoperability Framework ». Dans 2019 International Conference on Document Analysis and Recognition (ICDAR), 756-762. Sydney : IEEE. <https://doi.org/10.1109/ICDAR.2019.00126>.

Cottureau-Gabillet, Émilie. 2015. « Manuscrits de luxe et distinction sociale à la fin du Moyen Âge ». Dans Marquer la prééminence sociale, édité par Jean-Philippe Genet et E. Igor Mineo, 283-301. Histoire ancienne et médiévale. Paris : Éditions de la Sorbonne. <http://books.openedition.org/psorbonne/3352>.

Cottureau-Gabillet, Emilie. 2016. « Revealing Some Structures and Rules of Book Production (France, Fourteenth and Fifteenth Centuries) ». Dans Ruling the Script in the Middle Ages. Formal Aspects of Written Communication (Books, Charters, and Inscriptions), édité par Sébastien Barret, Dominique Stutzmann et Georg Vogeler, 129-163. Utrecht Studies in Medieval Literacy 35. Turnhout : Brepols.

Daille, Béatrice, Amir Hazem, Christopher Kermorvant, Martin Maarand, Marie-Laurence Bonhomme, Dominique Stutzmann, Jacob Currie et Christine Jacquin. 2019. « Handwritten text recognition and text segmentation adapted to manuscript books of hours ». Traitement Automatique des Langues, TAL et humanités numériques, 60 (3). ATALA : 13-36.

Hassner, Tal, Malte Rehbein, Peter A. Stokes et Lior Wolf. 2013. « Computation and Palaeography : Potentials and Limits ». Dagstuhl Manifestos 2 : 14-35. <http://dx.doi.org/doi:10.4230/DagMan.2.1.14>.

Hazem, Amir, Beatrice Daille, Christopher Kermorvant, Dominique Stutzmann, Marie-Laurence Bonhomme, Martin Maarand et Mélodie Boillet. 2020. « Books of Hours : the First Liturgical Corpus for Text Segmentation ». Dans Proceedings of The 12th Language Resources and Evaluation Conference, 776-784. Marseille, France : European Language Resources Association. <https://www.aclweb.org/anthology/2020.lrec-1.97>.

Hazem, Amir, Béatrice Daille, Dominique Stutzmann, Jacob Currie et Christine Jacquin. 2019. « Towards Automatic Variant Analysis of Ancient Devotional Texts ». Dans Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, 240-249. Firenze : Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-4730>.

Julien, Octave. 2016. « Délier, lire et relier ». Hypotheses 19 (1). Éditions de la Sorbonne : 211-224.

Ornato, Ezio. 1997. La face cachée du livre médiéval : l'histoire du livre vue par Ezio Ornato, ses amis et ses collègues. I libri di Viella 10. Roma : Viella.

Riva, Gustavo Fernandez. 2019. « Network Analysis of Medieval Manuscript Transmission ». Journal of Historical Network Research 3 (novembre) : 30-49. <https://doi.org/10.25517/jhnr.v3i1.61>.

Stutzmann, Dominique. 2019. « Résistance au changement ? Les écritures des livres d'heures dans l'espace français (1200-1600) ». Dans "Change" in medieval and Renaissance scripts and manuscripts. Proceedings of the 19th Colloquium of the Comité international de paléographie latine (Berlin, 16-18 September, 2015), édité par Eef Overgaauw et Martin J. Schubert, 97-116. Bibliologia 50. Turnhout : Brepols.

Stutzmann, Dominique, Jacob Currie, Béatrice Daille, Amir Hazem et Christopher Kermorvant. 2019. « Integrated DH. Rationale of the HORAE Research Project ». DH2019, Utrecht. <https://dev.clariah.nl/files/dh2019/boa/0192.html>.

Interactions avec la sirène de Cagniard-Latour en réalité virtuelle

Ronan Gaugne¹, Théophile Nicolas², Dominique Bernard³, Julie Priser⁴, Marion Lemaire⁴, Meven Leblanc⁵, Jim Pavan⁶, Tom Roy⁶, Raphael Dupont⁶, Julie Ayoubi⁶, Marie-Marie Le Cornec⁶, Valérie Gouranton⁶

1 : UMR 6074 Institut de recherche en informatique et systèmes aléatoires (IRISA), université Rennes 1

2 : Institut national de recherches archéologiques préventives (INRAP)

3 : Université Rennes 1, mission Patrimoine scientifique – commission Culture scientifique et technique

4 : Université Rennes 1

5 : École supérieure d'ingénieurs de Rennes, université Rennes 1

6 : INSA de Rennes, UMR 6074 Institut de recherche en informatique et systèmes aléatoires (IRISA), université Rennes 1

La sirène, due à Charles Cagniard de La Tour, est un instrument de mesure acoustique du XIX^e siècle qui a fourni une méthode d'évaluation de la fréquence des sons. Inventée en 1820, le premier appareil a été construit en 1845. Il permet de produire du son au moyen d'un mécanisme basé sur un disque percé d'orifices équidistants, actionné par un flux d'air. Un compteur associé au mécanisme permet de mesurer le nombre de tours du disque.

L'université Rennes 1 possède une telle sirène dans sa collection scientifique (<https://culture.univ-rennes1.fr/instruments-scientifiques>) (Bernard, 2018). L'équipe de conservation et de valorisation de cette collection s'est associée au projet de recherche pluridisciplinaire ANR-FRQSC INTROSPECT pour proposer une réplique numérique fonctionnelle de la sirène, dans un but de médiation scientifique.

Le projet franco-canadien INSTROSPECT, porté par l'INSA Rennes, associe des archéologues et des chercheurs en informatique. Il a pour objectif de proposer et concevoir des méthodes et outils innovants dans le domaine du patrimoine culturel, basés sur une combinaison d'imagerie numérique et de technologies 3D telles que la réalité virtuelle, la réalité augmentée ou l'impression 3D (Gaugne et Gouranton, 2019).

En archéologie, les processus scientifiques basés sur la Réalité Virtuelle (RV), la Réalité Mixte (RM) ou la Réalité Augmentée (RA) ont connu une augmentation considérable ces dernières années. L'archéologie virtuelle a été introduite pour la première fois par Reilly en 1990 et a été initialement présentée pour l'enregistrement des fouilles et la ré-excavation virtuelle à l'aide de technologies multimédias. De la même manière, Krasniewicz, en 2000, a proposé une infrastructure de visualisation à 360 ° pour aider les archéologues dans leurs travaux de recherche. Dans ce cas, l'archéologie virtuelle n'a pas été utilisée pour restaurer des connaissances, mais pour acquérir de nouvelles connaissances.

La XR (ou eXtended Realities) est un terme générique qui englobe la RV, RM et RA, elle est largement utilisée pour la valorisation du patrimoine culturel mais reste rarement utilisée à des fins scientifiques. Dans le cadre du projet INTROSPECT, nous montrons que les domaines de la XR et du patrimoine culturel peuvent bénéficier d'avantages mutuels déclenchés par des questions telles que :

- Comment la XR, considérée dans toute sa portée, de la science et de la technologie aux sciences humaines et aux sciences naturelles, peut-elle profiter à l'étude du patrimoine culturel ?
- Quels sont les défis scientifiques intrinsèques aux problématiques du patrimoine culturel ?

Les travaux du projet INTROSPECT visent en premier lieu à étudier et démontrer l'intérêt de ces technologies pour accompagner et améliorer les chaînes opératoires des archéologues (Nicolas, Gaugne, *et al.* 2018) ou des restaurateurs (Biron, Hurtin, *et al.* 2015), mais aussi pour en créer de nouvelles (Nicolas, Treyvaud *et al.* 2017, Lécuyer, Gouranton, *et al.* 2018, Gaugne, Nicolas *et al.* 2019). En complément de la démarche scientifique, les résultats obtenus bénéficient de l'apport des technologies 3D pour la valorisation

et la médiation scientifique (Gaugne, Porcier *et al.* 2018, Gaugne, Samaroudi, *et al.* 2018, Gaugne, Labaune-Jean *et al.* 2020).

Dans le cas de la sirène, l'objectif est de reconstituer un double numérique de l'appareil et de lui associer un contexte de mise en fonction dans un environnement de réalité virtuelle. Cette reconstitution a vocation à accompagner les actions de médiation organisées par l'équipe chargée de la conservation de la collection scientifique de l'université Rennes 1. Elle constitue également un défi pour nos travaux par l'application de ses méthodes à un objet scientifique multi-matériaux. Afin de réaliser cet objectif, deux étapes distinctes sont nécessaires, la numérisation et la contextualisation.

L'étape de numérisation consiste à produire des données numériques en 3D qui vont constituer la copie virtuelle de la sirène. Au-delà de la simple reproduction visuelle en trois dimensions de l'enveloppe externe texturée, il est nécessaire de représenter également les mécanismes internes de l'appareil afin de pouvoir simuler son fonctionnement. Pour cela, trois techniques ont été combinées : Les différentes pièces amovibles de la sirène ont été passées aux rayons X, en tomographie (Jankowski et Ferretti, 2010) (Nicolas *et al.* 2016), afin d'accéder à leur structure interne. Elles ont ensuite été numérisées avec un scanner à lumière structurée (Zhang, 2018). Enfin, une CAO du mécanisme de mesure a permis de compléter l'ensemble des données numériques collectées.

La contextualisation a pour objectif de proposer une simulation fonctionnelle de l'appareil et de l'intégrer dans un environnement virtuel interactif afin de proposer à l'utilisateur, un contexte d'utilisation. La remise en contexte d'un objet patrimonial est un aspect fondamental des réalités synthétiques (XR) (Gaugne et Gouranton, 2020) telles que la réalité virtuelle ou augmentée. Elle permet de dépasser la représentation visuelle d'un objet en lui associant des interactions fonctionnelles, pour simuler un fonctionnement, ou opérationnelles, pour simuler une chaîne opératoire (Gaugne & Gouranton, 2018). Dans le cas de la sirène, le contexte d'interaction consiste à mettre en fonction l'appareil en réalité virtuelle, dans un environnement qui simule l'expérience de mesure sonore.

Dans la reconstitution de la sirène de Cagniard en réalité virtuelle, les données produites associées à une simulation interactive permettent d'aller au-delà de la visualisation documentaire. La mise en contexte fonctionnelle reproduit pour l'utilisateur l'usage de l'appareil scientifique. Mais la représentation en réalité virtuelle permet également de dépasser cette reproduction réelle. Il est possible de rendre visibles des éléments et phénomènes non perçus par l'utilisateur dans le monde réel. Une augmentation de la visualisation de l'appareil, par exemple à travers des rendus en transparences, accompagnés d'une représentation des flux et de son impact sur le mécanisme de mesure assure une compréhension naturelle du fonctionnement de l'expérience originale.

Références

Bernard D (2018). « *Un trésor scientifique redécouvert : La collection d'instruments scientifiques de la faculté de sciences de Rennes (1840-1900)* », Rennes en Sciences, p. 88-89, 2018, EAN 9782490401024

Biron M., Hurtin S., Nicolas T., Tavernier C., (2015), La tomographie des objets archéologiques complexes et /ou altérés : outil d'identification, d'analyse et d'aide à la décision pour les mesures conservatoires, *In* Conservation-restauration des biens culturels, Cahier technique n° 22, Restaurer l'ordinaire. Exposer l'extraordinaire ? Journée des Restaurateurs en Archéologie (JRA), Arles 2014, p. 41-44

Gaugne R., F. Labaune-Jean, D. Fontaine, G. Le Cloirec, V. Gouranton. (2020) From the engraved tablet to the digital tablet, history of a fifteenth century music score. *Journal on Computing and Cultural Heritage*, Association for Computing Machinery, 2020, 13 (3), p. 1-18

Gaugne R., V. Gouranton. (2020) De la tomographie à l'interaction 3D en *eXtended Reality*, *Les nouvelles de l'archéologie*, 159 | 2020, 79-84.

Gaugne R., V. Gouranton. (2019) Le projet Introspect : archéologie et *eXtended Reality* (XR). *Les Rencontres du Consortium 3D SHS*, Florent Laroche ; Mehdi Chayani ; Xavier Granier ; Caroline Delevoie, Déc. 2019, Nantes, France.

- Gaugne R., V. Gouranton. (2019). 3D pour l'archéologie : quelles interactions, pour quoi faire ? – *Archéologie : imagerie numérique et 3D*, Sylvie Eusèbe ; Théophile Nicolas ; Valérie Gouranton ; Ronan Gaugne, Juin 2018, Rennes, France.
- Gaugne, R., T. Nicolas, Q. Petit, M. Otsuki, V. Gouranton. (2019), Evaluation of a Mixed Reality based Method for Archaeological Excavation Support. *ICAT-EGVE 2019 – International Conference on Artificial Reality and Telexistence – Eurographics Symposium on Virtual Environments*, Sep 2019, Tokyo, Japan
- Gaugne, R., S. Porcier, T. Nicolas, F. Coulon, O. Hays, Gouranton V. (2018) A digital introspection of a mummy cat. *Digital Heritage 2018 – 3rd International Congress & Expo, IEEE*, Oct 2018, San Francisco, United States.
- Gaugne R., M. Samaroudi, T. Nicolas, J.-B. Barreau, Gouranton V. (2018) Virtual Reality (VR) interactions with multiple interpretations of archaeological artefacts. *EG GCH 2018 – 16th EUROGRAPHICS Workshop on Graphics and Cultural Heritage*, Nov 2018, Vienna, Austria
- Jankowski A., G. Ferretti, (2010) Tomodensitométrie volumique : principe, paramètres, *Revue des Maladies Respiratoires*, Volume 27, Issue 8, 2010, Pages 964-969, ISSN 0761-8425, <https://doi.org/10.1016/j.rmr.2010.09.004>.
- Krasniewicz L. (2000) Immersive imaging technologies for archaeological research. *CAA : The 28th Conference on Computer Applications and Quantitative Methods in Archaeology*, 843: 163-169, 2000
- Lécuyer F., V. Gouranton, R. Gaugne, T. Nicolas, G. Marchand, *et al.* (2018). INSIDE Interactive and Non-destructive Solution for Introspection in Digital Environments. *Digital Heritage 2018 – 3rd International Congress & Expo, IEEE*, Oct 2018, San Francisco, United States.
- Nicolas T., R. Gaugne, C. Tavernier, E. Millet, R. Bernadet, Gouranton, V. (2018) Lift the veil of the block samples from the Warcq chariot burial with 3D digital technologies. *Digital Heritage 2018 – 3rd International Congress & Expo, IEEE*, Oct 2018, San Francisco, United States. p. 1-8
- Nicolas, T. & Treyvaud, G. & Favrel, Q. & Gaugne, R. & Gouranton, V.. (2017). Photogrammétrie, lasergrammétrie, tomodensitométrie : Des outils numériques pour la caractérisation des chaînes opératoires céramiques. *Archeopages*. 10.4000/arheopages.3483.
- Nicolas T., R. Gaugne, C. Tavernier, V. Gouranton, B. Arnaldi (2016). La tomographie, l'impression 3D et la réalité virtuelle au service de l'archéologie. *Les Nouvelles de l'archéologie*, Maison des Sciences de l'homme, 2016, p. 16-22
- Reilly P. (1990) Towards a virtual archaeology. In *Computer Applications in Archaeology*, pages 133-139. Oxford : British Archaeological Reports, 1990
- Zhang, S. (2018). High-speed 3D shape measurement with structured light methods : A review. *Optics and Lasers in Engineering*, 106, 119-131.

Katabase : à la recherche des manuscrits vendus

Simon Gabay^{1,2}, Ljudmila Petkovic, Alexandre Bartz³, Matthias Gille Levenson⁴, Lucie Rondeau Du Noyer

1 : Université de Genève

2 : Université de Neuchâtel

3 : École nationale des chartes

4 : UMR 5648 Histoire, archéologie et littératures des mondes chrétiens et musulmans médiévaux, Lyon

Les marchés de l'art, des livres ou des manuscrits sont tous relativement anciens, mais ne bénéficient cependant pas des mêmes outils pour la recherche. Des bases de données comme *ArtPrice*[1] existent pour les beaux-arts (peinture, sculpture...) et permettent de recenser les ventes. Des outils équivalents existent pour les livres anciens aux États-Unis [2], au Royaume-Uni [3], en Allemagne [4] ou en France [5].

Si certains index pour les ventes de livres anciens recensent bien les autographes, tous le ne font pas [6], et les publications apparues tardivement ne reviennent pas sur les ventes passées. La documentation est donc disparate et fragmentaire, concernant une ressource de premier ordre pour les collectionneurs, mais aussi pour les philologues en quête de sources, les historiens du livre ou les adeptes de la *Rezeptionsgeschichte* qui peuvent s'intéresser aux prix ou aux noms des collectionneurs.

Enjeu

Si les principaux problèmes posés par la numérisation de catalogues comme la *Revue des autographes* sont connus, tout comme les enjeux de la détection d'un manuscrit revenant plusieurs fois sur le marché, parfois sous une forme fragmentaire [8], il nous a paru important d'améliorer notre algorithme de classification. Ce dernier doit en effet être implémentable dans une application web disponible en ligne, tout en étant capable de traiter de grandes quantités de données avec un maximum de précision.

L'enjeu est donc la conception d'un algorithme de classification assez précis pour reconnaître un même document, mais assez souple pour s'accommoder de variations plus ou moins importantes.

Stratégie

Afin d'accélérer le traitement de l'information et d'alléger le poids des fichiers mis en ligne, l'encodage XML-TEI, qui n'est qu'un format pivot, est abandonné au profit du JSON.

Chaque fois que c'est possible une *string* est convertie en *integer* ou en *float* :

- Pour la longueur (*number_of_pages*) les documents incomplets sont ramenés à un nombre décimal (« une page et demie » → 1.5, « un quart de page » → 0.25...)
- Pour le format (*format*) le nombre de pliage est le chiffre retenu (« in-4^o » → 4, « in-folio » → 1...)
- Pour la date (*date*) on utilise le format ISO YYYY-MM-DD (« 3 mai 1645 » → 1645-05-03, « septembre 1736 » → 1736-09...)
- Le type de document (*term*) est converti en chiffre : ainsi L.a.s. (Lettre autographe signée) a le code 7, tandis que P.a.s. (Pièce autographe signée) a le code 2.

Les informations en JSON sont alors transformées pour faire une base de données orientée graphe, afin de faciliter la réconciliation des données.

Réconciliation

La transformation des données en graphe permet de simplifier le mécanisme de réconciliation : si chaque nœud représente un document vendu, il suffit d'ajouter une arête entre deux nœuds une fois atteint un certain degré de similarité.

Nous parlons de similarité et non d'identité stricte, car il n'est pas souhaitable de rechercher cette dernière : deux entrées différentes peuvent en effet renvoyer à un même document pour des raisons internes (deux fragments d'un même manuscrit) comme externes (une faute d'OCR). Il faut donc contourner ce problème via un algorithme de classification apte à gérer ces discrepances.

À partir de la liste des documents vendus, chaque entrée est comparée avec les autres. Cette comparaison se fait sur la bases des informations clefs standardisées dans le fichier JSON : pour chacune de ces informations, un système de bonus/malus est appliqué. Si le score obtenu est supérieur à 0.6, alors les entrées sont considérées comme renvoyant à un même manuscrit.

La valeur de ces bonus/malus a été trouvée de manière expérimentale, sur la base de tests unitaires évaluant l'efficacité de l'algorithme. Ces valeurs sont susceptibles d'évoluer avec l'ajout de nouveaux manuscrits.

Applications

Une application en ligne [<https://katabase.herokuapp.com>] s'appuie sur les données en JSON pour l'affichage des catalogues, qui sont disponibles à la lecture, et sur l'algorithme de classification afin de proposer un double mode de lecture des résultats pour une requête dans la base : la liste des ventes et la liste des manuscrits vendus.

Les données disponibles proviennent pour l'instant presque essentiellement de catalogues de vente à prix marqués, publiés dans le dernier tiers du XIX^e siècle à Paris par Gabriel Charavay (le détail précis des catalogues numérisés est disponible dans l'application).

En faisant tourner l'algorithme sur ces données préliminaires, nous pouvons déjà offrir quelques premiers résultats. Nous avons pu définir un ratio de retour sur le marché des manuscrits : pour 44 333 manuscrits vendus, 3 364 sont ont été vendus au moins deux fois, soit environ 7,5 %. À première vue, ces retours sur le marché sont marqués par une nette tendance baissière, notamment pour les manuscrits les plus chers, peu importe l'époque de l'auteur – la faible variation du franc à cette période et le de court laps de temps étudié permet par ailleurs une comparaison des prix malgré l'évolution du cours de la monnaie.

Recherches futures

Du point de vue philologique, la base de données ainsi que les capacités de classement développées pour l'application devraient permettre de retrouver plus aisément les sources des futures éditions, mais aussi de garantir l'authenticité des documents. Ces données devraient aussi être exploitables dans le cadre d'une approche distante du corpus afin d'étudier, par exemple, la construction du canon *via* la valeur marchande des auteurs.

Données et application

L'application web est disponible à l'adresse suivante : <https://katabase.herokuapp.com>.

Toutes les données utilisées pour ce projet sont disponibles en ligne à l'adresse suivante : <https://github.com/katabase>.

Bibliographie

[1] <https://fr.artprice.com>.

[2] *American Book-Prices Current*, New York, 1894/95-. ABPC tend avec le temps à répertorier de plus en plus de ventes européennes.

[3] *Book-Auction Records*, London, 1902-1997 et *Book Prices Current*, London, 1887-1952.

[4] *Jahrbuch der Auktionspreise für Bücher, Handschriften und Autographen*, Hamburg, 1950-. Au début Jahrbuch der Auktionspreise für Bücher und Autographen.

[5] *L'Argus mensuel du livre ancien et moderne*, Promodis, Paris, 1981-,

[6] *Autograph Prices Current*, London, 1914-1922

[7] Simon Gabay, Lucie Rondeau Du Noyer et Mohamed Khemakhem, « Selling autograph manuscripts in 19th c. Paris : digitising the Revue des Autographes », dans *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile : sfide e prospettive per l'Informatica Umanistica*, Milan, Italy, 2020 (Quaderni di Umanistica Digitale), p. 113-118, url : <https://hal.archives-ouvertes.fr/hal-02388407>.

[8] S. Gabay, L. Rondeau Du Noyer, Matthias Gille Levenson, Ljudmila Petkovic et Alexandre Bartz, « Quantifying the Unknown : How many manuscripts of the marquise de Sévigné still exist ? », dans *Digital Humanities DH2020*, Ottawa, Canada, 2020 (DH2020 Book of Abstracts), URL : <https://hal.archives-ouvertes.fr/hal-02898929> (visité le 23/11/2020).

L'observation (numériquement équipée) des problèmes publics

Waldir Lisboa Rocha¹

1 : Doctorant en sociologie, Centre d'étude des mouvements sociaux (CEMS), EHESS

À partir d'une étude de cas concernant la destitution de la présidente brésilienne Dilma Rousseff (2016) et de la manière dont nous nous sommes servi d'outils numériques pour la mener, nous aimerions soulever quelques questionnements sur les rapports entre la qualité de l'environnement numérique de travail du.de la chercheur.se, la temporalité de l'enquête et la manière dont celle-ci intervient dans l'histoire des problèmes publics (CEFAÏ, 1996, 2016 ; NEVEU, 1999, 2016) et *a fortiori* dans le processus démocratique. En reprenant John Dewey (1993, 2010 ; *apud* ISAAC, 2015) et en discutant des expérimentations socio-informatiques menées au sein de la *pragmatique des transformations* (CHATEAURAYNAUD, 2011, 2016, 2019 ; CHATEAURAYNAUD et DEBAZ, 2011, 2017), nous aimerions en outre présenter quelques arguments pour faire un plaidoyer en faveur d'une mise en place d'instances collaboratives et numériquement équipées d'observation continue de la réalité sociale qui puissent contribuer à « prendre soin » de la sphère publique (VOIROL, 2018).

La sociologie des problèmes publics a maintenant toute une histoire qui, de Becker (1963) et Blumer (1971) à Neveu (1999, 2015) et Cefaï (1996, 2016), en passant par Gusfield (1981), Pfohl (1977) et d'autres, a consolidé une tradition de recherche permettant de renouveler les approches des questions sociales à travers une attention accrue sur leurs formes de constitution et sur leurs processus d'émergence dans les arènes publiques. Ce champ de recherche, qui concerne une grande diversité d'objets et qui demande la mobilisation de différents savoirs, joue un rôle important dans la compréhension d'enjeux centraux de la contemporanéité, de la transformation de formes de mobilisation et des évolutions de la dispute pour l'attention publique. Pour ce faire, il demande une forte prise en compte de l'empirique à travers la mobilisation de différentes méthodes pour comprendre les acteurs et les institutions, les réseaux et les contextes où émergent les problèmes.

Tenant ces recherches comme une ressource théorique et méthodologique importante, nous nous sommes intéressé au processus de destitution de la présidente brésilienne Dilma Rousseff, mise à l'écart du pouvoir en 2016 et remplacée par son vice-président, Michel Temer. Il a alors été possible de repérer la manière dont différents problèmes publics se sont concurrencés, mais ont aussi convergé dans un processus complexe de formation d'arènes publiques multiples (CEFAÏ, 2016), imbriquées avec la destitution de Rousseff. En particulier, nous avons pu voir comment des problèmes qui avaient émergé des années auparavant (et qui ont bifurquer par la suite), comme la question de la corruption, ont été remobilisés en parallèle avec la problématique du « gonflement » de l'État pour faire de l'enlèvement de la présidente une cause publique majeur.

Nous aimerions dans cette communication présenter la manière dont les outils numériques ont été fondamentaux pour pouvoir traiter un objet dont la complexité est considérable, bien comme les différentes difficultés techniques qu'on du être affrontées pour rendre cela faisable. Nous voudrions en particulier expliciter trois points : premièrement, la façon dont Prospéro, logiciel d'analyse logico-sémantique fine, conçu pour traiter des dossiers complexes dans la durée (Chateauraynaud, 2005) et contribuant au développement de la *pragmatique des transformations*, a été notamment important pour saisir les configurations et reconfigurations de l'affaire ; deuxièmement, le travail de sélection et mise en place d'outils qui nous a permis de gérer et traiter les données et corpus, tout en inscrivant ces activités dans une logique collaborative et dans une visée de science ouverte ; troisièmement, comment nous avons, au fur et à mesure, été mené à rapprocher des outils de recherche et des outils de gestion et d'envisager la nécessiter de défracter l'environnement numérique du.de la chercheur.se. Ces différents éléments nous ont alors permis de

préfigurer un écosystème numérique dédié aux sciences sociales et aux humanités numériques qui soit à même de traiter d'enjeux concernant l'attention (CITTON, 2014) et d'actualiser les potentialités de l'écriture numérique (GUICHARD, 2017)

Références bibliographiques

- Cefaï, Daniel. 1996. « La construction des problèmes publics. Définitions de situations dans des arènes publiques ». *Réseaux* 14(75) : 43-66. DOI: 10.3406/reso.1996.3684.
- Cefaï, Daniel. 2016. « Publics, problèmes publics, arènes publiques... ». *Questions de communication* n° 30(2) : 25-64.
- Chateauraynaud, Francis. 2003. *Prospéro : une technologie littéraire pour les sciences humaines*. Paris : CNRS éditions.
- Chateauraynaud, Francis, et Josquin Debaz. 2011. Processus d'alerte et dispositifs d'expertise dans les dossiers sanitaires et environnementaux : Observatoire informatisé de veille sociologique. hal-00662501
- Chateauraynaud, Francis. 2014. « Trajectoires argumentatives et constellations discursives, Argumentative trajectories and discursive constellations ». *Réseaux* (188) : 121-158. DOI : 10.3917/res.188.0121.
- Chateauraynaud, Francis. 2019. « Petit traité de contre-intelligence artificielle. Retour sociologique sur des expérimentations numériques ». *Zilsel* 1(5) : 174-195.
- Citton, Yves. 2014. *Pour une écologie de l'attention*. Seuil.
- Dewey, John. 1993. *Logique : la théorie de l'enquête*. Paris, France : Presses universitaires de France.
- Dewey, John. 2010. *Le public et ses problèmes*. Paris, France : Gallimard.
- Guichard, Éric. 2017. « L'histoire et l'écriture numérique. Approche technique, politique, épistémologique ». *HAL* 19.
- Joseph, Isaac. 2015. « L'enquête au sens pragmatiste et ses conséquences. Vulnérabilité du public, observation coopérative et communauté d'exploration ». *SociologieS*.
- Neveu, Erik. 1999. « L'approche constructiviste des « problèmes publics ». Un aperçu des travaux anglo-saxons ». *Études de communication. langages, information, médiations* (22) : 41-58. DOI : 10.4000/edc.2342.
- Neveu, Érik. 2016. *Sociologie politique des problèmes publics*. Paris, France : Armand Colin.
- Voirol, Olivier. 2017. « Pathologies de l'espace public et agitation fasciste ». *Réseaux* (202 203) : 123-159. DOI : 10.3917/res.202.0123.

L'utilisation de l'outil Stylo dans le cadre de la revue *Sens public*

Antoine Fauchié¹, Margot Mellet¹, Nicolas Sauret^{1, 2, 3}

1 : Chaire de recherche du Canada sur les écritures numériques

2 : EA 7339 Laboratoire Dispositifs d'information et de communication à l'ère numérique (DICEN), Conservatoire national des arts et métiers, université Gustave-Eiffel, université Paris-Nanterre

3 : Chercheur au FabPart Lab à HumaNum Lab

Créée en 2003 par Gérard Wormser, la revue *Sens public* est une revue interdisciplinaire et internationale en Sciences humaines et sociales. Dès son origine, cette revue numérique s'est inscrite dans la pensée du collectif et le développement de conversation scientifique entre les différents acteurs de la production et diffusion de savoir : chercheur·e·s, éditeur·trice·s, traducteur·trice·s, créateur·trice·s. Cet engagement se traduit par la recherche d'un dialogue continu pour créer un espace public autour du savoir qui serait investi par un réseau d'intelligence (Lévy 1994).

À cette conception du savoir comme une dynamique commune s'ajoute un engagement dans la production de contenus scientifiques structurés et conformes aux formats standards libres et non-propriétaires. C'est par ces choix techniques et éthiques que l'éditeur de texte Stylo (<https://stylo.huma-num.fr>) a été conçu (Vitali-Rosati *et al.* 2020) et est désormais l'outil sur lequel se fonde la chaîne éditoriale de la revue. L'outil Stylo – pensé dans une démarche d'édition numérique (Epron et Vitali-Rosati 2018) et s'inspirant de nombreuses initiatives technologiques (Blanc et Haute 2018) – permet de produire notamment du HTML à partir de trois formats qui constituent aussi des standards pour l'édition savante : Markdown (pour l'écriture du corps du texte), YAML (pour l'agencement des métadonnées), BibTeX (pour la structuration des références bibliographiques). Dans l'idée de présenter cet outil autrement qu'au travers de ses modalités techniques, nous souhaitons, dans le cadre de cette communication, partager un retour d'expérience par les praticien·ne·s de l'outil dans le contexte plus particulier de la chaîne éditoriale de la revue *Sens public*. Stylo ne se positionne pas comme une alternative aux outils grand public et commerciaux (comme Microsoft Word), mais il aspire à promouvoir et encourager les usages des standards ouverts (notamment HTML, Markdown, XML). Stylo réalise cet objectif en implémentant des protocoles ouverts.

L'utilisation de Stylo advient tout au long de l'édition de l'article : de la proposition à la revue par l'auteur·e jusqu'à l'export de la version finale et éditée pour publication. La soumission pour la revue se fait donc à la condition d'une certaine adaptation des auteur·e·s et de leurs pratiques d'écriture définies majoritairement par l'usage de Word (Kirschenbaum 2016) à un nouvel outil qui favorise l'écriture et l'édition sémantique (Kembellec 2019). En plus d'un suivi des articles, la revue s'est donc engagée dans un accompagnement des auteur·e·s dans le cadre notamment d'une permanence hebdomadaire en ligne qui permet, en plus de constituer un soutien technique, d'établir une discussion autour des choix éditoriaux de *Sens public* et un dialogue direct entre les concepteurs et les utilisateurs de l'outil. Au-delà du mode édition de l'outil Stylo qui devient le lieu d'un travail commun de structuration du savoir, le mode *preview* qui mène à une visualisation web du contenu est, avec l'entremise de l'outil d'annotation Hypothesis, l'espace d'échanges à même le contenu visible. L'annotation est alors considérée comme un outil de dialogues divers (Sauret 2020). L'auteur·e a la possibilité de répondre et d'engager un dialogue dans le cadre de l'évolution de sa pensée et de son travail via les retours de l'évaluation ouverte par les pairs, des relectures par les membres du comité éditorial, des questions ou des demandes d'éditeurs. Le versionnage du travail, la possibilité adjointe d'une comparaison entre versions du texte, permet un suivi des différentes étapes de la chaîne éditoriale, et constitue également une vision progressive de la structuration du savoir qui se voit rythmée par les conversations successives.

Le fait d'un dialogue au sein même du discours scientifique entre l'auteur·e et les autres acteurs de l'édition modifie la perception et les pratiques éditoriales au sein de la revue pour favoriser le sentiment d'un travail

qui, au-delà d'être commun, s'inscrit dans un collectif ouvert et dans la (re)connaissance entre les individus de ce collectif et entre les différents praticien·ne·s de Stylo. Il nous semble déterminant de présenter, dans le cadre d'une communication, un retour d'expérience de l'utilisation de cet outil pour l'édition de la revue *Sens public* tout en contextualisant sa création et son usage dans l'objectif d'une conversation scientifique (Sauret 2018).

Blanc, Julie, et Lucile Haute. 2018. « Technologies de l'édition numérique ». *Sciences du Design* 8 (2) : 11-17. <https://www.cairn.info/revue-sciences-du-design-2018-2-page-11.htm>.

Epron, Benoît, et Marcello Vitali-Rosati. 2018. *L'édition à l'ère numérique*. Repères. Paris : La Découverte. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/20642>.

Kembellec, Gérald. 2019. « Semantic publishing, la sémantique dans la sémiotique des codes sources d'écrits d'écran scientifiques ». *Les Enjeux de l'information et de la communication* n° 20/2 (2) : 55. <https://doi.org/10.3917/enic.027.0055>.

Kirschenbaum, Matthew G. 2016. *Track changes : a literary history of word processing*. Cambridge, Massachusetts, États-Unis d'Amérique : The Belknap Press of Harvard University Press.

Lévy, Pierre. 1994. *L'intelligence collective : pour une anthropologie du cyberspace*. Sciences et société. Paris : La Découverte.

Sauret, Nicolas. 2018. « Design de la conversation scientifique : naissance d'un format éditorial ». *Sciences du Design* n° 8 (2) : 57-66. <https://www.cairn.info/revue-sciences-du-design-2018-2-page-57.htm>.

Sauret, Nicolas. 2020. « De la revue au collectif : la conversation comme dispositif d'éditorialisation des communautés savantes en lettres et sciences humaines ». Thèse de doctorat, Montréal, Canada : université de Montréal.

Vitali-Rosati, Marcello, Nicolas Sauret, Antoine Fauchié, et Margot Mellet. 2020. « Écrire les SHS en environnement numérique. L'éditeur de texte Stylo ». *Intelligibilité du numérique*, n° 1 (septembre). <http://intelligibilite-numerique.numerev.com/numeros/n-1-2020/18-ecrire-les-shs-en-environnement-numerique-l-editeur-de-texte-stylo>.

Le recours aux humanités numériques pour (re)constituer le dossier génétique d'une création collective

Johanne Melancon¹

1 : Université Laurentienne

L'année universitaire 2020-2021 marque le 50e anniversaire de la création collective de la Troupe de l'université Laurentienne (Sudbury, Ontario), *Moé, j'viens du Nord, 'stie !* (<https://letno.ca/programmation/moe-jviens-du-nord-stie-1970-1971/>), qui a mené à la fondation du Théâtre du Nouvel-Ontario (TNO) à l'été 1971. La pièce de théâtre attribuée à André Paiement constitue « la première brique du monument qui deviendra le théâtre franco-ontarien contemporain » (Beddows 2004, 12). Pour souligner cet anniversaire, le Programme de théâtre de l'université Laurentienne prépare, dans le cadre d'un cours et en collaboration avec le TNO, une création collective, *Quoi si, moé 'si j'viens du nord, 'stie* (<https://letno.ca/programmation/quoi-si/>). C'est dans ce contexte qu'un projet d'exposition virtuelle, faisant appel à la génétique et aux humanités numériques et incluant une dimension critique (Duval 2017, 15), a été élaborée avec comme principal objectif la conservation et la transmission du patrimoine culturel d'une communauté en situation minoritaire pour qui le théâtre a constitué l'élément déclencheur d'une nouvelle identité collective franco-ontarienne au début des années 1970.

Ce projet d'exposition virtuelle évolutive cherche à mettre en valeur les archives la création collective de 1970-1971, tant pour la genèse textuelle que la genèse scénique (Grésillon, Mervant-Roux et Budor 2010). S'il est vrai qu'il est toujours difficile de construire le dossier de genèse d'une œuvre performative (Brailion et Idmhand 2017 ; Grésillon et Thomasseau 2005, 32), l'expérience avec *Moé, j'viens du Nord, 'stie !* s'est avérée particulièrement difficile. En effet, il ne subsiste, dans les archives officielles, que très peu de traces de l'écriture et de la production de la pièce. Pour pallier l'absence de documents témoignant de cette création collective, il fallait donc entrer en contact avec ceux et celles qui avaient participé au projet pour tenter de retrouver d'autres documents, une recherche qui est toujours en cours. Cependant, une autre possibilité s'est rapidement imposée, soit la création de documents d'archives par des entrevues recueillant les témoignages des membres de la Troupe de 1970-1971 qui pourront être intégrées sous forme de capsules audio ou vidéo à l'exposition. Ainsi, les humanités numériques permettraient non seulement de partager des documents d'époque par l'inventaire, le catalogage et la numérisation (Laborderie 2018, 6) puis la mise en ligne sous forme d'exposition, mais également d'ajouter une nouvelle dimension en incluant un commentaire contemporain pour expliquer, contextualiser et documenter cette expérience collective hautement symbolique par ceux-là même qui l'avaient vécue.

De plus, l'exposition prévoit un volet présentant la création collective 2020-2021 du Programme de théâtre alors que le recours à la technologie numérique est désormais au cœur de la démarche de l'équipe de 2020-2021, entre autres avec la plateforme de communication Slack. Aussi, dans le cadre de la présentation du spectacle au printemps 2021, une table ronde réunira des membres de l'équipe de 1970-1971 et de celle de 2020-2021 qui aborderont leurs créations collectives respectives. L'un des axes retenus est le rôle de la technologie sur le processus de création puisque la production de 1970-1971 a été influencée par l'audio-visuel, une salle de spectacle à la fine pointe de la technologie ayant été inaugurée au printemps 1970 à la Laurentienne, alors que le processus de création et la production de 2020-2021 ne pouvait être réalisée sans le recours à la technologie numérique dans le contexte de la pandémie.

Ainsi, cette communication proposera les résultats d'une recherche qui porte d'abord sur le recours au numérique pour conserver et transmettre une œuvre fondatrice du patrimoine artistique et culturel du théâtre franco-ontarien contemporain, et sur les moyens qu'offre le numérique pour documenter à rebours un processus de création collective de façon à pallier l'absence de documents témoins. On pourra dès lors ajouter une dimension critique à l'exposition en interprétant ces nouveaux documents pour mieux

comprendre le processus de création. Cette démarche suscite cependant certaines questions. Ainsi, d'un point de vue épistémologique, les documents nouvellement créés doivent-ils être considérés comme des archives ? Comment doivent-ils être intégrés dans le dossier génétique ? Aussi dans quelle mesure le choix éditorial du recours à la création de nouvelles archives transforme-t-il le travail du commissaire d'exposition/généticien ? Par ailleurs, mettant à profit les possibilités offertes par le numérique et dans une perspective comparative, cette communication s'interrogera sur l'impact de la technologie – audio-visuelle en 1970-1971, numérique en 2020-2021 – sur le processus de création et de production, en s'appuyant principalement sur la discussion qui aura eu lieu entre les membres des deux équipes, tout en tenant compte des différents contextes.

Références bibliographiques

Beddows, Joël. 2004. « À l'origine du monument, une école de théâtre », préface dans André Paiement, *Les partitions d'une époque, Les pièces d'André Paiement et du Théâtre du Nouvel-Ontario (1971-1976)*, tome I, Sudbury, Prise de parole, p. 12-23.

Cécile Braillon, Fatiha Idmhand. Arts performatifs et processus de créations à l'ère numérique. Enjeux techniques et théoriques. *Revue d'histoire du théâtre*, Société d'Histoire du Théâtre, 2018, Études théâtrales et humanités numériques. hal-02546134

Duval, Frédéric. 2017. « Pour des éditions numériques critiques. L'exemple des textes français ». *Médiévales*, « Le texte à l'épreuve du numérique », 73 (2) : 13-29. doi : 10.4000/medievales.8165

Grésillon, Almuth, Marie-Madeleine Mervant-Roux et Dominique Budor (dir.). 2010. *Genèses théâtrales*. Paris : CNRS Éditions.

Grésillon, Almuth et Jean-Marie Thomasseau. 2005. « Scènes de genèses théâtrales », *Genesis* (Manuscrits-Recherche-Invention), n° 26 : 19-34. <https://doi.org/10.3406/item.2005.1362>

Laborderie, Arnaud. 2018. « Mettre en récit les données culturelles des bibliothèques numériques : exposition virtuelle et recherche-crédation. » *Colloque international sur les bibliothèques et archives à l'ère des humanités numériques*, Octobre 2018, Paris, France. hal-01891073

Le rôle des modalités linguistiques dans l'analyse automatique du style critique

Marguerite Bordry^{1,2}

1 : Observatoire de la vie littéraire (OBVIL), Sorbonne Université

2 : EA 1496 Littérature et culture italiennes, Sorbonne Université

L'analyse du style critique des critiques littéraires est loin d'être une nouveauté : nombreux sont les grands critiques, en particulier dans la littérature française, dont le style a été analysé. Nous nous proposons de mener une recherche de ce type en utilisant l'annotation automatique des modalités du discours critique à l'échelle d'un corpus entier, le *Mercur de France*, dont une édition électronique, rassemblant tous les articles consacrés à l'Italie entre 1890 et 1918, a été réalisée par le Labex OBVIL [1]. Une partie non négligeable de ces articles concerne la littérature : ils étaient d'autant plus importants que le *Mercur* jouissait d'un prestige considérable en Italie (Livi, 2012) et pouvait faire ou défaire le sort d'un écrivain.

L'objectif est d'analyser ce corpus selon les modalités linguistiques (appréciation, dépréciation, émotions, etc.) en se concentrant sur les « Lettres italiennes », la chronique consacrée à la littérature italienne. Elle est tenue par Remy de Gourmont, sous son nom jusqu'en 1896, puis sous le pseudonyme d'A. Zanoni jusqu'en 1897, avant d'être reprise par Luciano Zuccoli (1897-1904), Ricciotto Canudo (1904-1913) et Giovanni Papini (1913-1918). Si l'objet de leurs chroniques reste le même – recenser, positivement ou négativement, les publications les plus marquantes de la vie littéraire italienne –, leur style critique, lui, diffère fortement.

Le point de départ de ce projet est une analyse menée sur l'identification automatique de l'ironie dans la critique littéraire (Alrahabi, Bordry, 2020) : grâce à une approche symbolique à base de règles, il s'agissait d'étudier les phrases du *Mercur de France* doublement annotées, à la fois comme positives et négatives, une proportion importante d'entre elles étant ironiques. En marge de ce travail consacré spécifiquement à l'ironie dans la critique littéraire, nous avons constaté des résultats d'annotation très différents selon les chroniqueurs. Or ceux-ci vont bien au-delà des seules doubles annotations positives et négatives, qui étaient, à l'origine, le point de départ de l'étude.

Cela nous a conduit à formuler des hypothèses sur le style critique des différents chroniqueurs de littérature italienne du *Mercur de France*. Remy de Gourmont et Luciano Zuccoli utilisent souvent l'éloge paradoxal, en usant de la négation et des adverbes d'intensité pour affirmer leurs positions. En témoigne, chez ces deux chroniqueurs, une proportion assez forte de doubles annotations, qui – en apparence – se contredisent :

« Girolamo Rovetta confie à cette même Revue le dernier de ses trois mille romans, *La Signorina*, qui menace d'être intéressant et vulgaire, ce qui rentrerait parfaitement dans les habitudes de l'auteur. » (1900)

Cette phrase est doublement annotée « Appréciation intellectuelle » et « Dépréciation éthique », ce qui reflète l'ambiguïté voulue par Zuccoli, qui manie volontiers l'éloge pour mieux amener la critique sévère qui suit.

Ricciotto Canudo se distingue par un style critique emphatique, pompeux et marqué par des phrases grammaticalement complexes. C'est sous sa plume que l'on trouve le plus grand nombre d'annotations multiples à l'échelle d'une seule phrase :

« Par les rythmes vastes et sombres qui revêtent tous les mots et tous les gestes de ces personnes dramatiques, par la poésie subtile des expressions, par l'atmosphère ardente et vraiment musicale qui enveloppe toute la tragédie, et surtout le premier acte, par la force originale de certaines situations et par le choc violent, brutal, fatalement logique des passions, le *Flambeau sous le boisseau* est la plus belle des tragédies de Gabriele d'Annunzio. » (1905)

Ce passage, consacré à *La fiaccola sotto il moggio*, tragédie de Gabriele D'Annunzio (1905), est révélateur du style critique de Ricciotto Canudo. Outre deux annotations révélant l'appréciation, « Appréciation informative » et « Appréciation esthétique », l'outil d'annotation met aussi en évidence deux autres catégories, « Tristesse » et « Colère », qui renvoient à son analyse de la pièce.

Giovanni Papini, quant à lui, est aux antipodes de son prédécesseur : il est direct et se concentre sur les nouvelles générations d'écrivains italiens, qu'il veut faire découvrir aux lecteurs du *Mercure*. C'est d'ailleurs précisément la raison pour laquelle, suite à des polémiques internes, la chronique des « Lettres italiennes » lui revient en 1913 (Gogibu, 2017 ; Kalantis, 2017). Là encore, ces constatations se reflètent dans les annotations qui dominent chez Papini, largement positives. En témoigne cette observation formulée à propos d'Aldo Palazzeschi, dont il vient de citer le poème « *Gioco proibito* » (« Jeu interdit ») :

« Toute la première période de la poésie de Palazzeschi se trouve résumée dans ce contrepoint de sensibilité blanc et noir et dans cette musicalité qui n'a pas besoin de recourir aux mots sonores et aux rimes rares pour nous enchanter. » (1914)

Ce passage est annoté « Appréciation », catégorie que l'on retrouve à plusieurs reprises dans tout l'article. Pour Papini, il s'agit de se démarquer de son prédécesseur, jugé indifférent aux mouvements littéraires les plus récents.

L'objectif de ce travail est de présenter une étude fine des principaux traits du style critique de chaque chroniqueur de littérature italienne du *Mercure* en utilisant les outils développés par notre équipe, comme la plateforme Ariane [2], qui combine l'annotation sémantique et la navigation textuelle. Grâce à Ariane, qui permet la fouille de texte à des échelles différentes, de la phrase au texte, puis au corpus entier, nous entendons proposer une cartographie des annotations dominantes chez chaque chroniqueur afin de confirmer – ou d'infirmer – les hypothèses que nous avons formulées sur le style critique de chaque chroniqueur. À terme, en croisant ces annotations et la recherche de mots-clés, cette recherche pourra être affinée à l'échelle d'un seul mouvement littéraire, voire d'un écrivain, pour déterminer comment s'articule la critique à son égard formulée par les différents chroniqueurs du *Mercure de France*. En nous appuyant sur l'annotation automatique des modalités, nous entendons proposer une analyse qui soit à la fois fine, à l'échelle de chaque chroniqueur, et complète, à l'échelle d'un corpus entier, le but étant d'exploiter et de valoriser l'édition électronique réalisée dans le cadre du projet « La culture italienne dans le *Mercure de France* (1890-1918) » du Labex Obvil.

[1] <http://obvil.sorbonne-universite.site/corpus/mdf-italie/>.

[2] <https://obvil.huma-num.fr/ariane/mdf/search>.

Références

Alrahabi, Motasem, Bordry, Marguerite. 2020. « L'ironie dans la critique littéraire : quelques pistes pour un traitement automatique ». Dans *Humanités numériques et Digital Studies*. Montpellier.

Bordry, Marguerite. 2020. « Les lettres italiennes au miroir du *Mercure de France*. Une approche textométrique », in *Lectures du texte numérique : expériences du Labex Obvil*, sous la direction de Glenn Roe et Didier Alexandre, Paris, Classiques Garnier (à paraître).

Dotoli, Giovanni (éd.). 1978. Ricciotto Canudo. 1877-1977. *Atti del convegno internazionale nel centenario della nascita* (Bari – Gioia del Colle, 24-27 novembre 1977), Fasano, Grafischena, 1978

Gogibu, Vincent. 2017. « Du rififi au *Mercure*: Ricciotto Canudo, Remy de Gourmont et Arthur Rimbaud dans les « Lettres italiennes » du *Mercure de France* », *Ent'revues*, « La Revue des revues », 2017/2, n° 58, p. 64-75 (<https://www.cairn.info/revue-la-revue-des-revues-2017-2-page-64.htm>)

Kalantzis, Alexia. 2017. « Un réseau italien. Les écrivains du *Mercure de France* en Italie (1895-1910) », *Ent'revues*, « La Revue des revues », 2017/2, n° 58, p. 76-85. (<https://www.cairn.info/revue-la-revue-des-revues-2017-2-page-76.htm>)

Livi, François. 2012. *Italica. L'Italie littéraire de Dante à Eugenio Corti*, Lausanne, L'Âge d'Homme, 2012.

Livi, François. 2006. « Papini e Remy de Gourmont », in Cosimo Ceccuti (éd.), *Papini e il suo tempo*, Firenze, Le Lettere, 2006, p. 76-97.

Livi, François, Perrus, Claudette (dir.). 1997. *Paris-Florence (1900-1920), Aspects du dialogue culturel*, *Revue des Études italiennes*, tome XLIII, n° 3-4, juillet-décembre 1997.

Richter, Mario (éd.). 1969. *La formazione francese di Ardengo Soffici (1900-1914)*, Milano, Vita e Pensiero, 1969.

Wilfert-Portal, Blaise. 2001. « Le critique, la presse et la nation : Rémy de Gourmont au *Mercure de France* », 1890-1900, *Cahiers de l'Association internationale des études françaises*, 2007, n° 59. p. 281-301.

Le style polémique dans les correspondances de Voltaire

Motasem Alrahabi¹, Camille Koskas¹, Glenn Roe¹

¹ : Observatoire des textes, des idées et des corpus (ObTIC), Sorbonne Université

La correspondance est un lieu d'échange et de dialogue qui offre un espace privilégié à l'expression des controverses, des polémiques et des débats [1]. Dans une perspective d'analyse automatique, nous nous proposons d'étudier les stratégies dialogiques inhérentes au style polémique dans la correspondance de Voltaire. Numérisé dans le cadre du projet Electronic Enlightenment à l'université d'Oxford [2], le corpus de travail est composé d'environ 21000 lettres structurées en XML-TEI.

L'approche utilisée est symbolique et vise à annoter automatiquement les passages polémiques reflétant les postures et les positionnements des protagonistes autour de sujets débattus. En nous appuyant sur des travaux linguistiques autour de la subjectivité [3, 4, 5] et en traitement automatique des langues [6, 7, 8], nous avons mené une construction systématique de dictionnaires de patrons linguistiques à caractère polémique, organisés en quelques dizaines de classes : accords, désaccords, opinions, critiques, éloges, excuses, insultes, etc.

L'application finale permet de naviguer dans les lettres à travers les passages polémiques annotés. L'utilisateur a le moyen de filtrer les résultats par des mots clés et des métadonnées, et de les croiser avec des noms propres (personnes et lieux cités) automatiquement repérés à l'aide d'un modèle d'apprentissage automatique [9].

L'étude des résultats d'annotation nous a permis de mettre en avant la spécificité des marqueurs linguistiques observables qui accompagnent l'expression de la polémique dans le corpus. Dans cet espace de dialogue d'idées, nous avons pu constater chez Voltaire, à des degrés divers, une grande importance accordée aux éléments stylistiques de détournement du discours adverse : parodie, humour, sarcasme et satire. D'autres procédés annotés pourraient être caractéristiques du style polémique voltairien [10], notamment l'inscription de différentes voix opposées dans un espace polyphonique, et la manipulation savante de différentes formes du discours rapporté : le discours direct qui renforce l'aspect objectif du débat ; mais surtout le discours indirect où les propos de l'adversaire semblent d'abord tolérés pour vite être disqualifiés.

Les résultats des premières évaluations sont encourageants malgré certaines difficultés que nous soulignons : la nature très particulière de l'écriture épistolaire (orthographe ancienne, textes courts, changement thématique fréquent...) ; l'attribution des discours cités dans les lettres (distinguer les propos de l'écrivain de celui d'une personne citée [11]), les messages à caractère ironique ou allusif qui sont très difficiles à détecter automatiquement [12], etc.

Après une démonstration de l'outil sur le corpus de Voltaire, la présentation portera sur les caractéristiques du corpus (numérisation, structuration, modernisation...), sur la méthodologie de l'analyse automatique et son évaluation, et sur les perspectives.

Références

[1] Gavoille E., Guillaumont, F. (dir). *Conflits et polémiques dans l'épistolaire*, Presses universitaires François Rabelais, 2015.

[2] Cronk, N., (2019), « Electronic Enlightenment : recreating the Republic of Letters », in Burrows, S. and Roe, G., eds., *Digitizing Enlightenment : Digital Humanities and the Transformation of Eighteenth century Studies*. Oxford University Studies in the Enlightenment : 55-72.

[3] Kerbrat-Orecchioni C., (1980), « La polémique et ses définitions », p. 3-40, in : Gelas N., Kerbrat-Orecchioni C., eds, *Le Discours polémique*, Lyon, Presses universitaires de Lyon.

- [4] Banfield, A., (1982), *Unspeakable Sentences*, Boston, Routledge and Kegan Paul.
- [5] Culioli, A., (1999), *Pour une linguistique de l'énonciation*, Tome 2, Paris, Ophrys.
- [6] Balahur, A., Mihalcea, R., Montoyo, A., (2014), « Computational approaches to subjectivity and sentiment analysis : Present and envisaged methods and applications », *Computer Speech & Language*, Volume 28, Issue 1.
- [7] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011), « Lexicon-based methods for sentiment analysis ». *Computational linguistics*, 37(2) : 267-307.
- [8] Alrahabi M., (2016), « E-Quotes : un outil de navigation textuelle guidée par les annotations sémantiques », in *TALN 2016*, Paris, France
- [9] Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton et Christopher D. Manning. 2020. « Stanza : A Python Natural Language Processing Toolkit for Many Human Languages ». Dans *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics : System Demonstrations*, 101-108. Online : Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-demos.14>.
- [10] Ibtissem Skander. *Le Dialogue polémique dans l'œuvre philosophique de Voltaire – étude stylistique* Thèse de doctorat, université de Sfax (Tunisie), 28 octobre 2017.
- [11] Paret, S., O'Keefe, T., Konstas, I., Curran, J.R., Koprinska, I., (2013), « Automatically Detecting and Attributing Indirect Quotations ». *Empirical Methods in Natural Language Processing (EMNLP)*, Seattle
- [12] Hernández I., Rosso P., (2016), « Irony, Sarcasm, and Sentiment Analysis ». Chapter 7. *Sentiment Analysis in Social Networks*, F.A. Pozzi, E. Fersini, E. Messina, and B. Liu (Eds.), Elsevier Science and Technology, p. 113-128

Les humanités numériques pour valoriser le patrimoine minier

Eric Kergosien¹, Amélie Daloz¹, Bernard Jacquemin¹, Stéphane Chaudiron¹

1 : EA 4073 Groupe d'études et de recherche interdisciplinaire en information et communication (GERiCO), université de Lille

Les travaux présentés dans cette communication sont menés dans le cadre du projet ANR MémoMines[1], dont l'objectif est d'apporter des outils et méthodes scientifiques pour aider à la sauvegarde et à la mise en valeur du patrimoine minier des Hauts-de-France. En effet, un enjeu social important dans le domaine du patrimoine culturel et notamment celui des mines est lié à la collecte, l'analyse, la publication et la mise en valeur de l'histoire et de la mémoire collectives des acteurs du domaine, à l'oral comme à l'écrit. Dans ce sens, la formalisation de l'information sur le patrimoine culturel représente un véritable défi en raison de la diversité et de l'incomplétude des données. De plus, ces données sont hétérogènes et peuvent être trouvées dans différentes sources, en ligne ou hors ligne : bases de données, bibliothèques, musées, dossiers de presse, expertise des intervenants, etc. Cette diversité des ressources pose de nombreux défis tels que la documentation des données, la représentation, l'intégration et l'interopérabilité au sein d'une même base de connaissances.

Pour répondre à ces problématiques, nous proposons une méthodologie pour la construction d'une base de connaissances, que l'on nomme également ontologie de domaine du patrimoine minier (exploitation du charbon) circonscrit au territoire des départements du Nord et du Pas-de-Calais en France et sur une période allant du début de l'exploitation (vers 1750) jusqu'à nos jours avec par exemple la réhabilitation des anciens sites de production et des habitats liés à la mine. Nous souhaitons dans ce sens structurer et formaliser des connaissances caractérisant les acteurs, les lieux, les entités temporelles et les thématiques présentes dans les corpus documentaires oraux et écrits analysés, et l'ontologie produite doit alors nous permettre d'en faciliter l'accès.

La méthodologie que nous présenterons dans notre communication se décompose en 3 étapes, la première étant la formalisation des entités nommées (lieux, acteurs, entités temporelles) et d'un vocabulaire structuré décrivant le domaine des mines en une première ontologie du domaine des mines. Les entités temporelles correspondent aux références calendaires (dates et années). Concernant les entités nommées acteurs et lieux, nous nous appuyons sur les ressources produites par la Mission Bassin Minier, partenaire du projet qui gère pour l'état le label UNESCO. Concernant le vocabulaire du domaine, nous nous appuyons sur le thésaurus ThesoMines défini dans le cadre du projet (Daloz et Chaudiron, 2019), validé par des experts du domaine et disponible dans sa version 1.0 en accès libre[2] sur la plateforme OpenTheso mise à disposition sur les serveurs de la Très Grande Infrastructure de Recherche HumNum[3]. Le thésaurus, défini selon le standard SKOS (*Simple Knowledge Organization System*), est composé de 558 descripteurs et 764 non-descripteurs, et est structuré en quatre domaines (culture éducation, économie, environnement naturel, histoire) et cinq facettes (activité, bâti, gisement, matériel, opération) afin de couvrir aussi bien le patrimoine matériel qu'immatériel, selon la définition qu'en donne l'UNESCO.

Au sujet des formalismes, la plupart des tentatives pour résoudre les problèmes d'interopérabilité sémantique se concentrent sur la standardisation, voir la normalisation de formalismes communs tels que FRBR (Le Boeuf, 2005), FRBRoo (Doerr *et al.*, 2013), CIDOC CRM (Le Boeuf *et al.*, 2015), etc. Au regard de l'état de l'art et de nos premières expérimentations (Kergosien *et al.*, 2019), nous faisons le choix de formaliser l'ontologie produite en CIDOC CRM, modèle conceptuel de référence conçu pour modéliser les domaines du patrimoine culturel. Nous détaillerons à partir d'exemples concrets les choix de formalisation des connaissances provenant des ressources sélectionnées (thésaurus, liste de lieux, etc.).

La seconde étape consiste à décrire le corpus de documents oraux et écrits à partir de la première ontologie produite. Cette seconde étape correspond à l'instanciation (ou peuplement) du modèle ontologique. Nous

inscrivons ici nos travaux dans la politique d'ouverture des données qui a pris de l'ampleur au cours de la dernière décennie. Dans la continuité de la loi pour une République numérique (7 octobre 2016) renforçant la politique d'ouverture des données publiques, le ministère de la Culture s'est engagé dans une politique d'ouverture et de partage des données publiques, notamment pour permettre aux citoyens de participer à l'amélioration de la qualité des données. C'est ainsi que les Archives nationales, en lien étroit avec le service interministériel des Archives de France, viennent de publier un nouveau régime de réutilisation des informations publiques qu'elles conservent, qu'il s'agisse des inventaires et des corpus d'archives. Il s'agit de documents librement communicables à tous et sur lesquels des tiers ne détiennent pas de droits de propriété intellectuelle. Les informations doivent être librement communicables au sens de l'article L.213-1 du Code du patrimoine et elles ne doivent pas avoir été communiquées par autorisation ou par dérogation. C'est également dans ce cadre que les Archives nationales ont publié plusieurs corpus librement réutilisables sur la plateforme de données ouvertes du ministère : data.culturecommunication.gouv.fr. En 2019, le ministère de la Culture a dans la continuité de son programme lancé la plateforme ouverte du patrimoine POP, nouvel outil pour la gestion des données culturelles du ministère, comportant deux volets : (1) une plate-forme de production, qui permet le versement de métadonnées (images, photographies, informations descriptives ou réglementaires) ; (2) une plate-forme de diffusion de ces données, accessible au grand public[4]. Dans le cadre du projet, nous travaillons avec le partenaire INA pour intégrer notre ontologie de domaine sur la plateforme OKAPI [5] dédiée à l'indexation et l'analyse et la publication de corpus de données audiovisuelles. Afin d'explicitier cette deuxième étape d'instanciation, nous nous appuyons sur une analyse fine d'un premier entretien mené auprès d'un ancien mineur, pour montrer à la fois l'intérêt de notre démarche, et également les difficultés rencontrées liées à l'indexation, et notamment au choix du niveau de granularité de description retenu.

La troisième étape consiste à enrichir le premier modèle ontologique produit à partir de nouvelles connaissances présentes notamment dans les entretiens menés auprès des anciens mineurs (nouveaux termes, nouvelles entités nommées, etc.), puis de valider la base de connaissances obtenue avec les experts du domaine. Nous étayerons cette dernière étape à travers différents exemples tirés de notre travail.

L'ontologie produite sera prochainement rendue disponible dans le format OWL CIDOC CRM, afin de la rendre accessible au plus grand nombre. Notre démarche peut en effet être généralisée à d'autres patrimoines en danger et, dans ce sens, une documentation associée intégrant les schémas d'identification des entités que nous sommes en mesure de marquer à partir d'un document sera également proposée, pour tous ceux qui souhaiteraient utiliser le CIDOC CRM sur un cas concret lié à la valorisation du patrimoine.

[1] <https://memomines.hypotheses.org>

[2] Licence Creative Commons Attribution – Pas d'Utilisation Commerciale – Partage dans les Mêmes Conditions 4.0 International

[3] <https://opentheso.huma-num.fr/opentheso/api/theso/MemoMines>

[4] <https://www.pop.culture.gouv.fr/>

[5] *Open Knowledge Application Programming Interface*, <https://okapi.inalco.fr/mediascape>

Bibliographie :

Daloz, Amélie et Chaudiron Stéphane. 2019. Méthodologie de conception d'un thésaurus du domaine minier. In B. Jacquemin, M. Ghenima (Éds.), *La numérisation info-documentaire. Actes du 21^e Colloque international sur le Document Numérique* (p. 11-21). Europaia

Doerr, Martin, Gradmann Stefan, LeBoeuf Patrick, Aalberg Trond, Bailly Rodolphe and Olensky Marlies. 2013. *Final Report on EDM – FRBRoo Application Profile Task Force* [Final Report]. Europeana. https://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/EDM_FRBRoo/TaskfoApplication%20Profile%20EDM-FRBRoo.pdf

Kergosien, Eric, Daloz Amélie, Chaudiron Stéphane, Jacquemin Bernard. Le CIDOC CRM, un modèle ontologique pour représenter les connaissances du bassin minier des Hauts de France. Premiers résultats. *6th Hyperheritage International Symposium (HIS6) « La fabrique du patrimoine à l'ère numérique »*, nov 2019, Paris, France.

Le Boeuf Patrick. 2005.. *Functional Requirements for Bibliographic Records (FRBR) : Hype or Cure-All ?*. The Haworth Information Press (Binghamtown, NY), p. 316.

Le Bœuf, Patrick., Doerr Martin, Ore Christian-Emil and Stead Stephen Derek (Éds.). 2015. *Definition of the CIDOC Conceptual Reference Model (Version 6.2.1)*. ICOM/CIDOC Documentation Standards Group, CIDOC CRM Special Interest Group. <http://www.cidoc-crm.org/Version/version-6.2.1>

Les matériaux de recherche ethnographiques à l'épreuve de l'*open data*

Florence Revelin¹

1 : UMR Laboratoire interdisciplinaire Sciences, innovations, sociétés, CNRS, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, université Gustave Eiffel

La numérisation croissante des matériaux de recherche ethnographiques, qu'elle soit native ou produite lors des différentes étapes de traitement, a largement transformé les pratiques de fabrication de la recherche qualitative (Le Marec et Mairesse 2017). Combinées à cette évolution, deux nouvelles orientations réglementaires majeures touchent de près la fabrication de la recherche ethnographique, en lien avec les enjeux du traitement et du devenir des « données ». L'année 2018 marque un tournant avec d'un côté, l'adoption en mai du RGPD (Règlement Général sur la Protection des Données) qui constitue une refonte du cadre réglementaire lié à la protection des données personnelles en renforçant les sanctions en cas de manquements. De l'autre côté, le Plan national pour la Science Ouverte lancé au mois de juillet 2018 traduit à l'échelle nationale une injonction européenne à ouvrir les données de la recherche non soumises à un régime de protection.

Ce triple mouvement de numérisation croissante des matériaux de recherche, de réorganisation de la protection données personnelles et d'ouverture des données de la recherche par leur mise à disposition sur des entrepôts numériques dans le respect des principes *FAIR* (Wilkinson *et al.* 2016) recentre l'attention sur les processus de fabrication de la recherche qualitative et questionne l'évolution des pratiques des chercheurs et de la science produite. Ces évolutions nécessitent d'être appréhendées selon plusieurs plans complémentaires : épistémologique, éthique, juridique, pratique et technique.

Dans le cadre du programme de recherche PARDOQ – *Partage et protection des données qualitatives à l'ère du numérique – expériences, enjeux, stratégies*, nous avons mené entre septembre et décembre 2020 une enquête sur la manière dont ce triple mouvement affecte et interroge les communautés de recherche travaillant sur la base de matériaux ethnographiques, en entrant par le prisme du partage et de l'« ouverture des données » (*open data*). Partant de l'hypothèse que la grande hétérogénéité de ces matériaux qualitatifs est susceptible de mettre en lumière une diversité de situations et de problématiques, nous nous sommes interrogés sur les effets que ce mouvement produit sur les pratiques des chercheur.e.s et sur leurs rapports à leurs matériaux de recherche. L'objectif était de comprendre dans quelles mesures ils.elles intègrent et s'approprient ces enjeux, et quels problèmes ils.elles rencontrent. À partir d'un corpus de 26 entretiens semi-dirigés menés auprès de trois types d'interlocuteurs (des chercheur.e.s et enseignant.e.s-chercheur.e.s pratiquant l'ethnographie, des réseaux d'appui à la recherche et des prescripteur.trices de la science ouverte) notre enquête montre que l'ouverture des données, dans sa forme contemporaine dominante, ne s'inscrit pas dans une dynamique interne à la recherche qualitative en sciences sociales et que la conformation à ces cadres nécessite un travail d'ajustement considérable et multiforme de la part de ces communautés. Une majorité des chercheur.e.s et enseignant.e.s-chercheur.e.s interrogés ne s'approprie pas ou se sent peu concernée par ces injonctions. Néanmoins, ces communautés mettent en place un ensemble de dispositifs (pratiques et éthiques) visant à la fois à protéger leurs sources et à rendre accessible une partie de leurs matériaux de recherche. Elles déploient pour ce faire un large éventail de pratiques à la fois « ancrées » (en lien à la relation de confiance tissée avec les interlocuteur.trices sur le « terrain », souvent dans une perspective de restitution) et « bricolées » (à travers des formes artisanales de partages qui n'épousent pas nécessairement les standards ou modèles de l'*open data*). Par ailleurs, quelle que soit la qualité des ressources externes et réseaux d'appui dont il.elle dispose, le.la chercheur.e (enquêteur.trice) reste dans la pratique l'instance principale d'arbitrage et d'appréciation du statut des matériaux qu'il.elle recueille selon un *continuum* de configurations multiples entre l'« ouvert » et le « fermé ». L'enquête montre paradoxalement que ces questions sont peu discutées les arènes qui concernent ou incluent directement ces chercheur.e.s

enquêteur.trice.s, alors qu'elles sont plus vivement débattues au sein des communautés d'ingénieur.e.s et de réseaux d'appui à la recherche, qui plaident pour une meilleure intégration de ces enjeux dès les phases de conception de la recherche. Il existe également un décalage important entre les pratiques des chercheur.e.s et l'offre d'appui numérique, technique et juridique qui est déployée. Ce décalage relève de divers motifs dépassant la seule question des moyens alloués, parmi lesquels on retrouve une forme d'illisibilité d'un cadre en évolution, des projections différenciées quant aux enjeux et formes du partage, la manière dont les finalités de la recherche intègrent ou non l'objectif d'ouvrir les matériaux ethnographiques dans les phases de conception d'un projet, la difficile identification par les chercheur.e.s de l'offre d'accompagnement et d'appui (services TGIR de la HumaNum, rôle des DPD, services des MSH, etc.) qui tend à se diffuser préférentiellement par effets de réseaux, ou encore la nécessité d'un accompagnement au cas par cas propre au travail sur ces matériaux qualitatifs intriqués dans des enjeux éthiques et contextuels spécifiques. En croisant les retours d'expérience relevant de réseaux d'appui à la recherche et ceux de chercheur.e.s et enseignant.e.s-chercheur.e.s, nous proposons de revenir en détails sur les résultats d'enquête qui caractérisent ces décalages entre un appui à la recherche positionné sur un plan pratique, technique et juridique, alors que les enjeux majeurs pour les chercheur.e.s sont d'ordres éthique et épistémologique.

Enfin, l'enquête suggère que le retrait de ces communautés de recherche qualitative vis-à-vis du mouvement de *l'open data* ne peut être interprété uniquement à l'aune d'un déficit de compétences dans une période de transition. Il concerne un ensemble de paradigmes auxquels elles se réfèrent, soulevant des enjeux épistémologiques plus profonds, liés en particulier à la nature des matériaux de recherche ethnographiques qui, au-delà de leur hétérogénéité, embarquent la relation entre enquêté.e et enquêteur.trice. La collecte de ces « morceaux de réel » (Olivier de Sardan 1995) est imprégnée de subjectivité du regard du/de la chercheur.e, voire d'une forme d'intimité (Mohia 2008) indissociable des connaissances produites, qui rend difficile leur archivage ou ré-usage sous la forme de réductions en unités indépendantes et autonomes telles que le prévoient les entrepôts de données.

Le Marec, Joëlle, et François Mairesse. 2017. *Enquête sur les pratiques savantes ordinaires. Collectionnisme numérique et environnements matériels*. Le bord de l'eau. Lormont. <https://www.decitre.fr/livres/enquete-sur-les-pratiques-savantes-ordinaires-9782356875143.html>.

Mohia, Nadia. 2008. *L'expérience de terrain : pour une approche relationnelle dans les sciences sociales*. La Découverte.

Olivier de Sardan, Jean-Pierre. 1995. « La politique du terrain. Sur la production des données en anthropologie ». *Enquête. Archives de la revue Enquête*, no 1 (octobre) : 71-109. <https://doi.org/10.4000/enquete.263>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, *et al.* 2016. « The FAIR Guiding Principles for Scientific Data Management and Stewardship ». *Scientific Data* 3 (mars) : 160018. <https://doi.org/10.1038/sdata.2016.18>.

Modéliser le feuilleton. Créer et analyser une nouvelle archive patrimoniale avec les méthodes computationnelles

Pierre-Carl Langlais^{1, 2}

1 : EA 1498 Groupe de recherches interdisciplinaires sur les processus d'information et de communication (GRIPIC), université Paris IV – Paris Sorbonne

2 : EA 4209 Représenter, inventer la réalité du romantisme à l'aube du XXI^e siècle, université Paul-Valéry – Montpellier 3

Cette proposition de communication interroge « l'effet des méthodes computationnelles sur la recherche en SHS » et plus particulièrement en histoire culturelle. Elle présente un nouveau cadre méthodologique d'étude du roman-feuilleton de 1830 à 1950.

Le roman-feuilleton est une production emblématique des premières industries culturelles et médiatiques (Dumasy-Queffelec, Queffelec, et Queffelec 1989; Delafield 2016; Stein et Wiele 2019). Des décennies avant l'apparition des serials cinématographiques ou télévisés, le feuilleton introduit des dispositifs de narration sériels suivis par des centaines de milliers de lecteurs et de lectrices (Letourneux 2017). Les études du feuilleton ont été profondément renouvelées par le « tournant culturel » des études de presse ancienne depuis 2000 (M.-E. Thérénty et Vaillant 2001; M. Thérénty 2007; Kalifa *et al.* 2011; M.-È. Thérénty, Langlais, et Schuh 2020). Longtemps occultées au profit des structures économiques, sociales et politiques, les formes et pratiques d'écritures deviennent un terrain de recherche privilégié. Malgré ces efforts de recherche soutenus, l'étude du roman-feuilleton est pénalisée par la masse énorme des corpus : au tournant des XIX^e et XX^e siècles, la plupart des quotidiens nationaux et internationaux publient un voire deux feuilletons chaque jour. Les outils de consultations existants de la presse numérisée ne suffisent pas à saisir ce phénomène culturel global au-delà de quelques grands romans patrimonialisés.

De nouveaux projets français et européens (Europeana Newspaper, Newseye, Numapresse, Impresso) élaborent aujourd'hui des infrastructures de recherche « augmentée » de la presse ancienne dans la continuité de plusieurs travaux précurseurs dans le monde anglo-saxon (Cordell 2015; Mussell 2012; Underwood 2019). Le projet Numapresse s'est spécialisé dans la détection et l'analyse des genres journalistiques, avec le développement de modèles de classification « générationnels » (1840-1860, 1860-1880...). Ces modèles ont été appliqués à l'ensemble des archives de presse quotidienne nationale numérisées par Gallica. Les résultats de la classification permettent de construire rapidement des sous-corpus tels que l'ensemble des pages de cinéma de l'entre-deux-guerre. Ils font également émerger de nombreuses tendances pertinentes pour l'histoire culturelle longue de la presse comme, la spécialisation thématique de certains quotidiens, l'hebdomadairisation des formes journalistiques au début du XX^e siècle ou le développement précoce de certains genres, avant leur transformation en rubrique (Langlais 2021).

Le roman-feuilleton ne faisait pas initialement partie des trois genres journalistiques privilégiés par Numapresse (reportage, cinéma et sport) mais s'est finalement imposé comme un terrain privilégié pour des collaborations inédites avec les bibliothèques. Numapresse est depuis 2020, partenaire du futur laboratoire en humanités numériques de la BNF : la détection des romans-feuilletons est l'un des principaux enjeux de cette collaboration avec l'accès aux archives sous droits.

Ce projet mobilise plusieurs techniques en intelligence artificielle appliquée aux archives patrimoniales numérisées :

- Reconnaissance spatiale : L'objet éditorial du feuilleton est reconnu à partir d'une segmentation automatisée de la page du journal. Pour l'instant, notre approche privilégie un algorithme très rapide de détection de zones géométriques (la marge d'espace entre le feuilleton et le haut du journal) (voir l'illustration n° 1). Nous explorons également l'utilisation de modèles d'apprentissage profond (deep learning) entraînés à partir de l'annotation des zones du journal correspondant au feuilleton.

- Classification des textes : Les feuillets ne publient cependant pas que des romans en série mais aussi de nombreux autres textes : critiques théâtrales, chroniques scientifiques, billets d'humeur. L'identification du roman-feuilleton repose sur les modèles de détection des genres journalistiques développés par Numapresse (voir l'illustration n° 2).
- Identification du paratexte : Dans un dernier temps, le projet procède à l'identification des titres, des signatures et d'autres éléments paratextuels à partir d'une analyse croisée du vocabulaire et de la forme visuelle du texte (soit à partir d'une analyse géométrique ou d'une nouvelle classification par deep learning). L'identification des titres est notamment déterminante pour reconstituer la série des feuillets dans leur continuité.

Cette approche croisée a été pour l'instant appliquée avec succès à un titre de presse parisien, *Le Pays*. D'ici la conférence elle sera probablement généralisée à d'autres quotidiens.

En revenant sur cette collaboration, nous développerons une réflexion méthodologique sur les transformations de l'archive de presse induite par le déploiement des outils d'analyse computationnelle et/ou par la création de corpus de recherche (Bode 2018). La détection du roman-feuilleton fait en effet émerger des éditions voire des œuvres qui n'existaient pas. Plusieurs projets de recensions ont déjà mis en évidence qu'une large partie des textes publiés en feuillets (en particulier les plus courts) ne font pas l'objet d'une publication en monographie. Les méthodes computationnelles produisent des œuvres qui n'avaient jamais été cataloguées. Dans le cadre de Numapresse nous avons commencé à réfléchir à de nouvelles interfaces de lecture qui, tout en brisant la structure d'origine de l'archive de presse, restitue la continuité textuelle des feuillets.

Dans la dernière partie de notre présentation, nous présenterons plusieurs nouveaux champs de recherche ouverts par l'accès sans précédent à un grand corpus de feuillets :

- Recontextualisation de l'émergence du roman-feuilleton : la détection des reprises d'un corpus de 16 000 romans issus de Gallica montre que la presse commence à publier de plus en plus fréquemment des textes littéraires et des bonnes feuilles avant la date de naissance officielle du roman-feuilleton en 1836 avec la publication de *la Vieille fille* de Balzac (illustration n° 3). Nous avons ainsi pu identifier un court feuilleton d'Alphonse Karr publié en deux livraisons dans le *Figaro* en 1835.
- Organisation et structuration du feuilleton dans l'écosystème journalistique : la classification automatisée de l'ensemble du journal permet de situer le feuilleton par rapport à un espace éditorial très normé (illustration n° 4). Nous explorerons notamment l'hypothèse que les proximités spatiales (par exemple entre le feuilleton romanesque et le feuilleton boursier qui prend sous sa suite en dernière page) se traduisent par des interactions intertextuelles.
- Feuilletonisation/défeuilletonisation : Nous avons commencé à procéder à une comparaison systématique entre les versions éditées des feuillets et leur version publiées dans la presse en détournant des outils de détection automatisés des reprises de textes. Lorsque la publication intervient elle introduit des changements substantiels : les comparaisons semi-automatiques du projet Numapresse mettent en évidence des agencements significatifs avec des coupes, des rajouts, des changements de titres, ce qui suggère que les attentes de lecture d'une monographie ne sont pas les mêmes que celles d'une publication en feuilleton.
- Recyclage des feuillets : l'analyse préliminaire des reprises de feuilleton entre les journaux met en évidence une circulation fréquente des feuillets à succès d'un journal à l'autre. En particulier, plusieurs quotidiens des années 1880 et 1890 reproduisent de nombreux feuillets de la période classique des années 1840 et 1850. Dans une étude en cours d'écriture nous décrivons ce phénomène de patrimonialisation par analogie avec la daytime TV américaine des années 1980 et 1990 qui remet en circulation des séries alors vieillissantes.
- Évolution des thèmes et des formes d'écriture du feuilleton. Nous souhaitons mobiliser les méthodes appliquées par les cultural analytics anglo-saxonnes à de grands corpus de romans : détection de genres romanesques, de personnages, de lieux et de sujets. Cet axe encore préliminaire s'inscrit dans la perspective d'un enrichissement des nouvelles archives de feuilleton.

Bibliographie

Bode, Katherine. 2018. *A World of Fiction : Digital Collections and the Future of Literary History*. University of Michigan Press.

Cordell, Ryan. 2015. « Reprinting, Circulation, and the Network Author in Antebellum Newspapers ». *American Literary History* 27 (3) : 417-45. <https://doi.org/10.1093/alh/ajv028>.

Delafield, Catherine. 2016. *Serialization and the Novel in Mid-Victorian Magazines*. Routledge. <https://doi.org/10.4324/9781315608440>.

Dumasy-Queffélec, Lise, Lise Queffelec, et Lise Queffélec. 1989. *Le roman-feuilleton français au XIX^e siècle*. Presses universitaires de France.

Kalifa, Dominique, Philippe Régnier, Marie-Ève Thérénty, et Alain Vaillant, éd. 2011. *La civilisation du journal : Histoire culturelle et littéraire de la presse au XIX^e siècle*. Paris : Nouveau Monde Éditions.

Langlais, Pierre-Carl. 2021. « Classified News, Redefining the history of newspaper genre with supervised models ». In *Digital Newspaper : a new Eldorado for the historians*. De Gruyter.

Letourneux, Mathieu. 2017. *Fictions à la chaîne – Littératures sérielles et culture médiatique*. Paris : Le Seuil.

Mussell, James. 2012. *The Nineteenth-Century Press in the Digital Age*. Springer.

Stein, Daniel, et Lisanna Wiele, éd. 2019. *Nineteenth-Century Serial Narrative in Transnational Perspective, 1830s–1860s : Popular Culture–Serial Culture*. Palgrave Studies in Nineteenth-Century Writing and Culture. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-15895-8>.

Thérénty, Marie-eve. 2007. *La Littérature au quotidien. Poétiques journalistiques au XIX^e siècle*. Paris : Le Seuil.

Thérénty, Marie-Ève, Pierre-Carl Langlais, et Julien Schuh. 2020. *Fake News et viralité avant Internet. Les lapins du Père-Lachaise et autres légendes médiatiques*. Paris : Éditions du CNRS.

Thérénty, Marie-Eve, et Alain Vaillant, éd. 2001. *1836 : l'an 1 de l'ère médiatique. Etude littéraire et historique du journal La Presse d'Emile de Girardin*. Paris : Nouveau Monde Éditions.

Underwood, Ted. 2019. *Distant Horizons : Digital Evidence and Literary Change*. University of Chicago Press.

Moyen Âge et performances (MAP) : (modéliser un) répertoire de performances médiévales

Simon Gabay¹

1 : Université de Genève

Les spécialistes du théâtre médiéval ont depuis longtemps pour tradition de cataloguer les représentations. Le XIX^e siècle voit l'apparition des plus anciens répertoires, notamment pour les mystères [1] et le théâtre comique [2] – deux catégories sensées couvrir l'essentiel de la production. Cette pratique s'est perpétuée, et l'on dispose désormais de listes plus ou moins détaillées pour des villes (comme Compiègne [3]), des provinces françaises (comme la Savoie [4] ou le Dauphiné [5]) ou francophones (comme les Pays-Bas méridionaux [6]). Des projets similaires existent bien évidemment pour les autres domaines linguistiques, et notamment pour les îles britanniques (et non uniquement l'Angleterre, malgré le titre) avec le monumental projet des *Records of Early English Drama* (REED) [7].

Répertorier la pratique

Cet attachement à noter les représentations tient à la nature particulière du théâtre médiéval. La documentation se retrouve éparpillée dans les divers fonds : délibérations consulaires, échevinales ou capitulaires, divers registres de comptabilité... Il est donc difficile de la retrouver, contrairement à la documentation moderne qui est souvent attachée à des institutions/bâtiments comme le théâtre du Marais [8] ou la Comédie française [9] – les érudits du XIX^e (comme Piolin pour le Maine [10] ou Brouchoud pour Lyon [11]) restent ainsi l'une des meilleures sources du médiéviste. Si l'on ajoute à cela la complexité de déchiffrement pour des raisons de conservation comme de paléographie, on comprend aisément l'intérêt de ces répertoires.

De tels objets sont nécessairement d'excellents candidats à la numérisation. Une tentative pionnière a eu lieu en France avec la base *Théâtre et Performances en France au Moyen Âge* [12]. Cette dernière n'est cependant plus accessible, et le projet le plus actif est désormais la version numérique de REED lancée en 2017 : *eREED* [13]. Des projets équivalents existent pour des périodes plus tardives du domaine francophone comme la base *CESAR* pour l'Ancien Régime [14].

Choix techniques et modélisation

On peut observer deux politiques quant à la conception de ces bases. D'une part on trouve les bases de données relationnelles de type SQL comme *Théâtre et Performances* et *CESAR*, et de l'autre celles entièrement encodées en XML-TEI comme le corpus REED – un important travail de rétro-conversion ayant été mené [15]. Ce choix de la TEI, qui était déjà dix ans auparavant une recommandation de Matthieu Bonicel [16], précurseur des recherches au croisement des études numériques et théâtrales [17], se trouve être devenu une évidence avec l'évolution des recommandations du consortium [18] et le travail mené sur l'encodage des documents d'archive comme les registres de comptabilité [19], qui sont une des principales sources de l'histoire théâtrale [20].

L'encodage XML-TEI est plus en phase avec une approche philologique de la donnée, qui replace le document historique au centre plutôt que de tisser une multitude de liens : l'information n'est qu'une surcouche d'annotation sur la transcription, et le lien direct avec la source d'archive n'est ainsi pas rompu. Un tel modèle est plus à même de répondre aux exigences des médiévistes, car les spécificités du Moyen Âge font que, comme nous allons le voir, l'édition doit précéder l'extraction d'informations toujours

glissantes, car rétives à nos catégories d'analyses, notamment d'un point de vue linguistique et métalinguistique.

Il est à noter que le choix d'un encodage en TEI n'obère en rien l'interopérabilité avec d'autres bases, notamment via le RDF – le passage d'une technologie à l'autre ayant été fait plusieurs fois [21]. L'utilisation attentive de `@xml:id` sur les éléments clefs, notamment les entités nommées, devrait permettre de reconstituer efficacement des triplets afin de rendre ces données compatibles avec celles d'autres projets ayant fait des choix de modélisation différents.

Une approche philologique de la donnée

Nous avons « rétrospectivement voulu identifier comme “théâtre” du Moyen Âge » une chose « qui n'était pas – ou pas encore – conceptualisée comme tel » [22], l'époque médiévale ayant tendance à (ré)inventer ses propres catégories [23]. Pour prendre le cas des mystères, Ch. Mazouer explique bien que :

« Ce que nous appelons mystère – l'appellation est authentiquement médiévale, mais les *incipit* des manuscrits proposent bien d'autres titres pour ces textes dramatiques – n'a jamais été défini comme genre au Moyen Âge, mais nous le considérons comme genre dramatique religieux, avec un groupe homogène de textes, bien constitué et bien défini. » [24]

À l'inverse, des occurrences du terme « mystère » ne relèvent pas de ce que nous considérons aujourd'hui comme un « mystère », comme ici à Rouen :

Les rues avaient été « mieux tendues qu'elles ne furent oncques le jour du Sacrement. Et y avait à la porte Cauchoise draps où étaient les armes de France et d'Angleterre... et alla le roi à l'Eglise Notre-Dame... et criaient le peuple nouel, tellement que le roy dist que l'on cessât pour la noise que l'on lui faisait. Dans les rues se jouaient les Mystères et les regarda le roy ». [25]

Il est plus que probable qu'il s'agisse ici de tableaux vivants tenus dans le cadre d'une entrée royale – un sens de « mystère » qui ne peut être saisi qu'à la lecture du document complet.

Conclusion

La nature difficilement saisissable des formes et l'indétermination médiévale du métalangage littéraire plaide pour une traçabilité maximale de la source. L'objectif doit ainsi être la constitution d'une base de performances donnant à ce dernier terme une définition assez large pour recueillir l'essentiel des cas, mais aussi, corollaire logique, les moyens philologiques pour le chercheur de redéfinir des sous-corpus à l'intérieur de cet ensemble. Offrir un modèle orienté vers l'édition du document plutôt que vers une interprétation de l'information qu'il est supposé contenir devrait permettre de contenter les besoins contradictoires du médiéviste, tels que décrits par P. Zumthor :

« Nous ressentons de plus en plus clairement l'urgence de ne point couper la sémiotique du signe de l'interprétation du sens ; mais nous ne pouvons ignorer la tendance, profondément enracinée dans la pratique philologique (que rien encore n'a permis de dépasser), à refuser ce qui se détache de la lettre, à nier que le “reste” soit analysable, ou même digne de considération critique. » [26]

Data

Site : <https://maperformance.github.io>

Données : <https://github.com/MAPerformance>.

Bibliographie

[1] Louis Petit de Julleville, *Les Mystères*, 2 t., Paris, 1880.

[2] Id., *Histoire du théâtre en France : répertoire du théâtre comique en France au Moyen Age*, Paris, 1886.

- [3] Naomi Kanaoka, « La vie théâtrale à Compiègne entre 1450 et 1550 », Bibliothèque de l'École des chartes, 164.1 (2006).
- [4] Jacques Chocheyras, *Le Théâtre religieux en Savoie au XVIe siècle*, 1 t., Genève, 1971.
- [5] Id., *Le Théâtre religieux en Dauphiné du Moyen âge au XVIIIe siècle*, 1 t., Genève, 1975.
- [6] Katell Lavéant, *Théâtre et culture dramatique d'expression française dans les villes des Pays-Bas méridionaux (XVe-XVIe siècles)*, thèse, Amsterdam, UvA, 2007.
- [7] Alexandra F. Johnston, *Records of Early English Drama*, Toronto, Buffalo, 1976-2016.
- [8] S. Wilma Deierkauf-Holsboer, *Le Théâtre du Marais*, 2 t., Paris, 1954.
- [9] Sylvaine Guyot et Jeffrey S. Ravel, *Databases, Revenues, & Repertory : The French Stage Online, 1680-1793*, MIT Press, Boston, 2020.
- [10] Paul Piolin, *Le Théâtre chrétien dans le Maine au cours du moyen âge... Mamers*, 1891.
- [11] Claude Brouchoud, *Notice sur les origines du théâtre de Lyon : mystères, farces et tragédies, troupes ambulantes*, Molière, Paris, 1865.
- [12] Nadine Henrard, Darwin Smith, Matthieu Bonicel et Gabriella Parussa, *Théâtre et Performances en France au Moyen Âge*, 2009, URL : <http://arnoul.vjf.cnrs.fr/theatre> (disparu).
- [13] Sally-Beth Maclean, *Records of Early English Drama (REED) online*, 2017, URL : <https://reed.library.utoronto.ca>.
- [14] Jeffrey Ravel, Barry Russell et David Trott, *CESAR – Calendrier Electronique des Spectacles sous l'Ancien Régime*, URL : <https://cesar.huma-num.fr/cesar2>.
- [15] James Cummings, « Liturgy, Drama, and the Archive : Three conversions from legacy formats to TEI XML », *Digital Medievalist*, 2 (2006).
- [16] M. Bonicel, « Vers un répertoire en ligne des sources pour l'étude du théâtre médiéval », *SIBMAS*, Vienne, Autriche, 2006.
- [17] Id., *Arts et gens du spectacle à Avignon à la fin du Moyen Âge (1450-1550), une approche numérique*, thèse, Paris, Paris I, 2014.
- [18] J. Cummings, « Liturgy, Drama, and the Archive »..., §44.
- [19] Kathryn Tomasek et Syd Bauman, « Encoding Financial Records for Historical Research », *Journal of the Text Encoding Initiative – 6* (2013).
- [20] M. Bonicel, « Les modes de financement public des performances à Avignon à la fin du Moyen Age », *Studies in Early Modern France*, 13 : Spectacle (mars 2010), p. 31-40.
- [21] Michael O. Jewell, « Semantic screenplays : Preparing TEI for Linked Data », dans *Digital Humanities Conference 2010 – DH2010*, Londres, Angleterre, 2010.
- [22] Jelle Koopmans et D. Smith, « Un théâtre français du Moyen Âge ? », *Médiévales*, 59 (2010), p. 8.
- [23] Simon Gabay, *L'Acteur au Moyen Âge : L'Histrion et ses avatars en Occident de saint Augustin à saint Thomas*, thèse, Amsterdam, UvA, 2015.
- [24] Charles Mazouer, « Les mystères des XVe et XVIe siècles, carrefour des genres ? », *Eidôlon*, 97 (2011), p. 31.
- [25] Edouard Gosselin, *Recherches sur les origines et l'histoire du théâtre à Rouen avant Pierre Corneille*, Rouen, 1868, p. 13.
- [26] Paul Zumthor, *Parler du Moyen Âge*, Paris, 1980, p. 23.

POPP. Projet de reconnaissance optique des caractères des recensements de la population parisienne

Sandra Brée¹, Thierry Paquet², François Merveille³, Thomas Constum², Nicolas Kempf²

1 : UMR 5190 Laboratoire de recherche historique Rhône-Alpes (LARHRA)

2 : EA 4108 LITIS, INSA de Rouen, université de Rouen

3 : GED Campus Condorcet

POPP : un projet de science ouverte

Le projet POPP[1], dirigé par Sandra Brée (historienne, démographe CNRS au LARHRA et responsable de la partie SHS du projet), François Merveille (Bibliothécaire au GED-Campus Condorcet) et Thierry Paquet (Professeur d'informatique à l'université de Rouen et responsable de la partie informatique du projet) a obtenu un financement du GIS Collex-Persee [2] pour créer une vaste base de données (12 millions de personnes) basée sur les recensements nominaux de la population de Paris de 1926, 1931, 1936 et 1941 qui sont les seuls recensements de la population parisienne existants avant la fin du XX^e siècle. POPP est un projet novateur qui constituera une avancée majeure pour la recherche historique. En effet, la reconnaissance optique de caractères permet de lire des documents dactylographiés et même manuscrits, tels que les recensements parisiens, et donc de créer des bases de données quantitatives exceptionnellement importantes sans avoir à relever toutes les informations à la main.

Outre l'intérêt méthodologique promis par les progrès de la reconnaissance optique des caractères, la base POPP permettra de connaître la population parisienne pendant l'entre-deux-guerres dans son intégralité. L'analyse quantitative et statistique des populations est rendue particulièrement difficile en histoire en raison du manque de bases de données. L'indexation des recensements parisiens de 1926, 1931, 1941 et 1946 sera l'occasion d'une avancée majeure dans la connaissance de la population urbaine européenne, jusqu'alors peu étudiée.

Sources et méthode

Les recensements parisiens de l'entre-deux-guerres sont numérisés et mis en ligne sur le site des Archives de Paris. Les quatre recensements sont basés sur le même modèle, ce qui rend l'apprentissage plus aisé pour la machine. Chacun d'entre eux est composé de 50 000 vues, et donc environ 20 000 vues à travailler.

Les informations contenues dans les recensements de population permettent de connaître la composition de chaque ménage dans les 80 arrondissements parisiens (répartis en 20 arrondissements). Pour chaque individu, les informations suivantes sont fournies : nom, prénom, année et lieu de naissance, nationalité, état civil, situation du ménage (chef, épouse, enfant, travailleur domestique, etc.), ainsi que la profession et l'adresse.

L'équipe du laboratoire LITIS EA 4108 (Rouen, France) a développé un outil de lecture optique de plus en plus performant pour extraire des informations imprimées d'annuaires historiques afin de constituer des bases de données historiques. Ils ont déjà travaillé sur des données financières et travaillent maintenant pour la première fois sur des données de population. Dans cette continuité, le projet POPP est un véritable défi à relever puisque les recensements de population ne sont pas tapuscrits, comme les données boursières, mais remplis à la main.

Les premiers tests sont cependant concluants, d'une part parce que les recherches de Sandra Brée sur la population parisienne (Brée, 2015, 2017) lui ont permis de créer des bases de données (professions, noms, lieux) qui permettent de paramétrer les algorithmes de détection optique, et d'autre part parce que l'équipe

de Thierry Paquet possède également une expertise dans la reconnaissance de l'écriture manuscrite, et dans la mise en place de chaînes de traitement et de production de données structurées comme les archives de presse.

Concrètement, une fois les caractères lus optiquement, les données sont compilées dans un site web fonctionnel où les informations reconnues par la machine pourront être modifiées manuellement (grâce à une plateforme web collaborative). La machine est ensuite régulièrement « alimentée » en nouveaux termes afin d'enrichir de plus en plus son vocabulaire et de traiter d'autres parties du corpus. Une fois ce long travail terminé, une base de données sera créée pour être lisible sous un logiciel statistique.

Présentation

Nous proposons, pour le colloque Humanistica 2021, de présenter le projet POPP qui a débuté en septembre 2020. Nous commencerons par un état des lieux sur la recherche en histoire quantitative des populations et en démographie historique pour montrer l'intérêt de créer des bases de données informatiquement, notamment pour la population urbaine, et surtout à partir de la fin du XIX^e siècle car les villes sont très denses et les relevés particulièrement longs.

La sources, les recensements de la population parisienne, sera ensuite présentée ; ainsi que les problématiques et questions de recherche qu'elle permettra d'aborder.

Nous ferons ensuite un point sur les techniques de reconnaissance automatique de l'écriture manuscrite mobilisées pour les besoins du projet (Soulard *et al.*, 2019; Coquenot *et al.*, 2020). D'une part nous présenterons les différentes étapes nécessaires pour permettre à la machine d'apprendre une représentation optique des écritures. Les phases d'annotation des données et d'optimisation par la machine seront détaillées. D'autre part nous montrerons comment le modèle optique peut être complété par un modèle de langage contextuel adapté à chaque élément tabulaire, pour finalement permettre une lecture dirigée par la sémantique des tableaux de recensement.

La plateforme spécifique, PIVAN-POPP, développée par le LITIS pour la correction manuelle des données ocrisées par la machine sera ensuite présentée. Cette plateforme web constitue un élément essentiel pour la collaboration entre les partenaires en proposant une interface de visualisation IIIF avancée en mode image, et un éditeur/correcteur de transcriptions synchrone du visualiseur. Cette plateforme permet de réaliser des annotations massives indispensables aux techniques d'apprentissage machine, et de contrôler les résultats de reconnaissance automatique. Les liens entre l'équipe informatique du LITIS et les chercheurs en SHS et documentalistes pour la construction de la base de données seront alors concrètement explicités.

Références

Brée S. 2017, Paris l'inféconde, La limitation des naissances en région parisienne au XIX^e siècle, Paris, INED.

Brée S., 2015, La population de la région parisienne au XIX^e siècle, Démographie et Société n° 6

Coquenot D., Chatelain C., Paquet T., 2020, Recurrence-free unconstrained handwritten text recognition using gated fully convolutional network, 17th International Conference on Frontiers in Handwriting Recognition (ICFHR), p. 19-24.

Soullard Y., Swaileh W., Tranouez P., Paquet T., Chatelain C., 2019, Improving Text Recognition using Optical and Language Model Writer Adaptation, ICDAR, p. 1175-1180, Sydney.

[1] popp.hypotheses.org/

[2] <https://www.collexpersee.eu/projet/popp/>

Préserver et diffuser le patrimoine textile européen grâce aux technologies du Web sémantique. La réutilisation des données patrimoniales par le projet SILKNOW

Marie Puren¹, Pierre Vernus¹

1 : UMR 5190 Laboratoire de recherche historique Rhône-Alpes (LARHRA)

Le projet H2020 SILKNOW (Silk heritage in the Knowledge Society : from punched card to Big Data, Deep Learning and visual/tangible simulations) a pour but d'utiliser les technologies numériques pour faire connaître et valoriser le patrimoine soyeux européen (XV^e-XIX^e siècle). Ce patrimoine fragile et menacé – en témoignent par exemple les difficultés rencontrées par le Musée des Tissus à Lyon qui a été menacé de fermeture (Cf. Rykner, D (2015). « Les musées des Tissus et des Arts décoratifs de Lyon, des collections exceptionnelles en danger ». La Tribune de l'Art. <https://www.latribunedelart.com/les-musees-des-tissus-et-des-arts-decoratifs-de-lyon-des-collections-exceptionnelles-en-danger>) peut en effet grandement bénéficier des apports des nouvelles technologies numériques, plus particulièrement celles offertes par le Web sémantique, en matière de valorisation et de diffusion des données patrimoniales. D'une part, ces technologies permettent d'apporter une réponse aux difficultés rencontrées par des institutions patrimoniales petites en taille, et qui manquent de moyens financiers, humains et techniques pour valoriser des collections riches, mal connues du grand public – suivant le nouvel adage « ce qui n'est pas en ligne n'existe pas » (à ce sujet, on pourra lire : Ossensbach, G. (2015), « If it's not online, it doesn't exist ». IJHE, n° 5(1), p. 80-82). D'autre part, les grandes institutions patrimoniales généralistes, disposant de moyens plus conséquents, ne mettent pas toujours l'accent sur leurs collections textiles qui restent alors méconnues.

SILKNOW a créé un moteur de recherche exploratoire utilisant des graphes de connaissances (« knowledge graphs ») qui représentent les données contenues dans les catalogues d'institutions patrimoniales, et qui suivent une ontologie étendant le CIDOC Conceptual Reference Model ou CIDOC-CRM. Dans cette intervention, nous proposons de présenter :

- les différentes étapes de ce travail qui ont permis d'aboutir à la création de l'ontologie SILKNOW,
- l'ontologie elle-même, en mettant notamment l'accent sur les nouvelles classes et propriétés créées en collaboration avec des experts du domaine (historiens de l'art et du textile), des experts des humanités numériques et des informaticiens,
- le moteur de recherche qui utilise les graphes de connaissance suivant l'ontologie SILKNOW.

Cette présentation a non seulement pour but de présenter les résultats d'un projet de recherche, mais elle a aussi pour objectif de démontrer l'intérêt des technologies du Web sémantique pour valoriser les données patrimoniales, et de manifester l'importance d'une collaboration étroite entre experts du domaine et informaticiens dans ce domaine.

Dans une première partie, nous présenterons la démarche que nous avons adoptée pour développer cette ontologie. Nous montrerons pourquoi nous avons utilisé le CIDOC CRM, comment nous l'avons utilisé pour modéliser les métadonnées décrivant les objets conservés dans les collections patrimoniales, et quelles ont été les limites de ce modèle. Les classes et propriétés sélectionnées pour l'ontologie SILKNOW sont publiquement accessibles et documentées via OntoMe (Ontology Management Environment), un système de gestion d'ontologie développé par le Pôle Histoire Numérique du LARHRA.

Dans une deuxième partie, nous expliquerons comment nous avons exploité la flexibilité inhérente au CIDOC CRM, qui permet d'ajouter de nouvelles sous-classes et sous-propriétés pour exprimer des relations et propriétés plus spécifiques, sans modifier la structure de base du modèle. Nous donnerons quelques

exemples parmi les 23 nouvelles classes et 12 nouvelles propriétés CIDOC CRM que nous avons créées pour modéliser la sémantique complexe des données décrivant le processus de création et de production des textiles en soie. Ces nouvelles classes et propriétés, avec leurs définitions, sont accessibles au public via Ontome : <http://ontologies.dataforhistory.org/namespace/36>.

Dans une troisième partie, nous présenterons le moteur de recherche ADASilk, afin de montrer la puissance des technologies du Web sémantique en matière de valorisation du patrimoine. Nous présenterons notamment quelques scénarios d'utilisation, plus particulièrement centrés sur la découverte d'objets représentant une technique de tissage particulière, ou appartenant à des collections jusqu'alors inaccessibles sur le Web.

Bibliographie

ADASilk : <https://ada.silknow.org/fr>

Francesco Beretta. (s.d.) « A challenge for historical research : making data FAIR using a collaborative ontology management environment (OntoME) ». *Semantic Web – Interoperability, Usability, Applicability*, IOS Press, sous presse.

Le Boeuf, P., Doerr, M., Ore, C., Stead, S. (2015). *Definition of the CIDOC Conceptual Reference Model, Produced by the ICOM/CIDOC Documentation Standards Group, Suite by the CIDOC CRM Special Interest Group Version 6.2.1*. http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_6.2.1.pdf.

Marden, J., Li-Madeo, C., Whysel, N., & Edelstein, J. (2013). « Linked open data for cultural heritage ». *Proceedings of the 31st ACM International Conference on Design of Communication – SIGDOC '13. the 31st ACM international conference*. <https://doi.org/10.1145/2507065.2507103>

Portalés, C., Sebastián, J., Alba, E., Sevilla, J., Gaitán, M., Ruiz, P., & Fernández, M. (2018). « Interactive Tools for the Preservation, Dissemination and Study of Silk Heritage—An Introduction to the SILKNOW Project ». *Multimodal Technologies and Interaction*, 2(2), 28. <https://doi.org/10.3390/mti2020028>

Schleider, T. & Troncy, R. (2020). « Exploring the European Silk Cultural Heritage through the SILKNOW Knowledge Graph ». *Weaving Europe : Online International Conference on Silk heritage and Digital Technologies*.

SILKOW Ontology (2020). Ontome. <http://ontome.dataforhistory.org/profile/10>

SILKNOW Website : <https://silknow.eu/>

Saisir le buzz politique sur Internet : le cas Juan Branco autour de son livre *Crépuscule*

Amar Lakel¹, Franck Cormerais¹

1 : UR 4426 Médiation, information, communication, art (MICA), université Bordeaux-Montaigne

« Crépuscule, Macron et les oligarques. L'enquête-vérité. », c'est sur ce titre qu'une affiche noire sur bandeau jaune inondait les réseaux sociaux avec une vitesse et une répétition propre à la viralité numérique. L'auteur, Juan Branco, est de cette génération Y dite digital native, rompue aux réseaux sociaux. En quelques semaines, il devient un phénomène médiatique en apparaissant sur de très nombreux posts de réseaux comme Twitter, Facebook, YouTube. Des proches montent des groupes de « soutien » et de relais, le like et le partage assurent le « buzz » qui amène ses contenus à être visionnés des millions de fois. L'auteur répète à qui veut l'entendre que la puissance « médiatico-politique » a décidé de le faire taire, de le boycotter voire de l'ostraciser. Il en appelle au « peuple », il a une vérité à nous révéler sur la tyrannie secrète qui nous gouverne. Au nom de son invisibilité des médias traditionnels, il réveille les médias numériques s'assurant une visibilité massive et une vente de son livre qui le classe en tête des ventes durant plusieurs semaines. C'est donc sous le ton du pamphlet politique que l'auteur, dans une posture tribunicienne, organise une promotion de sa personne et de ses révélations au cœur d'un dispositif digital manifestant une maîtrise des codes sémiologiques du marketing 2.0. Juan Branco, Alain Soral, Etienne Chouard, mais aussi Michel Onfray sont les nouveaux intellectuels de la scène numérique contestataire. Blog, chaînes YouTube, omniprésence sur Twitter, nous voyons émerger une nouvelle figure de « l'intellectuel » dont les similarités laissent augurer la fabrication d'une nouvelle figure.

S'appuyant sur une méthodologie originale de constitution de « grand corpus numérique », nous avons utilisé les méthodes digitales mixtes (Cormerais et Lakel, 2019, Lakel & Le Deuff, 2017) pour analyser la stratégie de ce qui nous est apparu comme un parangon des nouvelles formes d'intervention. Parangon dont la fulgurance et la vitesse de production de contenus sur Internet manifestent un ensemble de caractéristiques révélatrices d'une économie nouvelle de la notoriété. Cette économie du « buzz » médiatique est un des défis paradigmatiques des humanités numériques. Vélocité, volume et variété, le buzz sur internet nécessite un agencement numérique radicalement nouveau source de nombreuses innovations dans la communauté des chercheurs en SHS. Le projet MyWebIntelligence [1] développé au sein de l'équipe E3D du MICA est aujourd'hui une de ces propositions. Sur une période allant du 01 octobre 2018 au 02 juin 2019, 2810 pages web sur 602 sites web qui ont participé à la controverse politique appuyée sur les réseaux sociaux et notamment 627 vidéos Youtube sur 286 chaînes. Notre analyse s'appuie sur une méthode d'analyse de contenus assistée par des algorithmes de classification pour dégager la structure des prises de position au sein des d'une controverse numérique. L'analyse cartographique des liens de citation vient compléter cette méthodologie de dévoilement d'une structure médiatique qui fut au service de ce buzz.

Nous réfléchissons dans une première partie à la corrélation entre économie médiatique et forme sémiotique d'intervention de l'intellectuel dans l'espace public médiatisé. L'arrivée d'internet comme nouveau média avait plutôt favorisé la figure de l'influenceur à celui de l'intellectuel tant semblaient manquer à l'amateur numérique les caractéristiques d'expertise et de sérieux nécessaire à cette appellation. Pourtant, notre thèse est de montrer dans une seconde partie qu'une génération d'influenceur numérique tente de faire émerger une figure hybride que nous définirons comme l'intellectuel numérique. Cette partie basée sur une étude de cas nous amène, par cette redéfinition, à ouvrir, dans une troisième partie, le champ à l'ensemble des espaces de controverse web et web 2.0 qui sont autant d'espaces qui assurent la co-construction assumée des nouveaux influenceurs de l'espace public. Dans ce cadre, apparaît la pertinence des Etudes digitales et de ses méthodes digitales hybrides (NLP, Classification, SNA) pour appréhender les espaces du web afin de faire émerger des objets de recherche originaux.

Bibliographie indicative

- Boyadjian, J., Olivesi, A., & Velcin, J. (2017). Le web politique au prisme de la science des données. *Réseaux*, n° 204(4), 9-31.
- Franck Cormerais, Olivier Le Deuff, Amar Lakel et David Pucheu, « Les SIC à l'épreuve du digital et des Humanités : des origines, des concepts, des méthodes et des outils », *Revue française des sciences de l'information et de la communication* [En ligne], 8 | 2016, mis en ligne le 23 mars 2016, consulté le 05 juillet 2019. URL : <http://journals.openedition.org/rfsic/1820> ; DOI : 10.4000/rfsic.1820
- Fabiani, J.-L. (2007). Disputes, polémiques et controverses dans les mondes intellectuels. *Mil neuf cent. Revue d'histoire intellectuelle*, n° 25(1), 45-60.
- Jeanneret, Y., & Souchier, E. (2005). L'énonciation éditoriale dans les écrits d'écran. *Communication & Langages*, 145(1), 3-15.
- Lemieux, C. (2007). À quoi sert l'analyse des controverses ? *Mil neuf cent. Revue d'histoire intellectuelle*, n° 25(1), 191-212.
- Muji, D. A., Bovet, A., Gonzalez, P., & Terzi, C. (2007). De la sociologie à l'analyse de discours, et retour. *Réseaux*, n° 144(5), 267-277.
- Ory, P., & Sirinelli, J.-F. (2004). *Les intellectuels en France : De l'affaire Dreyfus à nos jours*. Paris : Perrin.
- Poell, T., Abdulla, R., Rieder, B., Woltering, R., & Zack, L. (2016). Protest leadership in the age of social media. *Information, Communication & Society*, 19(7), 994-1014. <https://doi.org/10.1080/1369118X.2015.1088049>
- Rieder, B. (n.d.). *Membranes numériques : des réseaux aux écumes*. 11.
- Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). « From ranking algorithms to “ranking cultures” : Investigating the modulation of visibility in YouTube search results ». *Convergence*, 24(1), 50-68. <https://doi.org/10.1177/1354856517736982>
- Rueff, J. (2014). Penser le web politique avec Jacques Rancière. *tic&société*, (Vol. 8, n° 1-2). <https://doi.org/10.4000/ticetsociete.1423>
- Winock, M. (2014). *Le siècle des intellectuels* (éd. revue et augmentée, vol. 1.1). Paris : Éditions Points.

[1] <https://mywebintelligence.net/fr/home/>

Seconde traversée (*Défteros ploús*) : nouvelles et développements du projet d'édition collaborative numérique de l'*Anthologie palatine*

Margot Mellet¹, Marcello Vitali-Rosati^{1, 2}, Timothée Guicherd³

1 : Chaire de recherche du Canada sur les écritures numériques

2 : Université de Montréal

3 : BCIT Computing – École nationale supérieure de l'électronique et de ses applications, Vancouver

Un·e « usager·e » devant son écran lit son journal, regarde un film sur Netflix, une vidéo sur Youtube, cherche une information sur Wikipédia, réserve ses vacances sur Airbnb en rêvant d'une mer cristalline que lui présente une photo sur Google. Perdu·e dans sa navigation, l'usager·e se retrouve devant des vers en grec :

*Cypris est mon capitaine, Amour garde le cap en tenant
dans sa main le bout du gouvernail de mon âme
et le lourd souffle du désir me fait dériver,
car maintenant je nage dans une mer de garçons de toutes les races.*

Quel peut être le sens de cette épigramme au milieu de l'immensité de contenus que nous propose le Web ? Que signifie aujourd'hui lire un texte hellénistique (l'auteur de cette épigramme est Méléagre [I^{er} siècle av. J.-C.]) qui nous a été transmis grâce à un manuscrit du X^e siècle ? Comment le lecteur ou la lectrice peut contribuer à la transmission de l'épigramme et à faire en sorte que ce contenu continue à exister dans l'imaginaire collectif à venir ?

C'est à ces questions que nous avons consacré la première phase du projet d'édition collaborative numérique de l'Anthologie Palatine – un projet porté par la CRCEN¹ et financé par le CRSH-Savoir² – destiné à enrichir, valoriser et actualiser ce corpus classique fragmentaire. L'Anthologie palatine est un recueil qui rassemble plus de 4 000 épigrammes grecques allant de la période classique à la période byzantine (Cameron 1993). Considérant l'objet anthologique comme étant à la fois un objet de recherche spécialisé et un lieu d'intelligence et d'imaginaire collectif (Lévy 1994, 29; Doueïhi 2011 ; Vitali-Rosati 2017), le projet dans un premier temps visait à établir un espace ouvert en lecture comme écriture à tout utilisateur ou utilisatrice. Il est possible d'éditer (établir une version, proposer une traduction, aligner version et traduction), et de documenter l'épigramme (ajouter des mots-clefs, associer l'épigramme à un contenu externe).

Dans un premier temps du projet, nous avons constitué une API – une interface de programmation applicative, c'est-à-dire la mise à disposition de données structurées en suivant des protocoles et des standards – qui a été exploitée en plusieurs visualisations selon les différentes lectures/navigations dans l'univers anthologique. Cela a donné lieu à la plateforme Anthologia (<https://anthologia.ecrituresnumeriques.ca/>) qui permet l'édition collaborative, et à la plateforme POP (<http://pop.anthologiegrecque.org/>)³ qui propose des itinéraires sous la forme de parcours de lectures. De nombreux éditeurs et éditrices ont contribué au projet en éditant des transcriptions, des traductions, des liens vers des contenus externes qui font écho aux topoï des épigrammes. Cette première phase a été intensément documentée par de nombreuses communications (CRCEN 2018, 2018a, 2019a, 2019b) et de nombreux articles des participants (Vitali-Rosati 2015, 2017 ; Dumouchel 2018 ; Vitali-Rosati *et al.* 2020 ; Mellet 2020).

Nous inaugurons aujourd'hui une nouvelle phase du projet qui souhaite relever de nouveaux objectifs et défis techniques et éditoriaux. Le projet nous amène en effet à dépasser le cas de ce corpus particulier pour

penser et réaliser un espace numérique qui permette la découverte, l'étude comme les appropriations diverses de tout corpus fragmentaire ancien. La conceptualisation de cette plateforme – qui est réalisée en collaboration avec plusieurs partenaires travaillant à la diffusion de corpus anciens et fragmentaires dont Perseids, Perseus, Digital Milliet – est animée par les questions suivantes : comment décrire une épigramme ? Comment gérer les différentes versions, transcriptions, traductions ? Comment présenter et décrire un travail philologique ?

Notre structure de données est désormais complexe et riche, elle reflète la difficulté d'une organisation entre des modèles interreliés : un mot-clef peut en effet être lié à une épigramme, mais également à un auteur ou à d'autres modèles. Les anciennes données doivent également être migrées vers une nouvelle plateforme, ce qui constitue une réorganisation de la structure. Ce projet a par ailleurs une volonté de créer un méta-modèle générique, qui puisse être adapté et réutilisé pour d'autres corpus, tout en étant évolutif. La collaboration avec Perseids est donc un enjeu majeur pour la reproductibilité de ce modèle de données. Par ailleurs le développement technique de la plateforme fait émerger de nouvelles pratiques d'écriture numérique, la modélisation et les scénarios d'utilisation sont notamment rédigés au format Jupyter Notebook, dans une démarche de programmation éditoriale.

À la question de la structuration s'ajoute également le défi du multilinguisme : la plateforme en projet doit prendre en compte toutes les langues, que ce soit les traductions des épigrammes ou l'affichage de l'interface. La dimension collaborative pose quant-à-elle la question du statut de validation et de la mise en place de permissions différenciées pour l'édition ou la modification des données.

Afin de baliser le texte anthologique à des fins de recherche comme pour une plus large diffusion publique, nous travaillons actuellement sur des modèles de données et des ontologies du texte qui permettront de retranscrire l'unicité et la complexité des corpus anciens et fragmentaires au travers de l'édition numérique. Nous souhaitons dans cette communication présenter plus en détails les défis que nous avons rencontrés et les choix que nous avons fait pour les gérer : nous exposerons les états d'avancement, les modèles de données et les enjeux de construction de la nouvelle plateforme, en nous appuyant sur des exemples qui seront illustrés à l'aide de quelques jupyter notebook.

Bibliographie

- Cameron, Alan. 1993. *The Greek Anthology : From Meleager to Planudes by Rondo Etc* Cameron. OUP Oxford.
- CRCEN. 2018a. « Une API pour l'Anthologie grecque : repenser le codex Palatinus à l'époque du numérique ». CNRS – Paris.
- CRCEN. 2018b. « The Palatine Anthology project : an API allowing new interactions with the Greek epigrams ». University of Regina.
- CRCEN. 2019a. « AP POP : Anthologie Palatine, Plateforme Ouverte des Parcours d'imaginaires ». University of Vancouver.
- CRCEN. 2019b. « Palatine Anthology. Complexity for a digital research project ». Utrecht.
- CRCEN. s. d. « Projet Anthologie Palatine ». Documentation. Consulté le 14 décembre 2020.
- Doueihi, Milad. 2011. *Pour un humanisme numérique*. Paris : Seuil.
- Dumouchel, Suzanne. 2018. « Séance 2 EPHN : Marcello Vitali Rosati, université de Montréal : « Une API pour l'Anthologie grecque : repenser le codex Palatinus à l'époque du numérique ». Modéliser et virtualiser ». Billet. Consulté le 14 décembre 2020.
- Lévy, Pierre. 1994. *L'intelligence collective : pour une anthropologie du cyberspace*. Sciences et société. Paris : La Découverte.
- Mellet, Margot. 2020. « Penser le palimpseste numérique. Le projet d'édition numérique collaborative de l'Anthologie palatine ». Captures 5 (1).
- Vitali-Rosati, Marcello. 2015. « Midnight in Paris, l'Anthologie Palatine et la conjoncture médiatrice. Culture numérique ». BlogPost. 2015.

Vitali-Rosati, Marcello. 2017. « Pour une édition participative de l'Anthologie Palatine. Culture numérique ». BlogPost. 23 mars 2017.

Vitali-Rosati, Marcello, Servanne Monjour, Joana Casenave, Elsa Bouchard, et Margot Mellet. 2020. « Editorializing the Greek Anthology : The palatin manuscript as a collective imaginary ». *Digital Humanities Quarterly* 014 (1).

Notes

1. Chaire de recherche du Canada sur les écritures numériques
2. Conseil de recherche en sciences humaines
3. Plateforme ouverte des parcours d'imaginaires

Théâtre classique et difficultés numériques : quelques réflexions sur l'analyse dramatique assistée par ordinateur

Ioana Galleron¹

1 : UMR 8094 LATTICE, université Sorbonne nouvelle – Paris 3, CNRS

Peut-on étudier le théâtre français de la période classique (1600-1800) avec des méthodes computationnelles ? La réponse à cette question est évidemment oui, mais l'objectif de cette communication est de souligner l'existence de plusieurs difficultés et problèmes qui rendent cette étude moins évidente qu'on ne peut le penser a priori. L'intervention tracera ainsi une sorte de « plan de travail » pour le développement des études numériques sur le genre et la période envisagée, tout en contribuant à la réflexion sur les changements épistémologiques impliqués par le développement de telles études.

1. La question des corpus

Parmi les différents genres littéraires des XVII^e et XVIII^e siècles, le théâtre connaît sans aucun doute la meilleure couverture en termes de numérisation. En effet, le ratio entre les documents préservés sur papier, et ceux qui ont été numérisés ne serait-ce qu'en mode image est largement favorable au théâtre, comme je le montrerai à partir d'une étude réalisée sur la base du catalogue de la BNF et de différents entrepôts numériques. L'écart est encore plus important lorsqu'on compare les ressources qui peuvent être manipulées par ordinateur (formats txt et xml), et celles qui ne peuvent être qu'affichées pour une lecture humaine (visionnage sous forme de pdf ou lecture exclusive à travers une plateforme en ligne).

Toutefois, il ne s'ensuit pas de là que le spécialiste du théâtre est privilégié. En effet, lorsqu'on essaie d'utiliser les ressources numériques « machine readable » on se confronte à d'autres questions, liées à leur correction, à leur interopérabilité et, non en dernier lieu, à leur représentativité. En d'autres mots, la constitution d'un corpus, fondé sur des choix explicites et systématiques, est loin d'être facile dans le domaine du théâtre français des XVII^e et XVIII^e siècles.

2. La question des observables

Comme le suggère la section précédente, une approche de type « distant reading » est sinon impossible, du moins difficile dans le cas des études sur le théâtre des siècles classiques. Une approche « smart data » plutôt que « big data » semble donc, à l'heure actuelle, plus adaptée pour les études théâtrales françaises, d'autant plus qu'elle s'inscrit dans la mouvance actuelle de mise en place d'une « data-rich literary history » (Bode, 2018). En d'autres mots, le littéraire intéressé par l'analyse du théâtre doit envisager de créer ses propres observables (Landragin et Poudat, 2018), plutôt que d'utiliser des données pré-existantes.

Or, les initiatives d'enrichissement des ressources restent actuellement limitées. Pour ne prendre qu'un seul exemple, l'identification des entités nommées reste très imparfaite avec les outils actuellement existants, au théâtre comme pour d'autres genres littéraires. Le progrès ne peut venir que de l'intérêt des littéraires pour de tels chantiers, qui demandent au préalable un travail d'annotation manuelle. Mais de tels travaux suscitent plusieurs questionnements, dont je donnerai plusieurs exemples à partir des pièces de théâtre.

3. La question des perspectives

Cette partie prolonge la précédente, en prenant appui sur l'observation que la définition des observables et l'automatisation de leur identification ont toujours pris appui, jusqu'à présent, sur des curiosités et des initiatives venues d'autres disciplines que les lettres, dans lesquelles l'étude du théâtre français des XVII^e et

XVIII^e siècles est pourtant naturellement ancré. Pour le dire autrement, les littéraires se sont emparés de données et de méthodes élaborés par les linguistes, les informaticiens, parfois les géographes et les sociologues, mais se sont plus rarement attachés à l'opérationnalisation numérique de leurs propres objets d'études. On étudie les mots les plus fréquents, ou les mots-clés, parce que la linguistique a élaboré des protocoles pour ce faire, et que les littéraires ont pu tirer d'intéressantes conclusions pour leur discipline à partir de telles approches. On participe à l'effort de travail sur les entités nommées, parce qu'on a le sentiment qu'une nouvelle conceptualisation du personnage est possible dans le sillage de tels travaux. Mais il est plus rare que la démarche soit non pas celle d'un emprunt, mais plutôt celle d'une création – non pas dans le sens d'une création d'outil, ou d'interface, permettant d'étudier un aspect du texte littéraire, mais de modélisation d'un aspect du texte littéraire à étudier, menant à une réflexion sur la préparation éditoriale réaliser, puis à l'identification des outils appropriés pour mener cette étude. En d'autres mots, on assiste dans les études littéraires à la prévalence d'une pensée haptique (Rockwell, 2003) au détriment de l'émergence de « tinker toys » (W. McCarty, 2005). Sans dénier l'intérêt de la première (sur l'intérêt de laquelle v. Galleron et Idmhand, 2020), et sans opposer absolument les deux approches, il s'agit de proposer dans cette partie une réflexion sur la différence entre ce que les études théâtrales « traditionnelles » (pour les désigner avec une épithète commode, mais discutable) proposent comme programme d'étude du texte de théâtre, et ce que l'on peut actuellement réaliser grâce aux ressources numériques.

Bibliographie

Bode, Katherine. *A World of Fiction : Digital Collections and the Future of Literary History*. University of Michigan Press, 2018. DOI : 10.1353/book.59018.

Galleron, Ioana et Idmhand, Fatiha. « Why Go from Texts to Data, or The Digital Humanities as A Critique of the Humanities ». *Word and Text. A journal of Literary Studies and Linguistics*, vol. X/ 2020, p. 53-69, http://jls.upg-ploiesti.ro/site_engleza/No_1_2020.html

Poudat Céline, Landragin Frédéric, *Explorer un corpus textuel. Méthodes – pratiques – outils*. De Boeck Supérieur, « Champs linguistiques », 2017, DOI : 10.3917/dbu.pouda.2017.01.

McCarty, Willard. *Humanities Computing*. Basingstoke : Palgrave MacMillan, 2005.

Rockwell, Geoffrey. « What Is Text Analysis, Really ? ». *Literary and Linguistic Computing* 18.2 (2003) : 209-19. URL : <https://doi.org/10.1093/lc/18.2.209>.

Traduire sans trahir, la vocation des logiciels ou celle des chercheurs ? Programmes de recherche et programme informatique au carrefour des cultures et des méthodologies

Bastien Fond¹, Josquin Debaz²

1 : Université d'Augsbourg

2 : Groupe de sociologie pragmatique et réflexive (GSPR), École des hautes études en sciences sociales (EHESS)

Le numérique et ses outils facilitent-ils le rapprochement des écoles de pensée et le croisement des méthodes, au même titre qu'un partage au-delà des barrières linguistiques et culturelles ? Depuis 2018, une problématique à double-tranchant, d'ordre technique comme méthodologique, se fait jour au fil du rapprochement des programmes de recherche allemand et français que sont respectivement la sociologie de la connaissance appliquée au discours (Keller, 2005) et la sociologie des transformations (Chateauraynaud et Debaz, 2017). Au sein du projet ANR-DFG EnergiCorpus (université d'Augsbourg/EHESS Paris), nos deux équipes se sont engagées à confronter, enrichir et hybrider leurs approches. Plus précisément, cette mise en dialogue s'inscrit dans un rapprochement autour des programmes informatiques utilisés dans le cadre la socio-informatique (Chateauraynaud, 2003a), programmes notamment développés en coordination avec l'équipe française : Prospéro, Tiresias, etc. Or, au fil de la confluence des deux approches, plusieurs retours d'expérience s'esquissent, relativement aux enjeux de la compréhension sémantique en contexte, de la pluralité interprétative, ou encore de l'attention herméneutique portée à la finesse des expressions langagières.

Centrée thématiquement sur les transitions énergétiques, notre recherche a donné lieu à la constitution de *corpora* francophones et germanophones portant sur les débats publics des deux pays, ceci en vue d'une analyse comparée passant par le logiciel Prospéro (Fond *et al.*, 2019). Outre la difficulté à produire de tels *corpora* dans des contextes hétérogènes, les défis théoriques et méthodologiques inhérents à l'analyse computationnelle multilinguist(iqu)e s'avèrent majeurs et pluriels. D'une part, les environnements disciplinaire, culturel et linguistique du développement de la socio-informatique depuis les années 1990 influent en profondeur sur les orientations et les perspectives du projet. D'autre part, des défis de réduction technique émergent au fil de la transition logicielle d'un programme aussi herméneutique que la sociologie de la connaissance appliquée au discours (Scad), initialement développée par Reiner Keller à la marge des humanités numériques (Keller, 2007), et ultérieurement confrontée aux potentialités de l'analyse logicielle (Keller *et al.*, 2018).

De manière générale, la genèse et le développement de logiciels reflètent des questionnements, des problématiques de recherche, une culture scientifique, et par conséquent un sentier de dépendance qui peut potentialiser ou freiner une démarche d'exportation vers d'autres applications, d'autres sphères culturelles et d'autres langues. Si le logiciel Prospéro a initialement pris en charge des *corpora* textuels francophones et anglophones (Chateauraynaud et Torny, 1999 ; Boltanski et Chiapello, 1999 ; Chateauraynaud, 2003b), il a par la suite été adapté pour l'espagnol (Guerrero Bernal, 2012), l'italien (Marcodoppido, 2014), le tchèque (Smith, 2015), ou encore le portugais (Gouveia, 2016).

Certaines règles de codage linguistique s'avèrent cependant insatisfaisantes pour une traduction (trahison ?) vers l'allemand de son utilisation. Dans cette langue, le codage des noms propres à partir de leurs majuscules est par exemple dérouté par les noms communs, car ces derniers commencent également par une majuscule en allemand. Comment débusquer alors les protagonistes d'un corpus, ou encore reconstruire des jeux d'acteurs, si ce n'est même garantir un usage véritablement sociologique du logiciel ? Quels usages alternatifs – voire détournés – de Prospéro restent-ils envisageables en allemand ? En dépit de leur nature

fondamentalement contraignante, de telles difficultés demeurent particulièrement heuristiques, si tant est qu'elles soient par exemple abordées sous l'angle d'une réflexion plus large sur les enjeux de la convergence taxonomique (Desfriches-Doria *et al.*, 2020), ou sur les logiques de catégorisation algorithmique sous-jacentes au logiciel Prospéro.

Ceci étant, la conjonction méthodologique des équipes française et allemande s'avère facilitée par la nature de leurs approches de l'analyse de discours. Notamment, en s'éloignant de schémas et de méthodes de description clefs en main ou déléгатrices, les approches de Paris et d'Augsbourg insistent sur l'apprentissage du lecteur au fur et à mesure de son enquête et de ses lectures. Attentives autant à l'argumentation qu'à l'épistémologie des auteurs-acteurs, ces approches ont ainsi matière à se nourrir l'une de l'autre. Mais peuvent-elles pour autant confluer, et jusqu'à quel degré ?

Dans le contexte de l'arrivée à échéance de notre projet de recherche en fin d'année 2021, nous ambitionnons donc, à l'occasion de cette intervention, de proposer une réflexion sur la polyvalence logicielle. Il y sera question d'envisager ses écueils – tels du moins que nous les avons perçus et explorés depuis maintenant plusieurs années – et en retour de contribuer à l'exploration de certains contours du champ des humanités numériques. À l'aune des enjeux caractéristiques de la numérisation d'approches antérieurement constituées, telles que la Scad, qui apparaît et se popularise au début des années 2000, nous élargirons notre réflexion aux tensions qui émergent des évolutions des humanités numériques, au prisme des usages et des écoles de pensée (mélanges, déviations, ou encore hybridation méthodologiques).

Bibliographie indicative (par ordre chronologique)

Chateauraynaud, Francis et Didier Torny. 1999. *Les sombres précurseurs : une sociologie pragmatique de l'alerte et du risque*. Paris : EHESS.

Boltanski, Luc et Eve Chiapello. 1999. *Le nouvel esprit du capitalisme*. Paris : Gallimard.

Chateauraynaud, Francis. 2003a. *Prospéro : une technologie littéraire pour les sciences humaines*. Paris : CNRS.

Chateauraynaud, Francis. 2003b. « Les ressorts de la critique à l'échelle internationale Une lecture pragmatique de quelques épisodes de la guerre du Kosovo ». Dans *Historicité de l'action publique*, édité par Pascale Laborier et Danny Trom, 259-291. Paris : PUF.

Keller, Reiner. 2005. *Wissenssoziologische Diskursanalyse : Grundlegung eines Forschungsprogramms*. Wiesbaden : Springer.

Keller, Reiner. 2007 [2004]. *Diskursforschung : eine Einführung für SozialwissenschaftlerInnen (3. Aktualisierte Auflage)*. Wiesbaden : Springer.

Guerrero Bernal, Juan Carlos. 2012. « Adapter Prospéro à l'analyse de corpus en espagnol ». *Socio-informatique et argumentation*(blog). <https://socioargu.hypotheses.org/4087>.

Marcodoppido, Fabio. 2014. « Adapter Prospéro à l'analyse de corpus en italien : le travail effectué à partir d'un corpus de textes réglementaires et législatifs ». *Socio-informatique et argumentation* (blog). <https://socioargu.hypotheses.org/4763>.

Smith, Simon. 2014. « Mezi právnickým a žurnalistickým zdùvodňováním. Dovednosti administrátorů internetové diskuse ve slovenském deníku ». Communication présentée à *Úskalí žurnalistickéj a masmediálnej komunikácie v súčasnosti*, Nitra, 22 octobre.

Gouveia, Flávia. 2016. « Adaptation du Logiciel Prospéro pour l'analyse de corpus en langue portugaise (Portugais brésilien) ». *Socio-informatique et argumentation* (blog). <https://socioargu.hypotheses.org/5132>.

Chateauraynaud, Francis et Josquin Debaz. 2017. *Aux bords de l'irréversible : sociologie pragmatique des transformations*. Paris : Pétra.

Keller, Reiner, Anna-Katharina Hornidge, et Wolf J. Schünemann, éd. 2018. *The sociology of knowledge approach to discourse : investigating the politics of knowledge and meaning-making*. Londres : Routledge.

Fond, Bastien, Josquin Debaz et Martin Denoun. 2019. « Stakes and scales of the EnergiCorpus Project ». Communication présentée à *Transition stakes and scales. Dissenting arguments and conflicting prospects in French and German energy debates and practices*, Augsburg, 25-26 octobre.

Desfriches-Doria, Orélie, Josquin Debaz et Waldir Lisboa Rocha Filho. 2020. « Interoperability and discursive process about categories ». Communication présentée à *2nd Data and Digital Humanities Conference*, Hammamet, 15-17 octobre.

Un outil de dépouillement de sources archivistiques basé sur des technologies XML

Emmanuel Château-Dutier¹, Josselin Morvan²

1 : Université de Montréal (CRIHN)

2 : ANR Experts, CNRS

Depuis le Moyen Âge et probablement plus tôt, les autorités publiques confient à des personnes qu'elles estiment et qualifient compétentes l'action d'émettre un avis sur le savoir technique et scientifique, que ce soit dans le domaine gracieux comme contentieux. Cette recherche conduite dans le cadre d'un projet d'ANR vise à examiner, à partir d'un secteur économique majeur (celui du bâtiment à l'époque moderne), le mécanisme de l'expertise : comment la langue technique régulatrice et maîtrisée des experts s'impose à la société, comment leur compétence technique se convertit en autorité, voire parfois en « abus d'autorité » ? L'existence d'un fonds d'archives exceptionnel (A.N. Z1J) qui conserve l'ensemble des procès-verbaux d'expertise du bâtiment parisien de 1643 à 1792 nous permet de lancer une enquête pluridisciplinaire (juridique, économique et architecturale) de grande envergure sur la question de l'expertise qui connaît, à partir de 1690, un tournant particulier. En effet, les experts se divisent alors en deux branches différentes exerçant deux activités concurrentes, parfois complémentaires : les architectes et les entrepreneurs. La base de notre travail consistera d'abord à établir parallèlement deux corpus : d'une part, l'établissement d'un dictionnaire prosopographique des 234 experts exerçant de 1690 à 1790 à partir de sources diverses (manuscrites et imprimées) ; d'autre part, l'inventaire et l'analyse des procès-verbaux d'expertise sur la même période.

Ce travail repose principalement sur la constitution de deux corpus : d'une part, le dépouillement systématique de dix années de procès-verbaux conservés dans la sous-série Z1J des Archives nationales, d'autre part la constitution d'une base de données prosopographique sur les experts parisiens pour toute la période étudiée. La base de données prend la forme d'une application web réalisée avec des technologies XML libres et ouvertes qui soutient l'ensemble du travail du projet. Plusieurs formulaires dynamiques ont été développés avec XForms afin de permettre la mise à jour des dépouillements, de pouvoir renseigner la prosopographie et pour disposer de formulaires spécifiques pour le traitement de certaines sources telles que les inventaires après décès ou des annuaires. Cette application permet à l'équipe de partager les données au fur et à mesure de la réalisation du travail et prochainement de publier les résultats avant d'avoir terminé les dépouillements. L'utilisation de la base de données permet également de faire des requêtes riches dans les données pour des usages statistiques, des analyses de réseau ou pour produire des visualisations dynamiques.

La constitution de cet outil réclamait une réflexion approfondie sur la modélisation afin de produire des données qui soient pérennes et interoperables. Dans le cadre d'un partenariat avec les Archives nationales de France, l'équipe souhaite alimenter des instruments de recherche à partir des dépouillements réalisés pour lesquels il est nécessaire d'utiliser le standard ISAD-G et de communiquer les données au format XML-EAD du Conseil international des archives. Cependant, les questionnements spécifiques de notre recherche ne pouvaient pas être traités de manière satisfaisante avec ce format. Nous avons développé un schéma spécifique, compatible avec EAD, pour les dépouillements qui a fait l'objet de plusieurs ajustements à partir du travail sur les sources avant de pouvoir être implémenté informatiquement. Ce partenariat suggérait également d'emblée d'avoir recours au format XML-EAC-CPF issu du monde archivistique (norme ISAAR-CPF) pour la description des entités historiques en raison de sa cohérence avec EAD et des opportunités d'échange d'information possibles avec les Archives nationales qui utilisent actuellement ces standards pour ses référentiels d'acteurs.

La volonté de produire des données pérennes et interopérables et cette nécessité de fournir des notices au format XML nous ont naturellement guidées vers ces technologies pour le développement de notre plateforme de travail. L'application a donc été développée avec XQuery, un langage normalisé par le W3C, et la base de données XML native, libre et open source, BaseX. L'utilisation de XQuery et de son extension RESTXQ nous a permis de créer une application web dynamique, sans faire appel à aucun autre langage de programmation. C'est un langage qui peut facilement être maîtrisé par les chercheurs pour exprimer leurs requêtes à la base de données. Pour aller au bout de notre démarche tournée vers les technologies XML, nous avons fait le choix d'utiliser XForms pour l'intégration de tous nos formulaires (avec XSLTForms).

En l'absence de modèles adaptés, nous avons développé notre propre modèle pour le dépouillement des procès-verbaux d'expertise et des inventaires après-décès. Après avoir évalué les différentes solutions existantes pour le volet prosopographique, et en particulier l'extension de CIDOC-CRM, BioCRM, développée dans le cadre de *Reassembling the republic of letters*, nous avons finalement choisi d'utiliser Ric-CM pour la structuration de nos données, un modèle conceptuel développé par l'ICA (International Council on Archives). Malgré l'attrait de sa simplicité, l'ontologie BioCRM imposait de définir un vocabulaire pour la description des relations sociales (le travail autour de CRMsoc débutant seulement). Ric-CM présentait l'intérêt de s'inscrire dans la continuité des quatre standards archivistiques internationaux existants (ISAD-G, ISAAR-CPF, ISDF, ISDIAH) et d'être très très proche du format XML-EAC-CPF destiné à la description des acteurs historiques (personnes, institutions et groupes) et de leurs relations, tout en étant plus complet. Moyennant quelques aménagements, le format EAC-CPF nous permet de produire des données sémantiques en utilisant les XSLT développées par les Archives nationales de France.

Les trois formulaires pour le dépouillement des expertises, la saisie des données prosopographiques et l'analyse de inventaire après-décès sont maintenant pleinement opérationnels. Plusieurs milliers de procès-verbaux ont été dépouillés et sont reliés à des données prosopographiques sur plus d'une centaine d'experts. Le fait d'employer une pile de technologies unitaire a beaucoup facilité leurs développements. Surtout, nous disposons d'une entière maîtrise sur le produit, ce qui facilite beaucoup son adaptation aux besoins de la recherche. XQuery est un langage de choix pour préparer facilement toutes sortes de jeux de données en vue de l'analyse dans différents formats. Nous l'utilisons pour produire l'interface de consultation mais aussi des ressources au format GEXF ou JSON qui permettent de construire dynamiquement des analyses et des visualisations de réseau à partir de l'application avec des bibliothèques JavaScript comme D3.js ou bien Sigma.js. L'exploitation des données nous permet ainsi de préciser les logiques de répartition des affaires entre experts mais surtout de mener une étude diachronique des deux groupes professionnels, du réseau des acteurs et de sa constitution en rapport aux affaires.

Bibliographie

Barbot, Michela, et Robert Carvais. 2020. « Des archives pour analyser la ville et pour dessiner ses territoires ». *Histoire urbaine* 59 (3) : 63-84. <https://www.cairn.info/revue-histoire-urbaine-2020-3-page-63.htm>.

Barbot, Michela, et Robert Carvais. 2019. « Les livres sur le toisé et l'estimation en France et en Italie (XVI^e-XIX^e siècle) : circulations, continuités, ruptures ». Dans *Le livre technique avant le XX^e siècle : À l'échelle du monde*. Sous la direction de Liliane Hilaire-Pérez, Valérie Nègre, Delphine Spicq, et Koen Vermeir, 243-260. Histoire. Paris : CNRS Éditions. <http://books.openedition.org/editions-cnrs/27736>.

Carvais, Robert, et Valérie Nègre. 2015. « Parisian Surveyors (1690-1792) : Founding An Expert Corps ». Dans *5ICCH Proceedings*. 383-394. Chicago. <https://structurae.net/en/literature/conference-paper/parisian-surveyors-1690-1792-founding-an-expert-corps>.

Carvais, Robert. 2015. « Mesurer le bâti parisien à l'époque moderne. Les enjeux juridiques et surtout économiques du toisé ». *Histoire urbaine*, n° 43 (novembre) : 31-53. <https://doi.org/10.3917/rhu.043.0031>.

Carvais, Robert, et Valérie Nègre. 2018. « Experts and building assessments. An international comparison (thirteenth-twentieth century) ». *Verschiedene Materialien*. Dans *Building knowledge, constructing histories*. Sous la direction de Ine Wouters, Stephanie Van de Voorde, Inge Bertels, Bernard Espion, Krista de Jonge, et Denis Zastavni, Proceedings of the 6th International Congress on Construction History (6ICCH 2018), July 9-13, 2018, Brussels, Belgium. Boca Raton London New York Leiden : CRC Press, Taylor & Francis Group.

Carvais, Robert, Michela Barbot, Emmanuel Chateau-Dutier, et Valérie Nègre. 2018. « Pratiques des savoirs, entre jugement et innovation. Experts et expertises du bâtiment, Paris, 1690-1790. Premiers jalons pour une recherche ». Projet spontané Rapport final. Paris : Mission de recherche Droit & Justice. <http://www.gip-recherche-justice.fr/publication/pratiques-des-savoirs-entre-jugement-et-innovation-experts-expertises-du-batiment-paris-1690-1790-premiers-jalons-pour-une-recherche/>.

Delpu, Pierre-Marie. 2015. « La prosopographie, une ressource pour l'histoire sociale ». *Hypothèses* 18 (1) : 263-274. <https://www.cairn.info/revue-hypotheses-2015-1-page-263.htm>.

Dozo, Björn-Olav. 2008. « Données biographiques et données relationnelles. Notes théoriques pour une utilisation complémentaire des outils quantitatifs ». *CONTEXTES. Revue de sociologie de la littérature*, n° 3 (juin). <https://doi.org/10.4000/contextes.1933>.

Hotson, Howard, et Thomas Wallnig, dir. 2019. *Reassembling the Republic of Letters in the Digital Age : Standards, Systems, Scholarship*. Göttingen : Göttingen University Press. <https://doi.org/10.17875/gup2019-1146>.

Lemercier, Claire, et Emmanuelle Picard. 2012. « Quelle approche prosopographique ? » Dans *Les uns et les autres. Biographies et prosopographies en histoire des sciences*. Sous la direction de Laurent Rollet et Philippe Nabonnaud, 605-630. Presses Universitaires de Nancy ; Éditions Universitaires de Lorraine. <https://halshs.archives-ouvertes.fr/halshs-00521512/document>.

Nègre, Valérie. 2015. « The Toisés and the Emergence of New Technicians in Eighteenth-Century France ». Dans *Der Lehrbuchdiskurs über das Bauen*. Sous la direction de Uta Hassler, 112-121. Zürich : ETH.

Genet, Jean-Philippe. 1988. « L'informatique au service de la prosopographie : PROSOP ». *Mélanges de l'École française de Rome* 100 (1) : 247-263. <https://doi.org/10.3406/mefr.1988.2972>.

Tuominen, Jouni. 2016. « EMLO Prosopographical Data Model : Towards a Biographical Conceptual Reference Model ». STSM report, Cost Action IS1310. Reassembling the Republic of Letters. <http://www.republicofletters.net/wp-content/uploads/2017/03/cost-stsm-report-tuominen-jouni-2016-04-28.pdf>.

Références techniques

CIDOC-CRM <http://www.cidoc-crm.org>

BioCRM <http://seco.cs.aalto.fi/projects/biographies/>

CRMSoc <http://www.cidoc-crm.org/crmsoc/>

D3.js <https://d3js.org>

EAC-CPF <https://eac.staatsbibliothek-berlin.de>

EAD <https://www.loc.gov/ead/EAD3taglib/>

GEXF <https://gephi.org/gexf/format>

RESTXQ <http://exquery.github.io/exquery/exquery-restxq-specification/restxq-1.0-specification.html>

Ric-CM <https://www.ica.org/fr/records-in-contexts-modele-conceptuel>

Sigma.js <http://sigmajs.org>

XQuery 3.1 <https://www.w3.org/TR/xquery-31/>.

XForms <https://www.w3.org/TR/2009/REC-xforms-20091020/>.

XSLTForms <http://www.agencexml.com/xsltforms>

Utilisation d'approches automatiques pour la reconnaissance des expériences de lecture

François Vignale¹, Guillaume Le Noé Bienvenu², Guillaume Gravier², Pascale Sébillot²

1 : EA 4335 Langues, littératures, linguistique, Le Mans Université, université d'Angers

2 : UMR 6074 Institut de recherche en informatique et systèmes aléatoires (IRISA), université Rennes 1

Cette communication a pour but de présenter le rôle des techniques relevant de l'intelligence artificielle et du traitement du langage naturel dans la mise au point d'algorithmes de détection semi-automatique des expériences de lecture développés dans le cadre du projet READ-IT.

Au cours des dernières décennies, les connaissances sur l'histoire des pratiques de lecture ont considérablement augmenté au sujet des usages et des habitudes mais des questions fondamentales demeurent, telles que le « pourquoi » et le « comment » on lit. Grâce à l'exploration de sources numériques à la recherche de témoignages d'expériences de lecture, le projet READ-IT (Reading Europe Advanced Data Investigation Tool, <https://readit-project.eu>) vise à mieux comprendre ces phénomènes. Ce projet financé par le Joint Programming Initiative for Cultural Heritage (2018-2021) associe 5 partenaires de 4 pays (France, Royaume-Uni, Pays-Bas, République Tchèque).

En combinant différentes conceptions (Jauss 1982 ; Iser 1978) et en nous inscrivant dans une démarche fondée sur les sources, nous avons obtenu un modèle théorique et une ontologie (Reading Experiences Ontology, REO) proposant une description minimale où l'expérience de lecture est définie comme un phénomène temporel précédé de prémisses et suivis d'effets dans lesquels une personne interagit avec un contenu écrit par l'intermédiaire d'un médium (Antonini *et al.* 2019).

Pour répondre aux questions du « pourquoi » et du « comment » on lit, le projet READ-IT a fait apparaître des besoins importants en intelligence artificielle et plus particulièrement en traitement automatique du langage. Ces besoins entraînent entre autres la récupération en masse de données historiques et contemporaines ainsi que leur pré-annotation dans le but de détecter automatiquement dans les sources les passages contenant des témoignages de lecture ou des mentions d'œuvres d'art. Pour parvenir à ses objectifs, le projet READ-IT a mobilisé plusieurs technologies. Parmi elles, la reconnaissance des entités nommées (NER) qui est une tâche classique en traitement automatique des langues consistant à localiser et à associer les entités mentionnées présentes dans un texte dans des catégories prédéfinies telles que les noms de personnes, les organisations, les lieux, les œuvres d'arts... Des approches récentes, utilisées dans le cadre de READ-IT, permettent d'obtenir d'excellents résultats. Celles-ci se basent sur des modèles de langues pré-entraînés comme ELMo (Lample *et al.* 2016) ou BERT (Devlin *et al.* 2019)

La classification de textes a également été utilisée. Il s'agit du processus qui consiste à attribuer une catégorie à un texte en fonction de son contenu. Les approches en traitement automatique des langues pour cette tâche, comme pour les autres, se sont historiquement basées sur des méthodes à base de règles. Pour le projet READ-IT, les approches d'apprentissage automatique (machine learning) ainsi que les méthodes d'apprentissage profonds (deep learning) qui sont aujourd'hui considérées comme délivrant les meilleures performances ont été testées.

Après une série de campagnes d'annotation menées entre mars et septembre 2020, on peut livrer quelques résultats qui montrent la pertinence des approches retenues et qui permettent d'entrevoir des perspectives prometteuses.

En ce qui concerne la reconnaissance des entités nommées, la détection des mentions d'œuvres d'art fonctionne bien sur le plan qualitatif en utilisant les modèles BERT (Bidirectional Encoder Representations from Transformers) et plus particulièrement ontonotes pour l'anglais et Multilingual Cased pour les autres langues. Ces modèles de langues très complets permettent l'identification de mentions d'œuvre d'art dans

une centaine de langages, en plus d'identifier 18 autres types d'entités (PERSON, NORP, FACILITY, ORGANIZATION, GPE, LOCATION, PRODUCT, EVENT, WORK OF ART, LAW, LANGUAGE, DATE, TIME, PERCENT, MONEY, QUANTITY, ORDINAL, CARDINAL).

La mise au point de classifieurs pour la détection des témoignages de lectures dans les textes est une tâche plus complexe car elle requiert des données annotées en très grand nombre. Des commentaires de livres issus de réseaux sociaux de lecture francophone et anglophone Babelio et Goodreads ont été annotés lors de plusieurs campagnes qui se sont déroulées de mars à septembre 2020. Elles ont permis de mettre en lumière un faible taux d'accord inter-annotateurs qui sert à mesurer la cohérence des annotations produites (κ de Fleiss inférieur à 0,3). Ce constat est dû à la fois à la part considérable d'interprétation personnelle et la complexité des tâches qui consistaient à la fois en le balisage du début et de la fin des expériences de lecture et en l'identification des composantes des expériences, en respectant le modèle de données de READ-IT. Des approches d'apprentissage automatiques classiques ont été testées (td*idf/SVM, FastText) sur ces sources contemporaines, lesquelles ont donné des résultats très intéressants sur ce type de données avec un degré de précision très élevé mais un rappel (nombre de sources pertinentes) encore faible.

Cependant, la réutilisation de ces modèles sur des textes plus anciens annotés manuellement par le passé (correspondance de Joseph Conrad, Memories and Portraits de R. L. Stevenson) ont montré quelques faiblesses dans cette approche. En effet, les formes familières de commentaires issus d'un corpus web qu'un classifieur est en mesure d'apprendre sont très éloignées de celles que l'on retrouve dans des textes littéraires plus anciens et réciproquement, sans oublier le fait que l'accord inter-annotateur est probablement encore plus faible pour les sources historiques en raison d'une part encore plus importante laissée à l'interprétation personnelle.

Corpus	Babelio
Nombre de sources	87664
Nombre de sources contenant une expérience de lecture	2713
Ratio	3,1 %

Corpus	Goodreads
Nombre de sources	10000
Nombre de sources contenant une expérience de lecture	608
Ratio	6,1 %

Corpus	Conrad's Letters
Nombre de sources	273
Nombre de sources contenant une expérience de lecture	2
Ratio	0,7 %

Corpus	Stevenson's Memories
Nombre de sources	326
Nombre de sources contenant une expérience de lecture	2
Ratio	0,6 %

Par conséquent, afin de détecter les témoignages de lecture présents dans tous les types de textes (article/commentaire web, correspondances, essai littéraire, roman, ...), nous pensons délivrer un modèle basé sur des règles (présence d'action de lecture, identification du lecteur, présence de médium ...), moins précis que des algorithmes entraînés sur des données spécifiques, mais plus robuste sur des données

fortement hétérogènes. De plus, ces modèles permettent d'expliquer les caractéristiques qui ont permis la détection, lesquelles pourront être intégrées dans une interface, et ne nécessiteront qu'une faible maintenance sur le long terme, facilitant ainsi la durabilité du projet et sa transposition à d'autres phénomènes expérimentaux.

Références

- Jauss, Hans Robert. 1982. *Toward an Aesthetic of Reception*. Minneapolis : University of Minnesota Press.
- Iser, Wolfgang. 1978. *The act of reading : a theory of aesthetic response*. London : Routledge.
- Antonini, Alessio, Vignale, François, Gravier, Guillaume et Ouvry-Vial, Brigitte. 2019. « The Model of Reading : Modelling principles, Definitions, Schema, Alignments ». <https://hal-univ-lemans.archives-ouvertes.fr/hal-02301611>.
- Lample, Guillaume, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami et Chris Dyer. 2016. « Neural Architectures for Named Entity Recognition ». Dans *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, 260-270. San Diego, California : Association for Computational Linguistics. <https://doi.org/10.18653/v1/N16-1030>.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee et Kristina Toutanova. 2019. « BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding ». arXiv : 1810.04805 [cs], mai. <http://arxiv.org/abs/1810.04805>.

« Je pense que ça traite d'expérience de lecture, à voir... » : retour sur une expérience d'annotation collaborative

François Vignale¹, Guillaume Le Noé Bienvenu², Guillaume Gravier², Pascale Sébillot³

1 : EA 4335 Langues, littératures, linguistique, Le Mans Université, université d'Angers

2 : UMR 6074 Institut de recherche en informatique et systèmes aléatoires (IRISA), université Rennes 1

3 : INSA de Rennes, UMR 6074 Institut de recherche en informatique et systèmes aléatoires (IRISA), université Rennes 1

Cette communication a pour but de présenter une expérience d'annotation collaborative d'expériences de lecture menée avec des étudiants dans le cadre du projet READ-IT.

Au cours des dernières décennies, les connaissances sur l'histoire des pratiques de lecture ont considérablement augmenté au sujet des usages et des habitudes mais des questions fondamentales demeurent, telles que le « pourquoi » et le « comment » on lit. Grâce à l'exploration de sources numériques à la recherche de témoignages d'expériences de lecture, le projet READ-IT (Reading Europe Advanced Data Investigation Tool, <https://readit-project.eu>) vise à mieux comprendre ces phénomènes. Ce projet financé par le Joint Programming Initiative for Cultural Heritage (2018-2021) associe 5 partenaires de 4 pays (France, Royaume-Uni, Pays-Bas, République Tchèque).

En combinant différentes conceptions (Jauss 1982 ; Iser 1978) et en nous inscrivant dans une démarche fondée sur les sources, nous avons créé un modèle théorique et une ontologie (Reading Experiences Ontology, REO, <https://github.com/eureadit/reading-experience-ontology>) proposant une description minimale où l'expérience de lecture est définie comme un phénomène temporel précédé de prémisses et suivi d'effets dans lesquels une personne interagit avec un contenu écrit par l'intermédiaire d'un médium (Antonini *et al.* 2019).

Cependant, afin de dépasser le stade théorique et de commencer à disposer des outils de détection semi-automatique des expériences de lecture, il est apparu nécessaire de lancer une campagne d'annotation collaborative à partir d'une interface spécialement conçue pour le projet (<https://readit-it.hum.uu.nl>). Cette étape est essentielle pour le développement des algorithmes que READ-IT doit produire d'ici à son terme. Les résultats doivent donc être confrontés à l'expertise humaine afin d'améliorer leur performance, ceci dans un processus itératif. Elle a donc pour objectifs principaux : 1) de valider définitivement le modèle de données ; 2) de fournir suffisamment de sources annotées pour entraîner convenablement les détecteurs reposant sur des méthodes de traitement automatique du langage naturel ; 3) de tester et d'améliorer l'ergonomie de la plate-forme d'annotation ainsi que d'évaluer la définition et la clarté des concepts repris dans le guide d'annotation dans l'objectif d'une dissémination vers le grand public.

Cette campagne s'est déroulée sur 10 semaines de mars à mai 2020 avec la participation de 10 étudiants stagiaires de M2 Lettres et langues rémunérés à hauteur de 200 heures chacun. Les sources ont constitué en des commentaires de lecteurs en ligne extraits des réseaux sociaux de lecture francophone et anglophones Babelio et Goodreads (Rebora *et al.* 2019). Le corpus a été construit aléatoirement et les données ont été anonymisées. Les métadonnées relatives aux titres des ouvrages ont été masquées. Le corpus a été subdivisé en unités de 100 à 150 commentaires dont 30 % étaient partagés entre plusieurs annotateurs afin de calculer l'accord inter-annotateurs qui sert à mesurer la cohérence des annotations (Bayerl *et al.* 2011). Les stagiaires ont bénéficié de deux séances de formation en présentiel puis, en raison du confinement, de deux autres à distance. Ils se sont également appuyés sur des supports spécialement conçus pour cette campagne et sur un guide d'annotation. Des réunions hebdomadaires ont permis de suivre l'avancement du travail et de répondre aux difficultés d'organisation, de compréhension des consignes ou aux problèmes techniques sur la plateforme d'annotation. Des rendez-vous plus restreints (par groupes de 3 annotateurs) ont également eu lieu afin de mieux comprendre les raisons de certaines annotations.

Deux tâches ont été confiées : 1) le balisage des limites extrêmes des expériences de lecture contenues ou non dans les commentaires en ligne ; 2) l'identification des composantes de l'expérience de lecture selon les classes définies dans l'ontologie REO.

Le balisage consiste à repérer le début et la fin de l'action de lecture à l'intérieur du commentaire. Les éléments décisifs sont le sujet lecteur (pronoms personnels), les œuvres citées (entités nommées), les verbes d'action liés à la lecture (lire, dévorer ...), le vocabulaire lié au médium (livre, bouquin...) et des éléments contextuels sans que toutes ces conditions soient nécessairement réunies en même temps.

Les composantes de l'expérience de lecture à identifier correspondent à des classes existant dans le modèle de données et sont donc alignées avec l'ontologie REO. Schématiquement, elles appartiennent à 3 grands groupes :

- Reading agent (informations sur le lecteur)
- Reading resource (informations sur le texte lu)
- Reading process (informations sur la nature, le déroulé, les circonstances et les conséquences de l'expérience de lecture)

Au total, plus de 5 000 commentaires ont donné lieu à des annotations.

Les résultats de cette campagne sont contrastés. Le volume des données nécessaire pour entraîner les modèles de détection a été atteint et ces derniers ont donné des résultats déjà très encourageants (précision très satisfaisante mais rappel encore un peu faible) et l'ergonomie de la plate-forme s'est considérablement améliorée à la suite des remarques des stagiaires. En revanche, la qualité et la précision des annotations sont très hétérogènes. Sur ce point, la question de la motivation des annotateurs se pose et l'on ne peut exclure la présence de turkers (Cardon 2015) dans le groupe, ce qui a affecté la cohérence des annotations. Enfin, le taux d'accord inter-annotateurs a été très faible (kappa de Fleiss inférieur à 0,3). Ceci s'explique pour plusieurs raisons. Le concept d'expérience de lecture est encore flou et n'a pas de définition universellement admise pour le moment ce qui laisse une part très importante à l'interprétation personnelle et à la subjectivité, à la différence de la plupart des expériences d'annotation qui ont pour sujet des concepts clairs (entités nommées) ou la reconnaissance d'objets ou de formes dans des images par exemple.

Au total, ce bilan montre la nécessité de reformuler et de préciser la plupart des concepts de base en étant nettement plus explicite et en donnant des consignes claires si l'on veut ouvrir la plate-forme d'annotation à un public plus large et non spécialiste. Dans le même temps la formation des futurs utilisateurs, ainsi que leur accompagnement par des animateurs – qui ont ici été fortement perturbés par les exigences du confinement – que ce soit en présentiel ou à distance, doivent être au centre de l'organisation de ce type d'opération.

Références

- Antonini, Alessio, François Vignale, Guillaume Gravier et Brigitte Ouvry-Vial. 2019. « The Model of Reading : Modelling principles, Definitions, Schema, Alignments ». <https://hal-univ-lemans.archives-ouvertes.fr/hal-02301611>.
- Jauss, Hans Robert. 1982. *Towards an Aesthetic of Reception*. Minneapolis : University of Minnesota Press.
- Iser, Wolfgang. 1978. *The act of reading : a theory of aesthetic response*. London : Routledge.
- Bayerl, Petra Saskia et Karsten Ingmar Paul. 2011. « What Determines Inter-Coder Agreement in Manual Annotations ? A Meta-Analytic Investigation ». *Computational Linguistics* 37 (4) : 699-725. https://doi.org/10.1162/COLI_a_00074.
- Fort, Karen. 2016. *Collaborative annotation for reliable natural language processing : technical and sociological aspects*. London : John Wiley & Sons : ISTE.
- Rebora, Simone, Peter Boot, Federico Pianzola, Brigitte Gasser, J. Berenike Herrmann, Maria Kraxenberger, Moniek Kuijpers *et al.* 2019. « Digital Humanities and Digital Social Reading ». OSF Preprints. <https://doi.org/10.31219/osf.io/mf4nj>.
- Cardon, Dominique, Antonio A. Casilli. 2015. *Qu'est-ce que le digital labor ?* Bry-sur-Marne : INA éditions.

Présentations éclair

Adressbuch 2021 : la troisième vie de l'Annuaire des Allemands à Paris de 1854

Mareike Koenig¹, Gérald Kembellec^{1, 2}

1 : Institut historique allemand, Max Weber Stiftung

2 : EA 7339 Laboratoire Dispositifs d'information et de communication à l'ère numérique (DICEN), Conservatoire national des arts et métiers, université Gustave-Eiffel, université Paris-Nanterre

Le recensement de 1851 faisait ressortir la présence de 12.245 Allemands à Paris, avec une grande majorité d'artisans. Le projet « *Adressbuch 1854* », mené à l'Institut historique Allemand entre 2002 et 2006, valorisait à ce propos un ouvrage identifié dans les fonds de la bibliothèque de la ville de Paris. Quelque 4700 adresses étaient pointées dans un bottin encore consultable en archives. Le projet consistait à l'intégration des entrées de l'ouvrage dans une base de données consultable en ligne.

Environ 15 ans plus tard, le corpus est régulièrement consulté et fait l'objet de demande d'informations notamment de la part de généalogistes et des historien-ne-s menant une recherche prosopographique. Ce projet a été réalisé par des historien-ne-s qui s'équipaient d'autres pratiques que celles de leur discipline pour modéliser, analyser et diffuser leur objet de recherche, renouvelant ainsi leurs pratiques. Il s'agissait à l'époque d'un travail historique dans une mouvance innovante – pionnière même, si l'on considère que les humanités numériques en tant que concept problématisé, datent de 2008.

En 2021, nous reprenons ce projet, dans une double optique : premièrement, l'interface ne correspond plus aux canons de consultation, d'ergonomie, d'accessibilité et de sémantique du Web. Deuxièmement, dans une optique info-documentaire, ce corpus mérite une re-modélisation à visée humaniste pour correspondre aux sujets et méthodes des sciences humaines et sociales. Troisièmement, les données étaient sous copyright à l'époque lors du projet original et l'équipe souhaite aujourd'hui les libérer et les diffuser sur une licence ouverte. Quatrièmement, ce qui n'était pas possible à l'époque – à savoir une visualisation interactive des données sur une carte de Paris – est aujourd'hui techniquement aisément réalisable. Ainsi repensées, les données ré-agencées, documentées et surtout présentées et visualisées permettront d'offrir un matériau et une interface de consultation plastiques, modulables aux besoins des chercheurs et usagers qui seraient amenés à les consulter, mobiliser et re-travailler.

Dans le contexte d'une redocumentarisation des informations historico-sociologiques, il est bien sûr des écueils à éviter soigneusement pour éditorialiser un dispositif selon les canons des humanités numériques. Le premier défi est celui d'une modélisation en diachronie : les lieux, métiers et contextes socio-professionnels ont subi de profondes mutations. Heureusement, des thésauri historiques de métiers existent, parfois en plusieurs langues, mais il faut les aligner entre eux et avec le corpus. Le paysage de la ville de Paris, sa toponymie et son découpage administratif ont subi de profondes mutations en 170 ans. Dans ces conditions, il faut garder à l'esprit les mises en garde de J. Drucker sur le simple plaquage d'informations sur des cartes comme visualisation. Il devient alors indispensable de proposer une re-crédation de sens en respectant les règles de sémiologie graphique (de Bertin) et le bon sens narratif dans la représentation de la diachronie géographico-administrative. Enfin, dans une optique historienne, il est indéniable que la représentation des informations depuis une base de connaissances ne saurait se substituer à la mise à disposition d'une numérisation de qualité des éléments d'archive mobilisés. En effet, la gestion de la preuve, notion proche de la diplomatique, se base sur le lien étroit qu'il peut y avoir entre les informations et l'archive.

Source en archive

F. A. Kronauge, Adreßbuch der Deutschen in Paris für das Jahr 1854 oder vollständiges Adreßverzeichnis aller in Paris und seinen Vorstädten wohnenden selbständigen Deutschen in alphabetischer Ordnung, Paris 1854.

Littérature sur le projet initial

Mareike König, « Georg Kibler, Möbelbauer, rue de Charonne 39: Adreßbuch der Deutschen in Paris für das Jahr 1854 ». *Francia* 30/3 (2004), p. 145-158.

Mareike König, « Migration économique, exil politique et critique sociale : Les Allemands à Paris au XIX^e siècle ». *DeuFraMat*, 2002, <http://www.deuframat.de/fr/societe/structures-demographiques-migration-minorites/migration-economique-exil-politique-et-critique-sociale-les-allemands-a-paris-au-xixeme-siecle/tour-de-france-des-artisans.html>.

Éditer et transmettre le *Parallèle des Anciens et des Modernes* de Charles Perrault

Delphine Reguig¹, Emmanuelle Perrin²

1 : UMR 5317 Institut d'histoire des représentations et des idées dans les modernités (IHRIM), CNRS, université Jean Monnet (Saint-Étienne)
2 : Labex COMOD

La proposition de poster porte sur le projet d'édition numérique du *Parallèle des Anciens et des Modernes* de Charles Perrault, placé sous la responsabilité de Delphine Reguig (IHRIM), professeure de langue et littérature françaises du XVII^e siècle à l'université de Saint-Étienne. Elle s'inscrit dans l'un des principaux thèmes du colloque : la conservation et la transmission des patrimoines culturels, artistiques et scientifiques.

Paru en quatre tomes entre 1688 et 1697, ce texte central dans la Querelle des Anciens et des Modernes n'a jamais été édité, ni exploité dans son intégralité. Il met en scène un dialogue entre trois personnages : le Président, partisan des Anciens, l'Abbé, partisan des Modernes, et, entre les deux, le Chevalier, au tempérament enjoué. La discussion se déroule à Versailles, cadre privilégié d'observation du « Siècle de Louis XIV ».

Examinant les mérites respectifs des artistes, philosophes, savants antiques et contemporains dans l'architecture, la sculpture, la peinture, l'éloquence, la poésie, l'astronomie, la géographie, la navigation, la guerre, la philosophie, la médecine ou la musique, ce texte paraît emblématique d'une ambition, que nous appellerions aujourd'hui « interdisciplinaire », de refondation des savoirs.

Pour donner accès à ce texte fleuve voire labyrinthique, foisonnant de références, appelant largement l'iconographie et l'annotation, une édition électronique enrichie possède une véritable pertinence.

La transmission de ce texte se fonde par exemple sur un protocole éditorial destiné à en favoriser la lisibilité, avec la modernisation de la graphie et la régularisation de la ponctuation. Les aspects techniques liés à cette édition et à sa transmission s'appuient sur les principes FAIR (Facile à trouver, Accessible, Interopérable et Réutilisable). Le texte est encodé en XML-TEI P5, un format ouvert et standard pour l'édition numérique, soutenu par une large communauté internationale. L'ensemble des données XML est géré dans une base de données XML native (BaseX). Une édition HTML est générée directement dans la web application de BaseX en utilisant SynopsX. L'interface de consultation est un site web, hébergé par la TGIR Huma-Num à l'adresse : <https://parallele-anciens-modernes.huma-num.fr>.

La « plus-value » numérique apportée par l'encodage en XML-TEI est tout à fait significative. Il permet par exemple de rattacher chaque réplique à son locuteur et de filtrer les recherches selon les personnages. De même, le « typage » des différents éléments constitutifs du texte (préfaces, dialogues, pièces annexes, notes) offre la possibilité de mener des recherches à facettes. L'encodage autorise également à établir des liens vers des ressources documentaires (livres contemporains du *Parallèle* et numérisés sur Gallica) et iconographiques, car le texte de Charles Perrault contient d'amples descriptions du château de Versailles et de ses jardins. L'encodage des entités nommées (personnes et lieux), des œuvres (écrits, monuments, peintures, sculptures, art des jardins) et des sujets reflète la densité de ce texte avec plus de 620 noms de personnes, 130 toponymes, 340 œuvres et 300 sujets. La TEI offre une qualité et une finesse d'indexation qui va bien au-delà de la recherche plein texte, pour proposer différents points d'accès dans le texte. Pour favoriser l'interopérabilité et la réutilisation des données, les termes des index ont été normalisés selon les notices d'autorité de la BnF. Des liens avec de grands référentiels (BnF, IdRef, ISNI, VIAF) permet d'identifier ces termes sans ambiguïté et de les mettre en relation avec d'autres ressources publiées sur le Web.

L'annotation est un autre volet important dans la transmission de cette œuvre. L'équipe scientifique pluridisciplinaire est constituée de spécialistes de littérature, d'histoire des idées, de philosophie (en

particulier des sciences et de la médecine), d'histoire de l'art, de musicologie et d'histoire sociale. L'annotation comprend trois niveaux d'intervention. Elle propose des élucidations lexicales ou syntaxiques, des informations historiques ou contextuelles, des commentaires enfin permettant de situer et comprendre l'argumentation. Les quelque deux mille notes prévues témoignent encore une fois de la richesse de ce texte.

L'édition du *Parallèle* doit aussi servir de base à une redécouverte de Charles Perrault, dont la lecture et l'étude sont aujourd'hui limitées aux *Contes*. Ce projet est donc une première étape vers la reconstruction de la figure de Perrault comme un polygraphe à l'écriture éminemment politique.

Enfin cette œuvre est elle-même centrée sur la transmission puisqu'elle propose, au travers du regard critique porté sur l'héritage des Anciens, la définition d'une littérature et d'une pensée modernes.

Références bibliographiques

Sources primaires

La Querelle des Anciens et des Modernes [anthologie], éd. A-M. Lecoq, Paris, Gallimard, 2001.

BOILEAU Nicolas, *Traité du sublime, Réflexions critiques sur Longin*, dans *Œuvres complètes*, éd. F. Escal, Paris, Gallimard, « Bibliothèque de la Pléiade », 1966.

PERRAULT Charles, *Mémoires de ma vie*, précédé d'un essai d'A. Picon, *Un moderne paradoxal*, Paris, Macula, 1993.

Sources secondaires

« Charles Perrault », *Europe*, numéro spécial 739/740, nov.-déc. 1990.

D'un siècle à l'autre : Anciens et Modernes, éd. R. Duchêne, Marseille, CMR 17, 1987.

BAHIER-PORTE Christelle, dir., « Les Écrivains de la querelle : de la polémique à la poétique (1687-1750) », *Revue Fontenelle*, n° 9, 2012.

HARTOG François, *Régimes d'historicité. Présentisme et expérience du temps*, Paris, Seuil, 2003.

HARTOG François, *Anciens, Modernes, Sauvages*, Paris, Galaade Éditions, 2005.

NORMAN Larry F., *The Shock of the Ancient*, Chicago, The University of Chicago Press, 2011.

POULOUIN Claudine et BAHIER-PORTE Christelle, *Écrire et penser en Moderne (1687-1750)*, Paris, Champion, 2015.

VAN ESLANDE Jean-Pierre et NORMAN Larry F., « Modernités de Perrault », *Cahiers Parisiens/Parisian Notebooks*, vol. 4, 2008, p. 190-288.

Édition numérique

BURNARD Lou, *Qu'est-ce que la Text Encoding Initiative ?*, Marseille : OpenEdition Press, 2015 [en ligne : <http://books.openedition.org/oep/1237>].

GALLERON Ioana, DEMONET Marie-Luce, MEYNARD Cécile *et al.*, *Les Publications numériques de corpus d'auteurs – Guide de travail, grille d'analyse et recommandations*, 2018 [en ligne : <https://cahier.hypotheses.org/guides-juridiques/les-publications-numeriques-de-corpus-dauteurs>].

IDMHAND Fatiha et GALLERON Ioana, *Guide pour la FAIRisation des données des corpus d'auteurs*, groupe de travail Data_Cahier], 2020 [en ligne : <https://halshs.archives-ouvertes.fr/halshs-03037748/document>].

MOUNIER Pierre, *Les Humanités numériques. Une histoire critique*, Paris, Éditions de la Maison des sciences de l'homme, 2018.

PIERRAZZO Elena, *Digital Scholarly Editing : Theories, Models and Methods*, Farnham, Surrey, Ashgate, 2015 [disponible également en ligne : <https://hal.univ-grenoble-alpes.fr/hal-01182162/document>].

Essai de cartographie de médias imaginaires du XIX^e siècle

Philippe Ethuin^{1, 2}

1 : INSPÉ de l'académie d'Amiens, université de Picardie Jules-Verne

2 : UR 4426 Médiation, information, communication, art (MICA), université Bordeaux-Montaigne

Dans la *Théorie générale de la communication et de la communication*, Robert Escarpit met en évidence une « approche multilatérale qui a abouti au début du XX^e siècle à trois conquêtes essentielles et à peu près contemporaines :

1. Possibilité de créer des documents sonores (1878).
2. Possibilité de créer des documents visuels animés (1895).
3. Possibilité de diffuser l'information à distance sans la médiation d'un document (1897). » [Escarpit, 1976]

Ces conquêtes ont été précédées dans l'imaginaire social et littéraire de multiples préfigurations. En effet, au cours du XIX^e siècle, alors qu'apparaissent la télégraphie électrique, le téléphone ou le phonographe, des écrivains et des journalistes imaginent des extrapolations et des hybridations de ces dispositifs médiatiques émergents. Ces inventions, créées notamment dans la littérature, la caricature et la prospective, constituent un patrimoine immatériel à la fois culturel, artistique et scientifique, prenant parfois des formes matérielles.

Une centaine de ces inventions (couvrant la période 1789-1914) est rassemblée dans une base de données personnelle et divers outils numériques sont mobilisés permettant le repérage des inventions.

En 1760 Tiphaigne de la Roche dans son utopie *Giphantie* imagine un réseau de communication couvrant l'ensemble du monde et aboutissant à un globe servant d'interface :

« De petits canaux imperceptibles, replit le préfet, viennent de chaque point de la superficie de la Terre aboutir à ce globe. Son intérieur est organisé de manière que l'émotion de l'air qui se propage par les tuyaux imperceptibles et s'affaiblit à la longue reprend de l'énergie et redevient sensible. »

L'idée des tuyaux est reprise par Louis Desnoyers dans *Les Aventures de Jean-Paul Choppart* (1832) qui présente un réseau de tuyaux acoustiques permettant de livrer la musique à domicile. Mathieu Briancourt dans *Visite au phalanstère* (1848) fait des tuyaux acoustiques un moyen de communication avec les logophores.

Si les tuyaux acoustiques ont véritablement existé (notamment dans la marine et pour des usages domestiques) et que leur usage a perduré jusqu'à la généralisation du téléphone, les espaces couverts par les extrapolations sont bien plus grands.

L'histoire des médias a souvent livré un récit linéaire des inventions [Parikka, 2015]. Pourtant il existe « plusieurs niveaux de déterminations » [Foucault, 1969]. Parmi les déterminations issues de la fantaisie, de l'utopie ou du mythique, les médias imaginaires occupent une place particulière. La reproduction mécanique du son, la communication et la vision à distance au moyen de technologies ont été envisagées avant d'être inventées [Mombert 2019]. Les « médiamanies » sont accompagnées de réactivation de *topoi* qui ont certes des origines mythiques ou fantastiques [Huhtamo, 2019] mais qui ont pu avoir des « réalisations » dans la littérature ou la prospective sous des formes rationnelles.

Les études sur les médias prennent le plus souvent comme point de vue notre époque. Par exemple, Robert Escarpit présente un schéma de « La technologie de la communication » [Escarpit, 1976] ou Luc Tournès propose une « Une histoire de la musique enregistrée » dans *Du phonographe au MP3* (2008) qui aboutissent aux technologies en usage aujourd'hui.

Au XIX^e siècle, les contemporains avaient une grille d'analyse des inventions différente de la nôtre tissant des liens entre elles et s'appuyant sur des innovations antérieures. Par exemple un auteur affirme : « La phonographie sera pour la voix ce que la photographie est pour la figure, dont elle fait le portrait sans en dissimuler les rides, et même en les exagérant » [Leroux, 1878] et un autre : « Le téléphone est le télescope de l'acoustique ; le microphone en est le microscope. » [Aumont, 1878]. Ces rapprochements constituent l'univers conceptuel de l'époque.

L'analyse d'un corpus constitué de l'ensemble des articles parus dans la presse française annonçant en 1878 l'invention du phonographe au moyen d'Iramuteq permet d'appréhender l'imaginaire qui s'y attache à la fois sur les aspects techniques, généalogiques et prospectifs.

La décomposition des noms d'inventions imaginaires permet d'accéder à l'univers conceptuel des créateurs et de définir une périodisation. Du « micro-acoustico-télescope » (1809) au « théatographe » (1913) en passant par l'« historioscope » (1882), le « phonophotosténotypo-biographe » (1887) sans oublier le « téléphonoscope » (1883) d'Albert Robida, les noms d'inventions imaginaires accompagnent aussi l'évolution des dispositifs réels ou simplement projetés (comme le « télectroscope » (1878) décrit par le vulgarisateur scientifique Louis Figuier), souvent en les prolongeant ou en envisageant des usages inédits.

L'utilisation des fonctions d'Obsidian permet de construire un graphique relationnel entre ces différentes inventions imaginaires et de mettre en évidence les combinaisons de préfixes, racines et suffixes utilisées par les auteurs pour baptiser leurs appareils fictifs ainsi que les liens avec les inventions réellement envisagées (comme le paléophone de Charles Cros) ou effectivement réalisées (comme le téléphone de Graham Bell). La cartographie de la construction des mots peut ainsi être un moyen d'approcher l'imaginaire des médias de l'époque. Les noms donnés aux inventions imaginaires livrent des éléments prégnants du discours social. La distribution des éléments constitutifs (préfixes, suffixes, ...) de ces noms au moyen d'une cartographie doit mettre en évidence l'imaginaire social dans lesquels naissent ces dispositifs médiatiques inventés. Ces inventions fictives, qui préparent, accompagnent et prolongent l'émergence des nouveaux médias, et leur mise en relation, révèlent l'univers mental dans lequel elles ont été imaginées largement dominés par des inventions réelles (les technologies de communications mais aussi des inventions scientifiques parfois anciennes comme le télescope), l'écriture (creuset dans lequel les inventions imaginaires naissent) et l'utopie d'une communication universelle sans entraves aussi bien par-delà les distances que par-delà le temps.

Le corpus indique, dans la perspective médiarchéologique, des désirs, des fantasmes et des utopies communicationnels [Breton, 1997] qui ont pu devenir réalité ou qui continuent à sous-tendre des recherches technologiques contemporaines entre objets, télécommunications et réseaux [Musso, 1998].

Donner à voir « des histoires oubliées » c'est mesurer des potentialités qui y sont inscrites et suivre les « boucles incluant le nouveau dans l'ancien au sein des discours sur les media actuels » [Parikka, 2015].

Bibliographie

- Philippe Breton, L'utopie de la communication. Le mythe du « village planétaire », Paris, La Découverte, « Poche/Sciences humaines et sociales », 2004.
- Yves Citton, Médiarchie, Paris, Seuil, coll. « La couleur des idées », 2017.
- Robert Escarpit, Théorie générale de la communication et de la communication, Paris, Hachette Université, coll. « Langue-Linguistique-Communication », 1976.
- Michel Foucault, Archéologie du savoir, Paris, Gallimard, coll. « Bibliothèque des sciences humaines », 1969.
- Erkki Huhtamo, « L'envoutement de la télévision catoptrique. Archéologie des media, étude des topoi et traces attentionnelles », in Yves Citton et Estelle Doudet (dir.), Écologies de l'attention et archéologie des media, Grenoble, UGA Éditions, coll. « Savoirs littéraires et imaginaires », 2019.
- Jussi Parikka, Qu'est-ce que l'archéologie des media ?, Grenoble, UGA Éditions, coll. « Savoirs littéraires et imaginaires scientifiques », 2018.
- Sarah Mombert, « L'actualité de l'avenir. L'invention du phonographe dans la presse et la fiction », in Claire Barel-Moisan, Jean-François Chassay, Le roman des possibles. L'anticipation dans l'espace médiatique francophone (1860-1940), Montréal, PUM, coll. « Cavales », 2019.

Pierre Musso, Télécommunications et philosophie des réseaux. La postérité paradoxale de Saint-Simon. Presses Universitaires de France, coll. « La Politique éclatée », 1998.

Géo-visualisation des contenus de la Perséide Athar : le cas du Caire moderne. Un exemple d'intégration de données documentaires, sémantiques et spatiales au service des historiens

Bulle Tuil Leonetti¹, Hélène Bégnis², Julie Erismann

1 : Institut national d'histoire de l'art (INHA)

2 : UMS 3602 Persée, CNRS, École normale supérieure de Lyon

Initié dans le cadre de l'appel à projet lancé en 2018 par le CollEx-Persée pour susciter des initiatives en lien avec ses programmes de numérisation et de développement de services au chercheur, le projet intitulé « Géo-visualisation de contenus de la Perséide Athar : le cas du Caire moderne » a eu comme objectif la mise à disposition en ligne sur ladite Perséide un corpus iconographique thématique outillé.

Il a permis la structuration au niveau infra-paginaire d'un ensemble de cinq albums documentant la ville moderne du Caire au moyen de photographies originales ou issues de publications, ainsi que de coupures de presse. Cette structuration a été enrichie par un référentiel des toponymes mentionnés en légende ou identifiés, servant de point d'entrée pour la consultation dans les albums. Intitulé Modern Cairo Gazetteer, ce référentiel géré avec OpenTheso et hébergé par Huma Num sert de pivot pour alimenter un plugin QGIS et un WEB SIG afin de visualiser l'ensemble des données et leur lien vers Athar et Gallica.

L'enjeu scientifique et documentaire de ce projet a été de rendre compte du constant renouvellement urbain de la ville du Caire moderne grâce à un fonds encore inexploré car assez atypique. Du point de vue technologique, l'ambition a été de faire parler des données documentaires issues du travail de segmentation des albums photographiques (métadonnées à la fois en Dublin Core, MarcXml, MODS et TEI dans un container METS) alignées sur des données sémantisées (thesaurus en SKOS), des images issues d'un serveur IIIF (Gallica) et des données spatiales.

Les modalités de consultation de ce corpus sont multiples et constituent autant de points d'entrée complémentaires en réponse à des besoins particuliers :

- Accès en ligne et en libre accès aux albums photographiques sur la Perséide Athar et grâce à un moteur de recherche, un feuillage photographie par photographie pour chacun des albums et une liste des édifices produite grâce aux alignements vers le référentiel ad hoc.
- Consultation thématique des édifices grâce au référentiel toponymique qui décrit, géolocalise et source chaque édifice.
- Exploration experte du Caire moderne par le biais de ses édifices grâce un module de requête permettant d'interroger les différentes bases de données mises en œuvre et d'afficher ces résultats sur un fond cartographique géré dans QGIS.
- Exploration par des non-spécialistes grâce à une web application permettant de faire des recherches historiques à partir de ces mêmes données.

Ce projet reposait sur un partenariat engageant le laboratoire InVisu (CNRS/INHA), porteur du projet, l'UMS Persée (CNRS et ENS de Lyon), l'Institut français d'archéologie orientale – IFAO, et l'École normale supérieure de Géomatique – ENSG.

La médiation du patrimoine culturel et l'éducation : le rôle de la plate-forme numérique *e-learning* dans la conservation et la transmission de patrimoine

Radouane Mimoun El Khir¹, Abdelelah Ounana

1 : Centre d'études doctorales Lettres, sciences humaines et arts, université Mohamed-1^{er}, Oujda, Maroc

Le patrimoine culturel occupe une place majeure dans le développement social et économique des pays, il est aussi un motif qui reflète l'histoire et l'identité des nations[1].

La médiation dans le domaine culturel englobe un ensemble de moyens d'information et de communication au service d'une interprétation variée et enrichie du patrimoine. Les options de cette médiation sont aussi vastes que l'est l'esprit de création, mais elles restent toutefois limitées par une logique politique et institutionnelle[2].

La médiation du patrimoine culturel (MPC) est généralement définie comme un processus de facilitation de la communication entre les objets et le public dans ce sens, elle pourrait être envisagée sous divers aspects : comme une transmission, comme une transformation, comme une action ou, encore, comme un projet social qui vise la création des liens sociaux [3] ; les œuvres d'art, les objets culturels, les valeurs culturelles et les systèmes de symboles. La médiation du patrimoine culturel est une approche visant à stimuler la prise de conscience des « apprenants » de leur patrimoine proche, à favoriser l'accessibilité à des domaines moins connus et à encourager la capacité à penser et à agir de manière créative. Grâce à cette approche, les étudiants sont en mesure de se mettre en relation avec l'identité culturelle ou un objet culturel et de développer leur opinion à son sujet. Par conséquent, donner accès à ces objets et les expliquer sont des priorités essentielles de la médiation du patrimoine culturel.

Les établissements d'enseignement constituent un contexte privilégié pour l'accès à des produits du patrimoine culturel diversifiés et dynamiques à travers l'usage des plateformes numériques qui peuvent être des moyens à la fois d'apprentissages et de la communication et aussi à la conservation et la transmission du patrimoine, Cela permet de crier des possibilités de transition des connaissances « anciennes » à nos « temps modernes ».

Cette étude montre comment la MPC aborde les défis de la procédure d'apprentissage croissante et l'utilisation des Technologies de l'Information et de la Communication. Il décrit l'usage de la plateforme *e-learning* dans l'enseignement supérieur pour conserver et transmettre le patrimoine culturel du point de vue théorique et pratique.

Ce projet de recherche étudie comment les outils TIC, en particulier les plates-formes numériques, peuvent contribuer à améliorer l'éducation au patrimoine culturel.

En effet, les plateformes numériques offrent, d'une part, un accès plus facile et une vision multi-perspective des objets du patrimoine culturel et, d'autre part, peuvent également enrichir et améliorer l'éducation au patrimoine culturel grâce à l'adoption des méthodes pédagogiques innovantes.

L'objectif de ce projet est de déterminer comment les plates-formes peuvent apporter une valeur ajoutée à la pédagogie dite numérique pour la conservation et la transmission du patrimoine culturel.

Dans le cadre du colloque HUMANISTICA 2021 notre contribution sera consacrée à la description de la plateforme *E-learning* ainsi la méthodologie du travail que nous avons adopté pour la conservation et la transmission du patrimoine matériel, précisément l'architecture Ksourienne dans le Sud-Est du Maroc et fournit des exemples d'expériences innovantes dans ce domaine, telles que la transformation du Ksar en

musée virtuel, amenant ainsi le lecteur et le visiteur à réfléchir sur l'impact pédagogique qui peut découler de l'exploitation des potentialités de l'usage numérique.

[1] Salma Trabelsi, « Développement local et valorisation du patrimoine culturel fragile : le rôle médiateur des ONG : cas du Sud tunisien », 310.

[2] Selma Zaiane-Ghali, « Médiation culturelle pour la sauvegarde et la valorisation du patrimoine tunisien : des initiatives citoyennes », *Ethnologies*, 38, 1-2 (2017), 127-153.

[3] Nada Guzin Lukic, « Patrimoine, musée et médiation », 19.

Langues celtiques, typographie et hybridation

Clémence Jaron¹

1 : EA 7472 Pratiques et théories de l'art contemporain (PTAC), université Rennes 2

Le projet « Langues celtiques, typographie et hybridation » est un projet mêlant typographie et patrimoine linguistique. L'idée d'hybrider les lettres est une façon de faire apparaître les ressemblances et différenciations de retranscriptions écrites (variations selon les mots, élisions, remplacement de voyelles ou ajouts de consonnes) qui s'établissent entre trois langues cousines celtiques brittoniques que sont le breton, le gallois et le cornique.

Le recueil de données s'est concentré autour du vocabulaire géographique commun, lié au paysage maritime, des territoires qui accueillent ces langues : Bretagne, Cornouailles et Pays de Galles (Daniel Le Bris, Mael Jézéquel, ALCAM, 2018). À partir de ces trois dictionnaires en breton, Geriafurch (<https://geriafurch.bzh/fr?q=&d=frbr>), Favereau (<http://www.arkaevraz.net/dicobzh/index.php>) et Glosbe (<https://fr.glosbe.com>) ainsi que d'un dictionnaire en gallois de l'université du Pays de Galles (<https://geiriadur.uwtsd.ac.uk>) et le passage d'un traducteur français-anglais (<https://www.linguee.fr>) pour accéder à la traduction en cornique depuis l'anglais sur le site Gerlyver Kernewek (<https://www.cornishdictionary.org.uk>) j'ai pu faire émerger des résultats qui ont constitué mon lexique en prenant un soin particulier à la terminologie exacte des mots dans chaque langue. La lacune évidente d'un site permettant leur rapprochement (phonétique et scriptural) fut le point de départ de recherche d'un biais typographique pour les représenter. Il s'agit d'une représentation graphique et non d'une solution de traduction de ces langues.

Du côté de la recherche graphique, la création d'une typographie géométrique a été la première étape. L'usage de la géométrie, en typographie, permet d'identifier des formes communes à un certain nombre de lettres. À partir de l'hybridation géométrique des lettres (notamment dans l'alphabet latin) on a pu se rendre compte que l'on pouvait avoir une hybridation de ces alphabets et des langues (Frutiger, 2014 [1978-1981]). Ce choix a notamment été proposé avec trois versions qui diffèrent en épaisseur ('graisse' en terme technique typographique).

Une grille de fond a été utilisée car elle est utile comme base de construction typographique. La navigation sur la carte et l'action de l'utilisateur sur des points cliquables permet de faire apparaître les mots hybridés et leur traduction directement sur cette grille. Ainsi, une fois les langues hybridées, on peut deviner l'appartenance des lettres présentes dans les variations. Le résultat obtenu serait donc communiqué sur une interface web sous la forme d'une carte imaginaire, qui localise les spécificités linguistiques (Boussidan *et al.*, 2012).

Ce projet a pour vocation de recherche graphique d'apporter un nouvel axe d'exploration pour proposer un éclaircissement dans les études sur la visualisation graphique de la linguistique. En effet, la typographie permet une lecture immédiate de la proximité des retranscriptions écrites de ces mots qui permet de comprendre le lien qui s'établit entre ces langues, inexistant sur d'autres interfaces de traduction. La création d'une solution visuelle à l'expression de la linguistique pourrait ainsi devenir un véritable outil qui est transposable à toutes langues et/ou aire géographique.

Le projet ITHAC entre méthodes numériques, recherche multidisciplinaire et propositions pédagogiques

Diandra Cristache¹, Anne Garcia-Fernandez¹, Elisabeth Greslou¹, Arnaud Bey¹, Pascale Paré-Rey², Malika Bastin-Hammou¹

1 : UMR 5316 Arts et pratiques du texte, de l'image, de l'écran et de la scène (Litt&Arts), CNRS, université Grenoble-Alpes

2 : UMR 5189 Histoire et sources des mondes antiques, CNRS, université Jean-Monnet (Saint-Étienne), université Jean Moulin – Lyon 3, université de Lyon, université Lumière – Lyon 2, École normale supérieure de Lyon

Le projet ITHAC a pour objectif l'analyse, la traduction et l'exploration numérique des travaux savants du XVI^e siècle consacrés aux théâtre antique latin et grec [1]. Ce projet collaboratif rassemble un consortium de 20 enseignants-chercheurs, ingénieurs de recherche, doctorants et stagiaires des universités de Grenoble et de Lyon, relevant principalement de deux disciplines : l'histoire du théâtre de l'Antiquité à la Renaissance d'une part, les humanités numériques de l'autre. L'objectif de cette alliance est de rassembler, numériser, transcrire, traduire et explorer un corpus patrimonial délaissé jusqu'à aujourd'hui parce que physiquement inaccessible, étant dispersé dans une multiplicité de bibliothèques européennes ; et parce que rédigé principalement en latin, et trop vaste pour être abordé dans son entier par un seul individu, voire par un groupe d'individu. Un premier enjeu est ainsi de disposer d'un lieu unique où données et membres du projet pourront se retrouver pour étudier ces sources textuelles en tant que corpus.

Le corpus paratextuel du projet est numérisé, traduit en français et encodé en XML-TEI. Cela doit permettre non seulement sa diffusion, mais aussi le travail collaboratif de l'équipe à distance et sa lecture distante[2]. Du calendrier aux modalités de rencontres des différents *workpackages* en passant par l'accès à la bibliographie et à la formation des doctorants et stagiaires impliqués, tout est mis en œuvre pour permettre le travail d'une équipe explosée géographiquement sur un corpus explosé géographiquement. Ainsi, un laboratoire numérique [3] sous la forme d'une plateforme d'exploration et de visualisation des données évolutive est en cours de développement. Déjà opérationnelle pour les usages internes du projet, elle permet d'ores et déjà de partager les textes étudiés avec l'ensemble de l'équipe, ainsi que de tester certaines fonctionnalités de collaboration à distance.

Par ailleurs, ITHAC comporte une composante forte de formation universitaire en proposant des missions de stage aux étudiants de licence et de master. Les stagiaires sont chargé.e.s de transcrire des paratextes du latin ou du grec en partant d'éditions numérisées accessibles en ligne ; puis, d'effectuer des recherches autour de ceux-ci ; parfois, de les traduire ; enfin, d'encoder les descriptions, les transcriptions et les traductions au format XML-TEI. Les modalités d'encadrement des stagiaires sont adaptées en conséquence : ceux-ci sont non seulement encadrés par les enseignants-chercheurs en charge des différents auteurs anciens, mais aussi formés aux techniques d'encodage par les ingénieurs du projet. Ces missions de stage, partagée entre science et technique et enrichies par un double encadrement, ont pour effet de sensibiliser les étudiant.e.s à l'alliage délicat entre compétences humanistes et compétences numériques. Ainsi, le projet IThAC propose un terrain d'apprentissage et d'expérimentation pour les *digital humanists* de demain.

Soucieux de faire de « bonnes humanités numériques » en produisant un PGD et en suivant les principes FAIR, le projet ITHAC part cependant d'un principe fort : c'est le projet scientifique et ses différents objectifs qui guident le travail, de la constitution du corpus aux outils numériques mis en œuvre. La communication permettra d'aborder le difficile équilibre à trouver en fonction des moyens, du temps alloué, des forces disponibles et de la situation, entre la réalisation d'un projet répondant à un objectif scientifique précis et le respect de normes et de principes liés aux HN [4].

Dans cette communication, nous nous proposerons, tout d'abord, de présenter le projet IThAC en nous focalisant sur les progrès scientifiques et numériques réalisés dans ses premiers douze mois de fonctionnement ; puis, nous examinerons les problématiques méthodologiques liées à la mise en place de

solutions computationnelles adaptées à la collaboration entre sciences de l'antiquité et sciences du spectacle ; enfin, nous interrogerons le potentiel pédagogique du projet, partagé entre préparation à la recherche scientifique, formation aux méthodes numériques et sensibilisation aux équilibres complexes d'un projet multidisciplinaires en humanités numériques.

[1] Malika Bastin-Hammou et Pascale Paré-Rey, « ITHAC – L'Invention du Théâtre Antique dans le Corpus des paratextes savants du XVI^e siècle. Analyse, traduction, exploration numérique », *Anabases*, 32 | 2020, 185-190. <https://doi.org/10.4000/anabases.11307>

[2] Moretti (Franco), *Distant Reading*, London, New York, Verso, 2013.

[3] Anne Garcia-Fernandez, Isabelle Cogitore. Laboratoire numérique pour l'étude de paratextes : l'exemple de Tacitus On Line. *Digital Humanities 2019*, Jul 2019, Utrecht, Pays-Bas. hal-02336294

[4] Elena Pierazzo. *Digital Scholarly Editing : Theories, Models and Methods*. 2014. hal-01182162

Le projet « Démêler le cordel » : une bibliothèque numérique pour l'étude de la littérature éphémère espagnole du XIX^e siècle

Constance Carta¹, Elina Leblanc¹

1 : Université de Genève

Si les imprimés éphémères, produits dans toute l'Europe (et au-delà) depuis les débuts de l'imprimerie, se déclinent en de nombreux genres et formats, leur variante espagnole, les *pliegos de cordel*, consistent plus précisément en de courts textes de quelques pages (en général, quatre ou huit), de format in-quarto, imprimés sur du papier de piètre qualité et ornés de gravures sur bois, destinés à être vendus à l'unité à un public large incluant les personnes les moins instruites. La constance de leur format, du XV^e au XX^e siècle, en fait un genre éditorial, alors même que leurs contenus sont extrêmement variés : versions abrégées de romans ou de pièces de théâtre à la mode, récits d'événements réels ou fictifs d'importance nationale ou locale faisant la part belle aux bandits et aux assassins, poésies, chansons et comédies, ou encore écrits pieux. Longtemps dépréciés par la critique du fait de leur piètre qualité, tant en ce qui concerne leur facture que leur écriture et leurs contenus, on reconnaît depuis quelques décennies seulement leur importance dans de nombreux domaines, notamment la culture populaire, l'histoire du livre, la circulation des idées et des images, les processus de réécriture ou encore l'importance de la transmission orale, en ce qu'ils fournissent un pendant essentiel à l'étude de la grande Histoire et des œuvres littéraires canoniques.

La bibliothèque universitaire de Genève est le lieu de conservation d'une collection de près de mille feuillets de colportage espagnols ou *pliegos de cordel*. Notre projet, à la croisée de diverses sciences humaines et des humanités numériques, se centre sur un corpus de près de 400 de ces *pliegos* qui émanent des ateliers d'un seul et même imprimeur andalou de la deuxième moitié du XIX^e siècle : José María Moreno (Carta 2017, Palacios 2017)

Ce projet s'inscrit dans un courant de promotion et de valorisation numérique de la littérature de *cordel*, qui a conduit, au cours de ces dernières années, à l'émergence de nombreuses bibliothèques numériques qui lui sont dédiées, à l'image des projets *Mapping Pliegos*, *Impresos Populares Iberoamericanos* ou encore la *Biblioteca Virtual Cordel* [1]. Bien qu'aucun consensus n'ait à ce jour été trouvé pour représenter ce type de littérature, qui échappe aux standards des bibliothèques en raison de son hétérogénéité et de sa multiplicité (Sánchez Carretero 2000, Baena Sánchez *et al.* 2013, Hague 2015, Nieto 2015), la plupart de ces bibliothèques numériques tendent à se concentrer sur la description des *pliegos* en tant qu'objets. Par conséquent, dans le cadre de notre projet, si une attention particulière est accordée à la description des feuillets, l'accent est davantage mis sur l'exploration et l'analyse de leur contenu textuel et iconographique.

Après avoir OCRisé les *pliegos*, nous avons ainsi défini un modèle XML-TEI, adapté à l'hétérogénéité typologique et thématique des textes et permettant d'étudier les *pliegos* sous différents angles (littéraire, historique, sociologique et linguistique). Nous avons également extrait les gravures afin de créer un catalogue XML-TEI des tampons utilisés dans notre corpus. Ce travail conduira à une analyse des images dans leur relation aux textes, ainsi qu'en tant que tampons réutilisés, échangés et copiés entre imprimeurs au XIX^e siècle.

La TEI constitue la clé de voute de notre projet à partir de laquelle nous générerons notre site Web, à la fois bibliothèque numérique et outil de recherche, mais également des manifestes IIIF, afin de partager nos images avec d'autres projets, étudier les réimpressions d'un même *pliego* à travers plusieurs corpus, et comparer des gravures similaires, voire identiques. Le couple TEI-IIIF nous permet ainsi de proposer un modèle, qui sera l'objet de notre poster, pour représenter les textes des *pliegos*, leurs gravures et les liens qui les unissent. Il s'agit ainsi de gérer la complexité de la littérature de colportage, à la fois objet, texte et image.

Baena Sánchez, Francisco, Casas-Delgado, Inmaculada, Cuadros Palomares, M^a Teresa. 2013. « Una biblioteca digital de prensa antigua. El caso de las relaciones de sucesos sevillanas (siglo XVI-XVIII) ». *RIHC : Revista internacional de Historia de la Comunicación*, 1(1), 1-32.
<https://idus.us.es/handle/11441/24774;jsessionid=F483766C09F5F9E0124018D9C9DF5180>

Carta, Constance. 2017. « Primera aproximación a un parnaso popular decimonónico : la colección de pliegos de Carmona en la Biblioteca Universitaria de Ginebra ». In Abraham Madroñal y Carlos Mata Induráin (éds.), *El Parnaso de Cervantes y otros parnasos*, IDEA/IGAS, New York, p. 131-155.

Hague, Julien. 2015. « Les éphémères de l'âge de l'imprimé à l'ère numérique. Un champ disciplinaire en révolution ». *Fabula Colloques. Les éphémères, un patrimoine à construire*.
<https://www.fabula.org:443/colloques/document2925.php>.

Nieto, Philippe. 2015. « Cataloguer les éphémères. Quelques pistes de réflexion ». *Fabula Colloques. Les éphémères, un patrimoine à construire*. <https://www.fabula.org:443/colloques/document2896.php>.

Palacios, Belinda. 2017. « Índice de la colección de pliegos sueltos de la Biblioteca de la Universidad de Ginebra ». In Abraham Madroñal y Carlos Mata Induráin (éds.), *El Parnaso de Cervantes y otros parnasos*, IDEA/IGAS, New York, p. 285-303.

Sánchez Carretero, Cristina. 2000. « De historias y romances : las clasificaciones de los géneros editoriales y textuales en los pliegos de cordel ». In Luis Díaz Viana (coord.), *Palabras para el pueblo*, Vol. 1 (Aproximación general a la Literatura de Cordel), p. 429-486.

[1] *Mapping Pliegos* est un projet multi-partenarial conduit par la *Cambridge University Library*, la *Biblioteca Tomás Navarro Tomás del CSIC* et la *British Library* (<http://biblioteca.cchs.csic.es/MappingPliegos>). Il propose un portail numérique agrégeant l'ensemble des collections de *pliegos* numérisés disponibles en ligne.

Impresos Populares Iberoamericanos (<http://literaturaspopulares.org/ipm/w/Inicio>) et la *Biblioteca Virtual Cordel* (<http://cordel.edel.univ-poitiers.fr>) sont deux bibliothèques numériques dédiées aux *pliegos* de cordel. Elles sont portées par l'Universidad Nacional Autónoma de México pour la première, et par le Centre de Recherches Latino-Américaines de l'université de Poitiers pour la seconde.

Les bibliothécaires et leur rapport au livre numérique au Québec : vers une vision intégrative du patrimoine écrit

Christelle Pelbois¹, Marie-Claude Lapointe¹, Jason Luckerhoff¹

1 : Université du Québec à Trois-Rivières

Le livre numérique occupe encore, à l'heure actuelle, une place plutôt marginale au Québec, autant du point de vue économique que du point de vue social (Labrousse et Lapointe, 2020). Selon une enquête sur les pratiques culturelles, réalisée en 2014 par le Ministère de la Culture et des Communications du Québec (2016), alors que 86 % de la population québécoise ayant répondu à l'enquête utilise Internet à des fins personnelles, seulement 11.7 % des répondants déclarent lire des livres numériques un fois par mois et plus. Comparativement, 58.5 % des répondants déclarent lire des livres papier selon la même fréquence. Malgré des pratiques numériques répandues (MCCQ, 2016 ; CEFRIO, 2019) et l'accessibilité accrue à une lecture en tout temps et en tout lieu que le livre numérique rend possible, ce dernier ne semble pas connaître un franc succès auprès de tous les lecteurs. La conversion des lecteurs québécois de livres imprimés à la lecture numérique, les raisons pour lesquelles ces lecteurs ne lisent pas ou peu de livres numériques, et leur réserve face à l'usage de ces derniers sont actuellement encore relativement peu appréhendées. Dans un article récent, Labrousse et Lapointe (2020) avancent que non seulement les lecteurs de livres imprimés accordent à ceux-ci une signification profonde en tant qu'objets matériels, mais aussi que les contraintes d'usage associées aux livres numériques et l'expérience de lecture qu'ils procurent, expliquent certaines réticences d'usage de livres numériques par ces lecteurs. Ainsi, le livre numérique se définirait par ce qu'il n'est pas : un livre imprimé. Il se caractériserait aussi par ce qu'il ne permettrait pas ou peu, comparativement au livre imprimé : un rapport riche et affectif à l'objet-livre, une expérience sensorielle mobilisant le toucher, l'odorat, la vue et même l'ouïe (Labrousse et Lapointe, 2020).

Toutefois, une augmentation récente de l'utilisation du livre numérique en bibliothèque publique au Québec est actuellement constatée, notamment en raison du contexte sanitaire de la COVID-19, contexte ayant conduit à la fermeture plus ou moins prolongée des accès physiques aux bibliothèques. Cette augmentation actualise certaines questions relatives au livre numérique. En effet, si nous comprenons mieux désormais les usages et perceptions des lecteurs quant au livre numérique, la transformation de l'image des bibliothèques et de ses professionnels de l'information, dont la mission est encore au Québec fortement symbolisée par un lieu physique et des livres imprimés, demeure un objet d'étude peu exploré. Pour ces professionnels des bibliothèques publiques, parler du livre numérique ne se résume pourtant pas à l'opposer au livre imprimé, ou à confronter l'immatérialité de l'un à la matérialité de l'autre. Dans quelle mesure le mythe selon lequel « plus de livres numériques, c'est moins de livres imprimés en bibliothèque » est-il présent dans le quotidien des bibliothécaires ? Pour ces derniers, quelle question le livre numérique pose-t-il actuellement aux bibliothèques publiques ? Le livre numérique doit-il ou ne doit-il pas y être présent ? Comment peut-il l'être ?

L'objectif de notre contribution est d'apporter des pistes de réponses à ces questions. Nous avons à cet effet effectué, à l'automne 2020, une enquête auprès de bibliothécaires. Une cinquantaine d'entrevues individuelles et de groupe ont ainsi été animées, dans le cadre d'une approche inductive, auprès de bibliothécaires de diverses bibliothèques publiques. Nos résultats présentent une réalité du rapport au livre numérique en bibliothèque publique très nuancée. Ainsi, les participants à notre enquête ne perçoivent pas le livre numérique d'une façon réductrice l'opposant au livre imprimé. Le regard qu'ils posent est un regard pragmatique réconciliant l'apparente opposition entre deux types de support pour en proposer une vision intégrative de patrimoine écrit. À l'instar des résultats de Labrousse et Lapointe (2021), les participants perçoivent que le numérique peut faciliter l'accès circonscrit à certains contenus écrits. Les bibliothécaires invoquent aussi d'importantes contraintes techniques (connaissances informatiques ou technologiques) à

surmonter pour accéder au premier prêt numérique. Le principal effet du livre numérique sur la pratique quotidienne des bibliothécaires concernerait ainsi la modification de certains de leurs rôles. L'un de ces rôles est le soutien technique devant être apporté aux publics des bibliothèques afin d'accéder au livre numérique et un autre concerne la formation des publics des bibliothèques à la littératie numérique. Des questionnements principalement centrés sur les conditions de réalisation de leur mission auprès des publics sont présents. Ces questionnements traduisent de nombreux défis encore peu résolus par les participants : l'appropriation par les publics de ce nouveau type de support à la lecture, la valorisation de cette nouvelle collection numérique, notamment pour faire connaître aux publics l'existence d'un choix ; mais aussi le défi de la connaissance de ces publics, de leur besoins, préférences, attentes et pratiques. À la lumière des regards croisés de bibliothécaires, notre enquête permet d'enrichir l'appréhension du phénomène du livre numérique en bibliothèque publique au Québec en tentant de dépasser l'appauvrissement d'une discussion polarisée sur sa seule opposition au livre imprimé.

Références bibliographiques

Centre Facilitant la Recherche et l'Innovation dans les Organisations. 2020. « Portrait numérique des foyers québécois, édition 2019 ». NETendances, 10(4). Consulté le 3 janvier 2021. <https://transformation-numerique.ulaval.ca/wp-content/uploads/2020/09/netendances-2019-portrait-numerique-des-foyers-quebecois.pdf>

Labrousse, Marie et Marie-Claude Lapointe. 2021. « Perception du livre numérique auprès des lecteurs québécois de livres papier ». Sous presse.

Ministère de la Culture et des Communications du Québec (Canada). 2016. « Enquête sur les pratiques culturelles 2014 – Faits saillants de l'enquête ». Gouvernement du Québec. Consulté le 3 janvier 2021. https://www.mcc.gouv.qc.ca/fileadmin/documents/publications/Enquete_pratiques_culturelles/Bulletin-Survol-27-_2014.pdf

Ministère de la Culture et des Communications du Québec (Canada). 2016. « Enquête sur les pratiques culturelles 2014 – Recueil statistique ». Gouvernement du Québec. Consulté le 3 janvier 2021. https://www.mcc.gouv.qc.ca/fileadmin/documents/publications/Enquete_pratiques_culturelles/Recueil-Stat2014-Vol1.pdf

Paul Otlet en son réseau. Présentation de l'Otletosphère

Olivier Le Deuff¹, Arthur Perret¹, Clément Borel, Guillaume Brioude

1 : UR 4426 Médiation, information, communication, art (MICA), université Bordeaux-Montaigne

L'objectif de ce poster est de présenter le travail effectué pour représenter le réseau de Paul Otlet, avocat, pionnier de la documentation (Otlet, 1934) et membre actif des milieux pacifistes et associations internationales. Ce travail a été réalisé dans le cadre du programme ANR HyperOtlet. Les différentes activités d'Otlet lui ont permis de côtoyer un large nombre de personnalités scientifiques, artistiques et politiques (Blanquet, 2011) avec lesquels il a pu entretenir des correspondances, se lier d'amitié, ou collaborer sur des projets communs. Ces relations sont connues ou attestées par l'examen des archives de la correspondance de Paul Otlet du Mundaneum[1].

L'Otletosphère est une représentation interactive de ce réseau. Elle est désormais disponible en ligne dans sa version 2.0 : <http://hyperotlet.huma-num.fr/otletosphere/>. Ceci fait suite à une première expérimentation déjà décrite (Le Deuff *et al.*, 2019). Dans cette nouvelle version, nous avons entièrement refondu l'interface, la structuration des données ainsi que l'ensemble des bibliothèques de code utilisées.

Le projet développe une approche qui mêle la fiche et le graphe (Le Deuff, 2019). L'inspiration initiale provient du projet de représentation graphique de l'univers transmedia Marvel produite par l'équipe du journal Singapourien *The Straits Time*[2]. Celle-ci se présente comme une véritable constellation de super-héros, dessinée d'après leurs relations au sein des univers narratifs.

L'objectif du projet était de parvenir à montrer l'intérêt d'envisager de nouvelles approches graphiques en matière d'humanités digitales (Le Deuff, 2018). Il s'agissait également de répondre aux projections avant-gardistes de Paul Otlet en matière de représentation, d'organisation et d'accessibilité à la connaissance (Le Deuff, Perret, 2019)

Nous avons réalisé une base de données qui reprend les différentes personnes avec qui Paul Otlet est entré en relation, en qualifiant la nature de la relation suivant une ontologie dédiée.

Le dispositif est accessible sur tous les supports et permet différentes utilisations tant scientifiques que pédagogiques.

1. Présentation du dispositif

L'Otletosphère répond aux objectifs suivants :

- créer un graphique représentant les différents acteurs présents autour d'Otlet (nœuds) et les relations qui les lient (liens) ;
- pour chacun de ces acteurs, permettre l'accès, de manière interactive, à une fiche documentant des informations utiles à leur sujet, notamment pour mieux comprendre la finalité de leur présence autour de Paul Otlet ;
- permettre de filtrer l'affichage selon l'ontologie créée (par exemple, ne montrer que la sphère familiale de Paul Otlet, ou uniquement ses collaborateurs). L'ontologie comprend les catégories suivantes : Otlet, Famille, Collègues, Collaborateurs, Opposants, Œuvres, Événements, Institutions.

2. Dispositif technique. Bibliothèques de code mobilisées

L'équipe de développement a recouru aux bibliothèques suivantes :

JavaScript Vis.js (v7.10.2)

La bibliothèque JavaScript Vis.js nous a permis de réaliser la visualisation de la section *Réseau* grâce à deux de ses composants : *Network* et *DataSet*. La bibliothèque nous a permis de traiter les données (stockage, tri, circulation de données entre les fonctions), mais aussi de générer le canevas, son animation et enfin la gestion des événements en son sein (sélection, survol, zoom).

Fuse.js (v6.4.1)

La bibliothèque JavaScript Fuse.js nous a permis de mettre en place le moteur de recherche d'entité. Elle permet de parcourir des tableaux d'objets JavaScript, entre autres, avec un critère de recherche afin d'en extraire les objets les plus pertinents.

Gulp.js (v4.0.2)

Gulp.js est également mobilisé afin de faciliter la mise en œuvre du développement suivant une logique modulaire.

CSS Bootstrap Grid (v4.5.0) Nous avons utilisé la partie Grid CSS de la bibliothèque Bootstrap afin de réaliser facilement une interface flexible à la fois pour tablettes et ordinateurs de bureau.

3. Documentation et licence

L'Otletosphère est un logiciel open source (licence MIT) archivé et citable : DOI : 10.5281/zenodo.3981189

Une documentation est également disponible : <https://hyperotlet.github.io/otletosphere/>

Bibliographie

Blanquet, Marie-France. 2011. « Paul Otlet et le cercle de ses contemporains. Précurseurs, comme lui, des sciences de l'information. », dans *Schéma et schématisation, revue de bibliologie*, no73, p. 15-33.

Le Deuff, Olivier. 2019. « Hyperdocumentation, hyperville, hyperloin, jusqu'où peut-on étudier l'hypertexte ? ». *H2PTM'19. De l'hypertexte aux humanités numériques*. Londres : ISTE éditions, sous la dir. de Roxin, Ioan et al., Londres : ISTE éditions, p. 47-62.

Le Deuff, Olivier., David, Jean., Perret, Arthur., Borel, Clément. 2019. « Surfer dans l'Otletosphère. Des outils pour visualiser et interroger le réseau de Paul Otlet ». *H2PTM'19. De l'hypertexte aux humanités numériques*. Londres : ISTE éditions, sous la dir. de Roxin, Ioan et al., p. 65-76.

Le Deuff, Olivier., Perret, Arthur. 2019. « Paul Otlet and the Ultimate Prospect of Documentation. », *Proceedings from the Document Academy* : Vol. 6 , 1, article 14. DOI : <https://doi.org/10.35492/docam/6/1/9>.

Le Deuff, Olivier. 2018 « HyperOtlet : entre héritage et design prospectif pour les humanités digitales » : communication scientifique au *5th Hyperheritage International Symposium : Heritage and experience design in the digital age | Patrimoines et design d'expérience à l'ère numérique*. Mai 2018, université de Constantine 3, Algérie. <https://f-origin.hypotheses.org/wp-content/blogs.dir/4632/files/2018/11/Ledeuffhyperotlet.pdf>

Le Deuff, Olivier. 2019. « La fiche entre économie informationnelle et attentionnelle ». *CIDE 21. 21ème colloque international sur le document numérique. Économie(s) du document : pratiques et perspectives liées au numérique*. (Jacquemin Bernard ; Ghenima Malek, dir.), Avril 2019, Djerba, Tunisie. p. 137-144.

Otlet, Paul. 1934. *Traité de documentation : le livre sur le livre : théorie et pratique*, Mundaneum-Palais Mo

[1] <http://archives.mundaneum.org/>

[2] <https://graphics.straitstimes.com/STI/STIMEDIA/Interactives/2018/04/marvel-cinematic-universe-whos-who-interactive/index.html>

Penser la phase post-pandémique au supérieur : quels processus d'appropriation des technologies éducatives dans les pratiques enseignantes ?

Rachid El Ganbour¹, Samira Elouelji¹, Mohammed Boukare¹, Morad El Ganbour²

1 : Centre d'études doctorales Lettres, sciences humaines et arts, université Mohamed-1^{er}, Oujda, Maroc

2 : Communication, linguistique, éducation et humanités numériques (CLEHN), faculté des lettres et des sciences humaines, université Mohamed-1^{er}, Oujda, Maroc

Après avoir longtemps été un complément optionnel à l'université, l'enseignement en ligne connaît aujourd'hui un regain d'intérêt et est devenu une réalité plus qu'une exigence. La situation d'urgence (Gilliot, 2020) qui a accompagné la pandémie COVID-19 a entraîné un passage abrupt de l'enseignement présentiel à l'enseignement en ligne, déstabilisant ainsi les pratiques enseignantes, qui flottent entre deux positions : l'une conservatrice et l'autre innovante.

Dans ce contexte de complexité et d'incertitude (Morin, 2020), notre problématique est née de l'observation des pratiques enseignantes (Altet, 2000) dans un milieu naturel à la faculté des lettres et des sciences humaines (désormais FLSHO), qui est l'un des huit établissements affiliés à l'université Mohammed Premier (UMP) à Oujda, Maroc. L'étude s'est étalée sur une période de six mois, couvrant à la fois la phase de continuité pédagogique au cours des trois derniers mois de la session d'automne (2019-2020) et les trois premiers mois de la session de printemps (2020-2021).

En décrivant le dispositif dans lequel s'inscrivent les actions des enseignants et des étudiants, nous cherchons à savoir quel était l'impact de cette présence inattendue et obligatoire du numérique sur les pratiques et les usages au sein des disciplines liées aux sciences humaines et sociales ?

Sur le plan méthodologique, la recherche se veut une étude de cas (Creswell, 2007) qui s'inscrit avantagement dans un paradigme qualitatif. Elle vise l'examen des usages et des pratiques des enseignants et des apprenants des disciplines liées aux sciences humaines et sociales, et s'appuie sur un ensemble d'outils de collecte de données, en l'occurrence l'observation, les traces numériques sur la plateforme (Moodle) de la faculté des lettres et des sciences humaines et les retours d'expériences des enseignants et des apprenants.

L'analyse des données numériques se base sur les fonctionnalités statistiques de la plateforme de la faculté, les plugins installés par l'équipe technique du Centre Universitaire d'Innovation Pédagogique et Digital Learning (CUIPDIL), quant aux données narratives, elles sont traitées par des logiciels d'analyse (Voyant tools et Tropes).

Les principaux résultats de la recherche ont montré que les frontières délimitant les pratiques habituelles en présence et émergentes en ligne ne sont pas encore tracées de manière claire et précise. L'étude a dévoilé également la diversité des perceptions et des pratiques professionnelles liées au numérique depuis la crise sanitaire, des difficultés de nature différentes (techniques, méthodologiques, communicationnelles...), mais également l'ampleur d'une dynamique productive en ligne et des pratiques innovantes qui s'y rattachent. L'ensemble de ces conclusions nous a permis de lister les entrées susceptibles de constituer les éléments incontournables d'appropriation des technologies éducatives au supérieur.

Bibliographie sélective

Altet M. (2000). L'analyse de pratiques : une démarche de formation professionnalisante ?, dans Recherche et formation. Paris : INRP, n° 35.

Creswell, J.W. (2007). Qualitative inquiry & research design : Choosing among five approaches (2^e éd.). Thousand Oaks, CA : Sage Publications.

Gilliot, J.-M. (2020). Enseigner à distance en situation d'urgence – Ce n'est pas de l'enseignement en ligne, c'est une mutation ! Récupéré le 02.05. 2020, sur Techniques innovantes pour l'enseignement supérieur : <https://tips.wordpress.com/2020/04/10/enseigner-a-distance-en-situation-durgence-ce-nest-pas-de-lenseignement-en-ligne-cest-une-mutation/>

Lim, E. C., Oh, V. M., Koh, D. R., & Seet, R. C. (2009). The challenges of « continuing medical education » in a pandemic era. *Annals, Academy of Medicine, Singapore*, 38(8), 724-726. https://www.researchgate.net/profile/Vernon_Oh2/publication/26795469_The_Challenges_of_Continuing_Medical_Education_in_a_Pandemic_Era/links/5554430108ac6943a86f479f.pdf

Papi, C. (2020). Supporting teachers in the transition towards distance education : Challenges and means. *Journal of Learning for Development*, 7(3), 467-472.

Papi, C. (2016). De l'évolution du métier d'enseignant à distance. *STICEF*, 23. http://sticef.univlemans.fr/num/vol2016/03-papi-ensaccapp/sticef_2016_NS_papi_03.htm

United-Nations.(2020). Note de synthèse : L'éducation en temps de COVID-19 et après : https://www.un.org/sites/un2.un.org/files/policy_brief_-_education_during_covid-19_and_beyond_french.pdf

Reconnaissance et certification dynamique dans les dispositifs contributifs du secondaire : opportunité et levier de développement des habiletés numériques des élèves et des enseignants

Franck Bodin¹, Guillaume Viniacourt

1 : Atelier Canopé 94, réseau Canopé

Dans le cadre du déploiement de dispositifs pédagogiques contributifs et collaboratifs (Datasprints pédagogiques, hackathons pédagogiques, annotathon, Wikiconcours, CulturesRemix, Marathon presse, Numook ...), Réseau Canopé élabore un schéma opérationnel de reconnaissance de compétences entre pairs centré autour des habiletés numériques et des compétences psycho-sociales. La grille s'inspire des référentiels de l'OCDE et est formalisée sur la plateforme Open Badge Factory. Elle tend à expliciter les compétences implicites sur lesquelles reposent le succès d'une proposition pédagogique explorant les champs des humanités numériques à l'École. Elle se fixe le double objectif de proposer d'une part aux enseignant un cadre clair, utile à construire une visée de développement professionnel et d'autre part au collectif de se constituer en communauté de pratiques efficiente et à géométrie variable.

Références

OpenBadge Factory (consultation 15/12/2020) : <https://openbadgefactory.com/>

Serge Ravet. Réflexions sur la genèse des Open Badges (2017) : <https://doi.org/10.4000/dms.2043>

Étienne Wenger. Les communautés de pratique au service de la connaissance située et de l'apprentissage social (1998) : <https://www.cairn.info/les-grands-auteurs-en-systemes-d-information--9782376871309-page-462.htm>

Franck Bodin. La classe contributive (2018) : <https://medium.com/digital-praxis/quest-ce-que-la-classe-contributive-b18d888d627f>

Réseau Canopé. Développement professionnel des enseignants (consultation 15/12/2020) : <https://www.reseau-canope.fr/developpement-professionnel-des-enseignants.html>

UNESCO : Référentiel des compétences de l'UNESCO (2018) : https://fr.unesco.org/sites/default/files/Competency%20Framework_F.pdf

UNESCO : Référentiel des compétences TICE de l'enseignant (2016) : http://unesdoc.unesco.org/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_a27805f1-d7c3-42f5-ab36-5f9984a254cc?_=368966fre.pdf&to=71&from=1

SITT : Streaming, interactivité, téléperformance, transmédia. Un projet de sauvegarde du patrimoine chorégraphique par le développement d'outils technologiques spécifiques

Armando Menicacci¹, Nicolas Berzi, Marie Lavorel

1 : Hexagram

La recherche-crédation que nous nous proposons ici d'exposer a pris forme à partir d'un constat préoccupant. Un récent sondage du Regroupement Québécois de la Danse (décembre 2020) montre que 52 % des danseurs au Québec désirent entamer une reconversion professionnelle et abandonner leur métier. L'un des effets que la pandémie aura sur les arts de la scène, en particulier dans les états qui n'accompagnent pas les artistes par des dispositifs de sécurité sociale, sera la déperdition professionnelle et le ralentissement de l'excellence artistique. La distanciation sociale provoque la perte de ce que Bourdieu nomme le capital culturel, ce qui nous a menés à réfléchir aux conséquences néfastes sur la deuxième forme de capital culturel décrite par le sociologue : sa forme objectivée. En effet, la danse est un art avec une tradition orale importante et seulement une quantité négligeable de chorégraphies sont notées et conservées. Le répertoire vit essentiellement dans les mémoires corporelles de ses interprètes. La reconversion précipitée de nombre de danseurs pourrait entraîner un double appauvrissement en danse comme dans les autres secteurs scéniques : une atteinte inestimable à son capital culturel incorporé ; la fuite de ses protagonistes, entraînant de surcroît la perte potentielle de nombre d'œuvres du patrimoine chorégraphique.

Pour tenter de pallier la situation, nous avons développé et mis en place *SITT : Streaming, Interactivité, Téléperformance Transmédia*. Ce projet met en œuvre des expertises théoriques et pratiques au croisement entre théâtre, danse, muséologie, communications et nouvelles technologies, tout en réunissant déjà près d'une trentaine de partenaires diffuseurs et de compagnies de création et de production chorégraphiques québécoises. *SITT* deviendra leur interface numérique via le développement de trois extensions logicielles (en forme de blocs de programmation) intégrés au sein d'Isadora, logiciel largement répandu dans la gestion en temps réel de tout type de média pour les arts de la scène. Un premier bloc permet de faire de la téléperformance réunissant des interprètes créateurs distants au sein de la même image. En plus du traitement téléperformatif, le deuxième permettra d'intégrer à Isadora le streaming. Le troisième permettra aux interprètes d'avoir une idée précise de l'état émotionnel des spectateurs internautes grâce à un outil d'humanité numérique unissant l'intelligence artificielle à la *sentiment analysis*. Tous les outils développés par SITT seront distribués gratuitement.

Suite aux développements logiciel, nous allons former des artistes techniciens à l'usage de ces outils et qui assurera, ensuite la diffusion en streaming des œuvres chorégraphiques des structures de diffusion partenaires. Ce dispositif-réseau pourrait alors répondre non seulement aux nouvelles propositions stratégiques des financeurs, mais surtout à la réalité des créateurs, des interprètes et des diffuseurs qui n'ont ni les compétences en intégration numérique ni les ressources pour du support en informatique-réseau. Nous avons pensé cet ensemble d'actions de développement informatique, de formation et d'accompagnement à la webdiffusion intégrée aux humanités numériques en fonction de la crise dans le secteur de la danse, mais il nous est clair que tout le secteur de la création en arts vivants pourrait en bénéficier. Pour cela nous avons conçu un plan de transfert de connaissances détaillé et articulé autour de nos outils

Les cofondateurs de SITT, Armando Menicacci (artiste numérique et chercheur, Ph.D en danse et informatique, Paris 8) et Nicolas Berzi (metteur en scène, Ph.D en pratique scénique intermédiaire, UQAM),

en compagnie de Marie Lavorel (membre du board Humanistica et directrice du plan de transfert de connaissances de SITT, Ph.D en muséologie et communication, UQAM/UAPV) proposent à Humanistica l'une des premières démonstrations pratiques et commentées de ce projet de recherche-crédation hybride faisant dialoguer technologie numérique et arts vivants au service du patrimoine chorégraphique et scénique.

Ateliers de groupe de travail

Humanités numériques, éducation et formation

Thibaud Hulin¹, Aude Seurrat²

1 : EA 4661 Édition, littératures, langages, informatique, arts, didactique, discours (ELLIADD), université de Franche-Comté

2 : LabSIC, université Paris 13, Labex ICCA

Forme proposée : discussion avec les participants du colloque, à partir d'interventions des chercheur·e·s du groupe thématique numérique « Humanités numériques et éducation » (HUMANE, GT Num 7), porté par le Groupement d'intérêt scientifique Innovation, interdisciplinarité, formation (GIS 2IF) soutenu par la Direction numérique pour l'éducation (MNEJS).

Le GIS 2IF est une structure fédérative et interdisciplinaire qui met en réseau des chercheurs et des équipes de recherche travaillant sur les phénomènes d'innovation dans la formation. Elle se focalise en particulier sur les phénomènes d'innovation technologique, d'industrialisation de la formation et sur les interactions entre innovation et industrialisation.

Argumentaire. Alors que les pratiques de formation à distance se massifient, la recherche croisée en humanités numériques dans le champ de l'éducation et de la formation soulève des questions vives. Quelles sont les compétences numériques à enseigner des plus jeunes élèves jusqu'aux adultes en formation initiale et continue ? Quelles sont les conséquences du décloisonnement des savoirs opérés par ces enseignements ? Comment recenser, décrire et partager les expériences pédagogiques ? Quelle est la place et quelles sont les spécificités des sciences humaines et sociales dans ces enseignements ?

Pour répondre à ces enjeux actuels, les intervenants s'appuieront sur l'expérience acquise et les réflexions menées collectivement dans le groupe de travail : pour décrire et partager les pratiques enseignantes, pour cartographier la sémantique du champ des « humanités numériques, éducation et formation », enfin pour analyser la mise en œuvre d'enseignements en lien avec le contexte social et l'orientation de l'apprenant et pour analyser les conceptions du numérique à l'œuvre dans ces enseignements.

Les intervenants s'appuieront sur leur expérience de la recherche dans le domaine, et présenteront les avancées du projet HUMANE qui vise à fédérer les ressources pédagogiques et les recherches alliant humanités numériques, éducation et formation.

Références

Baron, G.-L., & Bruillard, E. (1996). *L'informatique et ses usagers dans l'éducation*. PUF.

http://mutatice.net/glbaron/lib/exe/fetch.php/baron_bruillard_livre1996-fprepub2.pdf

Baron, G.-L. (2020). Brèves réflexions sur les humanités numériques. *Frantice*, 17, sous presse. prépubl :

http://mutatice.net/glbaron/lib/exe/fetch.php/hn_baron_frantice_v3.pdf.

Drot-Delange, B., Pellet, J-Ph., Delmas-Rigoutsos, Y. et Bruillard, E. (2019). *Pensée informatique : points de vue contrastés*. Rubrique de la Revue STICEF. STICEF, 26 (1).

Collet, L., Hulin, T., & Rémond, É. (2021). Les enjeux de la formation de formateurs aux humanités numériques. In F. Paquienéguy & N. Pélissier, *Questionner les humanités numériques : Positions et propositions des SIC*. CPDirSIC.

Fiormonte, D. (2012). *Towards a cultural critique of the digital humanities*. Historical Social Research/Historische Sozialforschung, 59-76.

Hulin, T., & Petit, L. (2017). Pour une généalogie de la technique. *Revue française des sciences de l'information et de la communication*, 10. <https://doi.org/10.4000/rfsic.2649>.

Laferrière, T. (2020). La recherche en partenariat pour l'enseignement d'hier à demain. *Revue hybride de l'éducation*, 4(1), 94-115. <https://doi.org/10.1522/rhe.v4i1.974>

Lardellier, P., & Hulin, T. (Éds.). (2020). *Les nouvelles médiations pédagogiques. Design et innovation pédagogiques, expériences d'enseignement technologique*. L'Harmattan.

Les éditions critiques numériques : langues, textes et fragments

Alessi Robert¹, Marcello Vitali-Rosati²

1 : UMR 8167 Orient et Méditerranée, CNRS, Paris

2 : Université de Montréal

Organisation

Cette demi-journée se déroulera en deux temps. Chaque participant sera d'abord invité à présenter sur le sujet des remarques sous la forme d'une intervention courte de dix à quinze minutes, éventuellement soutenue par des diapositives. Un débat animé par R. Alessi et M. Vitali-Rosati sera ensuite proposé. Une fois le colloque terminé, les participants seront invités à envoyer une version écrite pour publication dans laquelle ils pourront développer leur pensée initiale tout en tenant compte des éléments que le débat aura pu ajouter.

Sujet proposé

Les grandes collections d'éditions critiques imprimées déterminent depuis le XIX^e siècle les standards de qualité dans la suite desquels il semble aujourd'hui naturel que les éditions numériques s'inscrivent. Celles-ci offrent cependant deux avantages majeurs : celui de fournir des textes à partir desquels il est possible réaliser des recherches, des corrélations ou encore des extractions de données et celui d'offrir un accès libre et à la fois respectueux de la propriété intellectuelle des éditeurs. Toutefois, la rigueur scientifique et la présence d'un appareil critique ne suffisent pas : pour prendre ici un exemple, l'encodage strict – et non fac-similaire – des données ne permet pas de produire naturellement des appareils critiques tels que ceux que les grands savants nous ont habitués à lire en latin. D'un autre côté, il ne faut pas oublier que la rigueur de l'encodage produit des éléments certes invisibles mais extrêmement précieux, tels que l'encodage précis des variantes textuelles, la segmentation des textes donnés dans différentes langues, les éléments de contexte des fragments, voire la corrélation des termes techniques d'une langue à l'autre. Le débat pourra porter aussi bien sur la définition de ces standards qu'il n'est pas aisé de réunir aujourd'hui dans les éditions numériques, que sur les méthodes et les langages à mettre en œuvre et à apprendre pour les atteindre. On tentera également de mesurer quelle est l'influence des méthodes d'encodage des données sur la présentation au lecteur des différents appareils de notes critiques.

Intervenants

Aurélien Berra (Paris), Bruno Bureau (Lyon), Emmanuel Château-Dutier (Montréal), Estelle Debouy (Poitiers), Michael Sinatra (Montréal).

Littérature et numérique : les sessions d'atelier du groupe CCLA

Antoine Fauchié¹, Margot Mellet¹, Marcello Vitali-Rosati^{1,2}

1 : Chaire de recherche du Canada sur les écritures numériques

2 : Université de Montréal

Durant l'année 2020, l'Association canadienne de littérature comparée (Canadian Comparative Literature Association) a proposé une série de sessions d'atelier pour échanger autour de la rencontre de la littérature et du numérique, dans le cadre de ses réflexions autour des matérialités comparatives (médias, littérature et théorie).

La formation de ce groupe de chercheur·e·s et d'étudiant·e·s en littératures survient au moment où les questions techniques d'écriture sont (encore) considérées comme devant rester en dehors du champ des littératures. Il est primordial, aujourd'hui peut-être encore plus qu'hier, d'interroger les formats et les outils d'écriture, collectivement. Les objectifs de ce groupe sont la mise en commun de recherches dans les domaines des littératures et des médias, l'organisation de sessions de travail et d'ateliers autour de ces domaines, ou encore la diffusion de manifestes ou de résultats de recherche. Ce groupe de recherche est initié et coordonné par Monique Tschofen, professeure à l'université de Ryerson (Canada) et Marcello Vitali-Rosati, professeur au Département des littératures de langue française de l'université de Montréal (Canada). Dans un désir d'inclusion les échanges de ce groupe se font en français et en anglais.

Les premières sessions ont été successivement consacrées aux enjeux de l'écriture (écriture et matérialité de l'environnement numérique), à l'écriture sémantique (avec le langage de balisage léger Markdown) et au système de gestion de versions de textes (avec Git, un logiciel habituellement utilisé en informatique). Les présentations et les enregistrements sont disponibles sur le site web du groupe : <https://ccla.digitaltextualities.ca>.

Cette séance du groupe CCLA, ouverte aux participant·e·s habituel·les mais aussi à toute personne intéressée, aura un triple objectif :

- rappeler les origines de ce groupe et ses objectifs ;
- réaliser une synthèse des premières sessions passées, grâce à un travail d'analyse réalisé en amont, et débattre sur la base des premiers résultats de cette synthèse ;
- planifier les prochaines sessions.

Ouvert à tous et à toutes, cette séance d'un groupe de travail sera aussi l'occasion de faire le lien avec la communauté francophone des humanités numériques. Durée : trois heures.

Personnes en charge de l'organisation :

- Antoine Fauchié, doctorant au Département des littératures de langue française de l'université de Montréal et responsable des projets de la Chaire de recherche du Canada sur les écritures numériques
- Margot Mellet, doctorante au Département des littératures de langue française de l'université de Montréal et coordinatrice scientifique de la Chaire de recherche du Canada sur les écritures numériques
- Marcello Vitali-Rosati, professeur au Département des littératures de langue française de l'université de Montréal et titulaire de la Chaire de recherche du Canada sur les écritures numériques

Réflexions sur l'évolution des formes et des structures argumentatives

Josquin Debaz¹, Orélie Desfriches Doria², Aymeric Luneau³

1 : Groupe de sociologie pragmatique et réflexive (GSPR), École des hautes études en sciences sociales (EHESS)

2 : EA 349 Laboratoire Paragraphe, université Paris 8 Vincennes – Saint-Denis

3 : UMR Laboratoire interdisciplinaire Sciences, innovations, sociétés, CNRS, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement, université Gustave Eiffel

L'étude de l'argumentation, transverse à de nombreuses recherches et disciplines, est au centre de l'analyse des corpus avec le logiciel Prospéro (Chateauraynaud, 2003). Le logiciel permet en effet d'identifier les formes et stratégies argumentatives présents dans les textes afin de suivre les processus agonistiques (Chateauraynaud et Debaz, 2013 ; Chateauraynaud et Doury, 2010). On utilise pour cela des « catégories » et des formules correspondant à des formes argumentatives typiques. Plusieurs questions se posent alors quant à la construction de ces catégories et formules.

La première concerne l'identification de formes argumentatives qui, parce qu'elles sont nouvelles ou propres à un dossier, ne figurent pas dans les dictionnaires de catégories.

Dans le même ordre d'idée, si l'on considère que la qualité « argumentative » d'un énoncé est relative au contexte, comment rendre compte des transformations subies par un type d'arguments au cours des différentes périodes d'un même corpus ou entre différents dossiers ?

Enfin, les catégories disponibles sous Prospéro ayant été construites à partir de corpus « contemporains », se pose la question de l'évolution historique des formes argumentatives. Argumente-t-on au XXI^e siècle comme on le faisait au XVI^e ou au XVIII^e ? Au-delà des catégories de Prospéro, jusqu'à quel point est-il pertinent de réduire ces variations historiques à quelques « types d'arguments » (Plantin, 1996 ; Perelman et Olbrechts-Tyteca 2008 [1958]).

Objet du groupe de travail

C'est pourquoi nous proposons d'organiser un « groupe de travail » de 4 heures pour aborder ces différentes questions. Si le point de départ de notre réflexion est le logiciel Prospéro et que l'atelier sera l'occasion de présenter la façon dont on utilise les « formules », notre objectif n'est pas de faire une formation sur le logiciel. Nous souhaitons confronter notre approche à d'autres manières – théoriques ou méthodologiques – d'appréhender les structures argumentatives dans les textes. L'atelier est également ouvert aux personnes travaillant ou envisageant de travailler sur l'argumentation à partir de corpus « non contemporains ». L'atelier sera de cette façon l'occasion de poser les premiers jalons d'une approche diachronique des formes et structures argumentatives et d'évaluer la pertinence d'un tel travail.

Public visé

Le public visé est composé de chercheurs en SHS ayant travaillé sur des corpus historiques ou contemporains, de spécialistes ou de personnes simplement intéressées par l'étude des formes argumentatives dans leur corpus.

Bibliographie

Chateauraynaud, Francis. 2003. *Prospéro, une technologie littéraire pour les sciences humaines*. Paris : CNRS.

Chateauraynaud, Francis et Josquin Debaz. 2013. « Scénariser les possibles énergétiques. Les gaz de schiste dans la matrice des futurs ». *Mouvements* 75 (3) : 53-69.

Chateauraynaud, Francis et Marianne Doury. 2010. « “Désormais...” Essai sur les fonctions argumentatives d’un marqueur de rupture temporelle ». *Argumentation et Analyse du Discours* 4 (avril)

Desfriches-Doria, Orélie, Josquin Debaz et Waldir Lisboa Rocha Filho. 2020. « Interoperability and Discursive Process about Categories ». Communication présentée à *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*, New York, 15-17 octobre.

Ducrot, Oswald. 1996. « La Pragmatique et l’étude sémantique de la langue ». Dans *Une école pour les sciences sociales*. Édité par Jacques Revel et Nathant Wachtel, 339-351. Paris : Cerf – EHESS

Doury, Marianne. 2004. « La classification des arguments dans les discours ordinaires ». *Langages* 154 (2) : 59-73.

Perelman, Chaïm et Lucie Olbrechts-Tyteca. 2008 [1958]. *Traité de l’argumentation. La Nouvelle Rhétorique*. 6th ed. Bruxelles : Éditions de l’université de Bruxelles.

Plantin, Christian. 1996. *L’argumentation*. Paris : Le Seuil.

Ateliers de formation

Atelier de formation TraduXio : une plateforme de traduction collaborative en ligne

Pascale Chazaly Elbaz¹, Philippe Lacour², Diana Lemay³

1 : Institut de management et de communication interculturels, Institut français de recherche sur l'Asie de l'Est (IFRAE), CNRS/INALCO

3 : Université de Brasília

4 : Plidam, INALCO

L'année 2020 a coïncidé avec la propagation d'une épidémie de Coronavirus, contre laquelle de nombreux gouvernements ont choisi de lutter en procédant à plusieurs reprises à un confinement de la population. Cet enfermement inédit a provoqué une adaptation rapide du monde enseignant au travail à distance, souvent de façon improvisée. Beaucoup de professeurs ont découvert à cette occasion le télétravail, avec ses normes spécifiques, synonymes de contraintes mais aussi de situations inédites, suscitant des réponses imaginatives et originales. Citons notamment les plateformes de visioconférences, adaptation du travail en cours et à la maison, modification du travail en groupe pour les élèves, nouveaux modes de lecture et d'annotation collaboratives, etc. Si les réflexions théoriques sur ces pratiques numériques et/ou collaboratives ne sont pas nouvelles (Lameul, Loisy, Charlier 2014 ; Elbaz 2016 ; Petit, Bouchardon 2017), elles ont pris soudainement une place décisive dans les pratiques pédagogiques du supérieur.

L'enseignement de la traduction ne fait pas exception. Il a dû se réinventer de façon très rapide, testant certaines solutions lors du premier confinement, les ajustant lors du second. Parmi les outils numériques, l'environnement TraduXio (<https://traduxio.org>) a pu être utilisé efficacement, parce qu'il permet non seulement de traduire de façon collaborative mais aussi de visualiser et de comparer différentes traductions concurrentes du même texte (y compris en cas de traduction multilingue). Ainsi, Pascale Elbaz a-t-elle mis en place un atelier de traduction collaborative sur des textes de presse (Elbaz 2020) puis sur une nouvelle littéraire avec des classes mixtes d'étudiants francophones et sinophones en troisième année de licence. Les bénéfices de cet apprentissage, qui s'inscrit dans la lignée d'expériences menées à différentes échelles (Cordingley, Frigau Manning 2017 ; Henkel, Lacour 2020) peuvent être résumés en trois points : sensibiliser les apprenants à l'importance de l'interprétation individuelle, développer le souci de cohérence d'un.e étudiant.e traduisant à l'autre (terminologie, temps, ton), souligner l'importance de la discussion en groupe-classe. Outre le fait que TraduXio permet à une équipe d'étudiants en France et en Chine de travailler sur un même support, et même si une préparation des textes et une distribution des tâches en amont reste nécessaire, certaines fonctionnalités de la plateforme ont facilité l'expérience de la traduction collaborative, notamment la segmentation manuelle des textes, un glossaire intégré et la recherche de concordances. Depuis septembre 2020 TraduXio est aussi utilisé dans le cadre des cours de traduction à l'Inalco, en troisième année de licence de slovaque.

Le présent atelier de formation a pour but de présenter la plateforme aux collègues intéressé.es et d'inviter les participant.es à la tester en direct. Les trois intervenant.es feront part de leurs découvertes, déceptions, innovations et suggestions touchant cet outil. Il et elles ont fait l'expérience de l'environnement en situation d'enseignement, avec différents types d'élèves, et auront à cœur de partager leurs découvertes, tout particulièrement concernant les aspects collaboratifs (gestion des équipes, répartition des rôles, résolution des différends, etc) mais également concernant les adaptations souhaitées de l'outil numérique. Les participants n'ayant pas fait l'expérience de la plateforme pourront bénéficier de ce retour d'expérience et poser les questions dans le but d'adopter cet outil.

Références bibliographiques

- Bouchardon, Serge, Petit, Victor. 2017. L'écriture numérique ou l'écriture selon les machines. Enjeux philosophiques et pédagogiques. Dans *Communication & langages* 2017/1 (n° 191), pages 129 à 148.
- Cordingley, Anthony, Frigau Manning, Céline. 2017. *Collaborative Translation : From the Renaissance to the Digital Age*. London : Bloomsbury. Part III Environments of collaboration.
- Elbaz, Pascale. 2016. Enseigner la traduction chinois/français à des locuteurs des deux langues. *Les Langues Modernes (Approches pratiques de la traduction)*, n° 2.
- Elbaz, Pascale. 2020. « Interactivité, co-construction, quand l'outil numérique dé-hiérarchise le rapport entre enseignant.es et étudiant.es en langue ». Dans *Distances apprivoisées. L'enseignement confiné des langues étrangères*, édité par Diana Jamborova Lemay et Louise Ouvrard, Paris, Éditions des Archives contemporaines, (à paraître).
- Henkel, D., Lacour, Philippe. 2020. « TraduXio, Collaboration Strategies in Multilingual Online Translation ». Dans *When Translation Goes Digital : Case Studies and Critical Reflections*, Desjardin R., Larssonneur C. and Lacour P. London : Palgrave.
- Lameul, Geneviève, Loisy, Catherine, Charlier, Bernadette. 2014. *La pédagogie universitaire à l'heure du numérique. Questionnement et éclairage de la recherche*. De Boeck Supérieur.
- Lemay, Diana. 2016. « Apprendre ou désapprendre à traduire ? Le slovaque en milieu universitaire ». Dans *Traduction et apprentissage des langues : entre médiation et remédiation*, dir. Héba Medhat-Lecocq, Delombera Negga et Thomas Szende, Paris, Éditions des Archives contemporaines.

Écriture et édition scientifiques avec l'éditeur de texte Stylo

Margot Mellet¹, Antoine Fauchié¹

1 : Chaire de recherche du Canada sur les écritures numériques

Stylo est conçu pour transformer le flux de travail numérique des revues scientifiques dans le domaine des Sciences humaines et sociales, c'est un éditeur de texte sémantique WYSIWYM (*What You See Is What You Mean*, Ce que vous voyez est ce que vous signifiez) pour l'écriture et l'édition scientifique. Il faut considérer Stylo comme un outil d'écriture sémantique qui vise à améliorer la chaîne de publication académique. La structuration du texte, via l'utilisation de balises sémantiques, est placée au tout début du processus de production de la connaissance, pour redonner la maîtrise du texte aux auteur·e·s.

Basé sur des outils et formats d'édition modulaires, *low-tech* et standard, tels que Markdown, YAML, BibTeX, Pandoc, Hypothes.is et Latex, Stylo intègre dans une seule interface les meilleures pratiques en termes de rédaction et de publication sur le web.

En mettant en œuvre les formats et les technologies de conversion déjà utilisés par la communauté, il permet une libre circulation des documents qui ne sont pas verrouillés dans un format particulier. Stylo a été pensé pour offrir un flux de travail complet et continu de la rédaction à la publication, en évitant la perte de données et de sens entre les différentes étapes d'édition.

Stylo regroupe des fonctionnalités telles que le partage, la gestion de versions, le suivi des modifications, la gestion des références bibliographiques, l'annotation pour les révisions, l'exportation multi-formats, les métadonnées alignées avec les autorités (LOC, Wikidata, ORCID, ...) et le balisage sémantique en ligne.

Dans cet atelier, nous présenterons comment Stylo est utilisé quotidiennement comme outil d'écriture et d'édition pour une revue scientifique (*Sens public*), et comment il peut être utilisé par les chercheurs et les étudiants, individuellement ou collectivement.

Les participants auront la possibilité de tester et d'éditer leur propre texte dans Stylo. Stylo est actuellement utilisé par une petite communauté grandissante, et il est disponible depuis 2020 parmi les services proposés par la TGIR Huma-Num (<https://stylo.huma-num.fr>).

Au cours de l'atelier, nous souhaitons aussi recueillir des commentaires sur l'outil : son interface, son ergonomie et sa convivialité, mais aussi sur les pratiques d'écriture scientifique que Stylo pourrait intégrer.

Durée : deux heures.

Lemmatiser des textes et corriger l'annotation grâce à l'apprentissage profond avec Pyrrha

Thibault Clérice^{1,2}, Ariane Pinche³, Jean-Baptiste Camps¹, Matthias Gille Levenson³, Simon Gabay⁴,
Lucence Ing¹

1 : EA 3624 Centre Jean-Mabillon, université Paris sciences et lettres, École nationale des chartes

2 : UMR 5189 Histoire et sources des mondes antiques, CNRS, université Jean-Monnet (Saint-Étienne), université Jean Moulin – Lyon 3, université de Lyon, université Lumière – Lyon 2, École normale supérieure de Lyon

3 : UMR 5648 Histoire, archéologie et littératures des mondes chrétiens et musulmans médiévaux, Lyon

4 : Université de Genève

L'analyse quantitative ou distante de données textuelles est centrale dans le champ des sciences humaines computationnelles, tout comme l'emploi de méthodes provenant du traitement automatique des langues. La lemmatisation et l'annotation morpho-syntaxique sont notamment utiles à de nombreux traitements très courants (modélisation de sujet, lexicométrie, etc.). Si les langues contemporaines sont la plupart du temps bien équipées, les langues anciennes ou médiévales, riches en variation (morphologique, graphique, etc.), et les langues peu dotées posent des difficultés réelles. La dernière génération de lemmatiseurs permet de remédier efficacement à ces problèmes, mais le passage d'outils à base de règles à des approches par apprentissage profond [1, 2] implique l'utilisation de grandes quantités de données d'entraînement, nécessitant un important travail de reprise et correction de données.

Toutefois, la plupart des interfaces pour faire de l'annotation morpho-syntaxique et de la lemmatisation sont construites pour une saisie de première main plutôt que pour de la correction [3, 4]. Pyrrha [5] a été développée en vue de répondre aux pratiques existantes d'utilisateur-trice-s (correction dans des tableurs, par exemple), tout en s'inspirant de fonctionnalités proposées par des applications similaires de corrections d'OCR (corrections par lots [6], validation par dictionnaire, etc.).

Cet atelier présentera aussi l'utilisation de Pie [7]. Pie est un étiqueteur de séquence pour les langues riches en variations, hautement configurable et fondé sur des réseaux de neurones. Pyrrha a été construit pour compléter l'étiqueteur afin de post-corriger le texte balisé. Cela rend possible ensuite de le réutiliser afin de faire croître les données d'entraînements, dans un cercle vertueux, ou bien simplement de les analyser.

Dans cette perspective, nous proposons de réaliser un atelier de 4h, abordant les points suivants :

- Utilisation du service de lemmatisation et d'annotation (langues actuellement traitées : différents états du français, du IX^e au XVIII^e siècles ; grec ancien et latin ; moyen néerlandais) ;
- Création de corpus, imports et exports en différents formats (tsv, XML-TEI) ;
- Correction au fil du texte ou par lots ;
- Moteur de recherche ;
- Bonnes pratiques et travail collaboratif (gestion de référentiels communs, suivi des interventions des différents correcteurs et correctrices, ...) ;
- Partage et archivage de données annotées.
- Organisation de groupes de travail pour le développement des modèles

Il serait utile de prendre contact avec les instructeurs avant l'atelier afin de s'assurer que les langues préférées des stagiaires sont incluses dans la configuration du cours. La liste actuelle comprend le français classique, moderne, l'ancien français, le latin classique et médiéval, et le grec ancien. L'ajout de langues supplémentaires est envisageable, si prévu suffisamment en avance et si les données existent.

Bibliographie

- [1] E. Manjavacas, A. Kádár et M. Kestemont, « Improving lemmatization of non-standard languages with joint learning », arXiv preprint arxiv:1903.06939 (2019).
- [2] H. Schmid, « Deep learning-based morphological taggers and lemmatizers for annotating historical texts », dans *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*, 2019, p. 133-137.
- [3] S. M. Yimam, I. Gurevych, R. E. de Castilho, and C. Biemann, « WebAnno : A flexible, web-based and visually supported system for distributed annotations, » in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics : System Demonstrations*, 2013, p. 1-6.
- [4] B. Almas and G. Höflechner, Arethusa : Annotation Environment. alpheios-project, 2018.
- [5] Thibault Clérice, Julien Pilla, and Jean-Baptiste-Camps, hipster-philology/pyrrha : 1.0.1. *Zenodo*, 2018.
- [6] T. Erjavec, « Architecture for Editing Complex Digital Documents, » in *Proceedings of the Conference on Digital Information and Heritage*. Zagreb, 2007, p. 105-114.
- [7] Enrique Manjavacas, Mike Kestemont, and Thibault Clérice, emanjavacas/pie v0.1.0. *Zenodo*, 2018.

Atelier partenaire

L'open content au musée : diffusion, contribution et innovation

Fabienne Martin Adam¹, Manuel Moreau¹, Karl Pineau²

1 : Musée de Bretagne

2 : Université de technologie de Compiègne

Installé à Rennes au sein de l'équipement culturel des Champs Libres, le musée de Bretagne a lancé en 2017 Des collections en partage, un portail de mise en ligne des collections avec une démarche active d'open content, qui se distingue dans le paysage muséal français par son ampleur et son ambition. Innovant dans sa philosophie, ce projet de diffusion en ligne des collections s'est inscrit dans le projet scientifique et culturel du musée, ainsi que dans les grandes orientations de la collectivité (Rennes Métropole) portant une politique culturelle, attentive à la diffusion et la transmission des savoirs, et au-delà, des ambitions d'innovation dans la relation aux usagers et d'ouverture militante des données.

L'atelier se déroulera en deux temps associant intervention puis échange avec les participants universitaires sur leurs besoins et retours d'usage.

Une première partie sera dédiée à la présentation des enjeux d'un portail des collections via l'historique de sa mise en œuvre au musée de Bretagne et ses perspectives d'évolution.

La seconde partie sera dédiée aux opportunités offertes par le développement de l'open content, notamment dans le cadre universitaire, de la recherche et du développement en évoquant notamment des projets en cours comme le captchan culturel. Il s'agira ici de montrer comment l'open content peut rencontrer à la fois des intérêts en termes de médiation culturelle (à travers la mise en avant de pratiques d'indexation professionnelles ou collaboratives), mais également de recherche académique en aidant à améliorer la qualité des données produites. Nous détaillerons donc le projet CaptchAN / OC-API, mené entre 2019 et 2021 et issu du hackathon des Archives nationales et dont nous présenterons les premiers retours de l'étude d'usage.

Intervenants :

Fabienne MARTIN ADAM - Chargé de l'inventaire et de la documentation au musée de Bretagne

Manuel MOREAU - Chargé prospective et innovation au musée de Bretagne

Karl PINEAU - Doctorant en sciences de l'information et de la communication à l'université de technologie de Compiègne et ingénieur de recherche chez l'éditeur Decalog

Programme du colloque

Lundi 10 mai 2021

HORAIRE	ÉVÉNEMENT
14:00 - 14:20	Ouverture du colloque (Salle bleue) - Martin Grandjean, Nicolas Thély, Marc Bergère, Fatiha Idmhand, Aurélien Berra
14:20 - 15:00	Conférence plénière (Salle bleue)
14:20 - 15:00	› L'intelligence artificielle pour décrire, indexer et explorer des collections multimédias : comment faire sans Google ? - Guillaume Gravier
15:15 - 16:45	Méthodes numériques (Salle bleue) - Fatiha Idmhand
15:15 - 15:45	› Contribution à l'histoire des méthodes de l'enregistrement archéologique de terrain. Etat d'avancement de travaux en cours sur l'étude d'archives de fouille par leur transcription numérique - Christophe Tufféry, Claudia Marinica, Maximilien Rioult, Yulin Xie
15:45 - 16:15	› Des prononciations et des cartes : une fenêtre ouverte sur la longue durée. Un exemple à partir de la langue bretonne - Tanguy Sollic
16:15 - 16:45	› Traduire sans trahir, la vocation des logiciels ou celle des chercheurs ? Programmes de recherche et programme informatique au carrefour des cultures et des méthodologies. - Bastien Fond, Josquin Debaz
15:15 - 16:45	Méthodes numériques (Salle rouge) - Nicolas Thély
15:15 - 15:45	› Automatisation des réseaux de personnages : proposition de cadre formel pour démystifier les algorithmes - Coline Métrailler
15:45 - 16:15	› Un outil de dépouillement de sources archivistiques basé sur des technologies XML - Emmanuel Château-Dutier, Josselin Morvan
15:15 - 16:45	Données et patrimoines (Salle jaune) - Aurélien Berra
15:15 - 15:45	› Image et texte : les livres d'heures manuscrits vus par l'intelligence artificielle - Dominique Stutzmann, Louis Chevalier
15:45 - 16:15	› Interactions avec la sirène de Cagniard-Latour en réalité virtuelle - Ronan Gagne, Théophile Nicolas, Dominique Bernard, Julie Priser, Marion Lemaire, Meven Leblanc, Jim Pavan, Tom Roy, Raphael Dupont, Julie Ayoubi, Marie-Marie Le Cornec, Valérie Gouranton
15:15 - 16:45	Méthodes numériques (Salle verte) - Martin Grandjean
15:15 - 15:45	› Seconde traversée (Déteros plous) : nouvelles et développements du projet d'édition collaborative numérique de l'Anthologie Palatine - Margot Mellet, Marcello Vitali-Rosati
15:45 - 16:15	› L'utilisation de l'outil Stylo dans le cadre de la revue Sens public - Antoine Fauchié, Margot Mellet, Nicolas Sauret
17:00 - 18:30	Données et patrimoines (Salle jaune) - Florence Thiault
17:10 - 17:15	› Éditer et transmettre le Parallèle des Anciens et des Modernes de Charles Perrault - Delphine Reguig, Emmanuelle Perrin
17:20 - 17:25	› Géo-visualisation des contenus de la Perséide Athar : le cas du Caire moderne. Un exemple d'intégration de données documentaires, sémantiques et spatiales au service des historiens - Bulle Tuil Leonetti, Hélène Bégnis, Julie Erismann
17:30 - 17:35	› Le projet « Démêler le cordel » : une bibliothèque numérique pour l'étude de la littérature éphémère espagnole du XIXe siècle - Constance Carta, Elina Leblanc

HORAIRE	ÉVÉNEMENT
17:40 - 17:45	› SITT Streaming Interactivité Téléperformance Transmédia Un projet de sauvegarde du patrimoine chorégraphique par le développement d'outils technologiques spécifiques - Armando Menicacci, Nicolas Berzi, Marie Lavorel
17:00 - 18:30	Méthodes numériques (Salle bleue) - Emmanuel Château-Dutier
17:00 - 17:30	› Comment exploiter un corpus à l'aide des technologies du Web sémantique ? Le cas de la correspondance d'Henri Poincaré. - Nicolas Lasolle, Pierre Willaime
17:30 - 18:00	› Dialogue interdisciplinaire en humanités numériques : le cas de l'évolution des formes discursives en critique d'art - Orélie Desfriches Doria, Josquin Debaz, Gérald Kembellec, Marie Gispert
18:00 - 18:30	› Le style polémique dans les correspondances de Voltaire - Motasem Alrahabi, Camille Koskas, Glenn Roe
17:00 - 19:00	Atelier de formation (Salle verte)
17:00 - 19:00	› Écriture et édition scientifiques avec l'éditeur de texte Stylo - Margot Mellet, Antoine Fauchié

Mardi 11 mai 2021

HORAIRE	ÉVÉNEMENT
09:00 - 13:00	Atelier de formation (Salle bleue)
09:00 - 13:00	› Lemmatiser des textes et corriger l'annotation grâce à l'apprentissage profond avec Pyrrha - Thibault Clérice, Jean-Baptiste Camps, Simon Gabay, Lucence Ing
09:00 - 13:00	Groupe de travail (Salle jaune)
09:00 - 13:00	› Réflexions sur l'évolution des formes et des structures argumentatives - Josquin Debaz, Orélie Desfriches Doria, Aymeric Luneau
09:30 - 11:30	Atelier (Salle rouge) - Morgane Mignon
09:30 - 11:30	› L'open content au musée : diffusion, contribution et innovation - Fabienne Martin Adam, Manuel Moreau, Karl Pineau
14:00 - 15:30	Méthodes numériques (Salle bleue) - Martin Grandjean
14:00 - 14:30	› Katabase: À la recherche des manuscrits vendus - Simon Gabay, Alexandre Bartz, Matthias Gille Levenson
14:30 - 15:00	› Du « Chantre tout divin » au « Ménétrier déconfit » : exploiter les sources anciennes d'un mythe pour mieux détecter les correspondances entre ses réécritures - Karolina Suchecka, Nathalie Gasiglia
15:00 - 15:30	› POPP. Projet de reconnaissance optique des caractères des recensements de la population parisienne - Sandra Brée, Thierry Paquet, Nicolas Kempf
14:00 - 15:30	Données et patrimoines (Salle jaune) - Aurélie Hess
14:00 - 14:30	› FAIRiser des données : état des lieux, barrières et choix .Une réflexion à partir des données des corpus d'auteurs ? - Ioana Galleron, Fatiha Idmhand
14:30 - 15:00	› Les matériaux de recherche ethnographiques à l'épreuve de l'open data - Florence Revelin
15:00 - 15:30	› Préserver et diffuser le patrimoine textile européen grâce aux technologies du Web sémantique. La réutilisation des données patrimoniales par le projet SILKNOW - Marie Puren, Pierre Vernus

HORAIRE	ÉVÉNEMENT
14:00 - 16:45	Groupe de travail (Salle verte)
14:00 - 16:45	› Littérature et numérique : les sessions d'atelier du groupe CCLA - Antoine Fauchié, Margot Mellet, Marcello Vitali-Rosati
15:45 - 16:45	Méthodes numériques (Salle bleue) - Emmanuel Château-Dutier
15:55 - 16:00	› Essai de cartographie de médias imaginaires du XIXe siècle - Philippe Ethuin
16:05 - 16:10	› Langues celtiques, typographie et hybridation - Clémence Jaron
16:15 - 16:20	› Paul Otlet en son réseau. Présentation de l'Otletosphère - Olivier Le Deuff, Arthur Perret, Clément Borel, Guillaume Brioude
16:25 - 16:30	› Le projet ITHAC entre méthodes numériques, recherche multidisciplinaire et propositions pédagogiques. - Diandra Cristache, Anne Garcia-Fernandez, Elisabeth Greslou, Arnaud Bey, Pascale Paré-Rey, Malika Bastin-Hammou
16:35 - 16:40	› Adressbuch 2021 : La troisième vie de l'Annuaire des Allemands à Paris de 1854 - Mareike Koenig, Gérald Kembellec
17:00 - 18:30	Assemblée générale d'Humanistica (Salle Humanistica) - Séance plénière

Mercredi 12 mai 2021

HORAIRE	ÉVÉNEMENT
09:30 - 11:00	Méthodes numériques (Salle rouge) - Aurélien Berra
09:30 - 10:00	› Contextualisation de la critique francophone d'art contemporain - William Diakité
10:00 - 10:30	› Utilisation d'approches automatiques pour la reconnaissance des expériences de lecture. - François Vignale, Guillaume Le Noé Bienvenu, Guillaume Gravier, Pascale Sébillot
10:30 - 11:00	› Saisir le buzz politique sur Internet : le cas Juan Branco autour de son livre "Crépuscule". - Amar Lakel, Franck Cormerais
09:30 - 12:30	Atelier de formation (Salle jaune)
09:30 - 12:30	› Atelier de formation TraduXio (plateforme de traduction collaborative en ligne) - Pascale Chazaly Elbaz, Philippe Lacour, Diana Lemay
11:15 - 12:30	Compétences numériques et formations (Salle bleue) - Florence Thiault
11:15 - 11:45	› « Je pense que ça traite d'expérience de lecture, à voir ... » : retour sur une expérience d'annotation collaborative. - François Vignale, Guillaume Le Noé Bienvenu, Guillaume Gravier, Pascale Sébillot
11:45 - 12:15	› L'observation (numériquement équipée) des problèmes publics - Waldir Lisboa Rocha
11:15 - 12:30	Méthodes numériques (Salle rouge) - Aurélie Hess
11:15 - 11:45	› IA et approches participatives dans les Humanités Numériques : de la conjonction à la coopération - Olivier Aubert, Guillaume Raschia, Benjamin Hervy
11:45 - 12:15	› Le rôle des modalités linguistiques dans l'analyse automatique du style critique - Marguerite Bordry

HORAIRE	ÉVÉNEMENT
14:00 - 15:30	Données et patrimoines (Salle jaune) - Nicolas Thély
14:00 - 14:30	› Le recours aux humanités numériques pour (re)constituer le dossier génétique d'une création collective - Johanne Melancon
14:30 - 15:00	› Les humanités numériques pour valoriser le patrimoine minier - Eric Kergosien, Amélie Daloz, Bernard Jacquemin, Stéphane Chaudiron
15:00 - 15:30	› Moyen Âge & Performances (MAP): (modéliser un) répertoire de performances médiévales - Simon Gabay
14:00 - 17:30	Groupe de travail (Salle verte)
14:00 - 17:30	› Les éditions critiques numériques: langues, textes et fragments - Alessi Robert, Marcello Vitali-Rosati
14:00 - 16:00	Groupe de travail (Salle bleue)
14:00 - 16:00	› groupe de travail « Humanités numériques, éducation et formation » - Thibaud Hulin, Aude Seurrat
16:00 - 17:30	Données et patrimoines (Salle jaune) - Martin Grandjean
16:00 - 16:30	› Exploration du théâtre alsacien à travers ses listes de personnages pendant la période 1870-1940 - Pablo Ruiz Fabo, Carole Werner
16:30 - 17:00	› Modéliser le feuilleton. Créer et analyser une nouvelle archive patrimoniale avec les méthodes computationnelles. - Pierre-Carl Langlais
17:00 - 17:30	› Théâtre classique et difficultés numériques : quelques réflexions sur l'analyse dramatique assistée par ordinateur - Ioana Galleron
16:15 - 17:30	Compétences numériques et formations (Salle bleue) - Fatiha Idmhand
16:25 - 16:30	› La Médiation du patrimoine culturel et l'éducation : Le rôle de la plate-forme numérique e-learning dans la conservation et la transmission de patrimoine. - Radouane Mimoun El Khir, Abdelelah Ounana
16:35 - 16:40	› Les bibliothécaires et leur rapport au livre numérique au Québec : vers une vision intégrative du patrimoine écrit - Christelle Pelbois, Marie-Claude Lapointe, Jason Luckerhoff
16:45 - 16:50	› Penser la phase post-pandémique au supérieur : Quels processus d'appropriation des technologies éducatives dans les pratiques enseignantes? - Rachid El Ganbour, Samira Elouelji, Mohammed Boukare, Morad El Ganbour
16:55 - 17:00	› Reconnaissance et certification dynamique dans les dispositifs contributifs du secondaire : opportunité et levier de développement des habiletés numériques des élèves et des enseignants - Franck Bodin, Guillaume Viniacourt
17:45 - 18:15	Conclusion du colloque (Salle bleue) - Nicolas Thély, Martin Grandjean, Aurélien Berra, Fatiha Idmhand, Michael Sinatra

lundi 10 mai 2021 >				
14:00	>14:00 (20min) Ouverture du colloque > Salle bleue			
	>14:20 (40min) Conférence plénière > Salle bleue			
15:00	>15:15 (1h30) Méthodes numériques <i>Fatiha Idmhand</i>	>15:15 (1h30) Méthodes numériques <i>Nicolas Thély</i>	>15:15 (1h30) Données et patrimoines <i>Aurélien Berra</i>	>15:15 (1h30) Méthodes numériques <i>Martin Grandjean</i>
	>16:00	>16:00	>16:00	>16:00
	>16:00	>16:00	>16:00	>16:00
	>16:00	>16:00	>16:00	>16:00
17:00	>17:00 (1h30) Données et patrimoines <i>Florence Thiault</i>		>17:00 (1h30) Méthodes numériques <i>Emmanuel Château-Dutier</i>	>17:00 (2h) Atelier de formation
	>18:00		>18:00	>18:00
18:00	>18:00		>18:00	>18:00
	>18:00		>18:00	>18:00
19:00	>19:00			
	>19:00			

< mardi 11 mai 2021 >		
09:00	>9:00 (4h) Atelier de formation	>9:00 (4h) Groupe de travail
10:00		>9:30 (2h) Atelier <i>Morgane Mignon</i>
11:00		> Salle rouge
12:00	> Salle bleue	> Salle jaune
13:00		
14:00	>14:00 (1h30) Méthodes numériques <i>Martin Grandjean</i>	>14:00 (1h30) Données et patrimoines <i>Aurélie Hess</i>
15:00	> Salle bleue	> Salle jaune
16:00	>15:45 (1h) Méthodes numériques <i>Emmanuel Château-Dutier</i>	>14:00 (2h45) Groupe de travail
	> Salle bleue	> Salle verte
17:00	>17:00 (1h30) Assemblée générale d'Humanistica <i>Séance plénière</i>	
18:00	> Salle Humanistica	

« mercredi 12 mai 2021		
09:00		
	>9:30 (1h30) Méthodes numériques <i>Aurélien Berra</i>	>9:30 (3h) Atelier de formation
10:00	> Salle rouge	
11:00	>11:15 (1h15) Compétences numériques et formations <i>Florence Thiault</i>	>11:15 (1h15) Méthodes numériques <i>Aurélie Hess</i>
12:00	> Salle bleue	> Salle rouge
13:00		
14:00	>14:00 (1h30) Données et patrimoines <i>Nicolas Thély</i>	>14:00 (3h30) Groupe de travail
15:00	> Salle jaune	>14:00 (2h) Groupe de travail
16:00	>16:00 (1h30) Données et patrimoines <i>Martin Grandjean</i>	>16:15 (1h15) Compétences numériques et formations <i>Fatiha Idmhand</i>
17:00	> Salle jaune	> Salle bleue
18:00	>17:45 (30min) Conclusion du colloque <i>Nicolas Thély, Martin Grandjean, Aurélien Berra, Fatiha Idmhand, Michael Sinatra</i>	
	> Salle bleue	