



CENTRE D'ÉTUDES
SUPÉRIEURES DE LA
Renaissance

Séminaire *Humanités Numériques* 2025-26

Nouveau séminaire
Humanités Numériques Tourangelles



Mercredi 26 février 2025

17h-19h - Tours, CESR - Salle Rapin
et en visioconférence

#7 Yes but... Can Large Language Models Identify Entities in Historical Documents?

Carlos-Emiliano González-Gallardo - CESR/LIFAT

The efficacy of large language models (LLMs) has greatly impacted the field of natural language processing, achieving state-of-the-art performance across various tasks, including named entity recognition (NER) for contemporary texts. However, the use of LLMs for NER in historical collections, such as newspapers and classical commentaries, remains underexplored. This gap presents significant challenges for Digital Humanities research, as historical texts often suffer from noise due to suboptimal storage conditions, errors in optical character recognition, and variations in spelling. During this talk, I will share findings and insights from an empirical evaluation that compares different Instruct variants of both closed and open models. This study aims to improve the understanding and application of NER in historical collections and its relevance in digital libraries. To achieve this, we employed prompt engineering through both deductive (guidelines provided) and inductive (guidelines absent) methodologies, using publicly available historical collections in English, French, and German, along with code-switching in Ancient Greek.

Carlos-Emiliano González-Gallardo is an associate professor of digital humanities at the CESR and the LIFAT at the University of Tours, where he actively participates in the PRIMA Project. He holds a Ph.D. in computer science and natural language processing (NLP) from the University of Avignon, specializing in automatic multimedia and multilingual summarization. He completed a post-doctoral fellowship at the University of La Rochelle, where his research concentrated on information extraction and layout analysis from ancient European press within the framework of the H2020 NewsEye European project.

Plus d'infos sur
cesr.cnrs.fr



Centre d'études supérieures de la Renaissance
59, rue Néricault-Destouches BP 12050 -37020 Tours Cedex 1
Contact - Elena Pierazzo / elena.pierazzo@univ-tours.fr



Grant n° 101142242