Comparing Named Entity Recognition in Classical and Modern Languages: Insights from Plutarch's *Life of Alexander*

This study investigates Named Entity Recognition (NER) in Plutarch's *Life of Alexander*, comparing results on the ancient Greek text processed using UGARIT Flair NER (Yousef, Palladino & Jänicke, 2022) with an English translation analyzed using NLTK's NER module (Bird et al., 2009). By applying automatic annotation methods to both versions, I evaluate the extent to which modern NER tools can accurately identify historical figures, locations, and organizations across languages, while also assessing the challenges posed by linguistic and translational differences.

The research addresses the following key questions:

- How well does UGARIT Flair NER perform on ancient Greek text compared to NLTK's NER on its English counterpart?

- What challenges arise when applying automatic semantic annotation to historical and classical texts?

- How do multilingual annotation practices influence entity recognition accuracy?

To analyze the texts, the Greek version was preprocessed with sentence segmentation and tagged using UGARIT Flair's transformer-based model, while the English translation was tokenized, tagged, and processed using NLTK's named entity chunking. The results reveal significant disparities in entity classification, particularly concerning mythological names and historical locations. UGARIT Flair demonstrates higher precision in ancient Greek, whereas NLTK struggles with ambiguous entities in translation.

For instance, UGARIT Flair correctly identified Ἀλέξανδρος (Alexander), Φίλιππος (Philip), and Ἀριστοτέλης (Aristotle) as PERSON entities, while NLTK misclassified them as geopolitical entities (GPE), especially in contexts where they were referenced in military and administrative settings. A striking example occurs in the sentence: "*Philonicus the Thessalian brought the horse Bucephalas to Philip, offering to sell him for thirteen talents; but when they went into the field to try him, they found him so very vicious and unmanageable, that he reared up when they endeavored to mount him, and would not so much as endure the voice of any of Philip's attendants.*" Here, NLTK incorrectly tags "Philip" as GPE, reflecting how modern NER models, trained on contemporary corpora, may conflate historical figures with place names.

Similarly, Ἔφεσος (Ephesus) was accurately recognized as a LOCATION in UGARIT Flair, while NLTK erroneously tagged "Ephesus" as an ORGANIZATION, as seen in the following example: "*Alexander was born the sixth of Hecatombæon, which month the Macedonians call Lous, the same day that the temple of Diana at Ephesus was burnt; which Hegesias of Magnesia makes the occasion of a conceit, frigid enough to have stopped the conflagration.*" This misclassification suggests that NLTK may interpret historical place names as institutional entities when appearing in religious or administrative contexts.

The findings in this study highlight the limitations of general-purpose NER models when applied to historical texts in translation and emphasize the necessity of specialized models for ancient languages, as suggested in Palladino & Yousef (2023), advocating for further refinement of

annotation frameworks to enhance accuracy and cross-linguistic interoperability in classical text processing.

Bibliography

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media.

Johnson, K. P., Burns, P. J., Stewart, J., Cook, T., Besnier, C., & Mattingly, W. J. (2021). The Classical Language Toolkit: An NLP framework for pre-modern languages. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing: System demonstrations* (pp. 20-29).

Palladino, C., & Yousef, T. (2024). Development of robust NER Models and Named Entity Tagsets for Ancient Greek. In R. Sprugnoli, & M. Passarotti (Eds.), 3rd Workshop on Language Technologies for Historical and Ancient Languages, LT4HALA 2024 at LREC-COLING 2024 - Workshop Proceedings (pp. 89–97). European Language Resources Association (ELRA). https://aclanthology.org/2024.lt4hala-1.11.

Yousef, T., Palladino, C., & Jänicke, S. (2022). Transformer-Based Named Entity Recognition for Ancient Greek. DOI:10.13140/RG.2.2.34846.61761.