

SunoikisisDC Summer 2022 Session 9

Natural Language Processing (NLP) for Historical Texts

Ludovic Moncla (INSA)
Katie McDonough (Turing)



Encyclopedic Geographical Discourse
Writing about Geography in France from the Enlightenment to the Age of Wikipedia

L'équipe

- [Denis Vigier](#), Université Lyon 2, ICAR UMR 5191 - coordinateur scientifique
- [Ludovic Moncla](#), INSA Lyon, LIRIS UMR 5205 - coordinateur scientifique
- [Thierry Joliveau](#), Université de Saint-Etienne EVS UMR 5600
- [Katherine McDonough](#), The Alan Turing Institute, London UK
- [Benoît Crabbé](#), Université Paris Diderot, LLF UMR 7110
- [Achille Falaise](#), CNRS, LLF UMR 7110
- [Olivier Kraif](#), Université Grenoble Alpes, LIDILEM

Four French Encyclopedias:

1. Diderot and d'Alembert's *Encyclopédie* (1751-72)
2. *La Grande Encyclopédie* (1885-1902)
3. *Encyclopédie Universalis* (2018)
4. *Wikipedia* (2018)

Diderot and d'Alembert's Encyclopédie

17 volumes of text articles published between 1751-1765

- 21% of total articles are classified as “Geography”

Authorship:

- Volume 1 & 2 : Diderot writes the most Geography articles
- Unsigned articles become the majority from volume 2 until volume 5 (1753-55)
- The chevalier de Jaucourt contributes articles from volume 3, but output increases significantly from volume 6 (1756),
- Jaucourt is the principal author of Geography articles from volume 7 until the end (1765)

Working with Humanities Data

Different (e.g. historical, multilingual, uncommon) data matters

LISSUS, (Géog. anc.) Ce nom, dans la géographie des anciens, désigne, 1°. une ville d'Illyrie en Dalmatie, sur les frontieres de la Macédoine, avec une citadelle qu'on appelloit acrolissus. Pline ajoute que c'étoit une colonie de citoyens romains, cent mille pas d'Epidaure. 2°. Lissus étoit un lieu de l'île de Crete, sur la côte méridionale, au couchant de Tarba. 3°. Lissus étoit cette riviere de Thrace qui fut tarie par l'armée de Xerxès, laquelle elle ne put suffire. Elle couloit entre les villes de Mésembria & de Stryma.

<https://artflsrv03.uchicago.edu/philologic4/encyclopedie1117/navigate/9/2770/>

≠

Lezhë

From Wikipedia, the free encyclopedia

This article is about the city in Albania. For other uses, see [Lezhë \(disambiguation\)](#).

Lezhë or **Lezha** (Albanian pronunciation: [lɛʒə]; Latin: *Lissus*) is a city in the Republic of Albania and the capital of the eponymous county and municipality.

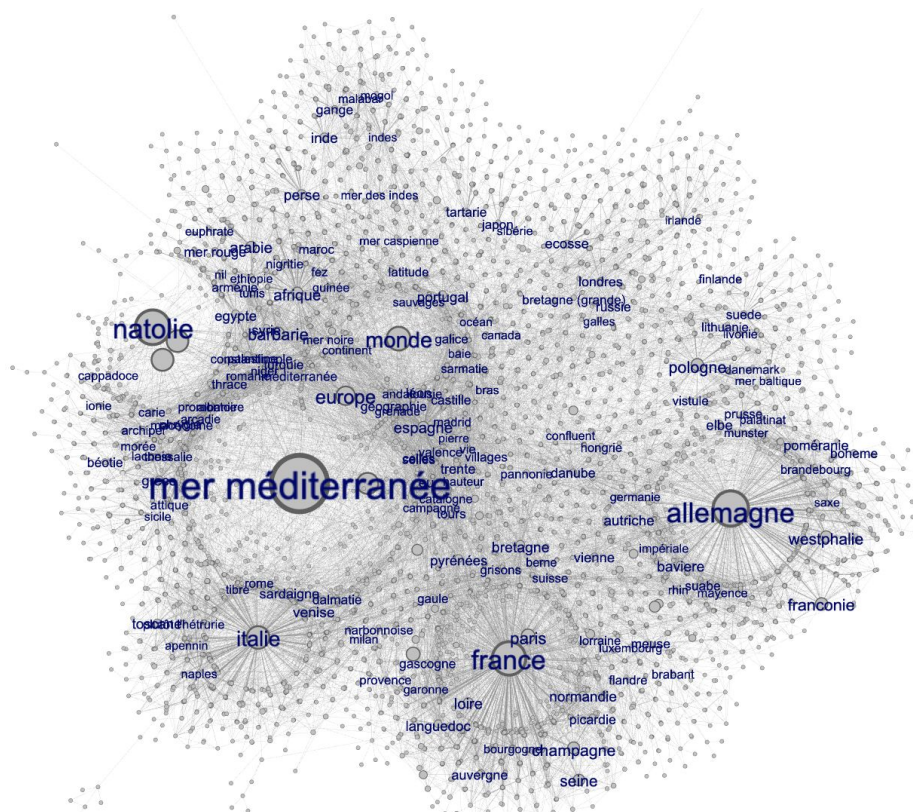
One of the main strongholds of the Labeatai,^[4] the earliest of the fortification walls of the city are of typical Illyrian construction and are dated to the late 4th century BC.^[5] The city was one of the main centres of the Illyrian kingdom.^[6] During the conflicts with Macedon it was captured by Philip V becoming the Macedonian outlet to the Adriatic Sea.^[7] The town was later recovered by the Illyrians. It was subjected to Rome after the Roman-Illyrian wars and the fall of Gentius' realm.^[6]

The total population is 65,633 (2011 census),^[8] in a total area of 514.97 square kilometres (199 square miles).^[9] The population of the former municipality at the 2011 census was 15,510.^[8]

Exploratory Data Analysis

Discovery happens when we look at data in new ways

Transforming *Encyclopédie* Spatial Named Entities into a graph → Identifying new patterns in Enlightenment geography writing



The *Encyclopédie* and the Classical Past

Geography articles describe:

- Celebrated places of antiquity, especially those discussed in well-known classical texts
- Biblical places
- Famous people

Their content includes debates about:

- Toponym spelling in different languages
- The location of places as referenced by historical and modern texts and maps