

Projets et Humanités Numériques

Giovanni Pietro Vitali

giovannipetrovitali@gmail.com | giovannni.vitali@uvsq.fr




Université de Versailles Saint Quentin en Yvelines
Université Paris-Saclay

La littérature du XIXe siècle au prisme des humanités numériques

Giovanni Pietro Vitali

L'analyse des textes du XIXe siècle par le biais des instruments informatiques simplifie la vie des chercheurs, car l'échange et l'exploitation des textes au format numérique sont gratuits, sans aucune limitation d'usage liée aux droits d'auteur ; ces textes sont faciles à consulter en ligne, grâce à des initiatives mondiales d'échange de la mémoire littéraire, comme le *Project Gutenberg* . Ces œuvres numériques donnent la possibilité de montrer des éditions ou des visualisations comprenant des parties significatives de texte.



About ▾ Search and Browse ▾ Help ▾


Quick search [Donation](#)

Welcome to Project Gutenberg


Project Gutenberg is a library of over 60,000 free eBooks

This is the new Project Gutenberg site See the [new website](#) page for information about currently known issues, and how to report problems or suggest changes.


Choose among free epub and Kindle eBooks, download them or read them online. You will find the world's great literature here, with focus on older works for which U.S. copyright has expired. Thousands of volunteers digitized and diligently proofread the eBooks, for you to enjoy.




The Third Little Green Man




Black Priestess of Varda




A Daughter of Witches




Die Uhr




Mágia




Életemből (II. rész)



Ricardo's Virus



Notes on Noses



Give Back a World

Distant  *Reading*

Distant  *Reading*



Distant Reading

Distant Reading COST Action European Literary Text Collection (ELTeC)

Il vise à créer un corpus plurilingue, principalement consacré à la période qui va de la fin du XVIIIe au début du XXe siècle.

- L'objectif de la première phase du projet est de créer six sous-collections (100 romans par langue) pour la période 1850-1920 ;
- Ensuite, cette première partie sera intégrée avec au moins quatre sous-collections ultérieures (100 romans par langue) pour la même période.
- Ces matériaux serviront de noyau central d'ELTeC, lequel sera intégré avec une troisième contribution d'au moins six collections ultérieures, en ajoutant encore d'autres langues qui n'avaient pas été considérées dans les deux premières contributions, mais cette fois pour la période précédente, de 1780 à 1850.

ELTeC Summary Page

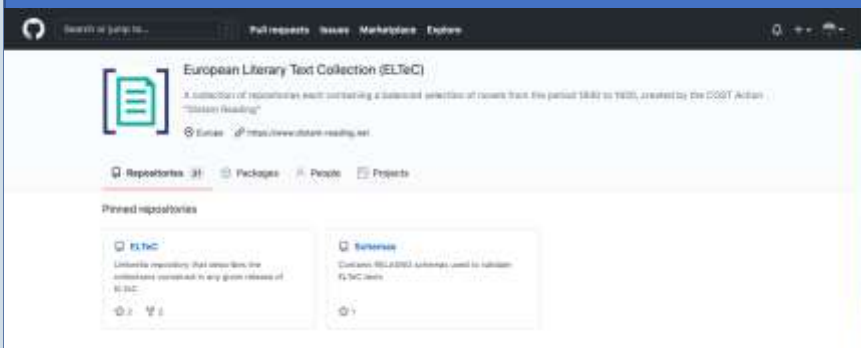
As part of the following summary statistics, this page provides data on human-readable abstracts of texts from Europe's earliest printed books in the European context. For full details, click on a language entry in the table below to see a list of titles with working in that language. Then click on the identifier of a title to see a more detailed view of the title as available in ELTeC. The original source that we found is a PDF of the title, which may be downloaded from there.

The PDF version gives the [ELTeC data](#) and the original text. The other column gives counts for each of the four dates shown, with counts in italics indicating that the version is uncertain. Working in progress. Right to see details of the title, click on the title in the table below.

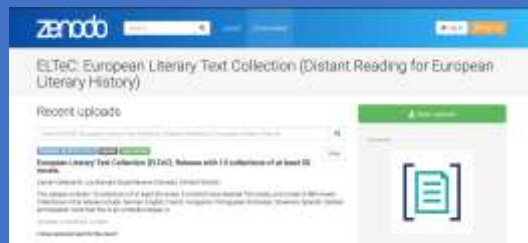
This column is a list of proposed human-readable and counts of the proposed and actual approaches. See below for the [ELTeC data](#).

Language	Title	Year	Words	Mean	LENGTH				TIME USED				REPERTORY COUNT			
					Points	3-10	11-20	More than 20	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
English	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
French	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
German	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
Italian	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
Latin	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
Portuguese	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
Spanish	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
Welsh	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
Yiddish	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899
Other	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899	1500-1599	1600-1699	1700-1799	1800-1899

Corpus:
<https://distantreading.github.io/ELTeC/>



Activites:
<https://github.com/COST-ELTeC>



Dépôt Données:
<https://zenodo.org/communities/eltec/?page=1&size=20>



Guidelines:
<https://distantreading.github.io/>

Outils



- <https://zenodo.org/>



- <https://github.com/>



- <https://hypotheses.org/>



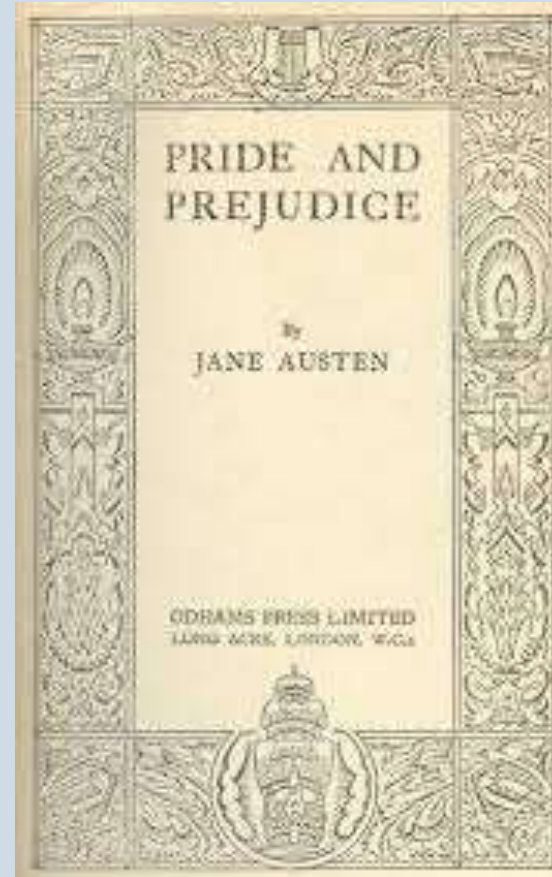
- <https://www.sciencesconf.org/>



- <https://hal.archives-ouvertes.fr/>



Jane Austen (1775-1817)



<https://janeausten.ac.uk/index.html>

Herman Melville (1819-1891)

<https://melville.electroniclibrary.org/>

ARCHIVE



BROWSE

EDITIONS



READ

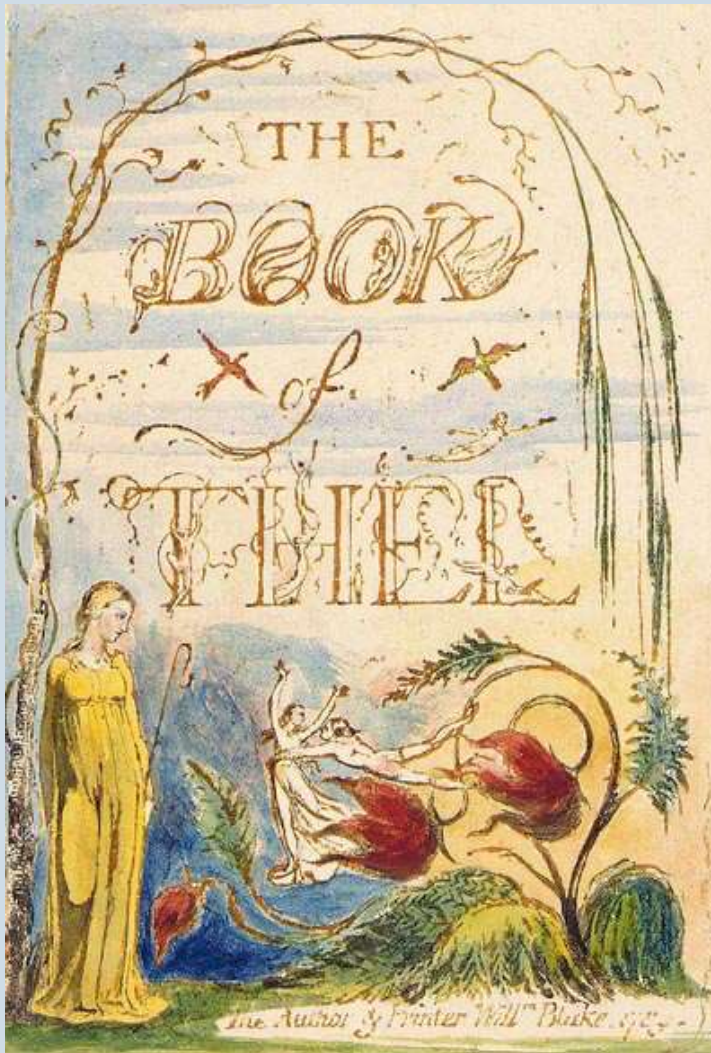
PROJECTS



CREATE

William Blake (1757-1827) Archive

<http://www.blakearchive.org/>



William Morris (1834-1896) Archive

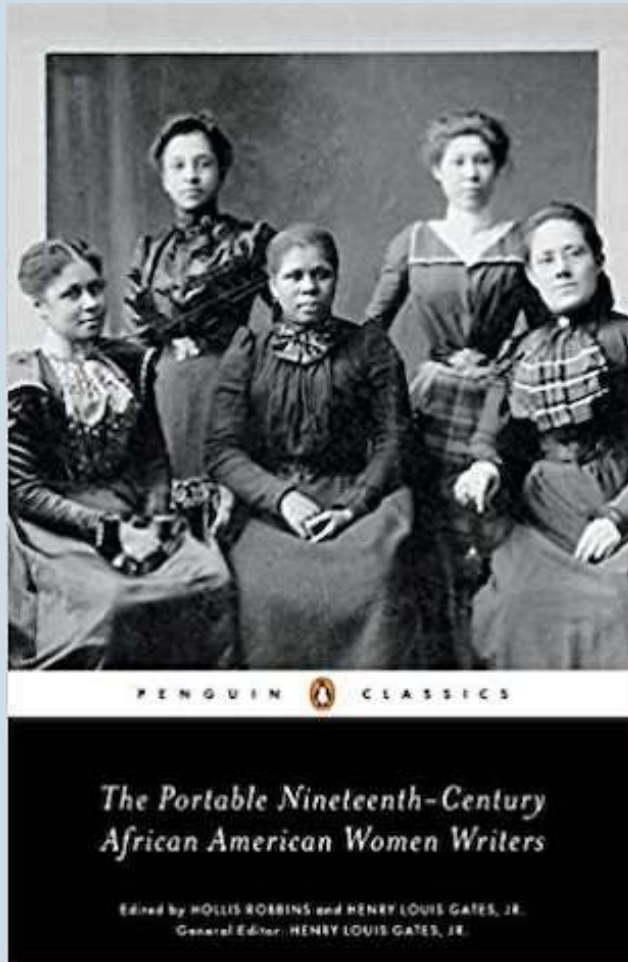
<http://morrisedition.lib.uiowa.edu/index.html>



William Morris, né le 24 mars 1834 à Walthamstow, Essex (aujourd'hui dans le borough londonien de Waltham Forest) et mort le 3 octobre 1896 à Hammersmith, Londres, est un fabricant, designer textile, imprimeur, écrivain, poète, conférencier, peintre, dessinateur et architecte britannique, célèbre à la fois pour ses œuvres littéraires, son engagement politique libertaire, son travail d'édition et ses créations dans le domaine des arts décoratifs, en tant que membre de la Confrérie préraphaélite, qui comptent parmi les sources du mouvement Arts & Crafts qui eut, dans ce domaine, l'une des influences les plus importantes en Grande-Bretagne au xxe siècle.

African American Women Writers of the 19th Century

http://digital.nypl.org/schomburg/writers_aa19/



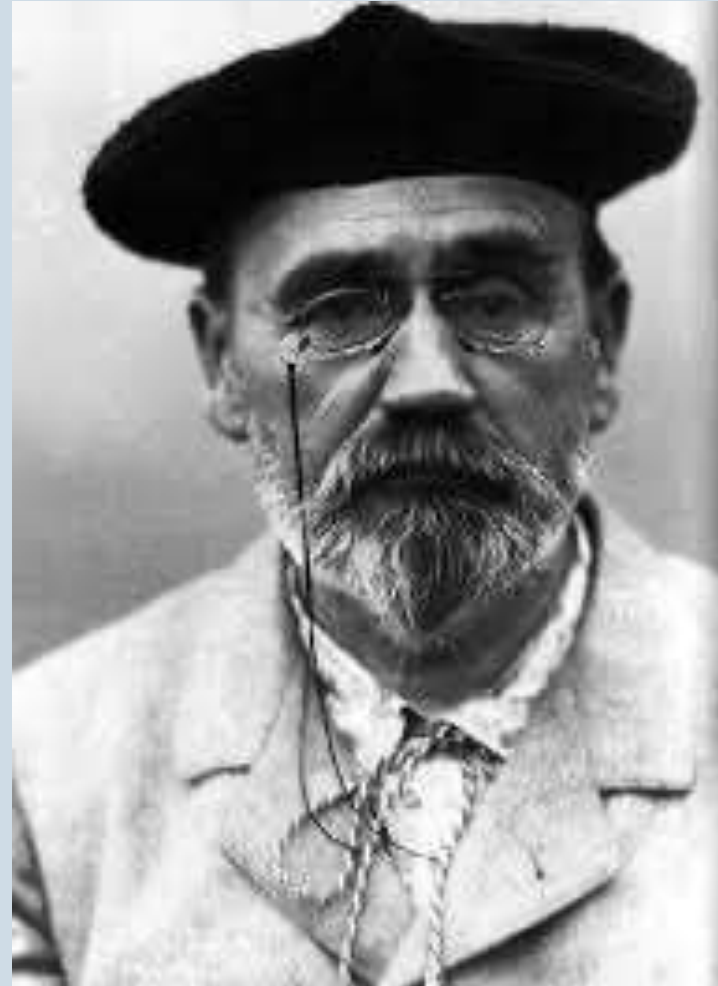
At the Circulating Library: A Database of Victorian Fiction, 1837-1901

<http://www.victorianresearch.org/atcl/index.php>



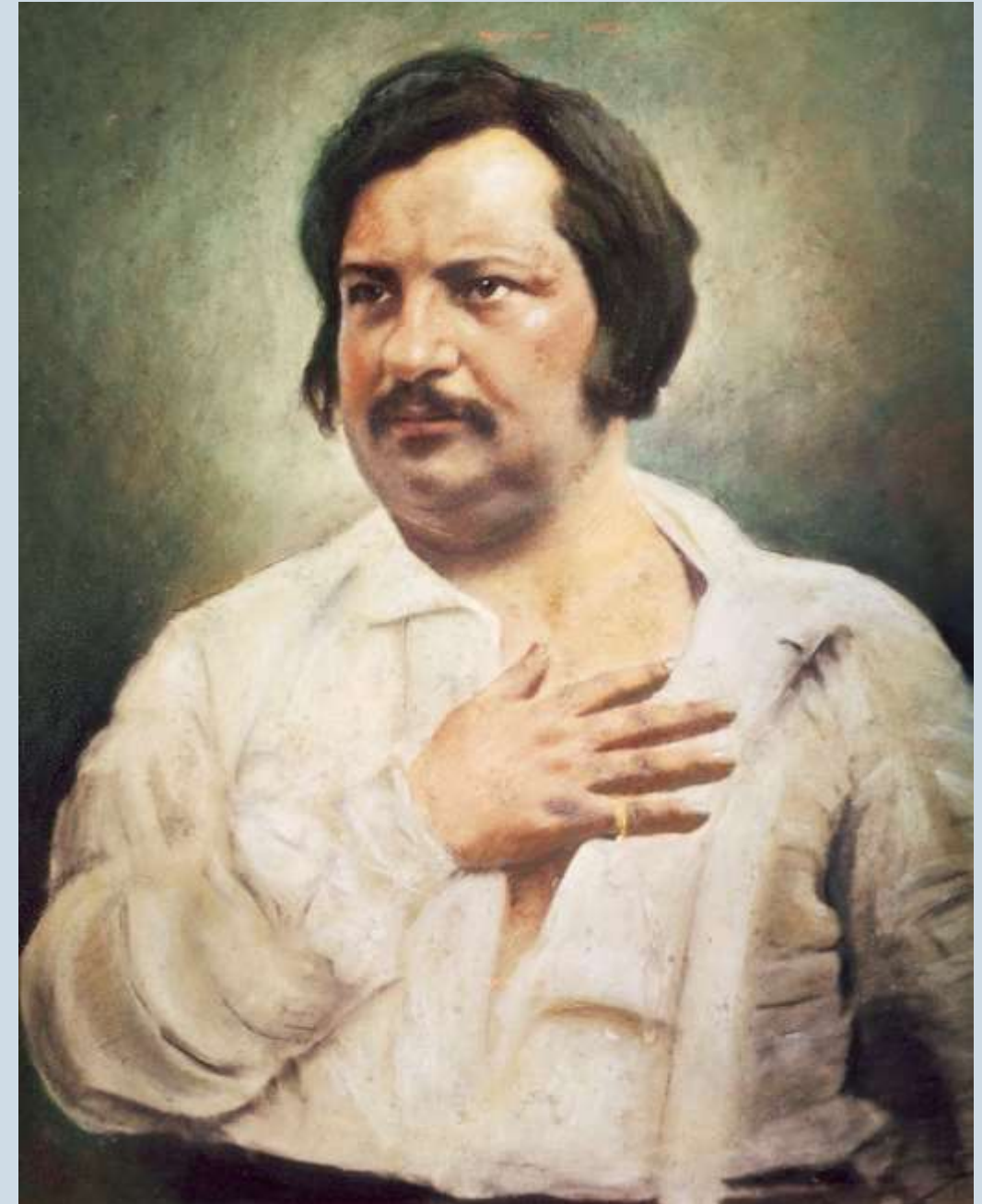
CORREZ - Édition des lettres internationales adressées à Émile Zola (1840-1902)

<http://eman-archives.org/CorrespondanceZola/>



Ebalzac (1799-1850)

<https://www.ebalzac.com/>



Au milieu de la rue Saint-Denis, presque au coin de la rue du Petit-Lion, existait naguère une de ces maisons précieuses qui donnent aux historiens la facilité de reconstituer par analogie l'ancien Paris. Les murs menaçants de cette hicoque semblaient avoir été hâriés d'hieroglyphes. Quel autre nom le fâneur pouvait-il donner aux X et aux V que traçaient sur la façade les pièces de bois transversales ou diagonales dessinées dans le badigeon par de petites bandes parallèles ? Évidemment, au passage de ~~ces~~ ~~les~~ voitures, chacune de ces solives s'agitait dans sa mortelle. Ce véralde édifice était surmonté d'un toit triangulaire dont aucun modèle ne se verra bientôt plus à Paris. Cette couverture, tordue par les intempéries du climat parisien, s'avançait de trois pieds sur la rue, autant pour garantir des eaux pluviales le seuil de la porte, que pour abriter le mur d'un grenier et sa lucarne sans appui. Ce dernier étage ~~était~~ construit en planches clouées l'une sur l'autre comme des ardoises, afin sans doute de ne pas charger cette frêle maison.

Par une matinée pluvieuse, au mois de mars, un jeune homme, soigneusement enveloppé dans son manteau, se tenait sous l'auvent d'une boutique qui se trouvait en face de ce vieux logis. Il examinait avec un enthousiasme d'archéologue. A la vérité,

CON. HUM. T. I.

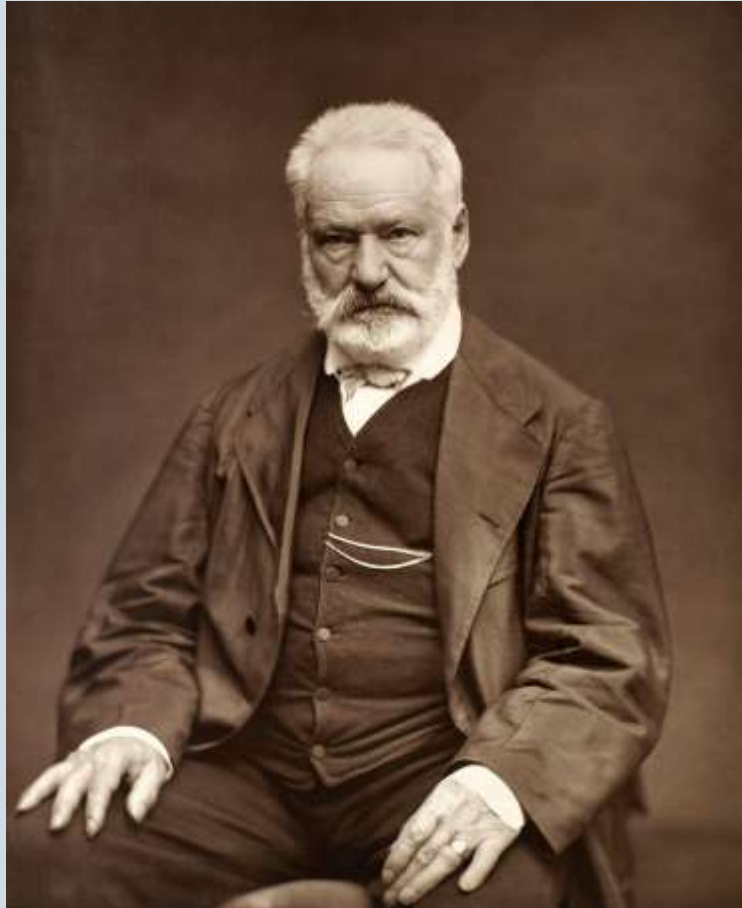
Mise en ligne de toute l'édition Furne

Nous avons un grand plaisir de vous annoncer que l'édition intégrale de *La Comédie humaine* est désormais enrichie de la version Furne. L'ajout d'une autre édition implique également la possibilité de comparer la version Furne et celle du Furne corrigé afin de rendre compte des corrections manuscrites effectuées par Balzac entre ces deux états.

En savoir plus...

Édition Nationale des Œuvres Complètes de Victor Hugo (1802-1885)

<https://hugo-en-images.nakalona.fr/>



Carto-Hugo

<http://cartohugo.elan-numerique.fr/>

Média19 & Numapress



<http://www.medias19.org/>



<http://www.numapresse.org/>

Literature



A Theatre of Places

Mapping 17th c. French Theatre

<https://hal.archives-ouvertes.fr/hal-02388411>

bit.ly/theatre_of_places

Data

N°	Titre	Auteur	Date	Genre	Actes	Vers	Téléchargements	
1	Théâtre Classique	(anonyme)	1722	L'actrice nouvelle	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
2	Théâtre Classique	(anonyme)	1739	Thélamire	Tragédie	5	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
3	Théâtre Classique	(anonyme)	1772	Le prix d'un moment	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
4	Théâtre Classique	(anonyme)	1772	Les bonnes gens	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
5	Théâtre Classique	(anonyme)	1772	Oh ! voilà bien le diable ! comédie	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
6	Théâtre Classique	(anonyme)	1784	Le club des dames ou le retour de descartes	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
7	CELLF. Bibliothèque dramatique	Abeille, Gaspard	1681	Lyncée, tragédie	Tragédie	5	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
8	Théâtre Classique	Aiguebelle Jean Dumas d'	1729	Pan et doris	Pastorale	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
9	Théâtre Classique	Aiguebelle Jean Dumas d'	1729	Polixène	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
10	Théâtre Classique	Aiguebelle Jean Dumas d'	1729	Trois spectacles	Prologue	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
11	Théâtre Classique	Aiguebelle Jean Dumas d'	1781	L'ivare amoureux	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
12	Théâtre Classique	Alan Robert	1711	L'épreuve réciproque	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
13	Théâtre Classique	Andrieux François Guillaume Jean Stanislas	1782	Anaximandre	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
14	Théâtre Classique	Anonyme		Farce nouvelle très bonne et fort joyeuse de la resurrection de jenin landore	Farce	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
15	Théâtre Classique	Anonyme	1770	Le portier du parnasse	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
16	Théâtre Classique	Anonyme	1778	Dialogue entre voltair et rousseau	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
17	Théâtre Classique	Anonyme	1785	La matinée du comédien français	Proverbe	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
18	Théâtre Classique	Annesime Louis	1759	Cendrillon	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
19	Théâtre Classique	Archambault Louis-François	1779	Janot chez la dégraisseur	Proverbe	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
20	Théâtre Classique	Archambault Louis-François	1779	Janot ou les battus paient l'amende	Proverbe	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
21	Théâtre Classique	Archambault Louis-François	1779	Ça n'en est pas ou tout ce qui reluit n'est pas or	Proverbe	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
22	Théâtre Classique	Archambault Louis-François	1780	Les étreintes de l'amitié	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
23	Théâtre Classique	Archambault Louis-François	1794	Hurlubreu ou tout de travers	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
24	Théâtre Classique	Archambault Louis-François	1794	Jocrisse changé de condition	Comédie	2	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
25	Théâtre Classique	Aristophane	1897	LES GUÉPES, COMÉDIE.	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
26	Théâtre Classique	Armand Huguenl Armand dit	1669	Le Cri de la nature	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
27	Théâtre Classique	Artaud Jean-Baptiste	1773	Le centenaire de moïse	Comédie	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
28	Théâtre Classique	Aude Joseph	1802	L'école tragique	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
29	Théâtre Classique	Aure François d'	1688	Digné	Tragédie	5	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
30	Théâtre Classique	Aure François d'	1689	Geneviève ou l'innocence reconnue	Tragédie	5	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
31	Théâtre Classique	Aure François d'	1670	Geneviève ou l'innocence reconnue. tragédie.	Tragédie	5	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
32	Théâtre Classique	Bacoulard François Thomas-Marie	1739	Le coligny	Tragédie	3	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
33	Théâtre Classique	Baptiste et du moulier	1885	Mascarade de la foire saint-germain	Mascarade	1	vers	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
34	Théâtre Classique	Barante Claude Ignace Brugière de	1696	Arlequin misanthrope	Comédie	3	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html
35	Théâtre Classique	Barante et dufrenoy	1897	Les fées ou les contes de ma mère l'oye.	Comédie	1	prose	TEI, epub, kindle, markdown, iamuteq, paroles, TXM, html, fragment html

Objectives

- Short term : create a map out of 40 plays
- Medium term : create training data for existing and future tools
- Medium term : create models for easily usable tools

Balancing the corpus

- 3 generations of authors
- 9 Authors (4 plays each)
- 3 genres : 17 tragedies, 10 comedies, 9 tragi-comedies

- Four are active in the first half of the century :



Pierre Du Ryer (1628-1655)

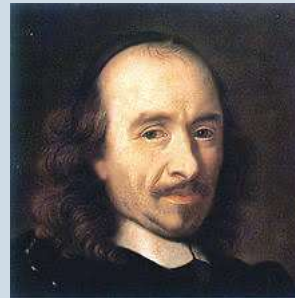


Georges de Scudéry (1631-1643)



Paul Scarron (1648-1658)

- One is in between these two groups :



Pierre Corneille (1629-1675)

- Four are active in the second half of the century :



Claude Boyer (1646-1697)



Thomas Corneille (1651-1696)



Molière (1655-1673)



Jean Racine (1664-1675)

Heterogeneity Looking for place names

- Different regions : from Asia to America
- Different periods : Biblical, classical, medieval, modern. .
- Different natures : cities, rivers, kingdoms, regions, seas
- Different linguistic origins : Greek, Italian, Spanish. . .

Complexities

For a non-specialist, it is not easy to distinguish toponyms from anthroponyms from toponyms (*Némée, Éryce*. . .) – and sometimes it is both (*mer Égée, i.e. Aegeus Sea* and not *Aegean Sea*)

(If it is complicated for a human being, it is also complicated for a NLP-based tool)

- Some places do not exist (*Chicuchiquizèque, Hell*)
- Some places cannot be located (*Vas*)
- Many place names are francised (*Gaiette*, cont. French *Gaète*, it. *Gaeta*)

NLP – Approaches

Combining methods

- Reading texts
- Using existing gazetteers
- Using POS-taggers and look for proper nouns
- Using NER(D) systems and look for

HER <https://github.com/alexerdmann/HER> [Alexander Erdmann - https://aclanthology.org/N19-1231/](https://aclanthology.org/N19-1231/)

- An active learning-based Named Entities Recognizer (NER) toolkit, which accelerates the process of manual annotation by pre-selecting sentences that are maximally informative to the model.
- It operates iteratively : a first limited wave of sentences are annotated to produce a first gazetteer, from which a second bigger wave is prepared for annotation, etc.
- In other words, HER does not guarantee the creation of an exhaustive gazetteer, but maximizes its exhaustiveness without reading entirely large corpora.

Efficiency

- With 100,000 pre-ranked tokens (out of c. 600,000) tagged with HER, we identified 86.2% of the Toponyms and 89.5% of the occurrences in our corpus.
- Place names represent 0.16% of the tokens of our corpus. Using POS-taggers, we can keep only proper nouns (c. 4% of the total), hoping that all toponyms are recognised
- (Not all POS-taggers are efficient enough to identify all place names, and unfortunately the less efficient are the most common among literary scholars)
- Using NER(D)-taggers, we can keep only named entities (c. 2% of the total) or even locations (c. 0.5% of the total), hoping that all toponyms are recognised

Test on one play, c. 16,500 tokens, 116 occurrences of place names, 18 types

	Proper nouns	Toponyms	%
TreeTagger	983	16	1.63%
MEIt	777	18	2.32%
Talismane	785	18	2.29%

Training NER

	Toponyms	Types
HER crf	17	12
HER bi-lstm	12	7
NER Stanford	11	6
Spacy	13 (+10 false)	8 (+3 false)

We have used additional data to increase the accuracy with Stanford NER and Spacy

	Toponyms	Types
NER Stanford	13	8
Spacy	13 (+3 false)	9 (+2 false)

A first Map

Objective

- I Not a static map, but a dynamic research environment
- I Easy to use and reuse, online and offline



A first
Map

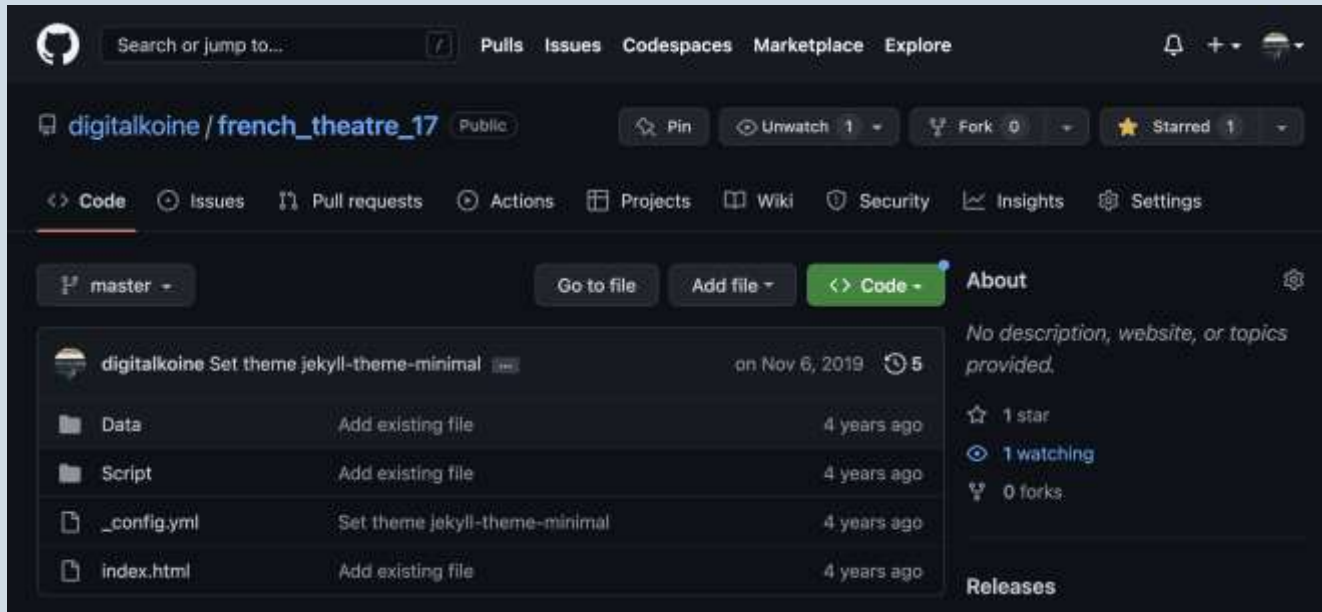


Start the presentation to see live content. Still no live content? Install the app or get help at PollEv.com/app

<http://bit.ly/2XwAaG1>

Github repository: bit.ly/french_theatre_17_github

Article link: bit.ly/theatre_of_places



A Second
Map



Map: bit.ly/french_theatre_17



Start the presentation to see live content. Still no live content? Install the app or get help at [PollEv.com/app](https://bit.ly/french_theatre_17)

Cultural horizon



Figure – Macro zones, the darker the more present in the corpus

Gravity centre



Genre

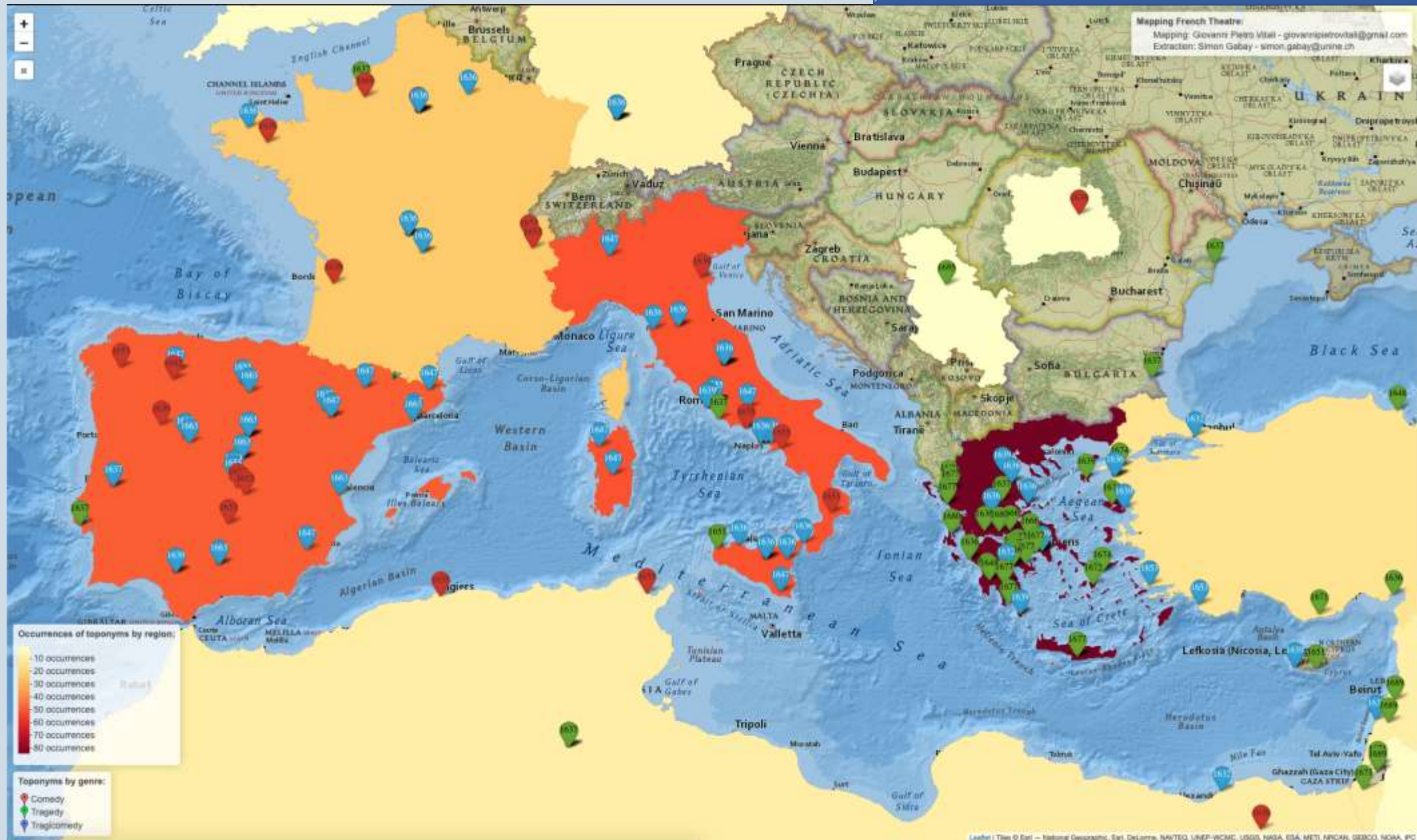
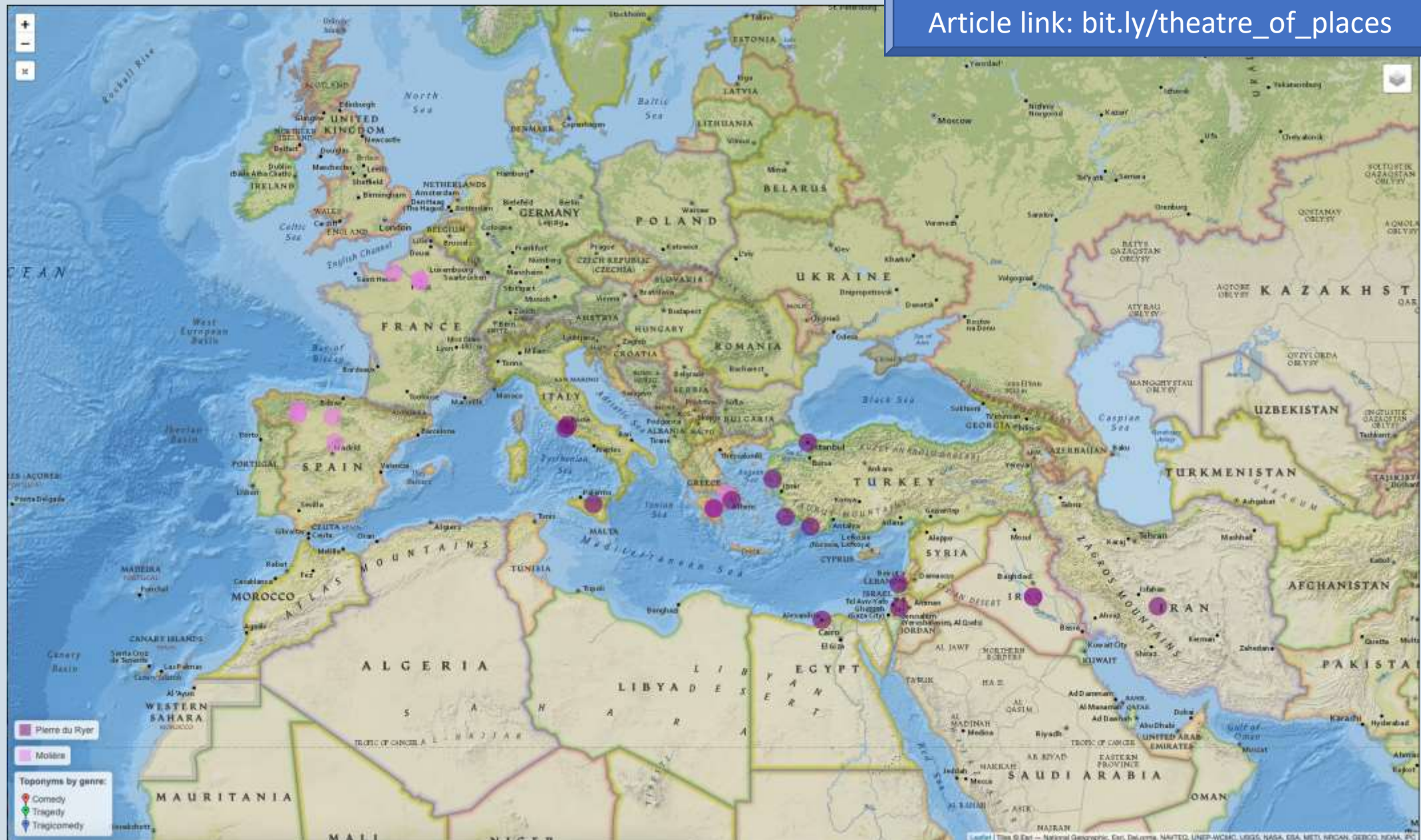
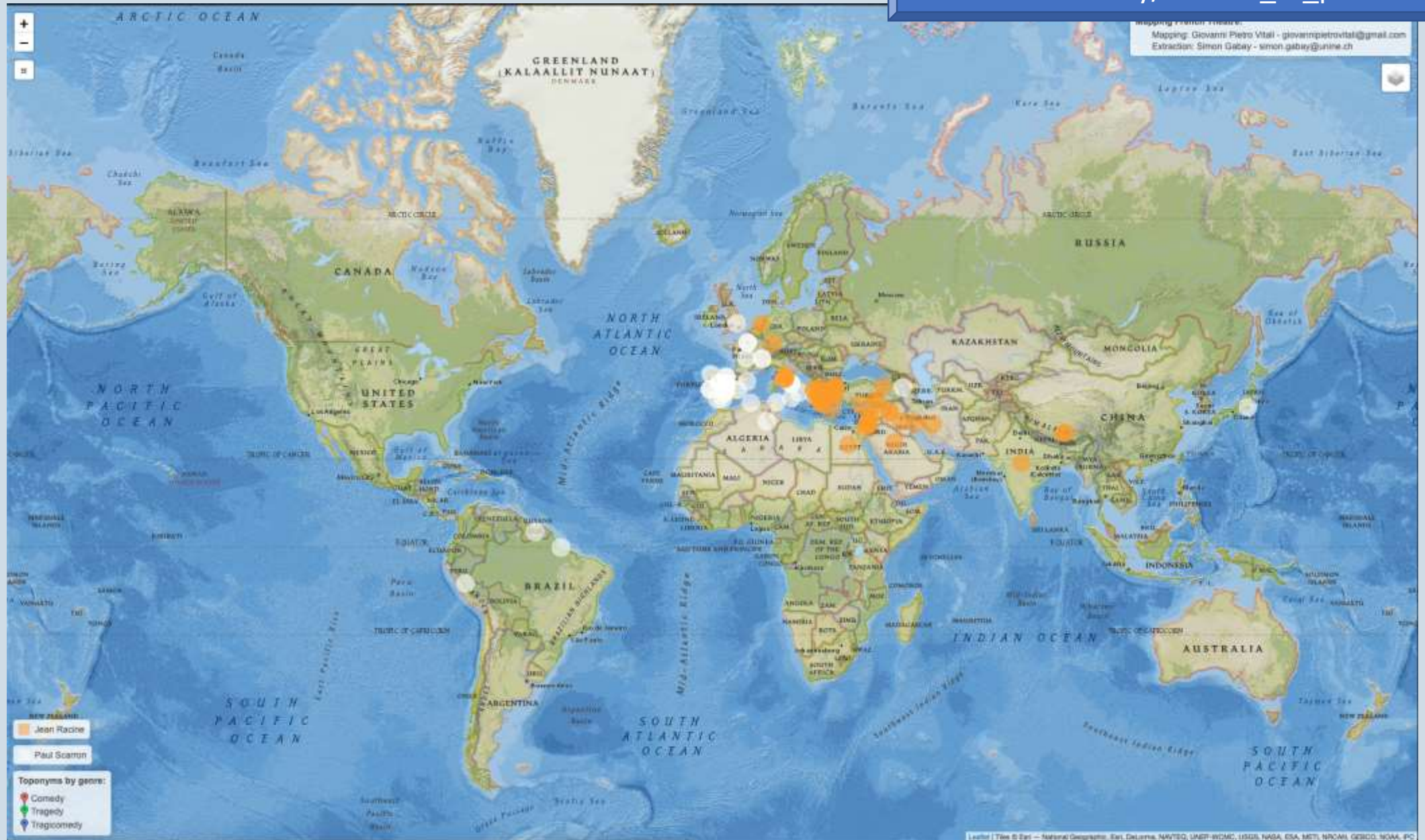


Figure – Comedies in red, tragedies in green, tragicomedies in blue



Style 1

Figure – Molière (pink) and Du Ryer (purple)



Style 2

Figure – Scarron (white) and Racine (orange)


```
# for details, see https://review.docs.microsoft.com/en-us/visualstudio/rva/sql-server
# Test code
library(RODBC)
channel <- odbcDriverConnect(dbConnection)
InputDataSet <- sqlQuery(channel, iconv(paste(readlines(
  "c:/proj/rproject1/rproject1/storedprocedure.query.sql",
  encoding = 'UTF-8', warn = FALSE), collapse = '\n'), from = 'UTF-8',
  to = 'ASCII', sub = ''))
odbcClose(channel)
OutputDataSet <- InputDataSet
```



Bibliography

- G. P. Vitali, *Voices of Dissent. Interdisciplinary Approaches to New Italian Popular and Political Music* (Oxford: Peter Lang: 2020), 364.
- G. P. Vitali, « Rethinking Rome as an Anthology: The Poeti Der Trullo's Street Poetry », *Umanistica Digitale* 3 (2018), 105-136
- G. P. Vitali, « Visualising Second World War Violence through an Atlas of Nazi-Fascist Repression », *DSH - Digital Scholarship in the Humanities* 37/2 (2020), 594-610
- G. P. Vitali, «Un viaje gráfico de crítica política: Análisis geográfico-temático de los diseños de Eneko de Las Heras sobre los periódicos Interviú y 20 Minutos.», *Artl@s Bulletin* 8/3 (2019), 86-110
- M. Reynolds and G. P. Vitali, «Mapping and Reading a World of Translations: Prismatic *Jane Eyre*», *Modern Languages Open*, 1/2 (2021), 23.
- G. P. Vitali et S. Gabay, «A Theatre of Places: Mapping 17th French theatre», *ACM-GIR'19*

bit.ly/voices_of_dissent_book

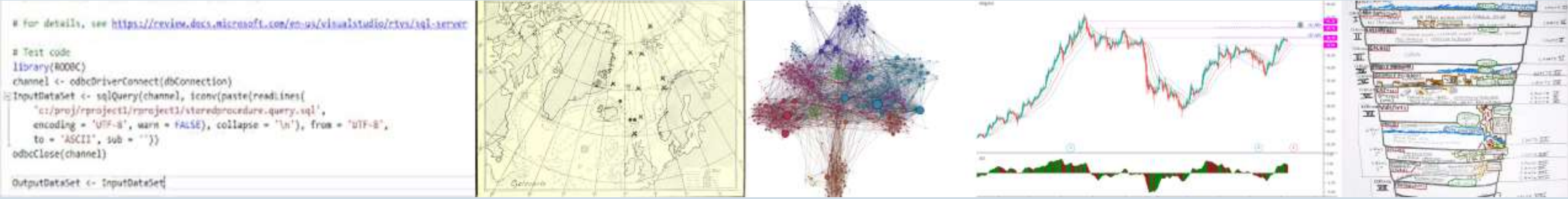
bit.ly/trullo_street_poetry

bit.ly/atlas_article

bit.ly/eneko_article

bit.ly/jane_eyre_mapping_article

bit.ly/theatre_of_places



Digital Humanities _end

Giovanni Pietro Vitali

giovannipietrovitali@gmail.com | giovannni.vitali@uvsq.fr

<https://github.com/digitalkoine> | <https://ucc-ie.academia.edu/GiovanniPietroVitali>



Université de Versailles Saint Quentin en Yvelines
Université Paris-Saclay