JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

[ COLLOQUE - L'IA pour toutes et tous, au service de la Santé et de l'Industrie : enjeux et défis ]

# IA et histoire du théâtre : un futur pour notre passé

Françoise Rubellin

Professeur - LAMO (EA 4276)

*francoise.rubellin@univ-nantes.fr*

JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

*[ IA et histoire du théâtre]*
**Françoise Rubellin (LAMO)**

- Présentation
  - Professeur de littérature française du XVIII$^{e}$ siècle.
  - Laboratoire LAMO (Littératures antiques et modernes)
- Travaux
  - Théâtre et musique XVIII$^{e}$ siècle
  - Hiérarchies culturelles
  - Humanités numériques
  - Projet ANR CIRESFI (*Contrainte et Intégration. Pour une réévaluation des spectacles forains et italiens* - 2014-2019)
- Interdisciplinarité
  - Collaborations avec le LS2N, avec des chercheurs en architecture, histoire de l'art, musicologie, art du spectacle…
  - Projet en cours (VESPACE) : reconstruction de théâtres du XVIII$^{e}$ s. en réalité virtuelle

UNIVERSITÉ DE NANTES

JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

*[ IA et histoire du théâtre]*

Françoise RUBELLIN (LAMO)

# Le projet RECITAL

# Registres de la Comédie Italienne

http://recital.univ-nantes.fr/

cethefi
CENTRE D'ÉTUDES DES THÉÂTRES DE LA FOIRE ET DE LA COMÉDIE-ITALIENNE

l'AmΩ
L'Antique, le Moderne

LS2N LABORATOIRE DES SCIENCES DU NUMÉRIQUE DE NANTES

MSH ANGE-GUÉPIN

Agence Nationale de la Recherche ANR

UNIVERSITÉ DE NANTES

JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

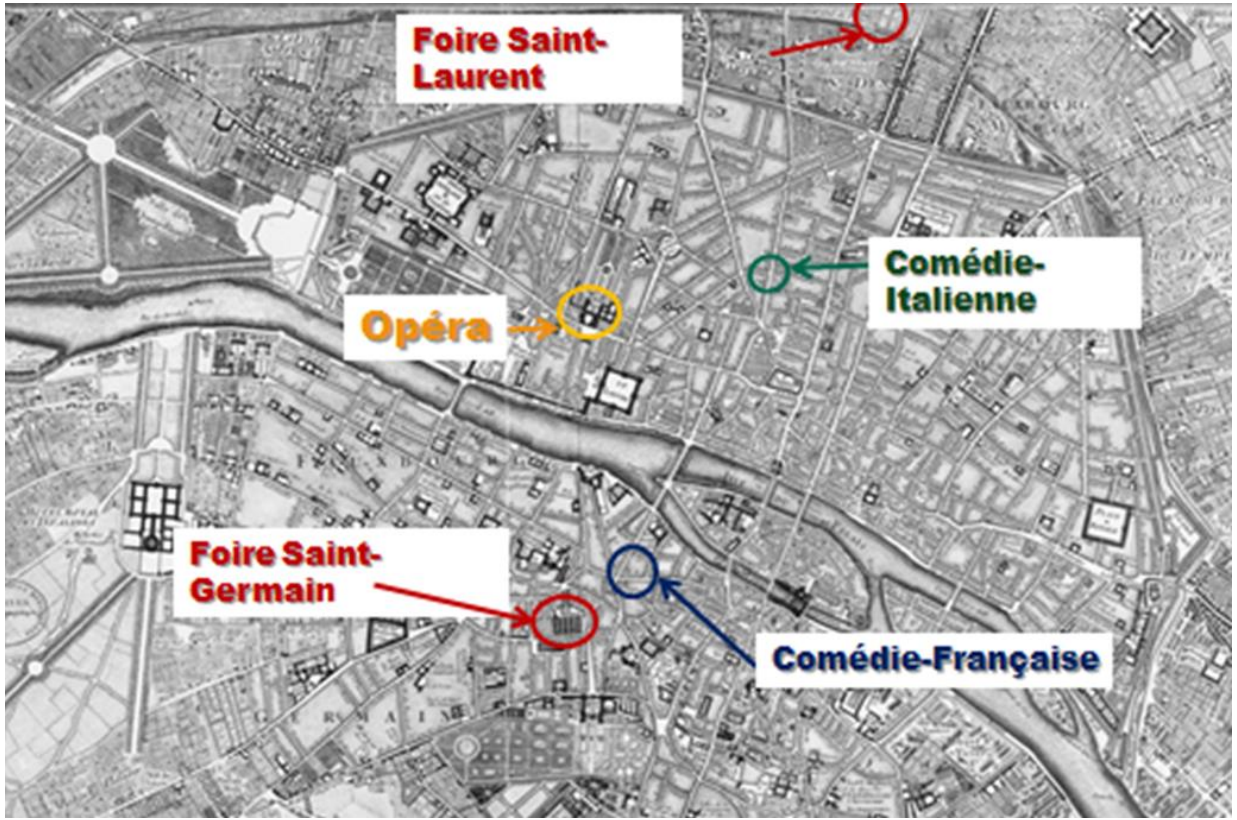
*[ IA et histoire du théâtre]*

Françoise RUBELLIN (LAMO)

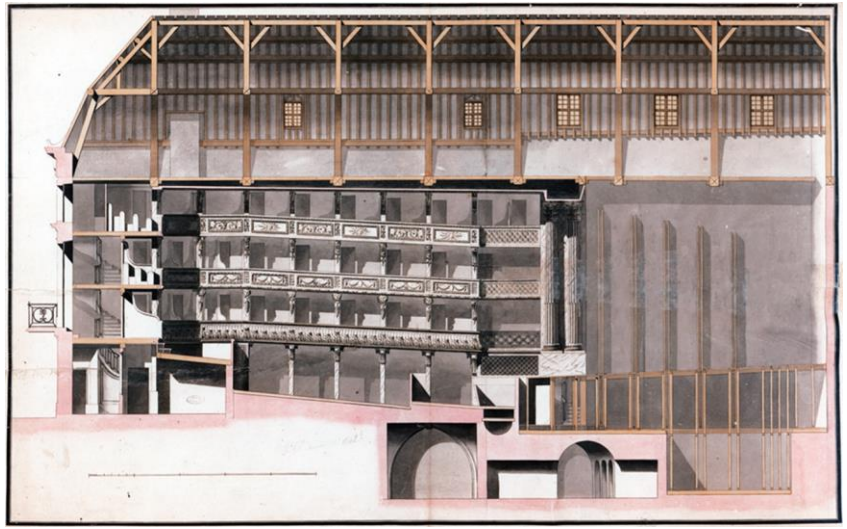
# La Comédie-Italienne

À Paris au XVIII$^{e}$ siècle

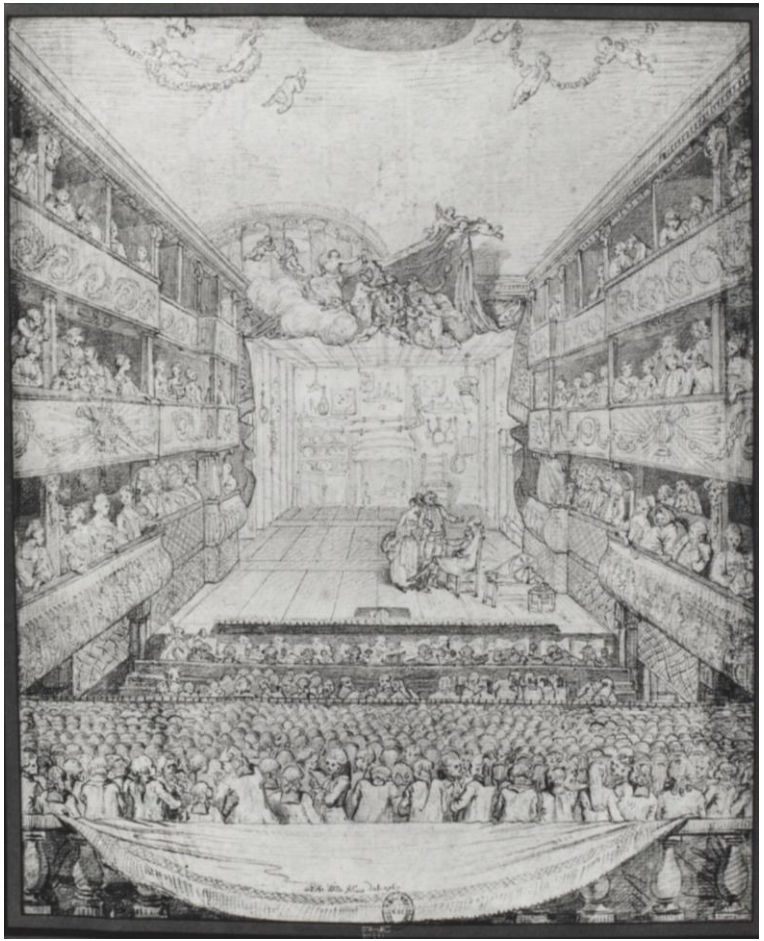
Claude Gillot, *Scène de ballet*, Alger, Musée national des Beaux-Arts



Plus de 1000 places
dans l'Hôtel de Bourgogne, où se tient la Comédie-Italienne



Louis-Alexandre Girault (1760), Archives nationales



Pierre A. Wille le jeune (1767), BnF, Gallica

JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

# Que sont les registres ?

64 registres

de 300 à 600 pages

soit 26 000 pages environ

Bibliothèque-musée de l'Opéra (© F. Rubellin)

JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

# Caractéristiques

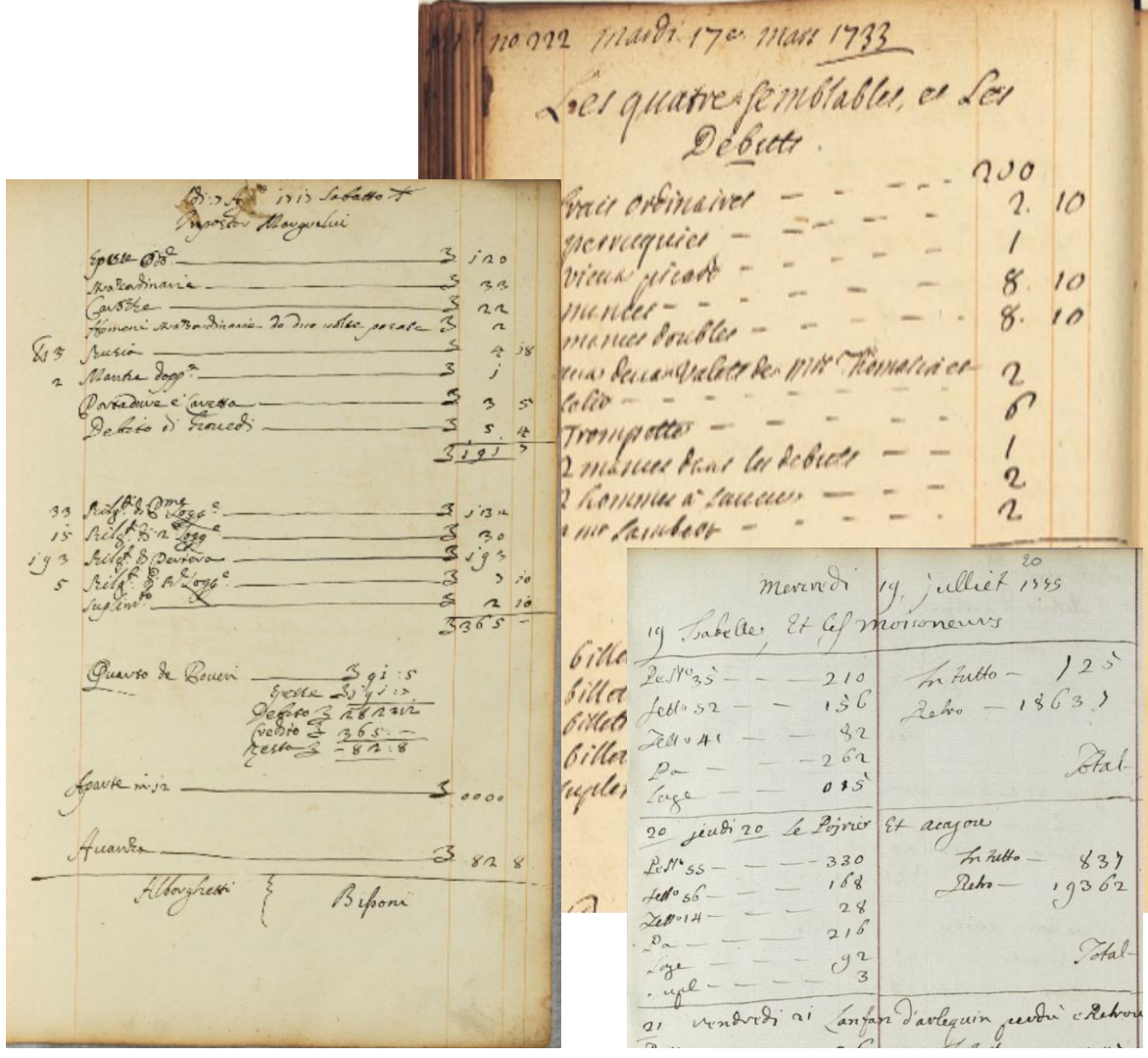
Écritures manuscrite

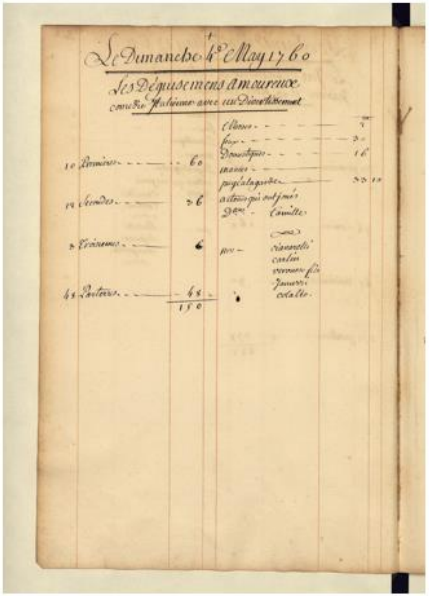
2 langues : italien (dialecte) / français

Système monétaire : Livre/sou/denier

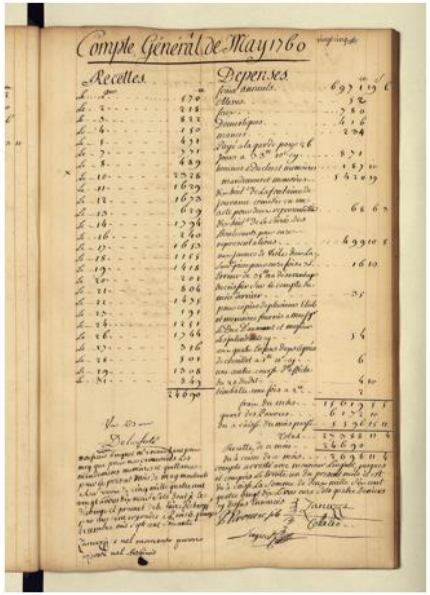
Plus de 7 rédacteurs différents

Règles comptables qui évoluent

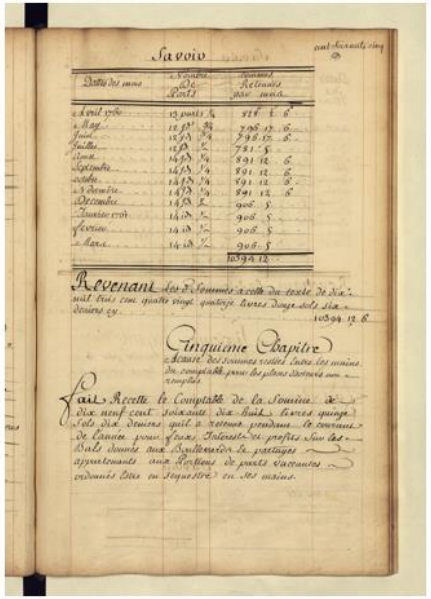




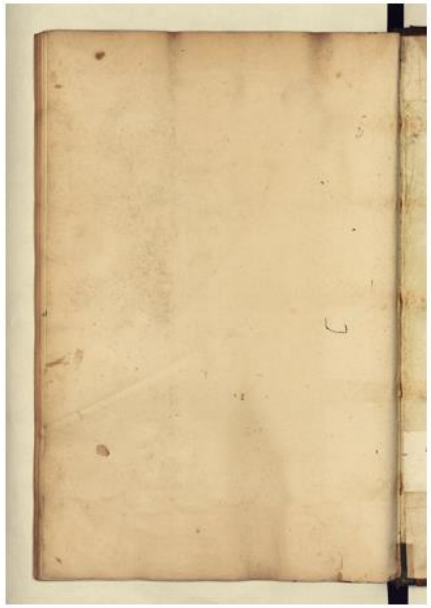
Comptes quotidiens
77%



Comptes mensuels
10%



Récapitulatif final
3%



Pages vierges
10%

Registre TH-OC-42 (1760-1761) - 373 pages

Françoise RUBELLIN (LAMO)

# Pour une histoire matérielle des spectacles

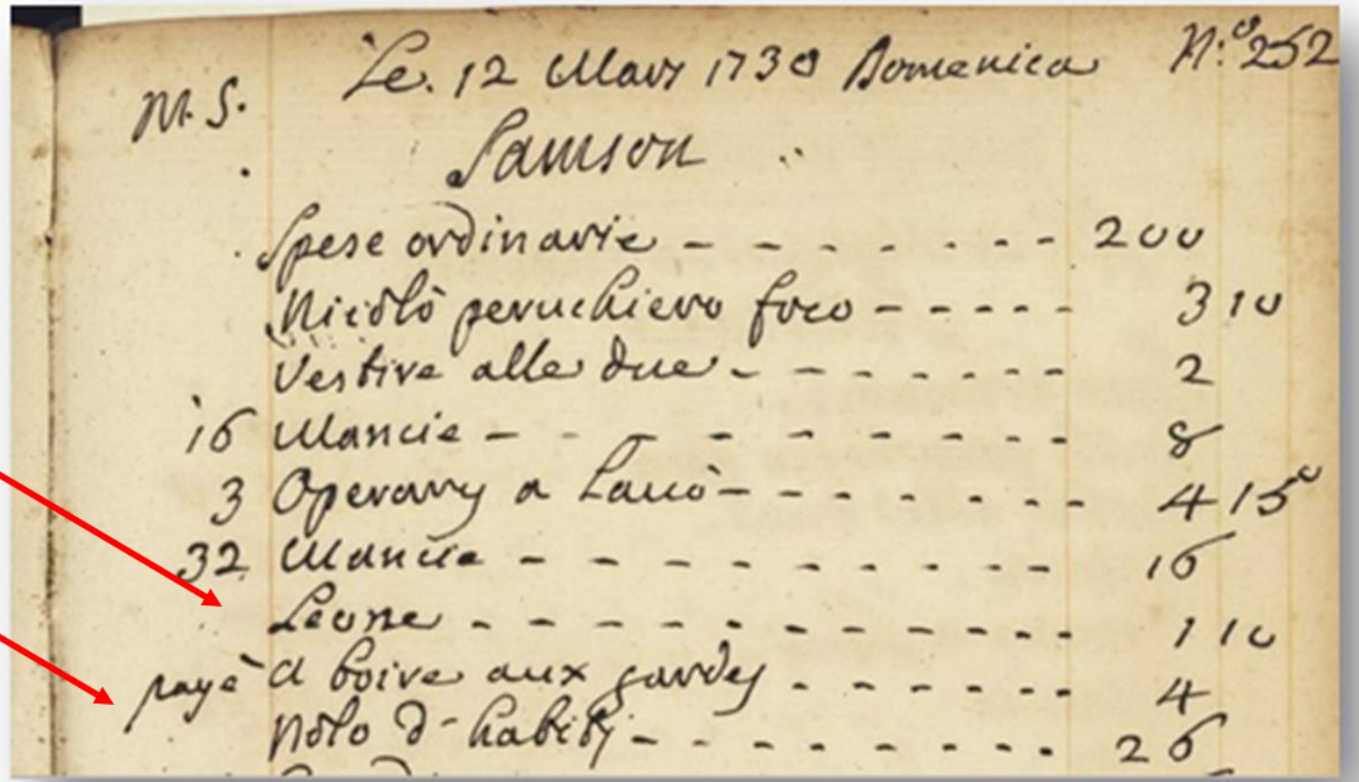
- Sur scène
  - Acteurs
  - Danseurs
  - Musiciens
  - Acrobates
- Dépenses matérielles
  - Accessoires
  - Décors
  - Lumières
- Petits métiers
  - Habilleuses
  - Porteurs
  - Domestiques
  - Concierges
  - Copistes
  - Limonadier
  - Cordonnier

etc.

# Des surprises

Un lion

Payé à boire aux gardes



# Un siècle de données (1717-1794) à numériser pour….

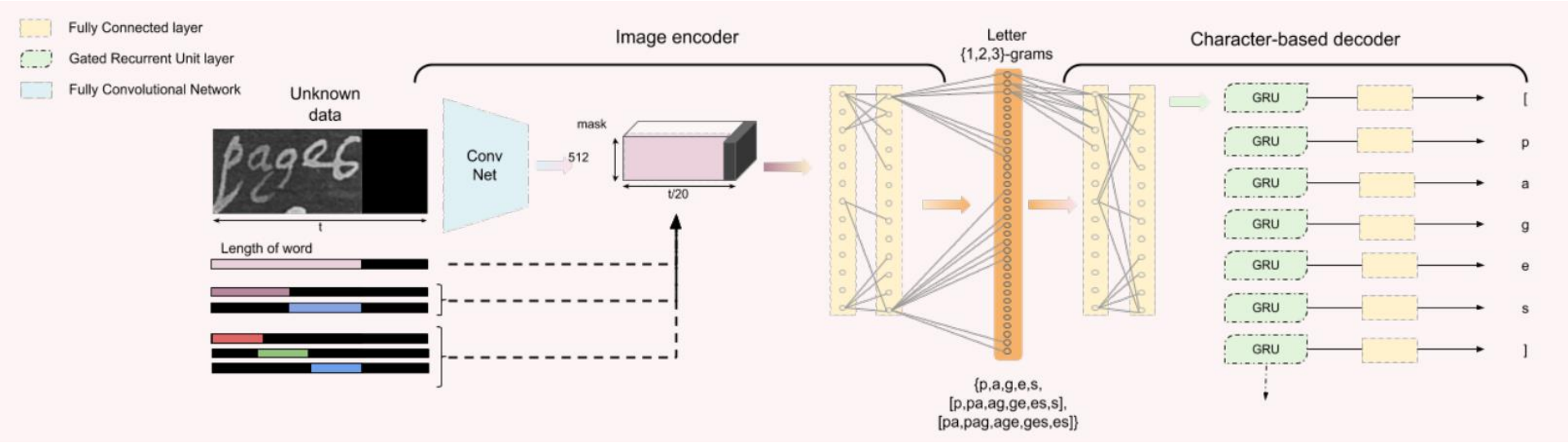
- systématiser les analyses
- faciliter les recherches ultérieures dans le corpus
- mutualiser les efforts de transcription
- pérenniser les données
- partager/réutiliser les données (FAIR)

# L'IA peut-elle nous aider ?

- Collaboration avec l'équipe IVC (Christian Viard-Gaudin et Harold Mouchère)
- Thèse d'Adeline Granet :

  Extraction d'information dans des documents manuscrits anciens

sous la direction d'Emmanuel Morin, co-encadrants Harold Mouchère et Solen Quiniou. Soutenue en décembre 2018

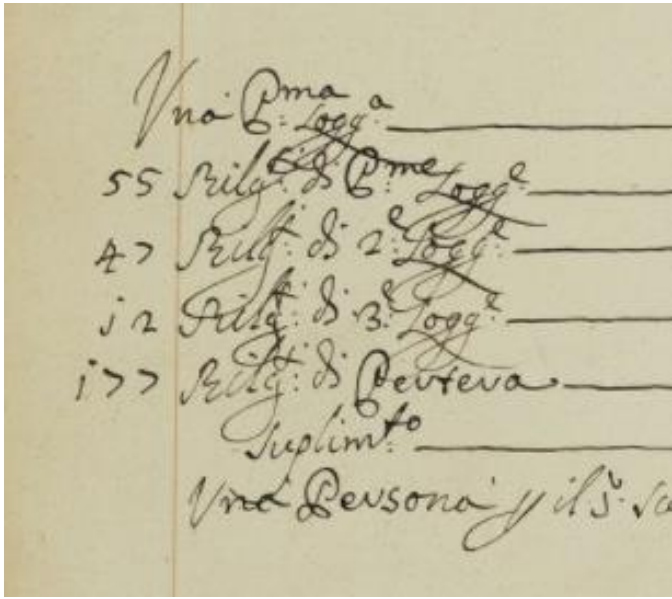



Adeline Granet ICFHR 2018 : “Encoder-Decoder Strategy for Transferable Handwriting Recognition”

- Utilisation des techniques d'IA prometteuse

- Mais nombreuses difficultés techniques :

  - données irrégulières
  - lignes et glyphes mal isolés
  - manque de données d'exemple pour l'apprentissage



Extrait d'une page de registre en italien, 1718.
Recette de la vente de billets (loges, parterre etc.)

# Autre projet : READ

- Projet READ (Recognition and Enrichment of Archival Documents)
- Projet européen H2020 (2016-2019)
- Première plateforme de reconnaissance de texte manuscrit, baptisée Transkribus (basée à Innsbruck, Autriche)
- Mais : demande 100 pages… des difficultés pour zoning, étiquetage, transcription, classification…

# Autre solution : le crowdsourcing

Plateforme développée par Guillaume Raschia (LS2N)
en collaboration avec Benjamin Hervy, Amélie Renard, Olivier Aubert.

- plus laborieux
- manuel
- permettant également l'appropriation du corpus
- adapté aux contenus irréguliers
- enjeux de validation / confiance dans l'information

[ IA et histoire du théâtre]
Françoise RUBELLIN (LAMO)

# La plateforme RECITAL

- Projet logiciel ScribeAPI issue du projet Zooniverse
- Annotation collaborative d'images
- Adapté à des enchaînements de tâches

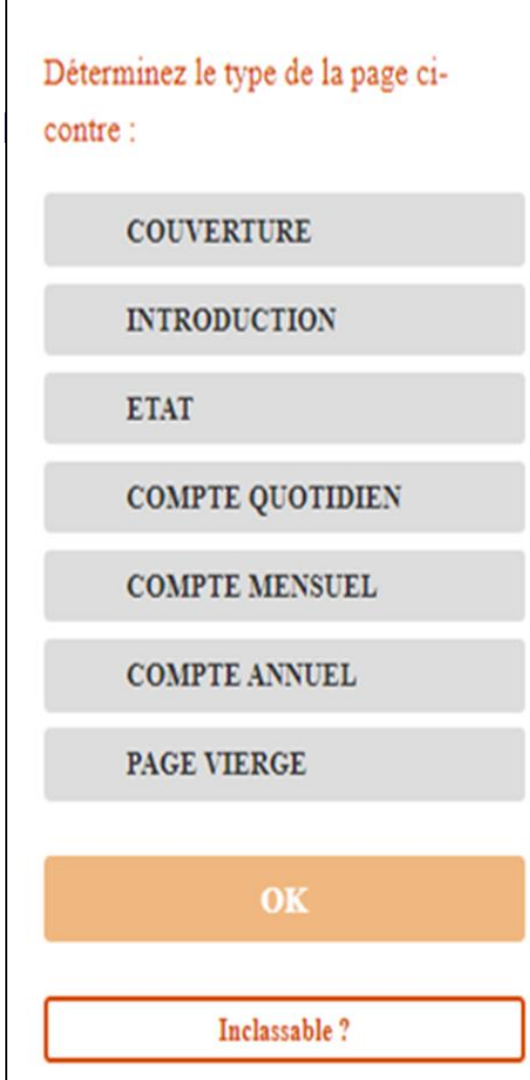
http://recital.univ-nantes.fr/

Trois opérations :

- Marquer (=localiser+catégoriser)
- Transcrire
- Vérifier

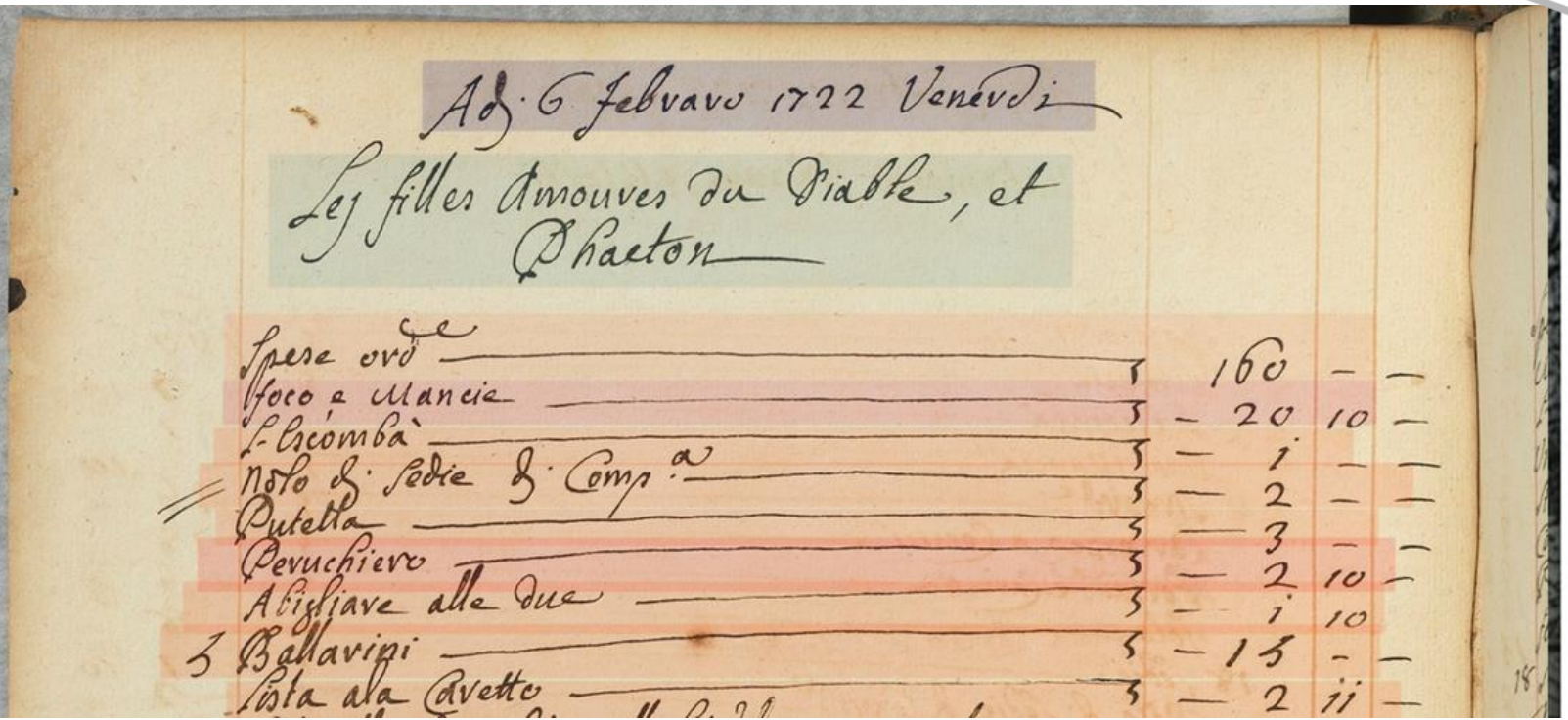
# Organisation du crowdsourcing :

Choisir parmi 7 types de page



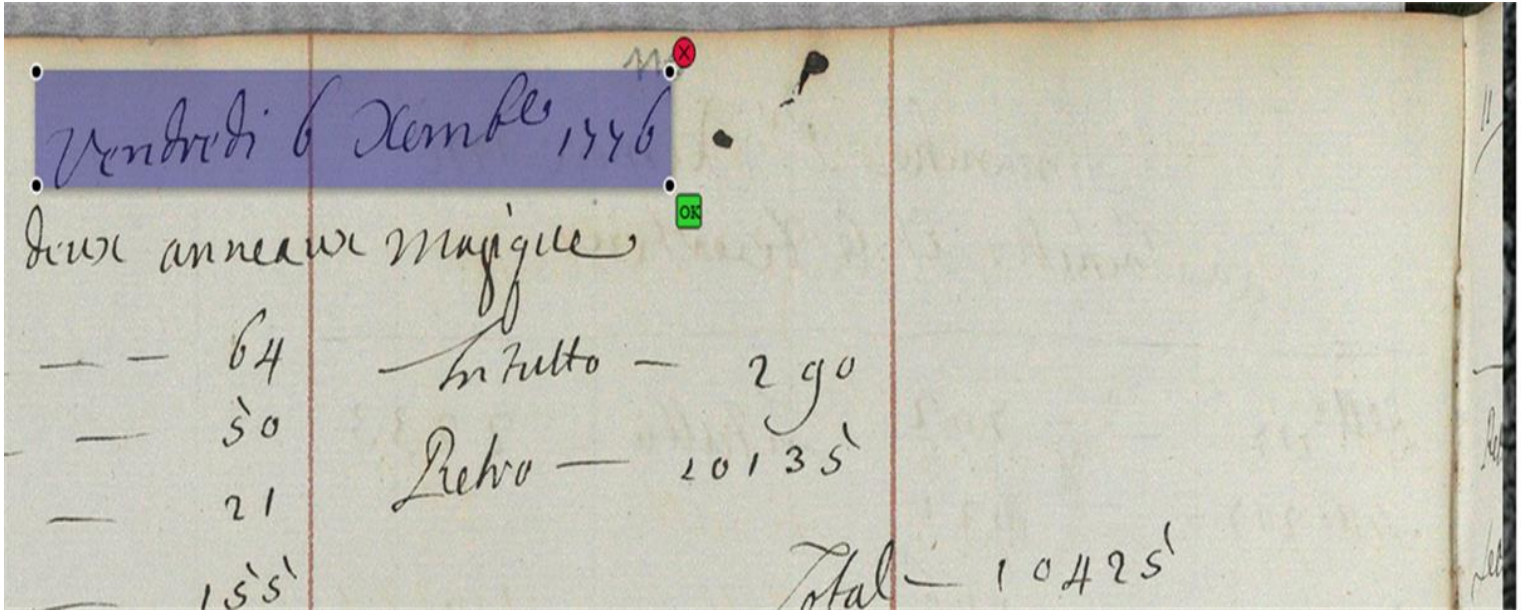
[ IA et histoire du théâtre]
Françoise RUBELLIN (LAMO)

# Organisation du crowdsourcing : 133 catégories d'information

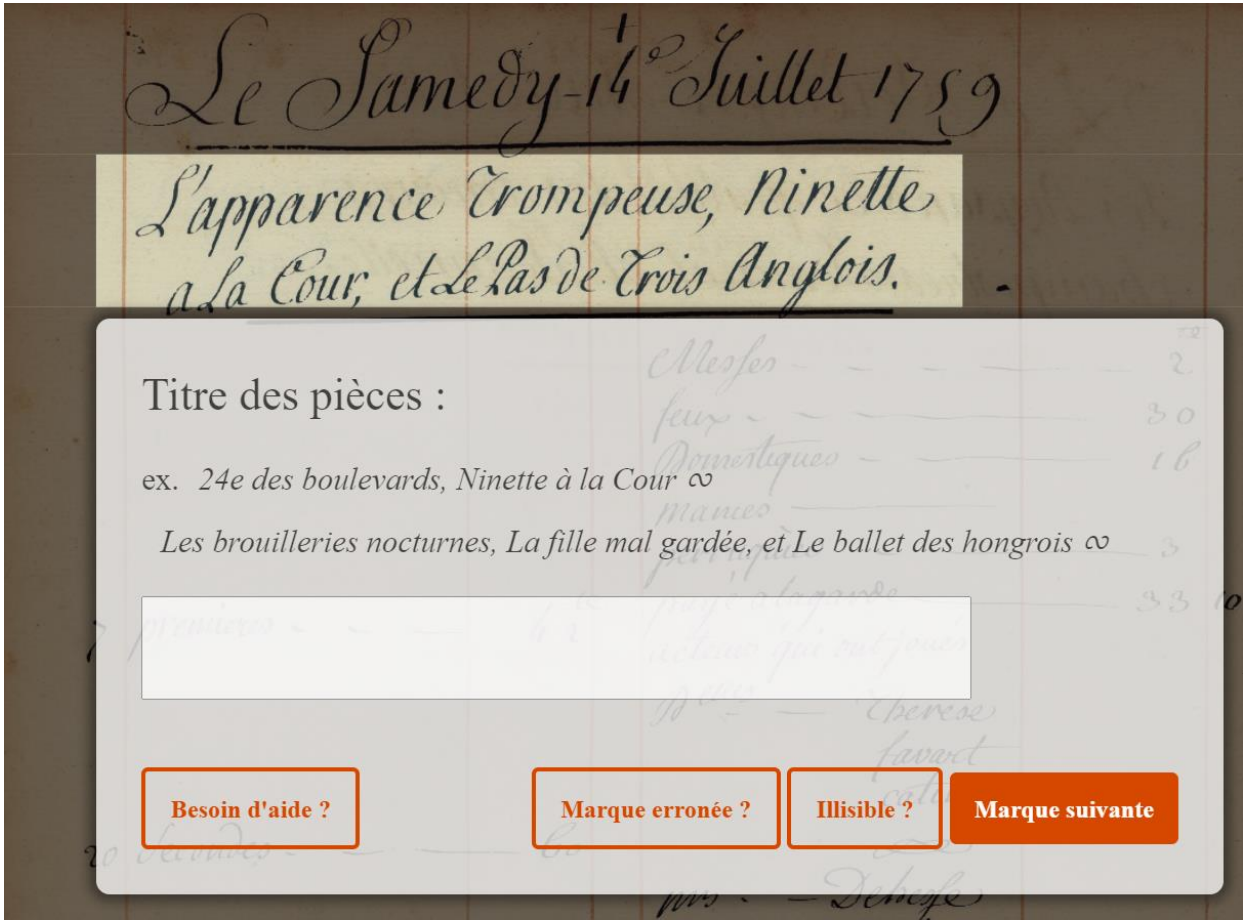


[ IA et histoire du théâtre]
Françoise RUBELLIN (LAMO)

# Marquage : exemple pour la date



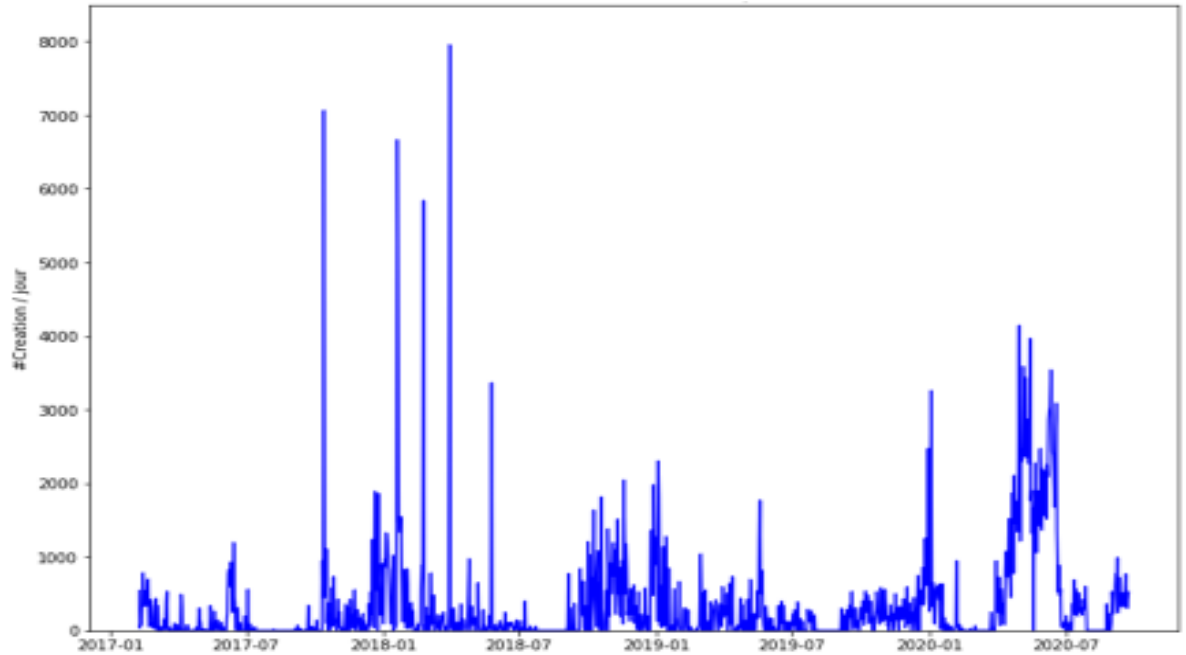
# Transcription



JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

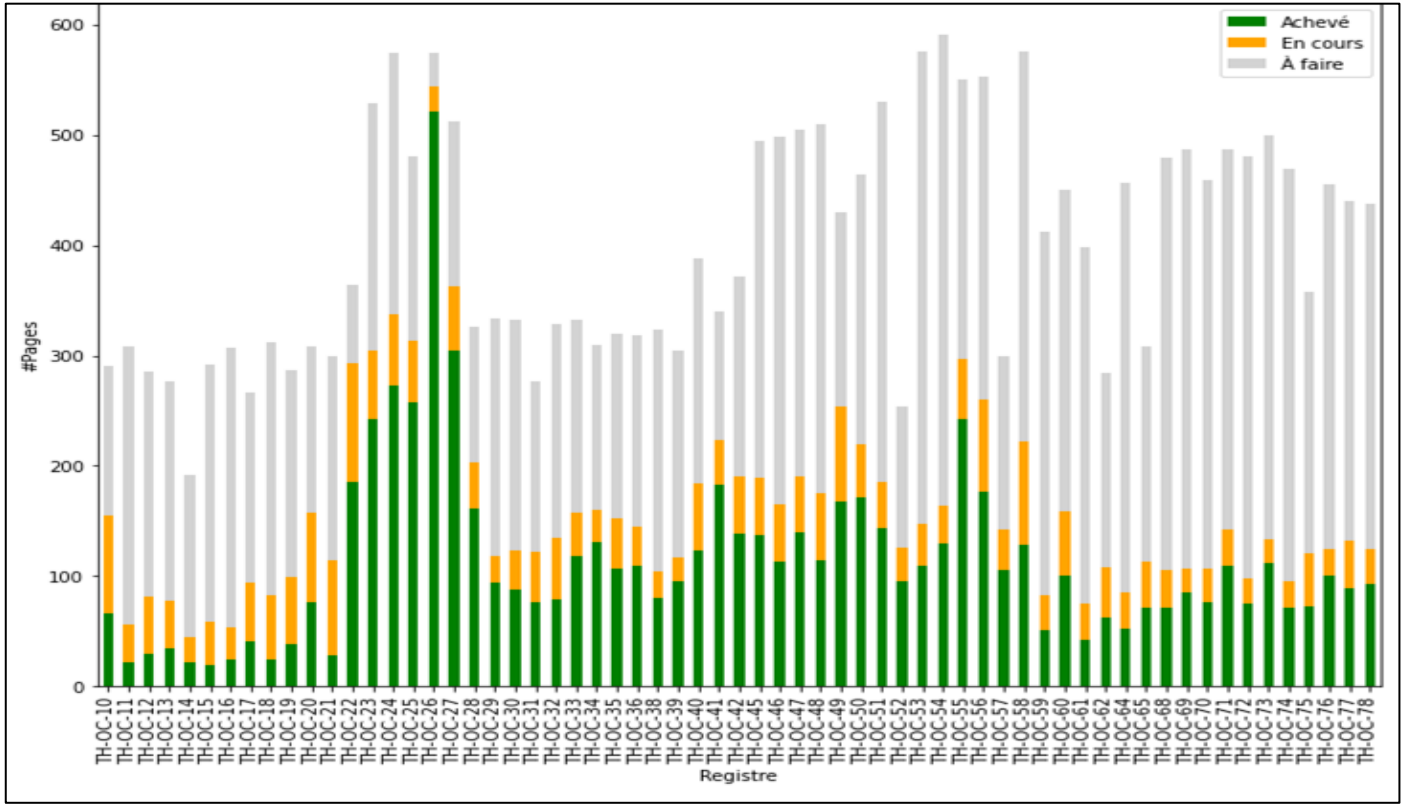
# Vérification

# Activité au cours du projet




Les pics :
réunions MSH et
1er confinement

# RECITAL : état d'achèvement automne 2020



JOURNÉES SCIENTIFIQUES
UNIVERSITÉ DE NANTES

# Exploitation informations

## Soirée du 1764-08-05

### Pièces jouées

| Titre | Confiance | Provenance |
|---|---|---|
| divertissement | 1.0 | record_linkage |
| Fermière | 0.668153104781061 | record_linkage |
| Bastien et Bastienne | 0.629940788348712 | record_linkage |

### Acteurs présents

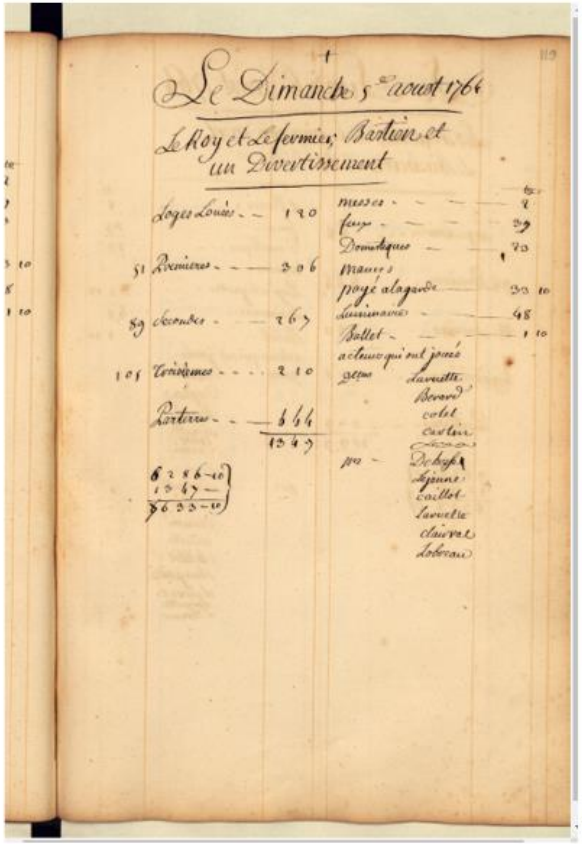
| Nom |
|---|
| Carlin |
| Colet |
| Lavuette |
| Berare |

### Informations budgétaires

Total des recettes : 1349

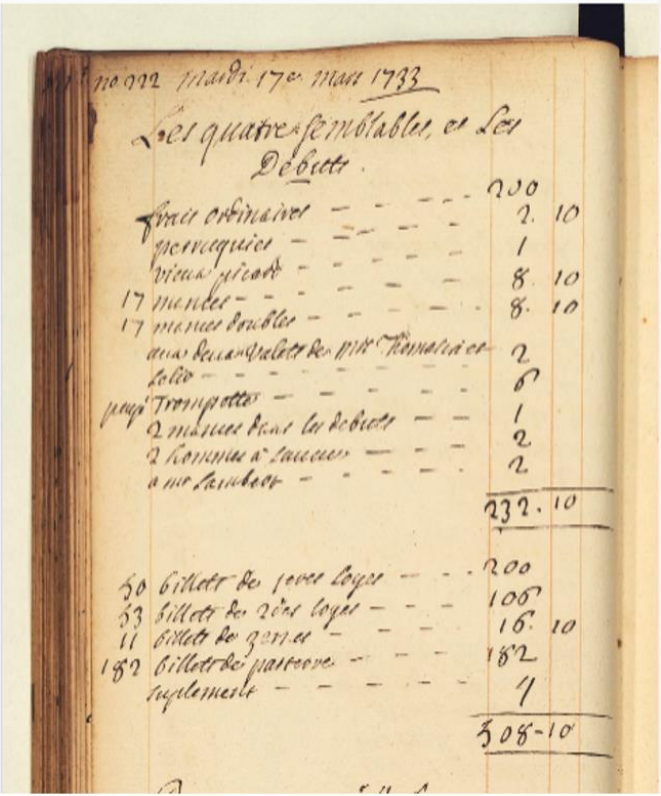
| Type | Montant | Quantite |
|---|---|---|

Total des dépenses : undefined



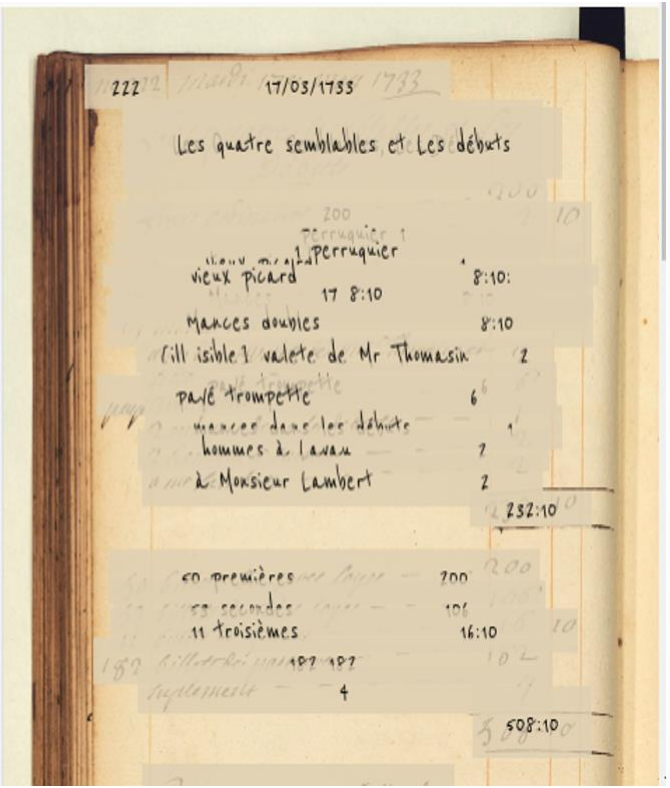
# Représentations substituts numériques (1)

| Box | Type | Valeurs |
|---|---|---|
| 770x99+453+109 | db_date | 17/03/1733 |
| 1149x189+325+189 | db_date | Les Quatre Semblables, Les Débu |
| 222x93+224+117 | db_page_number | 222 |
| 1177x115+290+1120 | db_first_box | 50 200 |
| 1180x113+303+1186 | db_second_box | 53 106 |
| 1280x95+292+1241 | db_third_box | 11 16:10 |
| 1254x99+247+1303 | db_floor | 182 182 |
| 1130x101+368+1361 | db_extra | 4 |



# Représentations substituts numériques (2)

| Box | Type | Valeurs |
|---|---|---|
| 770x99+453+109 | db_date | 17/03/1733 |
| 1149x189+325+189 | db_date | Les Quatre Semblables, Les Débu |
| 222x93+224+117 | db_page_number | 222 |
| 1177x115+290+1120 | db_first_box | 50 200 |
| 1180x113+303+1186 | db_second_box | 53 106 |
| 1280x95+292+1241 | db_third_box | 11 16:10 |
| 1254x99+247+1303 | db_floor | 182 182 |
| 1130x101+368+1361 | db_extra | 4 |




© Olivier Aubert

# Les problèmes du crowdsourcing

- évaluer la fiabilité des bénévoles : compétence et honnêteté
- évaluer la difficulté des tâches proposées
  - comment découper la tâche collective en micro-tâches indépendantes ?
  - quelle politique d'affectation des tâches aux bénévoles ?
- résoudre les divergences d'opinion entre bénévoles

Objectif : améliorer la qualité des données, réduire l'effort/le coût

*[ IA et histoire du théâtre]*
Françoise RUBELLIN (LAMO)

# L'IA est de retour… en support au crowdsourcing ?

- Recherche de consensus
- Apprentissage des profils de bénévoles
- Apprentissage de la difficulté des tâches

Exemple des saisies multiples d'utilisateurs pour un nom, Dehesse.

Jusqu'où peut-on se passer de la décision de l'expert ? Débat en cours !

En cours : Projet de 2 étudiants de Polytech en PRED (projet de R&D) sur l'IA appliquée aux données de RECITAL (encadrement Guillaume Raschia et Olivier Aubert)