

LE PROJET DES REGISTRES JOURNALIERS DE LA COMÉDIE-FRANÇAISE

Les humanités numériques, dialogue entre les mondes de la recherche et de la documentation

SARA HARVEY

Professeur à l'université de Victoria (Canada)

AGATHE SANJUAN

Conservatrice

Bibliothèque-musée de la Comédie-Française

Le projet des registres journaliers de la Comédie-Française¹ est un programme de recherche international bilingue, mais aussi interdisciplinaire et interprofessionnel. Il regroupe des institutions universitaires (Paris Ouest Nanterre– La Défense, Paris-Sorbonne, Harvard University, l'université de Poitiers, le Massachusetts Institute of Technology – à la fois son département d'histoire et un laboratoire d'humanités numériques) et une institution à vocation culturelle et patrimoniale, la bibliothèque-musée de la Comédie-Française. Sur un site internet dédié, le programme permet l'accès inédit aux archives de la Comédie-Française de 1680 à 1793 sous la forme de fac-similés numériques. À cette mise à disposition, s'ajoute la possibilité d'explorer ces archives patrimoniales grâce à plusieurs outils de recherche et de visualisation des résultats que sous-tend une base de données. Les registres considérés offrent une information journalière sur les pièces représentées (traditionnellement deux par soirée), le nombre de places vendues par catégorie de place, la recette de la soirée et quantité d'éléments factuels (présence de personnalités, débuts d'acteur, voyages à la cour, etc.). Les usagers visés se composent de cher-

cheurs (le site comprend non seulement des outils de recherche, mais propose aussi une valorisation de la recherche sous forme de publications en ligne), d'étudiants (un espace pédagogique est dédié au programme et vise à initier les professeurs à l'utilisation du site et de ses différents outils à des fins d'enseignement), de « curieux » et amoureux de la Comédie-Française (des vignettes encyclopédiques sont en cours de création et permettent une introduction virtuelle à l'univers des registres) et de praticiens du théâtre désireux de redécouvrir les textes du répertoire et les pratiques de programmation associées.

Si le public visé est hétérogène, les acteurs du projet forment eux-mêmes une communauté professionnelle multiple. Chercheurs, conservateurs, documentalistes et informaticiens ont construit ensemble un outil où chacun devait pouvoir retrouver ses compétences, sa technicité, ses intérêts de travail et de développement. Nous voudrions montrer ici que ce dialogue interprofessionnel a été fécond et constitue la clé de ce type de projet, mais qu'il suppose aussi de nouveaux modèles organisationnels à la fois dans la gestion de projet et dans le traitement de l'information.

¹ <http://cfregisters.org/fr/>

On ne joua point hier cause du depart de
M. Poisson et Raisin pour Fontainebleau 141:

Aujourd'hui dimanche 14. jour de Septembre 1681:

A *Iphigenie, et Les auberges.*

Theatre Cent billets 300^{tt}
Premieres Loges Quarante huit billets 144^{tt}
Amphiteatre
Secondes Loges Cent vingt trois billets 184^{tt} 10 s
Troisiemes Loges Vingt cinq billets 25^{tt}
Parterre Trois cens Vingt six billets 244^{tt} 10 s
Receu en tout 898^{tt}

Frais ordinaires 70^{tt} 7 s
Pensions & Loyers 30^{tt}
Frais extraordinaires, et chandelle de religieux 2^{tt} 11 s
Paye les Loyers et Pensions d'hier 30^{tt}
Pour les gagistes 7^{tt} 10 s
Pour l'affiche noire 4^{tt}
De Falque 11 s 6 d

PART Trente quatre livres dix sols 750^{tt} 750 s
Retire pour les pensions et Loyers 2^{tt} 13 s
Despence 898^{tt}

Registres journaliers de
la Comédie-Française
du 14 septembre 1681
(à gauche) et du
27 mars 1793.

COMEDIE FRANÇAISE.
RECETTE JOURNALIERE. Année 1792 à 1793.

277^e Représentation N^o 23 du Mardi 27 Mars 1793

Pour la clôture
Morceaux, tragédie, Service de la troisième représentation
Conteur seulement porté.

	Premieres Loges à 6 places, à 6 liv. ---		
	Premieres Loges à 5 places, à 6 liv. --		
	Premieres Loges à 4 places, à 6 liv. --		
	Loges au Rez-de-chaussée, à 5 pl. à 6 l.		
5	1 Loges au Rez-de-chaussée, à 5 pl. à 6 l.	30	
	Loges au Rez-de-chaussée, à 5 pl. à 6 l.		
20	5 Secondes Loges à 4 places, à 3 l. 15 s.	75	
6	7 Secondes Loges à 3 places, à 3 l. 15 s.	22	10
	Troisiemes Loges à 8 places, à 2 l. 10 s.		
	Troisiemes Loges à 6 places, à 2 liv. 10 s.		
	Troisiemes Loges à 4 places, à 2 liv. 10 s.		
2	1 Troisiemes Loges à 3 places, à 2 liv. 10 s.	7	10
	Troisiemes Loges à 4 places, à 2 liv. 10 s.		
	Quatriemes Loges à 4 places, à 2 l. 10 s.		
	Quatriemes Loges à 3 places, à 2 l. 10 s.		
31	31 Places de Premieres à 6 liv. ---	186	
100	104 Places de Galerie Et Secondes à 3 liv. ---	312	
182	182 Places de Parquet à 1 liv. 16 s. ---	297	12
40	64 Places de Troisiemes à 2 liv. ---	128	
80	80 Places de Paradies à 1 liv. 10 s. ---	120	
49			
	Total	1200	12

3265^{tt} 13 s

Accité par moi Contrôleur de la Caisse la Recette de ce jour
vingt deux cent mil sept cent quatre-vingt quatre
montant à la somme de deux cent livres treize sols.

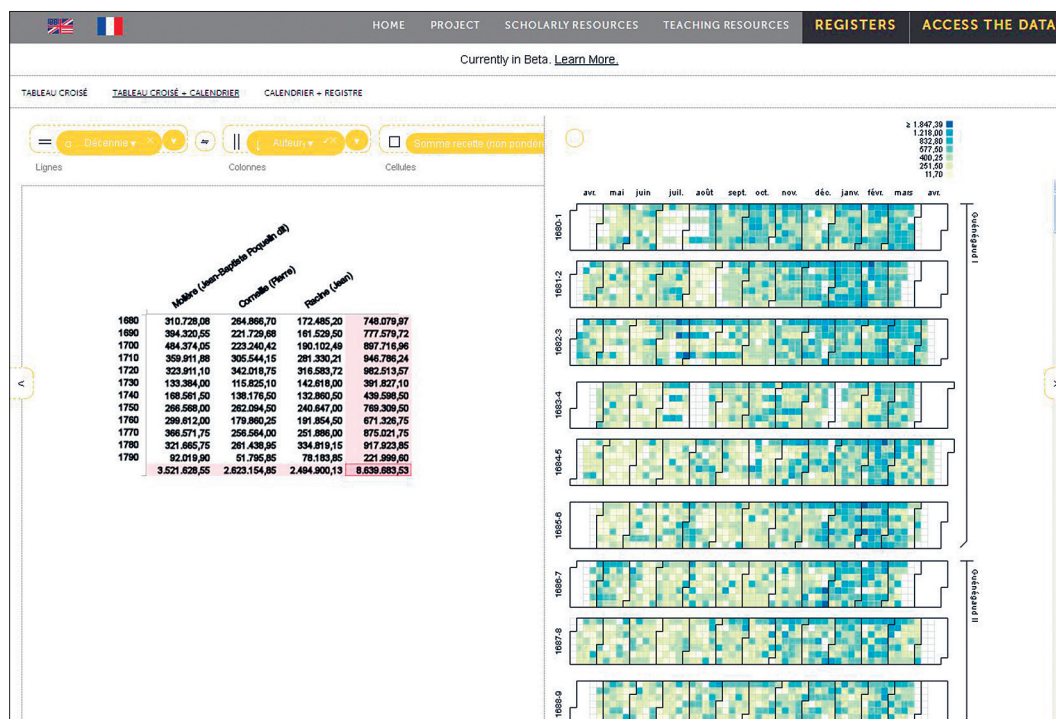
Dumoulin

LES HUMANITÉS NUMÉRIQUES EN CONTEXTE

L'un des défis de l'équipe a été de comprendre le sens et les enjeux du projet et de sa démarche en regard des humanités numériques. Sous ce chapeau que certains nomment un chapiteau², une problématique définitoire se dessine-t-elle et celle-ci permet-elle de mieux cibler l'expérience menée dans notre programme de recherche? En tenant compte à la fois du modèle expérimental qu'a constitué le projet des registres de la Comédie-Française mais aussi de nos lectures des principaux acteurs dans le domaine des humanités numériques en France³, nous choisissons de définir les humanités numériques⁴ en ces termes : il s'agit de pratiques et de dispositifs de recherche qui intègrent les nouvelles technologies au sein de projets issus du domaine des sciences humaines et sociales et de l'univers de la documentation. Liées à un ensemble de méthodes en partie exploratoires, liées également à des outils technologiques de recherche et de visualisation de la recherche, liées enfin à des moyens de communication et

de diffusion virtuels, les humanités numériques forment un environnement susceptible de créer des modalités nouvelles de production et de circulation de la documentation et du savoir. Un autre élément nous paraît essentiel en raison du fait que nous y sommes confrontés au quotidien : les humanités numériques sont non seulement de nature transdisciplinaire, elles sont aussi interprofessionnelles puisqu'elles reposent sur un dialogue entre les chercheurs œuvrant dans le domaine traditionnel des humanités et les spécialistes de l'informatique, mais aussi de l'information et de la documentation. Suscitées par l'émergence de la culture numérique, les humanités numériques supposent, à ce stade de leur développement, d'initier les chercheurs et les étudiants aux méthodes, pratiques et enjeux théoriques et réflexifs, d'une part; de réfléchir concrètement aux modèles organisationnels, d'autre part, c'est-à-dire aux *habitus* des communautés professionnelles concernées, mais aussi au schéma opératoire qu'exige la réalisation des projets, et enfin de s'interroger sur la pérennité de ces projets tant d'un point de vue technique que juridique et financier. Il im-

Exemple de visualisation d'une recherche.



² Voir à ce sujet Melissa Terras, «Peering inside the Big Tent : Digital Humanities and the Crisis of Inclusion», traduit par Pierre Mounier, «Un regard jeté sous le chapiteau : les humanités numériques ou la crise de l'inclusion», dans *Read/Write Book 2*, Open Edition Press, 2012. <http://books.openedition.org/oepp/2226>

³ Nous pensons principalement à : Frédéric Clavert et Serge Noiret, *L'histoire contemporaine à l'ère numérique*, Peter Lang, 2013; Marin Dacos et Pierre Mounier, *Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international*, Institut français, 2014; Milad Doueïhi, *Pour un humanisme numérique*, Seuil, 2011; Olivier Le Deuff (dir.), *Le temps des humanités digitales : La mutation des sciences humaines et sociales*, FYP éditions, 2014; Pierre Mounier (dir.), *Read/Write Book 2. Une introduction aux humanités numériques*, Open Edition Press, 2012; Bernard Stiegler (dir.), *Digital Studies. Organologie des savoirs et technologies de la connaissance*, FYP éditions, 2014.

⁴ Nous conservons l'expression «humanités numériques» plutôt que «digitales», sans souhaiter pour autant faire partie d'un débat, c'est par commodité.

porte donc de questionner le processus d'institutionnalisation de cette transdiscipline qui apparaît aujourd'hui aussi enthousiasmante que nécessaire, ne serait-ce que par sa nature exploratoire et collaborative, mais qu'il faut aborder avec prudence, rigueur et éthique.

FONDS D'ARCHIVES, DOCUMENT, DONNÉE : REDÉFINIR LE NIVEAU DE TRAITEMENT DE L'INFORMATION

Le projet des registres de la Comédie-Française porte sur des documents matériels peu connus des chercheurs en dehors du cercle restreint des historiens de la Comédie-Française. Par un abus de langage courant, on désigne les registres sous l'appellation de « collections » de la Comédie-Française. Pourtant, le rassemblement de ces registres ne procède pas d'une collecte mais bien d'une production organique de type archivistique⁵. De fait, ils constituent bien le « fonds d'archives » reflétant l'activité du théâtre, chacun étant un « document d'archives » à part entière au sein de ce fonds. Le niveau du « document » semble cependant moins opérant en contexte numérique, autant sur le plan du traitement documentaire que sur le plan de la recherche.

Un document renvoie à un ensemble formé par un support et une information. Or, le passage au numérique agit sur ces deux aspects : matériel par sa transformation sur un support numérique, immatériel car l'information va pouvoir être traitée avec des possibilités quasi infinies. Par ailleurs, le passage au numérique modifie la pratique scientifique d'exploration du document. Les nouvelles possibilités tendent à « transformer » le document en permettant de le manipuler (calculs, annotations en ligne). Le niveau d'analyse ne se fait plus au niveau du document, mais au niveau des « données » qu'il contient – le document est segmenté en unités d'information – et que l'on exploite grâce à des tableaux de données, des visualisations, des graphes, etc. Le produit de la recherche, obtenu grâce à l'exploitation numérique des données devient lui-même source – d'où la nécessité d'archiver aussi ces visualisations.

Ces actions sur le document – transformation et analyse au niveau de la donnée – nous renvoient à une autre dimension de la recherche en

contexte numérique : la manipulation de données de masse.

Traditionnellement, le métier de l'historien comme celui de l'archiviste ou du bibliothécaire ont en commun de gérer des masses de documents. Or, cette gestion de la masse change complètement avec le passage au numérique ; on traite non seulement en masse des documents, mais aussi les données qu'ils contiennent. Ce changement d'échelle du traitement du document est tout à fait paradoxal : la mise à disposition numérique devrait faciliter son traitement, mais les possibilités de l'outil numérique et l'accès immédiat et massif à la documentation disponible nous engagent au contraire à entrer dans le document – rendant le référencement plus riche mais plus complexe –, ce que ne pouvaient faire les catalogues, inventaires et répertoires d'autrefois. La perspective est donc vertigineuse pour le chercheur comme pour le professionnel de la documentation : pour l'un, en termes de possibilités de recherche, pour l'autre, en termes de finesse de traitement. On passe donc d'une documentation de masse à une masse de données.

Ce changement fondamental va modifier notre approche de la documentation et de son cadre de définition.

LE CORPUS EN TRANSFORMATION

La démarche archivistique est amenée à évoluer dès lors que l'on se place dans une perspective de traitement de l'archive par la technologie numérique. Or, cette démarche s'institue initialement en regard de procédés heuristiques qui rejoignent ceux des chercheurs. Pour les historiens du théâtre et de la vie culturelle des XVII^e et XVIII^e siècles, le fonds d'archives des registres est considéré comme une source massive à interpréter, un témoignage ou une preuve objective au service d'une question ou d'une hypothèse de recherche. Ce fonds présente en l'occurrence la particularité de constituer une source homogène que l'historien peut aborder comme un corpus déjà constitué. Ce terme clef de « corpus » exige quelques remarques, car dans le cadre de notre programme, cette notion s'est d'emblée imposée comme la plus évidente, mais elle s'est avérée aussi être la plus problématique⁶.

⁵ Sur la notion de « collection » en bibliothèque, voir David-Georges Picard, « Constituer une bibliothèque, constituer des savoirs », *Bulletin des bibliothèques de France (BBF)*, n° 1, 2011, p. 6-9. Disponible en ligne : <http://bbf.enssib.fr/consulter/bbf-2011-01-0006-001>

⁶ Notre programme fait symptomatiquement partie des projets Corpus de l'Agence nationale de la recherche (ANR).

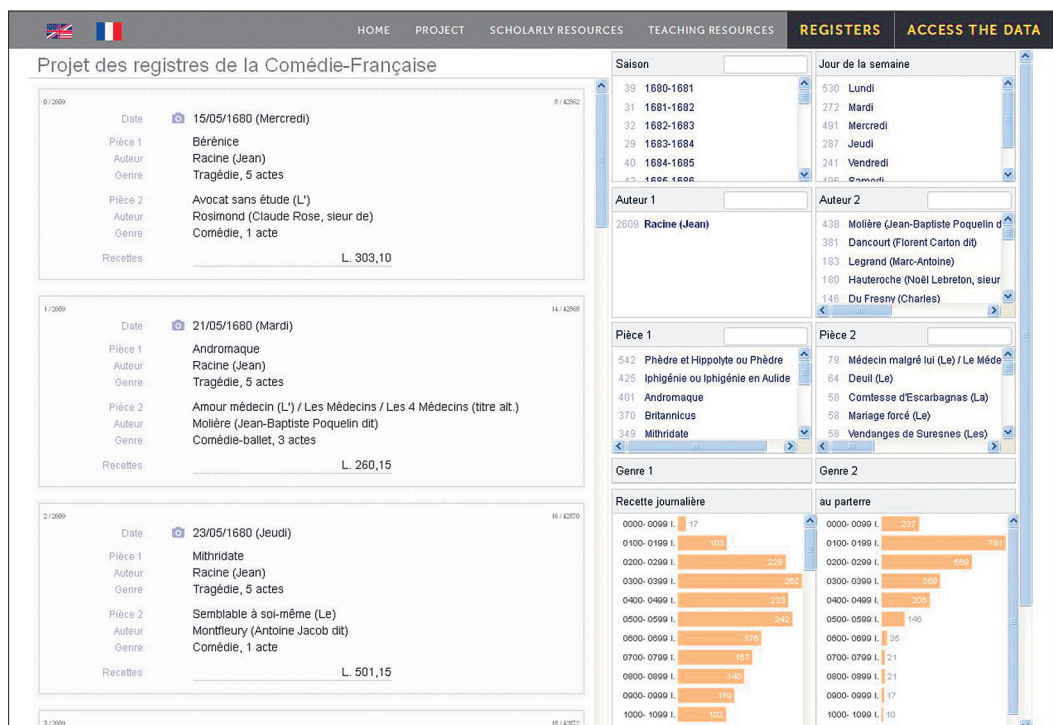
Les ouvrages d'archivistique consultés ne donnent aucune définition du terme « corpus », et pour cause, il s'agit d'un terme de vocabulaire propre à la démarche du chercheur. Le corpus constitue un ensemble exhaustif ou un échantillon de textes exemplaires et réunis selon des critères précis. Ceux-ci vont varier en fonction de la discipline, de l'étude scientifique à mener et de la méthode adoptée par le chercheur. Dans certains cas, le corpus est créé à des fins uniques d'édition. À cette mise à disposition de sources réunie en corpus se combine, dans d'autres cas, la mise en perspective des potentialités interprétatives des documents rassemblés. C'est à cette étape qu'un corpus de sources documentaires devient un corpus de données. Le projet des registres de la Comédie-Française repose précisément sur cette combinatoire : nous avons sélectionné une série de registres qui forme un rayonnage de documents sur notre site. Parmi cet ensemble documentaire, nous avons choisi d'explorer un échantillon : les registres journaliers dédiés aux recettes. Cet ensemble administratif a été choisi en fonction d'un critère essentiel à cette étape de notre projet, celui de la régularité formelle et structurelle des données sur la longue durée. Ces registres journaliers des recettes contiennent du reste suffisamment d'informations diverses à interpréter pour constituer un terrain de recherche à explorer. En outre, le registre journalier présente un intérêt technique et technologique évident pour des spécialistes de l'informatique en raison du fait que la régularité de la forme du document et de son contenu dans le temps se heurte à l'instabilité de certaines données. Cette rencontre entre stabilité et irrégularité représente un vrai défi que nous avons tenté de relever en créant une base de données évolutive, adaptée à notre corpus, et en intégrant une dimension qualitative et critique au sein même du quantitatif.

Ce passage du corpus de documents au corpus de données est primordial dans le processus de travail, autant pour les chercheurs que pour les spécialistes de la documentation, car c'est très précisément à cette étape que la démarche interprétative de l'historien entre en résonance avec les méthodes heuristiques de l'archiviste. Aux côtés des informaticiens qui doivent intégrer la singularité dans la quantité, les chercheurs et archivistes, eux, doivent bien baliser le cheminement vers l'interprétation pour éviter

les égarements et les erreurs potentiellement induits par l'automatisation générale. Dans notre cas particulier, le corpus des registres de recettes est en effet déjà constitué en un ensemble cohérent et les catégories de données à traiter se sont imposées par la forme même du document et la régularité des catégories dans le temps : des titres, des noms d'auteurs, des places, des recettes, autant d'informations qualitatives échelonnées sur la longue durée pour former un ensemble quantitatif massif. La préparation du terrain d'analyse que représente la construction de la base de données a très rapidement posé un problème descriptif et interprétatif aux archivistes comme aux chercheurs en raison de la discontinuité ou de l'instabilité de certaines catégories importantes qui interrogent la pertinence et la justesse des calculs obtenus sur nos données.

La transformation du contenu des fonds d'archives en données massives et calculables a eu pour premier effet de rendre nécessaire une enquête sur le sens même de certaines de nos données qui sont au fond des notions centrales pour une grande partie des recherches à mener. Dans notre cas, une analyse qualitative et préventive doit donc accompagner la démarche quantitative qui sera entreprise par les usagers, puisque la réalisation de nos recherches en dépend. C'est pour cette raison que nous entreprenons maintenant la création combinée d'un dictionnaire des données et d'un espace documentaire et encyclopédique visant à éclairer notre travail et à assurer rigueur et scientificité à notre démarche. Des termes clefs pour comprendre les registres comme « créations », « reprises », « saisons », « soirée », « loges », « parterre », etc., seront ainsi précisément définis, parce qu'ils se présentent comme des conditions d'interprétation des données chiffrées. Notre espace documentaire impose du reste d'interpeller d'autres corpus de données et notamment d'autres corpus numériques constitués. Est ainsi appelé à se développer un environnement de données constitué de corpus complémentaires qui s'éclairent les uns les autres grâce à une dynamique d'interopérabilité qu'il serait souhaitable de mettre en œuvre avec d'autres projets en histoire du théâtre. Notre corpus doit donc être ouvert, tant pour accueillir d'autres catégories de données comme les registres des dépenses et des feux qui font déjà

Exemple d'une navigation à facettes.



partie de notre rayonnage virtuel, que pour le partager avec d'autres corpus numériques. C'est dire que le périmètre heuristique du chercheur s'élargit et que nous devons intégrer de nouvelles compétences et *habitus* dans nos parcours d'enquête (accéder aux données disponibles, jongler avec les outils, apprendre à développer un esprit critique et mesuré face à l'automatisation). En outre, l'interopérabilité qui peut constituer l'une des ouvertures de nos corpus exige de mettre en œuvre des pratiques plus collaboratives et collectives, ce qui apparaît aussi comme un changement dans la manière de faire de la recherche et de constituer des équipes de recherche.

Tout projet de recherche part de la définition et de la délimitation de son corpus, or cette notion est fluctuante en contexte numérique : partant d'un corpus de documents, on se déplace vers un corpus de données, lui-même pouvant être intégrés à un corpus de métadonnées. De ce point de vue, on peut dire que notre corpus actuel n'est pas stabilisé et qu'il est potentiellement en expansion. Extension microscopique, d'une part : l'information étant segmentable à l'infini, on peut toujours aller plus loin dans l'ex-

ploitation des données et la finesse des questions posées. Extension macroscopique, d'autre part : notre corpus n'est finalement qu'un corpus parmi d'autres, que la mise à disposition virtuelle permet d'explorer plus facilement, d'élargir et de métamorphoser au contact d'autres corpus. Le corpus en contexte numérique est donc fondamentalement mouvant et ouvert. Ce sont désormais les moyens techniques, financiers et humains dont nous disposons pour numériser, saisir les données, développer des instruments de recherche, assurer une veille sur les projets, déterminer des liens d'interopérabilité avec d'autres programmes, qui nous limitent dans nos ambitions, et donc dans les extensions des formes de corpus.

CULTURE DU COMPROMIS ET INTERDISCIPLINARITÉ FÉCONDE

Chaque demande de traitement spécifique du corpus numérique fait l'objet d'un développement et doit donc être financée : telle est la limite des projets d'humanités numériques qui représentent un véritable changement de

