RECITAL
Combining crowdsourcing and expertise in Digital Humanities

FOSDEM - February 2021

Olivier Aubert
www.olivieraubert.net - contact@olivieraubert.net

in collaboration with Françoise Rubellin (LAMO - Univ. Nantes)
and Guillaume Raschia (LS2N - Univ. Nantes)

Context:
Comédie-Italienne history

RECITAL: Registres de la Comédie-Italienne

CETHEFI - Centre d'Études des Théâtres de la Foire et de la Comédie-Italienne

cethefi

Literature and history lab aiming at studying fairground theaters and Italian Comedy in Paris around XVIIIth century

Multiple approaches

- edition and study of unedited plays (Ciresfi)
- musical studies and database (Theaville)
- VR reconstitution of no longer visible theatres (VESPACE)
- accounting registers study (RECITAL)

What are we expecting to learn from accounting registers?

- performed plays titles
- sold tickets (by category)
- expenses (taxes, accessories, musicians)
- theatre accounting

We want to learn generalities, but also find out exceptional things/events

Data to re-think theatre history

Economic history

author’s royalties, subscriptions, accounting rules…

Social history

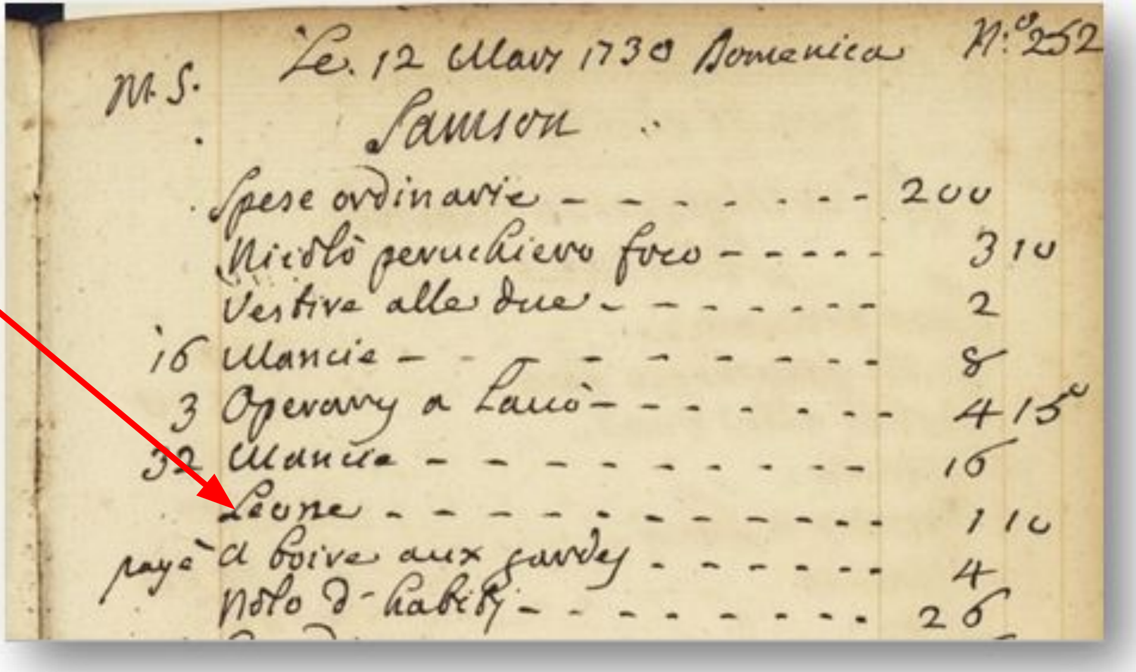
audience composition, placement, involved actors…

Hardware history

sets making, costumes…

Historical investigations...

Leone : Lion
(a turkey actually)



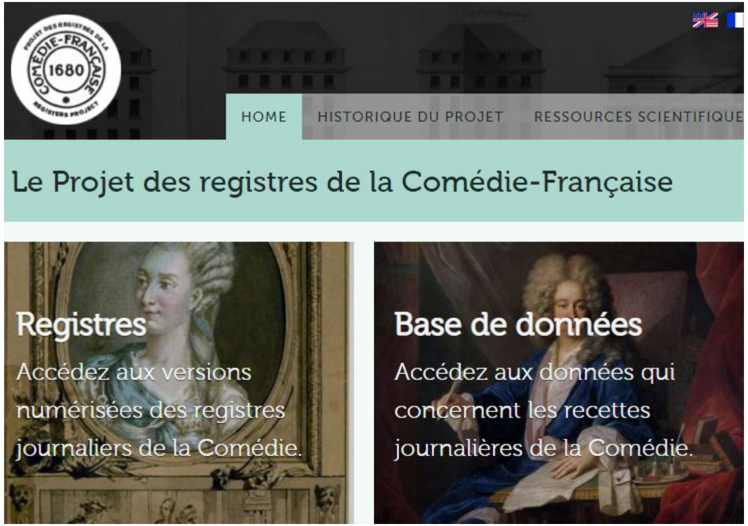
M. S. Le 12 Mars 1730 Domenica N:o 252

Samson

	Spese ordinarie	200
	Nicolò perucchiere foco	3 10
	Vestire alle due	2
16	Mancie	8
3	Operarj a Laccò	4 15
32	Mancia	16
	Leone	1 10
payé	a boire aux gardes	4
	Nolo d'habitj	2 6

A similar project is underway on Comédie Française registers

See https://www.cfregisters.org/



HOME | HISTORIQUE DU PROJET | RESSOURCES SCIENTIFIQUE

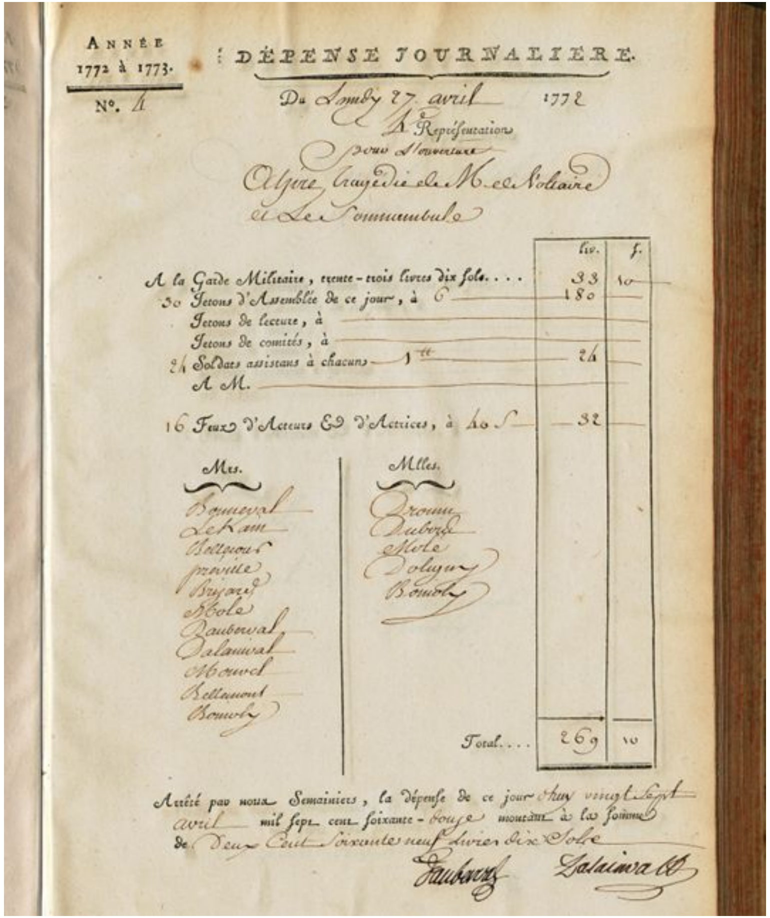
Le Projet des registres de la Comédie-Française

Registres

Accédez aux versions numérisées des registres journaliers de la Comédie.

Base de données

Accédez aux données qui concernent les recettes journalières de la Comédie.



ANNÉE 1772 à 1773.

N°. 1

DÉPENSE JOURNALIÈRE.

Du Lundy 27 avril 1772

1re Repréſentation pour l'ouverture

Alzire tragédie de M. de Voltaire et Le Somnambule

	liv.	ſ.
A la Garde Militaire, trente-trois livres dix ſols. . . .	33	10
30 Jetons d'Aſſemblée de ce jour, à 6	180	
Jetons de lecture, à		
Jetons de comités, à		
24 Soldats aſſiſtans à chacuns 1 [illegible]	24	
A M.		
16 Feux d'Acteurs & d'Actrices, à 40 ſ	32	
Total. . . .	269	10

Mrs.	Mlles.
Bonneval	Drouin
LeKain	Dubois
Bellecour	Molé
préville	Doligny
Brizard	Bouvet
Molé	
Dauberval	
Dalainval	
Monvel	
Bellemont	
Bouret	

Arrêté par nous Semainiers, la dépenſe de ce jour d'huy vingt ſept avril mil ſept cent ſoixante-douze montant à la ſomme de Deux Cent Soixante neuf livres dix Sols

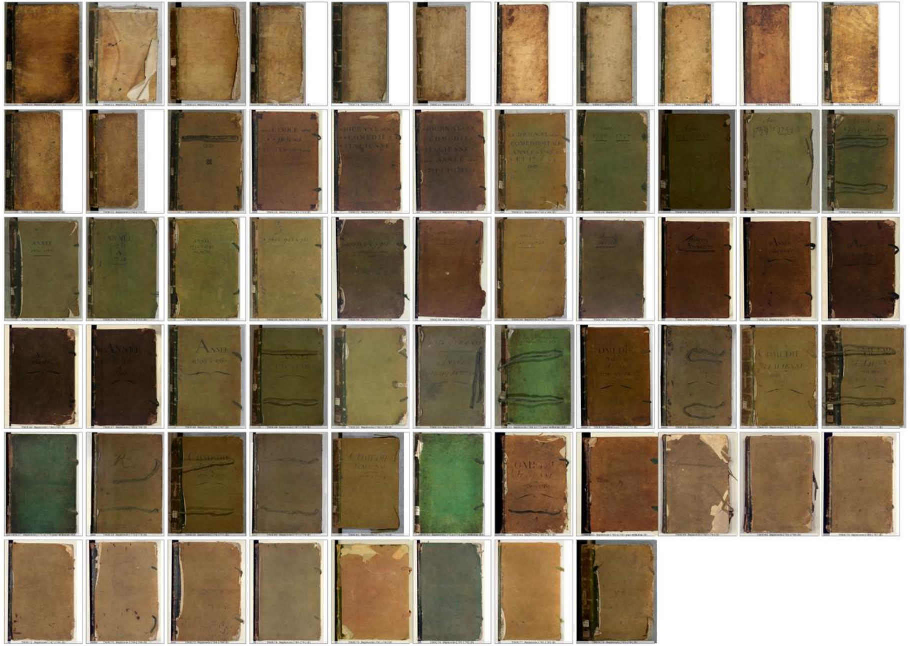
Dauberval Dalainval

The corpus

Accounting registers

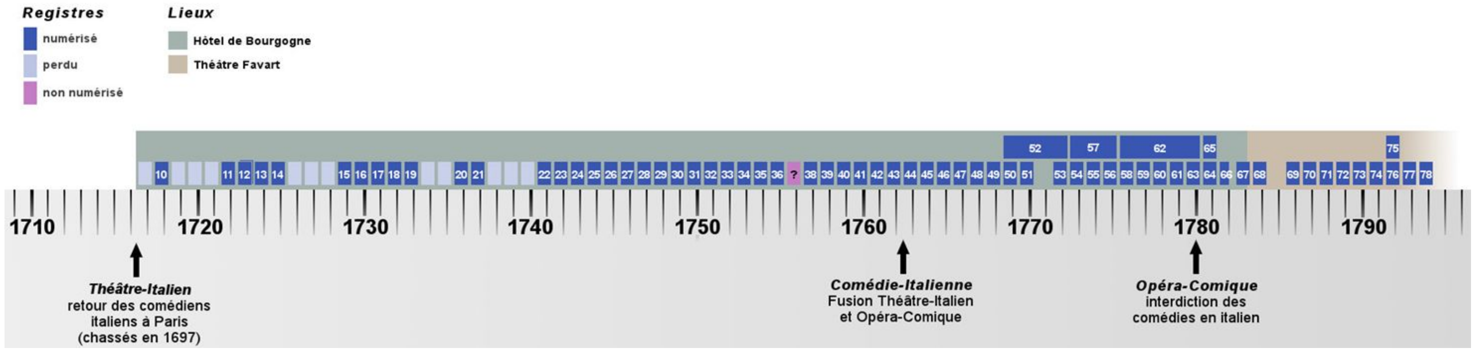
kept under the roof of the Bibliothèque-musée de l’Opéra

Available corpus



Registers for Comédie-Italienne - now available on Gallica

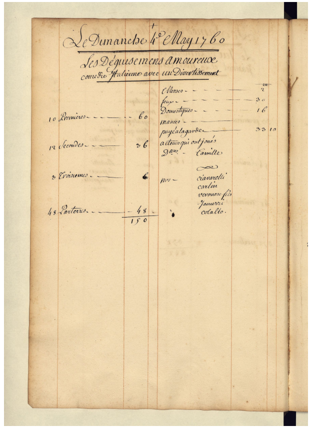
A century of accounting registers



(figure by Florent Coubard)

- 1 Register = 1 Theatre Season, from April to March
- digitized corpus: 1717-1794 (with 13 seasons missing)
- 64 registers of 300 to 600 pages
- about 26 000 pages

Anatomy of a register



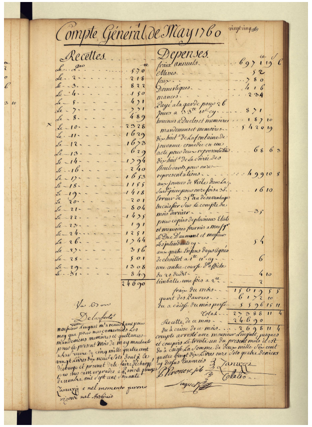
Le Dimanche 4e May 1760

Les Déguisemens amoureux

comedie Italienne avec un Divertissement

		Messes	2
		feux	30
		Domestiques	16
10 Premieres	60	manées	
		payé a la garde	33 10
12 Secondes	36	acteurs qui ont joués	
		Dlles Camille	
2 Troisiemes	6	mrs ciavarelli carlin veronese fils Zanuzzi colalto	
48 Parterres	48		
	150		

Daily accounts
77%



Compte Général de May 1760 vingt cinq

Recettes		Depenses	
Le 1er	570	fraix annuels	697 19 6
Le 2	218	Messes	52
Le 3	822	feux	780
Le 4	150	Domestiques	416
Le 5	471	manées	234
Le 7	771	Payé a la garde pour 26 jours a 33lt 10s cy	871
Le 8	489	sommes a Duclos et memoires	187 10
Le 10	2328	mandemens et memoires	5420 19
Le 11	1629		
Le 12	1673	dix huite de La fontaine de jouvance comedie en un acte pour deux representations	68 6 3
Le 13	629		
Le 14	1794	dix huite de La soirée des Boulevards pour onze representations	499 10 8
Le 16	240		
Le 17	1653		
Le 18	1155	aux joueurs de Viele dans La Surdepiece pour onze fois a 30s	16 10
Le 19	1418		
Le 20	201	Erreur de 35lt au désavantage du caissier sur le compte du mois dernier	35
Le 21	804		
Le 22	1475	pour copies de plusieurs Etats et memoires fournis a Monsgr Le Duc D'aumont et messieurs Les jntendants cy	54
Le 23	191		
Le 24	1251		
Le 26	1744		
Le 27	356	aux quatre Enfans dans Agnés de chaillot a 1lt 10s cy	6
Le 28	501		
Le 29	1308	une contre course d'affiche du 20 dudit	4 10
Le 31	849	timballe une fois a 2lt	2
	24690	fraix du mois	15619 5 5
		quart des Pauvres	6172 10
		du a caisse du mois passé	5596 15 11
		Total	27388 11 4
		Recette de ce mois	24690
		du a caisse de ce mois	2698 11 4

Vu Bon

Delafosté

monsieur Singuet m'a remis tant pour moy que pour mes camarades les mandemens memoires et quittances pour le present mois de may montant a La somme de cinq mille quatre cent vingt Livres dix neuf sols dont je le décharge et promet de le faire décharger par mes camarades a Paris le quinze decembre mil sept cent soixante

Zanuzzi e nel momento furono riposte nel Archivio

compte arresté avec monsieur Singuet jusques et compris le trente un du present mois il est du a caisse La somme de deux mille six cent quatre vingt dix Livres onze sols quatre deniers cy dessus énnoncés

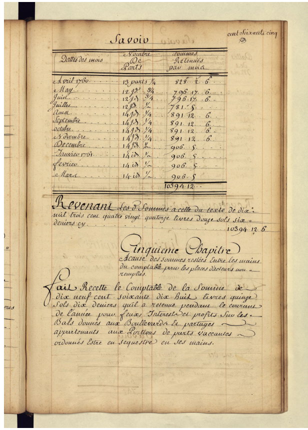
Zanuzzi

P Veronese fils

Colalto

Singuet

Monthly accounts
10%



cent soixante cinq

Savoir

Dattes des mois	Nombre De Parts	Sommes Retenües par mois
Avril 1760	13 parts ¼	828lt 2s 6d
May	12 jd 3/4	796 17 6
Juin	12 jd 3/4	796 17 6
Juillet	12 jd ½	781 5
Aoust	14 jd ¼	891 12 6
Septembre	14 jd ¼	891 12 6
octobre	14 jd ¼	891 12 6
Novembre	14 jd ¼	891 12 6
Decembre	14 jd ½	906 5
Janvier 1761	14 id ½	906 5
fevrier	14 id ½	906 5
Mars	14 id ½	906 5
		10394 12

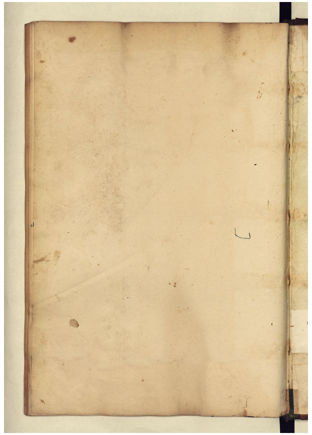
Revenant Les d. Sommes a celle du texte de dix mil trois cent quatre vingt quatorze livres douze sols six deniers cy 10394 12 6

Cinquieme Chapitre

Acause des sommes restées Entre les mains du comptable pour les places d'acteurs non remplies.

fait Recette le Comptable de la Somme de dix neuf cent soixante dix huit livres quinze sols dix deniers qu'il a retenue pendant le courant de l'année pour feux, Interests et profits sur les Bals donnés aux Boullevards Et partages appartenants aux Portions de parts vaccantes ordonnés Estre en sequestre en ses mains.

Final state + various
3%



Blank pages
10%

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

Original corpus characteristics

- Handwritten pages
- 2+ languages: French and Italian (various dialects)
- 7+ « bookkeepers »: from Alborghetti to Linguet
- Currency of the Ancien Régime : Livre-Sou-Denier
- All along the century
    - formal changes of balance sheets
    - updating of accounting rules

Digitizing approaches

Two complementary approaches

Automated (AI-based)

vs

Manual (expert+crowdsourcing)

AI automation

Usage of AI for segmentation and transcription

- PhD work by Adeline Granet during the project (2015-2018)
- Difficult corpus characteristics + lack of ground truth (bootstrap)

Off-the-shelf open solutions now exist for HCR

- Transkribus: mostly open-source, but recent paying model for using models in SAAS mode.
- eScriptorium (web interface for Kraken): open-source

Crowdsourcing approach

- more tedious, workforce-demanding
- fit for irregular contents
- more apt at identifying exceptional/interesting items
- a requirement for building the ground-truth
- issues of data quality evaluation and data validation

Crowdsourcing platform - ScribeAPI

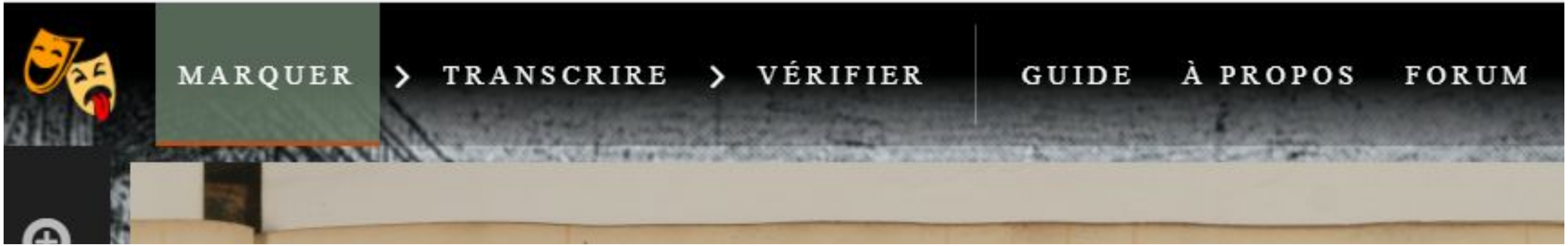
ScribeAPI software

issued from the Zooniverse project

MIT-licensed

Ruby-on-Rails platform

Crowdsourcing tasks in short

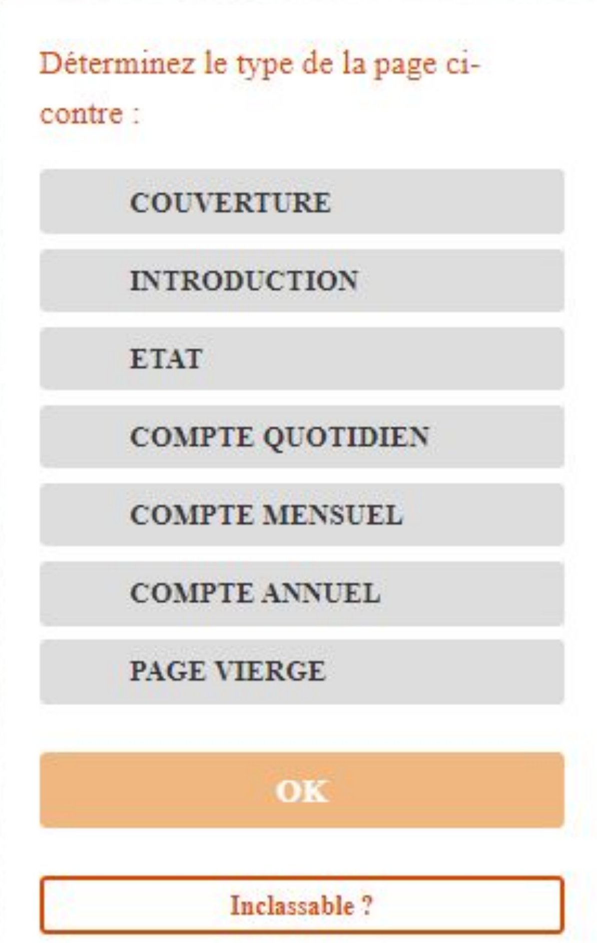


3 activities

- Mark (segment and categorize)
- Transcribe
- Verify

Crowdsourcing interface

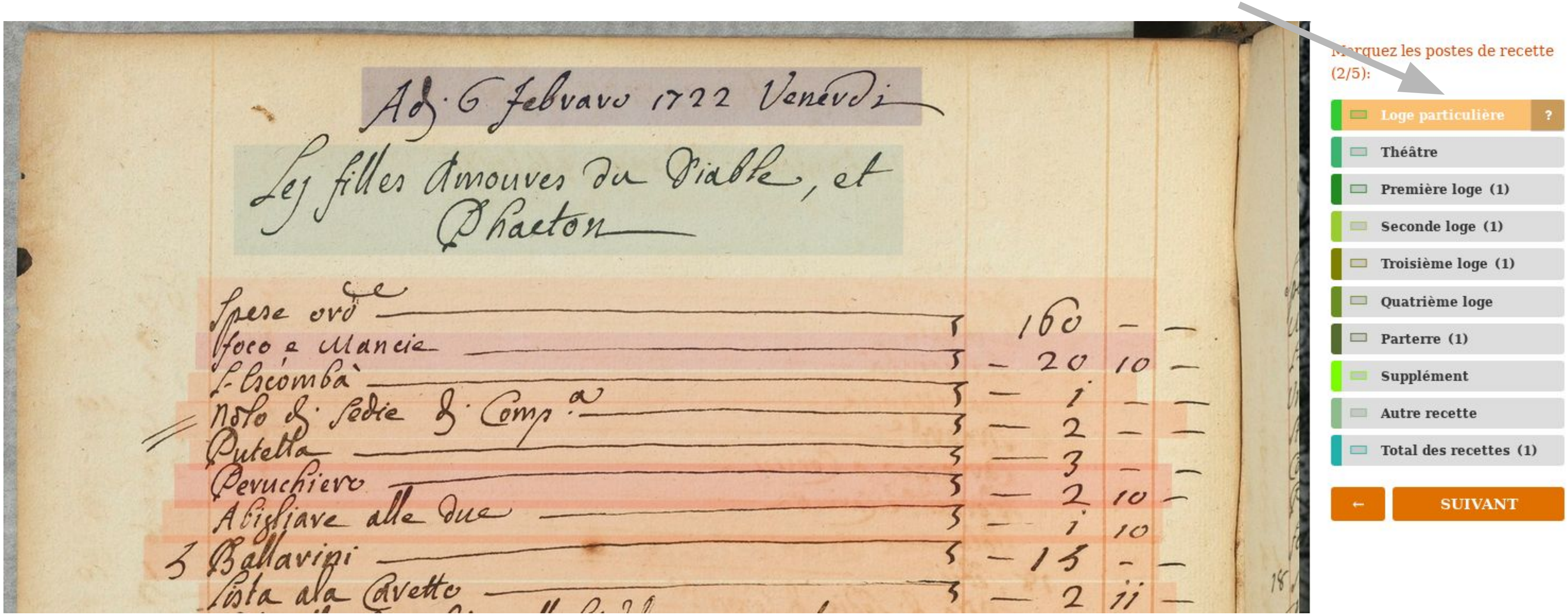
8 different page types



Déterminez le type de la page ci-contre :

Crowdsourcing interface - marking

133 data categories - displayed according to the page type



Adì 6 Febraro 1722 Venerdì

Les filles Amoures du Diable, et Phaeton

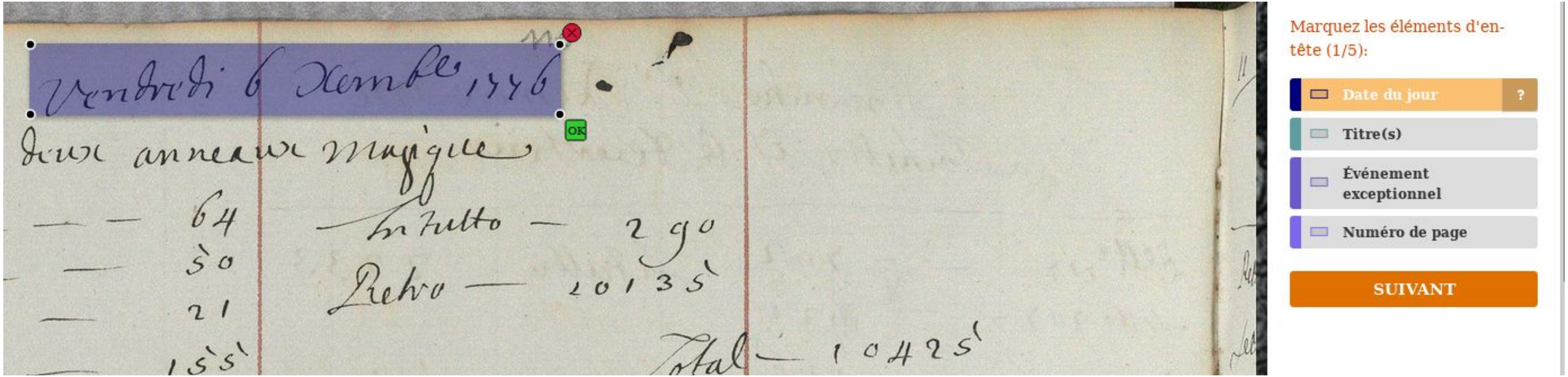
Spese ord.e	£	160	—	—	
Foco e Mancie	£	—	20	10	—
L'Ecombà	£	—	1	—	—
Nolo di Sedie di Comp.a	£	—	2	—	—
Putella	£	—	3	—	—
Peruchiero	£	—	2	10	—
Abigliare alle due	£	—	1	10	
3 Ballarini	£	—	15	—	—
Vista ala Cavetto	£	—	2	11	—

Marquez les postes de recette (2/5):

- □ Loge particulière ?
- □ Théâtre
- □ Première loge (1)
- □ Seconde loge (1)
- □ Troisième loge (1)
- □ Quatrième loge
- □ Parterre (1)
- □ Supplément
- □ Autre recette
- □ Total des recettes (1)

← SUIVANT

Crowdsourcing interface - marking



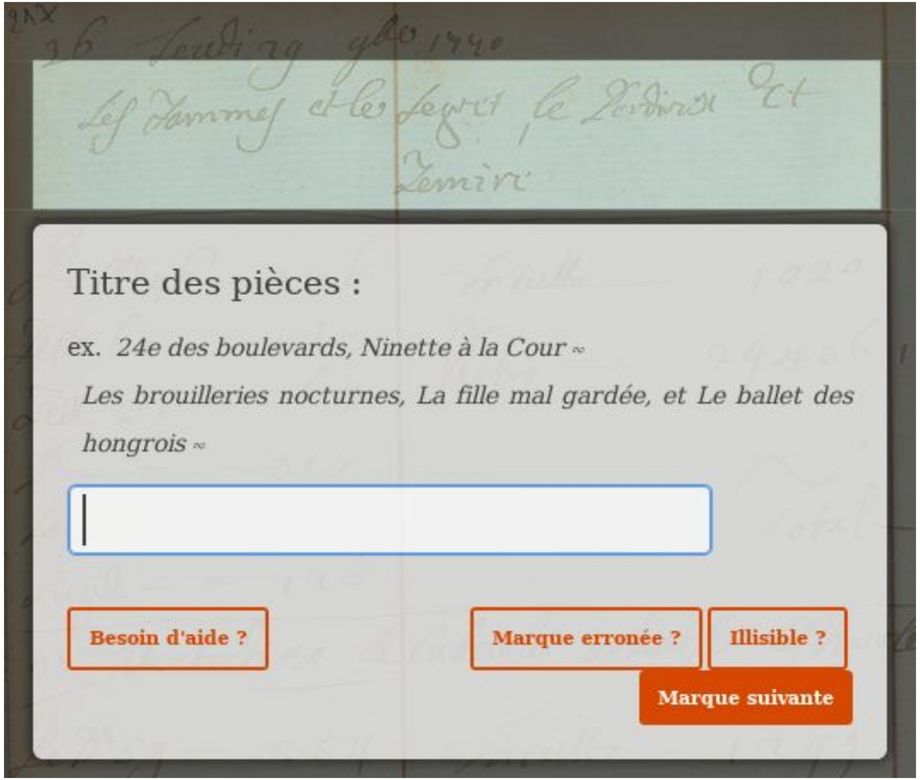
Vendredi 6 Xembre, 1776

deux anneaux magique

64	In Tutto	290	
50	Retro	20135	
21			
15s'		Total	10425'

Marquez les éléments d'en-tête (1/5):

Crowdsourcing interface - transcription



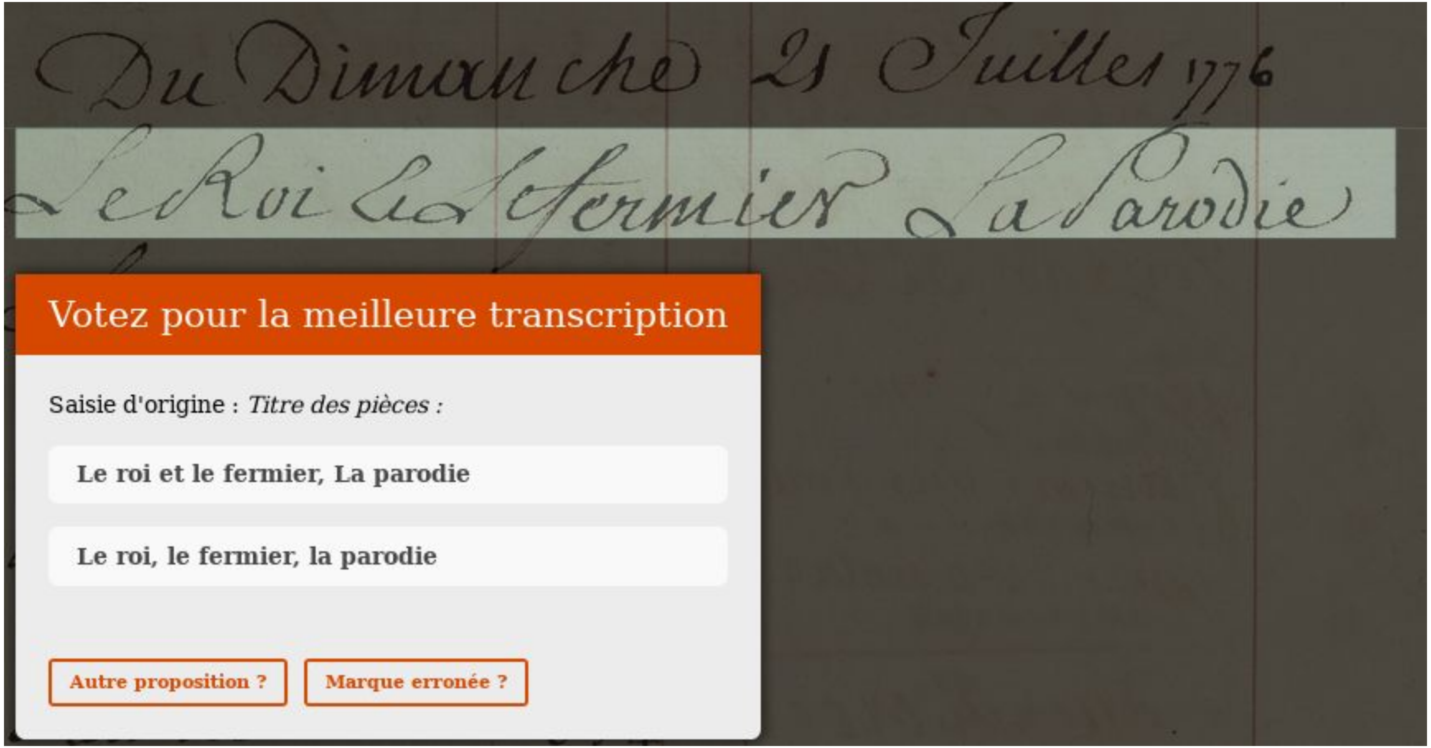
Titre des pièces :

ex. *24e des boulevards, Ninette à la Cour ~*

Les brouilleries nocturnes, La fille mal gardée, et Le ballet des hongrois ~

Les femmes et le secret, la perdrix et Zémire

Crowdsourcing interface - verification



Du Dimanche 21 Juillet 1776

Le Roi et le fermier La Parodie

Votez pour la meilleure transcription

Saisie d'origine : *Titre des pièces :*

Le roi et le fermier, La parodie

Le roi, le fermier, la parodie

ScribeAPI

Pros

- Open source
- Collaborative annotation of images
- Able to express chained task sequences (mark-transcribe-verify)
- Well thought interface (accessible to non-technical users)

Cons

- Complex architecture/model
- End-of-life reached in 2016

Open-Source and Technical debt

Official ScribeAPI - End-of-Life in 2016

Dependencies on old versions of Ruby, js/coffeescript, Mongo, unbuildable as-is now

Current server/dependencies dockerized

But not all hope is lost

effort by Utrecht University to update critical dependencies (unmerged pull request)

After crowdsourcing - ongoing work

- Data quality evaluation
    - first phase crowdsourced (peer-reviewed verification)
    - second phase automated
    - third phase manual
- Data cleaning and validation
    - custom tools, using external information (list of play titles, actors...)
    - dedicated user-interface (dashboard) for experts
    - integration of AI-based tools
- Data publication
    - FAIR principles
    - collaboration/links with other projects/data

If you want to contribute, welcome at...

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