



**HAL**  
open science

## **Comédie-Italienne de Paris (1716 - 1779) : étude du répertoire au moyen de la création d'une base de données**

Ada Desideri

### ► **To cite this version:**

Ada Desideri. Comédie-Italienne de Paris (1716 - 1779) : étude du répertoire au moyen de la création d'une base de données. Littérature générale [cs.GL]. Sorbonne Université, 2025. Français. ⟨NNT : 2025SORUL164⟩. ⟨tel-05559733v2⟩

**HAL Id: tel-05559733**

**<https://theses.hal.science/tel-05559733v2>**

Submitted on 19 Mar 2026

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

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



HAL Authorization

# Sorbonne Université

Ecole doctorale IV – Civilisations, cultures, littératures et sociétés

*Initiative Théâtre - PRITEPS / ELCI / ObTIC*

## **Comédie-Italienne de Paris (1716 – 1779) : étude du répertoire au moyen de la création d'une base de données**

Par Ada Desideri

Thèse de doctorat de Etudes Romanes Italiennes

Dirigée par Andrea Fabiano (directeur), Glenn Roe (co-directeur) et Emanuele  
De Luca (co-encadrant)

Présentée et soutenue publiquement le 17 décembre 2025 à 15h00 au Campus Pierre et Marie  
Curie, salle de conférence du SCAI, bâtiment Esclangon, 1er étage

4 place Jussieu, 75005, Paris

Devant un jury composé de :

Directeur de thèse : M. Andrea FABIANO, professeur des universités, Sorbonne Université

Rapporteuse : Mme Marine ROUSSILLON, professeure, Université Sorbonne Nouvelle

Rapporteuse : Mme Barbara SOMMOVIGO, professeure, Université de Pise

Examinatrice : Mme Camilla CEDERNA, professeure émérite, Université de Lille

Co-directeur : M. Glenn ROE, professeur, Sorbonne Université

Co-encadrant : M. Emanuele DE LUCA, maître de conférences, Université Côte d'Azur

*A nonna Anna, che ha tanto voluto per me ciò che a lei non é stato permesso.*

# Remerciements

Mes premiers remerciements vont aux stagiaires, Chloé Petridis, Chiara Protani, Lucas Gaiotti, Mahsa Saberi Moayed et Flavia Falbo, qui ont tou.tes travaillé sérieusement et de manière passionnée, donnant une contribution fondamentale au projet. Je remercie aussi spécialement Camilla Cantiani, qui s'est intéressée au projet et qui m'a amplement aidé même avant d'être intégrée comme ingénieure informatique. Sans leur travail la base de données n'aurait pas pu exister. Un ultérieur remerciement particulier est dédié à mon père, Umberto Desideri, qui a non seulement saisi des données par pure curiosité et par pur esprit de solidarité, mais qui m'a aussi toujours aidé à y voir plus clair dans le monde universitaire.

Ensuite je tiens à remercier toutes les personnes grâce à qui je me suis sentie moins seule dans ce parcours que l'on sait être psychologiquement éprouvant. Premièrement les doctorant.es en lutte, en particulier Adélaïde Allemand, Emma Belkacemi-Molinier, Marie De Toledo, Alexandre Detruit, Stella Garnier, Roméo Molina, ainsi que toutes les personnes qui luttent à l'Université, des syndicats, aux comités Palestine, aux participant.es et animateur.ices des ateliers SENS. Je remercie également de tout mon cœur les membres de l'équipe ObTIC et du SCAI avec qui on a partagé, non seulement des locaux, mais aussi des moments conviviaux ; je pense en particulier à Angélique Allaire, Johanna Cordova, Una Faller, Caroline Koudoro-Parfait, Ljudmila Petkovic et Adrien Tusseau. Je tiens à remercier ici aussi Bianca Vallarano et Camilla Cederna, pour avoir continué à m'intégrer dans le projet d'édition des inédits d'Elisa Chimenti et pour avoir partagé des moments humains et de recherche qui me motivent dans la carrière universitaire. Je remercie finalement et spécialement les membres du comité de suivi, Marguerite Bordry et Yoann Dupont, qui m'ont soutenu et aidé, avant et pendant ma thèse, techniquement et humainement.

Un remerciement spécial va évidemment à mes directeurs et encadrants, Andrea Fabiano, Emanuele De Luca et Glenn Roe, car rien n'aurait été possible sans eux, mais qui ont aussi valorisé mon travail et qui m'ont accompagnée, tout en me permettant d'avancer autonomement. Pour les aspects les plus techniques, je remercie encore une fois l'équipe ObTIC et en particulier Motasem Alrahabi et Valentina Fedchenko, mais aussi le SCAI, l'ISCD et SACADO, qui ont fourni du matériel, des locaux et de l'assistance pour accomplir ce travail. Pour le suivi informatique à l'avenir, je remercie le LabRadar, qui héberge nos plateformes et notre base de données gratuitement.

Pour leur disponibilité et pour leur intérêt dans ce projet, je remercie spécialement Renzo Guardenti, Barbara Sommovigo et Myriam Quatrini : les rencontres avec elleux ont été d'une aide fondamentale pour cette recherche. Pour son aide dans ma recherche, je tiens à remercier aussi Yann Trividic, qui a ouvert mes perspectives de recherches au design et à l'interaction humain-ordinateur, avec des conseils de lecture qui ont été fondamentaux dans la phase finale de l'écriture de cette thèse.

Pour leur soutien humain et matériel je remercie enfin Salah, tou.tes mes ami.es et toute ma famille : lorsque dans la thèse je parle de construction sociale de la connaissance, je pense en réalité, tout premièrement, à vous. En particulier, merci à Bernardo Agnelli e Cosimo Pantaleoni pour les conseils historiques et historiographiques.

## **Note sur les traductions**

Toutes les citations tirées de textes qui ne sont pas en français dans la bibliographie sont traduites. La citation dans la langue d'origine est reportée en note de bas de page. Toutes les traductions sont faites par la doctorante.

# Sommaire

Remerciements .....	3
Note sur les traductions .....	5
Sommaire .....	6
1 Introduction .....	8
1.1 Formaliser signifie interpréter .....	8
1.1.1 Modèles et sens .....	8
1.1.2 L'activité interprétative .....	11
1.1.3 Etat d'avancement du travail .....	14
1.2 Etat de l'art .....	16
1.2.1 Le Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle .....	20
1.2.2 Les sources .....	26
1.2.3 Etudes théâtrales et humanités numériques .....	29
1.3 Aspects méthodologiques .....	33
1.3.1 Concepts à désambiguïser .....	33
1.3.2 L'équipe .....	35
1.3.3 Procédure .....	37
2 Partie 1 : construction de la base de données .....	40
2.1 Préparer la numérisation .....	40
2.1.1 Objectifs de la numérisation .....	40
2.1.2 Méthodologie de la numérisation .....	48
2.2 Résultats de la numérisation .....	59
2.2.1 Présentation de la structure formelle des données .....	60
2.2.2 Conséquences de nos objectifs sur la modélisation .....	70
2.2.3 Présentation de l'interface de consultation .....	74
2.3 Interopérabilité .....	83
2.3.1 Préparer l'interopérabilité .....	83
2.3.2 Le projet Transnational Stages .....	84
2.3.3 Logiciels et formats utilisés .....	87
3 Partie 2 : utiliser la base de données .....	91
3.1 Utilisation de l'interface .....	91
3.1.1 Figure de la coquette .....	92
3.1.2 Parodie et Opéra-Comique .....	97
3.2 Utilisation de l'interface par des spécialistes .....	109
3.2.1 Déroulement des entretiens .....	110
3.2.2 Observations .....	110
3.3 Utilisation des données brutes .....	112
3.3.1 Lieux mentionnés dans les titres .....	113
3.3.2 Magie et métamorphoses .....	122
4 Partie 3 : discussion .....	130
4.1 Discussion des résultats .....	131
4.1.1 Théâtre et Comédie-Italienne .....	131
4.1.2 Transdisciplinarité .....	136
4.2 Questions méthodologiques et épistémologiques .....	143
4.2.1 Observations sur l'automation et sur l'IA .....	143
4.2.2 Epistémologie des Humanités Numériques .....	146
4.3 Vers des humanités numériques écologiques .....	158
4.3.1 De quelle écologie parle-t-on ? .....	158

4.3.2	Evaluation écologique de notre projet.....	160
5	Conclusion.....	168
5.1	Résumé des résultats.....	168
5.2	Développements futurs .....	170
	Bibliographie.....	173
	Annexes.....	191
	Plateformes pour les études théâtrales .....	191
	Liens vers la documentation.....	197
	Table des illustrations.....	198

# 1 Introduction

Que les données aient une importance capitale dans le monde d'aujourd'hui est un fait indéniable. Que leur récolte, exploitation et archivage engendre des problématiques scientifiques, sociales et écologiques est de plus en plus évident (CNRS, 2019 ; ROUSSILHE et BERTHOUD, 2021). Se questionner sur la pratique de leur production, ainsi que sur les résultats qu'ils génèrent, peut donc faire partie de nos priorités de recherche (BAILLOT et LIGOZAT, 2025). Dans les humanités numériques, le caractère situé et interprétatif des données est particulièrement évident, puisqu'elles sont utilisées dans des contextes de recherche où l'activité interprétative donne valeur aux résultats. Autrement dit, si le caractère interprétatif des données est vrai dans tous les contextes (et les différents scénarios générés par différentes modélisations des conséquences de la crise écologique le montrent dramatiquement), en travaillant avec les lettres l'activité interprétative est le point de départ du processus de création des données. Elle est même fondamentale et incontournable, car elle est la manière-même de procéder dans la recherche littéraire (CITTON et CUSSET, 2017).

## 1.1 Formaliser signifie interpréter

### 1.1.1 Modèles et sens

Problématiser la création des données est propre aux humanités numériques, qui mettent au centre de ce procédé la notion de *modèle* et de *modélisation*. Dans le volume *Digital Humanities. Metodi, strumenti, saperi*, coordonné par Fabio Ciotti (CIOTTI, 2023), la première partie des chapitres thématiques est dédiée à « Modelli, metamodelli e modellizzazione nelle Digital Humanities » (CIULA et MARRAS, 2023), car

« une des pratiques qui caractérise [les Humanités numériques] est la modélisation [...] qui implique la traduction des schémas conceptuels, des systèmes complexes de connaissance et de leurs objets, dans des modèles qui peuvent être traités computationnellement, ou dans des paradigmes opérationnels. »<sup>1</sup> (ibid., p. 51)

---

<sup>1</sup> «Una della pratiche caratterizzanti [le Digital Humanities] é la modellizzazione [...] la quale implica la traduzione di schemi concettuali, di sistemi complessi di conoscenza e relativi oggetti, in modelli processabili computazionalmente o in paradigmi operazionali.»

Autrement dit, modéliser permet de transformer les objets et méthodes de recherche en des formats numériques, donc exploitables avec des ordinateurs – notamment des métadonnées.

De manière plus ou moins explicite, un ensemble de données est accessible à l'être humain, via une machine, s'il sous-tend un modèle qui donne du *sens* à ses éléments. Anthony Masure écrit à propos de cette notion :

« s'il est compliqué d'en donner une définition précise, on remarquera tout de même que ce concept se réfère toujours à trois points clés : une fonction de représentation (un modèle est toujours modèle de quelque chose), une fonction de réduction (le modèle ne conserve que les propriétés les plus intéressantes du point de vue de la recherche), et enfin une fonction subjectivante (le modèle ne fonctionne que pour un contexte donné). » (MASURE, 2017, p. 41-42)

En effet, selon les contextes, la forme-même du modèle peut être très différente. Le modèle d'une édition numérique, par exemple, décrit l'ensemble des phénomènes à annoter et la manière de représenter le texte (MANCINELLI et PIERAZZO, 2020). Il se présente donc le plus souvent comme un document de plusieurs pages, avec les informations sur les formats et outils à utiliser, et avec la liste des phénomènes à annoter, avec des exemples. En revanche, le modèle d'une base de données est un schéma qui représente la structure des données (DESIDERI, 2024), et il est donc généralement constitué d'une image, associée à un document contenant une explication détaillée des éléments du schéma, sous forme de listes. Dans les deux cas, lorsque l'on applique le modèle à des données réelles, on attribue un sens à une certaine donnée en la posant à une certaine place du modèle. Dans le cas de l'édition numérique, s'il a été fait le choix d'annoter les lieux en utilisant le formalisme TEI (sur le formalisme TEI, voir BURNARD, 2015), un mot ou un ensemble de mots du texte acquiert la signification de *lieu* – pour l'analyse humaine et pour l'ordinateur – s'il est entouré de la balise <placeName>. Dans le cas d'une base de données d'une bibliothèque, une suite de mots sur la couverture d'un livre acquiert la signification de *titre* si elle est insérée dans la catégorie des titres de la base de données. Autrement dit, dans les humanités numériques : « modéliser est [...] une stratégie de recherche et une pratique [...] dans lesquelles on manipule et on crée des objets et des

représentations extérieures [...] et on donne une signification aux objets et aux phénomènes. »<sup>2</sup> (CIULA et MARRAS, 2023, p. 52)

Cette activité d’attribution de sens est fondamentale, et au même temps problématique, car le processus de modélisation, comme tout processus de formalisation, doit aboutir à des définitions non ambiguës et à une structure opérationnelle, afin de pouvoir dialoguer avec un ordinateur : « le numérique ne fait que manipuler (calculer) des éléments "discrets", c'est-à-dire appréhendés depuis leur forme syntaxique, et non pas depuis leur sémantique (ce qu'ils signifient) » (MASURE, 2017, p. 121-22). La tension entre le caractère insaisissable des données littéraires et historiques d’un côté, et le caractère non ambigu de la formalisation numérique de l’autre, a été source de contraintes fructueuses au sein du travail de création de la base de données pour le Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle, qui a constitué le centre de la recherche présentée dans cette thèse. Ce projet sera ainsi le terrain pour observer et problématiser le processus de création des données et leur exploitation.

L’appel au numérique dans l’étude du Répertoire de la Comédie-Italienne de Paris, par la création d’une base de données pour le XVIIIe siècle, répond à la volonté de faire des analyses quantitatives sur toute la programmation du théâtre à cette période, et de manière interactive. Cette volonté témoigne par ailleurs que la pratique des analyses quantitatives est éprouvée dans l’étude de ce Répertoire (voir par exemple FABIANO, 2018) : la création d’un nouvel outil s’insère de manière organique dans les recherches existantes et peut contribuer concrètement aux recherches futures. La création de cette base de données est aussi le résultat d’une disponibilité des sources en ligne et des possibilités de partage offertes par le réseau mondial.

Or, le fait qu’un travail de formalisation relève d’une activité interprétative peut être formulé comme un paradoxe, si l’on regarde, à priori, une base de données comme la nôtre : les analyses quantitatives sont considérées comme objectives, et le but de notre base de données est justement de faire des analyses quantitatives. Mais en lettres toute recherche est fondée sur une certaine interprétation, qui est nécessairement subjective. Nous voulons donc obtenir des résultats quantitatifs (donc objectifs) subjectifs (donc non objectifs) et cela amène au paradoxe (car on aurait des résultats au même temps objectifs et non objectifs). Poser la problématique

---

<sup>2</sup> “Modellizzare è [...] una strategia di ricerca e una pratica [...] in cui si manipolano e si creano oggetti e rappresentazioni esterne [...] e si dà significato a oggetti e fenomeni.”

de cette manière est sans doute simplificateur, car la valeur d'objectivité et subjectivité n'est pas la même dans des contextes différents. Néanmoins, formuler ce paradoxe aide à mettre en lumière la relation entre des paradigmes épistémologiques profondément différents – les lettres, les sciences humaines, les sciences formelles et les sciences expérimentales : quelle valeur peut avoir de quantifier une interprétation, du point de vue épistémologique ? En quelle mesure peut-on formaliser, non seulement nos objets, mais aussi le processus interprétatif ? Ces questions nous ont préoccupé tout le long des trois années de recherche, et constituent le fil rouge de ce texte, en complément de la problématique spécifique à notre recherche : qu'est-ce qu'une base de données *fait* à l'étude du Répertoire de la Comédie-Italienne de Paris ?

### 1.1.2 L'activité interprétative

Il faut tout d'abord définir qu'est-ce que cela signifie qu'*interpréter*. Dans le livre *Lire, interpréter, actualiser* (CITTON et CUSSET, 2017), Yves Citton, spécialiste d'humanités numériques, propose de nombreuses théories de l'interprétation, appliquées dans le domaine des études littéraires. Ce volume est focalisé sur l'analyse de textes littéraires, et certaines parties ne sont pas pertinentes pour aborder la création d'une base de données ; en revanche, les considérations sur l'interprétation et sur sa définition résonnent amplement dans notre contexte de recherche.

Afin de comprendre le mécanisme de l'activité interprétative lors de la lecture d'un texte, il y a deux éléments fondamentaux au départ : le livre et le lecteur.ice. Le livre est à son tour écrit par un.e auteur.ice, qui porte un contexte historique de création de l'œuvre, ainsi qu'une intention. Le lecteur.ice, de son côté, vit et agit dans des contextes qui définissent la ou les communautés interprétatives dont il ou elle fait partie, ce qui engendre un certain partage du sensible (RANCIERE, 2000). Par conséquent, parmi l'infinité potentielle d'interprétations possibles d'un texte donné, un.e certain.e lecteur.ice en fournira une qui est cohérente et pertinente au sein de sa communauté interprétative (CITTON et CUSSET, 2017). Autrement dit,

« parmi l'infinité virtuelle de « groupements », de « corrélations » et d'« équivalences » pouvant être repérés entre les indices objectifs fournis par le texte, chaque interprétation opère [...] un travail de *sélection*, en ce qu'elle ne construit qu'un sous-ensemble (fini et éminemment limité) de mises en rapport significatives. Toute émergence de signification est fondée sur ce

type de sélection, qui ne retient que *certaines* caractéristiques présentées par l'objet considéré – ces caractéristiques que la linguistique saussurienne qualifie de *pertinentes*. » (ibid., p. 67)

L'activité interprétative consiste donc, en premier lieu, en un travail de sélection des éléments qui sont significatifs pour le lecteur.ice. Cette conception de l'interprétation comme sélection permet de tenir compte de la matérialité des objets et des personnes qui participent à l'acte interprétatif, mais aussi de leur caractère situé : le texte contient un ensemble d'« indices objectifs », qui sont les mots qui le composent, dans l'ordre qui leur est donné par l'auteur.ice. Cela donne une valeur d'objectivité au texte, par le fait que toute personne reconnaîtra la présence d'un certain mot à un certain endroit du texte, sans ambiguïtés. Autrement dit, un mot est présent à un endroit du texte, indépendamment de l'interprétation qu'on donnera de ce mot à cet endroit du texte (le cas des textes anciens, reconstruits philologiquement, est bien sûr plus complexe).

Du point de vue du lecteur.ice, s'il, ou elle, a en théorie la liberté de dire ce qu'il veut ou elle veut, en pratique son analyse tiendra compte de l'agencement réel des mots (donc de la réalité matérielle du texte, de ce que le texte lui *suggère*), et respectera des méthodes et des règles, afin de fournir une interprétation qui soit acceptable au sein de ses communautés interprétatives. Dans ce sens, « *l'interprétation n'est rien d'autre qu'une structure de contraintes* (plus ou moins rigoureuses, plus ou moins exigeantes, plus ou moins explicites) que l'on se pose pour rendre compte d'un texte. » (ibid., p. 85)

Or, sélectionner et définir des contraintes, est aussi l'acte premier de l'activité de modélisation :

« c'est justement le processus de sélection qui représente l'un des aspects les plus significatifs et spécifiques de la modélisation dans les Humanités Numériques, autrement dit, il s'agit du processus analytique où l'on décide quelles sont les approches interprétatives et les choix techniques et technologiques les plus adaptés, afin de poursuivre dans la phase d'analyse plus descriptive et explicative. »<sup>3</sup> (CIULA et MARRAS, 2023, p. 57-58)

Dans les deux cas, l'activité de sélection vise à répondre à des questions précises et à fournir un certain type d'analyse : dans ce sens, l'activité de modélisation peut être assimilée à

---

<sup>3</sup> «É proprio nel processo di selezione che consiste uno degli aspetti più significativi e caratterizzanti della modellizzazione nelle Digital Humanities, ovvero il processo analitico decisionale degli approcci interpretativi e delle scelte tecniche e tecnologiche adeguate a seguire la fase di analisi più descrittiva ed esplicativa.»

une activité interprétative. Le strict lien entre interprétation et modélisation est par ailleurs explicité dans le chapitre de Ciula et Marras (ibid.) dans un schéma représentant le cycle de la modélisation (Figure 1) : le point de départ, est constitué par la « compréhension/interprétation » du « phénomène/objet culturel » que l'on veut modéliser.

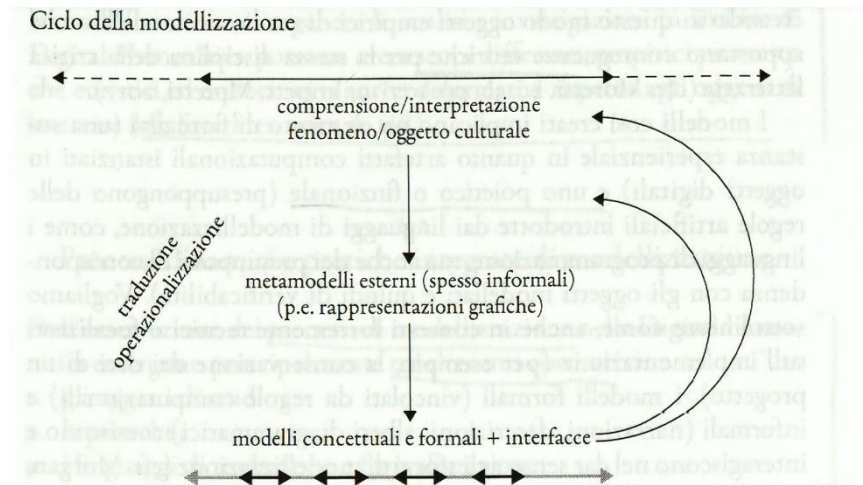


Figure 1 : Cycle de la modélisation tirée du chapitre de Ciula et Marras du volume *Digital Humanities. Metodi, strumenti, saperi*. (CIULA et MARRAS, 2023)

La rencontre entre un livre et un.e lecteur.ice, qui constitue le commencement de l'activité interprétative, n'advient généralement pas par hasard : quelque chose a mené la personne au texte qu'elle a devant ses yeux – le conseil d'un ami ou du libraire, une citation dans un autre livre, une couverture de sa couleur préférée mise en évidence dans sa librairie préférée... Un.e lecteur.ice choisit un livre pour une raison, parce qu'il ou elle espère que cela lui apporte quelque chose (du divertissement, un éclairage sur sa vie ou sur la vie de ses proches, des idées pour éduquer son enfant, par exemple) ; autrement dit, le lecteur.ice choisit un livre car il ou elle a des questions à lui poser (CITTON et CUSSET, 2017). Cela est tout aussi vrai lorsqu'un.e chercheur.euse rentre dans une interface de consultation, pour chercher des données pour sa recherche.

« Ce geste d'*interrogation*, qui constitue le moment premier de la pratique herméneutique, l'étymologie en suggère bien la double nature. En surface, il s'agit d'un simple *questionnement* porté sur l'œuvre [...], à travers sa question, l'interprète se borne à sélectionner un point de vue, un angle d'attaque, une dimension de pertinence. Si cette sélection est bien entendu d'ores et déjà décisive pour orienter le sens qu'on finira par reconnaître dans l'œuvre, elle paraît encore laisser parfaitement ouvert le spectre de réponses que le texte pourra apporter à la question qu'on

lui pose [...] Cette aptitude – louable, nécessaire et « scientifique » - d'ouverture d'esprit envers une multiplicité de réponses possibles ne paraît toutefois pas pouvoir être considérée comme étant la règle première et spontanée du jeu herméneutique. [...] La *rogation* ne se contente pas de poser une question pour se laisser surprendre par une réponse inattendue : elle consiste tout autant en une requête, une sollicitation, portant sur un objet précis et prédéterminé, que l'on cherche à obtenir. » (ibid., p. 95-96)

Ces considérations sur le geste d'interrogation sont particulièrement significatives dans le contexte d'une base de données : afin d'analyser les informations qu'y sont contenues, nous *interrogeons* les données, nous faisons des *requêtes*, de manière à obtenir les résultats dans la forme qui nous sert pour répondre à la question initiale. Or, cela implique que les données requêtées doivent pouvoir répondre à nos questions, c'est-à-dire que leur forme, leur *modèle*, doit être compatible avec nos objectifs de recherche. Ainsi, *interpréter consiste à sélectionner, modéliser et interroger les données*, d'une manière qui soit cohérente et pertinente aussi bien du point de vue de la base de données, que des objectifs de recherche et des communautés interprétatives de l'interprète. Le processus formel de modélisation des données est une activité interprétative dans la mesure où il constitue une « structure de contraintes » qui rend compte d'un objet d'étude (dans notre cas le Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle), et qui en saisit un angle spécifique (la chronologie de la programmation du théâtre), permettant donc de répondre à certaines interrogations (sur son évolution).

### 1.1.3 Etat d'avancement du travail

Afin d'observer l'ensemble de l'activité interprétative qui a guidé la création de la base de données pour le Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle, ce texte est divisé en trois chapitres. Le premier fournit une description détaillée du travail effectué et de ses deux résultats principaux, à ce jour : le modèle des données et l'interface de consultation de la base de données. Le deuxième chapitre montre des cas d'étude qui illustrent la manière dont les données peuvent être utilisées<sup>4</sup>. Le troisième et dernier chapitre discute des résultats obtenus dans les cas d'étude, afin de proposer des considérations épistémologiques sur la signification et sur la valeur de ces données au sein de la recherche sur la Comédie-Italienne de Paris, et plus généralement dans ce type de projet. La suite de l'introduction présente un état de l'art succinct sur les études sur la Comédie-Italienne de Paris et sur les recherches mêlant études théâtrales et

---

<sup>4</sup> Ces cas d'études ayant été effectués sur des données non vérifiées, ils doivent plus être considérés comme des exemples des types d'analyses que la base de données permet de faire, que comme des preuves.

humanités numériques, suivi d'une désambiguïisation de certains termes importants et de la présentation détaillée de la méthode de recherche suivie.

Cette thèse présente le travail de création de la base de données pour la Comédie-Italienne tel qu'accompli à un instant T, correspondant précisément à la publication d'une première version de l'interface de consultation des données. A ce moment, certaines tâches ont été totalement accomplies, certaines doivent être terminées et certaines autres sont à peine commencées – comme schématisé dans la Figure 2.

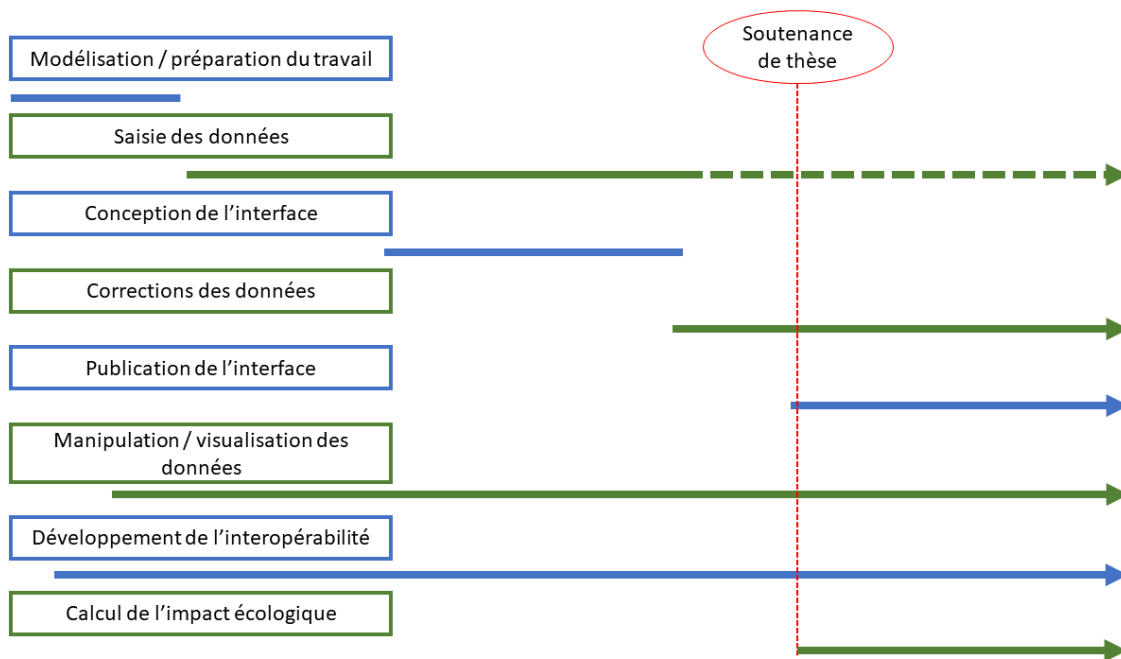


Figure 2 : développements des tâches pour la création de la base de données par rapport au moment de la publication de cette thèse.

La modélisation/préparation du travail concerne toutes les tâches informatiques nécessaires à la création de la base de données ; sa fin correspond donc au début de la saisie des données. La ligne décrivant cette dernière tâche présente le trait pointillé à la fin, pour indiquer que la saisie est déjà finie au moment de la soutenance de thèse, mais qu'il est toujours possible d'élargir davantage les données, en utilisant la structure fournie, mais aussi en ajoutant des éléments au modèle. La conception de l'interface, telle que publiée au moment de la soutenance de thèse, a été accomplie à l'intérieur des trois années de thèse, en parallèle à la saisie des données et à leur manipulation, tout en tenant compte du développement de l'interopérabilité. Conception de l'interface et modélisation sont les deux seules tâches qui se sont entièrement

déroulées au sein de cette recherche de doctorat. Il reste à accomplir tout ce qui concerne le calcul de l'impact écologique, qui est strictement lié au fonctionnement et à l'utilisation réels de l'interface. Certaines considérations théoriques sur l'écologie ont néanmoins guidé ce travail, comme détaillé dans la section sur l'écologie du troisième chapitre (Vers des humanités numériques écologiques).

Concernant la correction des données, elle a en partie été accomplie en parallèle à la saisie, lors des stages, mais la correction experte des représentations doit encore commencer. Les données devant encore être presque entièrement corrigées, tous les résultats obtenus à partir de celles-ci (qui sont présentés dans la Partie 2 : utiliser la base de données) sont susceptibles d'être infirmées par le simple accomplissement de cette activité. Néanmoins, présenter un bilan du travail à un moment décisif comme la première publication des données reste pertinent, car cela permet de tirer avantage de l'une des spécificités de la forme base de données, qui consiste en la possibilité de modifier les éléments qui la composent, tout en gardant sa cohérence formelle (FOLSOM, 2007). Autrement dit, la base de données *est faite pour* être modifiée et mise à jour. Ainsi, travailler sur des données non corrigées permet précisément de mieux les corriger, car les résultats intermédiaires peuvent en visibiliser des limites, voire aider à corriger des véritables erreurs. Dans ce sens, cette thèse a vocation, non seulement à donner les informations nécessaires pour compléter le travail, mais aussi, potentiellement, pour le développer.

## 1.2 Etat de l'art

Puisqu'il s'agit d'une thèse en humanités numériques, les domaines concernés sont multiples ; dans notre cas la recherche se situe au croisement entre études théâtrales et informatique. Or les études théâtrales sont à leur tour un champ de recherche transdisciplinaire, touchant à l'histoire, à la littérature, à la musique, à la danse, mais aussi à l'anthropologie ou à l'économie. Il est donc nécessaire de restreindre cet état de l'art à des recherches qui sont particulièrement proches de notre objet d'étude : nous nous focalisons sur la Comédie-Italienne de Paris au XVIIIe siècle et sur le fonctionnement général du système des théâtres parisiens d'Ancien Régime. Concrètement, la recherche s'est déroulée au sein d'une équipe<sup>5</sup>, structurée

---

<sup>5</sup> L'équipe restreinte est formée par les directeurs de cette thèse (Andrea Fabiano, Emanuele De Luca et Glenn Roe), les stagiaires (en ordre de chronologique Chloé Petridis, Chiara Protani, Lucas Gaiotti, Mahsa Saberi Maoyed, Flavia Falbo) et l'ingénieure informatique (Camilla Cantiani). Plus largement, ont contribué à la

autour d'expertises différentes et complémentaires : afin de garantir une profondeur scientifique sur tous les aspects du travail, chaque membre a contribué au dialogue transdisciplinaire depuis son point de vue. Cet état de l'art est fait depuis le point de vue de l'informaticienne qui a accompagné et coordonné la création de la base de données pour le Répertoire de la Comédie-Italienne de Paris ; par conséquent, les recherches mentionnées concernent les connaissances qui ont été utiles afin de réaliser ce travail, et n'ont pas vocation à être exhaustives.<sup>6</sup>

La rencontre entre études théâtrales et humanités numériques a produit de nombreuses recherches depuis plus de vingt ans. Celles-ci se sont particulièrement développées depuis la diffusion d'Internet, et se sont largement appuyées sur la possibilité de réunir des médias différents (textes, images, vidéos, sons) sur un même support (l'écran d'un ordinateur) :

« L'interconnexion entre les documents, le renvoi à des photographies de l'ouvrage, les discours de contextualisation, la disposition du texte ou la multiplication des instruments de recherche sont autant de moyens de réinscrire les documents dans leur contexte, de rendre visible la construction des savoirs et le travail d'interprétation, et de mettre à disposition des chercheuses et chercheurs les outils nécessaires à la critique des extraits présentés. » (ROUSSILON et SCHUWEY, 2020)

Les bases de données pour les études théâtrales représentent une partie fondamentale de ces productions : elles offrent la possibilité de consulter au même temps des matériaux différents, mais elles permettent aussi de réaliser des études quantitatives. Ces deux aspects, représentent une contribution fondamentale à un domaine de recherche transdisciplinaire, qui étudie des productions éphémères et aux moyens expressifs multiples, comme les pièces de théâtre.

Concernant plus strictement les humanités numériques, même s'il s'agit d'un champ difficile à définir et à cerner, les recherches qui s'attribuent cette étiquette sont aujourd'hui aussi nombreuses qu'il est possible de saisir des tendances et de distinguer des approches et des sous-domaines. Déjà en 2015, Citton présentait les humanités numériques comme composées de trois strates « appelées à se superposer entre elles bien davantage qu'à se remplacer l'une l'autre » (CITTON, 2015, p. 4). La première (*humanités numériques 1.0*) « explore[...] et

---

réalisation de ce projet de nombreux membres de l'équipe Obtic (particulièrement Motasem Alrahabi et Valentina Fedchenko), ainsi que le projet Lab Radar pour l'hébergement et la maintenance.

<sup>6</sup> Dans la suite du texte, l'utilisation du pronom personnel à la première personne servira à distinguer si le travail a été fait par la personne qui écrit ou par d'autres membres de l'équipe (qui seront le plus possible explicités).

expérimente[...] comment des moyens de traitement numériques peuvent enrichir notre accès, nos connaissances et notre compréhension de corpus (littéraires, artistiques, historiques) déjà constitués comme tels. » (ibid.) La deuxième strate (*humanités numériques 2.0*) exploite le numérique comme porteur d'un potentiel créatif pour les humanités, et prend en compte la subjectivité socio-politique du chercheur, ainsi que la dimension affective des choix de recherche. La troisième et dernière strate (*humanités numériques 3.0*) oriente son regard sur les processus de subjectivation de « corps humains éduqués au sein d'institutions sociales et d'appareils computationnels mis en réseau » (ibid., p. 6). Ces trois strates ainsi définies, en se superposant, orientent de plus en plus le regard sur la manière dont les êtres humains et les machines se construisent mutuellement.

Dans l'introduction au volume de 2023, *Digital Humanities. Metodi, strumenti, saperi*, Fabio Ciotti décrit aussi les humanités numériques comme composées d'une série de paradigmes qui ont apparu de manière successive au fil du temps mais qui ont entretenu une relation dialectique, « parfois opposée, parfois synthétique »<sup>7</sup> (CIOTTI, 2023, p. 24). Ces paradigmes définissent différentes visions du domaine des humanités numériques, engendrant des approches différentes de la part des chercheurs et différents types de questionnements épistémologiques. Parmi ces paradigmes, « le premier et le plus important [...] élaboré depuis les origines de ce champ d'études »<sup>8</sup> (ibid., p. 24), est celui que Ciotti appelle le « paradigme méthodologique », qui voit une unité méthodologique dans les humanités numériques, autour de la notion et de la pratique de la modélisation. Ce paradigme se décline en deux visions des résultats de la recherche. La première attribue un changement épistémologique à ce qui résulte de l'interaction entre informatique et humanités ; la deuxième borne le numérique à une position instrumentale, apportant des changements dans l'accès aux sources et aux résultats, sans engendrer des modifications profondes sur l'essence des recherches en humanités.

Un deuxième grand paradigme mentionné par Ciotti est celui « aux fondements pragmatistes et sociologiques »<sup>9</sup> (ibid., p. 26), qui réunit les humanités numériques en tant que communauté de pratiques (*community of practice*) : « une ample et diversifiée communauté de savants, qui s'intéressent, de différentes manières, à l'intersection entre numérique et sciences

---

<sup>7</sup> “Talvolta oppositiva talaltra sintetica.”

<sup>8</sup> “Il primo e più importante [...] elaborato fin dalle origini di questo campo di studi”.

<sup>9</sup> “Di impianto pragmatista o sociologico.”

humaines »<sup>10</sup> (ibid., p. 26). Ces « savants » partagent et développent ainsi des méthodes, des outils, des approches communes dans leur recherches. Ce deuxième paradigme a de son côté engendré toute une série de déclinaisons des humanités numériques qui focalisent leur attention sur des instances sociales et sur des variables sociologiques, en croisant leur chemin avec les études féministes (voir par exemple NYHAN, 2023) ou décoloniales (cet aspect a été largement abordé lors du colloque Humanistica 2025, par exemple par EL HADI, 2025), ou plus largement les études culturelles (voir par exemple PINOTTI et SOMAINI, 2016).

Ciotti mentionne en dernier une tradition italienne des humanités numériques, qui pourrait constituer un troisième paradigme par sa « forte spécificité intellectuelle, caractérisée par le refus d’une vision instrumentale de l’informatique et par un intérêt méthodologique et épistémologique répandu, même si pas majoritaire. »<sup>11</sup> (CIOTTI, 2023, p. 28) L’informatique est même prise dans sa signification originelle de « science théorique de la représentation et de l’élaboration de l’information »<sup>12</sup> (ibid.), et ces recherches valorisent ainsi l’apport des méthodes formelles à l’élaboration de données culturelles, avant même d’observer les potentialités du numérique des interfaces et de l’interactivité (voir notamment les publications de Tito Orlandi, par exemple ORLANDI, 2010). Le syntagme italien pour désigner les humanités numérique, *informatica umanistica*, a par ailleurs comme sujet l’informatique et comme attribut les humanités – de manière opposé au syntagme français ou anglais.

La problématique de cette recherche de doctorat porte sur les changements qui sont induits par la création d’une base de données dans l’étude du Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle, et se situe ainsi pleinement dans le domaine des humanités numériques. Mais, afin de couvrir les principales disciplines où notre recherche s’insère, la suite de cet état de l’art sera consacrée à notre objet d’étude spécifique, le Répertoire de la Comédie-Italienne de Paris. En particulier, il est constitué de trois parties : la première concerne les recherches sur la Comédie-Italienne de Paris réalisées et coordonnées par les membres de notre équipe spécialistes du sujet ; la deuxième les sources de ces recherches ; la troisième est dédiée aux publications scientifiques sur les humanités numériques pour les études théâtrales.

---

<sup>10</sup> “Un’ampia, diversificata e inclusiva comunità di studiosi che si occupano in vario modo dell’intersezione tra digitale e scienze umanistiche.”

<sup>11</sup> “Forte specificità intellettuale, caratterizzata dal rifiuto della visione strumentalista dell’informativa e dalla prevalente, sebbene non maggioritaria, impronta metodologica ed epistemologica.”

<sup>12</sup> “Scienza teorica della rappresentazione ed elaborazione formale dell’informazione.”

## 1.2.1 Le Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle

Concernant la Comédie-Italienne de Paris, cette recherche traite le XVIIIe siècle, et plus précisément la période 1716-1779, dates entre lesquelles il existe une véritable compagnie italienne à Paris, dont font partie des acteurs et actrices italianophones et qui dialogue avec la tradition dramaturgique de la péninsule (FABIANO et DE LUCA, 2023). L'ancienne troupe italienne ayant été chassée par Louis XIV en 1697, douze comédiens (plus précisément onze acteurs et actrices et un chanteur), guidés par Luigi Riccoboni arrivent en France à la demande du Régent, et jouent leur première pièce – *L'heureuse surprise* – en 1716 (DE LUCA, 2015). 1779 a été choisie comme borne finale car à partir de cette année il est interdit à la compagnie de jouer et chanter en italien : la compagnie italienne a, dans ce sens, disparu. A cette période du siècle, après le rachat du bail et du privilège de l'Opéra-comique en 1762, la compagnie joue presque uniquement des opéra-comiques. Néanmoins, la base de données et l'interface pour la consulter couvrent une période plus étendue, jusqu'à 1792, année qui correspond au moment où la compagnie change définitivement son nom en Théâtre de l'Opéra-Comique National. C'est aussi la période couverte par une des sources de données principales pour notre base de données : le répertoire de Clarence Brenner, *The Theatre Italien, its repertory, 1716-1793*, publié en 1961 (BRENNER, 1961). Le musicologue américain écrit à propos du choix de la période que

« l'année 1780 aurait pu être choisie comme date de fin pour ce travail, car à partir de cette année la compagnie n'a plus performé de pièces en italien, et ses acteurs italiens ont été écartés. Mais il s'agit des seuls changements [...] j'ai donc préféré poursuivre avec l'histoire du répertoire jusqu'à 1793, quand le théâtre devient officiellement le Théâtre de l'Opéra-Comique National »<sup>13</sup> (ibid., p. v).

Au niveau de l'interface de consultation des données, les années 1780-1792 sont par ailleurs actuellement cachées, en attendant de pouvoir les lier avec les données du futur Opéra-Comique Database : cela répond au même temps à une démarche scientifique de valorisation des travaux réciproques, mais aussi à des objectifs de sobriété écologique dans le numérique.

---

<sup>13</sup> “The year 1780 might have been taken as the terminal date of this work, for after that year the company no longer performed plays in Italian and its Italian actors were dropped. But these were the only changes [...] I have preferred therefore to proceed with the history and repertory to 1793, when the theater officially became the Théâtre de l'Opéra-Comique National.”

Un premier théâtre Italien à Paris se développe au XVII<sup>e</sup> siècle sous la protection du roi. Mais en 1697 Louis XIV chasse la compagnie :

« depuis sa naissance [au XVII<sup>e</sup> siècle], la Comédie-Italienne se positionne [...] sur le plan politique de l'instabilité entre son excentricité par rapport aux dramaturgies françaises et sa permanence officielle dans l'ordre royal du système théâtral parisien, entre la subversion de la poétique classique et l'acceptation de la protection du roi, qui en est à l'origine. Dans ce cheminement, l'expulsion de la troupe italienne ordonnée par Louis XIV est d'une certaine manière « génétiquement » inscrite, car le bouffon du roi se sait toujours consciemment balancer entre récompense et pendaison » (FABIANO, 2018, p. 25).

Au moment de l'arrivée de la nouvelle troupe italienne à Paris en 1716, le répertoire de la Commedia dell'Arte de l'ancienne troupe avait été approprié par les théâtres forains (GUARDENTI, 1995) :

« En 1716 [...] le dispositif des macrogenres [réglementé par les privilèges du roi] reste inchangé, car le périmètre qui avait appartenu à la Comédie-Italienne était resté occupé par les troupes des Foires [...] Riccoboni ne dispose donc que d'une marge de manœuvre réduite pour présenter un programme dramaturgique nouveau, différent, un théâtre italien reformé. » (FABIANO, 2018, p. 27)

Au sein de la stricte répartition des compétences des théâtres organisée par le roi, les Italiens doivent divertir. Or, Luigi Riccoboni est connu en Italie comme un acteur érudit, mais si « en Italie il avait interprété les héros tragiques des grands érudits, en France il n'intéressait son public que comme Lelio, l'Amoureux »<sup>14</sup> (SCHINO, 2002, p. 66). Sa tentative de réformer le théâtre italien par la création d'un nouveau répertoire tragique n'a pas de succès en France, où « après avoir goûté un concert de sonorités incomprises, le public réclame d'être amusé en français » (DE LUCA, 2011, p. 33). Ainsi, Riccoboni prend au début les distances avec l'ancienne troupe des italiens et avec les théâtres de la Foire, en congédiant Giovan Battista Costantini (acteur de l'ancienne Comédie-Italienne et ensuite entrepreneur forain) et en refusant les farces et les pièces du *Recueil Gherardi*. Mais déjà à la fin de 1717 il se retourne vers les forains, et en particulier vers Pierre-François Biancolelli, qui débute dans sa troupe en octobre,

---

<sup>14</sup> «In Italia aveva interpretato gli eroi tragici dei grandi letterati, in Francia sembrò interessante solo come Lelio, l'Innamorato.»

dans une volonté de francisation du Répertoire (DE LUCA, 2012). Pendant les étés entre 1721 et 1723, la Comédie-Italienne joue à la Foire Saint Laurent.

Ces changements stratégiques rapides reflètent deux réalités, par ailleurs fortement liées : la première est le caractère entrepreneurial des compagnies italiennes à l'époque de la naissance de la Commedia dell'Arte, cherchant un public pour vivre de leur art ; la deuxième est la formation hétéroclite des acteurs et des actrices. À propos de la naissance de la Commedia dell'Arte au XVI<sup>e</sup> siècle, Mirella Schino écrit :

« Commedia dell'Arte signifia, avant tout, l'autonomie des professionnels du spectacle, la vente au public, au prix d'un ticket, de ces spectacles qui étaient jusque-là mis en place uniquement lors des fêtes, surtout au carnaval. L'improvisation et les types (les masques) constituèrent les deux faces de la solution à un important problème commercial : la production rapide et variée de spectacles »<sup>15</sup> (SCHINO, 2002, p. 63).

La recherche de nouveaux publics pour vendre leurs spectacles mène les acteurs et actrices à voyager, dans les états italiens et à l'étranger. Ce caractère commercial persiste à l'époque de Riccoboni qui, selon Emanuele De Luca, « en acceptant de conduire sa troupe sur les tréteaux forains, fait preuve d'une conscience entrepreneuriale aiguë » (DE LUCA, 2011, p. 37).

Le besoin de renouveler le répertoire et de l'adapter à diverses circonstances s'accompagne avec une professionnalisation du jeu d'acteur :

« Les témoignages parlent de leur capacité [des acteurs et actrices] non seulement à improviser, mais aussi à jouer des comédies et des tragédies rédigées. Ils mettaient en scène des opéras, ils savaient danser. [...] Les compétences de l'acteur [italien] comprenaient celles du danseur, du chanteur, du musicien »<sup>16</sup> (SCHINO, 2002, p. 65).

Pendant la seconde moitié du XVIII<sup>e</sup> siècle à Paris cela fait toujours partie du bagage du « savoir-faire » des comédiens italiens :

---

<sup>15</sup> «Commedia dell'Arte significò, innanzitutto, l'autonomia dei professionisti dello spettacolo, la vendita al pubblico, al prezzo di un biglietto, di quegli spettacoli che fino a quel momento si allestivano solo in occasioni festive, soprattutto nel carnevale. L'improvvisazione e i tipi fissi (le maschere) costituirono le due facce della soluzione di un importante problema commerciale : le produzione rapida e variata di spettacoli.»

<sup>16</sup> «Le testimonianze parlano della loro capacità [des acteurs et actrices] non solo d'improvvisare ma anche di recitare commedie e tragedie scritte per esteso. Mettevano in scena opere in musica, sapevano danzare. [...] Le competenze dell'attore [italiano] comprendevano quelle del danzatore, del cantante, del musicista.»

« « sedentarisés » au sein de l'institution qui les accueille, les profils des comédiens [...] n'évoquent plus guère les acteurs professionnels itinérants, « souche » originaire et originelle de « leur » théâtre, mais ne sont pas moins rattachés aux modalités représentatives de leurs ancêtres. Le fantôme de la Commedia dell'arte, qui habite les canevas italiens du Répertoire de la Comédie-Italienne durant toute la seconde moitié du XVIIIe siècle, est l'empreinte d'une pratique traditionnelle qui se préserve tout en se renouvelant » (SPANU FREMDER, 2010, p. 12).

Les acteurs et actrices de la tradition italienne ont eu un rôle central dans la création des pièces et les frontières entre les dramaturges et les acteur.ices ne sont pas nettes (sur la Commedia dell'Arte et la dramaturgie italienne en France et en Europe voir FERRONE, 2024 et COMPARINI et FABIANO, 2019). Luigi Riccoboni, acteur, capocomico, dramaturge et théoricien du théâtre, en est un exemple. Mais au-delà de personnages comme Riccoboni qui incarnent les différents rôles, les pièces à l'improptu témoignent d'un procédé de véritable création collective des pièces :

« Quand on parle d'« improvisation » à propos de la Commedia dell'Arte on fait référence à des choses et à des pratiques multiples. Mais on fait surtout référence à une manière de produire les spectacles où chaque acteur compose les répliques de son rôle à partir d'un canevas réalisé par le dramaturge. Il ne s'agit donc pas d'improvisation telle qu'on pourrait l'entendre aujourd'hui et telle qu'on l'a cru longtemps : création spontanée devant le public »<sup>17</sup> (SCHINO, 2002, p. 63).

Dans sa thèse, Spanu décrit ce phénomène en relation à la troupe de la Comédie-Italienne dans la deuxième moitié du siècle qui nous intéresse :

« Ces spectacles, contrairement à ce que l'on pense, n'ont rien d'improvisé, hormis la maîtrise de ce qui se passe et ce qui s'écrit sur scène, ils sont le résultat des opérations de façonnage, des « concerts », comme les appellera Collalto, et de mise en œuvre d'une pratique et d'un savoir scénique ancien » (SPANU FREMDER, 2010, p. 18).

Plusieurs divisions internes au siècle peuvent être tracées pour rendre compte de l'évolution du Répertoire, mais, afin de décrire les aspects déterminants pour la création de la base de données, nous allons retenir uniquement 1762, année de l'acquisition du privilège de

---

<sup>17</sup> «Quando si parla di "improvvisazione" a proposito della Commedia dell'Arte si intendono molte cose, e molte pratiche. Ma si intende soprattutto un modo di produrre spettacoli in cui ciascun attore compone le battute della propria parte sulla base dello scenario realizzato dal drammaturgo. Non si tratta dunque di improvvisazione quale la si potrebbe intendere oggi e quale per lungo tempo fu creduta: creazione estemporanea davanti al pubblico.»

l'opéra-comique et de l'arrivée de Carlo Goldoni à la direction de la troupe. Pour la base de données c'est l'année à partir de laquelle moins de métadonnées sont disponibles sur les pièces, puisque le répertoire d'Emanuele De Luca s'arrête à cette date. Du point de vue du Répertoire, l'année 1762 marque un tournant vers la représentation musicale, qui débouchera sur la suppression des comédies en italien en 1779. Dans sa thèse, Silvia Spanu a étudié la période 1762-1780 comme exemplaire dans le processus d'hybridation de la tradition théâtrale italienne avec celle française :

« À l'intérieur du répertoire de la Comédie-Italienne d'après 1762, cohabitent des dramaturgies différentes, italienne et française, qui dialoguent constamment au sein d'un contexte productif particulièrement fertile qui permettra la naissance graduelle de genres hybrides, nés de la rencontre entre la dramaturgie italienne dell'arte, celle française de la farce, et celle musicale des opéras-comiques hérités de la foire : des opéras-comiques d'influence italienne et des comédies-parades imprégnées de dramaturgie des masques » (SPANU FREMDER, 2010, p. 35).

Ce processus concerne Carlo Goldoni aussi, par ailleurs connu pour sa réforme du théâtre en Italie (FABIANO, 2018) : « son arrivée à Paris s'insère dans un dessein de « musicalisation » de la Comédie-Italienne, fondé sur l'introduction des opéras buffe italiennes à côté des opéras-comiques de Favart et de ses alliés » (DE LUCA, 2011, p. 32). De Luca rappelle aussi que des six pièces de Goldoni jouées à la Comédie-Italienne avant l'arrivée du dramaturge, cinq font partie du répertoire musical. Deuxième réformateur du théâtre italien appelé en France, son nouveau théâtre ne répond pas non plus aux attentes du public de Paris. Dans l'ouvrage *La Comédie-Italienne de Paris et Carlo Goldoni*, Andrea Fabiano (FABIANO, 2018) retrace le parcours d'adaptation du dramaturge au contexte et aux attentes parisiennes, en observant ce phénomène sous le prisme de l'hybridation.

Dans la deuxième moitié du siècle, au sein du riche panorama des théâtres parisiens d'Ancien Régime, la Comédie-Italienne occupe une place particulière, intermédiaire, voire *instable* : il s'agit d'un théâtre subventionné par le roi mais pas privilégié, proche de la Foire. Plus précisément, Silvia Spanu écrit dans sa thèse :

« Contrepoids comique, s'exprimant par une expressivité exacerbée et par un comique grotesque, monstrueux et parfois obscène, la comédie des masques est également le théâtre du rire incontrôlable, voire « nerveux ». Ces particularités justifient la proximité de la commedia dell'arte avec les formes spectaculaires populaires fédérées sous le genre de la farce. En réalité,

la dramaturgie de la commedia dell'arte est au commencement la « voix » et la « voie » prioritaire de la circulation de la pratique théâtrale italienne migrée en France » (SPANU FREMDER, 2010, p. 27).

La position instable des comédien.nes italien.nes et de leur Répertoire dans le contexte théâtral parisien, se traduit dans un caractère expérimental de ce théâtre, surtout à partir de 1762 :

« la Comédie-Italienne est un laboratoire dramatique apte à accueillir des combinaisons fructueuses. Contexte favorable aux tentatives et aux expérimentations dramaturgiques, cette scène officielle apparaît clairement dans les critiques et écrits de l'époque comme un lieu « singulier », « mixte » ou « indéfinissable » » (SPANU FREMDER, 2010, p. 24)

Une publication collective récente, coordonnée par Andrea Fabiano et Emanuele De Luca (FABIANO et DE LUCA, 2023), réunit de nombreuses recherches axées sur l'étude du caractère expérimental de la Comédie-Italienne de Paris. Suivant les recherches des deux chercheurs, la base de données voudrait pouvoir accompagner l'étude du processus d'hybridation et les interactions fructueuses entre les traditions théâtrales françaises et italiennes, qui adviennent au sein du Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle.

Or, étudier le répertoire signifie orienter la recherche à la dimension spectaculaire du théâtre et se situer du point de vue des spectateurs (ROUGEMONT, 2001). En effet, les informations disponibles, ainsi que la manière dont elles sont traitées, permettent d'observer ce qui était proposé au public par la compagnie à cette époque. Cela est intéressant pour plusieurs raisons. Si d'un côté la programmation est très dépendante de la réaction du public, de l'autre côté celle des théâtres subventionnés était fortement réglementée par le roi : « si le répertoire d'un théâtre doit être considéré comme la « vitrine » de l'institution qu'il représente, il est, dans une acception plus large, le résultat d'une stratégie et d'une politique commerciale qui ont été étudiées et réfléchies pour répondre à des exigences d'ordre politique, social et esthétique » (SPANU FREMDER, 2010, p. 83). Ainsi,

« chargé d'une fonction esthétique, sociale et politique, le Répertoire d'un théâtre doit être considéré et étudié comme un objet dont la valeur est fonctionnelle et strictement proportionnelle au discours, au message qu'il veut canaliser. Mais le Répertoire détient et

sauvegarde également la mémoire du théâtre auquel il appartient » (SPANU FREMDER, 2010, p. 129).

Numériser le Répertoire de la Comédie-Italienne au XVIIIe siècle peut donc contribuer à l'étude de plusieurs aspects du théâtre dans sa dimension spectaculaire, sociale et politique (ROUSSILON et SCHUWEY, 2020).

### 1.2.2 Les sources

Si étudier le théâtre ancien signifie toujours étudier des objets éphémères et uniques, dont nous n'avons aucune trace sensorielle à part le matériel iconographique (sur l'étude du théâtre avec les images, voir les publications de Renzo Guardenti, par exemple GUARDENTI, 2021), les sources à dispositions peuvent néanmoins être différentes, par nature, par extension et par degré d'organisation. La Comédie-Française, par exemple, « théâtre le plus ancien et le plus institutionnel de France » (BIET et al., 2020, p. 13), bénéficie de registres complets et continus depuis le XVIIe siècle, tenant compte de l'activité du théâtre. La complétude et l'extension de ces documents permet une étude continue de la programmation et de la gestion économique du théâtre, avec des données comparables sur plusieurs siècles. Ces registres font par ailleurs l'objet d'un projet de base de données depuis plus que dix ans, avec une interface de consultation qui fournit de nombreuses fonctionnalités pour analyser et consulter les données<sup>18</sup>. Pour la Comédie-Italienne, dont la forme de l'institution a été plus instable, les registres n'ont pas été conservés pour toutes les années, l'historiographie et l'analyse littéraire souffrent d'un manque de textes et de documents. Ainsi, « évoquer les répertoires et les spectacles de la Comédie-Italienne des XVIIe et XVIIIe siècles signifie parler d'un phénomène théâtral et d'une bibliothèque dramatique qui ont perdu toute sorte de matérialité » (SPANU FREMDER, 2010, p. 13).

Les recherches auxquelles nous faisons référence sont le plus souvent basées sur des ensembles de sources de nature différente. Dans l'étude du Répertoire, une source incontournable est constituée par les Registres des Dépenses et Recettes de la compagnie : il s'agit de nombreux volumes manuscrits que la compagnie tenait pour rendre compte de ses activités. Ils mentionnent à minima la liste des pièces jouées chaque soir et les recettes de la soirée, sur certaines périodes ils contiennent plus de détails, surtout sur des aspects

---

<sup>18</sup> <https://www.cfregisters.org/#/>

économiques (détails des dépenses et des recettes, nombre de spectateurs, par exemple). Actuellement, ces Registres sont tous numérisés et consultables sur le site Gallica<sup>19</sup>. Néanmoins, certains Registres ont disparu, les comptes de certaines années ont été ainsi perdus (un schéma de la présence et absence des Registres est donné dans la deuxième partie, à la Figure 33).

Pour connaître la programmation du théâtre dans ces périodes-là, des sources largement mobilisées par les chercheurs.euses ont été les revues de l'époque, et particulièrement le *Mercur de France*. Une numérisation de cette publication est actuellement en cours, et un site de consultation avec certaines fonctionnalités est déjà en ligne<sup>20</sup> ; nous avons utilisé cette plateforme lors d'un stage (pour les détails sur le déroulement des stages voir la section Chaîne de traitement des informations), afin de remplir les métadonnées des pièces. Pour les pièces qui ont été éditées, l'édition constitue une ultérieure source qui peut être mobilisée. Néanmoins, si cela donne des informations spécifiques sur une certaine pièce, sa publication ne fournit pas de détails, à priori, sur son existence au sein du Répertoire. Enfin, étant la Comédie-Italienne un théâtre subventionnée par le roi, des informations sur la gestion de la compagnie existent dans les Archives Nationales – Silvia Spanu a en partie utilisé ce type de source dans sa thèse pour la deuxième moitié du siècle (SPANU FREMDER, 2010).

S'il a été possible de construire une base de données et une interface de consultation dans les trois ans d'une thèse de doctorat, c'est parce que les informations contenues dans les sources avaient déjà été réorganisées auparavant, dans deux répertoires : BRENNER, 1961, et DE LUCA, 2011. Ces deux documents sont écrits utilisant le français ancien et contemporain, et l'italien et l'anglais contemporains. Le répertoire de Brenner est dactylographié, celui de De Luca est en format PDF<sup>21</sup> et Excel. En revanche, les sources à partir desquelles ils ont été créés sont en grande partie des sources du XVIIIe siècle, dont les célèbres registres de la compagnie. Le nôtre est peut-être le premier travail sur le Répertoire de la Comédie-Italienne qui n'utilise pas directement les registres, ce qui a un avantage opérationnel évident : travailler sur ces documents demande de savoir déchiffrer des écritures manuscrites anciennes, en ancien

---

<sup>19</sup> Disponibles ici :

<https://gallica.bnf.fr/services/engine/search/sru?operation=searchRetrieve&version=1.2&query=%28dc.title%20all%20%22D%C3%A9penses%20et%20recettes%20journali%C3%A8res%20de%20la%20Com%C3%A9die-Italienne%22%29&keywords=D%C3%A9penses%20et%20recettes%20journali%C3%A8res%20de%20la%20Com%C3%A9die-Italienne&suggest=1>

<sup>20</sup> <https://bureauumercur.org/public/home/>

<sup>21</sup> Disponible ici : <https://api.nakala.fr/data/11280%2Fd28df67b/d7300be2848911f9e68c934cf26b6c0f55391ab1>

français parfois mélangé à des formes dialectales de l'italien. En revanche, en travaillant à partir des répertoires, nous avons des documents faciles à comprendre et à manipuler. De plus, le répertoire de De Luca nous fournit des métadonnées sur les pièces (genre, nombre d'actes, etc.) que le chercheur tire d'un grand nombre de sources, modernes et contemporaines, et qui enrichissent de manière remarquable la simple liste des représentations que fait Brenner dans son répertoire de 1961. Autrement dit, il s'agit d'un travail de synthèse précieux et complexe, contenant des informations complémentaires à celles de Brenner. Les conséquences épistémologiques de ce choix seront abordées dans la troisième partie.

La Comédie-Italienne, pour sa période d'activité au XVIII<sup>e</sup> siècle, fait l'objet d'un autre projet en humanités numériques, RECITAL<sup>22</sup>, visant à transcrire tout le contenu des registres de l'époque, afin de créer, dans un second temps, une base de données (pour une description du projet voir RUBELLIN et RASCHIA, 2020). La transcription de ces manuscrits modernes est ici faite de manière participative, et peut potentiellement être prise en charge par toute personne intéressée, grâce aux tutoriels et aux multiples vérifications, assurant une bonne fiabilité des données. La précision et l'exhaustivité de ce travail sont en revanche contrebalancées par le long temps nécessaire à son accomplissement. Mais ce n'est pas uniquement la précision et la rapidité du travail qui distingue ces deux projets : les phénomènes étudiés sont différents, car nous traitons les répertoires, alors que Récital traite les registres. Nos sources permettent de connaître les titres, le nombre de spectateur.ices du théâtre, et les recettes de la compagnie, jour par jour (grâce au répertoire de Brenner), mais aussi des informations synthétiques sur les pièces, comme leurs auteur.ices, le genre ou le nombre d'actes (grâce au répertoire de De Luca). Les registres ne donnent aucune caractéristique de la pièce jouée, mais fournissent en revanche plus d'informations sur l'activité du théâtre, comme les objets scéniques et les détails de certaines dépenses.

La période couverte par notre Répertoire se superpose avec un deuxième projet, celui de la base de données de l'Opéra-Comique<sup>23</sup> (Opéra-Comique Database), en conséquence du rachat du bail et du privilège de l'Opéra-comique en 1762. Pour cela, à partir de cette année, nous n'allons pas traiter le détail des opéras-comiques, en renvoyant à Opéra-Comique Database pour les pièces communes.

---

<sup>22</sup> <https://recital.univ-nantes.fr/#/>

<sup>23</sup> <https://www.ntnu.edu/opcom/database>

### 1.2.3 Etudes théâtrales et humanités numériques

De nombreuses bases de données, ou autres outils, pour les études théâtrales sont actuellement présentes sur Internet – en annexe un tableau qui résume leurs adresses et leurs sujets. Le contenu de ces plateformes est varié, les objectifs de recherche sont multiples : certaines valorisent l'accès aux textes (comme Molière 21 ou Naissance de la critique dramatique), d'autres sont plus orientées aux métadonnées et à leur visualisation (par exemple la célèbre base CESAR<sup>24</sup> ou le programme des Registres de la Comédie-Française<sup>25</sup>). Certains de ces projets datent du début de l'informatique personnelle et du premier internet :

« Cette adoption immédiate de la communauté indiqu[e que] le numérique [est] moins une rupture qu'une évolution pour l'histoire des spectacles. Comme l'histoire littéraire et l'histoire du livre, la discipline recourait depuis longtemps à des répertoires papier proches du modèle de la base de données. Lors d'une conférence donnée à Dartmouth en 2018, Jeffrey Ravel a rappelé les nombreuses similarités entre le travail des premiers historiens du théâtre français au XVIIIe siècle – les frères Parfaits – et les pratiques numériques. La nécessité de croiser des sources nombreuses et diverses pour documenter chaque représentation, ainsi que l'influence grandissante des sciences sociales sur la discipline, conduisait les chercheuses et les chercheurs à constituer leurs propres bases de données, sous forme de « fiches » ou de tableaux » (ROUSSILON et SCHUWEY, 2020).

Dans ce florilège de plateformes, une base de données pour le Répertoire de la Comédie-Italienne de Paris n'existait pas encore : cette thèse de doctorat comble donc, tout d'abord, ce manque. De manière plus spécifique, avec une base de données, une grande partie des informations contenues dans les répertoires de Brenner (BRENNER, 1961) et de De Luca (DE LUCA, 2011) peuvent être trouvées rapidement, avec plus de détails, mais peuvent aussi être visualisées dans des graphiques. Les détails du contenu de ces répertoires est fourni dans le premier chapitre, dans la section Données.

La visualisation ou autres manipulations des données sont souvent décrites comme étant une riche source de nouveaux points de vue et de nouvelles questions (c'est le cas dans la plupart des articles des numéros 4 et 5 de la *Revue d'Historiographie du Théâtre*). Cela concerne aussi bien la phase de modélisation que de visualisation, même si la dernière est plus évidente. Dans

---

<sup>24</sup> <https://cesar.huma-num.fr/cesar2/>

<sup>25</sup> <https://www.cfregisters.org/#/>

le contexte des recherches sur les études théâtrales et les humanités numériques, il est remarquable que, depuis 2017, la *Revue d'Historiographie du Théâtre* a publié trois numéros dédiés à la numérisation et aux humanités numériques. Le premier, coordonné par Ioana Galleron, aborde génériquement « Études théâtrales et humanités numériques », avec des articles concernant divers lieux, diverses époques et diverses approches, mentionnant un grand nombre de réussites, mais aussi un échec (GLORIEUX, 2017). Le but de ce numéro, tel que déclaré dans l'introduction par la coordinatrice, est de

« contribuer à l'accélération du mouvement [de numérisation] et à la dissipation des méfiances. Il ne s'attardera pas sur la position autoréflexive et critique que les humanités numériques ont su adopter, montrant qu'elles savent questionner les techniques, les approches et les présupposés qu'elles mobilisent, et qu'elles ont réussi à articuler les promesses du numérique avec les réflexes méthodologiques qui font la valeur des humanités en général : la construction réfléchie de la question de recherche, l'attention au contexte, la conscience critique et la volonté interprétative » (GALLERON, 2017).

Pour cela, elle aborde les avantages de l'utilisation du numérique dans les études théâtrales (accès aux informations, analyses quantitatives, renouvellement des questions de recherche et des interactions interdisciplinaires, fusion d'information de nature différente), ainsi que les limites montrées par les articles (appropriation des outils numériques, standards d'encodage des textes, interopérabilité, conception des interfaces).

Le deuxième numéro, sorti en 2020 et coordonné par Marine Roussillon et Christophe Schuwey, propose de dresser un bilan de l'écriture de l'histoire des spectacles avec des bases de données, avec cette fois une attention à la dimension auto-réflexive :

« Ni célébration, ni condamnation ; ni théorisation excessive, ni collection de cas disparates, il a été conçu comme une tentative de rassembler des chercheurs et des chercheuses autour de questions et d'expériences concrètes [...] Débattre du numérique, c'est en réalité questionner le sens, les méthodes et les objectifs de notre recherche » (ROUSSILON, SCHUWEY, 2020).

Pour cela, les coordinateur.ices ont proposé un questionnaire à six responsables de productions de bases de données pour le théâtre d'Ancien Régime, représentatives de diverses conditions matérielles de production (grandes ou petites équipes, financées ou pas), qui ont écrit leurs articles en suivant cette trame. À partir de ces réponses, ils font deux constats et définissent deux défis : les récits de créations de bases de données témoignent d'un changement d'échelle

dans leurs travaux, ainsi que d'un pouvoir d'« argumentation » de la base de données, opposé à l'idée d'une description neutre de l'objet modélisé. Concernant les défis, tout le monde est confronté à la recherche de financements pour assurer la pérennité des projets, ainsi qu'à la difficulté de conception de l'interface graphique, pour qu'elle soutienne l'« argumentation » sans suggérer une complétude des informations (nécessairement lacunaires par leur caractère historique). Presque toutes les contributions mentionnent aussi des temps très longs pour la production des données et des interfaces.

Le troisième, enfin, publié en 2024 et proposant des contenus multimédias, propose un focus sur le traitement des données concernant l'histoire du théâtre, et sur leur archivage, à partir du colloque de bilan des vingt ans de l'expérience pionnière de la base de données CESAR. Cette base de données, qui a beaucoup inspiré la recherche en études théâtrales et humanités numériques, est l'un des modèles de notre travail sous plusieurs aspects : premièrement pour l'approche chronologique de ce projet, qui est centré sur les métadonnées concernant la programmation des théâtres, malgré l'ajout successif de données iconographiques et de textes ; deuxièmement, notre outil de recherche et pour filtrer les informations s'inspire de celui présent sur la plateforme de CESAR ; et troisièmement pour son caractère collaboratif, qui reflète une certaine approche aux humanités numériques promouvant le partage et la mise en commun. En effet, la base CESAR naît comme une mutualisation des travaux personnels de Barry Russel, David Trott et Jeffrey Ravel, afin d'en faire profiter toute la communauté scientifique (SAKOVSKAIA-PANKEEVA, 2024). L'histoire de cette base de données est exemplaire aussi en ce qui concerne les difficultés à maintenir son existence matérielle : la première création, artisanale et hébergée sur des serveurs commerciaux, n'a pas été accessible pendant plusieurs années pour un manque de maintenance et pour des problèmes de droits ; aujourd'hui la plateforme est hébergée par le consortium Huma-Num<sup>26</sup> et sa maintenance et valorisation est prise en charge à l'Université de Grenoble. Grâce au travail de Marc Douguet (responsable actuellement du projet) et Théo Roulet, la base CESAR s'ouvre au web sémantique dans le cadre du projet Dramabase (DOUGUET et ROULET, 2024).

Le projet Registres de la Comédie-Française représente une autre source de publications sur le théâtre en France au XVIIIe siècle, ainsi qu'une source d'inspiration pour notre base de données. Le volume *Databases, Revenues, & Repertory: The French Stage Online, 1680–1793*, propose de nombreux exemples d'utilisation de la base de données et de l'interface du projet

---

<sup>26</sup> <https://www.huma-num.fr/>

de numérisation des registres de la Comédie-Française. Dans la postface (BIET et al., 2020), Biet, Harvey et Sanjuan commentent la diversité des publications du volume pour montrer « la puissance historique, narrative et inventive de ces archives lorsqu'elles sont mises au contact de l'approche quantitative » (BIET et al., 2020, p. 2). Cette diversité repose sur la diversité des approches des chercheur.euses lorsqu'ils abordent les données : « chacun a pu travailler à partir des datas disponibles en cherchant, à partir de ses propres préoccupations, la manière de les faire fonctionner » (BIET et al., 2020, p.5). Cela leur permet d'affirmer que :

« ces activités témoignent du fait que notre projet de recherche est plus qu'une simple méthode qui permettrait de faciliter la lecture et serait un instrument de recherche au service de traditions d'enquête qui resteraient inchangées ; il est ancré dans les humanités numériques, et plus largement dans la culture numérique, et il transforme notre rapport au savoir et à la médiation du savoir. » (BIET et al., 2020, p. 11).

Cela montre aussi la nécessité d'une documentation avec une explication approfondie sur la signification de la structure et des métadonnées, que ce projet fournit de manière précise et abordable dans la section « Ressources techniques ». Le site de consultation du projet Registres de la Comédie-Française fournit plusieurs autres types de ressources, dont certaines visualisations de données dans la partie « Outils de recherche », qui ont été une source d'inspiration pour les visualisations que nous proposons dans notre site, et pour la manière d'y accéder par thème.

En juin 2023 a eu lieu, à Paris, un colloque de bilan après dix ans du projet Registres de la Comédie-Française<sup>27</sup>. A côté de la présentation de recherches menées à partir de cet outil, un hackathon et des tables rondes ont été le cadre pour discuter des développements futurs, ainsi que des questions liées à l'interopérabilité et à l'écologie. Plusieurs débats se sont articulés autour des différences d'interprétation et de modélisation – même le concept de « personne » a été questionné – montrant que la difficulté à implémenter l'interopérabilité dépend de considérations théoriques, aussi bien que techniques.

---

<sup>27</sup> Le programme est consultable ici : <https://cellf.cnrs.fr/colloque/des-archives-aux-donnees-bilan-et-perspectives-du-programme-rcf/>

## 1.3 Aspects méthodologiques

### 1.3.1 Concepts à désambiguïser

Lors de nombreux échanges transdisciplinaires qui ont eu lieu dans les trois ans de projet doctoral, ont surgit des ambiguïtés relatives à certains termes utilisés, que cette section va expliciter. Afin de comprendre ces termes, nous utilisons le concept préalable de *chaîne de traitement numérique*. Il s'agit de la procédure, de la méthode, suivie dans un projet d'humanités numériques, afin d'obtenir l'objet informatique souhaité. Cela signifie, donc, transformer un objet du monde réel, en un objet manipulable avec un ordinateur. Ce deuxième objet doit donc avoir une forme, ou plus précisément un *format*, compatible avec les outils de traitement de l'information sur des machines : « un format numérique est fondamentalement une spécification de la façon dont l'information est organisée » (WITZ et PORTE, 2025). Cette spécification définit la manière dont l'information peut être manipulée, et permet donc de choisir ou de créer des logiciels pour cela.

#### 1.3.1.1 Document ou source ?

Dans cette recherche, nous parlons de *sources* pour faire référence à des documents historiques, qui contiennent les indices des informations que nous recherchons. La source, comme en histoire, doit pouvoir avoir la valeur de témoin, il s'agit donc généralement d'objets matériels et d'époque moderne si l'on étudie l'Ancien Régime. Un document est un revanche un objet générique, contenant de l'information qui n'a pas nécessairement la valeur de témoin. Dans une chaîne de traitement en humanités numériques, le document constitue le point de départ du travail. Il peut, par exemple, correspondre à une source, mais aussi à d'autres objets élaborés, comme des ouvrages critiques, des éditions, ou des répertoires, voire des images d'artefacts.

Dans notre cas, les sources sont les Registres de la Comédie-Italienne, auxquels on accède par leur forme numérisée par la BnF. Les documents, en revanche, sont les deux répertoires, celui de Brenner de 1961 et celui de De Luca, qui ont été le réel point de départ de notre chaîne de traitement. Dans le travail de De Luca, au contraire, les documents et les sources peuvent correspondre aux mêmes objets – en l'occurrence lorsqu'il s'agit des Registres, des revues et des éditions de l'époque.

### 1.3.1.2 De l'information à la donnée

La distinction entre *information* et *donnée* est un concept crucial pour les humanités numériques : à l'intérieur d'une chaîne de traitement, on transforme une *information* en *donnée*. Autrement dit, l'information représente l'objet ou la partie de l'objet à analyser ; la donnée correspond à la même information, transformée dans un format *machine-readable*, c'est-à-dire un format qui peut être manipulé avec un ordinateur. L'information est donc un objet multi-dimensionnel du monde réel, alors que la donnée correspondante saisit une certaine dimension, un certain caractère, de cet objet, et le fixe dans un format numérique (CIULA et MARRAS, 2023). La transformation d'une information en donnée se fait à travers une activité *d'encodage*, c'est-à-dire le passage d'un code à un autre code, d'un format à un autre format.

La transformation des informations en données est donc le cœur d'un projet d'humanités numériques, car il s'agit du moment crucial où l'on attribue une signification non ambiguë aux objets modélisés. Autrement dit, il s'agit du moment où l'on fixe les aspects interprétatifs qui nous intéressent, et où on leur donne la forme qui sera effectivement utilisées lors des analyses.

Cette conception de l'information est profondément différente, voire opposée, à la signification de ce terme dans le domaine informatique, comme m'ont fait remarquer les membres de l'équipe Ex)situ du LISN<sup>28</sup> lors d'un séminaire où j'avais été invitée en avril 2025. En effet, en informatique, l'*information* est l'unité de transmission d'un message ; dans ce sens, cela correspond à ce que nous qualifions de *donnée* dans les humanités numériques.

### 1.3.1.3 Répertoire

Le concept de répertoire est tout-à-fait central pour notre travail, mais sa signification est spécifique à chaque domaine. Dans les humanités de manière générale, le mot répertoire désigne un ensemble ordonné et homogène d'informations concernant un domaine spécifique. Les documents de Brenner (BRENNER, 1961) et de De Luca (DE LUCA, 2011) sont par exemple des répertoires. Il s'agit d'un type de catalogue où sont données certaines informations décrivant les objets, et qui permet souvent de les trouver (dans une bibliothèque ou des archives, par exemple). En informatique, un répertoire est un genre particulier de fichier, permettant d'accéder à d'autres fichiers, qu'il contient. C'est un synonyme de *dossier*. Il s'agit donc d'informations structurés, facilitant la recherche de fichiers. Cette signification est ainsi proche

---

<sup>28</sup> <https://www.lisn.upsaclay.fr/recherche/departements-et-equipes/interaction-avec-lhumain/ex-situ/>

du sens de ce mot en littérature, faisant référence à un document où sont présentes des informations synthétiques, organisées de manière à faciliter la recherche d'informations.

Dans le théâtre, un répertoire est un ensemble de pièces regroupées suivant un certain critère – par exemple l'ensemble des pièces jouées par une compagnie, ou à une certaine période, ou ayant un certain genre. Dans les théâtres parisiens du XVIIIe siècle, cela prend un sens spécifique : « au XVIIIe siècle on ne joue pas une œuvre mais un répertoire. [...] En un soir donné, seul l'Opéra ne présente qu'une seule œuvre – deux, trois ou quatre dans la semaine. Partout ailleurs, c'est au moins deux et fréquemment, aux foires et aux boulevards, quatre ou cinq pièces qui se suivent. Autant d'autres le lendemain » (ROUGEMONT, 2001, p. 54). Chaque théâtre a son répertoire, et sa spécialité, le public y va chercher un certain type de jeu et un certain type de pièces. Concernant la Comédie-Française, dans un article problématisant la signification du mot *répertoire*, Derek Miller donne cette définition : « le répertoire correspond à ce que, selon les registres, la troupe a joué, ou à ce qu'elle aurait pu jouer en fonction des textes dramatiques en sa possession » (MILLER, 2020). Dans ce sens, cette thèse de doctorat aborde l'objet *Répertoire* de la Comédie-Italienne, plus que son contenu. Dans ce texte, nous utilisons ce terme en lui attribuant des significations différentes, lorsque l'on parle des *répertoires* de De Luca et de Brenner, ou lorsque l'on parle du *Répertoire* de la Comédie-Italienne de Paris.

### 1.3.2 L'équipe

La réalisation de cette recherche s'est fondée sur un travail collectif, d'équipe : afin de garantir une expertise scientifique dans tous les domaines, de la Comédie-Italienne de Paris à la gestion de base de données. Nous avons voulu valoriser la complémentarité des rôles et le dialogue disciplinaire. Donc tout d'abord les membres relèvent de plusieurs disciplines : une équipe de recherche en études italiennes, une équipe de recherche en études théâtrales, et deux équipes d'humanités numériques. La direction couvre ces trois domaines. Andrea Fabiano et Emanuele De Luca ont représenté le regard expert sur la Comédie-Italienne ; concrètement, c'est en dialogue avec eux que j'ai conçu le modèle abstrait pour les données, et l'interface de consultation, afin de respecter le contexte historique, mais aussi afin de pouvoir étudier les phénomènes qui les intéressaient. Glenn Roe et l'équipe Obtic ont soutenu et aidé à la conception de la base de données et au premier déploiement de l'interface de saisie sur leurs serveurs. C'est à travers l'équipe Obtic que j'ai eu accès à Paris à des locaux pour travailler, à la BnF et au SCAI. Les deux premières stagiaires, Chloé Petridis et Chiara Protani ont aussi été

recrutés au sein de cette équipe. La réalisation des tâches les plus techniques du point de vue informatique a été confié à l'ingénieure informatique Camilla Cantiani, recrutée comme prestataire. La deuxième équipe d'humanités numériques, le LabRadar, intervient dans le déploiement de l'interface de consultation et dans sa maintenance. Ci-dessous (Figure 3) un schéma de l'équipe.

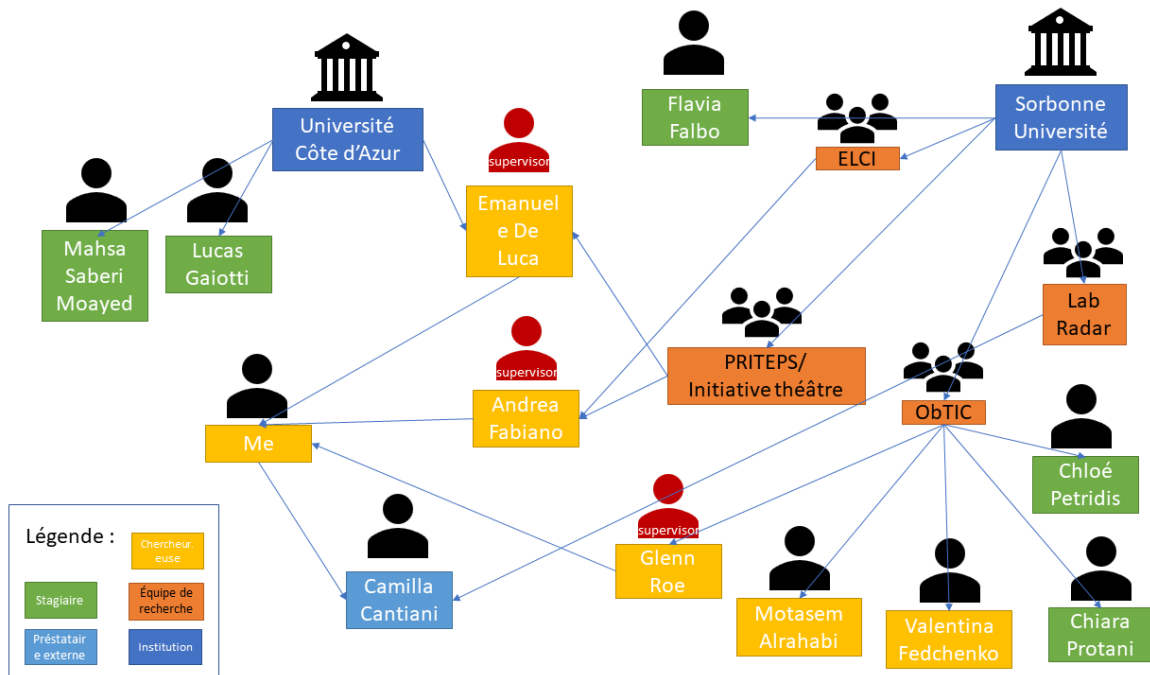


Figure 3 : schéma de l'équipe de recherche.

C'est à travers le dialogue entre les membres de l'équipe que les disciplines ont pu contribuer de manière spécifique à la réalisation du projet (MOUZA et al., 2023). Mon rôle, concrètement, a été celui de coordonner le travail, ce qui a demandé le choix d'une méthode et des outils pour la réaliser. Mais mon rôle a été aussi d'atteindre, avec les outils et méthodes numériques, des objectifs de recherche en lettres, respectant les finalités de recherche des membres de l'équipe spécialistes de Comédie-Italienne. Dans ce sens, je me retrouve dans une position privilégiée pour observer la transformation des informations en données, et pour observer le rôle de l'interprétation dans ce processus. La méthode a été choisie dans le but de mettre en lumière l'activité interprétative humaine dans la création des données. Cela ne répond pas uniquement à mes objectifs de recherche spécifiques, mais permet aussi de faire surgir les biais disciplinaires inhérents à tout travail comme le nôtre.

Travailler sur le *Répertoire* à partir de *répertoires* est un contexte particulièrement intéressant du point de vue des humanités numériques, puisqu'il permet d'observer une continuité des pratiques de recherche, avec une acquisition progressive d'outils et de méthodes informatiques. En effet, le travail fait par Brenner (BRENNER, 61) et De Luca (DE LUCA, 2011) peut être défini comme un travail informatique, dans le sens où les deux chercheurs ont systématisé les informations des registres dans une forme standardisée, facilitant leur exploitation et lui donnant un caractère automatique.

### 1.3.3 Procédure

La méthode suivie dans notre recherche est schématisée dans la Figure 4.

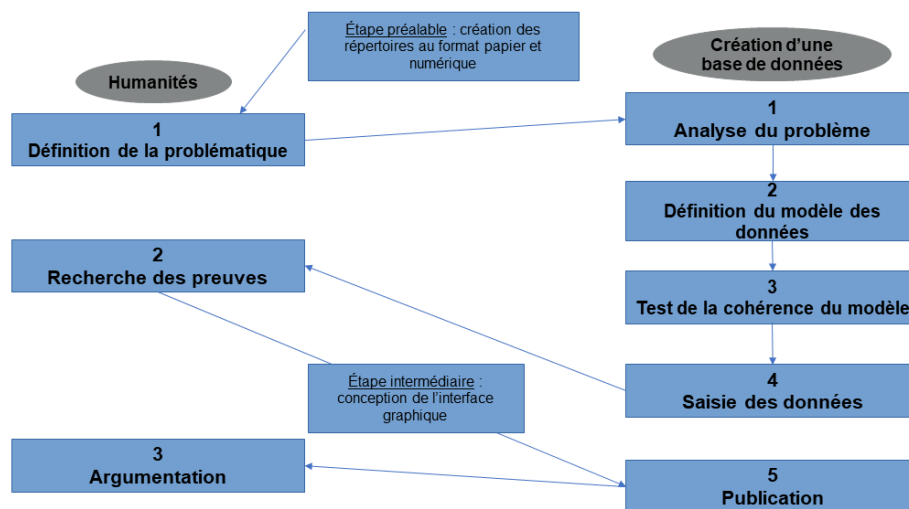


Figure 4 : Schéma de la méthode de construction de notre base de données.

Le procédé doit commencer et finir dans le champ littéraire, et passer aussi par toutes les étapes de création d'une base de données en informatique. Ainsi, afin de montrer et de valoriser l'aspect interprétatif du travail, le processus commence par la définition des résultats que l'on souhaite obtenir par la base de données, du point de vue des études théâtrales. Cela a été fait en créant un schéma des résultats souhaités, dont la proposition initiale a été faite à partir d'un document similaire, créé par Motasem Alrahabi pour le projet Opéra-Comique Database, et ensuite discutée avec Andrea Fabiano et Emanuele De Luca pour mieux l'adapter au Répertoire de la Comédie-Italienne. La liste des résultats souhaités à la fin du projet est donc le premier document que nous avons créé.

À partir de ce document, nous pouvons passer à l'analyse du problème du point de vue informatique, première étape dans la création d'une base de données. En plus de la liste des résultats souhaités, cette étape prévoit l'analyse de la forme et du format des informations et des données à traiter. Concrètement, à la fin de cette phase, nous avons créé des documents où sont recensées toutes les sources susceptibles d'être utiles, ainsi qu'une description schématique du contenu et de la forme des répertoires existants, en particulier des deux répertoires de Brenner et de De Luca.

Analyser le problème est considéré fondamental pour la définition du modèle logique des données, c'est-à-dire de la structure abstraite constituant la phase intermédiaire pour l'implémentation informatique. L'activité de définition du modèle se trouve à l'interface entre l'informatique et (ici) les études littéraires. Une explication détaillée du modèle est donnée dans le premier chapitre (section Présentation de la structure formelle des données).

Une fois le modèle défini de manière abstraite, il faut tester sa cohérence avant de l'implémenter (pour un bilan du travail de modélisation voir DESIDERI, 2024). Vérifier la cohérence du modèle, consiste en deux activités : d'un côté, vérifier que la structure permette d'obtenir les résultats souhaités, de l'autre côté vérifier que les informations ne soient pas redondantes et que la structure soit optimisée pour la recherche d'informations. Concrètement, la première vérification a été faite en définissant des fonctions (abstraites) qui éclaircissent les relations entre les entités ; par exemple, nous avons défini la relation entre une pièce et sa première, ou la relation entre des soirées et les genres des pièces jouées lors de ces soirées. Cela nous permet de vérifier qu'il est possible d'accéder aux informations nécessaires pour obtenir les résultats souhaités. La deuxième vérification, en revanche, se base sur des processus de normalisation, qui consistent en des transformations successives de la structure, afin d'éliminer les redondances et de la rendre donc plus efficace pour la recherche, et plus sûre en termes de mise à jour des données. Pour la réalisation de ce deuxième volet du test de cohérence, j'ai été aidée par l'ingénieure informatique Camilla Cantiani. Cette normalisation a mené à des changements importants dans la structure des données. L'enjeu principal a été de modifier le diagramme sans générer des incohérences d'interprétation, mais grâce à une analyse fine des besoins et des contraintes, le résultat final les respecte davantage.

La structure ayant été vérifiée et optimisée, nous pouvons l'implémenter afin de saisir les données. Pour l'implémentation et pour la création d'une interface de saisie, le travail de Cantiani a encore une fois été fondamental : utilisant la librairie Django de Python, elle m'a

fourni un code qui crée une base de données SQLite et qui permet d’y insérer les informations. À partir de sa première proposition, nous avons ensuite fait ensemble quelques modifications pour faciliter davantage le travail humain de saisie, et pour corriger les quelques erreurs dérivées d’un manque de clarté du modèle. La saisie des 450 pages du répertoire de Brenner a été assurée en grande partie par moi-même et par des stagiaires.

Avant de passer à celle qui est la dernière étape dans la création d’une base de données en informatique – c’est-à-dire la publication des données pour en permettre l’accès – nous pouvons déjà utiliser les données dans notre recherche des preuves. Concrètement, cela a été fait en générant des visualisations des données, avec le langage Python dans des notebooks Jupyter, qui sont accessibles dans un repository FramoGit<sup>29</sup>. Le travail sur les données nécessaire à produire ces visualisations, a contribué à la conception de l’interface graphique.

L’argumentation finale est ici représentée par le texte de cette thèse, mais aussi par le contenu et la forme de l’interface de consultation. En effet, celle-ci a été conçue suivant des objectifs, que sa réalisation finale reflète ; par exemple, l’approche chronologique à l’étude du Répertoire se reflète dans le grand nombre de visualisations basées sur des distributions temporelles. Aussi, parmi tous les choix possibles, nous avons défini trois thématiques pour accéder aux données : pièces, auteurs et premières. Si les deux premières sont universelles dans les études théâtrales, la troisième (celle sur les premières) se situe en continuité avec le répertoire de De Luca, qui est centré sur l’ajout de pièces au Répertoire.

---

<sup>29</sup> Les logiciels de type git permettent l’écriture collaborative, le partage et la création de versions d’un code. Un repository dans une plateforme de ce type représente un projet, et correspond à l’ensemble des fichiers qui le constituent.

Le repository associé à la base de données de la Comédie-Italienne de Paris est accessible ici <https://framagit.org/adadesideri/bdd-comit18>

## **2 Partie 1 : construction de la base de données**

Ce premier chapitre contient la description de la démarche suivie afin de construire, concrètement, la base de données. Cela concerne des activités de nature différente : analyse des besoins, compréhension du contexte historique, formalisation logique-mathématique, saisie des données, programmation informatique... Toutes les compétences qui sont réunies au sein de l'équipe (pour les détails sur l'équipe voir la section Aspects méthodologiques), sont ici mobilisées : c'est en les mettant en relation les unes avec les autres que le projet de numérisation peut respecter aussi bien les contraintes de la recherche en lettres que celles du développement informatique. C'est aussi dans ce dialogue que chaque membre peut valoriser son point de vue et ses intérêts au sein de l'équipe, formant un terrain fertile pour de nouveaux questionnements et de nouvelles recherches dans les domaines concernés (ROUSSILON et SCHUWEY, 2020 ; RUBELLIN et RASCHIA, 2020).

Dans ce chapitre, seront présentés les objectifs précis de la création de la base de données, afin de décrire les données qui ont constitué le cœur de la numérisation, et la chaîne de traitement des données. Poser ce cadre permettra de montrer les conséquences de ces objectifs sur les résultats de la numérisation, qui sont la structure des données et l'interface graphique de consultation du projet. Le chapitre aborde aussi des considérations sur l'interopérabilité, aspect incontournable de tout projet académique de numérisation.

### **2.1 Préparer la numérisation**

#### **2.1.1 Objectifs de la numérisation**

Les objectifs de la numérisation seront classifiés en trois catégories, présentées dans cet ordre :

1. Les objectifs disciplinaires, qui sont les plus proches de la Comédie-Italienne de Paris et concernent de manière directe les recherches faites dans ce domaine par les spécialistes de notre équipe ;
2. Les objectifs organisationnels, qui font référence à l'organisation du travail d'équipe et au déroulement de la recherche ;
3. Les objectifs sociétaux, qui contextualisent la recherche dans les problématiques actuelles d'accès à la connaissance et de développement des sciences et techniques.

Expliciter les objectifs de la recherche permettra de présenter les conséquences de ceux-ci sur la modélisation et d'évaluer, par la suite, en quelle mesure ils ont été atteints à la fin du projet. Mais cela est fait aussi avec l'objectif de rendre le plus transparent possible le cadre de la création technologique de notre base de données, afin d'explicitier le lien entre interprétation et modélisation logique : le développement technologique n'est pas conçu comme une activité neutre, mais au contraire comme une activité profondément liée au contexte social et économique (JARRIGE, 2022 ; BOOKCHIN, 2019, 2020). A notre échelle, comprendre ce lien revient à comprendre les relations entre les objectifs de la recherche et la forme des résultats que nous avons générés.

### **2.1.1.1 Objectifs disciplinaires**

Au sein de l'équipe, le Répertoire de la Comédie-Italienne de Paris est étudié avec une attention particulière au processus d'hybridation culturelle entre la tradition théâtrale française et italienne (FABIANO et DE LUCA, 2023). La base de données doit donc participer à cette étude, et le numérique peut pour cela jouer un rôle important. En effet, une base de données permet de lier des informations variées, multimédias, et d'analyser simultanément plusieurs variables. Une des attentes de la numérisation est donc de pouvoir modéliser et visualiser la complexité culturelle du Répertoire, afin de mettre en relation des informations de nature différente. Or, modéliser des contenus hybrides reste un défi en humanités numériques (FIORMONTE et DEL RIO RIANDE, 2023 ; VALLARANO et DESIDERI, 2024a), et nous tâcherons donc d'aborder ce sujet de manière critique : si le numérique a des avantages évidents en termes de puissance de calcul par rapport à un être humain, le traitement d'informations ambiguës demande plus de précautions (MOUZA et al., 2023 ; SAINT-RAYMOND, 2023).

Plus largement, s'insérant dans une longue tradition de recherche sur la Comédie-Italienne de Paris, le projet de doctorat constitue une étape organique dans l'étude du Répertoire de la compagnie. L'un des objectifs est de décrire, jour par jour, ce qu'elle proposait au public : dans ce sens, la recherche est focalisée spécifiquement sur la dimension spectaculaire des pièces du Répertoire, autrement dit à la valeur qu'elles ont en tant que spectacle, donné devant un certain public (ROUGEMONT, 2001). Comme il a été précisé dans l'état de l'art, les recherches qui représentent le fondement de cette étude mobilisent la notion, ainsi que l'objet *répertoire*. Dans cet écosystème, la nouvelle base de données veut permettre de réaliser des études sur la chronologie de la programmation : il s'agit par exemple d'observer quels choix sont faits jour par jour, ou sur une certaine période ; ou quels genres de pièces étaient le plus représentés à

un certain moment, et en relation à d'autres moments. Pouvoir obtenir des informations sur l'évolution chronologique de la programmation a des conséquences sur la manière dont les données sont choisies, modélisées et traitées.

A l'objectif d'étudier la programmation du théâtre de manière chronologique, s'ajoute la volonté de le faire de manière quantitative sur tout un siècle. En plus de l'intérêt d'étudier le XVIIIe siècle du point de vue historique et théâtral en France (ROUGEMONT, 2001), le *siècle* constitue une unité suffisamment large pour saisir des tendances, mais assez restreinte pour être traitée de manière intelligible – comme écrit Franco Moretti « un beau numéro rond, suffisamment étendu pour suggérer une échelle historique – mais pas assez ample à devenir incontrôlable »<sup>30</sup> (MORETTI et SOBCHUK, 2022, p. 94). Le numérique intervient ici de manière fondamentale pour ses avantages en termes de calcul : même lorsque les informations sont déjà partiellement structurées, et même si cela n'est pas strictement impossible, faire des calculs sur plusieurs variables, à la main, à partir de nombreuses données sur cent ans, reste inenvisageable en termes de temps et de risque d'erreurs.

Le numérique intervient aussi dans la volonté de se situer dans l'écosystème des projets de bases de données pour le théâtre. La possibilité de comparer, voire d'unifier, des jeux de données proches permettrait d'élargir davantage les informations accessibles sur le théâtre d'Ancien Régime. Or chaque projet a ses objectifs et ses moyens propres, ce qui génère une grande variété dans les solutions techniques adoptées. Actuellement, le projet Transnational Stages est mis en place afin de rendre réellement interopérables les données tirées du projet Registres de la Comédie-Française et de données analogues sur le Théâtre Royal de Copenhague, ainsi que d'autres projets similaires : Theater in Saint-Domingue 1764-1791, Theatromics, Onstage, et le nôtre sur la Comédie-Italienne. Deux réunions, ainsi que des prototypes de visualisations pour les données danoises, françaises et néerlandaises ont été réalisées à ce jour (août 2025). Une section de la sous-partie sur l'interopérabilité (Le projet Transnational Stages) sera donc consacrée à la mise en relation de notre base de données avec les autres bases de données participant à ce projet.

---

<sup>30</sup> «Un bel numero tondo, abbastanza esteso da suggerire una scala storica ampia – ma non così ampia da diventare incontrollabile.»

### 2.1.1.2 Objectifs organisationnels

Cette recherche prend aussi son sens dans une logique d'équipe, où l'on cherche à valoriser la continuité des pratiques et la complémentarité des connaissances : le but n'est pas de se poser en rupture avec des études déjà accomplies sur la compagnie par des spécialistes de théâtre, mais d'y intégrer des compétences et connaissances informatiques, pour atteindre certains objectifs spécifiques. L'expertise informatique et les potentialités qu'elle engendre ne se substituent pas aux connaissances historiques et littéraires ; elles peuvent les accompagner dans leur démarche. Réciproquement, la recherche informatique et ses outils évoluent dans ce contexte, où les applications demandent de manipuler un degré d'ambiguïté qui est habituellement éliminé dans les sciences formelles (MOUZA et al., 2023). En effet, la formalisation logique est fondée sur la désambiguïsation des informations, pour qu'elles soient traitées de manière déterministe. Dans les humanités numériques, les objets manipulés prennent leur valeur dans des processus interprétatifs de construction de la connaissance, au sein desquels l'ambiguïté peut jouer un rôle créatif.

Une forme de continuité est visée aussi dans la relation avec les recherches précédentes pour plusieurs raisons : partir de recherches déjà réalisées garantit une certaine fiabilité des données, ainsi qu'une cohérence méthodologique dans l'écosystème des recherches sur la Comédie-Italienne de Paris et sur le théâtre au XVIIIe siècle. En effet, construire des répertoires représente une pratique courante dans les études théâtrales et littéraires, qui n'a pas vocation à transmettre toute la complexité matérielle ni informationnelle des sources, mais a l'avantage de sélectionner les informations de manière experte et cohérente avec des objectifs de recherche. Ainsi, si suivre les choix qui ont été faits par d'autres chercheurs.euses fait courir le risque de répéter et amplifier leurs erreurs, voire de générer un biais d'auto-confirmation dans la nouvelle recherche, cela garantit en revanche une continuité méthodologiques et épistémologique aux résultats. En d'autres termes, utiliser des répertoires existants assure que les nouveaux résultats soient *comparables* avec des recherches précédentes, faites à partir de ces mêmes documents. La question de la comparabilité des données a d'ailleurs été longuement abordée lors de la deuxième rencontre (qui a eu lieu en fin avril 2025) du projet Transnational Stages : cette comparabilité est actuellement réalisable à travers les diverses plateformes avec un regard expert sur les données, mais pour être implémentée il est nécessaire d'évaluer de nombreux facteurs (de l'exhaustivité des sources, à la prise en compte du contexte national et historique). Le fait de sélectionner un ensemble cohérent d'informations sur la Comédie-Italienne de Paris

distingue une nouvelle fois notre démarche du projet RECITAL, qui vise à la création d'une base de données à partir des registres du XVIII<sup>e</sup> siècle et de toutes les informations qu'ils contiennent. Pour nous, traiter directement les sources historiques ne fait pas partie des objectifs fondamentaux. Au contraire, il est avantageux de partir de documents déjà élaborés, où les informations contenues dans les sources ont déjà été réorganisés à des fins de recherche.

La continuité des pratiques répond aussi à une volonté de valorisation du travail de chaque membre de l'équipe, car elle peut favoriser la réalisation créative des tâches, en facilitant la prise en main par chacun.e du travail qui lui est confié. Autrement dit, si le travail en contexte numérique est similaire à des activités que les chercheurs.euses accomplissent déjà dans leurs recherches, la prise en main des nouvelles activités informatiques sera facilitée (BOSCHETTI et DEL GROSSO, 2020). Par ailleurs, afin que tous les membres puissent gérer consciemment et de manière créative leur tâche, une condition préalable est de comprendre comment cela s'insère dans l'ensemble de la démarche, quels sont les enjeux et les objectifs du projet.

### **2.1.1.3 Objectifs sociétaux**

Outre des objectifs pertinents au sein du monde de la recherche, ce travail prend en compte le contexte sociétal de production des savoirs, avec une attention particulière à l'ouverture de la science et au respect écologique, qui constituent des problématiques majeures pour le développement des sociétés actuellement.

La science ouverte (*open-science*) est un mouvement international basé sur l'idée que le partage des connaissances améliore le développement scientifique : si tout le monde a accès aux publications scientifiques, non seulement leur diffusion sera élargie et leurs auteur.ices valorisé.es, mais elles pourront, surtout, être davantage vérifiées et développées (CNRS, 2016). Il s'agit aussi d'une stratégie de préservation : favoriser la copie décentralise le stockage et diminue ainsi les chances que le matériel se perde ou soit détruit de manière malveillante.

Lorsqu'une recherche présente des résultats basés sur un jeu de données, ces résultats peuvent être vérifiés en reproduisant la méthode de recherche sur les mêmes données. Afin de garantir la reproductibilité des résultats d'une telle publication, il est donc nécessaire de fournir les données, dans un format analogue à celui utilisé dans la recherche, ainsi que le code informatique ayant permis leur manipulation. Par conséquent, il est préférable de faciliter l'accès aux données lorsque leur publication n'a pas de restrictions légales ou éthiques (cela

concerne le plus souvent les sciences sociales, où les données peuvent porter sur des personnes vivantes, dont il est important de préserver la vie privée). Faciliter l'accès aux données, ne signifie pas seulement rendre accessibles les codes et les fichiers, mais aussi expliquer leur signification, le contexte et la méthode de leur production (WILKINSON et al., 2016). Pour cela, il est recommandé de suivre les principes FAIR, en déposant les données dans des entrepôts institutionnels, et en les décrivant avec des formats de métadonnées partagés, donc interopérables (MESRI, 2022). Le consortium Huma-Num fournit par exemple une infrastructure pour le partage des données en humanités numériques.

La page du site du gouvernement français qui présente le Deuxième plan national pour la science ouverte, donne la définition suivante :

« [La science ouverte] c'est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique.

Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et aux données, aux codes sources et aux méthodes de la recherche.

Elle permet à la recherche financée sur fonds publics de conserver la maîtrise des résultats qu'elle produit.

Elle construit un écosystème dans lequel la science est plus transparente, plus solidement étayée et reproductible, plus efficace et cumulative.

Elle induit une démocratisation de l'accès aux savoirs, utile à l'enseignement, à la formation, à l'économie, aux politiques publiques, aux citoyens et à la société dans son ensemble.

Elle constitue enfin un levier pour l'intégrité scientifique et favorise la confiance des citoyens dans la science. »<sup>31</sup>

En France, le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation a déjà réalisé un premier Plan national pour la science ouverte, en 2018-2021, ayant obtenu la création d'infrastructures qui facilitent et valorisent le développement de la science ouverte, comme les plateformes HAL et Research Data Gouv. Le deuxième plan (2021-2024) a eu comme objectif spécifique celui de « ouvrir l'ensemble du processus de la recherche, les publications bien sûr mais aussi les données et les codes sources »<sup>32</sup>, avec deux axes d'actions (sur quatre) focalisés sur ces aspects. Les Plans nationaux, mais aussi, par exemple, le CNRS

---

<sup>31</sup> <https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-de-la-science-ouverte-en-48525> consulté le 18 juillet 2025.

<sup>32</sup> <https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-de-la-science-ouverte-en-48525> consulté le 18 juillet 2025.

(CNRS, 2016), mettent en avant le rôle du numérique, comme levier dans ce mouvement d'ouverture de la science. Notre projet de base de données s'inscrit entièrement dans ce contexte, ayant comme résultat un site internet et des informations en format numérique, et étant financé sur fonds publics.

En ce qui concerne l'écologie, dans les dernières années, de nombreux travaux ont montré que le numérique a un impact écologique négatif sur la planète (pour un bilan récent des connaissances à ce sujet voir ROUSSILHE et al., 2023), impact qui est dû à la globalité du processus de construction, distribution et fin de vie des appareils, ainsi qu'à la consommation énergétique des terminaux et des datacenters (ou « méga-ordinateurs bétonnés en surchauffe », comme propose de les appeler la Quadrature du net suivant la suggestion du collectif marseillais des Gammars<sup>33</sup>). Si l'on entend le mot *écologie* au sens large, intégrant des préoccupations sociales à propos des conditions de vie des êtres humains (BOOKCHIN, 2020), les risques sociétaux engendrés par le numérique ne peuvent pas être ignorés, du phénomène de *digital divide* (VAN DIJK, 2017 ; HARGITTAI, 2003), à la manipulation de masse lors d'élections ou autres événements politiques majeurs (CECCOBELLI, 2017 ; CEPERNICH, 2017), aux changements qu'il induit dans le monde du travail (CASILLI, 2019 ; BERNARD, 2023) et de l'éducation (CITTON, 2014 ; EMIYA et al., 2023 ; RIPOLL, 2025). Ainsi, lorsque l'on travaille avec l'informatique, il est aujourd'hui tout-à-fait pertinent de se pencher sur ces questions, afin de s'assurer qu'un projet académique ne participe pas à la destruction écologique et à l'isolement social auxquels le numérique nous expose.

L'accessibilité de la science et l'écologie, telles qu'on les a abordées dans ce projet, participent d'une même vision de la technologie, qui est comprise au même temps comme fruit d'une certaine organisation sociale, et comme tendant à reproduire le système qui l'a générée (BOOKCHIN, 2019 ; GALLOWAY, 2023 ; JARRIGE, 2015, 2022). Autrement dit, une technologie n'est jamais neutre dans la mesure où elle reflète la société qui l'a générée, et elle a un impact car elle s'insère dans des dynamiques sociétales. Mais cela signifie aussi qu'il est possible de créer des technologies éthiquement vertueuses par rapport à la vision de la connaissance, voire du monde (BOOKCHIN, 1974), qu'une recherche souhaite transmettre. Une attention particulière est donc portée aux conséquences de nos choix lors du

---

<sup>33</sup> <https://www.laquadrature.net/2024/11/20/accaparement-du-territoire-par-les-infrastructures-du-numerique/> consulté le 19 août 2025.

développement d'une technologie, en termes écologiques aussi bien que d'accessibilité à la science.

Or, cette volonté a rencontré des obstacles au sein de l'équipe et n'a pas pu entièrement aboutir au moment de la fin de mon projet de doctorat. En effet, les trois directeurs, Andrea Fabiano, Emanuele De Luca et Glenn Roe se sont opposés à la mise à disposition libre d'un fichier contenant les données brutes, défendant une accessibilité de ce fichier sous demande, afin de protéger le travail, et en attendant d'avoir corrigé les données, tout en leur attribuant une licence Creative Commons. La publication de nos données ne pose pas de questionnement éthique, puisque les informations sont toutes repérables sur internet par ailleurs, et qu'elles concernent uniquement des événements passés et publics. En revanche, une partie des données est tirée d'un document de travail d'Emanuele De Luca, et ne peut donc pas être publiée sans son accord. Le choix de contraindre l'accès aux données par une demande qui doit être traité par un être humain, constitue un choix qui a été source de désaccord dans l'équipe et qui a été pris par majorité. Néanmoins, le résultat de la recherche reste partiellement accessible via l'interface de consultation, dans un format interactif pour l'utilisateur.ice.

Concernant la méthode et la procédure de travail, deux supports appuient la reproductibilité de la recherche : un carnet hypothèse lié au projet de thèse<sup>34</sup>, et un repository framaGit<sup>35</sup>. Dans le carnet est expliqué, étape par étape, toute la chaîne de traitement des informations, avec les détails qui permettent de l'adapter à d'autres contextes. Le repository framaGit héberge le code qui a été utilisé pour faire les expériences de visualisations, et pour d'autres manipulations des données qui ont été utiles à la création du site ou à l'écriture de la thèse. Nous fournissons aussi des explications de la structure des données (dans le carnet hypothèse, dans le repository framaGit, ainsi que dans l'interface de consultation), sans quoi les données brutes seraient difficilement exploitables. Le code utilisé pour les interfaces de saisie et de consultation est également en accès libre sur GitHub.

---

<sup>34</sup> <https://repcomit18.hypotheses.org/>

<sup>35</sup> <https://framagit.org/adadesideri/bdd-comit18>

## 2.1.2 Méthodologie de la numérisation

### 2.1.2.1 Données

A la lumière de toutes ces considérations, deux documents permettent de couvrir la plupart des besoins du projet, dans le but de créer la base de données : il s'agit du répertoire de Brenner de 1961 (BRENNER, 1961), résumant la plupart du contenu des Registres du XVIII<sup>e</sup> siècle qui nous intéresse (dates, titres, dépenses et nombre de spectateurs) ; et du répertoire d'Emanuele De Luca de 2011 (DE LUCA, 2011), qui synthétise plusieurs informations décrivant les pièces ajoutées au Répertoire de la Comédie-Italienne de Paris entre 1716 et 1762. Ces deux répertoires sont *informatisés* dans la signification première du terme : il s'agit en effet de documents qui facilitent l'automatisation de la recherche d'information, par leur forme schématique et par leur contenu homogène. Dans ce sens, chaque répertoire constitue une étape intermédiaire entre les sources historiques et la base de données, car il réorganise les informations des sources dans le but de faciliter la recherche et l'exploitation des informations. Néanmoins, les deux répertoires ont des contenus et des formats différents, dépendant du but du travail des deux chercheurs qui les ont respectivement créés, mais aussi des outils informatiques accessibles à une certaine époque (un résumé dans le tableau ci-dessous).

Répertoire	Année	Format	Objectifs	Actions possibles
<i>The Théâtre Italien. Its Repertory, 1716-1793</i>	1961	Papier dactylographié	Étude de l'économie du théâtre.	Recherche d'informations de manière chronologique.
<i>Le répertoire de la Comédie-Italienne de Paris (1716-1762)</i>	2011	PDF	Étude de la programmation du théâtre et du mécanisme d'insertion de pièces dans le Répertoire.	Recherche d'informations de manière chronologique ; recherche d'une suite de symboles dans le texte entier.
<i>Le répertoire de la Comédie-Italienne de Paris (1716-1762)</i>	2011	XSL (Excel)	Étude de la programmation du théâtre et du mécanisme d'insertion de pièces dans le Répertoire.	Recherche d'informations de manière chronologique ; recherche d'une suite de symboles dans le texte entier ; recherche d'une suite de symboles dans un champ précis ; filtrer, compter et afficher les données.

Malgré leur caractère *informatisé*, ces répertoires ne contiennent pas nécessairement des *données* à proprement parler, car uniquement le fichier Excel peut être entièrement manipulé avec des outils informatiques. Néanmoins, c'est ce travail informatique, fait par Brenner et De Luca, qui a permis la création de la base de données et l'accomplissement de cette nouvelle

étape dans l'étude du Répertoire : en combinant les deux répertoires nous pouvons connaître, jour par jour, quelles pièces ont été jouées et quelles étaient leurs caractéristiques. Les outils et méthodes informatiques permettent en plus d'analyser et de visualiser l'ensemble de ces informations, dans la limite des registres existants et des sources explorées.

Le contenu des deux répertoires est présenté de manière détaillée ci-dessous :

1. Le répertoire créé par Emanuele De Luca en 2011 contient les premières qui ont eu lieu entre 1716 et 1762, par ordre chronologique. Ce document est disponible sur internet en format PDF (Figure 5), et j'ai pu avoir le document de travail d'Emanuele De Luca en format Excel créé entre 2003 et 2008 (Figure 6), pour automatiser en partie l'acquisition des informations. Ce répertoire permet de connaître :

- le titre de la pièce et ses variantes ;
- la date de la première représentation ;
- le genre de la pièce ;
- son nombre d'actes ;
- sa langue ;
- son ou ses auteurs et autrices ;
- si la pièce est en prose ou en vers ;
- si la pièce a été éditée (et dans ce cas on a la référence de ou des éditions) ;
- si la source du texte est disponible (et dans ce cas on a des informations à ce propos) ;
- des chroniques de l'époque à propos de cette pièce, dans une note rédigée par le chercheur à partir des sources qu'il a utilisées.

2. Le répertoire créé par Clarence Brenner en 1961 (Figure 7) constitue une transcription partielle des registres de la Comédie-Italienne de Paris, de toute la période qui nous intéresse, dans un format contemporain. Ce répertoire permet de connaître :

- les titres des pièces représentées, jour par jour ;
- les recettes du théâtre, jour par jour (colonne « Receipts ») ;
- le nombre de spectateur.ices, jour par jour (colonne « Attendance »).

21 settembre 1716

**La Grotte de Scapin (La Gratta di Finocchio)**

Can. it., 3, it.  
In GAMBELLI 1997, pp. 347-350.

GUEULLETTE 1938, p. 79 annota che Scapino si era presentato in scena con il nome di Finocchio e recitava con la maschera di Scapin «ainsi qu'il faisait en Italie, où il était fort goûté, mais comme on était accoutumé à le voir se à visage découvert, on lui fit bientôt quitter le masque et reprendre le nom de Scapino». Scapino fu obbligato a levare la maschera a partire dalla seconda scena: cfr. PARFAICT 1767, vol. III, p. 30; D'ORIGNY 1788, vol. I, p. 36 e DESBOULMIERS 1769, vol. VII, p. 344. Una *Grotte di cavaliere* negli Scenari di Orasio il Calabrese: cfr. PANDOLFI 1988, vol. V, p. 328. Pandolfi rinvia a *La gelosia* contenuto nella Raccolta di Basilio Locatelli: cfr. *Int.*, p. 235 e cfr. la pubblicazione di quest'ultimo canovaccio in TESTAVERDE 2007, pp. 315-327. Cfr. GAMBELLI 1997, pp. 347-350 per *La gratta scchia dello Szenari* di Domenico Biancolli.

24 settembre 1716

**Lélio délinant par amour et Arlequin écolier ignorant (L'Hospitale di pazzi)**

Can. it., 3, it.  
*Argument* in BnF, Ms. f. fr. 9310, ff. 317r-317v.  
In GAMBELLI 1997, pp. 115-117 vi è un *L'hospitale di pazzi*.

GUEULLETTE 1938, p. 79: «Cette pièce est assez jolie, Lélio s'y dégage en chanteur de chansons. Nous avons deux comédies anciennes françaises en vers, de Bays, intitulées *L'hospitale des fous* très probablement de ce canova, dont l'auteur est inconnu. La commença con le commedia francese è probabile dai Parfaict che testimoniano solo la somiglianza del titolo con quello del nostro canovaccio. I Parfaict affermano invece: «voilà précisément le sujet de la Comédie d'Appasie de M. Desmarais, à quelques scènes différentes, entr'autres celle où Lélio se travestit en Chanteur des Chansons»: cfr. PARFAICT 1767, vol. III, pp. 269-270; quali, inoltre, riproducono in breve la trama. Cfr. anche D'ORIGNY 1788, vol. I, p. 36 e DESBOULMIERS 1769, vol. VII, p. 343. Un *L'hospitale di pazzi* nella Raccolta Casamariano: cfr. PANDOLFI 1988, vol. V, p. 353, ma con molte differenze rispetto al nostro canovaccio; cfr. anche la trascrizione del canovaccio in COTTICELLIHECK 2001, vol. II, pp. 175-178. Un *L'hospitale di pazzi* nello Szenari di Domenico Biancolli: cfr. GAMBELLI 1997, pp. 115-117, ma le scene non permettono di ricostruire la trama. Cfr. anche COURVILLE 1945, p. 60.

25 settembre 1716

**Lélio et Arlequin ravisseurs infortunés (Il Violatore deluso)**

Adattamento di Riccoboni L. di una commedia di Boccabadati  
Can. it., 3, it.

«Cette pièce qui est moderne est de la composition du docteur Bocca Badati, grand mathématicien italien». GUEULLETTE 1938, p. 79. Cfr. anche PARFAICT 1767, vol. III, p. 270; secondo D'ORIGNY 1788, vol. I, p. 36 si accollo fedelmente; cfr. COURVILLE 1945, p. 66 e n. 7.

1 ottobre 1716

**Arlequin peintre malade (Non vuol rivali amore)**

Can. it., 3, it.  
In PARFAICT, *Hist. An. Th. It.*, 1767, pp. 231-237; *L'Amour ne veut point de rivaux (Non vuol rivali Amore)*. In GAMBELLI 1997, pp. 239-245.

«Pièce très comique et très ancienne, dont on ne connaît pas la source»: GUEULLETTE 1938, p. 79. *Pièce qui rappresentata all'ancienne Comédie Italienne con il titolo L'Amour ne veut point de rivaux (Non vuol rivali Amore)* e portata alla Normille Comédie-Italienne con il titolo modificato poiché sparsi più convenevole: PARFAICT, *Hist. An. Th. It.*, 1767, pp. 231; cfr. anche DESBOULMIERS 1769, vol. I, p. 74, n. 1. GAMBELLI 1997, p. 240: «Nel 1744 e nel 1745 il canovaccio è stato ripreso con il titolo di *Arlequin peintre*. Vedi anche DESBOULMIERS 1769, vol. I, p. 74. Cfr. COURVILLE 1945, p. 59.

4 ottobre 1716

**Les Quatre Arlequins (Li Quatro Arlechini)**

Can. it., 3, it.  
In PARFAICT, *Hist. An. Th. It.*, 1767, pp. 192-205. In GAMBELLI 1997, pp. 183-194.

«Cette pièce est très ancienne. L'Arlequin (Thouassin) y faisait des tours de force extraordinaires. Il faisait, en dehors, le tour des premières, secondes et troisièmes loges. Le public qui l'aimait beaucoup et qui s'intéressait à sa vie, l'obligea à entracher ce jeu de théâtre qui lui paraissait trop pénible»: GUEULLETTE 1938, p. 80. Cfr. la breve trama riprodotte in PARFAICT 1767, vol. IV, pp. 319-320; cfr. DESBOULMIERS 1769, vol. I, pp. 66-70. Un *Li quattro Arlechini* tra i canovacci rappresentati alla corte dell'imperatrice Anna Ioannovna tra il 1733 e il 1735, dalla seconda e terza compagnia italiana in Russia: cfr. FERRAZZI 2000, p. 129. Un *Quattro Pollicioni simili* tra gli Scenari di Orasio il Calabrese: cfr. PANDOLFI 1988, vol. V, p. 336; un *Li quattro simili* nella raccolta Casamariano su cui cfr. COTTICELLIHECK 2001, vol. II, pp. 216-218. Per le altre varianti, meno vicine al nostro canovaccio, cfr. GAMBELLI 1997, p. 185. Sul canovaccio rappresentato dall'Antienne Troupe Italienne cfr. *Int.*, pp. 183-194 e PARFAICT, *Hist. An. Th. It.*, 1767, pp. 192-205. Cfr. COURVILLE 1945, p. 62, n. 1.

5 ottobre 1716

**Florentina femme fidèle et soldat par vengeance (Flaminia soldato per vendetta)**

Rumaneggiamento di un can. di Romagnesi M. A.  
Can. it., 3, it.  
*Argument* in *Nos. Th. It.*, 1733, vol. I, pp. 9-11. In PARFAICT, *Hist. An. Th. It.*, 1767, p. 326. In GAMBELLI 1997, pp. 479-488.

GUEULLETTE 1938: «Cette pièce, qui est fort ancienne, est fort bonne. Il y a, entre autres, trois scènes fort plaisantes, celle de la valise que l'on vole à Arlequin, de "qui va là ? - Ronda ordinata extraordinaria" et d'Arlequin, bléssé à la tête, qui mange sa cervelle. C'est, dans le manuscrit de Dominique, *Arlequin Soldat et Cavalier*. I PARFAICT 1767, vol. II, pp. 384-386 riproducono la trama esplicita nella terza *Lettre Intermédiaire sur la nouvelle Comédie Italienne* di M. de Charri (Nicolas Bourdin). Cfr. PARFAICT, *Hist. An. Th. It.*

Figure 5 : Capture d'écran d'un extrait du répertoire d'Emanuele De Luca en format PDF.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
10 settembre 1716		Les Équivoques d'amour	de can. it.	3				In GAMBELLI 1997, pp. 387-393.	Canovaccio tratto dalla pièce spagnola di Calderón Arle que todo es mi dama. Un canovaccio dallo stesso titolo fu rappresentato dagli Antenna italiani nel 1667: cfr. PARFAICT 1767, vol. II, pp. 430-433 ove riportato l'estratto di tre scene della pièce. A riguardo si vedano le parti settecentesche conservate nello Szenario di Domenico Biancolli, in GAMBELLI 1997, p. 387-393. Cfr. anche l'estratto della rappresentazione settecentesca che ne dà DESBOULMIERS 1769, vol. VII, pp. 319-320 e D'ORIGNY 1788, vol. I, p. 36. Un Prima di tutto la mia donna o sia giovenco del nostro nella Raccolta di Scenari rappresentati a Varsavia dal 1748 al 1754: cfr. PANDOLFI 1988, vol. V, p. 327. Cfr. COURVILLE II, p. 45 e p. 85 in relazione alle commedie sentimentali e ATTINGER 1993, p. 326.						
14 settembre 1716		Les Tuteurs trompés (I tappeti alessandrini)	(can. it.)	3				In PARFAICT, <i>Hist. An. Th. It.</i> , 1767, pp. 252-253; imitato sintonie di chi le loro mallesse». GUEULLETTE 1938, pp. 70-79. Questo canovaccio fu 260. Dal titolo è Leo rappresentato dagli Antenna italiani nel 1667 col titolo Leo tappeti. Nella rappresentazione del 1716 tappeti / Tappeti. In vi furono diverse corazzioni «dont la principale est d'avoir changés des Manés en Tuteurs». GAMBELLI 1997, pp. amuseuse de leurs punités apparemment pour se conformer à la délicate qui exigent en France la police des Spectacles. à la sévérité des Spectateurs, que l'adultère révolte, même au Théâtre». PARFAICT 1767, vol. V, pp. 585-592. Presenti nel resoconto dei Parfaict alcune battute di Antichino le francesi, riprese verosimilmente dal Cocu immaginare di Molière. Sulla rappresentazione settecentesca cfr. D'ORIGNY 1788, vol. I, p. 16. Si registra un canovaccio sul titolo <i>Li tappeti alessandrini</i> nel Teatro delle favole rappresentative di Flaminio Scala: cfr. MAROTTI 1976, pp. 260-269 (riedito in TESTAVERDE/CHANGELISTI 2007, pp. 87-97) ma che non ha nulla a che vedere con il nostro. Un <i>Li tappeti alessandrini</i> nella Raccolta Casamariano: cfr. PANDOLFI 1988, vol. V, p. 345, che sembra più simile al nostro; cfr. la trascrizione del canovaccio in COTTICELLIHECK 2001, vol. II, pp. 31-34. Corde anche con i tappeti ovvero Colatone Geseo tra gli Scenari pubblicati da Adolfo Bartoli: cfr. PANDOLFI 1988, vol. V, p. 322. Un <i>Les tuteurs de France</i> venne rappresentato in Russia, dalla prima compagnia di comici dell'Arte alla corte dell'imperatrice Anna Ioannovna nel 1731: cfr. FERRAZZI 2000, p. 124. Cfr. anche GAMBELLI 1997, p. 272 sulle relazioni del canovaccio con le opere di Molière, in particolare <i>Le Cocu immaginaire</i> . Vedi COURVILLE II, p. 54 in relazione ai rimandi e alle censure di Lelio Riccoboni.							
21 settembre 1716		La Grotte de Scapin (La can. it. Grotta di Finocchio)	(can. it.)	3				In GAMBELLI 1997, pp. 347-350.	GUEULLETTE 1938, p. 79 annota che Scapino si era presentato in scena con il nome di Finocchio e recitava con la maschera di Scapin «ainsi qu'il faisait en Italie, où il était fort goûté, mais comme on était accoutumé à le voir se à visage découvert, on lui fit bientôt quitter le masque et reprendre le nom de Scapino». Scapino fu obbligato a levare la maschera a partire dalla seconda scena: cfr. PARFAICT 1767, vol. III, p. 30; D'ORIGNY 1788, vol. I, p. 36 e DESBOULMIERS 1769, vol. VII, p. 344. Una Grotta di Meccolino negli Scenari di Orasio il Calabrese: cfr. PANDOLFI 1988, vol. V, p. 328. Pandolfi rinvia a <i>La Gelosia</i> contenuto nella Raccolta di Basilio Locatelli: cfr. <i>Int.</i> , p. 235 e cfr. la pubblicazione di quest'ultimo canovaccio in TESTAVERDE/CHANGELISTI 2007, pp. 315-327. Cfr. GAMBELLI 1997, pp. 347-350 per <i>La grotta vecchia dello Szenario</i> di Domenico Biancolli.						
24 settembre 1716		Lélio délinant par amour et Arlequin écolier ignorant (L'hospitale di pazzi)	can. it.	3				<i>Argument</i> in BnF, Ms. f. fr. 9310, ff. 317r-317v. In GAMBELLI 1997, pp. 115-117 vi è un <i>L'hospitale di pazzi</i> .	GUEULLETTE 1938: «Cette pièce est assez jolie, Lélio s'y dégage en chanteur de chansons. Nous avons deux comédies anciennes françaises en vers, de Bays, intitulées <i>L'hospitale des fous</i> très probablement de ce canova, dont l'auteur est inconnu. La commença con le commedia francese è probabile dai Parfaict che testimoniano solo la somiglianza del titolo con quello del nostro canovaccio. I Parfaict affermano invece: «voilà précisément le sujet de la Comédie d'Appasie de M. Desmarais, à quelques scènes différentes, entr'autres celle où Lélio se travestit en Chanteur des Chansons»: cfr. PARFAICT 1767, vol. III, pp. 269-270; quali, inoltre, riproducono in breve la trama. Cfr. anche D'ORIGNY 1788, vol. I, p. 36 e DESBOULMIERS 1769, vol. VII, p. 343. Un <i>L'hospitale di pazzi</i> nella Raccolta Casamariano: cfr. PANDOLFI 1988, vol. V, p. 353, ma con molte differenze rispetto al nostro canovaccio; cfr. anche la trascrizione del canovaccio in COTTICELLIHECK 2001, vol. II, pp. 175-178. Un <i>L'hospitale di pazzi</i> nello Szenario di Domenico Biancolli: cfr. GAMBELLI						

Figure 6 : Capture d'écran d'un extrait du répertoire d'Emanuele De Luca en format Excel.

1735		Receipts	Attendance
May 21	Mère confidente, Agnès de Chaillot	1,078	634
22	Faucon, Arlequin Hulla	288	192
23	Mère confidente, Billets-doux	383	261
25	Surprise de la haine, Agnès de Chaillot	356	197
26	Ecole des mères, Conte de fées, Billets-doux	3,587	1,270
28	Mère confidente, Conte de fées	1,934	865
30	Mère confidente, Conte de fées	1,150	527
Jun 2	*Arlequin médecin volant	50	31
4	Mère confidente, Conte de fées	1,457	673
5	Mère confidente, Conte de fées	499	344
6	Arlequin poli par l'amour, Veuve coquette, Ile de la folie	208	97
8	Mère confidente, Conte de fées	1,031	619
11	Mère confidente, Conte de fées	640	333
12	Amants réunis, <u>Fille raisonnable</u>	72	59
13	Mère confidente, Conte de fées	947	590
15	Mère confidente, Conte de fées	646	327
18	Mère confidente, Conte de fées	721	300
19	Double Inconstance, Bouquet	79	57
22	Mère confidente, Conte de fées	701	392
23	*Arlequin valet étourdi	106	71
25	Arlequin sauvage, Arlequin Hulla	252	115
26	Amour précepteur, Horoscope accompli	56	41
27	*Deux Léléos et les deux Arlequins, Ile de la folie	243	137
29	Surprise de l'amour, Arlequin toujours Arlequin	228	148
30	Heureux Stratagème, <u>Adieux de Mars</u>	770	395
Jul 2	Heureux Stratagème, Adieux de Mars	766	390
3	Amants ignorants, Triomphe de l'intérêt	89	69
4	Jeu de l'amour et du hasard, Adieux de Mars	631	297
6	Billets-doux, Sylphide, Adieux de Mars	654	343
7	*Arlequin bouffon de cour	97	72
9	Je ne sais quoi, Sylphide, Adieux de Mars	607	321
10	Embarras des richesses, Retour de tendresse	197	131
11	Ecole des mères, Arlequin Hulla, Adieux de Mars	470	261
13	Petit-maître amoureux, Adieux de Mars	342	172
16	Mère confidente, Adieux de Mars	326	154
17	Démocrite prétendu fou, Arlequin poli par l'amour	103	73
18	Heureux Stratagème, Adieux de Mars	220	128
20	Surprise de l'amour, Adieux de Mars	324	216
21	Naufrage	232	141
23	Frère ingrat, Retour de tendresse	151	96
24	Naufrage, Veuve coquette	194	121
25	Triomphe de l'intérêt, Enfants trouvés, Adieux de Mars	524	340
27	Billets-doux, Enfants trouvés, Adieux de Mars	160	83
28	*Arlequin et Mario (sic) valets dans la même maison	285	174
30	Quatre Semblables, Enfants trouvés	285	174
31	Double Inconstance, Adieux de Mars	194	138
Aug 1	Quatre Semblables, Ecole des mères	126	90
3	Femme jalouse, Paysans de qualité	120	68 (?)
4	Femme jalouse, Paysans de qualité	151	76
6	Jeu de l'amour et du hasard, Sylphide	173	109
7	Naufrage, Ile de la folie	101	85

Figure 7 : Capture d'écran d'un extrait du répertoire de Clarence Brenner.

L'acquisition des métadonnées du répertoire en format Excel de De Luca a constitué un premier remaniement des informations. En effet, lors de l'extraction automatique des informations, deux difficultés majeures ont été rencontrées. La première concerne les auteurs : pour chaque pièce, il y a une unique case qui contient la liste de tous les auteurs et autrices, éventuellement précédés par leur rôle (par exemple « mus. de » si la personne s'est occupée de la musique). Le nom d'un même auteur peut en plus avoir des formes différentes. Tout d'abord, il a donc été nécessaire de séparer les noms, afin de pouvoir créer les bonnes entrées dans le tableau des personnes. Le résultat a été un ensemble de fichiers CSV, contenant chacun une catégorie de personnes (dramaturges, compositeurs, chorégraphes etc.). Ces fichiers CSV ont été importés dans la base de données avec l'outil d'importation présent dans l'interface de saisie. Les noms des personnes ont par la suite dû être normalisés par moi-même et par la stagiaire Chiara Protani, afin d'éliminer les différentes formes présentes. La deuxième difficulté, analogue mais plus problématique, concerne les genres, qui doivent être uniformisés

afin d'étudier leur évolution au cours du siècle. En effet, suivant tous les détails qui sont donnés par De Luca, on obtient une quarantaine de genres différents. C'est à travers un dialogue avec Andrea Fabiano et Emanuele De Luca que nous avons finalement choisi les onze genres qui sont actuellement présents.

Or, traiter les données de ces deux répertoires laisse un vide important sur les métadonnées décrivant les pièces dont la première a eu lieu entre 1762 et 1780, car elles ne sont pas traitées dans le travail de De Luca. Avec une priorité sur l'attribution d'un genre et d'une langue, l'étudiante et traductrice Flavia Falbo a effectué un stage de deux mois afin d'étoffer ces données. Pour cela, elle a prioritairement utilisé des ressources disponibles en ligne, en particulier la plateforme de numérisation du Mercure de France et le portail Gallica, à côté d'une source traditionnelle (le deuxième célèbre répertoire de Brenner, *A bibliographical list of plays in the French language 1700-1789*, BRENNER, 1949) et de deux ouvrages critiques (l'édition d'Andrea Fabiano des *Scenari per la Comédie-Italienne* de Carlo Goldoni, GOLDONI, 2017, et la thèse de doctorat de Silvia Spanu, SPANU FREMDER, 2010)<sup>36</sup>. Le répertoire de Brenner de 1949 a par ailleurs fait l'objet d'un projet de numérisation coordonné par Anna Maria Raugei à l'Université de Pise, terminé en 2007 mais aujourd'hui indisponible suite à une attaque hacker contre les serveurs de cette université<sup>37</sup> - selon les informations obtenues lors d'échanges personnelles avec Barbara Sommovigo, collaboratrice du projet et professeure à l'Université de Pise. Ces informations avaient par ailleurs été exploitées par Jeffrey Ravel lors de la première création de la pionnière base de données CESAR<sup>38</sup> en 2001.

Ci-dessous (Figure 8), un schéma des sources utilisées pour saisir les données :

---

<sup>36</sup> La base BIBLIOFI peut être utilisée à l'avenir pour la recherche et la vérification d'une partie de ces informations.

<sup>37</sup> Cet article de blog témoigne de l'ancienne mise en ligne du répertoire : <https://www.fabula.org/actualites/24633/a-bibliographical-list-of-plays-in-the-french-language-1700-1789-le-catalogue-brenner.html> (consulté le 26 août 2025).

<sup>38</sup> Cette base de données est actuellement consultable ici : <https://cesar.huma-num.fr/cesar2/>

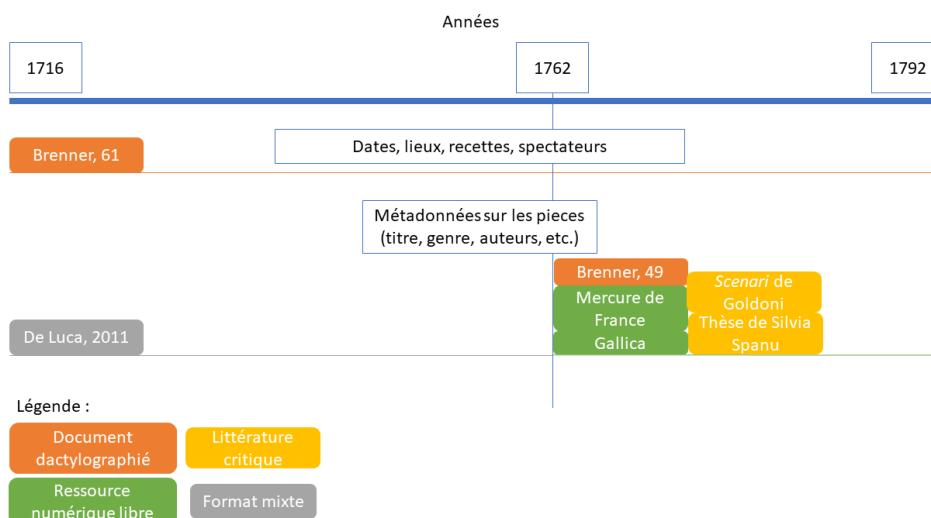


Figure 8 : schéma des sources utilisées pour saisir les données.

### 2.1.2.2 Créer une base de données

La méthode présentée dans cette section concerne la procédure informatique que nous avons suivi pour passer des deux répertoires de départ (décrits dans la section précédente) à l'interface de consultation – autrement dit, elle concerne la chaîne de traitement des informations. Concrètement, la méthode de construction d'une base de données en informatique a été adaptée à notre contexte spécifique.

En informatique, pour construire une base de données on suit les étapes suivantes :

1. Analyse du problème. On observe tout le contexte afin de le comprendre et de faire des choix pertinents : on analyse les données de départ, ce qu'on veut obtenir, les ressources à disposition, les contraintes matérielles du projet. Cette étape est nécessaire afin d'adapter la chaîne de traitement au contexte précis du projet, mais aussi pour en vérifier le bon avancement en définissant des objectifs à atteindre.
2. Définition du modèle des données. A partir des observations faites sur le contexte, on construit un modèle abstrait des données, c'est-à-dire un schéma où sont présents tous les éléments qui doivent être traités. Ce schéma représente la frontière entre les concepts utilisés par les humains et la structure que doivent prendre les informations pour être traitées numériquement. En logique mathématique, le modèle d'une théorie T est un

ensemble de règles et de symboles qui permettent de représenter le fonctionnement de T ; on peut donc utiliser les règles et les symboles du modèle pour étudier le fonctionnement de T. Pour les bases de données, on appelle modèle une structure décrivant les éléments à traiter et leurs relations. Le modèle est *abstrait* car il ne dépend pas d'un langage de programmation spécifique, mais peut au contraire être implémenté avec n'importe quel langage.

3. Test de la cohérence du modèle. Une fois le modèle créé, il faut tester sa *consistance*, ou cohérence, c'est-à-dire vérifier qu'il respecte les contraintes du projet, et que sa structure soit optimisée. L'optimisation du modèle d'une base de données se vérifie avec une procédure mécanique, consistant en une série de transformations de la structure jusqu'à atteindre un certain résultat, ou *forme*. Suivant le degré de performance et de sécurité que l'on souhaite obtenir, on parle de structure respectant la première, deuxième, troisième etc. *forme normale* (ce processus est appelé *normalisation*). Le respect des contraintes du projet est en revanche une procédure moins universelle et dont le déroulement dépend de la manière dont les contraintes ont été définies au début du projet.

4. Implémentation du modèle. Lorsque l'on a obtenu la forme définitive du modèle, il peut être implémenté dans le langage choisi : il acquiert ainsi une existence numérique matérielle. Le modèle abstrait est adapté aux contraintes du langage choisi, afin d'être manipulé et rempli avec les données.

5. Saisie des données. Le modèle implémenté permet donc d'accueillir les données. Ces dernières doivent s'adapter à la forme matérielle définie par le langage choisi. La saisie peut se faire à la main, ou bien de manière automatisée, c'est-à-dire avec une procédure informatique qui transforme les données dans la bonne forme.

6. Publication des données. La manière de publier les données dépend fortement des objectifs du projet et peut par conséquent être très variable : de la mise en ligne, libre ou non, des données brutes, à la création d'une API ou d'une interface graphique de consultation, avec ou sans la possibilité de télécharger les données.

### **2.1.2.3 Chaîne de traitement des informations**

Dans notre projet, l'analyse du problème a été faite à partir de l'étude des répertoires existants et de la problématique de recherche en lettres – à savoir, les changements induits par

la création d'une base de données pour l'étude du Répertoire de la Comédie-Italienne de Paris au XVIII<sup>e</sup> siècle. Un schéma (consultable à travers les annexes) avec les résultats que l'on souhaite obtenir de la numérisation a été créé au début du projet, en dialogue avec Andrea Fabiano et Emanuele De Luca, et à partir d'un document analogue réalisé par Motasem Alrahabi pour le projet Registres de la Comédie-Française. Notre schéma catégorise les résultats par qualitatifs et quantitatifs et les regroupe par type d'objet sur lequel on cherche (personnes, genres, composition des soirées, par exemple). Cette liste a permis de fixer des objectifs concrets depuis le début, et donc de vérifier que chaque nouvelle étape s'y approche, sans perdre la bonne direction. C'est au moment de l'analyse du problème que le choix d'utiliser une base de données relationnelle a été confirmé. Il s'agit de la forme la plus ancienne et la plus explorée de base de données, qui garantit, pour cela, une documentation ample sur le sujet. Mais ce choix assure aussi une plus grande pérennité des données dans l'écosystème numérique, et facilite l'interopérabilité à l'avenir.

Nous avons ensuite créé le modèle abstrait à partir du schéma avec les résultats souhaités, ainsi que de l'analyse des modèles de deux autres bases de données pour le théâtre, Registres de la Comédie-Française et Opéra-Comique Database. Tous les objectifs mentionnés au début du chapitre ont eu des conséquences sur la manière dont le modèle a été défini, à partir du choix de la convention symbolique à utiliser. En effet, la création d'une base de données relationnelle passe généralement par le dessin d'un modèle entité-association, qui prévoit la définition d'objets (entités) et des relations (associations) qui les caractérisent. Il existe plusieurs conventions graphiques pour dessiner ce genre de modèle, la plus courante étant celle de Chen, où les associations sont des losanges et les entités sont des carrés. Afin de nous aligner avec les projets proches dont nous avons la documentation à disposition (Registres de la Comédie-Française et Opéra-Comique Database), il a été créé aussi le modèle logique des données, où les entités sont des tableaux avec la liste détaillée des attributs, et où les relations sont des flèches sans label. Ce type de schéma a d'ailleurs l'avantage d'être relativement simple à concevoir, même s'il nécessite la manipulation de certains concepts mathématiques, comme la cardinalité, pour être bien défini (DESIDERI, 2024). Une description approfondie du modèle final est faite dans les sections suivantes.

La phase de modélisation a abouti à une première version du modèle (Figure 9), qui a été fixée après plusieurs propositions et discussions faites avec les spécialistes de la Comédie-Italienne. Cette première version du modèle a ensuite fait l'objet du processus de vérification

de la cohérence formelle ; le résultat est un nouveau schéma équivalent au modèle de départ (mises à part les entités concernant les activités des personnes, qui sont les moins prioritaires dans le traitement), mais ayant une forme saisie optimisée et sûre (Figure 10). Avec l'aide fondamentale de l'ingénieur informatique Camilla Cantiani, nous avons normalisé la structure pour obtenir la troisième forme normale, qui est la plus couramment utilisée.

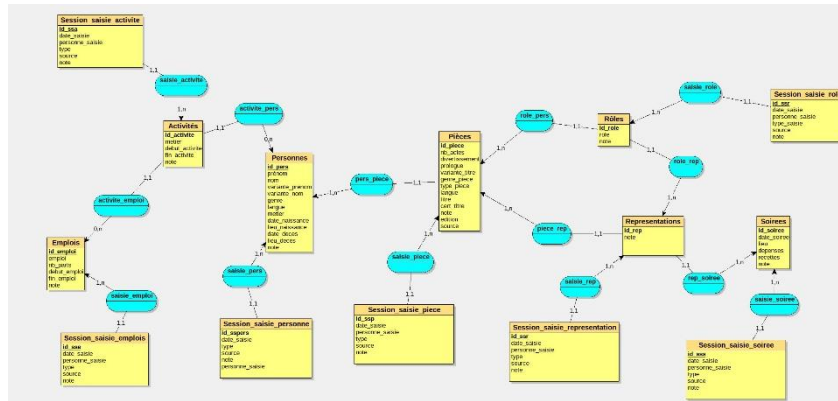


Figure 9 : Image du modèle avant la normalisation (modèle créé avec Looping).

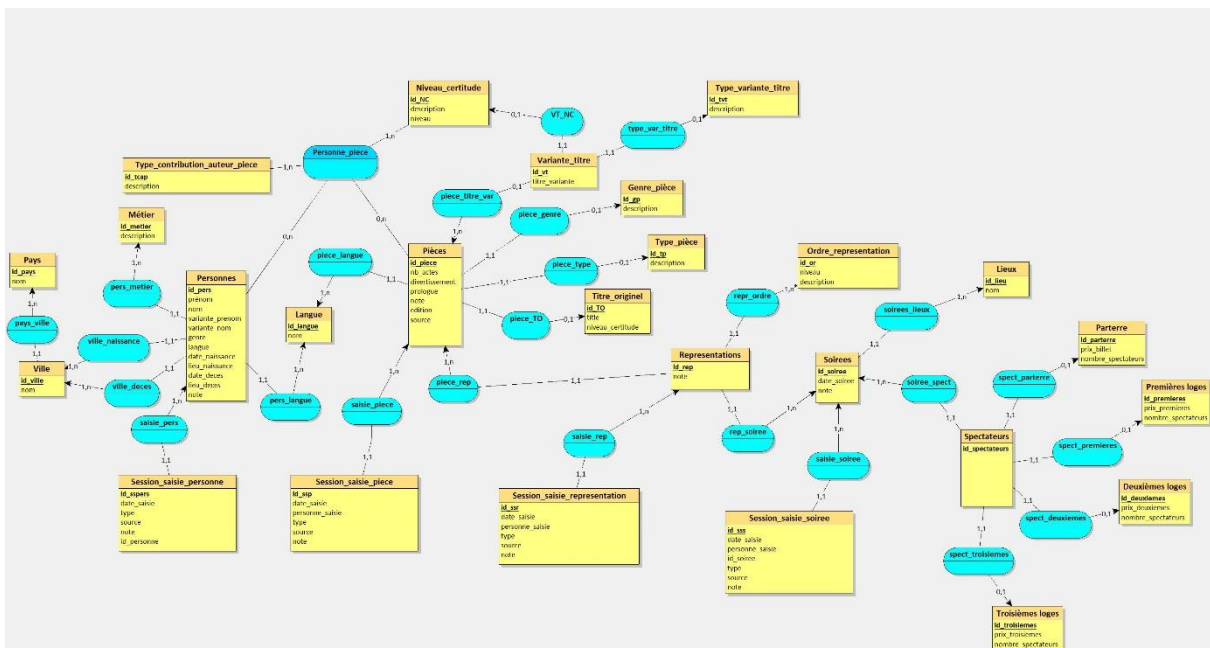


Figure 10 : Image du modèle implémenté (modèle créé avec Looping).

Après une expérimentation avec le logiciel nodegoat, l'implémentation du modèle a finalement été faite sur la base d'un code qui a été fourni par Camilla Cantiani avec la librairie Django de Python. Cela a permis d'avoir une interface adaptée aux documents qui devaient être saisis, ainsi qu'un lien entre l'interface de saisie et celle de consultation. En revanche, ce choix

a le désavantage de réduire la reproductibilité de la chaîne de numérisation, à cause de la nécessité de maîtriser Python pour adapter le code à son contexte. La véritable base de données, en format SQLite, est créée avec Django-même et directement liée à l'interface de saisie.

La saisie des données dans le modèle implémenté est une phase fondamentale pour plusieurs aspects : c'est le processus pendant lequel les informations se transforment, concrètement, en données, et deviennent donc des objets manipulables par des machines. Le plus souvent, il s'agit aussi de la partie la plus chronophage de toute la construction de la base de données. Lors de l'analyse du problème, il est apparu plus approprié à notre contexte de saisir les données à la main, plutôt qu'en automatisant la procédure. En effet, pour cela, il aurait fallu transcrire le document de Brenner avec un logiciel d'OCR, et ensuite faire correspondre les titres aux titres du répertoire de De Luca, les résultats des deux phases devant être corrigés. Or, les deux répertoires ne reportent pas les titres sous la même forme (abrégés et en français chez Brenner, tandis que longs et avec ajout de l'éventuel titre en italien chez De Luca), ce qui génère de nombreuses ambiguïtés. De plus, plusieurs pièces peuvent avoir des titres similaires, voire pareils ; c'est le cas par exemple des *Métamorphoses d'Arlequin* ou de *Italien marié à Paris*, dont existent trois pièces pour chacun de ces titres. Ainsi, ayant le choix entre passer un temps long à corriger les résultats de l'OCR et de l'association automatique des titres, et le passer à saisir les données à la main, cette dernière possibilité a été préférée, dans un souci de continuité des pratiques de recherche et de rapprochement aux sources.

Concrètement, la saisie des données a été assurée principalement par quatre stagiaires, dont j'ai coordonné le travail. L'étudiante en TAL Chloé Petridis, recrutée à Sorbonne Université, a travaillé au projet pendant trois mois et a saisi 120 pages du répertoire de Brenner. Elle a aussi produit des visualisations des données, qu'elle a présenté à la journée d'étude doctorale annuelle du PRITEPS en juin 2024. La doctorante en études théâtrales Chiara Protani, toujours recrutée à Sorbonne Université, a collaboré pendant six mois et en a saisi 110 environ. Elle a aussi vérifié que les métadonnées des pièces, insérées automatiquement, correspondent à la description qui en était faite dans les notes du répertoire de De Luca. Les étudiant.es en études théâtrales<sup>39</sup> Lucas Gaiotti et Mahsa Saberi Moayed à l'Université Côte d'Azur, ont tou.tes les deux fait un stage de deux mois et ont saisi chacun.e environ 30 pages, en partie directement depuis les Registres des dépenses et recettes de la Comédie-Italienne. Ils ont aussi saisi les recettes sur tout le siècle à partir du répertoire de Brenner. Le reste des 450 pages

---

<sup>39</sup> Master Nouvelles Écritures Théâtrales : Recherche et Création.

totales a été inséré principalement par moi-même (115 pages environ), sauf 50 pages qui ont été prises en charge de manière gratuite par un contributeur volontaire, Umberto Desideri – qui est mon père, chercheur et enseignant en ingénierie à l’Université de Pise en Italie. La répartition du travail de saisie du répertoire de Brenner est résumée dans un schéma en Figure 11. Un ultérieur travail de saisie a été effectué par l’étudiante en études italiennes Flavia Falbo, lors d’un stage de deux mois à Sorbonne Université. A partir principalement de ressources en ligne (Mercure de France et Gallica), elle a rempli le plus possible de métadonnées sur les pièces qui ne sont pas traitées par De Luca, avec une priorité pour l’attribution d’un genre et d’une langue.

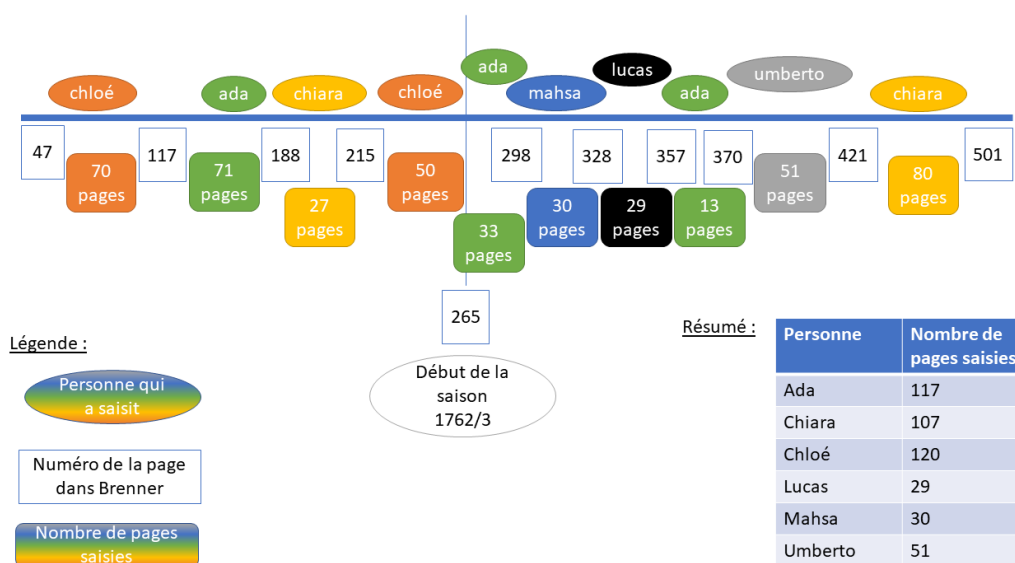


Figure 11 : répartition de la saisie du répertoire de Brenner de 1961.

Au niveau du choix des données, la priorité a été donnée aux dates et aux informations essentielles sur les pièces, comme l’auteur.ice, le genre, le nombre d’actes et la langue. En revanche, ce qui concerne l’activité des personnes travaillant dans le théâtre a été laissée pour une phase ultérieure : la base de données n’apportera ainsi qu’une petite contribution à l’étude de la vie du théâtre – ce qui est par ailleurs l’un des objectifs du projet RECITAL à l’Université de Nantes. Les textes complets des pièces ne sont pas non plus numérisés, pour deux raisons : outre le fait que cela n’est pas strictement nécessaire à l’étude chronologique du Répertoire, ce travail demanderait un genre de traitement des données de nature très différente par rapport à la forme base de données que nous avons choisie. En effet, étant donnée la situation des sources, chaque texte doit être repéré et numérisé de manière spécifique, mais toujours en suivant une même stratégie d’encodage afin que les différentes pièces puissent être traitées dans la même

chaîne. Des références aux éditions et aux sources sont néanmoins présentes dans notre base de données, afin d'en faciliter la recherche en bibliothèque ou dans les archives.

Les données saisies par les stagiaires et par moi-même vont finalement être corrigées de manière experte, par Andrea Fabiano et Emanuele De Luca, les spécialistes de la Comédie-Italienne de Paris de notre équipe. Cette correction est prévue pour fin avril 2026.

La dernière étape de la chaîne de traitement, une fois les données saisies et vérifiées, est la publication des données. Puisque nous nous adressons prioritairement aux chercheur.euses en humanités, la publication d'un choix de résultats en ligne, sous forme d'interface graphique de consultation, semblait nécessaire. La conception de ce site (réalisé par Camilla Cantiani et hébergé à Sorbonne Université dans un serveur géré par le projet LabRadar) est le résultat d'un dialogue virtuel entre les spécialistes de la Comédie-Italienne de Paris et l'ingénieure, dont j'ai assuré la médiation. Concrètement, l'architecture du site a été discutée avec Andrea Fabiano et Emanuele De Luca avant d'être présentée à Camilla Cantiani. Elle a ensuite travaillé en dialoguant avec moi, afin de proposer une première version de l'interface aux spécialistes de la Comédie-Italienne de Paris. La première version a été modifiée successivement, jusqu'à un compromis entre les attentes de ces-derniers, les difficultés techniques et l'emploi du temps. La base de données brute, en format CSV et SQLite, est disponible sur demande.

## **2.2 Résultats de la numérisation**

Le premier résultat du processus de numérisation a été le modèle abstrait des données, concrétisé dans le schéma entité-relation (Figure 12). Ce schéma a été écrit en utilisant deux formalismes équivalents, qui répondent à différentes exigences de communication transdisciplinaire : la notation de Chen (Figure 12) est la plus communément utilisée en informatique, alors que les cahiers des charges de Registres de la Comédie-Française et de Opéra-Comique Database utilisent des tables de relation (notre schéma dans cette notation est visible à la Figure 13). Ce dernier formalisme est d'ailleurs le plus intuitif lorsque l'on travaille avec des bases de données relationnelles, car elle prévoit moins de symboles et permet de visualiser les informations sous une forme utile pour retrouver les liens entre les informations dans les analyses des données.

## 2.2.1 Présentation de la structure formelle des données

### 2.2.1.1 Formalisme

Dans cette section, la structure de la base de données sera décrite à travers ses tables de relation.

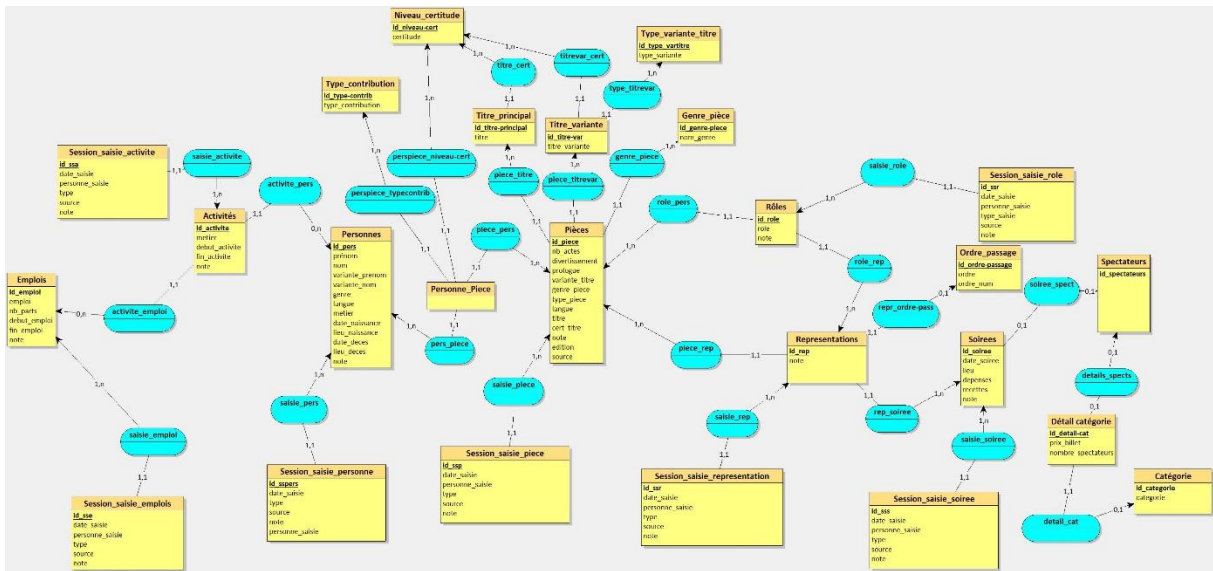


Figure 12 : Modèle abstrait complet représenté avec le formalisme de Chen. (Schéma créé avec Looping)

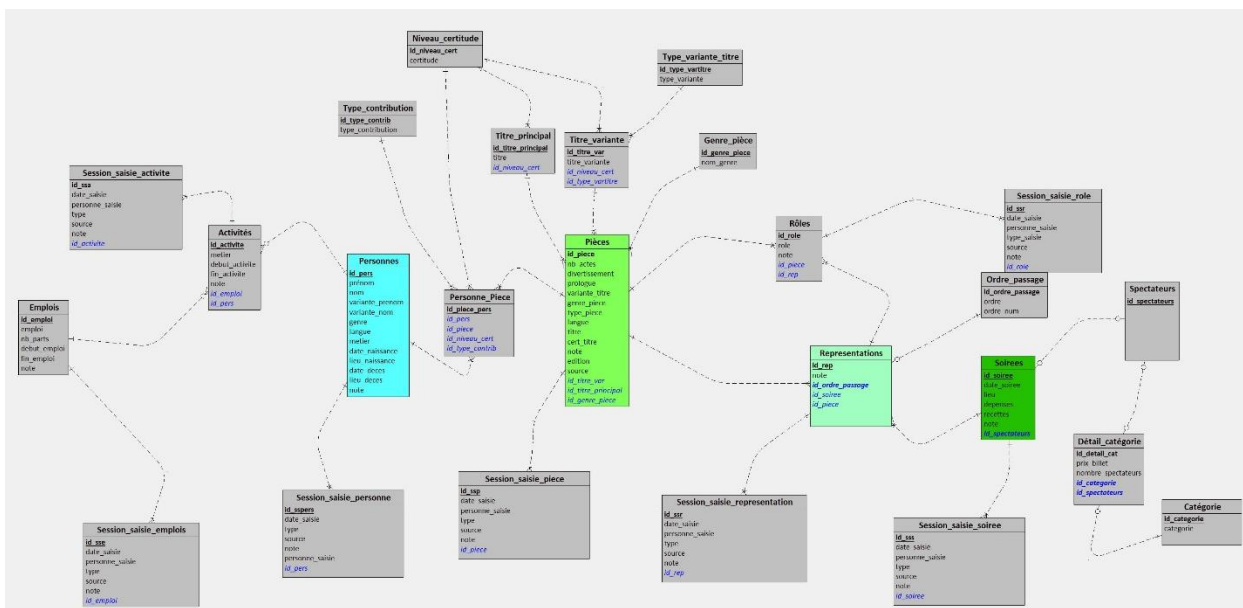


Figure 13 : Modèle abstrait complet représenté comme tables de relation (Schéma créé avec Looping)

Ce formalisme prévoit la définition d'entités sous forme de tableaux, ayant un titre et une liste d'attributs. Le titre correspond au nom de l'entité, et les attributs décrivent les caractéristiques de l'entité. Par exemple, l'entité soirée est décrite, à minima, par un attribut « date » et par un attribut « lieu ». A chaque attribut on peut associer une valeur, qui doit respecter certaines contraintes spécifiques (par exemple, une date doit être composée par trois numéros, le premier compris entre 1 et 31, le deuxième compris entre 1 et 12, et le troisième formé de quatre chiffres). Les objets réels sont donc concrètement décrits par l'ensemble des valeurs qu'on donne aux attributs ; chacun de ces ensembles de valeurs attribuées à un objet réel est appelé *entrée* de la base de données ; par exemple, la soirée du 25 mars 1723 à l'Hôtel de Bourgogne est décrite par l'entrée (date = "25-03-1723", lieu = "Hôtel de Bourgogne").

Un sous ensemble d'attributs (qui apparaissent sous-lignés) constitue la *clef primaire d'identification*. Lorsqu'aucun attribut n'est sous-ligné, l'identifiant seul constitue la clef primaire. Toute entrée a un attribut identifiant, qui permet de la retrouver sans ambiguïté et qui lui est attribué de manière automatique afin d'assurer qu'aucun identifiant ne se répète. Cet attribut ayant uniquement un rôle opérationnel dans la base de données, et étant toujours présent, nous l'avons omis de la liste des attributs des entités dans la description ci-dessous.

Pour chaque attribut, il est nécessaire de mentionner à minima le type de données dont il s'agit (numéro, texte, date, par exemple) et le nombre d'occurrences possibles. Lorsque le type de donnée est FK, cela signifie qu'il s'agit de l'identifiant d'une autre entité ; autrement dit, il s'agit d'une relation. Concernant le nombre d'occurrences de chaque attribut, quatre choix sont possibles : l'attribut peut apparaître

1. exactement une fois ;
2. une fois maximum (on l'indique avec le symbole "!" ) ;
3. une fois minimum (on l'indique avec le symbole "+" ) ;
4. zéro, une ou plusieurs fois (on l'indique avec le symbole "\*").

Les relations sont représentées par des flèches, dont la forme des extrémités décrit le nombre de relations qu'il peut y avoir entre les deux entités reliées. Lorsque le modèle respecte la troisième forme normale, seulement les deux premières possibilités apparaissent.

Ayant présenté le formalisme, la suite sera dédiée à une présentation formelle de ce diagramme, afin de présenter toutes les entités avec tous leurs attributs. Il n'est pas fondamental

de lire cette partie pour comprendre la section suivante, qui concerne les conséquences de nos objectifs sur le modèle, mais fournit les informations nécessaires pour en saisir les détails. Il est donc possible d'utiliser la section suivante uniquement comme référence pour approfondir les considérations de la suite du texte, sans devoir la lire entièrement. Après la présentation formelle du modèle, les pages suivantes seront consacrées à un commentaire de la structure, par rapport aux objectifs mentionnés au début du chapitre. Des exemples seront fournis afin de montrer de quelle manière, concrètement, notre contexte et notre point de vue ont façonné la structure abstraite des données. Les degrés de priorité peuvent s'observer par le niveau de détail des différents objets qui composent le modèle : par exemple, les informations concernant l'espace-temps et celles décrivant les pièces dans leur forme spectaculaire ont le niveau de détail le plus élevé, suivies par les informations sur l'activité du théâtre, et finalement par celles sur les éditions et les sources des pièces.

### 2.2.1.2 Les tables implémentées

Le cœur de nos informations sur le Répertoire est constitué par les quatre tableaux suivants. Il s'agit des entités que nous avons saisies en priorité :

**soirées<sup>40</sup> | représentations | pièces | personnes**

Les entrées du tableau soirées représentent les unités d'espace-temps où les pièces sont jouées. Autrement dit, chaque entrée du tableau **soirées** contient à minima une date et un lieu qui permettent d'indiquer où et quand des pièces ont été représentées.

Les entrées de type soirée ont les attributs suivants :

- date (date)
- lieu (string, default : Hôtel de Bourgogne)
- recettes (entier, !), indiquées en francs
- dépenses (entier, !), indiquées en francs
- spectateurs (FK, \*)
- note (texte, !), pour des communications internes

---

<sup>40</sup> Dans l'interface de consultation, les « soirées » correspondent aux « séances ».

Les soirées sont ultérieurement décrites par deux tableaux contenant les informations sur les spectateurs :

### **spectateurs | détail-catégorie | catégorie**

Chaque entrée du tableau **spectateurs** permet de lier une soirée à des détails sur le nombre de spectateurs par catégorie de billets.

Les entrées de type spectateurs ont uniquement l'attribut suivant, pour des raisons d'optimisation :

- soirée (FK)

Chaque entrée du tableau **détail-catégorie** donne les détails, pour une soirée, d'une certaine catégorie de billets que les spectateurs peuvent acheter (dans le parterre ou dans les loges).

Les entrées de type détail-catégorie ont les attributs suivants :

- spectateurs (FK)
- catégorie (FK)
- prix de l'unité (entier, !), indiqué en francs
- nombre de spectateurs (entier, !)

Chaque entrée du tableau **catégorie** indique le type de ticket dont il s'agit.

Les entrées de type catégorie ont uniquement l'attribut suivant :

- catégorie (string), à choisir dans l'ensemble fini {parterre, première loge, deuxième loge, troisième loge}

Chaque entrée du tableau **représentations** contient les données pour décrire la représentation d'une certaine pièce lors d'une certaine soirée. Concrètement, chaque représentation permet de lier une pièce et une date.

Les entrées de type représentation ont les attributs suivants :

- pièce (FK)
- soirée (FK)
- ordre de la représentation (FK, !), pour indiquer à quel moment de la soirée la pièce a été jouées, ou si elle a été jouée seule
- note (texte, !), pour des communications internes

Chaque entrée du tableau **pièces** contient les données pour décrire une certaine pièce du Répertoire.

Les entrées de type pièce ont les attributs suivants :

- titre principal (FK), en français, sans articles et sans sous-titres
- genre (FK, !)
- nombre d'actes (entier, !), à choisir dans l'ensemble fermé {1,2,3,4,5}
- type de pièce (FK, !), pour indiquer si la pièce est en prose ou en vers, ou mixte
- langue (FK, !)
- prologue (booléen, !), pour indiquer si la pièce a un prologue, à ne pas confondre avec le genre « prologue »
- divertissement (booléen, !), pour indiquer si la pièce a un divertissement, à ne pas confondre avec le genre « divertissement »
- édition (texte, !)
- source (texte, !)
- note (texte, !), pour des communications internes

La description des pièces demande un ultérieur ensemble de tableaux spécifiques aux pièces, pour des raisons d'optimisation de la structure des données. Il s'agit des tableaux suivants :

**titres-principaux | genre | type | variante-titre | type-variante-titre**

Chaque entrée du tableau **titres-principaux** indique le titre normalisé (en français, sans articles et sans sous-titre) d'une pièce.

Les entrées de type titres-principaux ont les attributs suivants :

- titre (string)
- niveau de certitude (FK, !)

Chaque entrée du tableau **genre** indique le genre d'une pièce, suivant une classification qui a été discutée par les spécialistes de la Comédie-Italienne de Paris de notre équipe.

Les entrées de type genre ont l'attribut suivant :

- genre (string), à choisir dans l'ensemble fini :
  - Comédie : pour les comédies jouées à partir de textes entièrement rédigés ;
  - Comédie à l'impromptu : pour les comédies jouées à partir d'un canevas ;
  - Opéra-comique : pour les pièces mêlant des parties parlées et chantées. Elles ne correspondent pas nécessairement aux pièces du répertoire de l'Opéra-comique, et incluent les comédies en vaudevilles et les comédies à ariettes ;
  - Tragédie : pour les tragédies ;
  - Tragi-comédie : pour les tragi-comédies ;
  - Parodie : pour les parodies parlés, chantés ainsi que pour les parodies-traduction. Nous distinguons les trois au niveau de la note relative à la pièce ;
  - Ballet : pour les ballets, sauf les ballets-pantomime et les ballets-divertissement ;
  - Ballet pantomime : pour les ballets pantomime ;
  - Divertissement : pour les divertissements, y compris les ballets-divertissement, mais en excluant les feux d'artifice ;
  - Prologue : pour les entrées de la base de données qui sont des prologues, à ne pas confondre avec les pièces ayant un prologue (dans ce cas on utilise l'attribut prologue de l'entité pièce) ;
  - Feux d'artifice : pour les spectacles de feux d'artifice.

Chaque entrée du tableau **type** indique si une pièce est en prose ou en vers, ou un mélange.

Les entrées de type type ont l'attribut suivant, analogue au tableau **genre** :

- type de pièce (string), à choisir dans l'ensemble fini {prose, vers, prose-vers}

Chaque entrée du tableau **variante-titre** indique un des éventuels autres titres, sous-titres, ou titres d'une partie de la pièce en question.

Les entrées de type variante-titre ont les attributs suivants :

- titre (string)
- niveau de certitude (FK, !)
- pièce (FK)
- type (FK), pour indiquer le type de variante dont il s'agit

Chaque entrée du tableau **type-variante-titre** indique de quelle sorte de titre secondaire il s'agit. Cet ensemble est ouvert mais il faut éviter le plus possible d'ajouter de nouvelles valeurs.

Les entrées de type type-variante-titre ont l'attribut suivant, analogue au tableau **genre** :

- type de variante de titre (string, **obligatoire**), à choisir dans un ensemble ouvert dont les valeurs actuellement disponibles sont {abrégé, orthographe différente, titre d'une partie, titre de la pièce parodiée, titre de la pièce reprise, titre du prologue, titre italien, variante française, variante italienne}

Chaque entrée du tableau **personnes** contient les données décrivant un.e dramaturge, acteur/actrice, danseur.e, un maître de ballet, ou autre personne travaillant dans le théâtre.

Les entrées de type personne ont les attributs suivants :

- prénom (string, !)
- variante du prénom (string, !)
- nom (string)
- variante du nom (string, !)
- métier (FK, !), pour indiquer le rôle principal de la personne dans le théâtre (si pas exemple une actrice participe souvent à l'écriture des pièces, son métier sera « actrice »)

- genre (string, !), à choisir dans l'ensemble fermé {femme, homme}
- langue (FK, !), pour indiquer la langue de sa première socialisation
- lieu de naissance (string, !)
- lieu de décès (string, !)
- date de naissance (date, !)
- date de décès (date, !)
- note (texte, !), pour des communications internes

Pour décrire les relations entre pièces et personnes, un simple lien ne suffisait pas, car cela ne permet pas de rendre compte de la complexité du processus de création des pièces propre à la Comédie-Italienne de Paris. Nous avons par conséquent les deux tableaux supplémentaires :

### **pièce-auteur | type-contribution**

Chaque entrée du tableau **pièce-auteur** permet de lier une pièce à une personne qui a contribué à son écriture.

Les entrées de type pièce-auteur ont les attributs suivants :

- pièce (FK)
- auteur (FK)
- niveau de certitude (FK, !)
- type de contribution (FK), pour indiquer si la personne est l'auteur principal de la pièce, ou si elle a contribué à une de ses parties

Chaque entrée du tableau **type-contribution** indique la manière dont une personne a contribué à la création de la pièce. L'ensemble des types de contributions est fermé, mais il existe une catégorie « autre » pour couvrir des cas qu'on n'aurait pas pris en considération auparavant.

Les entrées de type type-contribution ont l'attribut suivant, analogue au tableau **genre** :

- type de contribution (string), à choisir dans l'ensemble fermé {adaptateur/traducteur, auteur des schènes françaises, auteur du canevas, auteur principal, auteur secondaire, auteur vaudeville, autre, compositeur, maître de ballets}

Deux derniers tableaux complètent les informations sur les pièces et les personnes, et sont liés à différentes entités. Il s'agit des tableaux suivants :

### **langue | niveau-certitude**

Le tableau langue peut décrire la langue d'une pièce mais aussi la langue d'une personne.

Les entrées de type langue ont l'attribut suivant, analogue au tableau **genre** :

- langue (string), à choisir dans l'ensemble fini {italien, français, italien-français}

Le tableau niveau-certitude indique le niveau de certitude de l'attribution d'un titre ou d'un auteur à une pièce. Il peut donc être lié aux tableaux personne-pièce, titres-principaux et variante-titre.

Les entrées de type niveau-certitude ont les attributs suivants :

- description (string), à choisir dans l'ensemble fermé {haut, moyen, bas}
- niveau (nombre décimal), pour associer la description à une valeur numérique comprise entre 0 et 1.

### **2.2.1.3 Les tables à développer**

Deux tableaux supplémentaires sont nécessaires à la description de l'activité d'une personne :

### **rôles | activité | emplois**

Chaque entrée du tableau **rôles** décrit le rôle joué par un certain acteur ou actrice lors d'une certaine représentation.

Les entrées de type rôle ont les attributs suivants :

- rôle (string), pour décrire le rôle que la personne a pris lors de la représentation en question
- acteur (FK)
- représentation (FK)

Chaque entrée du tableau **activité** décrit une activité de travail d'une personne dans la compagnie, quel que soit le métier de la personne.

Les entrées de type activité ont les attributs suivants :

- personne employée (FK)
- emploi (FK, !)
- date de début (date)
- date de fin (date)

Chaque entrée du tableau **emplois** décrit l'emploi par part d'un acteur ou actrice lors de son activité.

Les entrées de type emploi ont les attributs suivants :

- emploi (string, !), pour décrire le type d'emploi dont il s'agit
- nombre de parts (nombre décimal, !)
- date de début (date, !)
- date de fin (date, !)

Il faut prévoir aussi des tableaux pour tenir trace de la saisie et des sources utilisées. Nous avons modélisé ces informations, mais finalement nous avons employé des entités internes à Django pour profiter de certaines fonctionnalités, qui nous permettaient d'automatiser l'insertion des données concernant la saisie, ainsi que d'avoir un historique des modifications faites sur les entités. Pour cela, nous renvoyons donc à la documentation de Django.

## 2.2.2 Conséquences des objectifs sur la modélisation

### 2.2.2.1 Spécificités de notre Répertoire

Afin de répondre aux objectifs concernant l'étude chronologique de la dimension spectaculaire du Répertoire de la Comédie-Italienne de Paris, il est nécessaire de modéliser une pièce comme l'*unité* qu'elle représente pour les spectateurs. En d'autres termes, une pièce n'est pas une représentation spécifique, ni un texte abstrait, ni édité, mais l'ensemble des textes abstraits, des représentations et des éventuelles éditions, auquel le public reconnaît une unité. Par conséquent, le niveau de détail est élevé dans la caractérisation de la pièce, tandis que les informations sur les éditions et les sources sont peu détaillées. Cela a mené, par exemple, à définir les variantes des titres et le type de variante dont il s'agit (abréviation, traduction, etc.). En revanche, les éditions sont décrites de manière simple, par du texte brut. Cela a comme conséquence, par exemple, le fait de limiter la possibilité d'étudier les dynamiques du passage de la pièce jouée au théâtre, à la pièce éditée et publiée. Néanmoins, cela reste réalisable dans des recherches ultérieures, et des informations sur les éditions et sources sont fournies, afin de faciliter la recherche des documents en dehors de la base de données.

Suivant la structure de Registres de la Comédie-Française et Opéra-Comique Database, la variable qui indique le lieu fait partie d'une entité soirée, correspondant à l'unité spatio-temporelle de la programmation du théâtre – le moment durant lequel une ou plusieurs pièces sont jouées, chaque jour. Puisque l'approche de notre recherche est chronologique, la présence d'objets pour décrire l'unité spatio-temporelle dans ce contexte est nécessaire. Concrètement, cette unité constitue un besoin pour différentes raisons : premièrement, suivant les registres du XVIII<sup>e</sup> siècle, Brenner transcrit les recettes de chaque soirée, donc, afin d'étudier ces données, il faut les réunir dans un même objet. De même, dans le but d'étudier les emplois des acteur.ices de la compagnie (cela est actuellement modélisé mais pas implémenté), il faut des objets pour tenir compte du temps, qui soient séparés des représentations des pièces. Techniquement, séparer les représentations de l'espace-temps revient aussi à optimiser la structure des données, suivant la première forme normale.

A partir des modèles de Registres de la Comédie-Française et de Opéra-Comique Database, des modifications ont dû être appliquées pour mieux représenter le processus collectif de création des pièces à la Comédie-Italienne de Paris, où les acteur.ices jouaient un rôle central (FABIANO, 2018). Pierre-François Biancolelli, que nous avons déjà mentionné, est recruté par

Luigi Riccoboni en tant qu'acteur bilingue (il peut jouer en italien et en français), mais aussi pour sa capacité à créer des pièces adaptées au goût du public français. La place des actrices est aussi centrale dans ce contexte : nombreuses d'entre elles au cours du siècle attirent le public avec leur jeu, mais aussi grâce à leur expertise dans le chant et/ou dans la danse, qu'elles mettent au profit du Répertoire. Pour tenir compte de ce cadre créatif, le type d'activité des personnes (dont font partie les acteur.ices) a été modélisé à deux endroits : premièrement, il est possible d'indiquer la fonction principale de la personne dans le théâtre, en lui attribuant un « métier » (entité **Metier**) ; deuxièmement, lorsqu'une personne a participé à la création d'une pièce, il est possible d'indiquer le « type de contribution » au niveau de la relation entre une pièce et une personne (entité **Type\_contribution\_auteur\_piece**). Ainsi, pour une certaine pièce ou pour un ensemble de pièces, il est possible d'étudier quels ont été.es les auteur.ices principaux.ales et secondaires, les auteur.ices d'éventuels prologues ou parties en français et italien dans les pièces multilingues, les compositeurs.ices et les maîtres de ballet. Il s'agit d'informations particulièrement intéressantes pour l'étude de l'hybridation du Répertoire. A ce jour, ces données ne sont pas complètes, et sont inhomogènes entre le début du siècle, où les détails sont connus grâce au répertoire d'Emanuele de Luca, et la fin du siècle, où les informations disponibles existent, mais sont moins systématisées.

### **2.2.2.2 Faire des études quantitatives**

La volonté de faire des études quantitatives sur l'évolution du Répertoire se répercute sur une série d'attributs. En premier lieu sur les dates des soirées, qui ne sont pas de simples suites de caractères mais des objets numériques plus complexes, représentant un moment exact dans le temps. Pour la plupart des langages de programmation, et en particulier dans Python, il existe des bibliothèques pour créer et gérer ce genre d'informations. La seule manipulation qui a été faite à ce propos, a été la transformation des dates des premières du répertoire d'Emanuele De Luca (en italien), en des dates dans un format acceptable pour Python : cela permet de vérifier automatiquement que les premières du répertoire de De Luca correspondent bien aux premières telles qu'elles ont été insérées dans la base de données.

La conséquence sans doute la plus problématique par rapport à l'objectif de produire des analyses quantitatives, concerne le choix de la classification pour les genres. Dans le répertoire d'Emanuele De Luca (DE LUCA, 2011), les genres des pièces sont attribués avec un grand niveau de détail, ce qui génère trop de variables. Or, si l'on affiche un grand nombre de genres différents, nos graphiques risquent de devenir illisibles. De plus, les genres indiquées

par De Luca ne reproduisent pas une classification universellement partagée, car ils reflètent au contraire le caractère hybride du Répertoire et la conséquente difficulté d'attribuer un genre unique à chaque pièce. Restreindre les genres indiqués par De Luca dans une classification moins riche a donc été une tâche problématique et avec un fort degré d'interprétation et de simplification. Concrètement, le choix de cette classification s'est déroulé ainsi : à partir de l'observation des genres reportés dans le fichier Excel de De Luca, ainsi que des genres utilisés dans les graphiques présents dans le livre *La Comédie-Italienne de Paris et Carlo Goldoni* (FABIANO, 2018), j'ai fait une première proposition de classification à Andrea Fabiano et Emanuele De Luca. Nous en avons ensuite discuté afin que je fasse une nouvelle proposition à évaluer ensemble. Ce procédé s'est répété jusqu'à atteindre la classification actuelle, qui comporte les 11 genres décrits ci-dessus (dans Les tables implémentées).

D'autres classifications auraient pu être pertinentes, mais nous nous sommes arrêtés à celle-ci, qui semblait représenter un bon milieu entre trop de détails, et trop de simplification. Elle permet aussi d'étudier certains phénomènes d'intérêt pour les spécialistes de la Comédie-Italienne de Paris dans notre équipe, comme le développement des parodies, l'arrivée de la pantomime et l'exploitation du genre « opéra-comique » avant l'achat du privilège de l'Opéra-Comique. Cette classification répond aussi à l'objectif organisationnel de maximiser la continuité des pratiques de recherche au sein de l'équipe, car elle reproduit en grande partie les catégories utilisés par Emanuele De Luca, en les simplifiant. Les autres attributs des pièces sont totalement repris tels quels du schéma de De Luca, à l'exception des variantes des titres, qui sont données par le chercheur dans la même cellule que le titre principal.

Toujours dans un souci de continuité des pratiques, nous n'avons modélisé les niveaux de certitude que pour les éléments et relations dont cette information était pertinente pour les spécialistes de la Comédie-Italienne de Paris de l'équipe. Cela a mené à pouvoir attribuer un niveau de certitude uniquement aux variantes des titres, à la contribution d'une personne à une pièce, et au rôle d'un.e acteur.ice lors d'une représentation spécifique.

Le modèle initial, décrit dans cette section, a en réalité été simplifié une première fois lors de l'implémentation, et une deuxième fois lors de l'analyse, afin de répondre à des objectifs écologiques de minimalisation des données à traiter. En effet, lors de l'implémentation, toutes les entités qui décrivent l'activité du théâtre (**Activite, Emploi, Role**) ont été ignorées, car ces données ne font pas partie de nos priorités. Ensuite, pour l'analyse, les données relatives à l'activité de saisie ont été exclues, car elles alourdissent de manière considérable le jeu de

données, sans rien apporter de fondamental à l'étude chronologique du Répertoire. Il reste possible de saisir les données dans l'interface de saisie et d'adapter ensuite la plateforme de consultation aux nouvelles données.

### 2.2.2.3 Comparaison avec d'autres bases de données pour le théâtre

Pour montrer les conséquences des objectifs par rapport au contexte plus large de bases de données pour le théâtre, il est possible de faire référence à un article de Jonathan Bollen (BOLLEN, 2016) qui, bien que datant de 2016, fait des considérations épistémologiques générales, qui sont toujours pertinentes. Le chercheur propose une comparaison des modèles de douze projets de bases de données pour le théâtre, dans le but de trouver des traits convergents. Ainsi, il identifie un ensemble d'entités qui apparaissent dans toutes ces modélisations, à savoir : lieux (« places »), personnes (« people »), compagnies (« companies »), performances (« performances ») et œuvres (« works »). Il explique cette convergence par le fait que « il est plus facile de transférer des données entre des projets qui partagent des modèles de données, de la même manière qu'un ensemble d'acteurs transmet des méthodes de travail communes à sa prochaine production »<sup>41</sup> (BOLLEN, 2016, p. 619).

Cette considération résonne largement avec notre projet : le modèle de données est construit en étudiant et en adaptant les modèles des bases de données Registres de la Comédie-Française et Opéra-Comique Database, et y rassemble donc profondément. De plus, la plupart des entités mentionnées sont effectivement présentes dans ces deux modèles et dans le nôtre, avec une exception importante pour la catégorie « compagnies », qui n'est pas pertinente dans des projets qui étudient, justement, le répertoire d'une certaine compagnie – en d'autres termes, il est superflu de créer une entité compagnie lorsque celle-ci est toujours la même. Une considération analogue peut être faite pour les lieux, qui sont bien présents dans notre modélisation, mais avec un degré de précision moindre que ceux des projets décrits par Bollen, où « le nom d'un théâtre est généralement une indication insuffisante »<sup>42</sup> (BOLLEN, 2016, p. 622). Dans la base de données pour la Comédie-Italienne, au contraire, le théâtre est indiqué uniquement comme une simple suite de caractères faisant référence à son nom - autrement dit, les coordonnées géographiques ne sont pas données, ni l'adresse (ancienne ou contemporaine) des lieux des soirées, mais simplement un nom (« Hôtel de Bourgogne », « Versailles »,

---

<sup>41</sup> "It is easier to move data among projects that share data models, in much the same way that an ensemble of actors carries shared ways of working into its next production."

<sup>42</sup> "A theatre's name alone is usually an insufficient designation."

« Fontainebleau », par exemple). Ce choix s'explique tout d'abord par le fait que la Comédie-Italienne joue presque toujours dans le même endroit (en l'occurrence il s'agit de l'Hôtel de Bourgogne), mais aussi pour d'autres raisons. D'un point de vue technique, puisque aucun affichage géographique des données n'est prévu, le lieu peut être représenté par une simple variable – son adresse, ancienne ou actuelle, et les coordonnées GPS sont des informations négligeables. En l'état, la base de données ne permet donc pas à elle seule de spatialiser les données sur une carte, c'est au chercheur.euse de trouver les détails supplémentaires qui l'intéressent. Une raison ultérieure dépend en effet du public auquel la base de données s'adresse : ce qui est envisagé est une utilisation de la base de données par des chercheur.euses en lettres et études théâtrales, qui peuvent donc connaître les emplacements des anciens théâtres. Il reste possible de lier les informations nécessaires à générer des cartographies dans un deuxième temps, pour la dizaine de lieux présents dans notre base de données.

Par rapport aux bases de données décrites par Bollen, dans la nôtre, l'entité qui renvoie le plus aux modélisations des « performances » est l'entité **représentation**, qui est constituée par la relation entre une soirée (donc une unité spatio-temporelle) et une pièce. Cette dernière correspond à l'entité abstraite, distincte et décrite à minima par un titre et un auteur, que Bollen appelle « work ». La dernière entité, « people », équivaut sans équivoque à notre entité « personnes ». Néanmoins, l'activité de chaque personne a été modélisée à deux endroits, comme il a déjà été expliqué au début de la section.

### 2.2.3 Présentation de l'interface de consultation

L'interface de consultation – disponible à l'adresse [repcomit18.mesu.sorbonne-universite.fr](http://repcomit18.mesu.sorbonne-universite.fr) – a été créée par Camilla Cantiani et est hébergée à Sorbonne Université, dans les serveurs du projet LabRadar, qui assure le suivi pour la maintenance. Cette interface, développée avec la librairie python Django, est liée à la plateforme pour la saisie, ce qui assure la mise à jour des données dans les visualisations.

#### 2.2.3.1 Le public de l'interface

L'interface a été conçue pour un public de chercheur.euses en humanités, et à partir des objectifs de recherche de notre équipe. Autrement dit, la plateforme ne s'adresse pas à un public générique qui ne connaît pas la Comédie-Italienne de Paris et qui voudrait avoir des informations introductives en ouvrant le site internet. L'interface est adaptée à une situation de

consultation où l'utilisateur.ice ouvre la plateforme afin de chercher des éléments de réponse à une question de recherche spécifique. L'intérêt de présenter une certaine visualisation à un certain endroit, a été évaluée à partir des recherches et des connaissances des spécialistes de la Comédie-Italienne de Paris de l'équipe. Les visualisations présentes, ainsi que la manière d'y accéder, reflètent ce cadre.

Comme la plupart des sites internet issus de projets de recherche, notre plateforme s'ouvre sur une page de description du projet et de l'objet d'étude (Figure 14). Le menu en haut de cette première page donne accès aux données et aux visualisations, à la structure des données, ainsi qu'à l'interface de saisie (via le bouton en haut à droite, qui mène à une page d'authentification).

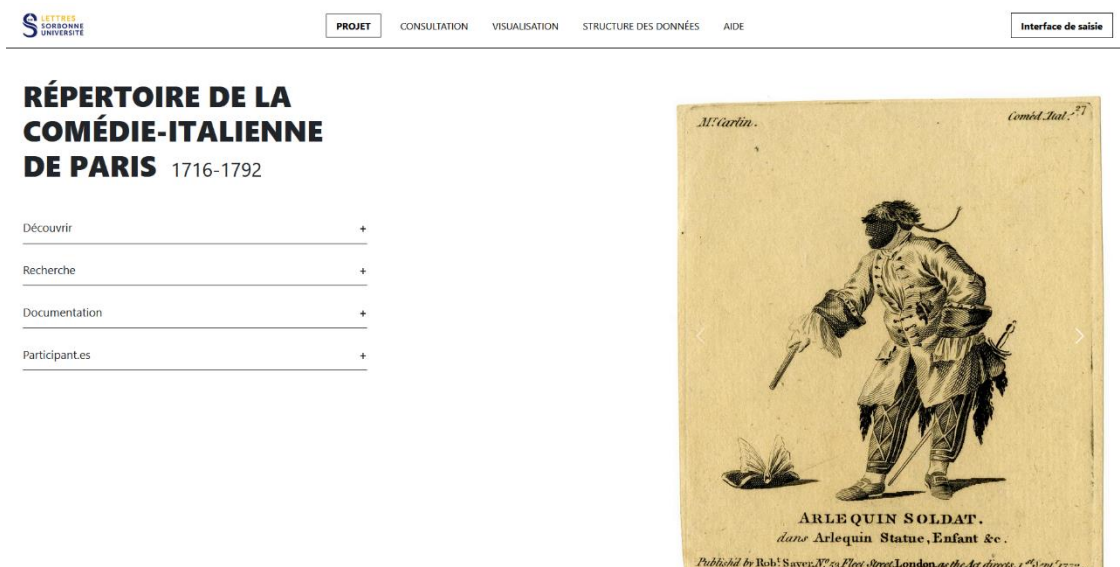


Figure 14 : page d'accueil de l'interface de consultation.

### 2.2.3.2 Consultation tabulaire des données

La page « consultation » sert pour afficher les données brutes, sous forme tabulaire (Figure 15). L'utilisateur.ice peut chercher parmi les pièces, les personnes ou les soirées, soit en effectuant une recherche simple sur tous les champs, soit à travers une recherche avancée avec des filtres adaptés. Cliquer sur une ligne du tableau renvoie à la fiche concernant cet élément (Figure 16 et Figure 17). La fiche contient une première visualisation des données (concernant les représentations de la pièce en question), permettant en plus la comparaison avec une ou plusieurs autres pièces (plusieurs exemples de comparaisons seront données lors des cas d'étude).

LETTRES SOBRIÈRE UNIVERSITÉ

PROJET **CONSULTATION** VISUALISATION STRUCTURE DES DONNÉES AIDE Interface de saisie

## Outil de Recherche

Cette partie de la plateforme permet de consulter les données relatives aux pièces, aux personnes et aux séances, sous forme tabulaire. Il est possible de chercher et sélectionner les données au moyen des filtres sur la droite. Chaque ligne (ou item) peut être cliquée pour ouvrir une fiche avec les détails. La fiche permet aussi de visualiser l'ensemble des représentations de la pièce, et de les mettre en relation avec d'autres pièces.

Œuvres Auteurs Séances

Œuvres  
30 résultat(s) sur 1266 total

Titre	Genre	Type	Langue	Auteurs	Variantes du titre	Édition
<a href="#">Arlequin Persée</a>	Parodie	Prose et vers	Français	Fuzelier		In Parodies 1731, vol. I, pp. 163-230. Confluita in Parodies 1738, vol. II, pp. 87-150.
<a href="#">Arlequin muet par crainte</a>	Comédie à l'improptu		Italien	Riccoboni L. <i>Auteur principal</i>	Il muto per spavento <i>Titre italien</i>	Argument in Nou. Th. It., 1733, vol. I, pp. 48-52. Tutto lo scenario in MAMCZARZ 1973, pp. 81-102. Argument in Nou. Th. It., 1733, vol. I, pp. 48-52. Tutto lo scenario in MAMCZARZ 1973, pp. 81-102.
<a href="#">Attendez-moi sous l'orme</a>	Comédie	Prose et vers	Français	Du Fresny		In GHERARDI 1700, vol. V, pp. 513-549. In GHERARDI 1741, vol. V, pp. 483-518.
<a href="#">Arlequin poli par l'amour</a>	Comédie	Prose	Français	Marivaux Mouret <i>Compositeur</i>		Paris, Veuve Guillaume, 1720. Confluita in Nou. Th. It., 1733, vol. II, pp. 1-56. Confluita in Nou. Th. It., 1753, vol. II, pp. 1-56. In DELOFFRE 1980, vol. I, pp. 71-109.
<a href="#">Arlequin maître de langues</a>						

Recherche par mots-clés :

Filtres disponibles :

Titre (original ou variante) :

Genre :

Nombre d'actes :

Type :

Langue :

Auteur :

Figure 15 : page « consultation » de l'interface de consultation.

LETTRES SOBRIÈRE UNIVERSITÉ

PROJET **CONSULTATION** VISUALISATION STRUCTURE DES DONNÉES AIDE Interface de saisie

## Outil de Recherche

Cette partie de la plateforme permet de consulter les données relatives aux pièces, aux personnes et aux séances, sous forme tabulaire. Il est possible de chercher et sélectionner les données au moyen des filtres sur la droite. Chaque ligne (ou item) peut être cliquée pour ouvrir une fiche avec les détails. La fiche permet aussi de visualiser l'ensemble des représentations de la pièce, et de les mettre en relation avec d'autres pièces.

Œuvres Auteurs Séances

Œuvres  
30 résultat(s) sur 1266 total

Titre	Genre	Type	Langue	Auteurs
<a href="#">Arlequin Persée</a>	Parodie	Prose et vers	Français	Fuzelier
<a href="#">Arlequin muet par crainte</a>	Comédie à l'improptu		Italien	Riccoboni L. <i>Auteur principal</i>
<a href="#">Attendez-moi sous l'orme</a>	Comédie	Prose et vers	Français	Du Fresny
<a href="#">Arlequin poli par l'amour</a>	Comédie	Prose	Français	Marivaux Mouret <i>Compositeur</i>
<a href="#">Arlequin maître de langues</a>				

**Arlequin muet par crainte**

Variantes du titre : • Il muto per spavento *Titre italien*

Auteurs : • Riccoboni L. *Auteur principal*

Genre : Comédie à l'improptu

Acti : 3

Type :

Langue : Italien

Édition : Argument in Nou. Th. It., 1733, vol. I, pp. 48-52. Tutto lo scenario in MAMCZARZ 1973, pp. 81-102. Argument in Nou. Th. It., 1733, vol. I, pp. 48-52. Tutto lo scenario in MAMCZARZ 1973, pp. 81-102.

Représentations : 108

Liens vers des ressources extérieures :

Ajouter une œuvre aux graphiques :

Recherche par mots-clés :

Filtres disponibles :

Titre (original ou variante) :

Genre :

Nombre d'actes :

Type :

Langue :

Auteur :

Figure 16 : exemple de fiche de consultation d'une pièce dans l'interface de consultation (page « consultation »).

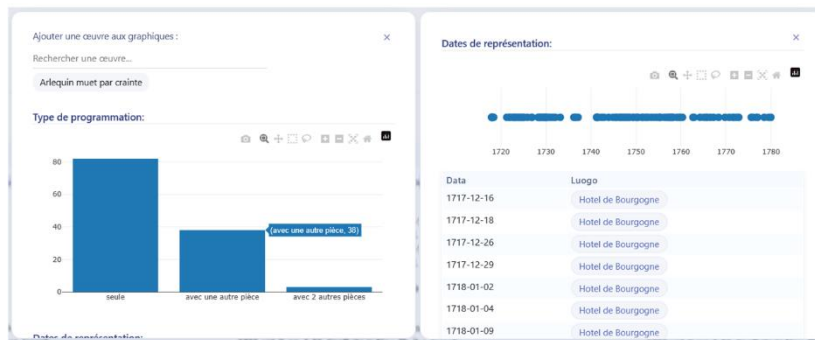


Figure 17 : détail de la fiche de consultation d’une pièce dans l’interface de consultation (page « consultation »).

### 2.2.3.3 Visualiser les données

Les visualisations de données sont accessibles via la page « visualisation » (Figure 18), qui s’ouvre sur trois choix thématiques : pièces, auteurs et premières. En cliquant sur « pièces », on accède à des graphiques qui concernent l’ensemble du Répertoire et de la programmation du théâtre au cours du siècle. Un premier histogramme (Figure 19) permet d’observer, suivant l’année ou la saison, l’évolution chronologique des genres représentés et des langues utilisées. Ce graphique est suivi par un deuxième qui met en relation les genres entre eux au moyen d’un heatmap (Figure 20) : plus il existe de soirées où deux genres sont représentés ensemble, plus la couleur du carré associé est foncée. La troisième visualisation de cette page est encore une fois un histogramme (Figure 21), qui met en relation les genres et les langues des pièces jouées, permettant d’observer le degré de multilinguisme pour chaque genre.

### Que voulez-vous analyser ?

Vous pouvez ouvrir les **graphiques** représentant le Répertoire de la Comédie-Italienne de Paris au XVIII<sup>e</sup> siècle, suivant une des **thématiques** ci-dessous :

Pièces Auteurs Premières

Figure 18 : page « visualisation » de l’interface de consultation.

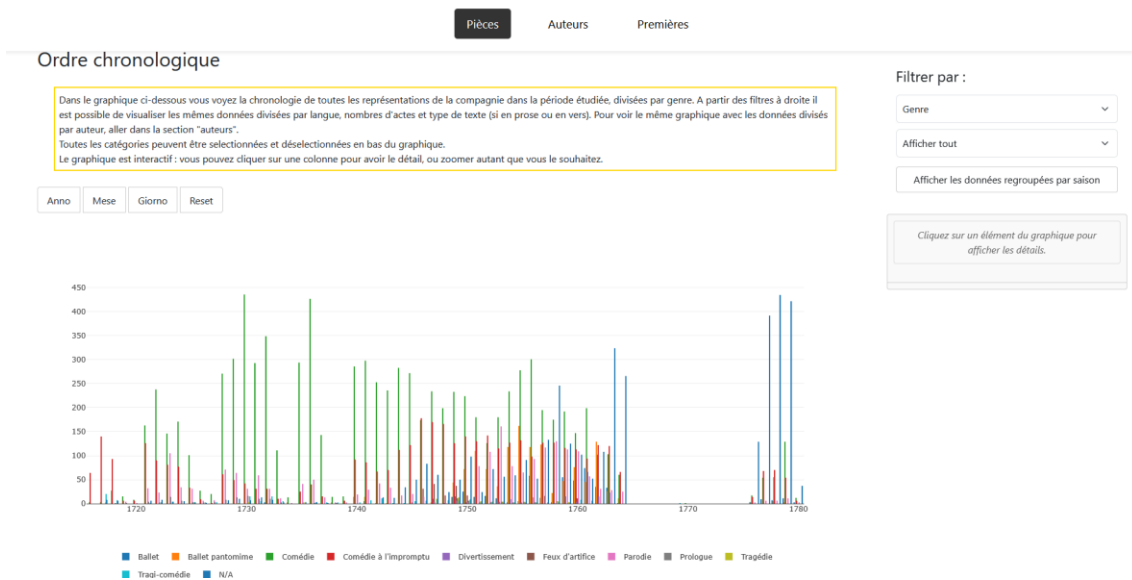


Figure 19 : premier graphique de la page « visualisation - pièces » de l’interface de consultation.



Figure 20 : deuxième graphique de la page « visualisation » de l'interface de consultation.

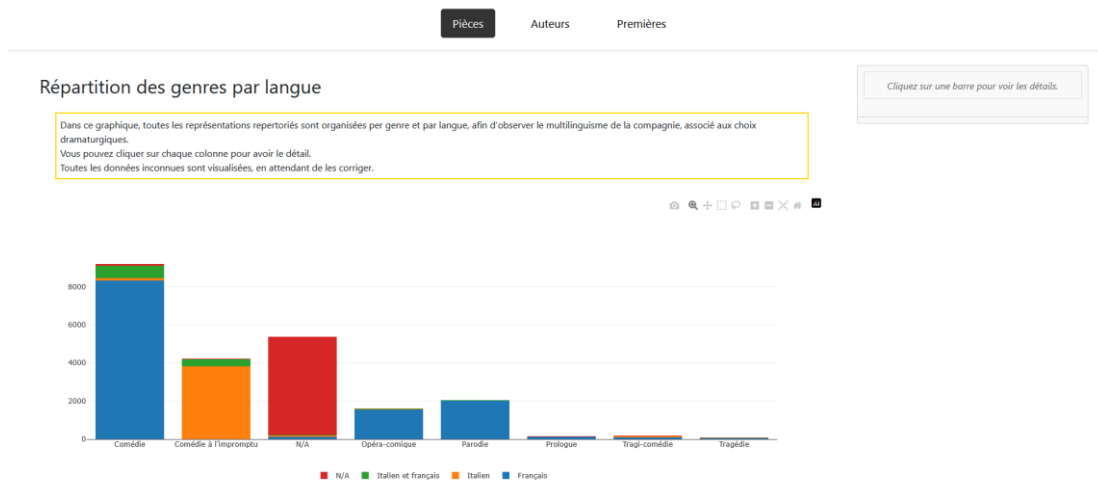


Figure 21 : troisième graphique de la page « visualisation » de l'interface de consultation.

Suivant la thématique « auteurs », l'interface affiche deux graphiques concernant les personnes qui ont contribué à la création des pièces. Le premier graphique (Figure 22) est un simple histogramme, sur le modèle de l'histogramme de la page « pièces », représentant le nombre de pièces d'un ensemble d'auteurs importants. L'ensemble des auteurs affichés peut être modifié, diminué ou augmenté. Le deuxième graphique (Figure 23) permet d'observer l'importance de la création collective au sein de la compagnie, avec un histogramme qui montre combien de pièces ont plusieurs auteurs et autrices.

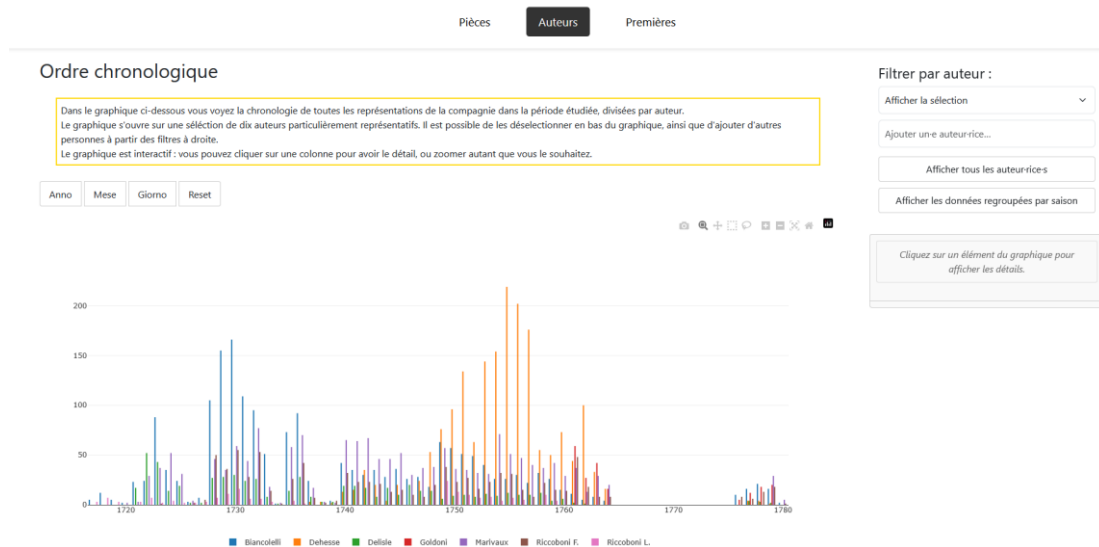


Figure 22 : premier graphique de la page « visualisation - auteurs » de l'interface de consultation.

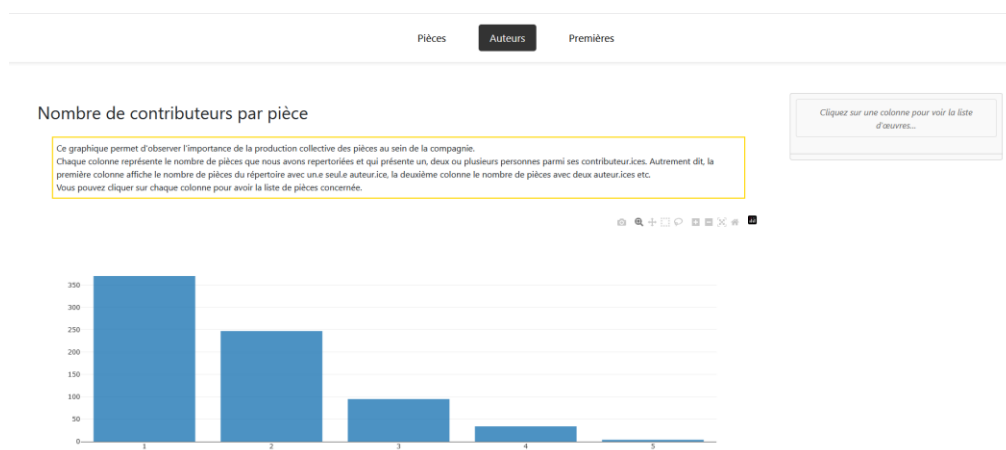


Figure 23 : deuxième graphique de la page « visualisation - auteurs » de l'interface de consultation.

Finalement, la thématique « premières » permet d'observer des phénomènes concernant l'ajout de pièces au Répertoire, à travers l'évolution de la programmation des premières. Encore une fois, la page s'ouvre sur un histogramme chronologique (Figure 24), où est affiché le nombre de premières par genre - avec la possibilité d'observer les données divisées par langue ou par auteur plutôt que par genre. Le deuxième graphique représente la relation entre genre et langue des premières (Figure 25), de manière analogue à la page de visualisation des pièces.

Un troisième histogramme sur cette page (Figure 26), permet d’observer si les nouvelles pièces étaient proposées au public seules et lors de soirées composites.

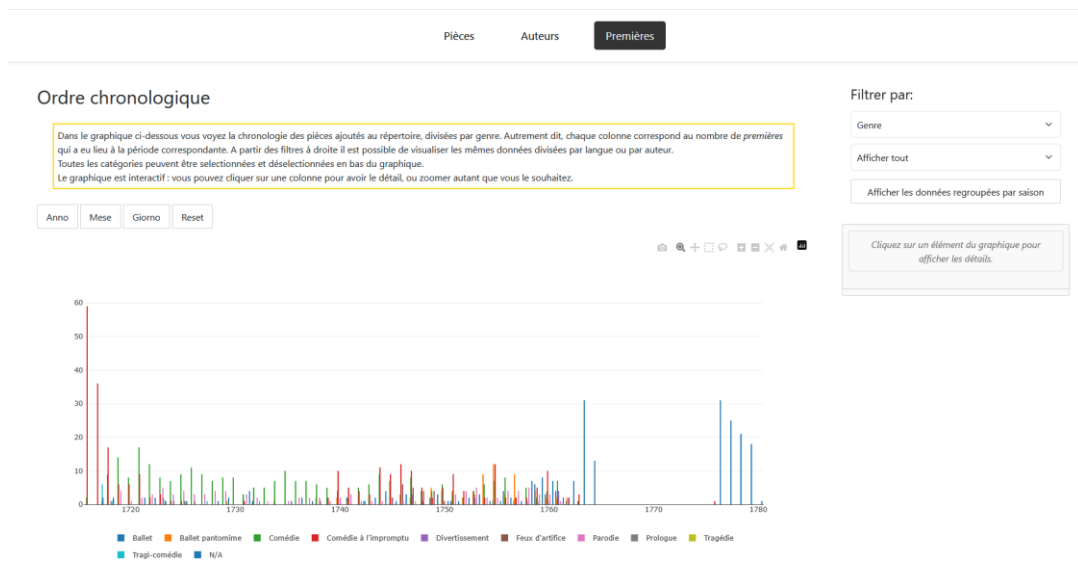


Figure 24 : premier graphique de la page « visualisation - premières » de l’interface de consultation.

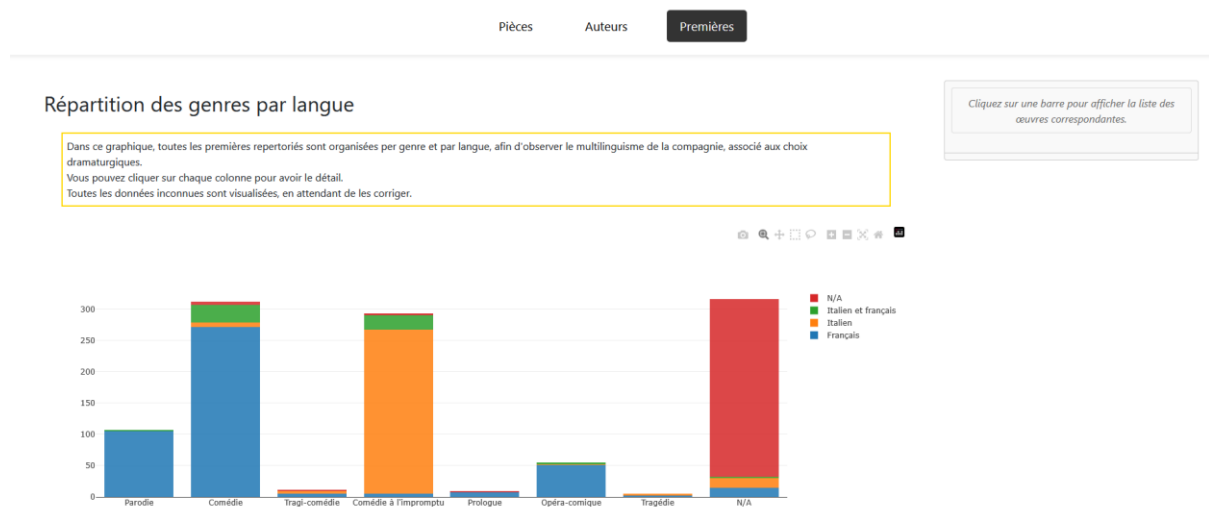


Figure 25 : deuxième graphique de la page « visualisation - premières » de l’interface de consultation.

### Type de programmation

Le graphique ci-dessous représente un détail supplémentaire pour étudier la dynamique de l'ajout des pièces au répertoire. La première colonne correspond au nombre de pièces qui, lorsqu'elles sont jouées la première fois, la séance leur est entièrement dédiée - autrement dit elles ont été représentées seules. La deuxième colonne correspond au nombre de pièces qui ont été jouées la première fois avec un autre pièce, etc. Cela donne des indices concernant les attentes sur la pièce en termes de succès face au public : la fait qu'une pièce soit jouée seule pour la première fois, dépend en partie de sa longueur, mais aussi des recettes attendues. Vous pouvez cliquer sur chaque colonne pour avoir la liste des pièces concernées.

Cliquez sur une colonne pour afficher la liste des œuvres ayant été programmées de cette manière.

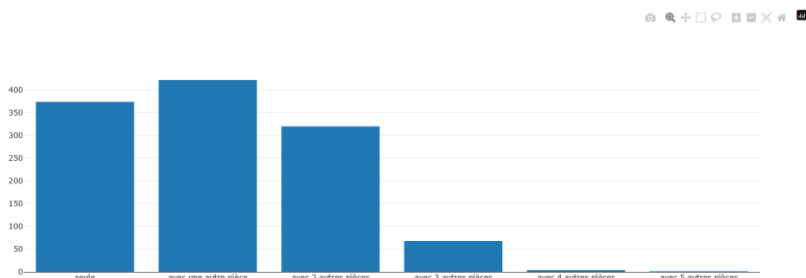
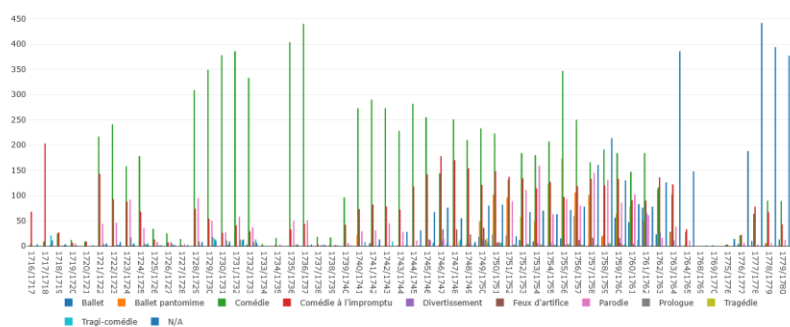


Figure 26 : troisième graphique de la page « visualisation - premières » de l’interface de consultation.

Tous les histogrammes des trois pages de la partie « visualisation » peuvent être affichés avec les données regroupées par saison, plutôt que par année, en cliquant sur le bouton dédié. Cela ajoute un graphique supplémentaire, avec ses propres fonctionnalités, analogue aux autres histogrammes, qui peut à son tour disparaître avec le bouton « masquer les données saisonnières ». Un exemple à la Figure 27.

### Par saison et genre



Filtrer par :

Genre

Afficher tout

Masquer les données saisonnières

Cliquez sur un élément du graphique pour afficher les détails.

Figure 27 : histogramme de toutes les représentations du Répertoire, divisées par genre et par saison théâtrale (page « visualisation – pièces »).

L’interface contient aussi une description de la structure des données et de la chaîne de traitement des documents, dans une page accessible via le menu en haut, ainsi qu’une page d’aide à la consultation. Ces informations enrichissent les informations sur le projet et rendent l’interprétation des visualisations plus transparente. Les informations contenues dans l’interface

de consultation sont à compléter avec les autres supports numériques : framaGit pour les visualisations expérimentales et pour les données brutes ; GitHub pour le code des interfaces de saisie et de consultation ; le carnet hypothèses pour les questions méthodologiques et organisationnelles.

## **2.3 Interopérabilité**

Voulant s’insérer dans le contexte des bases de données pour les études théâtrales, notre projet vise une certaine interopérabilité. Sa réalisation technique se concrétise depuis l’automne 2024, grâce au projet Transnational Stages, concernant la mise en relation de données issues de plusieurs bases de données pour le théâtre. Or, avant le lancement de ce projet d’interopérabilité, certaines mesures avaient été prises afin de faciliter la mise en relation de nos données avec d’autres bases de données proches, en particulier les Registres de la Comédie-Française et Opéra-Comique Database – dans une démarche d’interopérabilité théorique. Dans la suite de cette section, seront donc d’abord présentés ces efforts, avant de se focaliser sur une comparaison de notre base de données avec celles impliquées dans le projet d’interopérabilité.

### **2.3.1 Préparer l’interopérabilité**

Avant de commencer le processus de création de la base de données, il a été observé ce qui était possible (ou pas) de faire dans les autres plateformes. De plus, nous avons eu accès à la structure des données des projets Registres de la Comédie-Française et Opéra-Comique Database (grâce à Motasem Alrahabi), à partir desquels le nôtre a été construit. Afin de garantir la possibilité d’implémenter une interopérabilité qui n’était que potentielle dans cette phase du projet, la structure de la base de données a été décrite avec la plus grande précision : il ne s’agit pas seulement de donner son format et un schéma de la structure, mais aussi de préciser la signification de tous les éléments qui composent le modèle abstrait. Cela a comme but de faciliter des futurs alignements avec les autres modèles ou avec des ontologies, en réduisant ou en explicitant les ambiguïtés – qui sont en partie inévitables. La description de la structure peut être lue à plusieurs endroits, suivant un principe de redondance : dans un repository framaGit, dans un carnet hypothèses et dans l’interface de consultation.

En ce qui concerne les outils, au début le projet était orienté vers l’utilisation du logiciel nodegoat pour la saisie des données. Cela avait des avantages évidents : premièrement, il s’agit d’une plateforme pour créer des bases de données en histoire, ce qui était cohérent avec notre

approche chronologique. Deuxièmement, le processus pour insérer les données via la plateforme était en accord avec le déroulement de notre travail : on crée la structure abstraite, avant de pouvoir la remplir. Si ce chemin avait été poursuivi, la chaîne de traitement serait aujourd'hui probablement plus reproductible, mais le travail aurait été plus pénible, et la plateforme de consultation n'aurait pas pu dialoguer avec la plateforme de saisie. Finalement le choix a été fait d'utiliser une plateforme de saisie créée ad hoc par l'ingénieure informatique Camilla Cantiani. Cette plateforme est adaptée au genre de travail qui doit être accompli dans le projet, c'est-à-dire insérer, ligne par ligne, les informations contenues dans un répertoire (celui de Clarence Brenner de 1961), en faisant correspondre les titres avec les données présentes dans un deuxième répertoire (celui d'Emanuele De Luca de 2011). L'utilisation d'une librairie Python (Django) pour la création de cette application a aussi induit des modifications dans la structure logique, afin de l'adapter aux contraintes du langage. Mais à la différence des modifications nécessaires pour utiliser nodegoat, qui auraient changé la manière d'accéder aux informations chronologiques, ces adaptations ont été faites sur une partie de la structure qui n'était pas au centre de l'étude du Répertoire et qui permettait des fonctionnalités plus puissantes (mais utiles) que celles pensées dans la structure d'origine. Il s'agit en l'occurrence des entités décrivant l'activité des personnes assurant la saisie : Django permet de créer des utilisateurs avec des rôles et permissions différentes, simples à gérer, et aussi de reconstituer l'historique des versions successives d'un certain élément. Mais surtout, Django permet de lier la plateforme de saisie à la plateforme de consultation. Grâce à cela, le projet peut en partie continuer sans d'ultérieures interventions d'informaticien.es – au moins en ce qui concerne la saisie des informations, ainsi que la vérification et la correction des données actuellement présentes.

À défaut de pouvoir créer rapidement et simplement une plateforme permettant de faire des requêtes dans toutes les bases de données des nombreuses plateformes pour les études théâtrales, il est tout de même utile de les rassembler dans un même site. Une liste de projets numériques concernant les études théâtrales (ici en annexe) a donc été insérée dans le carnet hypothèses, afin de connaître facilement l'existence, les objectifs et les relations entre les différents projets, et de faciliter ainsi le travail des chercheur.euses.

### **2.3.2 Le projet Transnational Stages**

Concrétiser l'interopérabilité reste un objectif ambitieux dans le contexte des humanités numériques pour les études littéraires : on constate une multiplicité d'approches, engendrant

une multiplicité de modélisations, de chaînes de traitement, de visualisations, qui ne sont applicables dans des contextes différents qu’au prix de plus ou moins fortes adaptations (pour une approche critique à la problématique de l’interopérabilité voir WITZ et PORTE, 2025). Afin de dépasser ces limites, le projet d’interopérabilité Transnational Stages, porté par Ravel et Kallenbach, a commencé par l’organisation de deux réunions de discussion et de présentations des bases de données concernées, afin de prendre conscience et d’analyser les différences et les similitudes. Suite à ces deux réunions, j’ai créé un fichier partagé contenant un schéma avec les entités et les attributs des autres projets, qui ont une correspondance avec les éléments présents dans la base de données pour la Comédie-Italienne de Paris. Ce schéma, dont un extrait est visible à la Figure 28, a vocation à être corrigé et éventuellement amplifié par les porteurs.eures des autres projets.

N’ayant pas pu analyser en détail les modèles des autres projets, et le fichier devant encore être corrigé, il n’est pas possible d’en faire un commentaire approfondi. Mais on peut néanmoins remarquer, sans étonnement, que les entités qui correspondent le plus dans tous les projets sont celles qui décrivent les représentations, les pièces, les dates et les personnes.

	A	B	C	D	E	F	G	H
1	entité	attribut	RCF	Royal danish theater	Onstage	Theatronomi cs	Theater in St Domingue	Sceneweb
25	personne		personne, auteur, acteur	person	persons		auteur, compositeur, participant, bénéficiaire	person
26		prénom						
27		variante						
28		prénom						
29		nom						
30		variante nom						
31		métier	types d'acteurs (sociétaire/pensionnaire)	work contributor	role			functions
32		genre						genre
33		langue						nationality
34		lieu de naissance						
35		lieu de décès						
36		date de naissance						born
37		date de décès						dead
38		note						
39	pièce		pièce	works	play		pièce	artwork
40		titre				p_PerformanceTitle	titre	original title
		genre	genre pièce				genre	categories
		nombre					nombre	

Figure 28 : Extrait du fichier CSV qui met en relation les entités et attributs des bases de données concernés par le projet d’interopérabilité coordonnée par Ravel et Kallenbach.

Le choix de réunir des projets existants au sein d’une même interface de consultation reflète en grande partie la proposition de Witz et Porte (WITZ et PORTE, 2025) dans le numéro sur la sobriété numérique de la revue *Humanités Numériques*. Constatant une diversité toujours croissante au sein des projets de numérisation, même en tenant compte du grand nombre de formats et outils interopérables qui sont aujourd’hui disponibles, les auteurs préconisent l’utilisation d’interfaces communes séparées. Autrement dit, plutôt que de réunir dans une

grande base de données plusieurs jeux de données plus petits, chacun est aligné à une même interface de façade, qui ne modifie en rien les bases de données qui la composent. Cela permet de faciliter la maintenance, car cette technique réduit globalement le nombre d'alignements à réaliser pour mettre toutes les données en relation. Mais surtout :

« cette façade gomme les spécificités et détails d'implémentation des projets producteurs qui sont inutiles aux projets consommateurs, mais conserve et harmonise ce qui est indispensable. Ainsi, chaque base de données peut garder ses spécificités. Chacune voit la justification de son existence renforcée, car elle prend part à un écosystème plus grand que la somme de ses parties, car il permet de tirer le meilleur de toutes les données qui y transitent » (ibid.).

Transnational Stages se base sur une idée similaire : toutes les bases de données conservent leur autonomie et leur spécificité, et l'interface commune rend accessibles, depuis un même endroit, un ensemble de données comparables dans les divers projets. Cela a l'objectif ultérieur de faire connaître et de valoriser les différentes recherches et leurs spécificités.

Il reste par ailleurs possible de mettre en commun nos données avec celles d'autres projets, notamment la base CESAR et les recherches du Cethéfi (en particulier BIBLIOFI<sup>43</sup> et RECITAL). Une fois les données corrigées, il serait tout-à-fait envisageable de les intégrer dans le calendrier du projet CESAR ou dans le tableau des pièces<sup>44</sup> de BIBLIOFI. Or, par rapport à Transnational Stages, il s'agit dans ces deux cas d'adapter la forme de nos données à celle du projet hôte, et de la séparer totalement de la base de données d'origine, en perdant ainsi en partie le caractère comparatif qui est l'un des objectifs et intérêts principaux de Transnational Stages. Le choix de traiter le Répertoire de la Comédie-Italienne comme une unité reflète aussi l'approche de recherche des spécialistes de théâtre de l'équipe, et se distingue du Cethéfi qui aborde l'ensemble des théâtre forains et de la Comédie-Italienne. Notre base de données et la plateforme de consultation donnent au contraire une autonomie au Répertoire de la Comédie-Italienne, dont le fonctionnement institutionnel est davantage comparable à celui de la Comédie Française qu'à celui des théâtres de la Foire.

---

<sup>43</sup> <https://cethefi.org/bibliofi>

<sup>44</sup> <http://cethefi.org/bibliofi/pièces>

### 2.3.3 Logiciels et formats utilisés

Cette dernière section est consacrée à établir une liste des logiciels et des autres technologies utilisées, en décrivant les contextes d'utilisation et les raisons de leur choix. Globalement, nos décisions visent à maximiser au même temps l'utilisation de logiciels libres ou ouverts<sup>45</sup>, ainsi qu'à favoriser l'appropriabilité de la démarche. De manière cohérente avec nos objectifs, une attention particulière a été portée à assurer la continuité avec des projets proches d'humanités numériques pour les études théâtrales. Par conséquent, les technologies les plus nouvelles n'ont pas nécessairement été valorisées, en priorisant au contraire celles dont les applications sont nombreuses et de longue date.

#### 2.3.3.1 Modélisation des données

Pour la modélisation des données, il faut un logiciel adapté à la description de schémas entité-association. Or, techniquement, un tel schéma peut être dessiné avec n'importe quel logiciel qui supporte la création de formes géométriques, même des logiciels de bureautique comme PowerPoint. L'intérêt d'utiliser un logiciel adapté réside dans une plus grande simplicité lors de la création du schéma, mais aussi dans la possibilité de visualiser différentes formes du même schéma et de pouvoir générer du code permettant la création ou la manipulation de la véritable base de données associée au schéma.

Dans un premier temps, le modèle a été dessiné avec des logiciels en ligne, en particulier [db.diagram.io](https://db.diagram.io) et [dbdesigner.net](https://dbdesigner.net), suivant le choix fait dans les cahiers des charges de Registres de la Comédie-Française et de Opéra-Comique Database. Si ces sites ont l'avantage de pouvoir partager une version agrandissable du schéma avec un simple lien hypertexte, le nombre d'entités par schéma et d'autres fonctionnalités sont limitées dans la version gratuite.

Finalement, l'utilisation de Looping<sup>46</sup> a constitué le choix définitif : il s'agit d'un logiciel libre et gratuit développé à l'Université de Toulouse, permettant d'obtenir simultanément le modèle entité-association ainsi que le modèle logique. Looping n'est disponible que pour Windows, mais il est facilement exploitable sur Ubuntu, à travers l'émulateur Wine<sup>47</sup>. Aucun guide ou tutoriel gratuit n'est disponible sur le site du

---

<sup>45</sup> Pour la différence entre libre et open voir <https://www.gnu.org/philosophy/open-source-misses-the-point.html>

<sup>46</sup> <https://www.looping-mcd.fr/>

<sup>47</sup> <https://www.winehq.org/>

téléchargement ; néanmoins, il existe plusieurs tutoriels vidéo qui facilitent la prise en main. Le coût d'accès aux fonctionnalités reste élevé, car elles demandent la manipulation de concepts informatiques comme la cardinalité d'une relation et les types de données (sur la difficulté de manipuler ces concepts voir DESIDERI, 2024).

### 2.3.3.2 Saisie et manipulation des données

Pour la saisie des données, trois possibilités ont été prises en compte : Omeka<sup>48</sup>, Nodegoat<sup>49</sup> et la création d'une interface personnalisée. A premier abord, le logiciel qui répond le plus à nos exigences semblait être Omeka, gestionnaire de bibliothèques numériques adapté au contenu historique. Ce logiciel ouvert existe en plusieurs versions et a l'avantage de permettre la saisie directe des données, avant de créer la structure logique. Or cela n'était pas adapté à notre contexte et à notre méthode de travail, qui prévoyait de suivre la démarche informatique pour la création de la base de données, commençant par la définition d'un modèle de données. Pour cela, Omeka a été écarté.

Nodegoat est aussi un logiciel ouvert pour la création de bases de données en histoire, mais qui s'adaptait davantage à notre méthode. En effet, il est nécessaire de définir la structure des données afin de pouvoir les insérer. Comme Omeka, Nodegoat permet quelques visualisations de données ainsi que la possibilité de télécharger les données en différents formats. Afin de supporter plusieurs utilisateurs.ices – caractéristique fondamentale pour un projet d'équipe – ce logiciel doit être installé sur son propre serveur. Avec la version pour un.e seul.e utilisateur.ice, j'ai transformé les données du fichier Excel d'Emanuele De Luca et j'ai saisi une dizaine de pages du répertoire de Brenner.

Finalement, nous avons eu la possibilité de collaborer avec l'ingénieure informatique Camilla Cantiani qui m'a proposé à titre gratuit une interface de saisie qui était davantage adapté à notre contexte de travail. Cette interface est construite au moyen de la librairie Python Django<sup>50</sup>, et a été déployée, lors de la phase de travail, sur un serveur de ObTIC à travers Docker<sup>51</sup>. Le code est librement disponible sur GitHub sur demande, pour des raisons de sécurité de la base de données. Le code écrit avec Django crée au même temps la base de

---

<sup>48</sup> <https://omeka.org/>

<sup>49</sup> <https://nodegoat.net/>

<sup>50</sup> <https://www.djangoproject.com/>

<sup>51</sup> <https://www.docker.com/>

données, en format SQLite<sup>52</sup>, une interface de consultation et une interface d'administration du site, à laquelle il est nécessaire de se connecter pour saisir les données.

Le format SQLite est un format de bases de données léger et interopérable, qui ne nécessite pas d'un serveur pour fonctionner – trois caractéristiques pertinentes pour nos objectifs.

Au même temps que la saisie, nous avons commencé à faire des expérimentations de visualisations de données. Pour cela, nous manipulons le fichier SQLite contenant nos données sur des Jupyter Notebook<sup>53</sup>, avec les bibliothèques python suivantes :

- Sqlite3<sup>54</sup> pour manipuler le fichier SQLite ;
- Pandas<sup>55</sup> et Numpy<sup>56</sup> pour manipuler et analyser les données ;
- Datetime<sup>57</sup> pour la gestion des dates ;
- Matplotlib<sup>58</sup> et seaborn<sup>59</sup> pour les graphiques ;
- Plotly.express<sup>60</sup> pour les graphiques interactifs.

Toutes ces bibliothèques sont open-source et disposent d'une documentation riche.

Les Jupyter Notebook peuvent être créés et visualisés sur plusieurs plateformes, mais je me suis surtout basée sur VisualStudio Code, qui est un éditeur de code disponible pour les principaux systèmes d'exploitation (Microsoft, MacOS et Linux). Au moyen de nombreuses extensions téléchargeables gratuitement, ce logiciel peut s'adapter à plusieurs langages de programmation. Son installation est très simple et le logiciel propose automatiquement les extensions pour la lecture des fichiers ouverts. Les extensions nécessaires à utiliser les Jupyter Notebook varient selon les systèmes d'exploitation et selon ce qui est déjà présent sur l'ordinateur.

---

<sup>52</sup> <https://sqlite.org/>

<sup>53</sup> <https://jupyter.org/>

<sup>54</sup> <https://docs.python.org/3/library/sqlite3.html>

<sup>55</sup> <https://pandas.pydata.org/>

<sup>56</sup> <https://numpy.org/>

<sup>57</sup> <https://docs.python.org/3/library/datetime.html>

<sup>58</sup> <https://matplotlib.org/>

<sup>59</sup> <https://seaborn.pydata.org/>

<sup>60</sup> <https://plotly.com/python/plotly-express/>

### **2.3.3.3 Publication des données**

Les visualisations faites sur les Notebook Jupyter ont servi de base pour proposer des graphiques dans l'interface de consultation. Ces derniers ont été générés par Camilla Cantiani avec la librairie python Plotly. Camilla Cantiani a aussi créé l'interface graphique, dans le même environnement que l'interface de saisie, en se basant toujours sur la librairie Django et sur un déploiement serveur avec Docker. Sur demande, il est possible d'obtenir les données en format CSV ou SQLite, afin de s'adapter au contexte de travail des utilisateurs.

## 3 Partie 2 : utiliser la base de données

Après avoir observé, dans la première partie, le travail concrètement effectué lors du projet de doctorat, ce deuxième chapitre montre de quelle manière la base de données peut être utilisée. Les données sont accessibles de deux manières : des visualisations interactives sont proposées dans l'interface de consultation, mais les utilisateur.ices peuvent aussi demander l'accès aux données brutes et les manipuler, afin d'obtenir des résultats que l'interface ne peut pas générer. Dans le premier cas, l'utilisation des données est contrainte par les choix interactifs disponibles : il est possible de visualiser plusieurs histogrammes et certaines distributions temporelles des données, mais aucune carte ou autre visualisation géographique n'est fournie, par exemple. Dans le cas de l'exploitation des données brutes, leur utilisation est contrainte par la connaissance d'un langage de programmation ou de tout autre outil permettant la manipulation d'un fichier CSV ou SQLite (OpenOffice Calc, Excel, rawgraphs.io, ...).

Les trois parties de ce chapitre couvrent les différents contextes d'utilisation. Dans un premier temps, à partir d'un choix de deux articles, tirés de la dernière publication coordonnée par Andrea Fabiano et Emanuele De Luca sur la Comédie-Italienne de Paris (FABIANO et DE LUCA, 2023), l'interface de consultation sera utilisée pour contribuer aux résultats présentés. La section suivante est dédiée à une observation sur comment deux chercheurs (un doctorant en histoire d'Ancien Régime, et un spécialiste de la Comédie-Italienne en France) utilisent cette interface de manière exploratoire. La troisième section, dédiée à la manipulation du fichier brut, est composée aussi de deux cas d'étude : le premier basé sur le travail de recherche effectué avec les deux stagiaires de l'Université Côte d'Azur, et le deuxième concernant un article de l'*Apothéose d'Arlequin* sur les transformations dans les pièces de Goldoni et Veronese.

### 3.1 Utilisation de l'interface

Les deux sections suivantes seront dédiées à des cas d'étude, basés sur la comparaison entre des ensembles de pièces. L'interface de consultation, depuis les fiches de la section « consultation », permet d'afficher sur un même graphique toutes les dates de représentation d'une ou plusieurs pièces, ainsi que de voir s'il s'agit de représentations lors de soirées composites. Le premier cas d'étude utilise cette fonctionnalité pour approfondir la contribution de Camilla Cederna à propos de la figure de la coquette dans la dramaturgie de la Comédie-Italienne de Paris (CEDERNA, 2023). Le deuxième mobilise ces graphiques, ainsi que divers

histogrammes de la partie « visualisation », pour étudier les parodies et les opéra-comiques à partir d'un article de Pauline Beaucé (BEAUCE, 2023).

### 3.1.1 Figure de la coquette

La contribution de Camilla Cederna au volume collectif *L'Apothéose d'Arlequin* (CEDERNA, 2023) analyse l'évolution de la figure de la coquette dans le Répertoire de la Comédie-Italienne de Paris, en relation aussi à son rôle au sein de la Comédie-Française. En observant un ensemble de pièces, la chercheuse montre un changement de jugement sur cette figure, au sein de la Comédie-Italienne de Paris, entre la tradition du XVII<sup>e</sup> et le début du XVIII<sup>e</sup> siècle, s'accompagnant d'un changement des éléments spectaculaires des pièces. La publication d'une monographie à ce sujet est prévue prochainement, et j'ai pu avoir accès à l'inédit (CEDERNA, à paraître), que la chercheuse a partagé avec moi lorsque je l'ai contactée en travaillant à ce cas d'étude.

Selon Cederna, suivant la tradition du XVII<sup>e</sup> siècle, à la Comédie-Italienne de Paris au début du XVIII<sup>e</sup> siècle la coquette apparaît d'abord comme un personnage positif, symbolisant « le théâtre comme jeu » et valorisant « la créativité et l'habileté des acteurs et actrices » (CEDERNA, 2023). Cela représente une défense du savoir-faire théâtral italien, en opposition à la tradition française, fondée sur la parole et sur la déclamation de faits moraux et exemplaires. La coquetterie est ainsi ce qui distingue et défend le jeu hétéroclite des comédiens italiens, des règles traditionnelles des comédiens français.

Mais au-delà des dynamiques de compétition entre les théâtres parisiens d'Ancien Régime, la coquette a représenté aussi « une forme de contre-pouvoir, une stratégie de rébellion, à la fois de défense et d'attaque, dans le cadre de la guerre des sexes » (ibid.) au sein de la Comédie-Italienne de Paris (CEDERNA, à paraître). Cela est le cas dans deux des pièces analysées en détail par Cederna : *La fausse coquette* et *Panurge à marier ou la coquetterie universelle*. La première est une pièce tirée du *Recueil Gherardi*, qui se situe donc en continuité avec le répertoire de l'ancienne troupe italienne à Paris. La deuxième, « dominée par le merveilleux, [...] comporte plusieurs références mythologiques et une forte composante métathéâtrale, et témoigne du goût pour la musique et pour le spectaculaire, typique de la tradition foraine et de la *commedia dell'arte* » (CEDERNA, 2023).

En revanche, dans *La veuve coquette*, troisième pièce analysée dans l'article, la coquette est « présenté[e] comme ayant un comportement déréglé et méritant, au mieux une correction, au pire une punition » (ibid.) Il s'agit d'une pièce ajoutée au Répertoire peu de temps après les deux premières, mais où la coquetterie devient un comportement repréhensible, symbole d'une société qui perd de vue la moralité au profit des apparences. Cederna montre ainsi, à travers la succession de ces pièces, une adhésion croissante du Répertoire italien à une moralité misogyne, utilisant des figures féminines pour représenter des phénomènes de débauche. La coquette n'est ainsi qu'une femme déplorable (comme c'est le cas au sein de la Comédie-Française), car elle ne respecte pas le rôle qui lui est traditionnellement assigné en société (par son âge, son statut ou simplement son genre). La figure perd donc son caractère contestataire et novateur, pour être réduite à un comportement moralement condamnable. Est-ce que nous pouvons observer ce changement de paradigme si l'on regarde toutes les représentations de ces trois pièces ? Dans notre interface il est possible de comparer les dates de toutes les représentations des trois pièces analysées par Camilla Cederna (*La fausse coquette*, *Panurge à marier* ou *la coquetterie universelle* et *La veuve coquette*), afin d'observer quels paradigmes prévalent à différentes époques.

En cherchant le mot « coquette » dans l'outil de recherche des pièces, on trouve six résultats : *La fausse coquette*, *Panurge à marier*, *La veuve coquette*, *La coquette fixée*, *La fermière coquette* et *La coquette de village*. Sauf *La fermière coquette*, elles sont toutes mentionnées dans l'article de Camilla Cederna ; les quatre premières sont présentes aussi dans le répertoire d'Emanuele De Luca. *La fermière coquette* a été jouée quinze fois en novembre et décembre 1756<sup>61</sup>, alors que *La coquette de village* a été représentée une seule fois, le 19 septembre 1771. *La coquette fixée* a été en revanche jouée 76 fois entre 1752 et 1780, et fait partie des pièces présentant une « stratégie de correction » (ibid.) de la coquetterie.

Concernant les pièces analysées dans l'article de Cederna, le Répertoire ne présente qu'une seule représentation de *La fausse coquette*, le 9 avril 1720, sans d'autres pièces en programme pour ce jour-là. *Panurge à marier*, en revanche, apparaît cinq fois dans les séances, entre 1720 et 1722, dont quatre fois accompagnée par une autre pièce. *La veuve coquette* est parmi les trois celle qui a été la plus jouée : 112 fois, entre 1721 et 1748, le plus souvent lors

---

<sup>61</sup> La première de cette pièce, selon les informations saisies depuis Brenner, aurait eu lieu le 14 novembre 1756 et elle devrait donc faire partie du répertoire d'Emanuele De Luca. Or ce dernier ne mentionne aucune nouvelle pièce entre le 11 octobre (*L'imagination*) et le 17 novembre 1756 (*Le charlatan*).

de séances où sont jouées deux pièces. Ci-dessous (Figure 29) un graphique avec la chronologie des représentations des trois pièces.

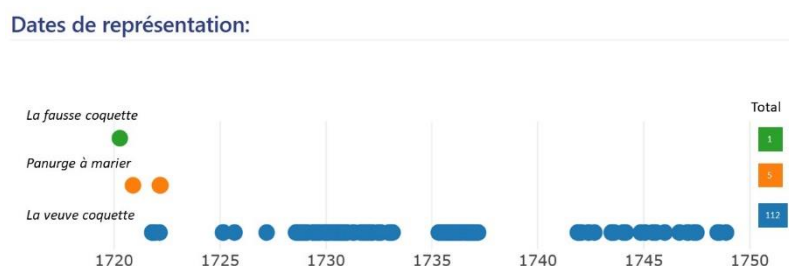


Figure 29 : chronologie des représentations de *La fausse coquette*, *Panurge à marier* et *La veuve coquette*.

Ce graphique peut être séparé en trois phases :

1. entre le 9 avril 1720 et le 12 mars 1722 il y a des représentations des trois pièces (Figure 30) ;
2. entre le 13 mars 1722 et le 18 février 1725 aucune pièce n'est représentée (Figure 31) ;
3. à partir du 19 février 1725 et jusqu'à la dernière représentation de *La veuve coquette* le 20 novembre 1748, il n'y a de représentations que de cette dernière (Figure 32).

### Dates de représentation:

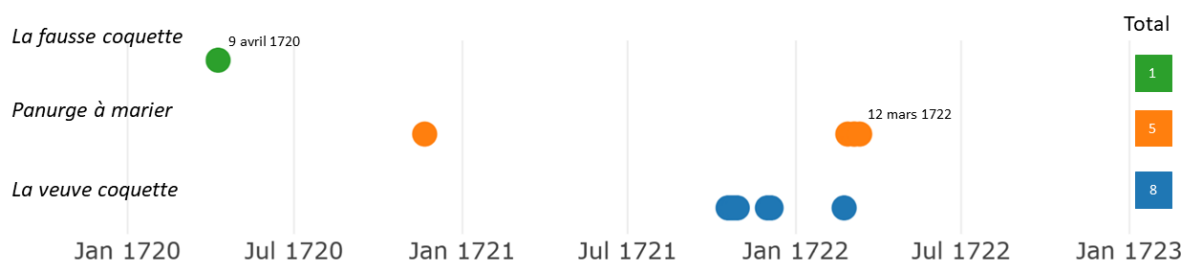


Figure 30 : détail de la première phase.

### Dates de représentation:

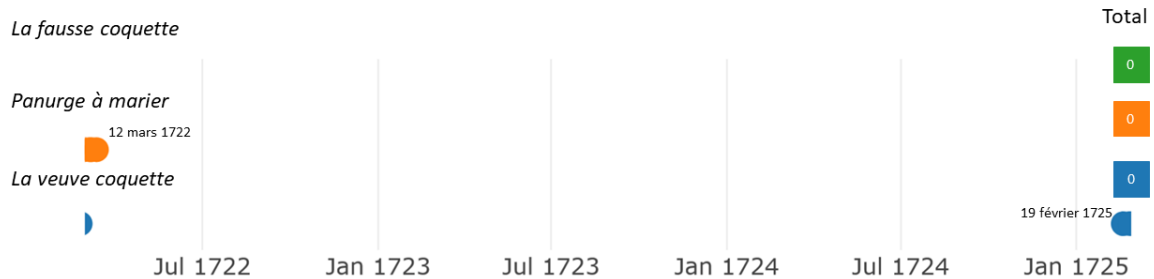


Figure 31 : détail de la deuxième phase.

### Dates de représentation:

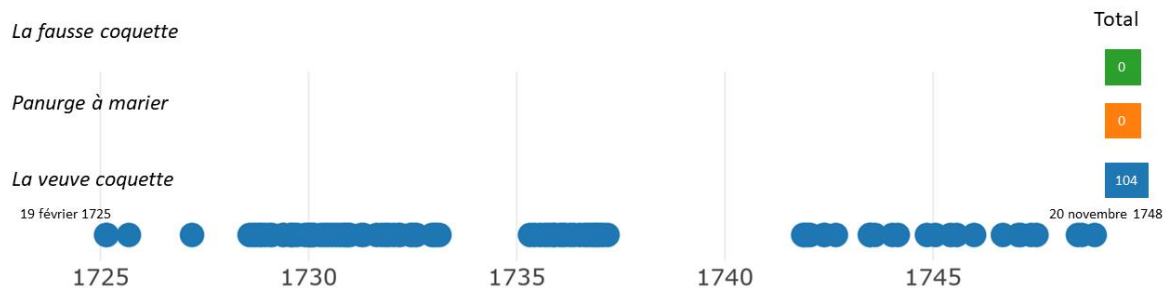


Figure 32 : détail de la troisième phase.

Par rapport à la disponibilité des sources (les registres de la compagnie, pour un schéma général voir la Figure 33), la première phase est marquée par une absence partielle de ces documents, qui ont été perdus pour la saison 1720/21. La représentation de *La fausse coquette* du 9 avril 1720 et celle de *Panurge à marier* du 21 novembre 1720 correspondent aux premières, connues grâce à l'étude des revues. A partir de la saison 1721/22, le Répertoire mentionne quatre représentations de *Panurge à marier* et huit de *La veuve coquette*, et à partir de la saison 1722/23 jusqu'à 1724/25 plus aucune représentation de ces trois pièces n'est répertoriée, alors que les registres sont disponibles.

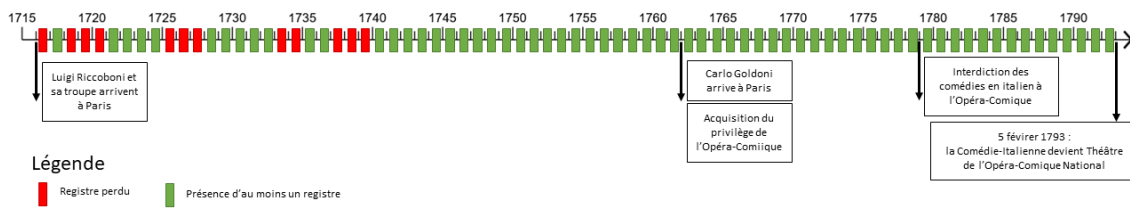


Figure 33 : présence et absence des Registres de la compagnie au XVIIIe siècle.

La troisième phase est aussi caractérisée par une absence partielle des registres. Néanmoins, une forme d'affirmation du succès de *La veuve coquette* peut être observée : elle reste la seule pièce documentée sur toutes ces années, avec un nombre de représentations considérable. Un plus grand succès de *La veuve coquette* est par ailleurs documenté par Cederna à travers les revues de l'époque. *Panurge à marier* « fut sifflée lors de la première, seuls le prologue et le premier acte de cette comédie burlesque eurent un peu de succès. Ce qui poussa l'auteur à supprimer le deuxième et le troisième acte dès la deuxième représentation. » (ibid.), alors que *La veuve coquette* « eut beaucoup de succès et fut représentée plusieurs fois jusqu'en 1748. Son succès est confirmé par d'Origny [...] Gueullette exprima un jugement positif à l'égard de la pièce [...] Totalement négatif, en revanche, est l'avis de d'Argenson. »<sup>62</sup> (ibid.).

#### Dates de représentation:

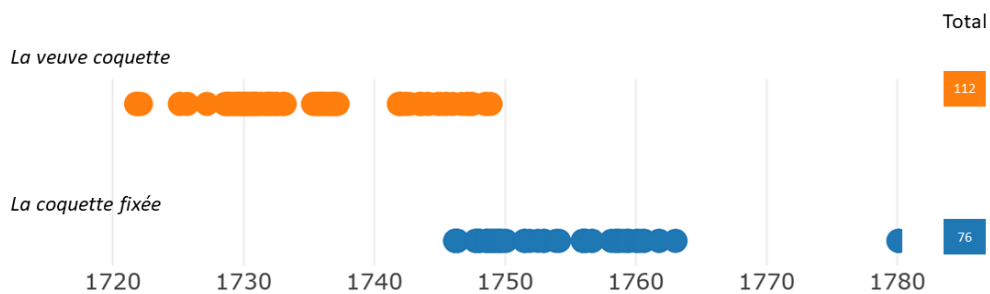


Figure 34 : représentations de *La veuve coquette* et de *La coquette fixée*.

L'étude qualitative de Camilla Cederna semble ainsi être confirmée par l'observation de notre base de données, avec l'affirmation de la pièce relevant du paradigme basé sur la correction ou la punition de la coquette. Si l'on étend l'observation à *La coquette fixée* (Figure

<sup>62</sup> Cederna tire les dates des représentations du répertoire de Brenner de 1961.

34), dont de nombreuses représentations sont répertoriées, ces pièces apparaissent tout au long du XVIII<sup>e</sup> siècle. L'étude du Répertoire suggère en plus l'existence de trois phases dans l'affirmation du paradigme de la correction de la coquette, dont la première pourrait correspondre au moment de l'expérimentation dramaturgique et spectaculaire, et la troisième à celui du développement du paradigme qui a eu le plus de succès. Mais que se passe-t-il dans la deuxième phase (1722-1725), lorsqu'aucune représentation de ces trois pièces n'est présente, avant que seule *La veuve coquette* reste active dans le Répertoire ? Après cette étude quantitative (réalisée à partir d'une recherche en lettres), le retour à une analyse qualitative et historiquement informée permettrait de conclure sur ces hypothèses ainsi que d'approfondir la signification de la deuxième phase.

Par ailleurs, dans l'introduction au répertoire PDF d'Emanuele De Luca (DE LUCA, 2011), le chercheur mentionne un épisode significatif pour l'histoire de la compagnie en ce qui concerne le rôle des femmes : le 27 avril 1729, c'est pour la première fois une actrice, Silvia, qui fait un compliment au public en ouverture des spectacles. Le compliment de Silvia de 1729 met explicitement en relation le nouveau rôle des femmes au sein de la société théâtrale, où elles s'émancipent en vertu des principes d'égalité. Selon De Luca, cet épisode marque une nouvelle époque pour la compagnie, qui est devenue depuis peu (suite au départ des Riccoboni) une société d'acteurs, régie sur l'égalité entre ses membres. En effet, depuis l'établissement de la troupe à Paris en 1716 et jusqu'en 1729, un article du statut de la Comédie-Italienne (cité par De Luca) limitait grandement la participation des femmes aux assemblées de la compagnie.

Ainsi, si d'un côté les pièces analysées véhiculent une critique misogyne de la société à travers la figure négative de la coquette, de l'autre côté les femmes ne seraient plus exclues du pouvoir en ce qui concerne l'organisation de la troupe. Ces deux tendances ne sont pas a priori incompatibles, mais il est tout de même pertinent de se demander si cette forme de prise de pouvoir des femmes au sein de la compagnie ne se reflète pas dans des choix dramaturgiques. Notre interface pourrait fournir des pistes de réponses à cette problématique, si l'on élargit la comparaison à des pièces dont le sujet concerne les femmes et qui soient représentatives des différentes visions du rôle des femmes dans la société.

### **3.1.2 Parodie et Opéra-Comique**

Ce deuxième cas d'étude est focalisé sur l'un des aspects les plus problématiques de notre travail : la classification des genres des pièces. Dans un souci de lisibilité des graphiques,

nous avons défini seulement une dizaine de genres (définis dans la partie Les tables implémentées) qui peuvent être attribués aux pièces. Néanmoins, comme toute classification (BOWKER et LEIGH STAR, 2023), elle est arbitraire dans la mesure où elle dépend des objectifs et des présupposés de la recherche. Ce cas d'étude permettra, dans ce sens, d'observer certaines limites de notre modélisation.

Parmi les choix qui ont été faits, un même genre « parodie » réunit les parodies parlées, chantées et les parodies-traduction (le type de parodie, si connu, est indiqué au début de la note entre crochets, et n'est donc pas visible dans l'interface de consultation), mais celles-ci sont distinguées des « opéra-comiques ». Cela reflète un contexte historique, mais aussi des objectifs de recherche. En effet, les opéra-comiques peuvent être un type spécifique de parodie, qui acquière une autonomie au cours du siècle devenant un véritable genre dramatique (BEAUCE, 2013) : le choix de les distinguer des autres parodies relève justement de la volonté de montrer l'affirmation de ce nouveau genre. Une autre raison, opérationnelle, de cette distinction concerne l'existence de la base de données Opéra-Comique Database : avoir un genre « opéra-comique » favorise la mise en relation des deux projets, car cela permet de savoir immédiatement que les pièces ainsi catégorisées à partir de 1780 font aussi partie de l'autre base de données.

L'étude des parodies et des opéra-comiques constitue par ailleurs un intérêt particulier concernant la Comédie-Italienne de Paris : « si l'histoire de la parodie d'opéra ne se réduit en rien à son fonctionnement à la Comédie-Italienne, en revanche elle passe par celle du compagnonnage avec ce théâtre : c'est là que le genre gagne en légitimité, trouve les moyens de sa théorisation, de sa pérennisation » (BEAUCE, 2023). Beucé justifie cette affirmation avec l'appui d'une étude quantitative sur les créations de parodies à la Comédie-Italienne de Paris et dans d'autres théâtres au XVIII<sup>e</sup> siècle, ainsi que sur des considérations concernant les deux éditions (de 1731 et de 1738) des *Parodies du nouveau théâtre italien*. Notre base de données, si elle ne contient aucune information sur les théâtres de la Foire ni sur les autres scènes prises en compte par la chercheuse, permet néanmoins d'étendre l'étude à toutes les représentations de la Comédie-Italienne, et de les mettre en relation avec les créations. Pour ce cas d'étude, l'interface de consultation est utilisée pour approfondir les phénomènes étudiés par Beucé, d'abord concernant les éditions, ensuite concernant la programmation. Cela permettra aussi de dégager certaines limites de notre modélisation des genres des pièces.

### 3.1.2.1 Les éditions des parodies

« Les *Parodies du nouveau Théâtre italien* répondent à une volonté précise : compléter l’anthologie du répertoire de la Comédie-Italienne [...] stratégie financière de la part de l’éditeur, mais qui répond aussi au goût du public, des lecteurs, et correspond à une vogue du genre sur la scène. L’hégémonie des parodies d’opéras est frappante dans la première édition de 1731, en trois volumes : vingt-quatre pièces éditées dont quinze parodies d’opéras, cinq parodies de tragédies, deux parodies d’intermèdes italiens, deux comédies critiques. En 1738, la seconde édition des *Parodies* comporte un volume de plus, pour trente et une pièces dont vingt et une parodies d’opéras, huit parodies de tragédies, une parodie de comédie et une parodie d’intermède italien » (ibid.).

Ci-dessous un tableau avec les titres des pièces des deux éditions, toutes présentes dans notre base de données (en gris celles décrites comme étant des « parodies d’opéra » dans l’édition de 1738).<sup>63</sup>

1731	1738
<i>Œdipe travesti</i>	<i>Œdipe travesti</i>
<i>La rupture du carnaval et de la folie</i>	<i>La rupture du carnaval et de la folie</i>
<i>Artémire</i>	<i>Artémire</i>
<i>Hercule filant</i>	<i>Hercule filant</i>
<i>Arlequin Persée</i>	<i>Arlequin Persée</i>
<i>Le serdeau des théâtres</i>	<i>Le serdeau des théâtres</i>
<i>Parodie, tragédie</i>	<i>Parodie, tragédie</i>
<i>Agnès de Chaillot</i>	<i>Agnès de Chaillot</i>
<i>Noces d’Arlequin et de Silvia ou Théthys et Pelée travestis</i>	<i>Noces d’Arlequin et de Silvia ou Théthys et Pelée travestis</i>
<i>Le chaos</i>	<i>Le chaos</i>
<i>Amadis le cadet</i>	<i>Amadis le cadet</i>
<i>Armide</i>	<i>Armide</i>
<i>Le mauvais ménage</i>	<i>Le mauvais ménage</i>

<sup>63</sup> L’édition de 1738 disponible sur Gallica contient en réalité 32 pièces. Suivant la description des pièces présentes dans cette édition, et pour garantir la cohérence du genre attribué à chaque pièce dans les deux éditions, j’ai trouvé seulement dix-neuf parodies d’opéra dans celle de 1738. De plus, si les deux « comédies critiques » de l’édition de 1731 sont *Le serdeau des théâtres* et *Parodie, tragi-comédie*, comme je déduis de l’édition de 1738, la distribution des genres telle que décrite par Pauline Beaucé n’est pas cohérente entre les deux éditions.

<i>Momus exilé</i>	<i>Momus exilé</i>
<i>Arlequin Atys</i>	<i>Arlequin Atys</i>
<i>Le chevalier errant</i>	<i>Le chevalier errant</i>
<i>Arlequin Roland</i>	<i>Arlequin Roland</i>
<i>Arlequin Bellérophon</i>	<i>Arlequin Bellérophon</i>
<i>La méchante femme</i>	<i>La méchante femme</i>
<i>Arlequin Tancredi</i>	<i>Arlequin Tancredi</i>
<i>Alceste</i>	<i>Alceste</i>
<i>Le joueur</i>	<i>Le joueur</i>
<i>Don Micco et Lesbina</i>	<i>Don Micco et Lesbina</i>
<i>Hésione</i>	<i>Hésione</i>
	<i>Les enfants trouvées</i>
	<i>Bolus</i>
	<i>Arlequin au Parnasse ou les folies de Melpomène</i>
	<i>Médée et Jason</i>
	<i>L'île de la folie</i>
	<i>Pyrame et Thisbé</i>
	<i>La bonne femme</i>
	<i>Phaéton</i>

Peut-on observer des changements dans la programmation de ces pièces avant et après les deux éditions ? Des éléments de réponse à cette question seront donnés à partir des graphiques qui affichent la chronologie des représentations des pièces concernées. La Figure 35 représente les pièces de l'édition de 1731 ; pour des raisons de lisibilité, et puisque toutes les parodies d'opéra de la première édition sont aussi dans la deuxième, la Figure 36 ne contient que celles ajoutées à l'édition de 1738.

## Dates de représentation:

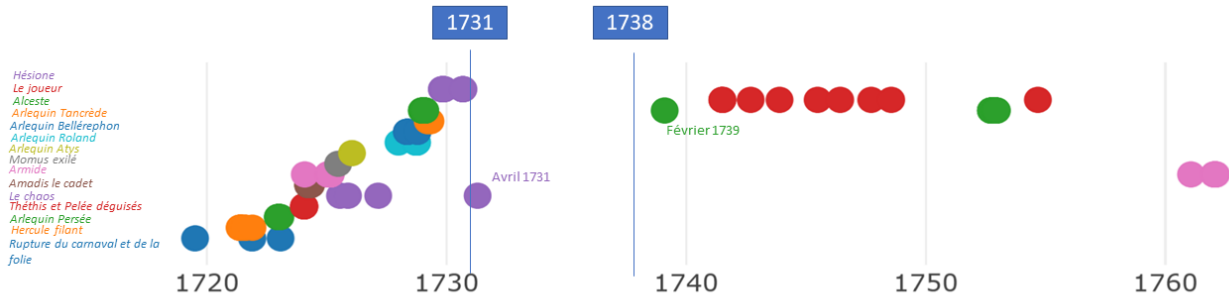


Figure 35 : Représentations des parodies d’opéra présentes dans l’édition des *Parodies du nouveau théâtre italien* de 1731.

## Dates de représentation:

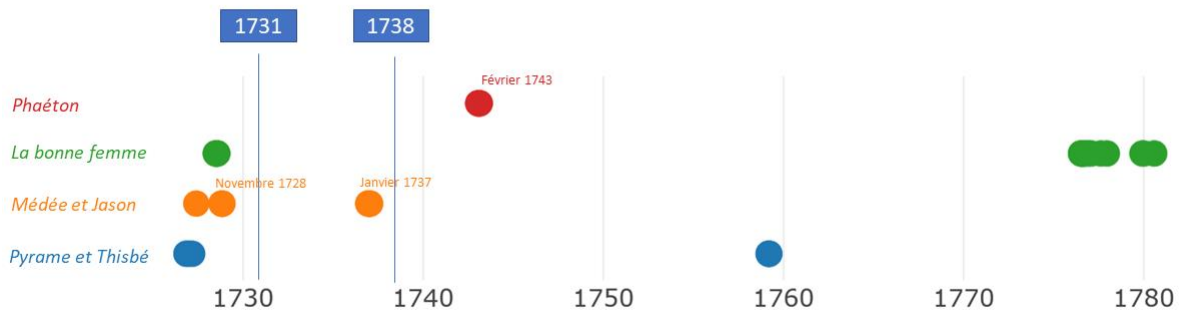


Figure 36 : Représentations des parodies d’opéra ajoutées à l’édition des *Parodies du nouveau théâtre italien* de 1738.

Encore une fois, il a été nécessaire de recourir à ce genre de graphique, car les pièces sont toutes catégorisées comme des « parodies » dans la base de données, par conséquent il n’est pas possible de distinguer des types différents de parodies au niveau des visualisations dans l’interface (à ce stade du travail). Il s’agit d’une première limite de notre catégorisation : s’il est possible de retrouver le type de parodie (parlé, chanté ou traduction) au niveau de la note textuelle associée à la pièce, cette information ne peut pas être manipulée au niveau du genre dans l’interface. Autrement dit, à l’heure actuelle, il est nécessaire d’avoir le fichier brut de la base de données et de savoir le manipuler, afin de pouvoir distinguer automatiquement différents types de parodies. Pour ce cas d’étude, des pièces qui sont décrites comme des

« parodies d’opéra » dans leur édition, sont étiquetées dans la base de comme de simples « parodies », ce qui montre explicitement l’ambiguïté de la division en genres (ROUGEMONT, 2001 ; BEAUCE, 2013) et la non-universalité des choix de modélisation.

Malgré ces limites, un même phénomène peut être observé dans les deux visualisations (Figure 35 et Figure 36). En effet, les représentations des pièces éditées semblent s’interrompre (ou presque) entre les deux éditions, et reprendre pour un nombre limité de titres après 1738. Des dix-neuf pièces considérées, seulement six continuent d’être jouées après 1740, équivalant à moins qu’un tiers des titres édités ; si l’on se limite aux pièces de l’édition de 1731, cette proportion descend à un cinquième (trois pièces sur quinze sont toujours jouées après 1740). Est-ce du hasard, ou est-ce que cela suggère une certaine dynamique entre pièce jouée et pièce éditée ? La réponse à cette question ne peut pas être donnée uniquement au moyen d’une étude quantitative, mais celle-ci peut fournir des éléments de réponse, ainsi que des pistes de recherche pour répondre de manière plus précise.

Une première recherche de preuves peut être faite au niveau du genre « parodie » : est-ce que les autres parodies ont aussi été moins jouées entre 1731 et 1738, ou est-ce que ce phénomène concerne uniquement les parodies d’opéra des deux éditions des *Parodies* ? Les graphiques des représentations vont d’abord être étendus à toutes les pièces éditées en 1738 (Figure 37), avant de passer à l’observation de l’évolution du genre dans le graphique chronologique de la section « pièces » (Figure 38).

### Dates de représentation:

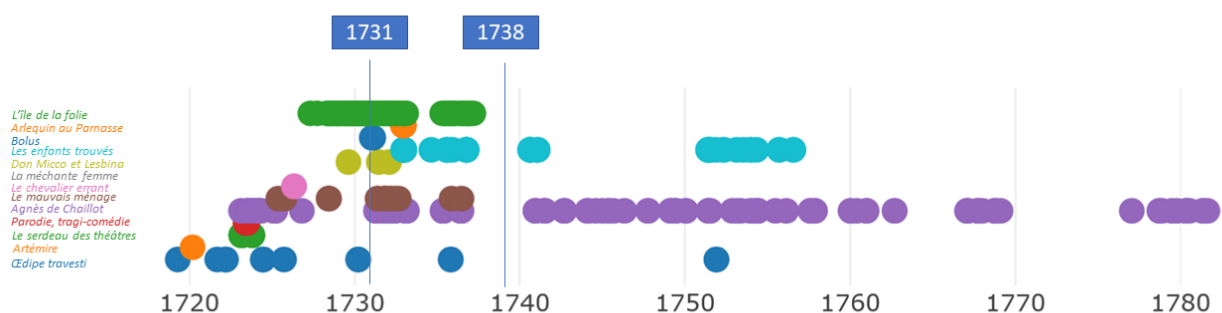


Figure 37 : Représentations des pièces éditées dans les *Parodies du nouveau théâtre italien* (1738), non catégorisées comme des « parodies d’opéra » per l’éditeur.

Ce nouveau graphique (Figure 37) ne présente pas d'interruptions dans la programmation des titres entre 1731 et 1738, mais, dans ce cas aussi, seulement trois pièces restent à l'affiche après 1740, ce qui équivaut à moins qu'un quart de l'ensemble. Une plus courte interruption des représentations est en réalité présente, entre mars 1737 et août 1740, mais cela correspond à une absence des registres (Figure 33), et elle n'est donc pas significative. Au contraire, le manque de représentations des pièces dans les deux premiers graphiques (Figure 35 et Figure 36) ne peut pas s'expliquer uniquement par une lacune dans les sources, qui sont disponibles sur la majorité de la période entre 1731 et 1737.

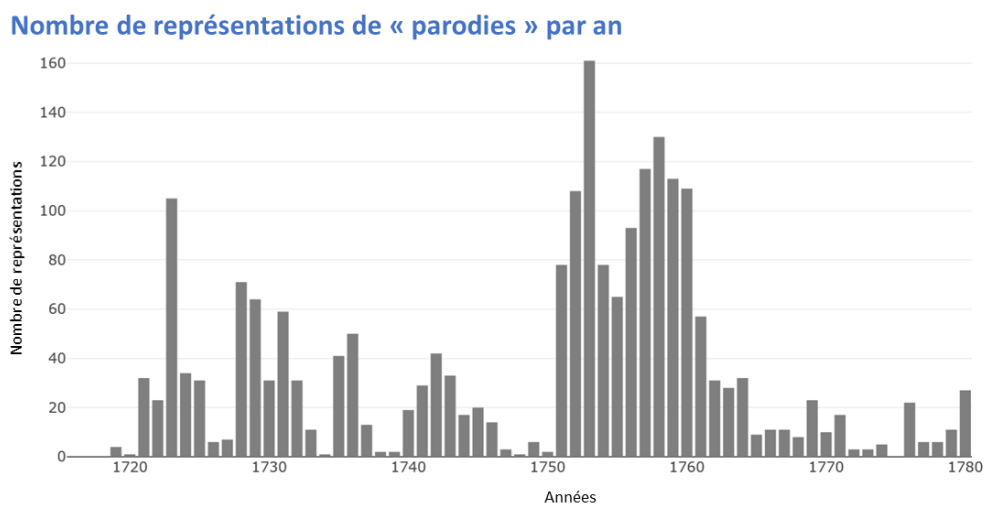


Figure 38 : Nombre de représentations par an des pièces catégorisées comme des « parodies » dans notre base de données.

Si l'on observe toutes les représentations des pièces qui sont catégorisées comme des « parodies » dans notre base de données (Figure 38), globalement, les diminutions dans le nombre des dates correspondent aux lacunes dans les registres (Figure 33). Entre 1745 et 1751, la parodie est interdite aux italiens (DE LUCA, 2011 ; DE LUCA, 2015), ce qui justifie une nette diminution de ce genre entre ces deux dates. Or, le graphique de la Figure 38 permet uniquement de connaître le nombre de représentations des pièces catégorisées comme « parodies », mais ne donne aucune information sur la persistance de chacune dans le Répertoire, ni sur le type de parodie dont il s'agit. Néanmoins, dans l'interface, il est possible d'afficher les titres des pièces en question dans un cadre à côté du graphique, en cliquant sur la colonne dont on cherche les détails. Cette liste de titres permet d'élargir l'étude des représentations, au moyen de l'interface ou du fichier de la base de données.

Un phénomène ultérieur qu'il serait nécessaire d'aborder pour étudier la dynamique entre éditions et représentations, concerne l'analyse d'autres éditions : est-ce que, pour d'autres pièces de la Comédie-Italienne de Paris, on observe une interruption des représentations autour du moment de l'édition ? Est-ce que, pour d'autres ensembles de pièces éditées, seulement certaines restent actives dans le Répertoire après la publication du texte ? Pour chercher des réponses précises, une possibilité consiste à établir des corpus représentatifs de l'ensemble des genres et couvrant le siècle entier, afin d'analyser les pièces concernées via l'interface de consultation. Mais surtout, ce cas d'étude devrait être confirmé par des arguments historiques, qui justifieraient réellement un phénomène de disparition des pièces du Répertoire, lorsqu'elles sont éditées. A propos des deux volumes des *Parodies*, Beaucé se demande : « Quelle est la valeur de ces éditions ? Elles fixent une pratique et un répertoire, pérennisent un genre. Les éditions permettent aussi les reprises de telle ou telle pièce en province ou à l'étranger. » (BEAUCE, 2023) Est-ce que l'édition d'une pièce – en fixant et en diffusant un texte qui est jusqu'à ce jour uniquement accessible à travers des acteurs.ices sur une scène – signifie la fin de sa vie dans le théâtre où elle a été créée ? Ou est-ce qu'éditer une pièce sert pour garder en vie un spectacle qui a perdu l'intérêt du public ?

A l'état actuel les observations faites dans cette étude de cas pourraient en réalité ne pas dépendre de la présence d'éditions, mais simplement d'une progressive naturelle perte d'intérêt du public, ou d'un changement dans la stratégie dramaturgique dû à des événements politiques ou culturels extérieurs. En particulier, les parodies ont une durée de vie nécessairement limitée, car elle dépend de la pièce parodiée (DE LUCA, 2015). Avec la base de données, il est possible de mettre en relation des parodies avec les représentations des pièces sources, qui peuvent être trouvées dans d'autres base de données, comme CESAR, les Registres de la Comédie-Française ou les Archives de l'Opéra-Comique. Connaissant les dates des représentations, ces bases de données permettent de comparer les périodes auxquelles sont jouées les pièces et leurs parodies.

### **3.1.2.2 Programmation des parodies et des opéra-comiques**

Cette dernière section montre ce que l'interface peut apporter comme précision concernant l'affirmation de l'opéra-comique comme genre dramatique. Mais avant de rentrer dans le vif de l'analyse, quelques considérations sur les données utilisées par Pauline Beaucé : pour l'analyse quantitative, elle se base sur des données concernant environ 260 titres de parodies, qu'elle a elle-même recensé, sur une période « entre 1709, date de la première parodie dramatique d'opéra au XVIIIe siècle sur un théâtre forain, et 1791 » (BEAUCE, 2023). Cela

est le résultat d'un travail de recherche personnelle de nombreuses années, dont l'article ne précise pas la méthode de collecte, ni de modélisation, des données. Les métadonnées recueillies pour chaque pièce peuvent en partie être déduites à partir des visualisations de données (l'année et le lieu de la première, par exemple, doivent avoir été recensés pour produire le premier graphique sur le nombre de créations de parodies dramatiques d'opéras dans les divers théâtres parisiens, BEAUCE, 2023), mais ne sont pas explicitées. De l'autre côté, notre base de données est exhaustive par rapport aux sources disponibles, et elle est construite avec une méthode de travail collective, décrite en introduction. En revanche, elle couvre une période plus courte, qui va de 1716 à 1780. Par rapport à ce cas d'étude, nous allons donc considérer que notre base de données contient un sous-ensemble des données recueillies par Beaucé – autrement dit, les deux jeux de données forment une intersection, comme schématisé dans la Figure 39.

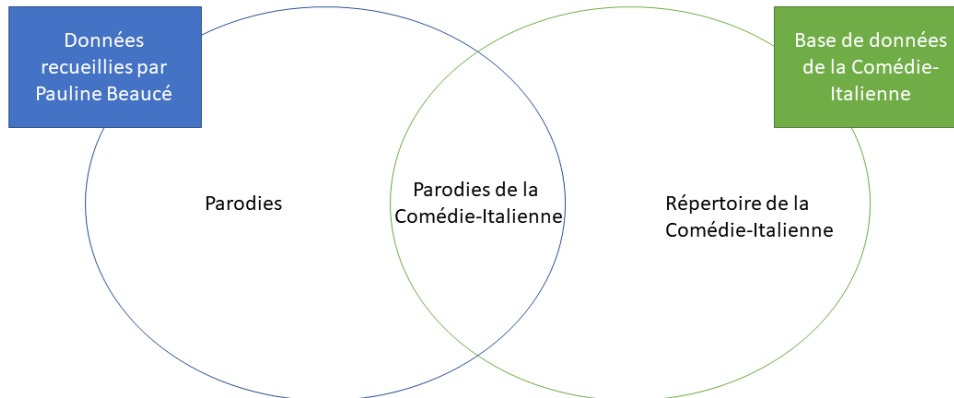


Figure 39 : Intersection schématisant la relation entre les données recueillies par Beaucé et notre base de données.

Pour l'affirmation de l'opéra-comique comme genre, la décennie 1765-1774 est cruciale : à cette période

« la Comédie-Italienne n'accueille aucune création de parodie d'opéra. Comme l'a notamment expliqué Andrea Fabiano, cette période voit le recul des parodies d'opéras, porteuses de vaudeville, au profit du nouvel opéra-comique, ceci afin d'assurer la refonte du répertoire de la

Comédie-Italienne. En fait, si aucune parodie d'opéra n'est créée à la Comédie-Italienne durant cette décade, de nombreuses pièces de ce genre sont reprises et continuent d'alimenter le goût pour les vaudevilles et le comique parodique. » (BEAUCE, 2023)

Est-ce que dans notre base de données il est possible d'observer ces phénomènes ? Les données telles que nous les avons créées, permettent-elles d'observer un processus d'affirmation de l'opéra-comique sur les parodies au sein de la Comédie-Italienne de Paris ? Des éléments de réponse à cette question peuvent être trouvés dans l'interface de consultation, en analysant les graphiques de la section « visualisation » qui concernent la distribution chronologique des représentations (section « pièces ») et des premières (section « premières ») par genres, avec une attention particulière aux années 1765-1774. Finalement, les données inconnues (qui peuvent être visualisées dans l'interface) seront prises en compte afin d'évaluer les résultats.

#### Représentations des parodies et des opéra-comiques par an

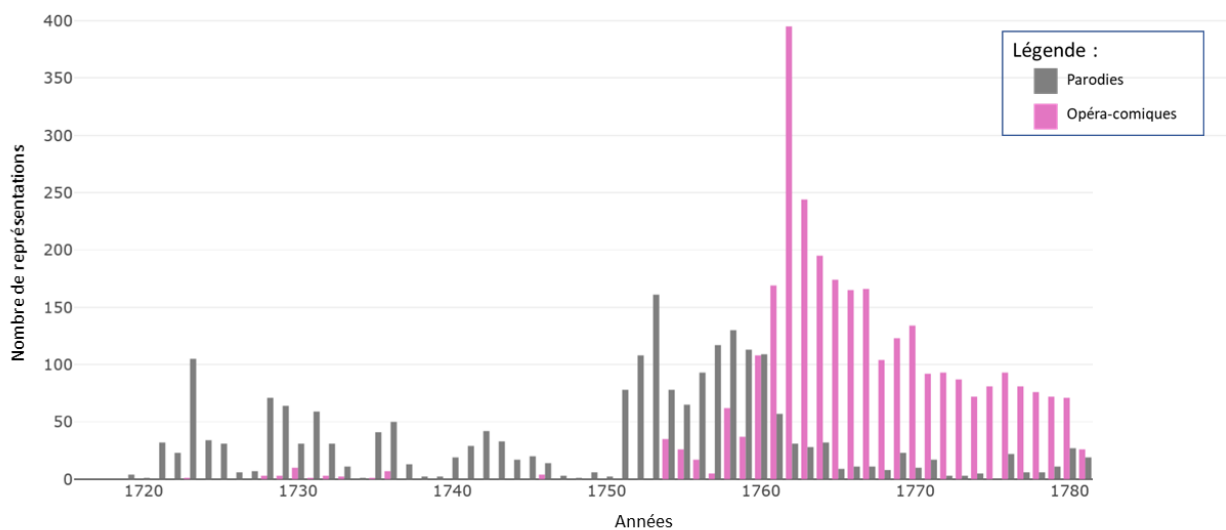


Figure 40 : Nombre de représentations de parodies et d'opéra-comiques par an.

## Créations de parodies et d'opéra-comiques par an

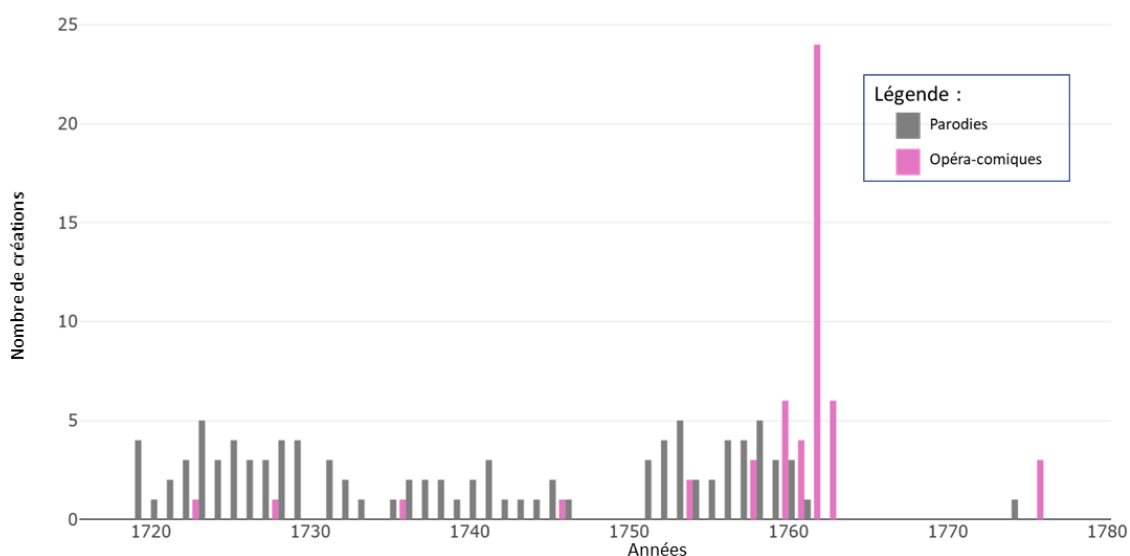


Figure 41 : Nombre de créations de parodies et d'opéra-comiques par an.

Ces deux derniers graphiques (Figure 40 et Figure 41) confirment la citation ci-dessus. En effet, aucune création de parodie n'est répertoriée entre 1762 et 1773, alors que les représentations persistent et que les opéra-comiques connaissent leur pic de représentations et de créations en 1762. Exception faite pour les années 1769-1772, les représentations des pièces catégorisées comme des « parodies » ne font que diminuer à partir des années 1760. Dans la Figure 40, entre 1750 et 1762 on observe une véritable inversion de tendance concernant les représentations : par rapport à la première moitié du siècle, il y a une évidente augmentation du nombre de dates où sont jouées des parodies, avec un maximum absolu de 161 représentations en 1753 ; ensuite, le nombre de représentations reste relativement élevé (par rapport à la première moitié du siècle). Or, à partir de 1754, on observe une nette augmentation des pièces catégorisées comme « opéra-comique », qui n'étaient presque pas présentes dans la première moitié du siècle. A partir de cette année-là, globalement, le nombre de représentations d'opéra-comiques augmente, jusqu'à atteindre le même niveau que les parodies en 1760, et à le dépasser largement en 1761, lorsqu'il y a 169 dates avec un opéra-comique, contre 57 avec une parodie (chiffre qui reste élevé par rapport à la première moitié du siècle, où l'histogramme des parodies dépasse les 50 représentations seulement quatre fois, en 1723, 1728, 1729 et 1731).

## Représentations des parodies et des opéra-comiques par an (avec données inconnues)

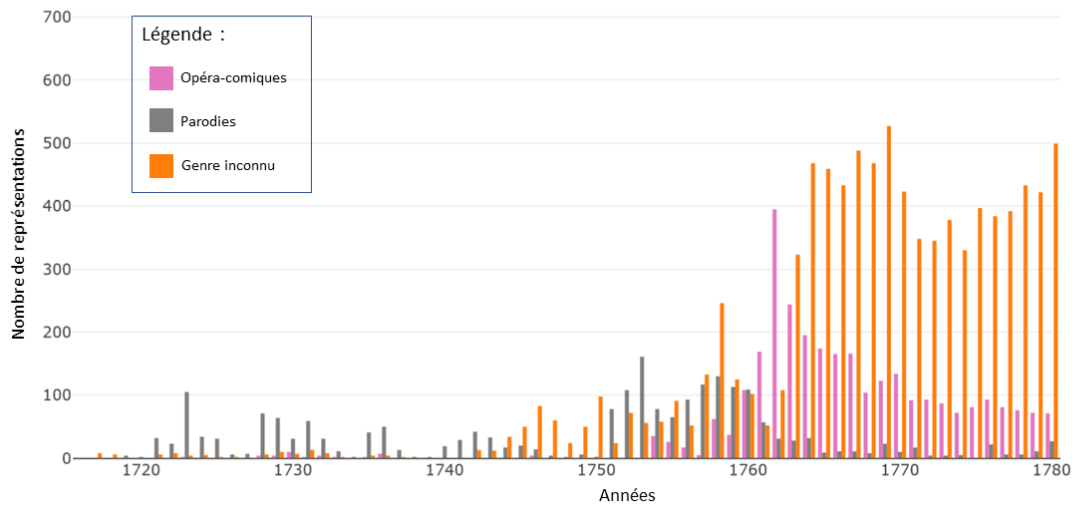
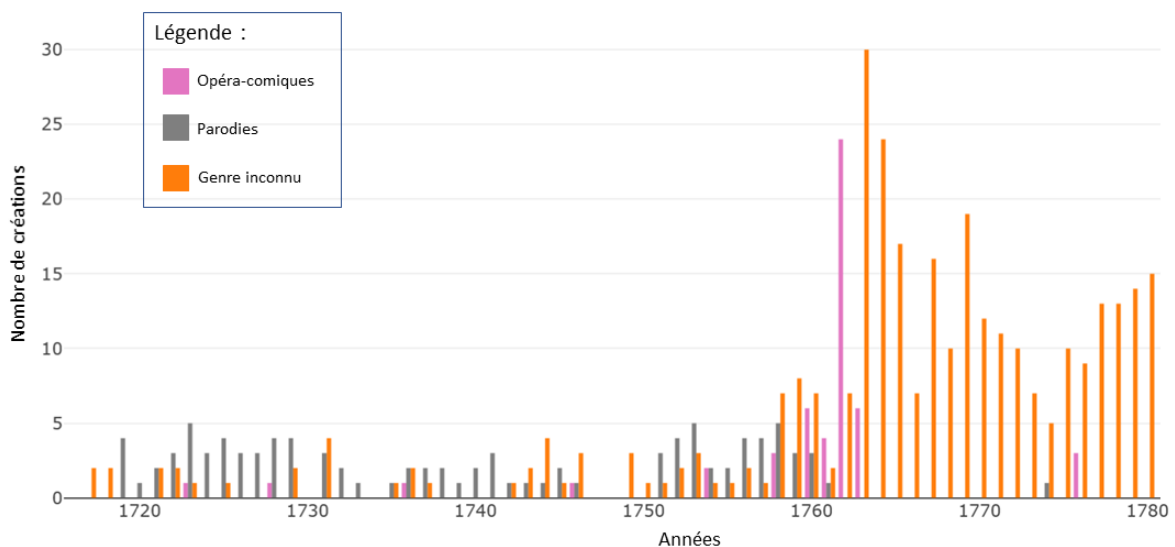


Figure 42 : Nombre de représentations de parodies, d’opéra-comiques et de pièces dont on ne connaît pas, à ce jour, le genre<sup>64</sup>.

## Créations de parodies et d’opéra-comiques par an (avec données inconnues)



<sup>64</sup> Les colonnes orange correspondent à des pièces présentes dans la base de données mais auxquelles nous n’avons pas encore pu attribuer un genre. A priori nous ne savons donc pas de quelles pièces il s’agit. Il est tout de même possible de cliquer sur les colonnes afin de connaître la liste de pièces auxquelles la colonne fait référence.

Figure 43 : Nombre de créations de parodies, d'opéra-comiques et de pièces dont on ne connaît pas, à ce jour, le genre.

Néanmoins, les données inconnues restent très nombreuses, surtout dans la deuxième moitié du siècle, comme le montrent la Figure 42 et la Figure 43. La quantité de données non catégorisées est si élevée que l'attribution d'un genre à toutes les pièces pourrait changer remarquablement la forme des deux graphiques précédents (Figure 40 et Figure 41). Afin de conclure sur toutes les observations que nous venons de faire, il sera donc nécessaire d'avoir des données plus fiables.

### 3.2 Utilisation de l'interface par des spécialistes

Après avoir donné des exemples d'utilisation de la base de données à travers l'interface, cette section sera dédiée à observer comment des spécialistes en humanités abordent la plateforme de consultation. Pour des raisons de temps et d'organisation, ce test a été réalisé avec deux chercheurs sur une version non finalisée de l'interface. Les chercheurs ont été choisis pour représenter des profils différents, mais au même temps réellement susceptibles de s'intéresser aux données. La première personne est un spécialiste de la Comédie-Italienne en France, Renzo Guardenti, professeur d'études théâtrales à l'Université de Florence ; la deuxième est un doctorant en histoire française d'Ancien Régime à l'Université de Milan, Bernardo Agnelli. Les recherches de Guardenti sont très proches de celles des membres de notre équipe (voir par exemple GUARDENTI 1990 et 2023) ; il connaît surtout en profondeur le répertoire d'Emanuele De Luca, dont il a suivi la toute première réalisation, durant la direction du mémoire de Master 2 (*tesi di laurea magistrale*) de De Luca, à l'Université de Florence. Les recherches d'Agnelli portent sur la police et sur la presse d'Ancien Régime en France. Ils sont donc, tous les deux, susceptibles d'utiliser notre interface de consultation pour leurs recherches, mais avec des objectifs et des méthodes différentes. Le choix de ces deux chercheurs permet de couvrir par ailleurs des profils différents en termes de familiarité avec les humanités numériques : Renzo Guardenti a dirigé la publication de la base de données iconographique Dyonisos (GUARDENTI, 2021), travail pionnier en humanités numériques et études théâtrales, qui n'est actuellement plus disponible à cause de l'obsolescence des technologies utilisées. Bernardo Agnelli mène en revanche des recherches d'archive avec une approche micro-historique et n'a jamais participé à un projet d'humanités numériques.

### 3.2.1 Déroulement des entretiens

Les deux entretiens se sont déroulés séparément et ont duré entre une et deux heures. La consultation a été faite depuis mon ordinateur professionnel, puisque l'interface n'était pas encore consultable en ligne, et s'est déroulée ainsi : après une brève introduction générale de ma part sur l'état de l'interface et des contenus présents dans la version finale (qui avait vocation à substituer les images et le texte qui n'étaient pas encore présents dans la page initiale), les deux chercheurs ont exploré la plateforme librement. A la fin de leur exploration, je leur ai montré les fonctionnalités qu'ils n'avaient pas remarqué afin qu'ils les observent. Chaque entretien s'est terminé sur une discussion de bilan général de l'outil.

D'une façon générale, les deux chercheurs ont approché l'interface de manières différentes, à partir de leurs attentes : Guardenti, connaissant en profondeur le répertoire de De Luca et plus largement notre Répertoire, s'attendait à trouver certaines informations. Il a cherché des données spécifiques liées à ses recherches, et il a mis très rapidement l'interface en relation avec la version PDF du répertoire de De Luca, ainsi qu'avec la base Dyonisos (consultable avec un émulateur depuis son ordinateur professionnel). Agnelli, en revanche, a utilisé l'interface de manière plus exploratoire et en observant des périodes qui l'intéressaient par ailleurs (celle des années autour de la Révolution de 1789 par exemple). Autrement dit, Guardenti s'est focalisé sur le Répertoire, alors qu'Agnelli s'est concentré sur le siècle.

### 3.2.2 Observations

Après une première observation et les premières recherches dans la page de consultation des données, Guardenti ouvre le répertoire d'Emanuele De Luca, dont il trouve que l'accès aux informations est très facile, malgré le manque d'interactivité. Cela lui permet de compléter les métadonnées de l'interface avec la note rédigée par De Luca, afin de pouvoir suivre « au fur et à mesure la fortune scénique d'une pièce ». Il considère par ailleurs que les graphiques qui affichent la manière dont une pièce est représentée (si seule, avec une ou plusieurs pièces) ne sont pas sans intérêt, car ils permettent de saisir la force ou la faiblesse d'une pièce. Si au début de l'exploration il avait exprimé le besoin de connaître les sources (autre raison pour laquelle la confrontation avec le répertoire de De Luca est nécessaire), il revient sur cette critique après l'exploration des graphiques dans la partie de visualisation et après avoir explicité que le fonctionnement de la recherche se fonde uniquement sur le traitement de métadonnées.

Néanmoins, plusieurs informations manquent du point de vue de l'historien du théâtre, outre la note de De Luca – qui a vocation à apparaître à l'avenir, dans une forme appropriée au médium base de données. Premièrement, la possibilité de voir si les pièces ont été représentées seules ou dans des soirées composites devrait être possible sur une période donnée, et pas uniquement sur une pièce et sur l'ensemble des premières. Cela permettrait d'observer la force ou la faiblesse du Répertoire entier au cours du siècle. Deuxièmement, les informations et les possibilités de filtrage par éléments musicaux ne sont pas présentes, alors qu'elles seraient très pertinentes pour ce Répertoire. Il n'est par exemple pas possible de voir si une pièce présente des vaudevilles, à moins de faire appel aux notes d'Emanuele De Luca avec le fichier base de données brute. Troisièmement, il n'est pas automatique d'accéder à la somme d'un sous-ensemble de pièces, ce qui constitue une information fondamentale pour l'étude quantitative. En effet, dans aucun graphique il n'est possible de sélectionner un ensemble de colonnes pour connaître le nombre d'occurrence total ; pour cela, il faut compter à la main la somme des valeurs de chaque colonne.

Cette dernière critique est partagée par Agnelli, qui a longuement cherché, en vain, une fonctionnalité pour connaître la somme des valeurs de plusieurs colonnes, ou pour agréger sur une même colonne les informations d'une année (autrement dit une fonction de *stack*). Cette dernière possibilité était en effet initialement prévue, mais n'étant pas gérable directement avec le framework utilisé par l'ingénieure, sa réalisation a été reportée à des développements ultérieurs du projet, car elle aurait été chronophage. Il est tout de même possible, en regardant les histogrammes complets, d'avoir une idée de la quantité relative de représentations pour chaque année. Afin de chercher l'année où le plus de pièces ont été ajoutées au Répertoire, Agnelli a donc d'abord regardé le Répertoire entier, et a ensuite zoomé sur la période centrale, présentant le plus grand nombre de colonnes hautes (en proportion à l'ensemble). De cette manière il est possible de choisir un petit ensemble d'années avec une grande productivité. Le chercheur a ensuite regardé le détail de tous les graphiques sur ces années-là. Cela l'a amené à se poser de nombreuses questions sur le fonctionnement du théâtre : les raisons de la disparition des auteurs.ices du Répertoire, pourquoi à certains moments il y a une augmentation de la production de nouvelles pièces, ou comment fonctionnait la production de tous ces Arlequins, par exemple. Avec Guardenti, il partage aussi un manque en ce qui concerne les sources utilisées pour la création des données, manque qui va être en partie comblé par les informations introductives dans la page d'accueil.

Les deux chercheurs ont un avis similaire aussi en ce qui concerne certains aspects qu'ils trouvent les plus enthousiasmants de l'interface. Malgré le besoin de compléter les informations avec d'autres sources, ressources et textes qui se trouvent en dehors de l'interface (sur internet ou non), notre base de données contient toutes les informations essentielles pour mener une étude sur le Répertoire ; non seulement pour compter les données donc, mais aussi pour les lier entre elles. Les données présentes permettent d'étudier des tendances sur plusieurs aspects : le poids des auteurs, la permanence d'une pièce dans le Répertoire, la gestion économique du théâtre, pour citer des phénomènes mentionnés par Guardenti et Agnelli.

En particulier, Guardenti considère que les données sur les dépenses, recettes et spectateurs sont très sensibles, et qu'un accès via une base de données représente un véritable avantage, par rapport à chercher cette information dans les registres ou les répertoires, et surtout lorsque ces résultats sont mis en relation avec d'autres sources et d'autres types d'informations sur les pièces. Par rapport au répertoire de De Luca, cela représente une ouverture à l'étude de la dimension productive de la compagnie, et il s'agit pour Guardenti d'une différence fondamentale par rapport au répertoire de 2011. Cette amélioration s'ajoute aux deux plus évidentes : la possibilité de voir toutes les dates d'une pièce (non seulement la première), ainsi que l'extension de nos données au siècle entier.

En conclusion, si Agnelli considère, globalement, que l'outil est intuitif et opérationnel, pour Guardenti les données et les visualisations, non seulement constituent une contribution potentielle à ses recherches en cours, mais lui ont même suggéré de nouvelles pistes de recherche ou de nouvelles manières pour répondre aux anciennes questions. Malgré le manque d'explications et de contextualisation, l'outil semble donc avoir un potentiel expressif et interactif qui peut être utile, voire productif, dans des contextes de recherche différentes. Cela ne pourra néanmoins pas être évalué exhaustivement avant la mise en ligne et la diffusion de la plateforme.

### **3.3 Utilisation des données brutes**

La section précédente montre comment il est possible d'utiliser la base de données depuis les fonctionnalités offertes dans l'interface de consultation. Mais d'autres types d'analyses sont envisageables, en manipulant directement les données brutes, en format CSV et SQLite. Cela demande évidemment l'intermédiaire d'autres logiciels, que ce soit Excel ou

un langage de programmation comme Python, adapté à la lecture et à la manipulation du format de fichier choisi.

Les deux cas d'études suivants sont basés sur ce genre d'utilisation de la base de données. Le premier est fondé sur une activité de recherche qui s'est déroulée au sein du stage effectué par Mahsa Saberi Moayed et Lucas Gaiotti à l'Université de Nice : cela montre, par ailleurs, un aspect positif de la saisie manuelle, qui est le rapprochement au corpus et aux données, particulièrement bénéfique pour les membres de l'équipe qui ne sont pas spécialistes du sujet. Pour le deuxième cas d'étude, nous reviendront à la publication *L'apothéose d'Arlequin* (FABIANO et DE LUCA, 2023), dont sera approfondie la contribution de Giovanna Sparacello sur les canevas de magie de Veronese et de Goldoni (SPARACELLO, 2023).

### 3.3.1 Lieux mentionnés dans les titres

Ce premier exemple a été réalisé avec Lucas Gaiotti et Mahsa Saberi Moayed, dans le cadre de leur stage à l'Université Côte d'Azur, à partir d'une proposition de Mahsa<sup>65</sup> d'approfondir certaines pièces à partir des lieux qu'elles mentionnent. Finalement, nous avons décidé de commencer par visualiser les lieux qui sont mentionnés dans tous les titres des pièces. Pour cela, nous avons procédé ainsi : par l'intermédiaire d'un script, j'ai créé un fichier contenant tous les titres principaux des pièces (et uniquement ceux-ci) ; à partir de cette liste, Lucas et moi avons rempli un fichier CSV avec des informations sur les lieux trouvés : les coordonnées GPS, une catégorie et les informations sur la pièce en question (titre et identifiant du titre dans la base de données). Grâce à ce fichier il est possible d'afficher les lieux sur une carte et les distinguer suivant les catégories. Nous avons prévu initialement d'insérer aussi la première et la dernière date de représentation, mais nous n'avons pas pu compléter cette partie, pour des raisons organisationnelles<sup>66</sup>.

Le choix des catégories a constitué une première activité de recherche : Lucas a lu attentivement la liste des titres, afin de faire une proposition de classification. Il m'a envoyé un premier fichier avec des exemples associés à chaque catégorie, et je lui ai fait une proposition de modification à partir de celui-ci. Finalement, la classification est la suivante :

– lieux imaginaires :

---

<sup>65</sup> Malheureusement, elle n'a pas pu participer à la réalisation de la recherche pour des raisons personnelles.

<sup>66</sup> Cela peut constituer l'objet d'un stage ultérieur.

- divins (pour les lieux associés à des dieux ou des divinités, même mythologiques) ;
- inatteignables (pour des endroits réels mais qui ne peuvent pas être atteints par les êtres humains) ;
- inventés (pour les lieux imaginaires de la littérature) ;
- lieux réels :
  - quartier (pour les unités qui composent les villes et les villages) ;
  - ville (pour les villes et les villages) ;
  - région (pour des unités qui composent des pays) ;
  - pays (pour les territoires avec un statut politique) ;
  - continent.

	A	B	C	D	E	F	G	H
1	id_titre	titre_piece	lieu	latitude	longitude	1ere_rep	derniere_rep	catégorie
2	750	Arlequin dans l'île de Ceylan	Sri Lanka	7.5554942	80.7137847			région
3	12	Pantalon banqueroutier vénitien	Venise	45.4371908	12.3345898			ville
4	16	Bélphégor ou la Descente d'Arlequin aux Enfers	Enfers					divin
5	24	Arlequin corsaire africain	Afrique	11.5024338	17.7578122			continent
6	30	Agnès de Chaillot	Chaillot	48.8689495	2.2920996			quartier
7	414	Italien marié à Paris	Paris	48.8588897	2.320041			ville
8	585	Arcadie enchantée	Arcadie	37.4290886000	22.4235420000			région
9	99	Renaud de Montauban	Montauban	44.0175835000	1.3549991			ville
10	140	Naufrage au Port-à-l'Anglais	Port-à-l'Anglais	48.8000886	2.4090452			quartier
11	148	Jugement de Paris	Paris	48.8588897	2.320041			ville
12	153	Don Gaston de Moncade	Moncade	43.4919068	-0.7709902000			ville
13	157	Fragments italiens	Italie	42.6384261	12.674297			pays
14	730	Arlequin Empereur dans la Lune	Lune					inatteignable
15	174	Retour de la foire de Bezons	Bezons	48.9250016	2.2105491			ville
16	178	Amours de Vincennes	Vincennes	48.8474508	2.4396714			ville
17	179	Momies d'Egypte	Egypte	26.2540493	29.2675469			pays
18	182	Aventures de la rue Quincampoix	rue Quincampoix	48.8608983	2.3504541			quartier
19	198	Foire Saint-Germain	Saint-Germain	48.8547407	2.3242155			quartier
20	212	Terres australes	Terres australes					inventé

Figure 44 : extrait du fichier CSV avec les titres mentionnés dans les pièces.

Une fois ce fichier CSV créé (un extrait à la Figure 44), les données peuvent être visualisées sur une carte géographique, distinguées par catégorie. La répartition générale des lieux par catégories est représentée dans la Figure 45. Les lieux réels sont majoritaires dans notre échantillon. Les villes représentent presque la moitié de nos données ; l'ensemble des catégories *ville*, *région*, *pays* et *continent* couvrent la quasi-totalité de nos données, en excluant uniquement douze lieux sur les 97 répertoriés pour cette recherche.

Répartition des lieux mentionnés dans les titres  
par catégorie

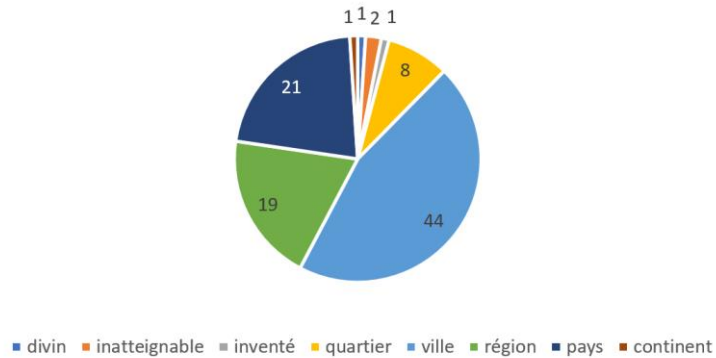


Figure 45 : répartition des lieux mentionnés dans les titres des pièces du Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle, par catégorie de lieu. (Graphique créé sur Excel)

La version interactive de la carte, créée à partir du fichier CSV, est disponible à ce lien : <http://u.osmfr.org/m/1294439/>. Cette carte ne peut pas afficher les lieux imaginaires. Dans la première visualisation des lieux réels que nous avons générée (Figure 46, Figure 47, Figure 48), l’affichage change selon la catégorie attribué (les carrés vert foncé sont les continents, les gouttes vert foncé sont les pays et celles vertes plus claires les régions, les villes sont les ronds bleus et les quartiers sont les épingles). Lorsque l’on clique sur un point quelconque, on affiche le titre de la pièce associée à ce lieu, et il est possible d’obtenir le détail de l’ensemble des données dans le menu à gauche.

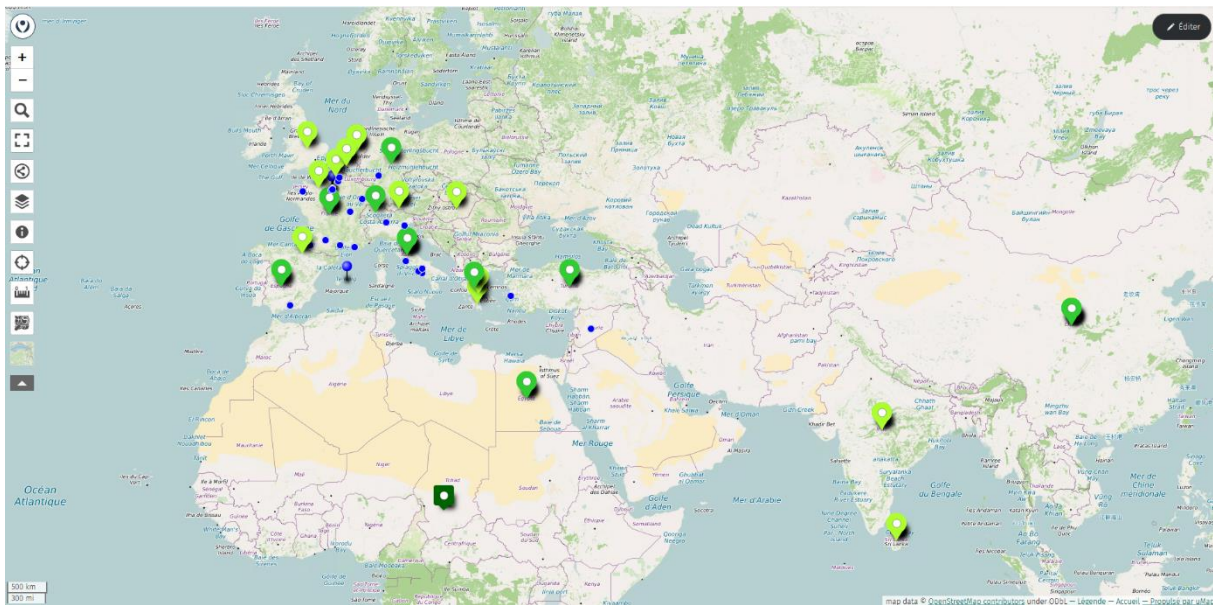


Figure 46 : carte avec les lieux mentionnés dans les titres du Répertoire.

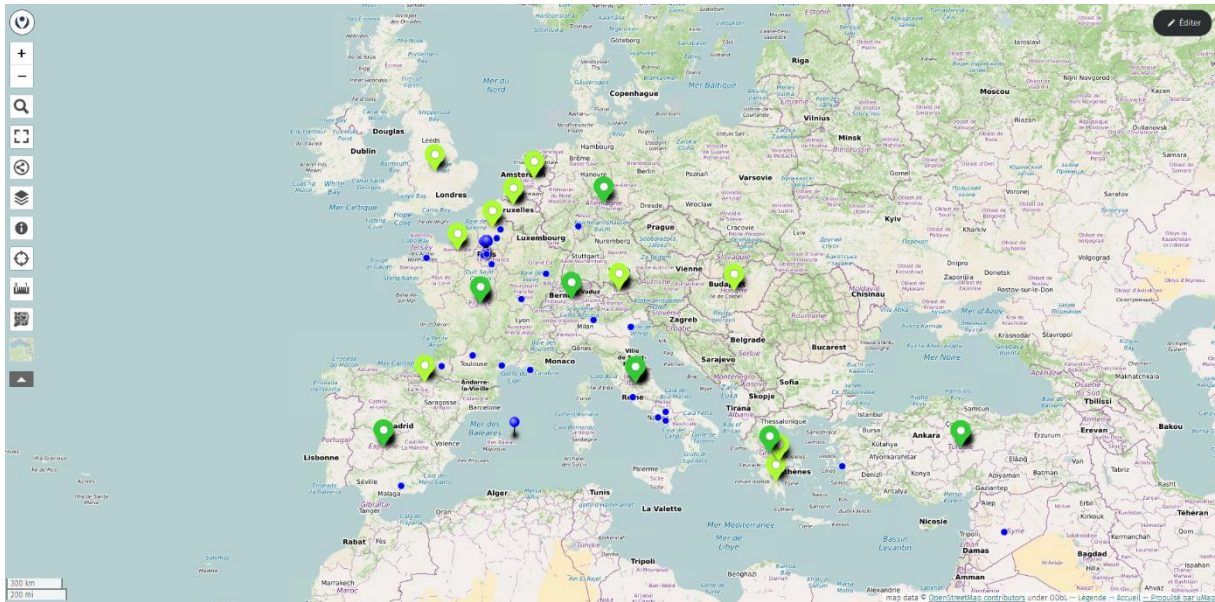


Figure 47 : carte avec les lieux mentionnés dans les titres du Répertoire (zoom centré sur l'Europe).

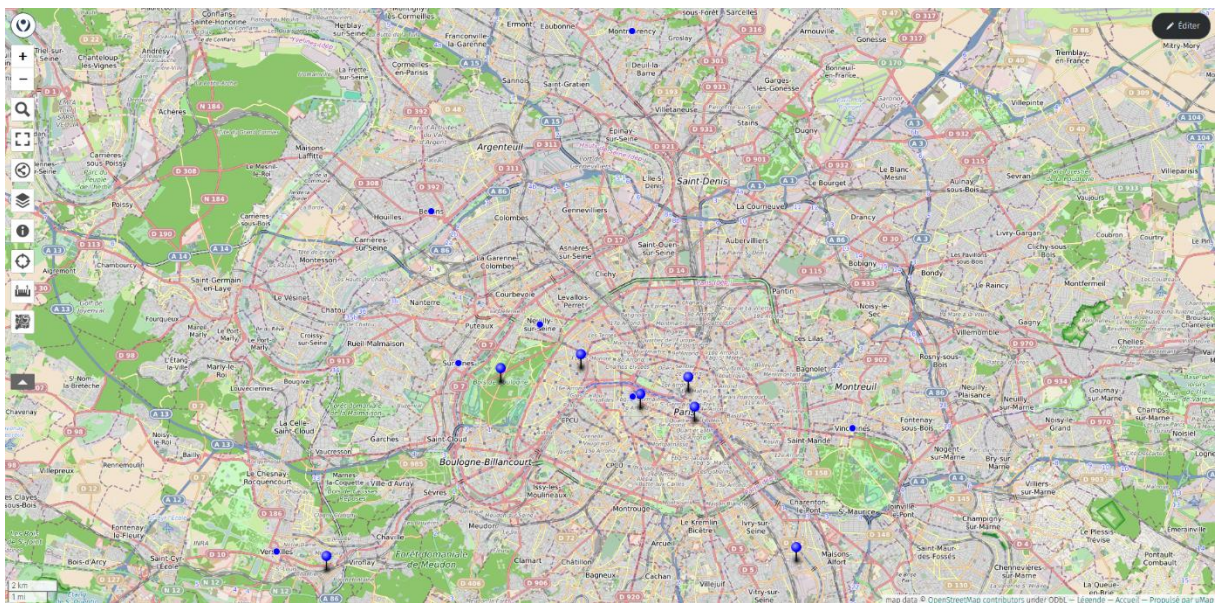


Figure 48 : carte avec les lieux mentionnés dans les titres du Répertoire (zoom centré sur Paris).

Cette première visualisation a une limite importante pour notre contexte : si un lieu est mentionné plusieurs fois, les points se superposent et il est possible de lire uniquement un titre où ce lieu est cité (comme c'est le cas pour le lieu Paris en tant que ville, Figure 48). Afin d'indiquer qu'un lieu apparaît plusieurs fois, nous avons expérimenté deux types de

visualisation des données : par zone de chaleur ou heatmap (Figure 49), et par cluster proportionnel à la distance de la carte (Figure 50, Figure 51, Figure 52).

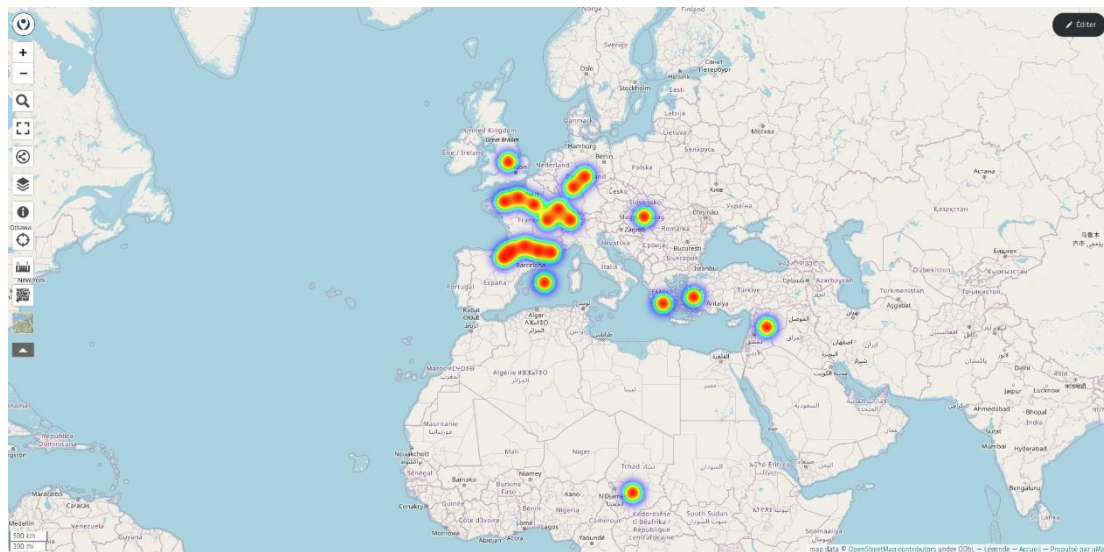


Figure 49 : affichage des lieux mentionnés dans les titres du Répertoire par zones de chaleur (heatmap).

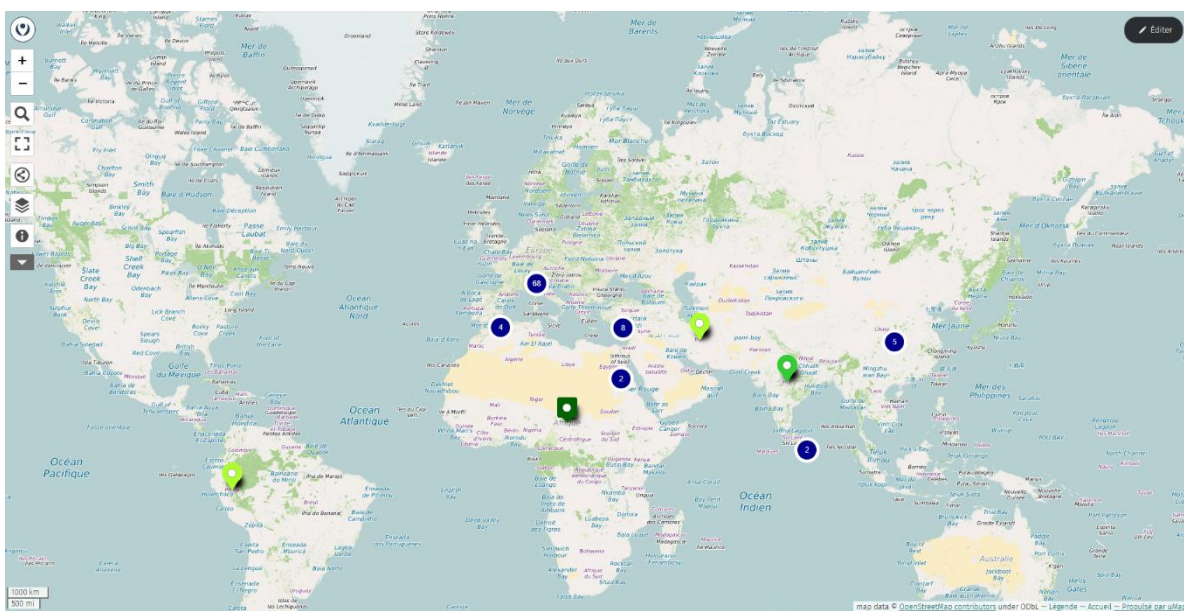


Figure 50 : affichage des lieux mentionnés dans les titres du Répertoire par clusters.

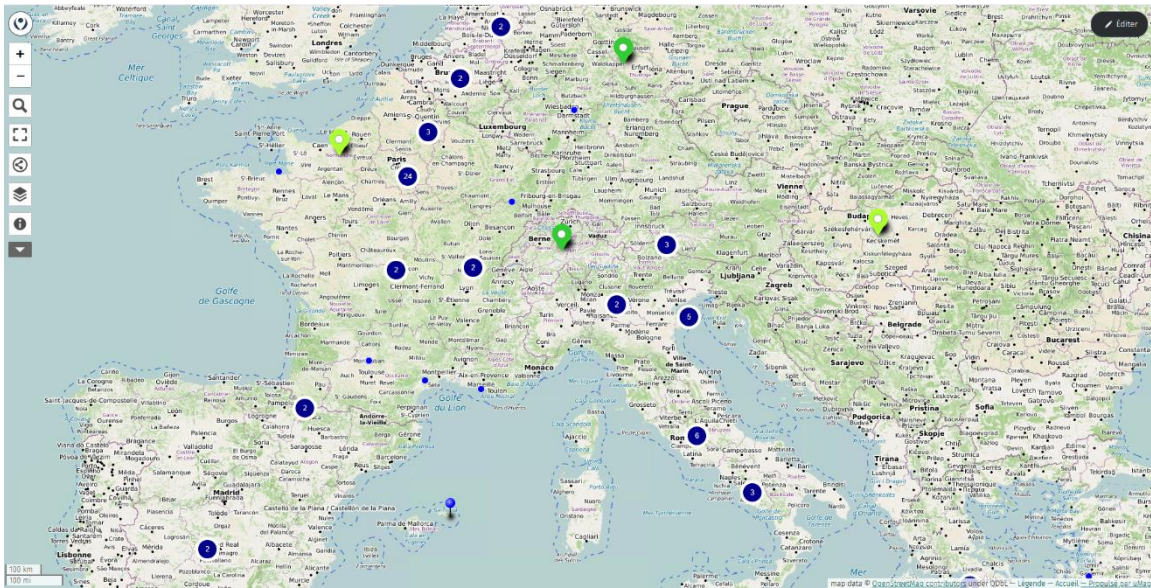


Figure 51 : affichage des lieux mentionnés dans les titres du Répertoire par clusters, centré sur l'Europe.

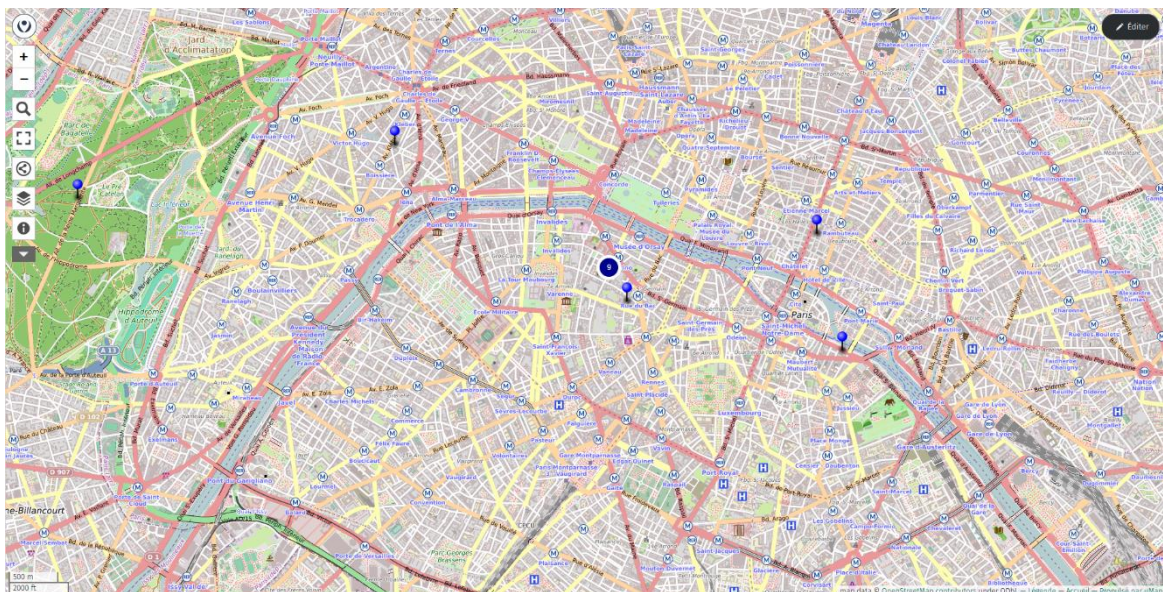


Figure 52 : affichage des lieux mentionnés dans les titres du Répertoire par clusters, zoom sur Paris.

La visualisation par zones de chaleur a l'avantage de montrer à coup d'œil les zones qui sont citées et leur importance dans l'ensemble de données, grâce aux couleurs et à la dimension des spectres affichés. En revanche, il n'est possible d'obtenir aucun détail sur les données, comme le titre ou l'emplacement exacte du point qui est cartographié. Le type de lieu dont il



qu’ensembles de données, avant de s’intéresser à leur détail. Une telle analyse est pertinente et peut être intéressante dans notre contexte de recherche, compte tenu de la présence fréquente d’éléments exotiques dans les pièces de notre Répertoire, mais doit être complété par des études sur le contenu des pièces.

La carte telle que nous l’avons créée, montre premièrement une plus forte concentration de lieux mentionnés en Europe et autour de la Méditerranée, par rapport au reste du monde, ce qui n’est pas étonnant vu l’emplacement géographique et la tradition culturelle de la compagnie. En faisant un zoom centré sur la Méditerranée (Figure 54), la carte présente une plus grande concentration de points en France et en Italie. Si l’on regarde le détail des clusters dans ces deux régions par rapport au reste de la zone dans l’image, ces deux pays présentent un plus grand nombre de points faisant référence à des endroits précis qu’à des zones étendues. Autrement dit, il y a un plus grand nombre de lieux catégorisés comme des villes et des quartiers, que comme des régions, des pays ou des continents. Cette considération est vraie aussi si l’on regarde la zone européenne par rapport au reste du monde, où plus aucune ville ni aucun quartier n’apparaît.

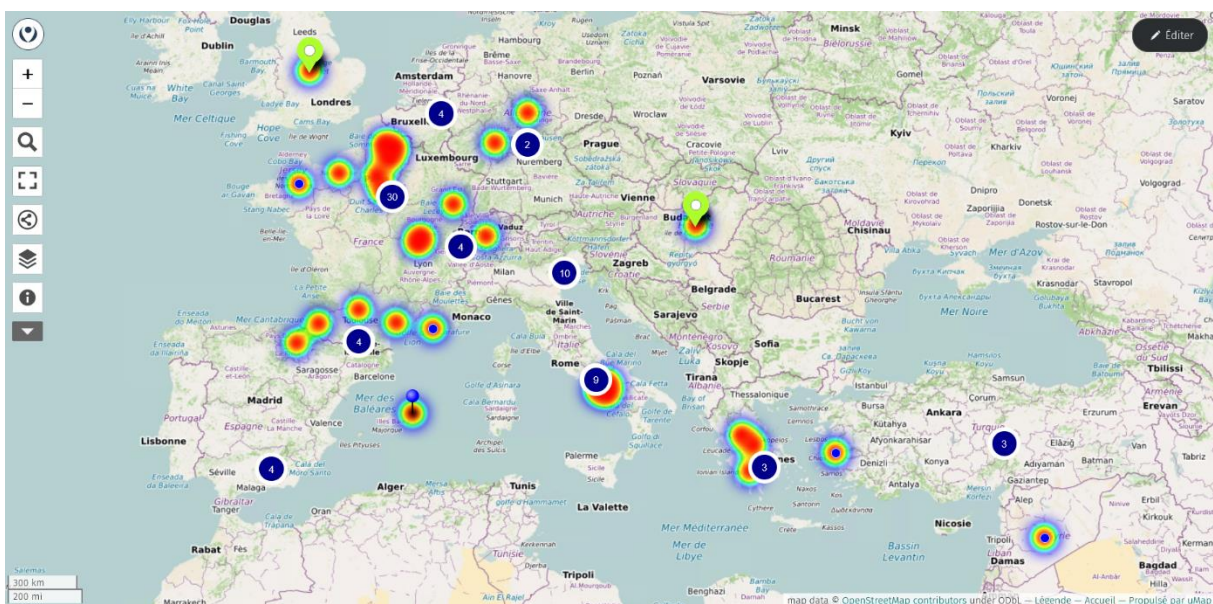


Figure 54 : lieux mentionnés dans les titres, zoom centré sur la Méditerranée.

Ainsi, plus l’on s’approche du lieu d’activité de notre compagnie, plus les lieux mentionnés dans les titres vont être circonscrits et précis : seulement Paris et ses alentours représentent vingt-quatre points sur la carte, alors que l’Afrique est génériquement mentionnée une seule fois (dans *Arlequin corsaire africain*), la Chine cinq fois, le Sri-Lanka et l’Egypte

deux fois. Il existe aussi une mention à l'ancien empire Moghol (*Arlequin grand Mogol*), se situant dans l'actuelle région indienne. Les deux derniers points qui restent en dehors de l'espace européen se situent dans l'actuel Iran et au Pérou, et se réfèrent à la même pièce, les *Indes dansantes*. Celui au Pérou est par ailleurs le seul point présent dans toutes les Amériques.

Or l'affichage de ces deux derniers lieux n'est pas à priori évident, et la carte ne les présentait pas au début (voir par exemple la Figure 46). En effet, le mot *Indes* au XVIII<sup>e</sup> siècle ne décrit pas l'actuelle Inde, ni sa région, mais il s'agit au contraire d'un terme générique qui se réfère à des endroits lointains et exotiques, sans distinction entre Asie et Amériques (il s'agit d'un terme similaire à ce que signifie en Europe le terme *orient*, un espace indéfini situé au sud ou à l'est de l'espace européen, comprenant l'Afrique, le Levant et l'Asie entière, comme décrit notamment dans SAID, 2021). Dans ce sens, attribuer la catégorie *lieu imaginaire inventé* à ce titre aurait été tout-à-fait pertinent, si l'objectif de recherche concernait une critique de la manière dont les lieux sont décrits, par exemple. Mais le but est d'observer l'imaginaire géographique de notre Répertoire, et il est possible de connaître les lieux réels auxquels la pièce se réfère avec ce terme, à partir de la pièce *Indes galantes*, dont elle est la parodie. Ainsi, le titre *Indes dansantes* apparaît trois fois dans le fichier Excel, avec trois lieux différents (Turquie, Iran, Pérou), afin de pouvoir tenir compte de l'intégrité de l'imaginaire géographique auquel les titres des pièces du Répertoire renvoient.

Cette répartition des points sur la carte suggère des questions sur la distance culturelle et sur le potentiel suggestif qui était attribué aux lieux lointains : la carte présente neuf points en Asie (y compris l'Iran), trois points en Afrique et un seul point dans toutes les Amériques, ce dernier représentant seulement une partie d'une pièce. Pourquoi cette répartition ? Est-ce que, dans l'imaginaire d'un.e spectateur.ice de la Comédie-Italienne de Paris au XVIII<sup>e</sup> siècle, les Amériques sont *plus lointaines* que la Chine, en termes culturels ? Ou est-ce qu'au contraire les Amériques n'apparaissent presque pas car elles ne sont pas *assez* exotiques ? *Quid* de l'Afrique, avec ses deux points en Egypte et son point générique sur le continent ? Que représente l'Afrique pour les comédien.nes italien.nes ? Répondre à ces questions fournirait des informations non seulement sur le caractère hybride du Répertoire, mais plus globalement sur la représentation de l'altérité dans l'Ancien Régime.

Afin de donner des éléments de réponse convaincants, il serait tout d'abord nécessaire de se référer à la littérature critique sur ce sujet et à l'histoire culturelle, comme dans les cas d'étude précédents, afin d'enrichir l'analyse avec des éléments historiques et contextuels. De

plus, ces considérations peuvent encore être complétées par nos données : la base de données contenant les dates des représentations connues, il est possible de voir à quels moments du siècle apparaît chaque lieu, ainsi que de connaître le nombre de représentations et d'évaluer par conséquent le succès des pièces. Cela permettrait d'analyser plus en détail l'imaginaire géographique mobilisé au sein de la stratégie de programmation de la Comédie-Italienne de Paris. L'étude serait enfin complétée par une analyse du contenu des pièces, qui permettrait de tenir compte de tous les lieux mobilisés dans notre Répertoire, sans se limiter aux titres.

### 3.3.2 Magie et métamorphoses

Lors de la saisie des données ma curiosité a été attirée par la présence récurrente du champ lexical de la magie et du surnaturel dans les titres des pièces : connaissant de manière superficielle le siècle des Lumières, il s'agit d'un phénomène étonnant. Seulement en cherchant dans les titres des pièces la parole « fée », la base de données renvoie dix résultats<sup>67</sup>, presque tous joués plus que cinq fois (sauf *Arcadie enchantée*, connue aussi comme *Arlequin jouet des fées*), certaines représentées pendant plus que dix ans. Et les références à la magie et au surnaturel ne se limitent pas à la présence de fées. Dans le livre d'Andrea Fabiano sur *La Comédie-Italienne de Paris et Carlo Goldoni* (FABIANO, 2018) on retrouve par exemple des éléments surnaturels dans la pièce *Le bon et le mauvais Génie*, où les actions des personnages sont orchestrées par l'action magique des deux êtres surnaturels. Dans sa contribution à *l'Apothéose d'Arlequin* (FABIANO et DE LUCA, 2023), aussi, Giovanna Sparacello (SPARACELLO, 2023) focalise l'analyse d'éléments magiques sur la présence de transformations, ou métamorphoses, qu'elle distingue en extérieures (lorsque la transformation est visible sur un acteur ou sur la scène) et intérieures (lorsque la transformation concerne la psychologie ou le comportement d'un personnage).

Or cette présence d'éléments magiques peut s'expliquer en réalité par la spécificité du répertoire italien et par les attentes du public vis-à-vis de ses acteurs et actrices. En effet, dans la grande variété de pièces créées et jouées à la Comédie-Italienne de Paris, le fantastique et le merveilleux trouvent aussi leur place, sous différentes formes : bagues magiques, enchantements, transformations, fées, génies et autres créatures imaginaires. Ainsi, malgré le

---

<sup>67</sup> Cette recherche a été effectuée dans l'interface de saisie. Dans cette partie de la plateforme, lorsque l'on fait une recherche d'un mot dans les pièces, l'algorithme de recherche renvoie les pièces dont l'un des titres contient le mot recherché. Autrement dit, la recherche est faite uniquement sur les titres, et non sur l'auteur, le genre ou l'édition, par exemple.

« mépris presque unanime [...] des "philosophes" pour les Italiens et ce qu'ils jouent » (ROUGEMONT, 2001, p. 252), les actrices et acteurs italien.nes véhiculent, avec leur jeu, une approche à la vie plus sensible et plus corporelle :

« En s'appuyant sur la perception propre aux spectateurs et aux critiques français, qui décèlent à priori une constante baroque dans chaque réalisation artistique italienne, qu'elle soit théâtrale, littéraire ou figurative, les comédiens italiens à Paris structurent leur production autour de l'idée de la surprise, du mélange, de l'ambiguïté et de la fragmentation dans son opposition à la cohérence rationnelle assumée par la Comédie-Française et l'Académie Royale de musique » (FABIANO, 2018, p. 24).

Autrement dit, les féeries, les métamorphoses et autres phénomènes merveilleux font partie de l'essence-même du Répertoire italien. Après avoir analysé plusieurs pièces contenant des transformations, dans la deuxième moitié du siècle (1744-1770), Sparacello conclut que :

« L'usage que Veronese et Goldoni ont fait de la métamorphose permet d'indiquer les deux voies suivies par les canevas de magie à la Comédie-Italienne dans la période 1744-1770. La voie traditionnelle et spectaculaire, où la métamorphose est avant tout physique et où le rythme des transformations, que ce soit des personnages ou des lieux, est soutenu, et une voie différente, où les personnages, touchés par une métamorphose intérieure, font migrer la comédie vers les territoires du drame » (SPARACELLO, 2023).

Dans notre base de données, il est possible de chercher toutes les dates documentées de pièces représentant les deux paradigmes, afin d'observer leur évolution dans le temps. Quel paradigme rencontre le plus le succès du public ? Même si une conclusion ne peut pas être fournie uniquement au moyen des données, elles peuvent suggérer des éléments de réponse. Analogiquement aux cas d'étude de la première section (Utilisation de l'interface), les représentations des pièces mentionnées dans l'article de Sparacello seront comparées, afin d'observer comment évoluent les deux voies des canevas de magie, en élargissant les données au siècle entier (Sparacello aborde la période 1744-1770). Afin d'évaluer le degré de succès face au public, ces données seront à leur tour comparées avec les valeurs moyennes du Répertoire entier.

Sparacello analyse uniquement deux pièces avec transformations (ou métamorphoses) internes – *Les folies de Coraline* et *La bague magique* – alors qu'elle en mentionne une dizaine présentant des transformations (ou métamorphoses) externes : *Coraline esprit follet*, *Arcadie*

*enchantée, Coraline fée, Arlequin et Scapin morts vivants, Coraline magicienne, Les fées rivales, Arlequin génie, Le prince de Salerne, Les métamorphoses d'Arlequin, Les nouvelles métamorphoses d'Arlequin et Arlequin esprit follet.* Ci-dessous un tableau résumant le nombre de représentations présentes dans la base de données pour chacune de ces pièces.

<b>Titre</b>	<b>Nombre de représentations</b>
<i>Les folies de Coraline</i>	55
<i>La bague magique</i>	17
<i>Coraline esprit follet</i>	40
<i>Arcadie enchantée</i>	24
<i>Coraline fée</i>	18
<i>Arlequin et Scapin morts vivants</i>	7
<i>Coraline magicienne</i>	49
<i>Les fées rivales</i>	23
<i>Arlequin génie</i>	16
<i>Le prince de Salerne</i>	161
<i>Métamorphoses d'Arlequin</i>	Ambigu
<i>Nouvelles métamorphoses d'Arlequin</i>	Pièce non présente
<i>Arlequin esprit follet</i>	13

*Le prince de Salerne* est sans ambiguïté la pièce la plus représentée, avec 161 répliques présentes dans la base de données. Des autres pièces dont cette valeur est connue, à part *Arlequin et Scapin morts vivants*, on a répertorié plus que dix représentations chacune. Afin d'évaluer en quelle mesure ces valeurs peuvent représenter un succès pour la compagnie, il est nécessaire de les comparer avec des données générales sur le Répertoire entier : minimum, maximum, moyenne, médiane, mode et écart-type, à minima.

Ces valeurs ont été calculées au moyen de Python, en manipulant directement les données brutes avec un Jupyter Notebook (disponible dans le repository framagit du projet). Les résultats de ces calculs sont résumés dans le tableau ci-dessous :

<b>Donnée</b>	<b>Nombre de représentations</b>
Moyenne	26
Médiane	8

Mode	1
Ecart-type	47
Minimum	1
Maximum	347

Il est possible de visualiser la distribution des nombres de représentations pour toutes les pièces dans la Figure 55. Dans ce graphique les pièces sont individuées avec leur identifiant sur l'axe des abscisses, pas avec leur titre, pour des raisons de lisibilité, et puisque cette information est négligeable dans un graphique qui a comme but de saisir des tendances générales. A chaque identifiant sur les abscisses correspond un point bleu qui définit sur l'axe des ordonnées le nombre de représentations de la pièce en question. Le graphique montre aussi des lignes correspondant, du bas vers le haut, à la médiane (en rouge), à la moyenne (en vert) et à l'écart-type sur la moyenne (en vert avec trait fin).

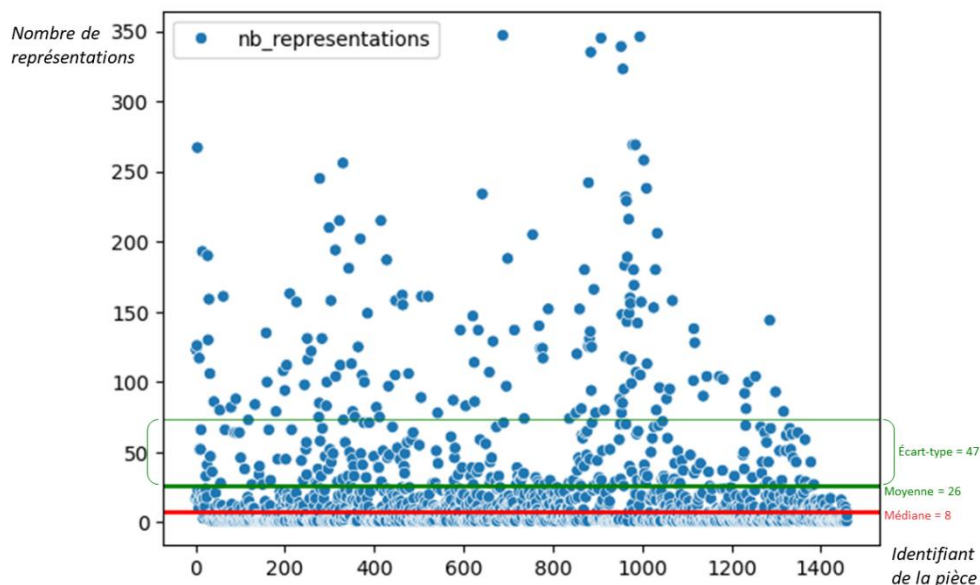


Figure 55 : Distribution du nombre de représentations de toutes les pièces de la base de données. (Scatter plot généré avec seaborn).

Comme l'indiquaient déjà la moyenne, la médiane et la mode, la plupart des pièces du Répertoire ont été représentées peu de fois : sur un total de 1460 pièces, la moitié a été représentée moins que huit fois, avec une moyenne générale de 26 représentations par pièce. La mode égale à un indique le nombre de représentations le plus fréquent ; autrement dit, en tirant aléatoirement une pièce du Répertoire, il y a plus de chances de trouver une seule représentation, que toute autre valeur plus grande que l'unité.

*Arlequin et Scapin morts vivants* est ainsi la seule pièce, parmi celles mentionnées par Sparacello, qui se trouve en-dessous de la médiane (chiffre marqué en rouge dans le tableau). Six pièces se trouvent entre la médiane et la moyenne (chiffres marqués en orange dans le tableau) et trois se situent dans l'écart type de la moyenne (chiffres marqués en vert dans le tableau). Suivant nos calculs, *Le prince de Salerne* est en revanche bien au-dessus de l'écart type de la moyenne, avec ses 161 représentations répertoriées : par rapport à cette donnée, il est possible d'affirmer que *Le prince de Salerne* a représenté un véritable succès à la Comédie-Italienne, et il s'agit d'une pièce présentant des métamorphoses externes. Au total, en somme, des onze pièces dont le nombre de représentations est connu, dix sont au-dessus de la médiane, dont quatre sont aussi au-dessus de la moyenne. Ces données suggèrent qu'il s'agit d'un échantillon qui a connu le succès du public, dans la mesure où seulement une pièce se trouve en-dessous de la médiane, et qu'une autre se situe au-dessus de l'écart-type de la moyenne. Pour démontrer sans ambiguïté cette affirmation, il faudrait calculer la probabilité d'obtenir cet ensemble de nombres de représentations, lorsque l'on tire aléatoirement onze pièces de la base de données (c'est-à-dire qu'il faudrait construire un modèle statistique *neutre*).

En revanche, la comparaison entre le succès des deux paradigmes, interne et externe, est plus délicate, car les deux échantillons ne sont pas comparables (deux pièces d'un côté et onze de l'autre, neuf si l'on ne considère que les pièces dont on connaît les dates sans ambiguïté). En observant d'abord les représentations des pièces avec transformations externes (Figure 56), jusqu'à 1760 des répliques sont présentes pour la majorité de ces pièces. Ensuite, seule *Le prince de Salerne* reste, et *Arlequin esprit follet* apparaît à la fin des années 1770 avec ses treize représentations répertoriées.

## Dates de représentation:

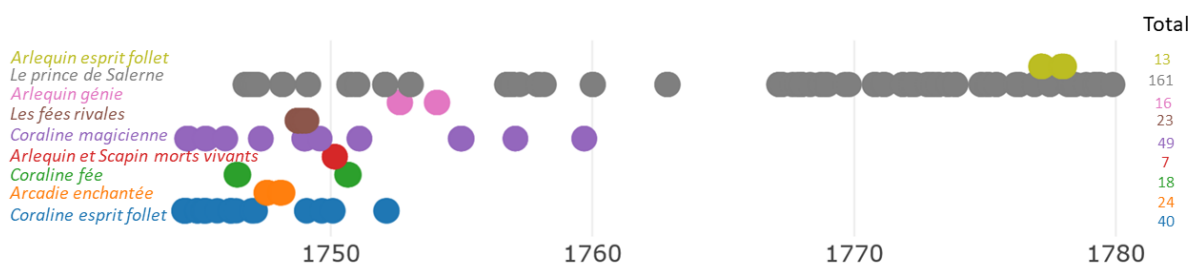


Figure 56 : Représentations des pièces avec métamorphoses externes mentionnées dans l'article de Sparacello (SPARACELLO, 2023).

Puisque *Le prince de Salerne* a eu un succès qui va bien au-delà de la moyenne du Répertoire, ce graphique (Figure 56) suggère deux hypothèses :

1. le paradigme des transformations extérieures n'a plus assez de succès à partir de la moitié du siècle, mais *Le prince de Salerne* continue d'être représentée car elle suscite l'enthousiasme du public, pour des raisons liées à sa forme spectaculaire ou dramaturgique spécifique ;
2. le paradigme des transformations extérieures est expérimenté et s'affine entre les années 1740 et 1760, pour aboutir à l'affirmation de *Le prince de Salerne* comme meilleur représentant de ce choix spectaculaire, permettant de réduire l'effort créatif et d'assurer l'affluence du public.

Si l'on élargit l'observation des représentations aux pièces avec transformations intérieures, on remarque que sur aucune période il n'y a de représentations des deux pièces ensemble. *Les folies de Coraline*, dont le succès dépasse la moyenne, a été jouée à la même période où apparaissent la majorité des pièces à transformations externes (entre les années 1740 et 1760). *La bague magique*, en revanche, a surtout été jouée après 1770, parallèlement seulement à *Le prince de Salerne* et à *Arlequin esprit follet*, et semble avoir connu moins de succès.

## Dates de représentation:

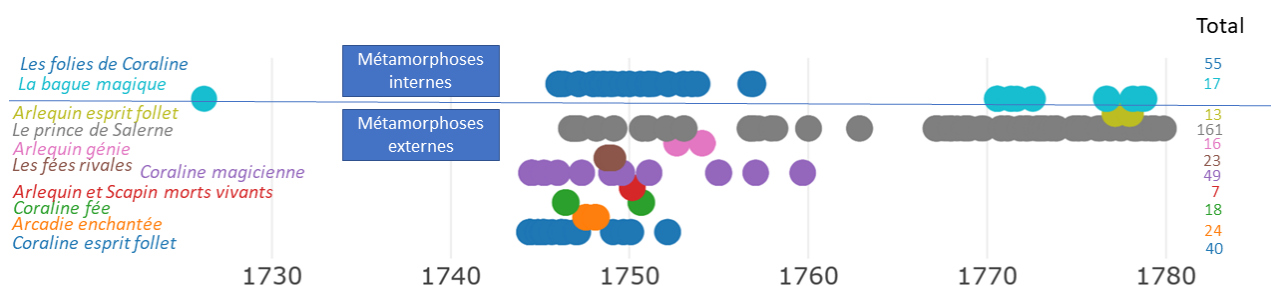


Figure 57 : Représentations de toutes les pièces avec métamorphoses mentionnées dans l'article de Sparacello (SPARACELLO, 2023)

Par rapport aux hypothèses émises après l'analyse des pièces avec transformations externes, l'observation de ces deux pièces ferait pencher vers la première possibilité : *Les folies de Coraline* ferait partie de la phase d'expérimentation et d'affirmation du paradigme extérieur, et *La bague magique* serait une tentative de revitalisation du paradigme intérieur, qui rencontre un succès moyen. Or, le nombre de représentations de *Les folies de Coraline* se situe dans la même plage que *Coraline esprit follet* et *Coraline magicienne* – entre la moyenne et l'écart-type de la moyenne : si l'on exclut le succès de *Le prince de Salerne*, les trois pièces les plus jouées de notre échantillon mentionnent l'actrice Maria Anna Veronese (Coraline) dans le titre. Cela suggère une hypothèse ultérieure : le succès d'un certain paradigme dépend plus de la virtuosité des acteurs et actrices, que de la forme des choix spectaculaires en tant que tels. Par ailleurs, comme suggéré par Sparacello lors d'échanges personnelles, la représentation de la folie permet à l'actrice de montrer ses capacités, en performant un jeu instable et variable, et représente un motif théâtral récurrent. Un autre élément contextuel que la chercheuse met en lumière est le projet dramaturgique de Goldoni dans la direction d'une comédie de caractère, qui n'aura pas le succès attendu devant le public parisien (FABIANO, 2018).

En conclusion, ces analyses suggèrent que le paradigme des transformations extérieures, avec ses éléments fortement spectaculaires, s'affirme au cours du siècle et rencontre le succès du public, avec la persistance de *Le prince de Salerne* dans le Répertoire. Néanmoins, cette conclusion peut être biaisée par la forme de l'échantillon, car il ne représente pas les deux paradigmes de manière comparable et exhaustive, mais aussi parce qu'il contient trois pièces dont le succès dépend de la faveur du public envers le jeu de l'actrice Maria Anna Veronese.

Mais l'affirmation ou l'infirmité de toutes ces hypothèses, ne peut être faite qu'en élargissant l'étude à d'autres pièces, afin d'avoir deux échantillons comparables pour les deux paradigmes, ainsi qu'en enrichissant l'analyse des incontournables éléments historiques et littéraires.

Ces pages montrent donc en premier lieu les potentialités de l'analyse quantitative, mais montrent surtout la nécessité de mobiliser au même temps l'étude qualitative afin de conclure sur des hypothèses. En effet, l'analyse des données a permis de mieux contextualiser l'évaluation du succès des pièces, mais elle a aussi suggéré des pistes de recherche que la numérisation du Répertoire permet de suivre. Néanmoins, un manque d'approfondissement du contexte historique limite ces cas d'étude à des considérations *à distance* sur l'objet étudié, et empêche de conclure de manière exhaustive.

## 4 Partie 3 : discussion

Cette troisième partie est consacrée à la discussion du travail présenté dans les deux premières. Après avoir décrit le travail effectué et montré des cas d'étude, il faut à présent évaluer les résultats obtenus : qu'est-ce que notre base de données apporte à l'étude du Répertoire de la Comédie-Italienne de Paris ? Afin de donner une réponse à cette question, le chapitre ne se limitera pas à considérer l'objet d'étude Comédie-Italienne, mais élargira les considérations à l'activité de production des données en elle-même.

Ayant une formation plus informatique que littéraire, mais mêlant les deux, cette recherche se situe d'un point de vue informaticien, observant ce que les humanités numériques *font* aux données. Il s'agit donc plus d'une recherche en *informatica umanistica* qu'en humanités numériques, si l'on accepte qu'une distinction entre ces deux termes peut être donnée par différents points de vue : « si dans "humanités numériques" le sujet sont les "humanités", dans la formulation italienne [*informatica umanistica*] le sujet est l'informatique »<sup>68</sup> (CIOTTI, 2023, p. 20) Ciotti exhorte aussi à prendre

« au sérieux la signification du terme "informatique" qui dérive de son étymologie attestée : la contraction entre "information" et "automatique", c'est-à-dire l'étude de processus d'élaboration automatique de l'information, ce qui rend central les concepts d'information et de son élaboration, et périphérique celui des machines physiques (et de la façon dont elles sont construites) avec lesquelles on peut faire cela »<sup>69</sup> (ibid.)

Parmi les différentes approches des humanités numériques dans les études littéraires, cette recherche prend donc un point de vue informatique observant les lettres, et s'insère dans une longue recherche sur la Comédie-Italienne de Paris dans l'Ancien Régime qui a porté sur la manière dont les traditions théâtrales française et italienne s'y sont rencontrées et ont évolué au sein de ce théâtre (FABIANO et DE LUCA, 2023). La création de la base de données pour le Répertoire représente une contribution à ce projet plus large, où ma position a été celle de

---

<sup>68</sup> «Laddove in "Digital Humanities" il soggetto sono le "Humanities", nella formulazione italiana [*informatica umanistica*] lo è l'*informatica*»

<sup>69</sup> «Sul serio l'accezione del termine "informatica" che deriva dalla sua etimologia attestata: la contrazione tra "informazione" e "automatica", ovvero studio dei processi di elaborazione automatica dell'informazione, che rende centrale il concetto di informazione ed elaborazione della stessa, e periferico quello delle macchine fisiche (e del modo in cui sono costruite) con cui ciò si può fare.»

*l'informaticienne travaillant dans les humanités (informatica-umanista on dirait en italien) : une intermédiaire, ou une sorte d'interprète, entre le monde littéraire et le monde informatique.*

L'incompréhension entre personnes travaillant dans les deux domaines, informatique et littéraire, est bien sûr loin d'être absolue, ni indépassable. Au contraire, la communication interdisciplinaire est souvent mentionnée comme étant source de nouvelles questions, de nouveaux points de vue et de nouvelles recherches (voir par exemple RUBELLIN, RASCHIA, 2020). Travailler dans une équipe interdisciplinaire peut nourrir les recherches de tous les domaines concernés, des lettres à l'informatique. Ainsi, un rôle que peut avoir une informaticienne-humaniste est celui de regarder ce qui se passe et ce qui se joue dans ces échanges, observer et analyser comment les disciplines se changent mutuellement.

Dans la première section de cette partie, seront discutés les résultats obtenus en relation aux recherches précédentes sur la Comédie-Italienne de Paris, avant l'élargir la discussion au contexte de travail transdisciplinaire, ayant permis la réalisation du projet. Les sections suivantes aborderont des questions méthodologiques et épistémologiques, autour de l'automation et de l'IA, de l'épistémologie des humanités numériques et finalement de l'écologie.

## **4.1 Discussion des résultats**

### **4.1.1 Théâtre et Comédie-Italienne**

Les deux chapitres précédents ont permis de voir, en pratique, comment la base de données, et en particulier l'interface de consultation, peuvent être utilisées pour approfondir des recherches sur la Comédie-Italienne de Paris. Tous les cas d'étude se sont arrêtés au constat de la nécessité d'intégrer des connaissances historiques ou littéraires, afin de conclure sur les hypothèses. Franco Moretti constatait déjà cette dynamique en 2008, dans son célèbre livre *Graphes, cartes et arbres* (MORETTI et al., 2008) : les recherches présentent des « aller-retour » entre lecture à distance et lecture rapprochée des phénomènes qu'il étudie.

Travailler sur de centaines, voire de milliers, de métadonnées, même lorsqu'elles ne sont pas tout-à-fait complètes, permet néanmoins d'observer les phénomènes sur une plus grande échelle. Rubellin soutient plus fortement que « les requêtes et les statistiques permettent d'avoir des données fiables, et non plus des impressions subjectives dont regorgent certains

livres qui, à partir de trois ou quatre textes et non deux cents, formulent des généralités non vérifiées » (RUBELLIN, 2020, p. 235). Or, en réalité, cela est vrai lorsque « les requêtes et les statistiques » s'insèrent dans une recherche plus large et sont cohérentes avec la méthode de travail : comme pour le cas d'étude concernant les parodies, les données de Pauline Beaucé étant profondément différentes des nôtres, il a été pertinent de décrire leur relation, avant de pouvoir comparer ses résultats avec notre base de données. De plus, il n'a été que partiellement possible de conclure, pas uniquement par manque de données, mais aussi par manque de méthode – en particulier par besoin de nouveaux corpus, et la constitution d'un corpus est un travail tout-à-fait interprétatif.

Dans ce sens, le seul fait de travailler sur de grandes quantités de données, si cela permet toujours d'élargir l'échelle des observations, ne garantit pas en soi d'éliminer toutes les « impressions subjectives » :

« les méthodes permettant de faire de l'œuvre littéraire un ensemble de faits textuels ordinaires, examinables empiriquement dans leur connexion à d'autres faits textuels ou références, et susceptibles de comparaisons, supposent [...] le recours à un savoir philologique et historique ancien avant toute expérimentation, et exposent [...] tout savoir empirique aux aléas des conditions de sa constitution » (GEFEN, 2015, p. 5).

Ces « aléas » sont d'autant plus importants si l'on se penche sur les « boîtes noires » utilisées pour automatiser le travail : une « boîte noire » est un outil dont on connaît l'input et l'output, sans savoir comment passer du premier au second (MASURE, 2017). Tout logiciel propriétaire a ce fonctionnement, pour garantir le secret industriel. Cette question sera ultérieurement abordée dans la section sur l'automatisation et l'IA.

Concernant les données traitées, la base de données construite lors du projet de doctorat, ne contient pas de texte ni d'image, mais elle permet d'agréger des informations différentes, provenant de supports et d'époques différentes, qui peuvent participer à la reconstitution de la dimension spectaculaire de notre Répertoire profondément lacunaire. Spanu Fremder propose et suit dans sa thèse une méthode pour reconstituer la dimension spectaculaire des pièces du Répertoire de la Comédie-Italienne :

« La représentation fictive de ces spectacles « invisibles » se fait en trois temps, qui vont de la création jusqu'à la représentation et comprennent les phases de « composition », de « performance » et enfin de « synthèse ». À l'intérieur de cette structure interprétative, il est

possible de lire la genèse et l'évolution des spectacles, à travers les multiples typologies de sources dont nous disposons, qui, en raison de leur hétérogénéité, ne peuvent en aucun cas être pesées au moyen des mêmes paramètres. Les registres du théâtre, et en particulier celui des inventaires des accessoires, décors et personnages, s'inscrivent dans la phase de « composition », tandis que les récits des performances d'acteurs et enfin les résumés des spectacles, appartiennent respectivement aux phases de « performance » et « synthèse ». Seuls l'assemblage et l'utilisation combinée de ces matériaux documentaires conduit à la « recomposition » du répertoire italien de la Comédie-Italienne » (SPANU FREMDER, 2010, p. 15).

Les deux répertoires qui ont été numérisés afin de créer la base de données (BRENNER, 1961 et DE LUCA, 2011) couvrent les trois phases de la recomposition d'un spectacle : celui de Brenner concerne la « composition », alors que celui de De Luca se situe dans la « performance » et dans la « synthèse ». Or, l'assemblage et la combinaison de matériaux qui sont faits dans la base de données sont tout-à-fait différents de celles effectuées par Spanu Fremder pour reconstituer la théâtralité de certaines pièces, mais peuvent participer à compléter ce genre d'analyse. En effet, ni nos données ni nos visualisations ne permettent à elles seules de reconstituer une pièce ancienne, mais proposent l'accès à des informations sur le Répertoire, pour toutes les phases de l'existence d'un spectacle<sup>70</sup>.

En dehors des études théâtrales, la base de données peut plus largement être mobilisée dans les études historiques, car le Répertoire est le résultat de dynamiques complexes entre la Cour royale, le peuple et les acteurs.ices du théâtre (un schéma à la Figure 58). Agrégeant des données concernant non seulement les pièces, mais aussi l'économie du théâtre (les dépenses et recettes, les abonnements), les spectateurs<sup>71</sup> et les récits de la presse (à travers les notes d'Emanuele De Luca<sup>72</sup>), la base de données peut trouver une utilité dans des recherches en histoire sociale et culturelle, y compris en histoire culturelle des élites.

---

<sup>70</sup> Le projet RECITAL couvre, en revanche, de manière complète tout ce que l'on peut connaître à travers les Registres sur la phase de « composition ». Les deux bases de données sont donc, en quelque sorte, complémentaires.

<sup>71</sup> Même si ces données sont actuellement très partielles, la structure de la base de données permet d'ajouter des informations précises sur le nombre et le type de spectateurs (parterre/loges).

<sup>72</sup> Les notes ne sont actuellement que partiellement accessibles, car leur forme n'est pas adaptée au medium base de données, mais il est envisageable, non seulement d'adapter leur forme, mais aussi de créer un système d'annotation permettant d'analyser de manière automatique et précise ces textes.

En réunissant des documents très différents du point de vue historiographique (le répertoire de Brenner et celui de De Luca), créés à cinquante ans de distance, avec des outils et des objectifs de recherche spécifiques, la base de données fournit des informations concernant toutes les entités qui participent à la création du Répertoire : les acteurs et actrices, mais aussi la Cour royale, la presse et le peuple, destinataire final de la stratégie du roi et des acteur.ices. La base de données réunit donc dans un même endroit des informations sur l'intérieur et sur l'extérieur du théâtre, sur la production et sur la réception des spectacles : même si les données ne sont pas exhaustives, elles racontent la dimension politique de la production artistique et culturelle d'un théâtre d'Ancien Régime, mettant au premier plan la relation entre culture, arts et société.

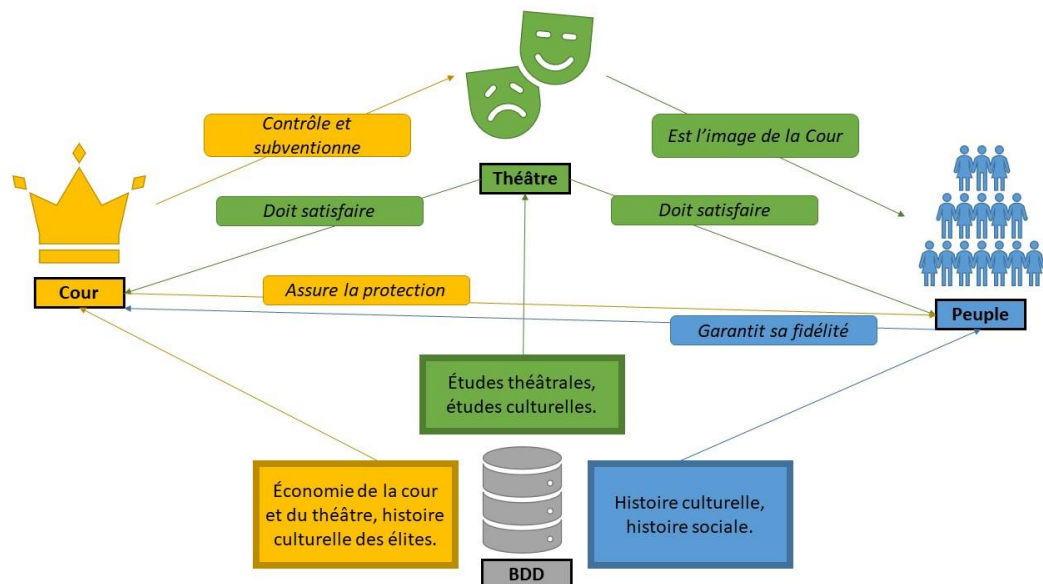


Figure 58 : Fonctionnement sociétal du théâtre et contributions de la base de données.

A propos plus spécifiquement du corpus numérisé, un positionnement inverse a été pris par rapport aux projets qui ont l'ambition de traiter un maximum d'informations sur un sujet. Concernant le projet Society Theater Project, Maria Teodora Comsa écrit que

« la pratique la plus pertinente consiste à inclure autant d'aspects que possible dans le schéma des données, tout en sachant que certaines catégories ne pourront jamais être complétées, soit parce que les données ne sont pas disponibles, soit parce que la quantité de sources s'étend au-delà du corpus établi, soit parce qu'elles ne seront peut-être pas utiles dans la visualisation. Tout compte fait, il est préférable d'avoir des données à éliminer qu'à ajouter lors des phases finales d'un projet numérique » (COMSA, 2020, p. 212).

Dans notre projet, ce choix a été fait au niveau du modèle abstrait, qui est apte à représenter une bonne partie de l'activité de la compagnie, mais qui n'a été finalement implémenté que pour les aspects qui allaient être effectivement traités – puisque, étant donné notre contexte de travail restreint, saisir des données qui se révèlent ne pas être exploitables peut avoir un coût considérable.

Le fait de décider à l'avance d'un ensemble de phénomènes et d'informations à traiter semble en effet pertinent pour plusieurs aspects. Même si cela ne permet pas de saisir toute la complexité historique de notre objet (et on pourrait se demander s'il existe vraiment un ensemble d'informations qui le permettrait), un ensemble de données limité mais cohérent peut plus facilement être repris en main, avec des risques moindres d'accomplir un travail qui n'est finalement pas utilisable. D'une certaine manière, circonscrire les données rapproche leur traitement aux capacités humaines d'élaboration des informations. Dans notre cas, se concentrer sur la programmation du Répertoire chronologiquement a permis d'avancer dans le travail, tout en ayant des données cohérentes dans les différentes étapes : nous avons tout d'abord saisi toutes les informations année par année jusqu'à 1750 et vérifié les métadonnées des pièces, avant de nous concentrer sur la fin du siècle ; la possibilité de saisir les dépenses et recettes ne s'est ainsi présentée qu'une fois ce travail terminé, et a été rapidement accomplie. Le traitement détaillé des informations concernant les personnes a aussi été reporté à une phase ultérieure du projet, puisque cela demandait l'utilisation de sources et d'une méthode de recherche des informations très différentes. Mais surtout, traiter un maximum d'informations ne garantit pas en soi de traiter toutes les informations qui sont utiles dans un certain contexte : j'ai moi-même fait l'expérience de cela dans le projet d'édition numérique du corpus de l'écrivaine Elisa Chimenti (VALLARANO, DESIDERI, 2024). Lors de la première expérimentation d'interface d'édition, un maximum de phénomènes avait été encodé, générant une complexité de traitement très grande, sans qu'on puisse, par exemple, relier entre elles les différentes graphies des noms des personnages – alors que cette fonctionnalité aurait été utile dans plusieurs recherches.

Du point de vue historiographique, le choix d'un ensemble de données cohérent et circonscrit facilite l'évaluation des lacunes et des manques dans les données, car leurs limites sont connues précisément. D'ailleurs, c'est l'interprétation humaine qui fait parler les données :

« ce que peut [...] mesurer l'ordinateur n'est pas le "réel historique", mais une partie extraite, formalisée, construite de celui-ci [...] un ensemble réduit et fini de données extraites de l'ensemble de données infinies qu'est le "réel historique [...] il faut [donc] se prémunir de

l'illusion référentielle qui consisterait à croire que l'ordinateur peut procéder à des calculs sur des sources brutes et produire naturellement des connaissances historiques sur le "réel passé" ou même le "réel historique" sans intervention de l'historien » (MOUNIER, 2018, p. 108)

À la lumière de ces considérations, tout en étant précautionnés sur la manière d'afficher les données pour qu'elles ne suggèrent pas des interprétations contradictoires et erronées par rapport à ce qu'on connaît de la réalité historique, et tout en décrivant le plus précisément possible notre modèle, notre recherche confie en les capacités interprétatives des utilisateur.ices pour la manipulation et visualisation des données, afin de les interpréter dans leurs contextes de recherche.

Les cas d'étude décrits présentent néanmoins un grand manque : nous n'avons pas exploré la possibilité de manipuler au même temps notre jeu de données avec d'autres données comparables, comme par exemple celles des Registres de la Comédie-Française. Dans le cas d'étude concernant les parodies, seulement les résultats ont été comparés de manière superficielle, sans manipuler les données de Beaucé. Or, permettre la manipulation et l'utilisation des données dans des contextes différents constitue l'un des grands intérêts d'un traitement de données comme le nôtre. Tout en étant attentif à la signification des données et donc à la manière dont deux jeux de données peuvent être comparés, leur forme numérique permet en elle-même, théoriquement, de traiter toutes les informations sur un même support (modulo des éventuelles transformations de format, qui peuvent être en réalité très complexes), ce qui représente un changement fondamental par rapport aux recherches traditionnelles, en termes d'accès à la connaissance.

#### **4.1.2 Transdisciplinarité**

Avec la création d'une base de données, si le résultat ne peut pas être une représentation complète et exhaustive de l'objet historique, nous nous attendons au moins à une description précise et désambiguïsée de l'aspect qui est étudié de cet objet, puisqu'un certain pouvoir de communication est reconnu à la formalisation logique-informatique (MOUNIER, 2018 ; MOUZA et al., 2023). Du point de vue opérationnel, le besoin de formalisation a rendu nécessaire de préciser de manière non ambiguë ce que représente chaque table, chaque relation et chaque attribut, ce qui a permis de rendre visibles des biais disciplinaires (comme le montre la section sur les Conséquences des objectifs sur la modélisation). La désambiguïsation et la visibilité des biais peuvent par ailleurs être favorisées par un contexte de travail multi-

disciplinaire : si ce type de travail existait bien sûr dans les lettres avant la diffusion des humanités numériques – les études théâtrales en témoignent – elles en changent l'échelle, car les équipes sont plus nombreuses, et les domaines concernés sont plus éloignés méthodologiquement et épistémologiquement (MOUNIER, 2018). Le fait de devoir communiquer entre spécialistes de domaines différents, connaissant des méthodes différentes et utilisant des vocabulaires différents, nécessite d'explicitier les termes, les objectifs, les enjeux. Cela est source d'interactions fructueuses entre ces spécialistes, dont l'une des conséquences est l'explicitation des termes : même si les recherches ne sortent pas entièrement de leurs cloisons disciplinaires, devoir expliciter a l'avantage de rendre visibles les présupposées, et donc de travailler en montrant sa propre position à l'intérieur de la discipline (MOUZA et al., 2023).

Le contexte de travail en humanités numériques a d'autres conséquences sur l'organisation des recherches en ce qui concerne les activités humaines. La portée large des projets (dans notre cas, traiter le Répertoire de la Comédie-Italienne de Paris sur un siècle entier), les méthodes adoptées et l'agrandissement des équipes engendrent une plus grande division du travail, par rapport aux recherches traditionnelles. En effet, rendre les informations calculables représente un temps de travail très long, qui est le plus souvent divisé parmi plusieurs personnes pour pouvoir être accompli. Cela représente une caractéristique cruciale d'un projet d'humanités numériques comme le nôtre, qui le distingue de manière fondamentale des recherches traditionnelles. En effet, à l'étude pointue, personnelle et experte d'un corpus, on substitue un écosystème de tâches demandant des types différents d'expertises et une certaine complémentarité. Dans ce sens, la possibilité de standardiser et donc diviser et externaliser le travail est probablement le changement le plus déterminant dans le but de faire des analyses quantitatives à grande échelle en lettres. En d'autres termes, les outils numériques ne réduisent pas le temps de travail en soi, mais favorisent un partage du travail parmi plusieurs personnes qui l'accomplissent en parallèle, ou qui sont chargées de différentes tâches complémentaires (GEFEN, 2015 ; MOUNIER, 2018).

Or, travaillant dans un contexte d'études littéraires, même les tâches les plus répétitives, comme la saisie des données, demandent un certain degré d'interprétation (NYHAN, 2023). En effet, dans notre cas, le travail de saisie a consisté en une activité de synthèse non négligeable, car la personne qui saisit travaille sur trois supports : le document dactylographié de Brenner, le document en format Excel de De Luca et l'interface de saisie des informations. Cette activité

de synthèse demande un travail intellectuel d'organisation, de mémorisation, et de compréhension des notes rédigées par Emanuele De Luca. Il est nécessaire d'organiser la manière dont on insère les informations dans la base de données, page par page, par rapport à la forme du logiciel pour la saisie, qui permet de saisir les représentations de plusieurs façons (en l'occurrence, en créant une représentation après l'autre, ou à l'intérieur d'un objet soirée, en cherchant les noms des pièces, ou à l'intérieur d'un objet pièce, en cherchant les dates des soirées). Pour donner un exemple des différentes organisations possibles de ce travail, Chloé Petridis a procédé en individuant tout d'abord les différentes pièces sur chaque page, puis en saisissant toutes les représentations, pièce par pièce. Chiara Protani et moi, avons au contraire saisi les représentations à partir des soirées, ligne par ligne du répertoire. En d'autres termes, Chloé Petridis a ajouté des dates aux pièces, alors que Chiara Protani et moi avons lié les pièces à des dates. Umberto Desideri a suivi une méthodologie intermédiaire, consistant à saisir les pièces soirée par soirée, sauf si une pièce est jouée seule pendant plusieurs soirées de suite, ce qui le mène à ajouter ces dates à l'objet pièce.

Une double activité de mémorisation est aussi nécessaire : il faut retenir pour un temps court les pièces d'une soirée (ou les dates d'une pièce) afin de les saisir, mais il faut aussi retenir des informations plus complexes sur les pièces dans un souci de désambiguïsation. Pour faire face aux nombreuses ambiguïtés du Répertoire, il est parfois nécessaire de mémoriser, parmi plusieurs pièces au titre similaire, celle qui est jouée à un certain moment. Cette dernière activité de mémorisation dépend de la compréhension des notes de De Luca, dont la forme peut rendre complexe l'accès aux informations importantes pour plusieurs raisons. Premièrement, ces notes mentionnent les sources avec des codes, fonctionnels à la transparence de la recherche et à l'utilisation de filtres sur Excel, mais rendant moins fluide la lecture du texte. Deuxièmement, ces notes mélangent italien et français, ancien et moderne, ce qui demande la maîtrise de trois langages différents. Si le passage fondamental des informations aux données est assuré tout au long du travail de formalisation, c'est lors de l'activité de saisie que la transformation a véritablement eu lieu. Autrement dit, même la tâche la plus répétitive génère du *sens* à l'intérieur d'une chaîne de traitement numérique.

Un ultérieur aspect intéressant de cette nouvelle division du travail, est que cela rend visible le caractère collectif et social de la production de connaissance (CITTON et CUSSET, 2017), même dans les études littéraires où l'image de l'intellectuel seul avec ses livres stéréotype la manière dont les recherches sont menées. Lors du premier colloque auquel j'ai

assisté en tant que doctorante, après l'intervention d'un grand chercheur en études théâtrales, la modératrice remercie avec ironie la « générosité » intellectuelle des créateurs de bases de données, en faisant remarquer qu'une véritable recherche en humanités se fonde sur une interprétation personnelle et originale, et sur une connaissance profonde des sources et de la littérature : « moi, moi, je, je, moi » c'est ce qu'elle attend d'un résultat d'une recherche en lettres. Si même notre recherche insiste sur l'interprétation personnelle comme valeur dans les recherches en humanités numériques, ces dernières montrent aussi la dimension collective de la production de connaissances. Cela est mis en lumière non seulement par la composition interdisciplinaire des équipes, mais aussi par la manière dont les corpus sont choisis et par les enjeux de disponibilité des sources. En effet, le choix des corpus inscrit les chercheurs dans des traditions et des communautés interprétatives, qui ont progressivement avancé sur un certain sujet – phénomène dont notre projet est une illustration exemplaire. La méthode traditionnelle consistant à présenter un état de l'art au début d'une recherche témoigne, par ailleurs, de l'importance accordée au fait de se situer de manière cohérente et constructive par rapport aux connaissances acquises par la communauté dans un certain domaine, bien avant le développement du numérique.

Concernant la disponibilité des sources, elle représente un facteur clé dans les projets en humanités numériques : le fait de pouvoir repérer certains documents peut favoriser leur présence dans des corpus (le projet RECITAL est réalisable car les Registres de la Comédie-Italienne de Paris sont entièrement numérisés sur Gallica), et leur indisponibilité ou fragilité peut justifier un nouveau projet de numérisation (le projet d'édition numérique du corpus d'Elisa Chimenti répond, entre autres, à la volonté de préserver ces documents mal conservés et non consultables – à propos de ce projet voir par exemple VALLARANO et DESIDERI, 2024a et 2024b). Les projets s'inscrivent ainsi explicitement et de manière dynamique dans un certain écosystème documentaire, profondément social : tout nouveau projet est utile ou pertinent s'il sait se situer de manière créative et constructive par rapport à ce qu'on sait et à ce qu'on veut savoir.

Afin de visualiser la configuration humaine et institutionnelle qui a permis la réalisation de notre projet, j'ai adapté à ce contexte la proposition qu'Arthur Perret fait dans un article sur la documentation de la construction sociale de la connaissance avec des graphes documentaires (PERRET, 2023). Ces derniers sont des « ensemble[s] de documents ou fragments de documents interreliés », que nous pouvons concrètement créer avec le logiciel Cosma, conçu

par Perret-même. Pour construire un tel graphe suivant la proposition du chercheur, on peut remplir des fiches (qui sont ici des fichiers au format Markdown) et les annoter avec une annotation légère en YALM, Markdown et Wikitext pour exprimer les liens entre les fiches. Cela permet de visualiser le graphe documentaire dans une page HTML. Afin de rendre compte de la construction sociale de la connaissance au moyen d'un graphe documentaire, Perret propose d'intégrer des fiches qui représentent le « milieu de savoir », c'est-à-dire l'ensemble des personnes, objets et phénomènes qui sous-tendent la production et l'existence d'un savoir. Il s'agit donc notamment des espaces réels et virtuels qui permettent à la connaissance de s'organiser, mais aussi des objets techniques qui existent dans ces espaces, et des personnes qu'y agissent ; on peut ajouter aussi d'autres « éléments mémorables », c'est-à-dire des éléments qui permettent de se souvenir d'un phénomène, et dont la présence décrit le phénomène-même, modifiant par-là la manière de l'approcher.

L'intérêt de cette méthode de construction de graphes documentaires est multiple selon Perret :

1. Le graphe aide l'individu qui le crée à réactiver les liens entre les connaissances contenues dans le graphe ;
2. Le graphe permet d'observer le positionnement intellectuel de l'individu qui l'a créé ;
3. Le graphe représente un système d'organisation des connaissances (SOC) subjectif, itératif et construit par composition de concepts.

Cette proposition nous intéresse davantage pour observer la construction sociale de la connaissance que comme aide à la mémorisation, et notre application sera donc focalisée sur cet aspect. Ainsi, afin de mieux observer les positionnements des membres de l'équipe dans le projet, mais surtout afin de se rapprocher davantage d'un système d'organisation des connaissances, des contraintes ont été ajoutées au graphe. Concrètement, la classification des fiches a été conçue avant la création des fiches (il y a donc des documents, des personnes, des institutions, des équipes, des activités, des domaines et des objets numériques, et seulement les documents, les personnes, les équipes et les domaines peuvent avoir d'autres types associés, de la manière indiquée dans la Figure 59). L'autre contrainte est que chaque type de fiche peut contenir des liens uniquement vers un ensemble précis de types de fiches, comme représenté dans le schéma de la Figure 59. Cela définit un ensemble de types autonomes, et des ensembles de types contraints, qui peuvent être associés à certains types autonomes pour les caractériser.

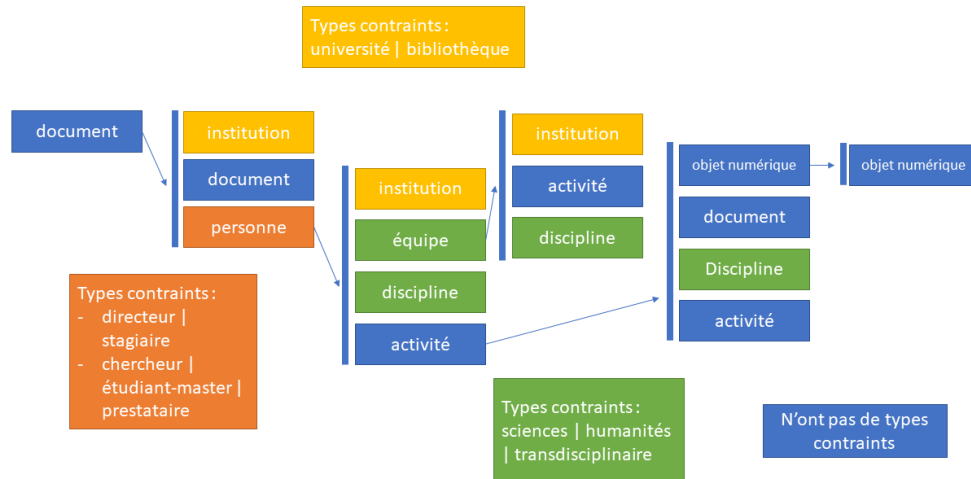


Figure 59 : contraintes appliquées dans le logiciel Cosma.



Figure 60 : graphe représentant notre activité, créé avec le logiciel Cosma.

Or, les contraintes que nous avons choisies sont fortement adaptées à notre contexte, mais il serait envisageable de définir une grammaire commune aux humanités numériques. De cette manière il serait possible de mettre en relation les constructions de la connaissance dans différents projets, au moyen de graphes comme celui généré pour notre projet (Figure 60). Cette

modélisation fait apparaître trois groupes d'activités : vérification et saisie des données d'un côté ; programmation, développement web, hébergement web et maintenance du côté opposé ; dans une position intermédiaire il y a les analyses des données, la modélisation et la conception de l'interface. Les sources et les personnes relevant du domaine littéraire se situent proches des activités de saisie et vérification ; alors que Camilla Cantiani (l'ingénieure informatique) est la seule personne proche des activités plus informatiques. Les personnes relevant des humanités numériques se situent sur un axe central.

Si la position des activités et des personnes semble donc cohérente avec le déroulement réel de la recherche, la position des objets numériques est en revanche à discuter. La version Excel du répertoire d'Emanuele De Luca, en particulier, devrait être dans une position plus centrale, surtout en relation au répertoire PDF. La position relative de l'interface de saisie, de la base de données, de l'interface de consultation et de l'architecture du site internet semblent, de leur côté, fonctionnelles. Cela dépend probablement des contraintes, autrement dit de la modélisation : même si le répertoire de De Luca est tout-à-fait central pour le projet, il n'est en revanche lié qu'à l'activité de saisie des données et a donc peu de liens dans le graphe, donnant une place plus centrale à la version PDF.

Malgré ces incohérences, qui peuvent être réduites avec de nouveaux ensembles de contraintes et avec la confrontation avec d'autres projets, ce graphique permet d'observer le travail de recherche dans ses aspects organisationnels et dans sa construction sociale. Les nœuds du graphe peuvent par ailleurs être cliqués, pour obtenir leur description et la liste de tous les liens, entrant et sortant. Les fichiers permettant de générer ce graphique étant conservés dans un repository Framagit ouvert, il s'agit d'un véritable outil pour comprendre notre recherche, qui s'ajoute à tous les documents et plateformes concernant la base de données dans ses aspects les plus techniques. Rendre compte de la dimension collective de nos projets nous situe par ailleurs dans

« une certaine aspiration politique de démocratisation des savoirs [qui] semble inhérente au mouvement global des humanités numériques [...] Malgré l'association imaginaire des humanités numériques à des grosses machines et à des gros budgets suspectés d'emprises impérialistes, les praticiens insistent à présenter leur travail comme relevant d'un bidouillage (*hacking*) fortement artisanal qui induit un redéploiement des pratiques interprétatives, tout au contraire d'entraîner leur liquidation dans des flux de données gérées en traitement automatisé » (CITTON et CUSSET, 2017, p. 439-40).

## 4.2 Questions méthodologiques et épistémologiques

### 4.2.1 Observations sur l'automatisation et sur l'IA

Visant une humanisation de l'informatique (opposée à une informatisation de l'humain), nous avons réduit le plus possible l'utilisation des logiciels qui fonctionnent avec l'intelligence artificielle, afin d'aborder l'automatisation de manière transparente et créative. Utiliser l'intelligence artificielle pose aussi des questions éthiques, liées à la manière dont ces logiciels sont produits et fonctionnent (CRAWFORD, BURY, 2022). Mais avant même de déclarer un positionnement éthique par rapport à l'intelligence artificielle, nous pouvons questionner l'automatisation dans notre cadre concret de travail.

Notre base de données est construite principalement à partir de deux documents, dont l'un est numérisé (le répertoire de DE LUCA, 2011, dont une version en PDF est disponible pour la consultation), et l'autre est dactylographié et structuré avec des colonnes (le répertoire de BRENNER, 1961). En apparence, ces deux documents se prêtent donc bien à un traitement totalement automatisé : un logiciel d'OCR pourrait être utilisé pour transcrire le deuxième document (en s'attendant à obtenir des bons résultats vue sa forme et son état de conservation), avant d'écrire un programme qui insère automatiquement les dates en les associant aux entrées de De Luca. Automatiser la procédure aurait demandé à deux reprises de corriger les erreurs générées par des programmes : premièrement la sortie de l'OCR, et ensuite les résultats des associations entre les deux documents. Ainsi, le gain de temps par rapport à insérer les informations à la main n'est pas évident a priori. Nous avons donc décidé de procéder avec un travail manuel, pour des raisons méthodologiques.

Premièrement, les algorithmes ne sont pas toujours transparents et le fonctionnement de l'intelligence artificielle l'est encore moins. Cela peut être un frein pour toute recherche, si l'on veut décrire les opérations qui sont faites à chaque étape du projet, mais l'est d'autant plus si, comme dans notre cas, l'un des objectifs est justement d'observer et d'analyser la procédure. Minimiser la présence de boîtes noires informatiques peut donc être un critère pertinent de choix. Néanmoins, il a semblé acceptable d'employer des bibliothèques Python open-source par leur grande simplicité d'apprentissage et de prise en main, même si elles peuvent favoriser une

utilisation de l'informatique comme boîte noire<sup>73</sup>, car, étant ces bibliothèques open-source, il est toujours possible de reconstituer le détail des opérations qui sont faites sur l'input afin d'obtenir l'output, car l'information est accessible. Cela n'est pas toujours le cas avec l'intelligence artificielle, surtout s'il s'agit de *deep learning*, où la machine est programmée pour faire des ensembles d'éléments, mais ne sais pas à priori selon quels critères deux éléments vont dans le même ensemble ; les résultats de la machine sont ensuite corrigés par des humains pour qu'elle recommence la procédure à partir des corrections, jusqu'à obtenir un taux d'erreur suffisamment bas. Ainsi, le fonctionnement général est connu, mais il n'est pas possible de savoir exactement comment, ni selon quels critères, la machine choisit. Comprendre ses critères de choix demande d'émettre des hypothèses et de chercher à les vérifier par *rétro-ingénierie*. Ainsi, dans un souci de transparence, si la tâche et la quantité de données à traiter le permettent, il est pertinent de pencher du côté d'une utilisation minimale de ces logiciels dans les recherches en lettres.

Deuxièmement, un logiciel d'intelligence artificielle fonctionne bien si ses résultats ont été corrigés par des humains de nombreuses fois. Or ce travail est le plus souvent invisibilisé :

« des recherches en sociologie sur le « travail du clic » montrent que l'automatisation reconfigure la division mondiale du travail, plus que ce qu'elle libère l'humanité du labeur pénible (CASILLI, 2019 ; LE LUDEC, 2024) En effet, ce dernier est confié à des travailleur.euse.s peu qualifié.e.s, qui accomplissent des micro-tâches sous-payées, afin de permettre aux machines de mieux simuler le travail humain. L'invisibilisation du travail du clic accompagne les humanités numériques depuis l'expérience pionnière de Roberto Busa, avec l'Index Thomisticus (NYHAN, 2023) : on attribue souvent la réalisation de la célèbre numérisation de l'œuvre de Thomas d'Aquin au génie de Busa, alors que ce long travail a impliqué, non seulement l'expertise d'ingénieurs d'IBM (MOUNIER, 2018), mais aussi le travail de dizaines de femmes qui ont créé, pendant près de 15 ans, les fiches perforées, c'est-à-dire les objets qui représentaient concrètement les textes » (DESIDERI et al., 2025).

Or, la masse de données sur la Comédie-Italienne de Paris à traiter est grande mais n'est pas gigantesque, et le projet vise une certaine précision dans le traitement : les erreurs ne sont pas statistiquement négligeables, comme elles peuvent l'être dans des calculs d'occurrences de mots ou dans la reconnaissance d'entités nommées au sein de corpus constitués de millions de

---

<sup>73</sup> En effet, pour utiliser une fonction de ces bibliothèques, il suffit de savoir quelles informations lui donner (input) afin d'obtenir le résultat (output) souhaité.

mots. Les deux documents de départ sont le fruit d'un travail de reconstitution historique et de réorganisation d'informations, qui a composé avec les phénomènes de lacune et de manque typiques du travail d'historien.ne, générant d'ultérieures lacunes et manques : Brenner fait des erreurs en transcrivant les titres des pièces, et hérite les manques et les ambiguïtés des sources historiques sur lesquelles il travaille ; De Luca accomplit de son côté un remaniement des informations de sources disparates et organise leur forme afin de pouvoir filtrer les informations avec les outils d'analyse d'Excel. Mais surtout, Brenner fait un effort de normalisation des titres vers le français, alors que De Luca les développe et les accompagne du titre en italien. Ainsi, si nous devons reconstituer ces liens à la main, la simple transcription du document de Brenner perd en grande partie son intérêt. Devant le choix entre employer un temps long à saisir directement les données, ou à corriger les résultats de l'automatisation, la première solution a été préférée.

Troisièmement, le travail répétitif de saisie n'est pas sans intérêt, comme déjà remarqué plusieurs fois depuis le début de ce texte : il demande une activité intellectuelle spécifique, et il permet en plus de s'approcher de l'objet d'étude depuis un angle particulier. En d'autres termes, un travail répétitif n'est pas seulement mécanique. Dans son livre sur le travail invisibilisé des femmes pour la construction de l'Index Thomisticus (NYHAN, 2023), Nyhan montre comment l'activité de création physique des cartes perforées, qui leur était confiée, demande la mobilisation de plusieurs genres et niveaux d'expertises, alors qu'elles étaient employées comme travailleuses non qualifiées. Elle affirme même que « la dévalorisation de ce travail ne peut reposer sur la seule évaluation de son caractère interne »<sup>74</sup> (NYHAN, 2023, p. 34). Une idée similaire est défendue aussi dans l'article « Les mains dans les bases de données » des anthropologues Dagiral et Peerbaye (DAGIRAL, PEERBAYE, 2012), qui porte sur un tout autre type de base de données (à savoir, la base de données pour les maladies rares Orphanet).

La répétition permet de remarquer des phénomènes, de se souvenir des noms des pièces les plus récurrentes, de se poser des questions... Cette activité résonne avec les récits des recherches dans les archives, dans une perspective micro-historique : à l'intérieur de la répétition, de l'ordinaire, une certaine intuition peut se développer, permettant de remarquer des schémas, et donc aussi ce qui sort de la norme et de ces schémas. Arlette Farge écrit que « le goût de l'archive passe par ce geste artisan, lent et peu rentable, où l'on recopie les textes, morceaux après morceaux, sans en transformer ni la forme, ni l'orthographe, ni même la

---

<sup>74</sup> "The devaluing of this work cannot have been based on an evaluation of its internal nature alone."

ponctuation » (FARGE, 1997, p. 25) et qui permet de plonger dans un monde ancien, dont l'archive constitue l'une des seules traces qui restent. Elle revient ensuite sur cette idée de la répétitivité :

« À partir de cette lecture obstinée s'organise le travail. [...] Cela commence tout doucement par des manipulations presque banales sur lesquelles il est finalement rare de réfléchir. Pourtant, en les accomplissant, se fabrique un objet nouveau, se constitue une autre forme de savoir, s'écrit une nouvelle « archive » »

qui rend « possible une autre narration du réel » (FARGE, 1997, p. 78-79). Or Farge prend explicitement les distances de l'informatique peu avant la première citation, affirmant que « au temps de l'informatique, ce geste de copie, à peine peut se dire » (FARGE, 1997, p. 24). En effet, l'approche de la micro-histoire est tout-à-fait différente par rapport à celle quantitative, à laquelle elle fait référence dans la citation : pour la première il s'agit de faire revivre des vies oubliées, par le peu de traces que les autorités ont laissé d'elles, alors que la deuxième cherche des macro-structures. Les méthodologies ne sont pas les mêmes non plus. Mais ce que montre le procédé de la micro-histoire – tel que décrit par Arlette Farge, et par exemple aussi par Jérémie Foa dans *Tous ceux qui tombent* (FOA, 2021) – est qu'il existe des gestes répétitifs qui génèrent du sens. En informatique, cela est le cas aussi. Il n'est donc pas toujours pertinent de confier ce travail aux machines.

#### 4.2.2 Epistémologie des Humanités Numériques

Le point de vue informatique qui a été pris pour approcher les données, situe ce travail dans ce que Franco Moretti décrit, dans son dernier livre (MORETTI, 2022), comme des humanités numériques qui priorisent l'attention à la « morphologie », plutôt qu'au contenu, car on observe les « opérations et interactions des éléments structurels »<sup>75</sup> pour les poser « *entre l'observateur et le monde* »<sup>76</sup> (ibid., p. 108). Étudier la structure des productions informatiques met aussi l'humain au centre de la recherche, puisque cela permet d'observer comment et pourquoi les êtres humains font usage des machines pour accomplir une activité interprétative.

Dans cette recherche, la base de données intervient comme un appui à des hypothèses qui sont faites sur la Comédie-Italienne par des chercheur.euses spécialistes de la matière, mais

---

<sup>75</sup> «Operazioni e interazioni degli elementi strutturali.»

<sup>76</sup> «*Tra l'osservatore e il mondo.*» (En italique dans le texte d'origine.)

aussi comme un objet d'étude : elle n'est pas acceptée comme un outil neutre, qu'il suffirait d'appliquer à un sujet pour qu'il révèle des phénomènes plus ou moins attendus. Si on imagine un spectre des recherches en humanités numériques littéraires qui va des éditions numériques (où les aspects qualitatifs sont prédominants) au traitement automatique de corpus (orienté à la quantification), notre base de données se situerait dans une position centrale, parmi ces deux extrêmes. Dans les éditions numériques, les détails sont importants, et l'intervention humaine donnant la signification des phénomènes ne peut pas être remplacée (ou ne peut l'être que partiellement) par le travail des machines. Même si la quantité d'informations est très grande, il est nécessaire d'accomplir un travail précis afin d'attribuer du sens au texte en entier, ainsi qu'aux phénomènes qu'on a décidé de traiter. Dans les études quantitatives qui prévoient le traitement de grands corpus, en revanche, on travaille avec des masses de données où l'on tend à maximiser la précision, sans pouvoir l'atteindre totalement, et où les erreurs peuvent être rendus statistiquement négligeables. Dans les éditions numériques, l'ordinateur est davantage utilisé pour les possibilités qu'il ouvre en termes de visualisation et de connexion entre les informations, et a ainsi la fonction de suggérer un récit d'une manière spécifique au nouveau médium (par exemple avec une lecture non linéaire, ou avec la possibilité de voir au même temps plusieurs versions d'un texte). À l'inverse, dans les études quantitatives des corpus, l'ordinateur est davantage utile pour ses apports en termes de vitesse de calcul et donc pour la possibilité de traiter des quantités de données qui ne sont pas saisissables avec les méthodes quantitatives traditionnelles. Or dans un cas comme le nôtre, ces deux aspects sont déterminants : la base de données permet de mettre en relation, sur un même support, des informations qui étaient auparavant sur des supports différents, mais aussi de faire des calculs sur ces données qui étaient possibles mais déraisonnables sans l'appui d'une machine (les calculs de distances et de fréquences, par exemple, sont des analyses qu'on peut accomplir à la main en théorie, mais en pratique cela serait long, complexe et avec un grand risque d'erreurs).

Franco Moretti, célèbre pour avoir défendu le concept de *distant reading* (MORETTI, 2013), consistant à étudier la littérature à travers la manipulation de grandes quantités de données avec des allers-retours entre quantitatif (loin des textes) et qualitatif (proche des textes), a publié en 2022 un livre où il revient sur le travail précédemment accompli, où il prône un retour à la théorie :

« "Data driven" signifie deux choses : qu'une grande quantité de données peut aiguiller de manière puissante la recherche, et cela est vrai ; et que la recherche-même peut être guidée par les données, et cela est faux. Les outils avec lesquels on travaille sur les données dépendent

toujours d'une théorie : s'il n'y a pas de théorie, sa place sera prise, fatalement, par des lieux communs ambiants »<sup>77</sup> (MORETTI, 2022 , p.12-13)

Et à la fin du livre il revient sur cette idée :

« Avec l'arrivée des bases de données et des algorithmes, l'aspect statistique a pris des dimensions gigantesques, et sa capacité à générer des résultats - à toujours sortir quelque chose de nouveau et d'inattendu de la masse de données - l'a amené à dominer la manière concrète dont la recherche avance. On ne percevait aucun besoin de théorie. Et donc, il n'y a eu aucune théorie. »<sup>78</sup> (MORETTI, 2022, p. 146-47)

Autrement dit, l'utilisation d'outils et méthodes informatiques doit être organisée de manière cohérente et en continuité avec les objectifs de recherche en lettres, afin de garantir une continuité méthodologique, car « l'interprétation d'un phénomène constaté n'est possible que par la compréhension du contexte de son avènement » (MOUNIER, 2018, p. 105), comme remarque Pierre Mounier dans un chapitre sur « La fin de la théorie ». Il s'agit par ailleurs d'une question souvent abordée dans les recherches en humanités numériques. Selon Ciotti (CIOTTI, 2023), privilégier les questions méthodologiques et épistémologiques est caractéristique de l'un des paradigmes qui ont construit les humanités numériques, et le plus ancien. On constate aussi une plus grande présence de publications autour de l'épistémologie dans les humanités numériques dans les années 2010-2015 (par exemple CITTON, 2015, GEFEN, 2015, mais aussi les articles du volume *Reading Graphs, maps, trees: responses to Franco Moretti*, GOODWIN et HOLBO, 2011) qu'aujourd'hui, où ces évaluations sont faites au sein des différents projets et sont plus rarement abordées de manière générale (en témoigne GALLERON, 2017). Parmi les chercheur.euses en humanités numériques, le fait que l'utilisation de l'informatique change le statut épistémologique des recherches en lettres est un constat aujourd'hui largement partagé, et les chercheur.euses observent donc plutôt comment cela se manifeste concrètement dans les différents domaines (en témoignent les numéros 4, 5 et 9 de la *Revue d'Historiographie du Théâtre* dédiés aux études théâtrales et aux humanités numériques).

---

<sup>77</sup> “"Data driven" vuol dire due cose : che una gran quantità di dati può agire come un pungolo potente alla ricerca, il che é vero; e che la ricerca stessa può essere guidata dai dati, il che é falso. Gli strumenti con cui si lavora sui dati dipendono sempre da una teoria: se una teoria non c'è, a prenderne il posto saranno, fatalmente, dei luoghi comuni che circolano nell'aria.”

<sup>78</sup> “Con l'avvento dei database e degli algoritmi, l'aspetto statistico si é [...] ingigantito, e la sua capacità di generare risultati – di tirar fuori sempre qualcosa di nuovo e di inaspettato dalla massa dei dati – lo ha portato a dominare il concreto procedere della ricerca. Efficientissima e piena di sorprese, la nuova pratica sembrò bastare a se stessa. [...] Non si sentiva alcun bisogno di teoria. E dunque, non c'è stata alcuna teoria.”

Dans un article sur « Les enjeux épistémologiques des humanités numériques » (GEFEN, 2015), Gefen soutenais en 2015 que « le geste épistémologique le plus fort des travaux de Moretti est d'avoir tenté d'objectiver, pour les abstraire, les différentes variables définissant les œuvres littéraires (genres, formes, etc.) afin de permettre des changements d'échelle d'analyse » car ce travail fait « interroger des corpus qui ne hiérarchisent pas à priori les entités informationnelles » (GEFEN, 2015, p.6) et donc d'observer les objets sans *pré-jugé*. Or, avant d'être calculées et visualisées, les informations de départ doivent en réalité être analysées et modélisées, afin d'être traitées de manière automatique. Puisque ce travail a un caractère interprétatif, une valeur d'objectivité ou de plus grande vérité ne peut pas être attribuée à priori aux résultats de la modélisation, par rapport à des informations historiques multidimensionnelles. Au contraire, « il s'agit [...] d'incarner des idées abstraites, de les « formaliser » si l'on donne à ce verbe le sens de « mettre en forme » » (MOUNIER, 2018, p. 79). Par exemple, lorsque l'on doit définir ce qu'est une pièce dans un modèle de données, il est nécessaire de clarifier (et figer) s'il s'agit de l'édition d'une pièce, ou du texte « virtuel » (terme employé par SPANU FREMDER, 2010) qui était joué par les acteur.ices, ou bien s'il s'agit de l'ensemble texte-performance etc., et quelles sont les relations entre ces objets. Si l'on veut tenir compte de toutes ces dimensions, le modèle doit être suffisamment complexe pour les décrire, et pour décrire leurs relations.

Parmi les publications concernant l'épistémologie des humanités numériques, l'essai de Cosma Shalizi (SHALIZI, 2011) dans *Reading Graphs, maps, trees* (volume de réponse à un livre de Moretti, publié en français avec le titre *Graphes, cartes et arbres : Modèles abstraits pour une autre histoire de la littérature*) est particulièrement éclairant à propos des méthodes quantitatives. Le statisticien y soutient que « si l'on se donne la peine de collecter des données quantitatives, il faut les utiliser de manière quantitative. L'abstraction mathématique (quantitative ou autre) n'a pas de valeur en soi, mais pour les déductions qu'elle nous permet de faire, lorsque les outils appropriés sont appliqués »<sup>79</sup> (SHALIZI, 2011, p. 119). Il ne suffit pas de construire des modèles et des expériences selon des méthodes et des théories, mais il faut aussi prouver le lien de causalité entre les phénomènes qu'on met en relation, ce qui est particulièrement difficile en littérature car, malgré son approche marxiste, Shalizi admet que « la culture n'a pas à exprimer ou à refléter l'ordre social. Je soupçonne que Moretti serait déçu

---

<sup>79</sup> "If one is going to go to the trouble to collect quantitative data, one ought to use it quantitatively. Mathematical abstraction (quantitative or otherwise) is not valuable for its own sake, but for the inferences it lets us make, when the proper tools are applied."

si c'était le cas, par exemple, pour les genres des romans. Moi aussi. Mais cela doit être vérifié. Une façon de le faire serait d'essayer de développer de bons modèles neutres et de voir si, et où, ils échouent »<sup>80</sup> (SHALIZI, 2011, p. 127). Un modèle neutre en statistique sert pour vérifier que nos résultats ne sont pas dus au hasard et qu'une certaine valeur dépend effectivement d'un certain paramètre. Le principe est de créer un modèle où l'on fait varier les paramètres, pour voir ce qui change en relation à chaque paramètre, ou en relation à un ensemble de paramètres. À ma connaissance, ce genre de vérification n'a été appliquée dans une recherche en humanités numériques que dans l'article de Cosma Shalizi. Cela représente un manque de rigueur méthodologique du point de vue du statisticien. Mais cela suggère aussi que les données dans les humanités numériques ne sont pas toujours pertinentes à analyser avec les méthodes statistiques, et cela est le cas pour notre Répertoire, où elles sont trop homogènes et pas assez étendues.

Les statistiques se basent sur le principe qu'observer un phénomène de très nombreuses fois permet de détecter une forme de régularité dans le hasard. Shalizi explique cela clairement :

« la grande masse des détails de l'histoire à petite échelle est essentiellement aléatoire ou, plus exactement, les liens entre eux sont aussi alambiqués et imbriqués que les détails eux-mêmes (c'est l'une des façons de définir le hasard, mathématiquement). Mais à plus grande échelle, le hasard s'estompe, laissant place à des régularités plus simples et plus facilement compréhensibles par des esprits finis, et plus fiables. En tant que statisticien et physicien statisticien, je suis le dernier à ne pas être d'accord : « En fait, toute la valeur épistémologique de la théorie des probabilités repose sur le fait que les phénomènes aléatoires à grande échelle, dans leur action collective, créent une régularité stricte et non aléatoire » [(GNEDENKO, KOLMOGOROV, 1954)]. »<sup>81</sup> (SHALIZI, 2011, p. 131)

---

<sup>80</sup> “Culture doesn’t have to express or reflect the social order. I suspect Moretti would be disappointed if this were the case for, say, genres of novels. Well, so would I. But this needs to be checked. One way would be to try to develop good neutral models, and see whether, and where, they break down.”

<sup>81</sup> “The vast mass of details in small-scale history are essentially random, or, more exactly, the connections among them are as convoluted and involved as the details themselves. (This is one way to define randomness, mathematically.) But looking at larger scales, the randomness averages out, leaving regularities which are simpler and more nearly comprehensible by finite minds, and more reliable. As a statistical physicist and a statistician, I am the last to disagree: “In fact, all epistemological value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, nonrandom regularity” [(GNEDENKO, KOLMOGOROV, 1954)].”

Cela implique que si l'échelle n'est pas assez grande ou si les données sont trop homogènes, même si une forme de régularité semble se dégager dans les visualisations de ces données, il n'est pas possible de conclure uniquement au moyen des résultats statistiques.

Or, si des résultats quantitatifs ne sont pas vérifiés avec les méthodes traditionnellement employées pour cela, il faut d'autres méthodologies, qui permettent de reconstruire ce lien de causalité et pour pouvoir affirmer qu'un graphique *démontre* un phénomène. Suivant la tradition des recherches en lettres, chaque recherche en humanités numériques propose sa propre méthodologie avec ses critères d'inférence, déclarés de manière explicite au début de la recherche. L'intuition donnée par une connaissance profonde du sujet de leurs recherches, garantit le plus souvent aux chercheur.euses de pouvoir établir ces liens de causalité. Peut-on, pour cela, considérer que des recherches en lettres comme la nôtre (ayant trop de données pour une analyse purement qualitative, mais pas assez pour utiliser les statistiques) relèvent de la méthode scientifique expérimentale ?

La méthode scientifique expérimentale se base sur l'idée de produire une expérience pour qu'elle démontre une certaine hypothèse. En d'autres termes, si cette expérience donne ce résultat, alors l'hypothèse est correcte. Pour donner un exemple célèbre : Galilée étant convaincu que la lumière voyage avec une vitesse qui n'est pas infinie (et qui est donc mesurable) construit une expérience où la lumière doit parcourir une longue distance (la distance entre deux collines), pour observer un décalage entre la transmission de la lumière de la première colline à sa réception dans la deuxième colline. S'il observe un décalage, alors la lumière voyage à une vitesse finie. Mais le décalage ne se produit pas dans l'expérience de Galilée. Aujourd'hui nous savons que les collines n'étaient pas assez distantes l'une de l'autre et que l'intuition de Galilée était correcte, mais, lui, il doit admettre de ne pas avoir obtenu les résultats qui lui permettaient de démontrer son hypothèse. Or, le fait de ne pas avoir observé un décalage ne démontre pas non plus que la vitesse de la lumière est infinie. Ainsi, suite à l'expérience manqué de Galilée, deux options se présentent : ou bien la vitesse de la lumière est infinie, ou bien la distance entre les deux collines n'était pas assez grande pour observer un décalage dans son parcours (ce qui donnerait par ailleurs des informations intéressantes sur l'ordre de grandeur de la vitesse). Si en revanche son expérience avait réussi, il aurait pu affirmer assurément que la lumière a une vitesse finie – et aurait même pu calculer sa valeur approximativement. Ce qui est le plus intéressant du point de vue des humanités numériques est que la possibilité que la distance entre les collines ne soit pas suffisamment grande pour

observer un décalage, a ouvert la voie à d'autres expériences, où les scientifiques ont inventé des moyens pour augmenter la distance entre le point de diffusion et le point de réception, jusqu'à l'expérience de Fizeau en 1849 qui a permis la première mesure de la vitesse de la lumière. Cet exemple illustre que « ce qui est important dans la démarche expérimentale, ce n'est pas tant la validation d'hypothèses par des résultats positifs que la génération de nouvelles questions par des résultats négatifs » (MOUNIER, 2018, p. 63). Tout cela reflète la manière de procéder en humanités numériques littéraires, où l'on modélise des objets d'études afin d'en étudier des caractéristiques spécifiques ; les simulations statistiques, de leur côté, rendent compte du fonctionnement à large échelle d'un système et peuvent faire partie de la démarche si cela est pertinent.

Mais penser la scientificité des humanités numériques uniquement en relation aux sciences formelles et expérimentales ne permet pas de prendre en compte tous les enjeux épistémologiques, comme argumente Pierre Mounier dans son histoire critique des humanités numériques (MOUNIER, 2018) :

« Si le traitement en masse de données entraîne une mutation profonde de la méthode scientifique, son application aux disciplines des humanités ne contribue pas à rapprocher celles-ci du modèle classique des sciences de la nature, comme on l'entend souvent, mais à l'en éloigner. En passant des sciences humaines aux humanités numériques, c'est toute la relation des humanités aux sciences qui est redéfinie sur des bases qui ne sont pas celles d'une convergence directe, mais plutôt d'une mutation plus générale du fait de l'utilisation universelle des technologies numériques. » (ibid., p. 17)

Suivant Johanna Drucker, il reconnaît un pouvoir « spéculati[f] des humanités par l'expérimentation technologique » (ibid., p. 18), qui les différencie d'autres usages du numérique. Si le numérique est en réalité utilisé dans d'autres contextes comme méthode exploratoire (par exemple, on crée des modèles numériques en biologie pour observer les variations d'un écosystème selon la variation de diverses variables), la dimension spéculative est en lettres le caractère central des recherches et n'a pas le même objectif que dans les sciences expérimentales. En effet, pour ces-dernières, observer des variations sert pour confirmer, figer, une hypothèse, on recherche une forme de vérité (*le mieux qu'on peut connaître à un certain moment, avec les théories et les instruments à disposition*) ; alors que la dimension spéculative des productions des humanités numériques est donnée par des *interprétations qui suggèrent d'autres interprétations* (pour reprendre des expressions utilisées dans CITTON et CUSSET,

2017) : on ne montre pas une vérité absolue mais un point de vue – ce qui est loin d’être sans intérêt.

Notre recherche montre avec force à quel point objectifs de recherche, modélisation et visualisation sont fortement interconnectés au niveau épistémologique. La proposition de Schuwey dans *Interfaces* (SCHUWEY, 2019), consistant à définir la structure de l’interface de consultation au début d’un projet en humanités numériques, part aussi de cette hypothèse.<sup>82</sup> Drucker, connue pour ses recherches centrées sur les aspects visuels, aborde dans plusieurs publications le lien entre la visualisation de phénomènes et le discours qu’on veut porter sur ces mêmes phénomènes (par exemple DRUCKER, 2014 et 2011). Chercheuse et artiste, elle a critiqué la simplification qui est faite dans les visualisations des données en humanités numériques. L’exemple des correspondances dans le projet Mapping the Republic of Letters est connu : l’envoi d’une lettre d’un personnage choisi vers un interlocuteur est ici représenté comme une ligne légèrement courbée, entre le point de départ et le point d’arrivée, dans une forme similaire aux cartes représentant les parcours des avions. Cela suggère une simplicité dans le voyage de la lettre entre le point de départ et le point d’arrivée : la critique de Drucker porte sur le fait que cette suggestion ne permet pas de voir la réalité historique du voyage d’une lettre au XVIII<sup>e</sup> siècle. Nos visualisations ont essayé de tenir compte de ces propositions concernant la représentation graphique de l’espace et du temps. Par exemple, des visualisations par saisons ont été ajoutées à celles par année (comme suggéré aussi par Myriam Quatrini), afin de mieux rendre compte de l’unité réelle des informations, mais aussi pour visualiser la présence ou l’absence des registres – qui est par ailleurs exprimée par une bande représentant l’absence de ces documents, et donc l’absence d’informations sur le Répertoire (Figure 33). Néanmoins, les pistes que Drucker donne vont dans la direction d’une trop grande granularité dans la représentation, ce qui peut poser des difficultés au niveau de la complexité de calcul (éviter le traitement de données inutiles fait partie des principes pour un numérique écologique, tels que définis par le minimal computing, par exemple dans RADIO, 2020 ; RISAM et GIL, 2022, qui sera approfondi dans la section sur l’écologie), mais aussi d’un point de vue

---

<sup>82</sup> Je pense néanmoins que cette proposition est à prendre de manière flexible, afin d’éviter un problème potentiel au niveau méthodologique : s’il est toujours pertinent de définir les objectifs d’une interface de consultation au début d’un projet numérique, il faut aussi que sa forme exacte puisse changer selon les découvertes qui sont faites au cours de la recherche. On peut se rendre compte au cours de la recherche qu’un aspect qu’on voulait aborder dans l’interface est finalement trop complexe, ou qu’il ne donne pas d’informations intéressantes pour le public choisi ; ce genre de situation doit pouvoir modifier le projet d’interface initial.

méthodologique (comme abordé dans la première section de ce chapitre, à propos du choix d'un ensemble restreint de données).

Le lien fort entre objectifs, modélisation et visualisation dans les humanités numériques génère un environnement favorable aux biais d'auto-confirmation. En effet, en modélisant un objet pour un certain objectif, et suivant une certaine théorie, on peut s'attendre à ce que les résultats tendent à davantage confirmer cette interprétation. Dans l'introduction à la partie « Humanités numériques et théâtres de société » du volume *Espaces des théâtres de société*, Christophe Schwuey écrit que

« les humanités numériques ne génèrent pas de réponses automatiques. C'est en les utilisant pour répondre à des questions concrètes qu'elles démontrent leur efficacité. En proposant de nouveaux modes de consultation des données, elles suscitent de nouvelles questions et nous donnent le moyen d'y répondre. Avant le "big data" ou les graphiques complexes, le plus grand apport des humanités numériques réside peut-être dans cet ensemble varié d'interfaces entre l'humain et les données (des textes, des graphiques, des recherches, des éditions en ligne, etc.) qui suggèrent au premier de nouvelles façons de lire et de penser. » (SCHUWEY, 2020, p. 208)

Cela a pour conséquence que « la base de données n'est pas seulement un outil pour la recherche. Elle est bien une forme d'écriture de l'histoire, au même titre qu'un ouvrage critique » (ROUSSILON, SCHUWEY, 2020), comme Schwuey défend aussi longuement dans son livre *Interfaces* (SCHUWEY, 2019). Cela s'oppose à l'idée que « la base de données, comme Manovitch a soutenu [dans *Le langage des nouveaux médias* (MANOVITCH, 2010)], est l'ennemie de la narration »<sup>83</sup> (FOLSOM, 2007). S'il est vrai qu'une absence de hiérarchie peut être vue dans la manière dont une base de données organise les informations (GEFEN, 2015), et que ces données peuvent potentiellement être utilisées pour différents objectifs (donc à l'intérieur de plusieurs récits), les données ne sont accessibles aux êtres humains qu'à travers une forme de visualisation des informations (ne serait-ce que par un tableau ou une liste). Ainsi, si la base de données n'est pas une narration à elle seule, elle doit tout de même faire partie d'un récit pour que ses données soient saisissables par des êtres humains. Le fait que les données ne parlent pas en elles-mêmes est vrai dans tout contexte, mais la dialectique entre donnée et récit a une place particulière dans les humanités : « dans les sciences sociales et les humanités, nous sommes mis au défi de promouvoir une expressivité qui convienne aux connectivités que

---

<sup>83</sup> "Database, as Manovitch has argued [dans *Le langage des nouveaux médias* (MANOVITCH, 2010)], is the enemy of narrative."

nous explorons, et qui soit à la fois vigoureuse et rigoureuse » (ROSE et ROBIN, 2019, p. 29). Comme argumente aussi Yves Citton dans *Lire, interpréter, actualiser* (CITTON et CUSSET, 2017), la valeur des études littéraires réside dans le développement des capacités interprétatives humaines, qui sont fondamentales pour vivre en société, mais aussi pour imaginer et construire des sociétés souhaitables. L'interprétation du monde et l'invention de récits qui en décrivent des aspects, apparaît ainsi comme complémentaire avec une approche scientifique basée sur la mesure et le calcul (même lorsque le récit est basé sur la mesure et le calcul). Dans le cas de la Comédie-Italienne, en réunissant des éléments caractérisant les divers points de vue sur le Répertoire (la compagnie, le peuple et le cour royale), la base de données raconte la relation entre production artistique et culturelle, pouvoir politique et société.

À la lumière de toutes ces considérations, les analyses et les visualisations réalisables avec nos données sur le Répertoire de la Comédie-Italienne ne permettent pas à proprement parler de démontrer des phénomènes, mais plutôt de faire des observations sur une quantité de données qui est très difficilement manipulable avec des méthodes non automatisées, afin de suggérer des hypothèses. En reprenant un schéma produit dans la thèse de doctorat de Tanguy (TANGUY, 1997, Figure 61) qui décrit l'organisation d'un système interactif, des analyses et des visualisations comme les nôtres se situent au niveau de la couche sémiotique, qui est une « sorte d'intermédiaire, où se projettent les actions et conceptions humaines, qui doit donc être porteuse de grande souplesse, et d'interaction forte avec l'utilisateur » (TANGUY, 1997, p. 26). Les analyses et visualisations sont en effet le résultat de la représentation des données faite au niveau du modèle (qui correspond à la couche logique), elles suggèrent des phénomènes à l'utilisateur, qui les interprète. L'activation de mécanismes inférentiels, dont la couche logique assure la rigueur, correspond au processus de modélisation, fait à partir des résultats qu'on voulait obtenir.

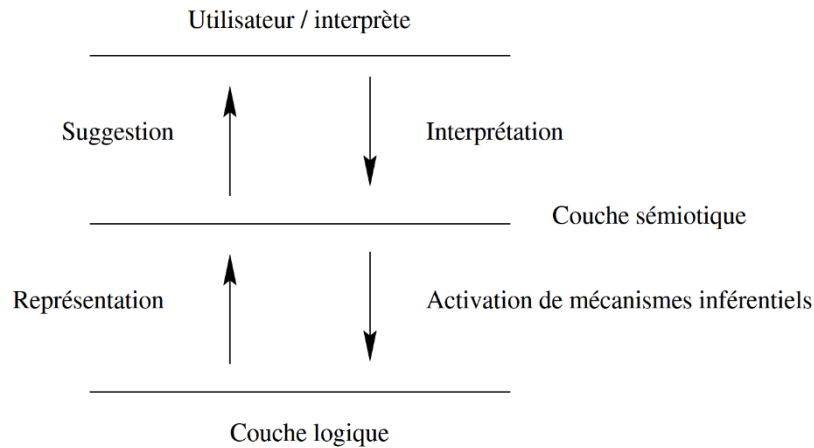


Figure 61 : Schéma décrivant l'organisation d'un système interactif. Source : TANGUY, 1997, p. 26.

Une formalisation logique-mathématique des interactions est proposée par Lecomte et Quatrini (LECOMTE et QUATRINI, 2010), à partir de la Logique Linéaire de Girard et de la Ludique. La formalisation logique-mathématique a historiquement comme objectif d'étudier les démonstrations : quand et pourquoi une certaine suite d'arguments constitue une *preuve* ? Pour aborder cette question, il existe plusieurs théories et plusieurs langages, permettant de modéliser des processus inférentiels et d'en étudier les caractéristiques.

« La Ludique, s'émancipant à priori des notions de preuves et de formules, propose de redéfinir la logique, à partir des éléments plus primitifs sur lesquels se base l'interaction. Ces objets, appelés *desseins* ont été obtenus au terme à la fois d'une déconstruction de l'objet preuve comme support de l'interaction et d'une généralisation de l'objet preuve comme processus » (ibid., p. 38).

Dans le but de formaliser le processus interprétatif dans un projet d'humanités numériques, ce positionnement théorique semble approprié : le processus interprétatif fonctionne sur une interaction entre une machine et un.e chercheur.euse, mais cette interaction ne constitue pas la démonstration elle-même. Ce qui nous intéresse de l'interaction c'est le processus interprétatif qu'elle sous-tend. A partir de cet article ainsi que du schéma de l'interaction de Tanguy (Figure 61), il serait tout-à-fait envisageable de définir un ensemble de desseins qui décrivent les interactions entre machines et êtres humains dans une recherche comme la nôtre. De cette manière, il serait possible d'en étudier le déroulement, avec les méthodes formelles de la logique-mathématique.

Ainsi, depuis un point de vue informatique, la portée épistémologique majeure de la modélisation des données dans les humanités réside dans la possibilité, à travers des processus de formalisation, de rendre l'interprétation aussi un objet d'étude manipulable. En reprenant le célèbre ouvrage de Willard McCarty *Humanities Computing* (MCCARTY, 2013), Mounier écrit que

« ce n'est pas la puissance de calcul de l'ordinateur qui est véritablement intéressante ; c'est plutôt la contrainte imposée par le traitement numérique de l'information au chercheur.euse qui le conduit à devoir formaliser tout à la fois le matériau sur lequel il travaille (ce qui le transforme en donnée) et le raisonnement qu'il mobilise pour traiter ces données. [...] L'activité intellectuelle qui se trouve au centre de cet usage heuristique de l'informatique en sciences humaines, McCarty la situe au niveau de la notion de « modélisation » » (MOUNIER, 2018, p. 60).

Concernant des projets numériques en histoire, Mouza, Lamassé et Rygiel écrivent en effet que

« il est assez vite apparu que les enjeux n'étaient pas seulement d'appropriation d'un vocabulaire technique. Il était également nécessaire d'explicitier les implicites d'un jugement expert, les modes de gestion de l'incertitude, les hypothèses construites pour faire face à celle-ci, bref la logique d'un raisonnement historien, cependant que les informaticiens à leur tour étaient saisis de doutes, menant à la construction de projets de recherche. » (MOUZA et al., 2023)

Ils soutiennent ainsi la nécessité de modéliser le processus interprétatif de recherche de l'historien.ne qui reconstitue la lacune et le manque, car « si l'incertitude imprègne notre compréhension des mondes historiques et du réel, le but des systèmes d'information étant de modéliser le monde réel, alors, les systèmes d'informations doivent composer avec l'incertitude. » (ibid.)

Finalement, notre base de données change à plusieurs endroits la manière d'étudier le Répertoire. Premièrement, cela concerne l'organisation du travail de recherche, entre transdisciplinarité et division des tâches : les résultats apparaissent ainsi comme le fruit d'un travail collectif, composé de nombreuses contributions personnelles, et montrent que, même dans les études littéraires, la production de connaissance a un caractère social. Deuxièmement,

cela concerne la production de preuves scientifiques : la grande quantité de données que nous avons produit, développe les recherches précédentes sur la Comédie-Italienne de Paris, en permettant de mieux contextualiser les phénomènes, mais surtout en suggérant des hypothèses. Autrement dit, ce travail montre que les humanités numériques n'apportent pas nécessairement un avantage en termes de fiabilité, et que les données ne constituent pas une *preuve* en elles-mêmes ; elles changent néanmoins profondément la manière dont les preuves sont cherchées et dont les résultats sont produits. Ces changements semblent aller dans la direction d'une collectivisation dans la production de la connaissance littéraire.

### 4.3 Vers des humanités numériques écologiques

#### 4.3.1 De quelle écologie parle-t-on ?

A notre époque de désastre environnemental et de destruction du vivant, des considérations écologiques sont fondamentales dans les recherches qui concernent le numérique, puisque ce dernier est parfois salué comme une solution à la crise, mais est étudié aussi comme faisant partie du problème. La deuxième option semble néanmoins plus convaincante : les dégâts générés par le numérique, de l'extraction des ressources au démantèlement des terminaux, ne font qu'augmenter, alors que les effets bénéfiques de la numérisation sur les écosystèmes sont encore à démontrer (ROUSSILHE et al., 2023). Cela n'exclue pas les humanités numériques car,

« sans solution durable pour la production d'énergie, et tant que les conditions d'extraction des métaux précieux demeureront indécentes, les humanités numériques feront partie du problème. L'enjeu [écologique] questionne également la contradiction entre le modèle de financement de la recherche par projets, à durée limitée, et les logiques numériques. La pérennité des outils et des publications numériques exige des équipes stables, alors même que la précarisation et l'instabilité du monde académique s'accroissent. » (ROUSSILON, SCHUWEY, 2020)

Deux types de considérations sont à faire en ce qui concerne l'écologie et les projets numériques : il faut des chiffres pour évaluer l'impact avec des paramètres partagés (comme l'empreinte carbone ou la consommation énergétique), mais il faut aussi caractériser les activités en termes théoriques et politiques. Or, aucune activité du projet de doctorat n'a été dédiée au calcul de l'impact matériel du projet, car la tâche est loin d'être évidente (COURANT, 2025) et nous avons préféré confier cela à des recherches ultérieures, plutôt que de donner des

chiffres qui risquent de suggérer des conclusions erronées. De plus, notre projet reste restreint en termes de données : le fichier SQLite ne fait même pas 40 Mo, pour environ 1550 pièces, 41 053 représentations et 27 300 soirées. De plus, ces moins que 40 Mo sont en réalité majoritairement remplis par des données sur l'activité des personnes et sur les versions successives d'une entité, qui ont vocation à être exclues avant l'analyse des données sur le Répertoire, générant un fichier d'environ 20 Mo. En termes de complexité théorique de calcul, le traitement des métadonnées est par ailleurs généralement plus rapide que la recherche dans des textes, car la structure base de données permet de retrouver un élément par son identifiant en un temps unitaire, sans parcourir des suites de données. Néanmoins, au moment du développement de l'interface de consultation, afin d'optimiser la génération des graphiques, l'ingénieure a pris des mesures pour réduire davantage le temps de calcul. Concrètement, les données sont chargées dans l'interface une seule fois (au moment où celle-ci est ouverte) et préparées pour l'affichage, afin d'être traitées en local sans les aller-retours continus entre client et server : cela permet d'élaborer les données plus rapidement tout en dépensant moins d'énergie. L'avantage en termes d'économie d'énergie reste à quantifier.

Travaillant dans un domaine relativement peu énergivore comme la littérature, il a semblé plus intéressant de se concentrer sur des aspects théoriques et sur ce que les humanités peuvent apporter comme contribution dans le débat sur un numérique soutenable, voire écologique :

« bien que les sciences humaines soient souvent considérées comme sobres *a priori*, par opposition à d'autres domaines notoirement plus coûteux pour l'environnement comme l'investigation spatiale, elles sont pourtant partie prenante d'une organisation financière et théorique de la recherche qui doit entrer dans la réflexion pour la transition écologique des sociétés comme des savoirs » (ibid.).

Les humanités peuvent apporter des connaissances pour penser le numérique du futur, complémentaires et fondamentales par rapport aux calculs d'impacts et autres chiffres. La suite de cette partie est donc dédiée à situer, du point de vue théorique, notre projet dans les théories écologiques, et se base sur un article (DESIDERI et al., 2025) qui a paru dernièrement et que j'ai co-écrit avec deux autres doctorant.es de disciplines différentes. Dans cet article, paru dans la revue *Humanités Numériques*, nous proposons une approche basée sur l'écologie sociale de Murray Bookchin (BOOKCHIN 1974, 2019, 2020) et où nous soutenons l'idée que les humanités, et particulièrement les humanités numériques, ont un rôle à jouer dans le

développement d'un numérique écologique. Dans la suite de cette section, seront repris les trois axes proposés pour penser un projet d'humanités numériques écologique (à savoir, les effets indirects du numérique, l'accessibilité sans GAFAM et les conditions de travail), afin d'évaluer à quel point et dans quelle mesure notre projet s'y approche. Une grande partie du texte dans cette section est reprise avec peu de modifications de la partie de cet article dont j'ai assuré la conception et la rédaction.

À partir des théories et pratiques des low-tech (ROUSSILHE et MATEUS, 2023 ; BIHOUIX, 2014) et du minimal computing (RISAM et GIL, 2022 ; WYTHOFF, 2022 ; CHAN et SAYERS, 2022 ; GIL, 2015), l'innovation n'est pas à entendre uniquement comme innovation technique, qui utilise en priorité les technologies les plus à la pointe ; au contraire, il s'agit de centrer le développement de technologies sur l'être humain et sur ses capacités de réappropriation de techniques et de méthodes (RISAM, GIL, 2022 ; CHAN et SAYERS, 2022). Dans un contexte d'études littéraires, cela signifie aussi se poser en continuité avec les pratiques des chercheur.euses, afin d'intégrer les nouveaux aspects à l'écosystème de la connaissance qui existe déjà, sans nécessairement provoquer un bouleversement. Cela repose sur une conception sociale de l'écologie, qui ne voit pas les êtres humains comme mauvais pour l'environnement en tant que tels, mais qui reconnaît un caractère nocif à certaines organisations sociales qui fonctionnent sur l'oppression aussi bien des humains que des non-humains (BOOKCHIN 2020). Dans cette perspective, la compréhension et le soin des écosystèmes et du vivant s'accompagne avec une amélioration des conditions de vie humaines. Ainsi, s'il n'est pas raisonnable de reconnaître au numérique le pouvoir de sortir l'humanité de la crise écologique (IZOARD, 2020), nous envisageons tout de même la possibilité de construire un numérique qui soit véritablement écologique – et non seulement moins impactant en termes d'exploitation des ressources.

#### **4.3.2 Evaluation écologique de notre projet**

Les estimations de l'impact du numérique au niveau mondial sont assez variables (entre 2% et 4% des émissions globales en 2020 selon ARCEP, 2022 et ROUSSILHE et al., 2023). Comme il n'existe pas de mesure partagée pour le calcul des impacts environnementaux, à l'exception de l'empreinte carbone, cette variabilité s'explique par la diversité des phénomènes pris en compte, et les études à ce sujet peuvent être difficilement comparées (ROUSSILHE, BERTHOUD, 2021 ; MARIETTE et al., 2022). Il est néanmoins possible de saisir des tendances : entre autres facteurs, l'augmentation du nombre de services en ligne et de la qualité

des informations qui circulent sont des facteurs susceptibles d'augmenter la dépense énergétique globale (ARCEP, 2022 ; ROUSSILHE et al., 2023). Il semble alors nécessaire de se demander, tout d'abord, si la présence de nos productions sur Internet est pertinente. Selon le public visé et les objectifs du projet, une mise en ligne de toutes les informations pourrait ne pas être toujours souhaitable. L'accès continu aux applications peut aussi être questionné, comme l'a fait le Low Tech Magazine<sup>84</sup> avec son site accessible uniquement quand le panneau solaire qui l'alimente est en mesure de fournir assez d'énergie. Le site est par ailleurs statique, ce qui réduit la consommation d'énergie par rapport à une application dynamique (RIPOLL et BARESCH, 2024).

Pour le projet de base de données pour la Comédie-Italienne de Paris, mettre à disposition les informations pour le public international de chercheur.euses en études théâtrales était l'un des objectifs du projet, auquel il n'était pas possible de renoncer. Cependant, nous avons traité un ensemble cohérent et relativement restreint d'informations, afin de diminuer la complexité de calcul et l'espace nécessaire au stockage. À partir du modèle initial qui permet de décrire une bonne partie de l'activité de la compagnie, une première réduction a eu lieu au moment de l'implémentation du modèle pour la saisie : les entités concernant l'activité des acteur.ices n'ont pas été créées, car leur saisie n'était pas strictement nécessaire à l'étude de l'évolution chronologique de la programmation, et demandait une recherche d'informations en dehors du cadre des répertoires. Une deuxième réduction a été faite au moment de l'exploitation des données pour l'interface de consultation, par l'élimination de toute information liée aux personnes ayant saisi, ce qui répond à des soucis de confidentialité aussi. En effet, ces informations auraient permis de connaître quelles données ont été saisies par qui et à quel moment, sans pouvoir totalement les anonymiser. De plus, nous nous sommes rendues comptes (lors de la transformation du format de SQLite à MySQL effectué pour une expérience qui n'a pas abouti), que ces données constituaient une portion très grande de toutes les données de la base à la fin de la saisie, et que les éliminer était donc tout à fait pertinent afin de réduire la complexité de traitement. Procédant ainsi, à chaque phase du projet, uniquement les informations utiles et pertinentes ont été traitées.

Concernant les serveurs utilisés, l'utilisation de services universitaires a été privilégiée, par rapport à faire appel à des hébergeurs proposant des solutions écologiques, dans un souci de pérennisation des résultats et de coûts d'utilisation des services. En l'occurrence le site sera

---

<sup>84</sup> <https://solar.lowtechmagazine.com/fr/> (consulté le 12 août 2024).

hébergé à Sorbonne Université et maintenu dans le cadre du projet Lab Radar. Par ailleurs, l'avantage de se procurer un serveur et un panneau solaire uniquement pour notre projet ne semble pas évident, si l'on considère les impacts de la construction, du transport et de la fin de vie de ces appareils, ainsi que le coût pour leur fonctionnement et maintenance. Il serait en revanche envisageable de mutualiser ces efforts pour plusieurs projets d'humanités numériques et études théâtrales, par exemple.

Le site avec l'interface de consultation n'est pas non plus statique, car un fonctionnement dynamique est nécessaire afin de faire des requêtes à la base de données. Or, nous n'avons pas renoncé à la forme base de données, qui était un des objets de l'étude et qui représente un format adapté au concept de répertoire. Néanmoins, les données ne sont chargées depuis le serveur qu'une seule fois par utilisation, réduisant la communication client-server, qui est plus énergivore qu'une utilisation en local. Lorsque les données sont chargées dans l'interface, elles sont directement associées aux informations nécessaires pour leur affichage, ce qui réduit davantage le temps global de traitement des informations (les données sont traitées une fois au début, et sont ensuite accessibles en un temps unitaire).

Concernant l'accessibilité, si Internet et le numérique permettent en théorie de communiquer de manière libre et extrêmement rapide depuis tout endroit de la planète, cette universalité théorique a de nombreuses limites matérielles, en particulier la géographie de l'infrastructure d'Internet (CARDON et al., 2023), le phénomène de digital divide (HARGITTAI, 2003 ; VAN DIJK, HACKER, 2003), les politiques de censure et de souveraineté (GLASZE et al., 2023 ; OEVER, 2021a). Ces trois facteurs sont profondément liés : le déploiement de l'infrastructure matérielle est régi par des intérêts économiques, et les applications sont davantage orientées vers un certain public et certains usages (FIORMONTE et DEL RIO RIANDE, 2023). Le numérique tel qu'on le connaît est en fait façonné par des monopoles (OEVER, 2021b ; GALLOWAY, 2023), dont les conséquences font apparaître Internet plus comme un service vendu par un relativement petit nombre d'entreprises à un large ensemble de la population mondiale, que comme un produit de la liberté humaine (OEVER, 2021a). Cela ouvre de nombreuses questions concernant l'accessibilité, dont deux aspects nous semblent particulièrement saillants pour les humanités numériques : l'appropriabilité des technologies, d'un côté, et le sentiment d'intuitivité des services, de l'autre.

Afin de maximiser l'appropriabilité des technologies qu'on développe, tout en restant dans le réseau mondial avec ses contradictions, ses limites et ses contraintes, des choix peuvent

être faits concernant le type de logiciels et de technologies qui dépendent le moins possible des monopoles. L'utilisation de logiciels propriétaires a donc été réduite au maximum, en favorisant autant que possible l'utilisation de logiciels libres. Dans un souci de maximisation de l'indépendance du projet par rapport au réseau numérique, il est possible d'envisager la création d'une clef USB contenant notre base de données et un logiciel pour y accéder, qui tournerait sur la clef même. Une telle clefs USB pourrait être déposée dans les bibliothèques de l'université et consultée comme les autres objets que la bibliothèque met à disposition. Elle pourrait aussi être clonée sur une nouvelle clef. De cette manière, elle peut être utilisée sur n'importe quel ordinateur : il faudrait, pour cela, installer un système d'exploitation directement sur la clef USB avec le programme, pour que ce dernier puisse tourner, non seulement indépendamment d'Internet, mais aussi du système d'exploitation de l'ordinateur et de sa version. Cela a néanmoins des limites évidentes : premièrement, sur chaque clés USB, on n'aurait potentiellement pas la même version des données, si ces dernières sont modifiées par l'utilisateur.ice. Si une nouvelle version de la base de données est publiée, il faudrait aussi mettre à jour toutes les clefs pour qu'elles la prennent en compte. Deuxièmement, la transmission d'un tel objet reste plus complexe que l'accès à un site internet, malgré les considérations sur l'accessibilité qu'on vient de faire.<sup>85</sup>

Le caractère de vente du réseau d'Internet, se répercute aussi sur l'autre aspect mentionné, qui concerne le sentiment d'intuitivité lors de l'utilisation des services. Marcello Vitali-Rosati a développé cette problématique dans trois articles de son blog Sens Public (VITALI-ROSATI, 2021a, 2021b, 2021c). On pourrait résumer ainsi sa position : si l'on définit l'intuition comme « un rapport immédiat avec le réel » (VITALI-ROSATI, 2021c), alors l'existence d'un numérique « intuitif » signifie qu'il n'y a qu'une seule manière de l'appréhender, en l'occurrence celle promue par les GAFAM. Car, « cette immédiateté ne peut qu'être basée sur l'unicité : il y a une et une seule manière pour saisir ce que le monde est en soi [...]. Cela implique qu'il y a un et un seul bon outil et que le travail de conception devrait juste essayer de se rapprocher le plus possible de cet outil idéal » (VITALI-ROSATI, 2021c). Anthony Masure fait des considérations similaires à propos du design, lorsqu'il écrit que

---

<sup>85</sup> Une solution intermédiaire consisterait à pouvoir télécharger, depuis github, le logiciel présent sur internet avec la base de données, afin de consulter les données même lorsque la connexion est absente. L'utilisation d'un tel code reste relativement complexe et cette solution n'est pas adoptable si la base de données n'est pas librement accessible.

« cet intérêt pour l'efficacité de la technique, qui peut virer dans les cas extrêmes dans la fascination aveugle, relègue bien souvent, malheureusement, les questions propres au design et à l'esthétique à un rôle d'embellissement et de décoration d'interfaces « standardisées » qui ne peuvent être que « personnalisées » à la marge, de la même façon que l'on paramètre un « thème » de blog » (MASURE, 2017, p. 30).

Il critique ainsi les grandes plateformes pour le design, qui réduisent « la chaîne graphique [...] à un seul acteur qui concentre tout le pouvoir dans un schéma pyramidal » (MASURE, 2017, p. 70). Il prône ainsi pour une approche créative et diversifiée au design d'interfaces, faisant appel à la complexité des capacités interprétatives humaines.

Cette approche de l'intuition permet d'aborder une problématique fréquente concernant l'accessibilité, qui consiste à dire que les outils free ou open source seraient moins accessibles, car moins intuitifs pour des personnes peu familières avec l'informatique. Si l'intuition correspond à l'unicité promue par les GAFAM, et si l'accessibilité signifie utiliser des services rapidement, sans se soucier de leur fonctionnement, cette problématique reste une véritable impasse (CHAN, SAYERS, 2022). Pour la dépasser, suivant Chan et Sayers, l'accessibilité pourrait être conçue comme un équilibre à trouver entre la simplicité d'utilisation (qui reste incontournable) et le potentiel d'appropriabilité des outils numériques qu'on développe. Cela implique de substituer la centralité de cette – peut-être inexistante (VITALI-ROSATI, 2021b) – « intuition » numérique par l'effort de produire des explications lorsque l'on implémente un nouvel outil ou qu'on adapte un ancien. Ainsi, l'explication du processus de formalisation des données et d'implémentation des applications, devient un point incontournable pour un projet en humanité numériques, afin que les utilisateur.ice.s soient en mesure de s'approprier des outils, mais aussi dans un souci d'interdisciplinarité (MOUZA et al., 2023). Le contexte transdisciplinaire des humanité numériques peut d'ailleurs favoriser une transmission des connaissances entre informatique et lettres, mais ce type de dialogue peut aussi être propice à la conception d'outils et de chaînes de traitement avec des phases d'apprentissage relativement simples – il est par exemple pertinent de construire les outils à partir des habitudes de travail des chercheur.euses en humanités (BOSCHETTI, DEL GROSSO, 2020).

Dans notre projet, la méthodologie est restée le plus proche possible des études littéraires, et le plus possible en continuité avec l'activité de création des répertoires que nous avons numérisé, tout en intégrant la méthode informatique de création de base de données, comme décrit dans la partie sur la méthodologie de recherche (Aspects méthodologiques). Un

carnet Hypothèse avec les informations sur le processus de création de cette base de données est aussi disponible sur internet, présentant une description des différentes étapes, le but étant de pouvoir s'approprier et reproduire le processus dans un autre contexte similaire. Les bénéfices de ce carnet par rapport à son objectif ne sont néanmoins pas estimables à ce jour.

Si l'amélioration globale des conditions de vie des êtres humains est l'un des objectifs de l'écologie sociale, une approche écologique et sociale des humanités numériques prend aussi en compte les bonnes conditions de travail, et valorise tout type de tâche, des plus créatives aux plus répétitives. Théorie et pratique se construisent en effet mutuellement, et même les tâches les plus répétitives peuvent en réalité demander un haut degré d'interprétation et d'expertise (DAGIRAL et PEERBAYE, 2012 ; NYHAN, 2023). Deux aspects nous semblent particulièrement significatifs concernant les conditions de travail et les humanités numériques : premièrement, l'invisibilisation des tâches les plus répétitives et les moins qualifiées, commune à tout le numérique ; deuxièmement, la manière dont le travail est organisé dans les laboratoires de recherche. L'invisibilisation du travail est par ailleurs fortement dépendante des processus d'externalisation et de standardisation des tâches dans la recherche.

Avec le développement du numérique, on assiste à une division du travail davantage standardisée, hiérarchisée et aliénante, ainsi qu'à une externalisation des tâches vers des pays où la main-d'œuvre est moins chère qu'en France. Ces mêmes phénomènes peuvent être observés, à plus petite échelle, dans le monde de la recherche, où une division du travail se met souvent en place, et surtout lorsqu'il s'agit de traiter de longs corpus, ou de grandes bases de données (COURANT, 2025). Ainsi, un travail qui était traditionnellement accompli par un.e chercheur.euse seul.e, est partagé en des tâches plus ou moins longues et complexes, qui peuvent être confiées à un ensemble de personnes, selon leurs compétences. Si, suivant Marx, la division du travail est la condition pour une mise en place efficace de la coopération, il ne s'agit pas d'un outil d'émancipation des travailleur.euse.s en soi, selon la manière dont cette division est organisée. Observer ces considérations sous le prisme de l'écologie sociale (BOOKCHIN, 1974), nous suggère qu'il est possible de diviser le travail de manière à réduire les aspects hiérarchiques et aliénants qui peuvent le caractériser. Des considérations analogues peuvent être faites sur la standardisation : si elle participe à l'aliénation des travailleur.euse.s – en les éloignant davantage du produit de leur travail et de son sens – au même temps, organiser le travail de manière standardisée peut faciliter l'implication d'acteurs différents dans un projet de recherche. Il s'agit donc de comprendre ce qu'on standardise et pourquoi, et de faire en sorte

que tou.te.s les participant.es aient conscience de l'organisation du travail, pour pouvoir agir en conséquence.

Comme déjà développé dans la section sur l'automatisation et l'IA (Observations sur l'automation et sur l'IA), aucun logiciel d'intelligence artificielle n'a été utilisé, mais la base de données existe grâce au travail des stagiaires, qui ont saisi une grande portion des données. Suivant la réglementation universitaire, les stages de moins de deux mois n'ont pas été gratifiés. En termes d'externalisation, nous n'avons pas fait appel à des prestataires sous-payé.es, mais nous avons employé une développeuse informatique, Camilla Cantiani, pour la réalisation de l'interface de consultation. Cela répond à des contraintes évidentes concernant l'expertise nécessaire pour mettre en ligne une plateforme et pour la rendre efficiente. Néanmoins, Camilla Cantiani a réalisé aussi l'interface de saisie, par intérêt personnel et sans avoir aucune compensation pour ce travail. Lors de cette première contribution, elle a aussi rendu la structure de la base de données optimale, ce qui a demandé un dialogue avec moi sur la signification des objets modélisés : cela repose sur une activité interprétative, et basée sur ses compétences spécifiques.

Pour conclure sur cet aspect, qu'est-ce que les humanités numériques, et plus en particulier les humanités numériques littéraires, apportent à la construction d'un numérique écologique ? A la lumière des considérations précédentes, elles constituent en premier lieu une ouverture à la variabilité et à la complexité humaines. Cela peut favoriser un développement technologique fondé sur la créativité, et ancré vertueusement dans une production de la connaissance à échelle humaine et valorisant l'interprétation. Par rapport à un numérique pensé uniquement en termes d'efficacité et de rapidité, cela déplace l'attention vers la manière dont les technologies informatiques interagissent avec les écosystèmes de connaissance humains, et avec des écosystèmes tout-court. Autrement dit, les humanités numériques aident à construire des récits et des outils afin d'imaginer (et construire) un numérique écologique, en dialogue avec la complexité propre aux humains.

« Le danger principal du numérique ne vient pas du numérique lui-même, mais des types de pressions sous lesquelles s'opère son déploiement. Sa merveilleuse puissance computationnelle repose sur la rapidité avec laquelle les flux électriques parviennent à circuler au sein de câbles intercontinentaux. De plus en plus, selon une dynamique qui, bien en-deçà du numérique, remonte à l'industrialisation initiée dès le XVIIIe siècle, le passage par un système nerveux humain apparaît comme une cause de ralentissement coûteux et comme une source de problèmes

difficiles à contrôler. Il s'ensuit une double tendance, elle aussi multiséculaire, qui pousse, d'un côté, à court circuiter toute intervention humaine pour confier les tâches à une automatisation intégrale et, d'un autre côté, à n'intégrer les subjectivités humaines dans les processus de production et de communication qu'au sein de tâches étroitement circonscrites et paramétrées » (CITTON et CUSSET, 2017, p. 459).

Il s'agit alors d'opérer un déplacement d'enjeu : non pas standardiser les humains pour les faire dialoguer avec le numérique, mais déstandardiser le numérique pour contribuer à la créativité humaine (VITALI-ROSATI, 2024).

## 5 Conclusion

### 5.1 Résumé des résultats

Après une introduction autour des concepts de modélisation et d'interprétation, et de leurs liens, cette thèse a présenté tout le déroulement de la recherche qui a permis la création et l'exploitation d'une base de données pour le Répertoire de la Comédie-Italienne de Paris au XVIII<sup>e</sup> siècle. La première partie décrit les résultats matériels obtenus, du point de vue numérique : le modèle des données et l'interface de consultation. Cela a permis de montrer de quelle manière ces résultats dépendent du contexte de recherche. Grâce à la base de données brute et à l'interface de consultation, il est possible de faire des analyses sur les données recueillies, dont la deuxième partie contient des exemples. Cela a permis de montrer comment les données peuvent être utilisées, concrètement. Dans la troisième partie sont finalement discutés les cas d'étude, en abordant la production de connaissance de divers points de vue : l'objet d'étude (la Comédie-Italienne de Paris), l'organisation du travail et les méthodes de démonstration scientifiques.

Pour répondre à la problématique sur les effets d'une base de données dans l'étude du Répertoire, ce travail confirme en premier lieu diverses considérations partagées dans le domaine des études théâtrales et des humanités numériques. Premièrement, les informations ont été agrégées de sources différentes : même si la possibilité, propre au numérique, de mettre sur un même support plusieurs médias n'a pas été entièrement exploitée, notre modèle réunit des informations tirées de documents aux objectifs et aux formes différentes. Cela permet de contribuer à la reconstruction historique des phénomènes spectaculaires liés au Répertoire en question, dans le domaine des études théâtrales, mais aussi plus généralement en histoire (culturelle, sociale et des élites), en littérature ou en économie. Deuxièmement, plus largement, l'agrégation d'informations différentes et l'utilisation de plus grandes quantités de données, permet de mieux contextualiser les phénomènes qu'on étudie : même lorsque les données sont lacunaires, les outils numériques permettent d'être plus complets, voire exhaustifs, par rapport aux sources disponibles. Troisièmement, le processus de formalisation qui rend les informations calculables, par le caractère non ambigu du modèle, a comme effet de visibiliser des biais disciplinaires, du côté informatique comme littéraire.

L'action de *modéliser* afin de rendre *computable* est centrale dans tout le processus. A propos du terme computation, Yves Citton écrit :

« contrairement à *cogitare*, qui désigne la puissance logique de la pensée, *putare* met plutôt l'accent sur sa dimension d'évaluation et de croyance : *com-putare*, c'est rassembler, sélectionner et calculer afin d'asseoir une croyance – autrement dit, *computer*, c'est *collecter au sein d'une estimation*, bien davantage qu'aligner au sein d'un enchaînement propositionnel » (CITTON et CUSSET, 2017, p. 458).

Ces considérations rejoignent largement nos conclusions : pour faire des analyses quantitatives il faut donner une certaine forme aux informations, pour qu'elles soient manipulables à travers une machine, et cette forme dépend des objectifs du projet, de ce que nous souhaitons que les données représentent. Cette forme est loin d'être la seule possible : au contraire, elle est choisie afin de contribuer à un récit. Ces considérations éloignent nos résultats des attentes d'objectivité qui sont souvent attribuées au numérique. S'il est indéniable qu'une plus grande quantité de données permet de préciser des résultats obtenus sur moins d'informations, il n'est pas possible d'affirmer aussi aisément que les visualisations produites *démontrent* nos hypothèses, par le seul fait qu'elles s'appuient sur plus de données. Pour cela, elles doivent prendre part à des méthodes composites de recherche, qui articulent les méthodologies littéraires avec les procédures des sciences formelles et expérimentales, afin de démontrer véritablement les hypothèses émises. Autrement dit, du point de vue épistémologique, nos données, et surtout nos visualisations, ne sont pas des *démonstrations*, mais constituent des *preuves*, des *indices*, qui prennent sens au sein d'une démonstration. Le concept de *suggestion* semble ainsi fortement pertinent pour décrire la manière dont les visualisations de données participent à la construction d'une nouvelle connaissance, en humanités numériques.

« Traduire les affects en chiffres et les chiffres en nombres ne suffit à constituer aucun vrai savoir ni aucun vrai pouvoir. Le moment crucial, subjectivant, incontournable pour donner sens à la computation, est celui de l'actualisation comme *application* ( $x =$  Alfonse van Worden ;  $y =$  Pèdre Velasquez) » (ibid., p. 451), avec l'exemple tiré du *Manuscrit trouvé à Saragosse* de Jan Potocki. Dans notre projet, cette attribution de sens se fait à plusieurs endroits : du moment où une certaine colonne du répertoire de Brenner (BRENNER, 1961) devient un champ dans la base de données, au moment où une certaine couleur dans un histogramme de l'interface de consultation devient un genre ou un auteur. La contribution de la base de données du Répertoire

aux recherches sur la Comédie-Italienne de Paris fonctionne, donc, en articulant deux fonctions : d'un côté l'activité d'attribution de sens, que le chercheur.euse fait sur des objets formels ou graphiques (en l'occurrence un modèle de données et des visualisations sur une interface de consultation de la base de données) ; et de l'autre côté le potentiel suggestif que ces objets exercent sur le chercheur.euse, en tant qu'être humain et individu.

Si « une machine perfectionnée n'est pas une machine isolée de l'homme, fonctionnant sans lui de manière autonome. C'est même tout le contraire » (MASURE, 2017, p. 143), alors les projets d'humanités numériques peuvent représenter des prototypes de « machines perfectionnées », dans la mesure où leur potentiel ne s'exprime qu'en présence d'un humain qui interroge les données avec son regard critique. Autrement dit, l'automatisation et la computation n'ont un *sens* que si un humain les regarde pour les comprendre, au sein de son contexte culturel et social.

## 5.2 Développements futurs

Concernant la suite du projet, de nombreuses tâches restent à accomplir pour compléter et améliorer la base de données et l'interface de consultation. A partir du schéma représentant l'accomplissement des tâches (Figure 2), et par rapport aux objectifs annoncés, l'accomplissement de certaines activités est prioritaire – ces activités ont été listées dans les carrés bleus dans la Figure 62. La plupart peuvent être divisées en plusieurs tâches susceptibles d'être réalisées lors de stages de licence et master, pour des étudiant.es en sciences ou en lettres.<sup>86</sup> Certaines demandent des compétences scientifiques ou techniques plus approfondies. Un des objectifs de ce texte est de transmettre les informations nécessaires afin que l'équipe puisse prendre en main la suite du projet, de manière cohérente et informée.

---

<sup>86</sup> Pour ces activités, des fiches de stage sont fournies en annexe.

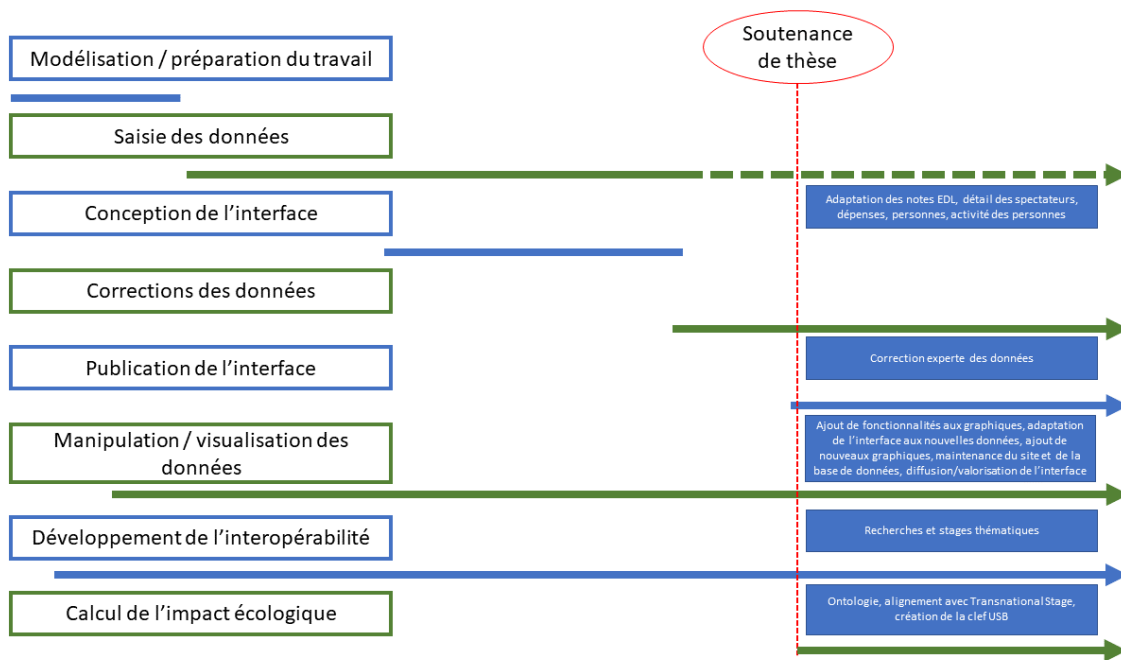


Figure 62 : développement des tâches nécessaires à la réalisation de la base de données, avec listes des activités qui restent à faire.

Pour cela, ces activités ont en premier lieu été intégrées au graphe de connaissance représentant la dimension sociale de la construction de la recherche (Figure 60), en ajoutant l'état de développement de l'activité (terminé en vert, à commencer en rouge, ou partiellement accompli en orange). La partie du graphe dont les nœuds correspondent uniquement aux activités est représentée dans la Figure 63.

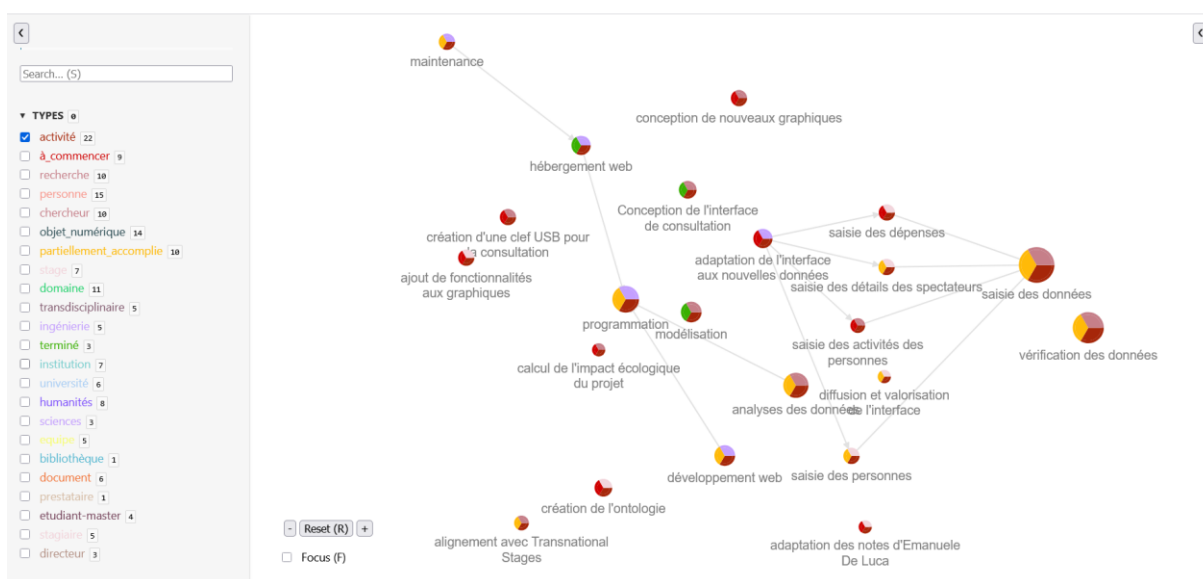


Figure 63 : extrait du graphe représentant notre activité, créé avec le logiciel Cosma.

Dans ce graphe seulement trois activités sont terminées (en vert), alors qu'il y a dix activités partiellement accomplies (en orange) et neuf à commencer (en rouge). La tâche qui doit être accomplie en première est la correction experte des données (nœud « vérification des données » dans le graphe), afin de pouvoir exploiter les graphiques avec une plus grande certitude. L'attribution d'une priorité aux autres tâches va dépendre des disponibilités humaines et économiques de l'équipe, ainsi que des futurs objectifs de recherche : il vaut mieux d'abord compléter les données, en réalisant toute la saisie des données manquantes, ou bien il est préférable de valoriser l'accès aux données actuellement présentes, en intégrant des fonctionnalités importantes aux graphiques, et en expérimentant de nouvelles visualisations de données ? Répondre à cette question constitue le premier pas afin de faire évoluer notre base de données.

Si le graphe de la Figure 63 peut donner une image du travail présenté dans cette thèse comme d'une création incomplète, en réalité il s'agit plutôt d'un travail qui est toujours en cours, toujours susceptible d'être amélioré et augmenté. Comme déjà remarqué en introduction, ce caractère évolutif exploite et valorise une des spécificités des objets numériques, qui est la possibilité d'apporter des modifications à tout moment. Cet aspect est souvent regardé avec une certaine préoccupation : si un texte ou une image sur internet peuvent être changés n'importe quand, alors ils ne peuvent pas être des sources scientifiques fiables – d'où la nécessité de mentionner la date de consultation lorsque l'on cite un objet numérique.

Or, si cela provoque une certaine insécurité dans l'utilisation des sources numériques, ce caractère changeant peut au contraire constituer une garantie de fiabilité scientifique, car il est toujours possible de corriger les erreurs et d'intégrer de nouvelles connaissances. Dans ce sens, du point de vue épistémologique, pour une base de données comme la nôtre, cela peut être vu comme un rapprochement de nos résultats des sciences expérimentales, car les métadonnées représentent ainsi le *meilleur* état de la connaissance sur ce sujet, étant donné les recherches faites et les instruments à disposition. Ainsi, notre production, bien qu'incomplète, ou plutôt grâce à son incomplétude, se présente comme un objet qui ne peut pas être statique, qui a besoin de rester en activité afin d'exprimer au mieux sa valeur.

## Bibliographie

- Alexandre, O. (2023). *La tech : Quand la Silicon Valley refait le monde*. Éditions du Seuil.
- Amarger, F., & Granger, C. (2023). “ The « Comédie-Française » Registers as Linked Open Data : From Heterogeneity to Quantitative RDF Data ”. *Communications in Computer and Information Science*, 1789, 77-88. [https://doi.org/10.1007/978-3-031-39141-5\\_7](https://doi.org/10.1007/978-3-031-39141-5_7)
- Ambriola, V., & Gervasi, V. (1997). An environment for cooperative construction of natural-language requirement bases. *Proceedings 8th Conference on Software Engineering Environments*, 124-130. <https://ieeexplore.ieee.org/abstract/document/591824/>
- Andreoli, J.-M. (1992). Logic Programming with Focusing Proofs in Linear Logic. *Journal of Logic and Computation*, 2(3), 297-347. <https://doi.org/10.1093/logcom/2.3.297>
- ARCEP. (2022). *Pour un numérique soutenable*. [https://www.arcep.fr/fileadmin/cru-1677573101/user\\_upload/12-22-version-francaise.pdf](https://www.arcep.fr/fileadmin/cru-1677573101/user_upload/12-22-version-francaise.pdf)
- Badin, F., & Mallarino, D. (2025). Les données, le numérique et l’Anthropocène. *Humanités numériques*, 11. <https://doi.org/10.4000/1498t>
- Baillot, A., & Ligozat, A.-L. (2025). Introduction. Sobriété numérique. *Humanités numériques*, 11. <https://doi.org/10.4000/1498x>
- Beaucé, P. (2013). *Parodies d’opéra au siècle des Lumières : Évolution d’un genre comique*. Presses universitaires de Rennes. <https://doi.org/10.4000/books.pur.80603>
- Beaucé, P. (2023). Chronique d’un compagnonnage singulier : La parodie dramatique d’opéra à la Comédie-Italienne au XVIIIe siècle. In *L’apothéose d’Arlequin* (p. 161-174). Sorbonne Université Presses.
- Beretta, F. (2023). Données ouvertes liées et recherche historique : Un changement de paradigme. *Humanités numériques*, 7, Article 7. <https://doi.org/10.4000/revuehn.3349>
- Bernard, S. (2023). *Uberusés : Le capitalisme racial de plateforme à Paris, Londres et Montréal*. PUF.

- Besson, R. (2023). La place des sources documentaires dans l'écriture de l'histoire à l'ère du numérique : Le cas de l'Encyclopédie raisonnée des techniques du cinéma. *Les Cahiers de Framespa. e-STORIA*, 42, Article 42. <https://doi.org/10.4000/framespa.13901>
- Bieser, J. C. T., Coroamă, V. C., Bergmark, P., & Stürmer, M. (2024a). The greenhouse gas (GHG) reduction potential of ICT : A critical review of telecommunication companies' GHG enablement assessments. *Journal of Industrial Ecology*, 28(5), 1132-1146. <https://doi.org/10.1111/jiec.13524>
- Bieser, J. C. T., Coroamă, V. C., Bergmark, P., & Stürmer, M. (2024b). The greenhouse gas (GHG) reduction potential of ICT : A critical review of telecommunication companies' GHG enablement assessments. *Journal of Industrial Ecology*, 28(5), 1132-1146. <https://doi.org/10.1111/jiec.13524>
- Biet, C., Harvey, S., & Sanjuan, A. (2020). Postface – Le Programme RCF, de l'archéologie à la futurologie. *Databases, Revenues, & Repertory: The French Stage Online, 1680–1793*. <https://doi.org/10.21428/671d579e.ad4280ba>
- Bihouix, P. (2014). *L'âge des low tech : Vers une civilisation techniquement soutenable*. Edition du Seuil.
- Blom, F. R. E., Nijboer, H., & Van Der Zalm, R. (2020). ONSTAGE, the Online Data System of Theatre in Amsterdam from the Golden Age to Today : Arts and Media. *Research Data Journal for the Humanities and Social Sciences*, 5(2), 27-40. <https://doi.org/10.1163/24523666-00502003>
- Bollen, J. (2016). Data models for theatre research : People, places, and performance. *Theatre Journal*, 68(4), 615-632.
- Bookchin, M. (1974). *Vers une technologie libératrice*. Librairie Parallèles. <https://fr.theanarchistlibrary.org/library/murray-bookchin-vers-une-technologie-liberatrice>

- Bookchin, M. (2019). *Pouvoir de détruire, pouvoir de créer : Vers une écologie sociale et libertaire*. l'Échappée.
- Bookchin, M. (2020). *L'écologie sociale : Penser la liberté au-delà de l'humain*. Éditions Wildproject.
- Boschetti, F., & Del Grosso, A. M. (2020). L'annotation di testi storico-letterari al tempo dei social media. *Italica Wratislaviensia*, 11(1), 65-99. <https://doi.org/10.15804/IW.2020.11.1.03>
- Bowker, G. C., & Leigh Star, S. (2023). *Arranger les choses : Des conséquences de la classification*. Éditions de l'EHESS.
- Bowker, G., Star, S. L., Boullier, H., Cardon, V., Dagiral, É., Kotras, B., Peerbaye, A., Zabban, V., Cardon, V., Dagiral, É., & Peerbaye, A. (2023). *Arranger les choses : Des conséquences de la classification*. Éditions de l'EHESS.
- Brenner, C. (1947). *A bibliographical list of plays in the French language 1700-1789*. Ann Arbor, Mich. : printed by Edwards brothers. <https://data.bnf.fr/temp-work/eb9c3695caa22822aff7554c6d90e173/>
- Brenner, C. (1961). *The Theatre italien, its repertory, 1716-1793*. University of California press. <https://data.bnf.fr/temp-work/592c6f574ab3abc7220afc8474838297/>
- Burnard, L. (2015). La TEI et le XML. In *Qu'est-ce que la Text Encoding Initiative ?* OpenEdition Press. <https://doi.org/10.4000/books.oep.1298>
- Cafiero, C. (2021). *Compendio del Capitale*. Independently published.
- Cardon, D., Parasie, S., & Ricci, D. (2023). *Atlas du numérique*. Presses de Sciences Po.
- Casilli, A. A. (2019). *En attendant les robots : Enquête sur le travail du clic*. Éditions du Seuil.
- Casilli, A. A., Tubaro, P., Le Ludec, C., Coville, M., Besenval, M., Mouhtare, T., & Wahal, E. (2019). *Le Micro-Travail en France. Derrière l'automatisation, de nouvelles précarités au travail?* [PhD Thesis, Projet de recherche DiPLab]. <https://hal.science/hal-02139528/document>

- Ceccobelli, D. (2017). *Facebook al potere : Lo stile della leadership al tempo dei social media*. Maggioli editore.
- Cederna, C. (2023). Les scènes de coquettes, entre tradition et innovation (1708-1721). In *L'apothéose d'Arlequin* (p. 61-74). Sorbonne Université Presses.
- Cederna, C. (à paraître). *La coquette : Femme rebelle et actrice. Métamorphoses de modèles et de genres, entre le théâtre italien et français (XVIIe-XVIIIe siècles)*.
- Cepernich, C. (2017). *Le campagne elettorali al tempo della networked politics*. GLF editori Laterza.
- Chan, T., & Sayers, J. (2022). Minimal Computing from the Labor Perspective. *Digital Humanities Quarterly*, 016(2).
- Ciotti, F. (Éd.). (2023). *Digital humanities : Metodi, strumenti, saperi* (1a edizione). Carocci editore.
- Citton, Y. (2014). *Pour une écologie de l'attention*. Éd. du Seuil.
- Citton, Y. (2015). Humanités numériques. Une médiapolitique des savoirs encore à inventer. *Multitudes*, 59, 171-180.
- Citton, Y., & Cusset, F. (2017). *Lire, interpréter, actualiser* (Nouvelle édition augmentée). Éditions Amsterdam.
- Ciula, A., & Marras, C. (2023). Modelli, metamodelli e modellizzazione nelle Digital Humanities. In F. Ciotti (Éd.), *Digital Humanities. Metodi, strumenti, saperi* (p. 51-65). Carrocci.
- CNRS. (2016). *Plaquette Science-Ouverte*. [https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/Plaquette\\_Science-Ouverte\\_18112019.pdf](https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/Plaquette_Science-Ouverte_18112019.pdf)
- Collège des données de la recherche du Comité pour la science ouverte. (2022). *Partager les données liées aux publications scientifiques. Guide pour les chercheurs* [Guide]. Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.
- Comparini, L., & Fabiano, A. (2019). *Dictionnaire Goldoni*. Classiques Garnier.

- Comsa, M. T. (2020). Bâtir une base de données de théâtres de société. Défis et applications. In *Espaces des théâtres de société. Définitions, enjeux, postérité* (p. 209-218). Presses universitaires de Rennes.
- Cotte, M. (2023). Intégrité et authenticité d'un objet patrimonial numérisé : L'approche déontologique du projet Reseed. *Humanités numériques*, 7, Article 7.  
<https://doi.org/10.4000/revuehn.3476>
- Courant, E. (2025). La sobriété numérique en mode projet : Quelles stratégies ? *Humanités numériques*, 11. <https://doi.org/10.4000/1498q>
- Crawford, K., & Bury, L. (2022). *Contre-atlas de l'intelligence artificielle : Les coûts politiques, sociaux et environnementaux de l'IA*. Éditions Zulma.
- Curien, P.-L. (2005). Introduction to linear logic and ludics, part I. *Advances in polymer science*, 34(5), 513-544.
- Curien, P.-L. (2006). Introduction to linear logic and ludics, part II. *Advances in polymer science*, 35(1), 1-44.
- Dagiral, É., & Peerbaye, A. (2012). Les mains dans les bases de données: Connaître et faire reconnaître le travail invisible. *Revue d'anthropologie des connaissances*, 61(1), 191-216.  
<https://doi.org/10.3917/rac.015.0229>
- De Luca, E. (2011). *Il repertorio della Comédie-Italienne di Parigi (1716-1762) / Le répertoire de la Comédie-Italienne de Paris (1716-1762)*. IRPMF.  
<https://api.nakala.fr/data/11280%2Fd28df67b/d7300be2848911f9e68c934cf26b6c0f55391ab1>
- De Luca, E. (2012). La circulation des acteurs italiens et des genres dramatiques dans la première moitié du XVIIIe siècle. In S. Chaouche, D. Herlin, & S. Serre (Éds.), *L'Opéra de Paris, la Comédie-Française et l'Opéra-Comique* (p. 241-254). Publications de l'École nationale des chartes. <https://doi.org/10.4000/books.enc.909>

- De Luca, E. (2015). « *Un uomo di qualche talento* » : François Antoine Valentin Riccoboni (1707-1772) *vita, attività teatrale, poetica di un attore-autore nell'Europa dei Lumi*. Fabrizio Serra editore.
- Delacourt, M. (2020). Les chrétiens d'Allah, une recherche vivante : Une base de données de 1600 renégats à partir des archives de Lucile et Bartolomé Bennassar. *Les Cahiers de Framespa. e-STORIA, Hors-série 1*, Article Hors-série 1. <https://doi.org/10.4000/framespa.7889>
- Desideri, A. (2024). Formaliser pour mieux partager : Bilan du travail de modélisation pour la BDD du répertoire de la Comédie Italienne de Paris (XVIII siècle). *Humanistica 2024*. <https://hal.science/hal-04572773>
- Desideri, A., Détruit, A., & Vallarano, B. (2025). Écologie sociale et humanités numériques : Intégrer la poésie à la science. *Humanités numériques, 11*. <https://doi.org/10.4000/1498r>
- Douguet, M. (s. d.). Introduction. *Revue d'Historiographie du Théâtre, 9*. Consulté 17 décembre 2024, à l'adresse <https://sht.asso.fr/introduction-8/>
- Douguet, M., & Roulet, T. (s. d.). CESAR et le web des données : Le projet Dramabase. *Revue d'Historiographie du Théâtre, 9*. Consulté 17 décembre 2024, à l'adresse <https://sht.asso.fr/cesar-et-le-web-des-donnees-le-projet-dramabase/>
- Driscoll, M. J., & Pierazzo, E. (2016). *Digital Scholarly Editing : Theories and Practices*. Open Book Publishers. <https://doi.org/10.11647/obp.0095>
- Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly, 5*(1). <https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>
- Drucker, J. (2014). *Graphesis : Visual forms of knowledge production*. Harvard university press.
- Ducloy, J. (s. d.). Humanités assistées par ordinateur, un exemple avec la Chanson de Roland. à paraître. [https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS\\_%282023%29\\_Ducloy\\_%28proposition%29](https://wicri-demo.istex.fr/Wicri/Europe/ChansonRoland/fr/index.php/HIS_%282023%29_Ducloy_%28proposition%29)

- El Hadi, W. M. (2025, avril 25). *Préservation du patrimoine culturel : Un défi pour les humanités numériques*. Humanistica 2025, Dakar.
- Evrard, H. (2020). *La pollution numérique : État des lieux de la prise de conscience de son impact et mesure d'efficacité relative de diverses techniques d'influence pour la résorber*. Louvain School of Management, Université catholique de Louvain. [p://hdl.handle.net/2078.1/thesis:24302](https://hdl.handle.net/2078.1/thesis:24302)
- Ezquerro, M. (2010). De l'hybridation féconde. *Cahiers de Narratologie. Analyse et théorie narratives*, 18, Article 18. <https://doi.org/10.4000/narratologie.6005>
- Fabiano, A. (2018). *La Comédie-Italienne de Paris et Carlo Goldoni : De la commedia dell'arte à l'opéra comique, une dramaturgie de l'hybridation au XVIIIe siècle*. PU Paris-Sorbonne.
- Fabiano, A., & De Luca, E. (Éds.). (2023). *L'apothéose d'Arlequin*. Sorbonne Université Presses. <https://sup.sorbonne-universite.fr/catalogue/civilisations-cultures-litteratures-etrangeres/e-theatrum-mundi/lapothese-darlequin>
- Farge, A. (1997). *Le goût de l'archive*. Éd. du Seuil.
- Ferrone, S., & Decroisette, F. (2024). *La commedia dell'arte : Actrices et acteurs italiens en Europe, XVIe-XVIIIe siècle*. Sorbonne Université presses.
- Fiormonte, D., & del Rio Riande, G. (2023). Dalla periferia dell'impero. Digital Humanities e diversità culturale. In F. Ciotti (Éd.), *Digital Humanities. Metodi, strumenti, saperi* (p. 351-372). Carrocci.
- Flécher, M. (2019a). Des inégalités d'accès aux inégalités de succès : Enquête sur les fondateurs et fondatrices de start-up. *Travail et emploi*, 159(3), 39-68.
- Flécher, M. (2019b). Les start-ups, des entreprises "cool" et pacifiées ? Formes et gestion des tensions dans des entreprises en croissance. *Nouvelle Revue du travail*, 15. <https://doi.org/10.4000/nrt.5930>

- Flécher, M. (2023). Produire l'auto-engagement au travail. Modalités du travail d'organisation dans une start-up en croissance. *La Revue de l'Ires*, 109(1), 3-34.  
<https://doi.org/10.3917/rdli.109.0001>
- Flipo, F. (2025). L'impératif de la sobriété numérique : Enjeux et définitions. *Humanités numériques*, 11. <https://doi.org/10.4000/1498s>
- Foa, J. (2021). *Tous ceux qui tombent : Visages du massacre de la Saint-Barthélemy*. La Découverte.
- Folsom, E. (2007). Database as Genre : The Epic Transformation of Archives. *PMLA*, 122(5), 1571-1579.
- Fuchs, C., & Horak, E. (2008). Africa and the digital divide. *Telematics and Informatics*, 25(2), 99-116. <https://doi.org/10.1016/j.tele.2006.06.004>
- Galleron, I. (2017). Études théâtrales et humanités numériques : Une introduction. *Revue d'Historiographie du Théâtre*, 4. <https://sht.asso.fr/etudes-theatrales-et-humanites-numeriques-une-introduction/>
- Galloway, A. R. (2023). *Incomputabile* (I. Collettivo, Trad.). Meltemi.  
<https://www.meltemieditore.it/catalogo/incomputabile/>
- Garcia-Fernandez, A., & Greslou, E. (s. d.). Données relationnelles hétérogènes en SHS. *Revue d'Historiographie du Théâtre*, 9. Consulté 17 décembre 2024, à l'adresse <https://sht.asso.fr/donnees-relationnelles-heterogenes-en-shs/>
- Gefen, A. (2015). Les enjeux épistémologiques des humanités numériques. *Socio. La nouvelle revue des sciences sociales*, 4, Article 4. <https://doi.org/10.4000/socio.1296>
- Geoffrey, R. (2013). Is Humanities Computing an Academic Discipline ? In *Defining Digital Humanities* (p. 13-34). Ashgate Publishing Company.

- Geslot, J.-C., & Rebolledo-Dhuin, V. (2020). Du livre imprimé au Web sémantique : Le projet du Dictionnaire des éditeurs français du xixe siècle. *Humanités numériques*, 2, Article 2. <https://doi.org/10.4000/revuehn.426>
- Gil, A. (2015). The User, the Learner and the Machines We Make. *Minimal Computing*. <https://github.io/mincomp/thoughts/2015/05/21/user-vs-learner/>
- Giovacchini, J. (s. d.). *De la source à l'image : Y a-t-il une philologie numérique ? | Le goût de l'archive à l'ère numérique*. Consulté 17 août 2025, à l'adresse <https://gout-numerique.net/table-of-contents/collecter-et-mettre-en-donnees/de-la-source-a-limage-y-a-t-il-une-philologie-numerique>
- Glasze, G., Cattaruzza, A., Douzet, F., Dammann, F., Bertran, M.-G., Bômont, C., Braun, M., Danet, D., Desforges, A., Géry, A., Grumbach, S., Hummel, P., Limonier, K., Münßinger, M., Nicolai, F., Pétiñiaud, L., Winkler, J., & Zanin, C. (2023). Contested Spatialities of Digital Sovereignty. *Geopolitics*, 28(2), 919-958. <https://doi.org/10.1080/14650045.2022.2050070>
- Glorieux, F. (2017). Les confidents du théâtre classique. Essai d'enquête quantitative. *Revue d'Historiographie du Théâtre*, 4. <https://sht.asso.fr/les-confidents-du-theatre-classique-essai-denquete-quantitative/>
- Gnedenko, B. V., & Kolmogorov, A. N. (1954). *Limit Distributions For Sums Of Independent Random Variables*. <http://archive.org/details/gnedenko-kolmogorov-limit-distributions-for-sums-of-independent-random-variables>
- Goldoni, C. (2017). *Scenari per la Comédie-Italienne* (A. Fabiano, Éd.). Marsilio.
- Goodwin, J., & Holbo, J. (2011). *Reading Graphs, maps, trees : Responses to Franco Moretti*. Parlor press.
- Guardenti, R. (1990). *Gli italiani a Parigi : La comédie italienne (1660 - 1697); storia, pratica scenica, iconografia*. Bulzoni.

- Guardenti, R. (1995). *Le fiere del teatro : Percorsi del teatro forain del primo settecento; con una scelta di commedie rappresentate alle foires Saint-Germain e Saint-Laurent (1711 - 1715)*. Bulzoni.
- Guardenti, R. (2021). L'achive Dionysos : Une approche méthodologique à l'iconographie théâtrale. *Studia Universitatis Babeș-Bolyai - Dramatica*, 66(2), 29-46.
- Guardenti, R. (2023). *Atlante iconografico : La Commedia dell'arte*. Titivillus.
- Guyot, S., & Ravel, J. S. (2020a). *Databases, Revenues, & Repertory : The French Stage Online, 1680–1793*. <https://cfrp.mitpress.mit.edu/>
- Guyot, S., & Ravel, J. S. (2020b). Introduction – Histoire du théâtre et humanités numériques. *Databases, Revenues, & Repertory: The French Stage Online, 1680–1793*. <https://doi.org/10.21428/671d579e.e139b5e7>
- Hargittai, E. (2003). The digital divide and what to do about it. In *New economy handbook* (p. 821-839). Derek C. Jones.
- Harvey, S., & Sanjuan, A. (2016). *Le projet des registres journaliers de La Comédie-Française*.
- Izoard, C. (2020). *Merci de changer de métier : Lettres aux humains qui robotisent le monde*. Éditions de la Dernière lettre.
- Jarrige, F. (2015). Révolutions industrielles : Histoire d'un mythe. *Revue Projet*, 349(6), 14-21. <https://doi.org/10.3917/pro.349.0014>
- Jarrige, F. (2022). *On arrête parfois le progrès : Histoire et décroissance*. l'Échappée.
- Kesäniemi, J., Koho, M., & Hyvönen, E. (2022). Using Wikibase for Managing Cultural Heritage Linked Open Data Based on CIDOC CRM. In S. Chiusano, T. Cerquitelli, R. Wrembel, K. Nørvåg, B. Catania, G. Vargas-Solar, & E. Zumpano (Éds.), *New Trends in Database and Information Systems* (p. 542-549). Springer International Publishing. [https://doi.org/10.1007/978-3-031-15743-1\\_49](https://doi.org/10.1007/978-3-031-15743-1_49)

- Le Ludec, C. (2024). *Des humains derrière l'intelligence artificielle. La sous-traitance du travail de la donnée entre la France et Madagascar* [PhD Thesis, Institut Polytechnique de Paris].  
<https://hal.science/tel-04605206/>
- Lecomte, A., & Quatrini, M. (2010). Pour une étude du langage via l'interaction. *Mathématiques et Sciences Humaines*, 189, 37-67. <https://doi.org/10.4000/msh.11640>
- Lejeune, E. (2021). *Médiévistes et ordinateurs : Organisations collectives, pratiques des sources et conséquences historiographiques (1966-1990)* [These de doctorat, Université Paris Cité].  
<https://www.theses.fr/2021UNIP7095>
- Lejeune, E. (2023). Médiévistes et ordinateurs. Organisations collectives, pratiques des sources et conséquences historiographiques (1966-1990). *Les Cahiers de Framespa. e-STORIA*, 42, Article 42. <https://doi.org/10.4000/framespa.14441>
- Loty, L. (2005). Pour l'indisciplinarité. *Studies on Voltaire and the Eighteenth Century, The Interdisciplinary Century; Tensions and convergences in 18th-century Art, History and Literature*, edited by Julia Douthwaite and Mary Vidal, 245-259.
- Mancinelli, T., & Pierazzo, E. (2020). *Che cos'è un'edizione scientifica digitale*. Carocci Editore.
- Manovitch, L., & Crevier, R. (2010). *Le langage des nouveaux médias*. les Presses du réel.
- Mariette, J., Blanchard, O., Berné, O., Aumont, O., Carrey, J., Ligozat, A., Lellouch, E., Roche, P.-E., Guennebaud, G., Thanwerdas, J., Bardou, P., Salin, G., Maigne, E., Servan, S., & Ben-Ari, T. (2022). An open-source tool to assess the carbon footprint of research. *Environmental Research: Infrastructure and Sustainability*, 2(3), 035008. <https://doi.org/10.1088/2634-4505/ac84a4>
- Masure, A. (2017). *Design et humanités numériques*. Éditions B42.
- McCarty, W. (2014). *Humanities computing*. Palgrave Macmillan.
- McCarty, W., & Short, H. (2002). *Mapping the Field*. European Association for Digital Humanities, Pisa. <https://eadh.org/publications/mapping-field>

- Miller, D. (2020). Commentaire – Définir le « répertoire ». *Databases, Revenues, & Repertory: The French Stage Online, 1680–1793*. <https://doi.org/10.21428/671d579e.06b137f3>
- Moretti, F. (2013). *Distant reading*. Verso.
- Moretti, F., Dobenesque, É., & Jeanpierre, L. (2008). *Graphes, cartes et arbres : Modèles abstraits pour une autre histoire de la littérature*. Les Prairies ordinaires.
- Moretti, F., & Sobchuk, O. (2022). *Falso movimento : La svolta quantitativa nello studio della letteratura*. Nottetempo.
- Morse, C. (2020). Réseaux performatifs. Les bases de données et l’exploration des théâtres de société. In *Espaces des théâtres de société. Définitions, enjeux, postérité* (p. 219-224). Presses universitaires de Rennes.
- Mounier, P. (2018). *Les humanités numériques : Une histoire critique*. Éditions de la Maison des sciences de l’homme.
- Mouza, C. du, Lamassé, S., & Rygiel, P. (2023). De l’histoire numérique à l’histoire données ? *Les Cahiers de Framespa. e-STORIA*, 42, Article 42. <https://doi.org/10.4000/framespa.14374>
- Nicolet, A. (Éd.). (2022). Enjeux autour de l’impact environnemental des humanités numériques. *Journal of information sciences*, 21(2), 31-41.
- Nyhan, J. (2023). *Hidden and devalued feminized labour in the digital humanities : On the Index Thomisticus project 1965-67*. Routledge.
- Oever, N. ten. (2021a). The metagovernance of internet governance. In *Power and Authority in Internet Governance*. Routledge.
- Oever, N. ten (2021b). “This is not how we imagined it” : Technological affordances, economic drivers, and the Internet architecture imaginary. *New Media & Society*, 23(2), 344-362. <https://doi.org/10.1177/1461444820929320>
- Orlandi, T. (2010). *Informatica testuale : Teoria e prassi* (1. ed). GLF editori Laterza.

- Perret, A. (2023). Documenter la dimension sociale du travail de la connaissance : Une approche hypertextuelle. *13ème Colloque international d'ISKO-France*. <https://hal.science/hal-04271502>
- Peter Wittenburg (Ed. ), Margareta Hellström (Ed. ), Carlo-Maria Zwölf (Ed.), Abroshan, H., Asmi, A., Bernardo, G. D., Couvreur, D., Gaizer, T., Holub, P., Hooft, R., Häggström, I., Kohler, M., Koureas, D., Kuchinke, W., Milanese, L., Padfield, J., Rosato, A., Staiger, C., Uytvanck, D. V., & Weigel, T. (2018). *Persistent identifiers : Consolidated assertions* [Text/pdf]. <https://doi.org/10.15497/RDA00027>
- Pinotti, A., Somaini, A., & Pinotti, A. (2016). *Cultura visuale : Immagini, sguardi, media, dispositivi*. Giulio Einaudi editore s.p.a.
- Ponzetto, V. (2020). *Espaces des théâtres de société : Définitions, enjeux, postérité*. Presses universitaires de Rennes.
- Radi, S. (2014). Chapitre II. Les djinns. In *Surnaturel et société : L'explication magique de la maladie et du malheur à Khénifra, Maroc* (p. 41-60). Centre Jacques-Berque. <https://doi.org/10.4000/books.cjb.482>
- Radio, E. (2020). Minimal Computing and Ontologies. *Minimal Computing*. <https://github.io/mincomp/thoughts/2020/07/21/minimal-ontology/>
- Ramsay, S. (2013). Who's In and Who's Out. In M. Terras, J. Nyhan, & E. Vanhoutte (Éds.), *Defining Digital Humanities* (p. 239-242). Ashgate Publishing Company.
- Rancière, J. (2000). *Le partage du sensible : Esthétique et politique*. Fabrique : Diffusion Les Belles Lettres.
- Ripoll, É. (2025). Sensibiliser les étudiants à la sobriété numérique dans la gestion de projets en humanités numériques. *Humanités numériques, 11*. <https://doi.org/10.4000/1498v>
- Ripoll, É., & Baresch, A. (2024, mai). Repenser le site internet avec le collectif Greening DH. *Humanistica 2024*. <https://hal.science/hal-04588587>

- Risam, R., & Gil, A. (2022). Introduction : The Questions of Minimal Computing. *Digital Humanities Quarterly*, 016(2).
- Robinson, P. (2002). What is a critical digital edition? *Variants: The Journal of the European Society for Textual Scholarship*. <https://doi.org/10.5281/ZENODO.6533168>
- Rose, D. B., Robin, L., & Schaffner, M. (2019). *Vers des humanités écologiques*. Wildproject.
- Rosso, S. (2023, avril 20). *Foxconn revoit les salaires à la baisse dans son usine de Shenzhen*. Siècle Digital. <https://siecledigital.fr/2023/04/20/foxconn-salaires-baisse-usine-shenzhen/>
- Rougemont, M. de. (2001). *La vie théâtrale en France au XVIIIe siècle*. H. Champion.
- Roussilhe, G., & Berthoud, F. (2021). Le numérique face à la crise environnementale. *Alternatives Non-Violentes*, 199(2), 11-13. <https://doi.org/10.3917/anv.199.0011>
- Roussilhe, G., Ligozat, A.-L., & Quinton, S. (2023). A long road ahead : A review of the state of knowledge of the environmental effects of digitization. *Current Opinion in Environmental Sustainability*, 62, 101296. <https://doi.org/10.1016/j.cosust.2023.101296>
- Roussilhe, G., & Mateus, Q. (2023). *Perspectives low-tech (Divergences)*.
- Roussillon, M., & Schuwey, C. (Éds.). (2020). Écrire l'histoire des spectacles avec des bases de données. *Revue d'Historiographie du Théâtre*, 5.
- Rubellin, F. (2020). *Le théâtre du XVIIIe siècle et les humanités numériques*. <https://shs.hal.science/halshs-02931469>
- Rubellin, F., & Raschia, G. (2020). Redécouvrir les théâtres de la Foire et de la Comédie-Italienne avec les bases THEAVILLE et RECITAL. *Revue d'historiographie du théâtre*, 5. <https://shs.hal.science/halshs-03119590>
- Sahle, P. (2016). 2. What is a Scholarly Digital Edition? In M. J. Driscoll & E. Pierazzo (Éds.), *Digital Scholarly Editing : Theories and Practices* (p. 19-39). Open Book Publishers. <https://books.openedition.org/obp/3397>

- Said, E. W., & Galli, S. (2021). *Orientalismo : L'immagine europea dell'Oriente* (16. ed). Feltrinelli.
- Saint-Raymond, L. (2023). Digital Viau : Histoire et aléas d'un projet en humanités numériques. *Les Cahiers de Framespa. e-STORIA*, 42, Article 42. <https://doi.org/10.4000/framespa.14121>
- Sakhnovskaia-Pankeeva, A. (2024). La création et la constitution du site CESAR. *Revue d'Historiographie du Théâtre*, 9. Consulté 17 décembre 2024, à l'adresse <https://sht.asso.fr/la-creation-et-la-constituton-du-site-cesar/>
- Santos, I., Vieira, R., Trojahn, C., Rocha, L., & Cuenca, E. C. (2022). Megalithism Representation in CIDOC-CRM. In S. Chiusano, T. Cerquitelli, R. Wrembel, K. Nørnvåg, B. Catania, G. Vargas-Solar, & E. Zumpano (Éds.), *New Trends in Database and Information Systems* (p. 550-558). Springer International Publishing. [https://doi.org/10.1007/978-3-031-15743-1\\_50](https://doi.org/10.1007/978-3-031-15743-1_50)
- Schino, M. (2002). *Profilo del teatro italiano : Dal XV al XX secolo*. Carocci.
- Schuwey, C. (2019). *Interfaces : L'apport des humanités numériques à la littérature*. Éditions Livreo-Alphil.
- Schuwey, C. (2020). Humanités numériques et théâtres de société. In *Espaces des théâtres de société. Définitions, enjeux, postérité* (p. 205-208). Presses universitaires de Rennes.
- Shalizi, C. (2011). Graphs, Trees, Materialism, Fishing. In *Reading Graphs, maps, trees : Responses to Franco Moretti*. Parlor press.
- Siemens, R. (2016). Communities of practice, the methodological commons, and digital self-determination in the Humanities. *Digital Studies/Le champ numérique*, 0(0). <https://doi.org/10.16995/dscn.31>
- Spanu Fremder, S. (2010). *Le répertoire et la dramaturgie de la Comédie-Italienne de Paris durant la seconde moitié du XVIIIe siècle* [Thèse de doctorat en études italiennes (sous la direction d'Andrea Fabiano)]. Université Paris-Sorbonne.

- Sparacello, G. (2023). Les canevas de magie à la Comédie-Italienne : Métamorphoses et autres mutations magiques de Veronese à Goldoni. In *L'apothéose d'Arlequin* (p. 175-186). Sorbonne Université Presses.
- Tanguy, L. (1997). *Traitement Automatique de la Langue Naturelle et interprétation: Contribution à l'élaboration d'un modèle informatique de la Sémantique Interprétative* [Phdthesis, Université de Rennes 1]. <https://shs.hal.science/tel-01322692>
- Terras, M. M., Nyhan, J., & Vanhoutte, E. (2016). *Defining digital humanities : A reader*. Routledge, Taylor & Francis group.
- Valentin Emiya, Julien Lefèvre, Frédéric Olive, Pierre-Alain Reynier, & Corentin Travers. (2023). Introduire les enjeux environnementaux et sociétaux du numérique en L3 informatique: *Bulletin 1024, 21*, 83-92. <https://doi.org/10.48556/SIF.1024.21.83>
- Vallarano, B., & Desideri, A. (2024a). Il progetto digitale su Elisa Chimenti (LAI-ALEEF) : La problematicità di un profilo mediterraneo tra reti e frontiere. In A. Di Silvestro & D. Spampinato (Éds.), *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD2024* (p. 15-20). Quaderni di Umanistica Digitale. <https://doi.org/10.6092/unibo/amsacta/7927>
- Vallarano, B., & Desideri, A. (2024b). Two Experiments and a New Model for the Digital Processing of Elisa Chimenti's Corpus (Naples 1883—Tangier 1969). *International Journal of Information Science and Technology*, 8(2), Article 2. <https://doi.org/10.57675/IMIST.PRSM/ijist-v8i2.252>
- Van Dijk, J. (2017). Digital divide : Impact of access. *The international encyclopedia of media effects*, 1-11.
- Van Dijk, J., & Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. *The information society*, 19(4), 315-326.

- Vertongen, C. B. F. et M. (2022). Qu'est-ce qu'un projet d'édition critique numérique ? [Billet].  
*DLIS*. <https://doi.org/10.58079/nt40>
- Vitali-Rosati, M. (2021a, mars 18). *Ce qui pourrait être autrement : Éloge du non fonctionnement*. Culture Numérique. Pour Une Philosophie Du Numérique. <http://blog.sens-public.org/marcellovitalirosati/cequipourrait/fonctionnement.html>
- Vitali-Rosati, M. (2021b, mars 25). *Ce qui pourrait être autrement : Intuitif?* Culture Numérique. Pour Une Philosophie Du Numérique. <http://blog.sens-public.org/marcellovitalirosati/cequipourrait/intuitif.html>
- Vitali-Rosati, M. (2021c, mars 28). *Ce qui pourrait être autrement : Réflexions ultérieures sur l'intuition*. Culture Numérique. Pour Une Philosophie Du Numérique. <http://blog.sens-public.org/marcellovitalirosati/cequipourrait/reflexionsintuition.html>
- Vitali-Rosati, M. (2024). *Eloge du bug : Être libre à l'époque du numérique*. Zones.
- Weber, W. (2020). Les répertoires parallèles de l'Opéra et de la Comédie-Française à la fin de l'Ancien Régime. *Databases, Revenues, & Repertory: The French Stage Online, 1680–1793*.  
<https://doi.org/10.21428/671d579e.37bc13ac>
- Wild, N., & Charlton, D. (2005). *Théâtre de l'Opéra-Comique Paris : Répertoire 1762-1972*. Mardaga.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ...
- Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Witz, R., & Porte, G. (2025). Interopérabilité des bases de données scientifiques : Une sobriété de façade. *Humanités numériques*, 11. <https://doi.org/10.4000/1498w>

Wythoff, G. (2022). Ensuring Minimal Computing Serves Maximal Connection. *Digital Humanities Quarterly*, 016(2).

## Annexes

### Plateformes pour les études théâtrales

Nom	Adresse	Description
César (Calendrier Électronique des Spectacles sous l'Ancien Régime et sous la Révolution)	<a href="https://cesar.huma-num.fr/cesar2/">https://cesar.huma-num.fr/cesar2/</a>	Base de données et banque d'images, concernant le théâtre français des XVII <sup>e</sup> et XVIII <sup>e</sup> siècles.
RCF (Registres de la Comédie- Française)	<a href="https://www.cfregisters.org/#/">https://www.cfregisters.org/#/</a>	Base de données des registres de la Comédie-Française. La plateforme permet de consulter et télécharger les données, ainsi que d'analyser les données au moyen d'outils de recherche interactifs.
Archives de l'Opéra- Comique	<a href="https://dezedo.org/dossiers/archives-opera-comique/data">https://dezedo.org/dossiers/archives-opera-comique/data</a>	Répertoire de nombreux documents relatifs à l'Opéra-Comique de Paris entre 1800 et le début du XXI <sup>e</sup> siècle, consultables à partir de la liste des pièces jouées. À chaque pièce sont associés des

		métadonnées, des extraits, des événements et des fichiers numérisés (manuscrit, livret, affiche, enregistrement, procès-verbal de censure etc.).
Corago	<a href="https://corago.unibo.it/">https://corago.unibo.it/</a>	Répertoire et archive du mélodrame italien entre du XVIIe au XXe siècle.
ArpreGo – Achivio del Teatro Pregoldoniano	<a href="https://www.usc.gal/goldoni/">https://www.usc.gal/goldoni/</a>	Ensemble d'archives du théâtre italien précédant l'époque de Carlo Goldoni, dont l'objectif est d'étudier la réforme mise en place par celui-ci. Le projet met à disposition des informations sur une liste de pièces recensées, ainsi que des éditions critiques pour certaines de celles-ci, sans les numériser en une base de données.
La Grange	<a href="https://comedie-francaise.bibli.fr/index.php">https://comedie-francaise.bibli.fr/index.php</a>	Portail documentaire de la Comédie-Française (il ne s'agit donc pas d'une base de données), et inclut un catalogue des collections ainsi que la numérisation de nombreux documents de

		différents types (œuvres d'art, maquettes, photographies...). Il propose aussi des visites d'expositions virtuelles.
Merveilles de la cour	<a href="https://merveilles17.humanum.fr/apropos.html">https://merveilles17.humanum.fr/apropos.html</a>	Bibliothèque numérique des écrits sur les divertissements de cour au XVIIe siècle. Les liens entre les différentes sections concernent toutes les parties de la bibliothèque (œuvres, personnages, événements, techniques, lieux). Les informations pour chaque élément sont complétées par des liens vers des ressources externes (par exemple les documents sources numérisées par la BnF).
Chronopéra	<a href="http://chronopera.free.fr">http://chronopera.free.fr</a>	Répertoire de l'Opéra de Paris entre 1749 et 1989, qui regroupe, sous forme de base de données, des informations synthétiques sur les représentations de l'Opéra dans la période choisie, accessibles selon différents axes de recherche.

RECITAL	<a href="https://recital.univ-nantes.fr/#/">https://recital.univ-nantes.fr/#/</a>	Plateforme de transcription participative des registres de la Comédie-Italienne au XVIIIe siècle.
BIBLIOFI	<a href="https://cethefi.org/bibliofi/">https://cethefi.org/bibliofi/</a>	Bibliothèque numérique du Cethefi, contenant plusieurs ressources : des numérisations avec métadonnées de documents de Fuzeliers issus de collections du fond de la Library of Congress de Washington ; une base d'autographes pour le XVIIe et XVIIIe siècle d'auteurs, copistes et collectionneurs du théâtre de la Foire et de la Comédie-Italienne ; un tableau avec des métadonnées de pièces.
AMAtI (Archivio Multimediale Attori Italiani)	<a href="https://amati.unifi.it/app/#/">https://amati.unifi.it/app/#/</a>	Archive numérique avec les informations sur les acteurs italiens, ayant joué du XVe au XXIe siècle.
Theater in Saint- Domingue 1764-1791	<a href="https://www.theatreinsaintdomingue.org">https://www.theatreinsaintdomingue.org</a>	Base de données du théâtre public de Saint-Domingue, entre 1764 et 1791.

Theatronics	<a href="https://www.theatronics.com/">https://www.theatronics.com/</a>	Base de données des théâtres de Londres au XVIIIe siècle.
Onstage	<a href="https://www.vondel.humanities.uva.nl/onstage/">https://www.vondel.humanities.uva.nl/onstage/</a>	Base de données du théâtre public d'Amsterdam de 1637 à 1772.
Revolutionary Opéra-Comique	<a href="https://obtic.sorbonne-universite.fr/projet/revolutionary-opera-comique/">https://obtic.sorbonne-universite.fr/projet/revolutionary-opera-comique/</a>	Projet à venir de base de données de l'Opéra de Paris du 1750 au 1815.
Royal Danish Theater Database	<a href="https://artex.au.dk/">https://artex.au.dk/</a>	Base de données du Théâtre Royal de Copenhague.
Théâtres de société	<a href="https://theatresdesociete.unil.ch/">https://theatresdesociete.unil.ch/</a>	Base de données des représentations des théâtre de société en Europe entre le XVIIIe et le XXe siècle.
Naissance de la critique dramatique	<a href="https://ncd17.unil.ch/">https://ncd17.unil.ch/</a>	Base de données d'extraits textuels du XVIIe siècle sur le théâtre du point de vue des spectateur.ices.
Théâtre classique	<a href="https://www.theatre-classique.fr/">https://www.theatre-classique.fr/</a>	Plateforme pour le théâtre classique fournissant différents types de consultations des données, à travers les textes ou les métadonnées.
Molières 21	<a href="https://moliere21.cnrs.fr/">https://moliere21.cnrs.fr/</a>	Plateforme de contextualisation

		d'extraits textuels des pièces de Molière.
Mercure galant	<a href="https://obtic.huma-num.fr/obvil-web/corpus/mercure-galant/">https://obtic.huma-num.fr/obvil-web/corpus/mercure-galant/</a>	Editions numériques du Mercure Galant entre 1672 et 1710.
Theaville	<a href="http://www.theaville.org/">http://www.theaville.org/</a>	Base de données des parodies dramatiques d'opéra au XVIIIe siècle, avec extraits musicaux.
PerformArt Database	<a href="https://performart.huma-num.fr/#">https://performart.huma-num.fr/#</a>	Base de données du théâtre aristocratique romain au XVIIe et XVIIIe siècle.
Projet Molière	<a href="https://obtic.huma-num.fr/obvil-web/corpus/moliere/">https://obtic.huma-num.fr/obvil-web/corpus/moliere/</a>	Editions numériques de textes et autres ressources liés à Molière.
Trasnational Stages	<a href="https://transnationalstages.net/">https://transnationalstages.net/</a>	Projet d'interopérabilité de plusieurs bases de données pour le théâtre, dans le monde.

## Liens vers la documentation

Documents à numériser :

- Répertoire d'Emanuele De Luca (version PDF) :  
<https://api.nakala.fr/data/11280%2Fd28df67b/d7300be2848911f9e68c934cf26b6c0f55391ab1>
- Répertoire de Brenner, 1961 (photographies personnelles du dactylographié conservé à la BnF Mitterand, salle X) : <https://drive.proton.me/urls/4HDZV0VDYR#iKsGRAney7lh>

Plateformes :

- Interface de saisie (identifiants nécessaires pour accéder) :  
<http://134.157.57.237:8000/login/?next=/>
- Interface de consultation : [repcomit18.mesu.sorbonne-universite.fr](http://repcomit18.mesu.sorbonne-universite.fr)
- Carnet hypothèses : <https://repcomit18.hypotheses.org/>

Documents du travail informatique :

<https://drive.proton.me/urls/NZ9VBSMDAM#2wyodqJRg9Pb>

Fiches de stages pour les tâches à réaliser :

<https://drive.proton.me/urls/4E06FF9CWC#5zGrSAPsOpFa>

Repositories avec le code :

- FramaGit (expérimentations, résultats des stages, informations sur l'équipe) :  
<https://framagit.org/adadesideri/bdd-comit18>
- GitHub (interface de saisie et de consultation) :  
<https://github.com/kamuttini/ComedieltalienneParis>

## Table des illustrations

Figure 1 : Cycle de la modélisation tirée du chapitre de Ciula et Marras du volume <i>Digital Humanities. Metodi, strumenti, saperi.</i> (CIULA et MARRAS, 2023) .....	13
Figure 2 : développements des tâches pour la création de la base de données par rapport au moment de la publication de cette thèse.....	15
Figure 3 : schéma de l'équipe de recherche. ....	36
Figure 4 : Schéma de la méthode de construction de notre base de données.....	37
Figure 5 : Capture d'écran d'un extrait du répertoire d'Emanuele De Luca en format PDF... ..	50
Figure 6 : Capture d'écran d'un extrait du répertoire d'Emanuele De Luca en format Excel. ....	50
Figure 7 : Capture d'écran d'un extrait du répertoire de Clarence Brenner.....	51
Figure 8 : schéma des sources utilisées pour saisir les données.....	53
Figure 9 : Image du modèle avant la normalisation (modèle créé avec Looping). ....	56
Figure 10 : Image du modèle implémenté (modèle créé avec Looping).....	56
Figure 11 : répartition de la saisie du répertoire de Brenner de 1961. ....	58
Figure 12 : Modèle abstrait complet représenté avec le formalisme de Chen. (Schéma créé avec Looping).....	60
Figure 13 : Modèle abstrait complet représenté comme tables de relation (Schéma créé avec Looping).....	60
Figure 14 : page d'accueil de l'interface de consultation.....	75
Figure 15 : page « consultation » de l'interface de consultation.....	76
Figure 16 : exemple de fiche de consultation d'une pièce dans l'interface de consultation (page « consultation »). ....	76
Figure 17 : détail de la fiche de consultation d'une pièce dans l'interface de consultation (page « consultation »).....	77
Figure 18 : page « visualisation » de l'interface de consultation. ....	78
Figure 19 : premier graphique de la page « visualisation - pièces » de l'interface de consultation. ....	78
Figure 20 : deuxième graphique de la page « visualisation » de l'interface de consultation... ..	79
Figure 21 : troisième graphique de la page « visualisation » de l'interface de consultation. ..	79
Figure 22 : premier graphique de la page « visualisation - auteurs » de l'interface de consultation. ....	80
Figure 23 : deuxième graphique de la page « visualisation - auteurs » de l'interface de consultation. ....	80
Figure 24 : premier graphique de la page « visualisation - premières » de l'interface de consultation. ....	81
Figure 25 : deuxième graphique de la page « visualisation - premières » de l'interface de consultation. ....	81
Figure 26 : troisième graphique de la page « visualisation - premières » de l'interface de consultation. ....	82
Figure 27 : histogramme de toutes les représentations du Répertoire, divisées par genre et par saison théâtrale (page « visualisation – pièces »).....	82
Figure 28 : Extrait du fichier CSV qui met en relation les entités et attributs des bases de données concernés par le projet d'interopérabilité coordonnée par Ravel et Kallenbach.....	85
Figure 29 : chronologie des représentations de <i>La fausse coquette</i> , <i>Panurge à marier</i> et <i>La veuve coquette</i> . ....	94
Figure 30 : détail de la première phase. ....	95
Figure 31 : détail de la deuxième phase. ....	95
Figure 32 : détail de la troisième phase.....	95

Figure 33 : présence et absence des Registres de la compagnie au XVIIIe siècle. ....	96
Figure 34 : représentations de <i>La veuve coquette</i> et de <i>La coquette fixée</i> . ....	96
Figure 35 : Représentations des parodies d’opéra présentes dans l’édition des <i>Parodies du nouveau théâtre italien</i> de 1731. ....	101
Figure 36 : Représentations des parodies d’opéra ajoutées à l’édition des <i>Parodies du nouveau théâtre italien</i> de 1738. ....	101
Figure 37 : Représentations des pièces éditées dans les <i>Parodies du nouveau théâtre italien</i> (1738), non catégorisées comme des « parodies d’opéra » per l’éditeur. ....	102
Figure 38 : Nombre de représentations par an des pièces catégorisées comme des « parodies » dans notre base de données. ....	103
Figure 39 : Intersection schématisant la relation entre les données recueillies par Beaucé et notre base de données. ....	105
Figure 40 : Nombre de représentations de parodies et d’opéra-comiques par an. ....	106
Figure 41 : Nombre de créations de parodies et d’opéra-comiques par an. ....	107
Figure 42 : Nombre de représentations de parodies, d’opéra-comiques et de pièces dont on ne connaît pas, à ce jour, le genre. ....	108
Figure 43 : Nombre de créations de parodies, d’opéra-comiques et de pièces dont on ne connaît pas, à ce jour, le genre. ....	109
Figure 44 : extrait du fichier CSV avec les titres mentionnés dans les pièces. ....	114
Figure 45 : répartition des lieux mentionnés dans les titres des pièces du Répertoire de la Comédie-Italienne de Paris au XVIIIe siècle, par catégorie de lieu. (Graphique créé sur Excel) ....	115
Figure 46 : carte avec les lieux mentionnés dans les titres du Répertoire. ....	115
Figure 47 : carte avec les lieux mentionnés dans les titres du Répertoire (zoom centré sur l’Europe). ....	116
Figure 48 : carte avec les lieux mentionnés dans les titres du Répertoire (zoom centré sur Paris). ....	116
Figure 49 : affichage des lieux mentionnées dans les titres du Répertoire par zones de chaleur (heatmap). ....	117
Figure 50 : affichage des lieux mentionnés dans les titres du Répertoire par clusters. ....	117
Figure 51 : affichage des lieux mentionnés dans les titres du Répertoire par clusters, centré sur l’Europe. ....	118
Figure 52 : affichage des lieux mentionnés dans les titres du Répertoire par clusters, zoom sur Paris. ....	118
Figure 53 : aperçu initial de la carte avec les lieux mentionnés dans les titres. ....	119
Figure 54 : lieux mentionnés dans les titres, zoom centré sur la Méditerranée. ....	120
Figure 55 : Distribution du nombre de représentations de toutes les pièces de la base de données. (Scatter plot généré avec seaborn). ....	125
Figure 56 : Représentations des pièces avec métamorphoses externes mentionnées dans l’article de Sparacello (SPARACELLO, 2023). ....	127
Figure 57 : Représentations de toutes les pièces avec métamorphoses mentionnées dans l’article de Sparacello (SPARACELLO, 2023) ....	128
Figure 58 : Fonctionnement sociétal du théâtre et contributions de la base de données. ....	134
Figure 59 : contraintes appliquées dans le logiciel Cosma. ....	141
Figure 60 : graphe représentant notre activité, créé avec le logiciel Cosma. ....	141
Figure 61 : Schéma décrivant l’organisation d’un système interactif. Source : TANGUY, 1997, p. 26. ....	156
Figure 62 : développement des tâches nécessaires à la réalisation de la base de données, avec listes des activités qui restent à faire. ....	171
Figure 63 : extrait du graphe représentant notre activité, créé avec le logiciel Cosma. ....	171

## **Comédie-Italienne de Paris (1716 – 1779) : étude du répertoire au moyen de la création d'une base de données**

### Résumé :

La thèse présente la création d'une base de données relationnelle pour représenter et étudier le répertoire de la Comédie-Italienne de Paris au XVIII<sup>e</sup> siècle (1716-1779). Observant ce travail du point de vue des humanités numériques, une double réflexion est menée : d'un côté on analyse les conséquences de l'utilisation d'une technologie comme la base de données au sein des études théâtrales, et en particulier dans l'écosystème des recherches sur la Comédie-Italienne de Paris. De l'autre côté on étudie la création-même des données, afin d'évaluer plus largement la valeur épistémologique des résultats obtenus. La première partie présente concrètement le travail effectué : des objectifs de recherche, aux productions informatiques (en l'occurrence il s'agit du modèle abstrait des données et de l'interface de consultation du projet), à l'interopérabilité. La deuxième partie montre de quelle manière la base de données et la plateforme de consultation peuvent être utilisées, avec quatre cas d'études, ainsi qu'une observation de l'utilisation de l'interface par deux chercheurs en humanités. La troisième partie est dédiée à la discussion des résultats de la recherche, avec des approfondissements sur les aspects méthodologiques et épistémologiques, ainsi que sur l'écologie en contexte numérique. La thèse soutient globalement que, en dépit du caractère autonome qui est souvent attendu des productions numériques, celles-ci se fondent en réalité sur une activité de sélection, où l'humain attribue un sens à des symboles afin de les interpréter. Ce travail défend ainsi l'idée que ce type d'interaction entre l'être humain et l'ordinateur est d'un intérêt fondamental pour nos sociétés actuelles : les humanités numériques apparaissent ainsi comme un champ de recherche qui peut favoriser des productions informatiques socialement et écologiquement vertueuses.

Mots clés : [Comédie-Italienne de Paris, XVIII<sup>e</sup> siècle, répertoire, humanités numériques, base de données relationnelle, écologie, épistémologie, automation, transdisciplinarité, technologie et société, études théâtrales, études italiennes]

## **Comédie-Italienne of Paris (1716 – 1779) : studying the repertoire through the creation of a database**

### Abstract:

The PhD thesis presents the creation of a relational database to represent and study the repertoire of the Comédie-Italienne de Paris in the 18th century (1716-1779). Looking at this work from a digital humanities perspective, a two-pronged approach is taken: on the one hand, we analyse the consequences of using technology such as databases in theatre studies, particularly in the ecosystem of research on the Comédie-Italienne de Paris. On the other hand, the creation of the data itself is studied in order to evaluate more broadly the epistemological value of the results obtained. The first part presents the work carried out in concrete terms: from research objectives, to IT productions (in this case, the data model and the project's consultation interface), to interoperability. The second part shows how the database and consultation platform can be used, with four case studies and an observation of how two researchers in the humanities approach the interface. The third part is devoted to discussing the research results, with in-depth analysis of methodological and epistemological aspects, as well as ecology in a digital context. The thesis broadly argues that, despite the autonomous nature often expected of digital productions, they are in fact based on a process of selection, whereby humans attribute meaning to symbols in order to interpret them. This work thus defends the idea that this kind of interaction between humans and computers is of fundamental interest to our current societies: digital humanities thus appear as a field of research that can promote socially and ecologically virtuous computer productions.

Keywords: [Comédie-Italienne de Paris, XVIII century, repertoire, digital humanities, relational database, ecology, epistemology, automation, transdisciplinary, technologies and society, theatre studies, Italian studies]