Transformer-Based Named Entity Recognition for Ancient Greek

Yousef, Tariq
info@tariq-yousef.com
Leipzig University

Palladino, Chiara
chiara.palladino@furman.edu
Furman University

Jänicke, Stefan
stjaenicke@imada.sdu.dk
University of Southern Denmark

Introduction

The identification and classification of names in ancient texts, especially Classical ones like Homer's Iliad , is an essential task to support further text processing, but also simple reading facilitation and the creation of reading environments (Blackwell / Crane 2009). Typically, readers of Ancient Greek or Syriac face the challenge of a complex and sometimes obscure language, but also of very distant cultural references, which go back to events and traditions that may not be immediately comprehensible to a non-specialist. Names are a substantial part of this challenge. Texts like the Bible or the Iliad contain hundreds, if not thousands, of names, many of which appear only once or twice, and even language specialists are sometimes unable to recognize certain places, people, or events immediately.

An automatic Named Entity Recognition (NER) model for this category of texts can enormously facilitate the task of a reader, by extracting, classifying, and linking ancient names to available resources (Kemp 2021) and by using those names to design applications that encourage different approaches to textual and historical exploration (Barker / Terras 2016).

Nevertheless, in the domain of ancient languages, NER is a complicated task. The lack of adequate infrastructure and annotated data is the main obstacle to developing reliable NER pipelines, as many of these languages have scarce (or none at all) services for annotation, lemmatization, morphosyntactic analysis, and named entity classification. Some resources of this kind have been developed, for example, for Ancient Greek and Latin (Burns 2019). However, the lack of annotated texts in the original languages still makes Named Entity Recognition challenging task for many types of texts (Erdmann et al. 2016).

This paper presents our work on training two automatic NER models for ancient Greek using transformer-based models. The models classify the entities into three categories, namely, Person, Location, and Miscellaneous and achieved promising results on test and evaluation datasets. The models are available on Hugging Face [1,2].

Related Work

(Blackwell / Crane 2009) and (Babeu et al. 2007) emphasize the importance of processing historical texts for names in the broader field of digital infrastructures for reading and annotation, such as the Perseus Digital Library [3] and the Scaife Viewer [4]. The Classical Language Toolkit (CLTK) is the largest Python library to perform NLP tasks on ancient languages, including NER (Johnson et al. 2021): the lack of adequately annotated datasets for most corpora, however, is a fundamental hindrance to the high performance of this task (Palladino et al. 2020). Other efforts have been made starting from large annotated datasets of specific sources, using semantic annotation platforms and Machine Learning (Berti 2019).

Data

The data used for model training are collected from two resources, *The Deipnosophists* of Athenaeus of Naucratis [5] , which have been annotated semi-automatically in the context of the Digital Athenaeus project [6] using INCEpTION annotation platform [7].

The entities are classified into ten categories (Table 1). However, We reduced them to three main categories, PERson, LOCation, MISCellaneous [8].

Table 1: An overview of the training data set.

	Class	Frequency	3 Classes	Frequency
Odyssey	Person	2.469	PER	2.469
Place	698	LOC	698	
Deipnosophists	Person	12.424	PER	12.424
Place	2.305	LOC	2.305	
Ethnic	3.548	MISC	6.735	
NoClass	2.263	MISC		
Group	681	MISC		
Title	206	MISC		
Festival	20	MISC		
Month	8	MISC		
Language	7	MISC		
Constellation	2	MISC		
Total				24.631

Moreover, we used texts from *Odyssey* annotated manually by *Chiara Palladino*. The dataset contains 3.167 entities as PERsons and LOCations only. The main limitation of the dataset is that almost all entities are single-token entities which will affect the model performance in detecting multiple-tokens entities.

We also created an evaluation dataset to evaluate the models' performance. The dataset consists of 50 paragraphs randomly selected from the Book 5 of the *Histories* of Herodotus. The texts were annotated manually using Recogito [9] and contain 351 entities.

Training & Results

The recent advances in neural networks and language modelling using transformers (Vaswani et al. 2017), and the availability of large pretrained language models allowed researchers to achieve state-of-the-art performance on various NLP tasks, including NER.

In the experiments, we used two transformer-based ancient Greek language models, namely, *Ancient Greek Alignment*[10] (GRC_A) (Yousef et al. 2022a, Yousef et al. 2022b), an XLM-R-based multilingual model fine-tuned on the translation alignment task, and *Ancient Greek BERT*[11] (GRC_B) (Pranaydeep et al. 2021), a BERT-based monolingual model fine-tuned on POS tagging and morphological analysis tasls. We used *Flair* framework [12] (Akbik et al. 2019) to train and fine-tune the models using 75% of the data for training, 12.5% for testing, and 12.5% as development dataset. We trained the models 10 epochs and used Conditional Random Field (CRF) for prediction.

The training results presented in table 2 show that GRC_A outperforms GRC_B regarding the F1-score on all classes. Overall, both models achieved over 82% F1 on all classes. However, their performance on PERson is significantly higher than LOCation and MISC because the training dataset contains more PERson than other classes.

Table 2: Training and evaluation results.

		GRC_A	GRC_B				
		Precision	Recall	F1-score	Precision	Recall	F1-score
Training	PER	93.39%	96.33%	94.84%	91.24%	94.45%	92.82%
MISC	84.69%	92.50%	88.42%	80.92%	83.17%	82.03%	
LOC	89.55%	77.32%	82.99%	86.86%	78.35%	82.38%	
Evaluation	PER	90.48%	91.94%	91.20%	96.43%	87.10%	91.53%
MISC	89.29%	94.34%	91.74%	92.00%	86.79%	89.32%	
LOC	82.69%	65.15%	72.88%	80.00%	84.85%	82.35%	

The evaluation on *Histories*'s texts showed that GRC_B achieved higher precision and lower recall on PERsons and MISCs, but significantly higher recall and lower precision on LOCations. Nevertheless, there is still room for performance improvement by fine-tuning training hyperparameters and expanding the training dataset by manually annotating more texts or generating the annotations automatically with our model and then correcting them manually.

Qualitative Evaluation

Overall, the two models we tested performed with similar accuracy, but with some individual differences. GRC_B miscategorized 25 entities, and omitted 22 entities, while incorrectly extracting 3 non-entities. GRC_A miscategorized 34 entities, but only omitted 5 entities and never extracted words incorrectly. GRC_A performed considerably worse on location names: out of 70 manually annotated entities 23 were mislabelled, mostly as MISC (GRC_B only mislabelled 10 in total). GRC_B performed worse on the MISC category, where 18 entities out of 156 were not recognized as such (GRC_A only made 6 mistakes with this category). Personal names were equally difficult for both, where GRC_B mislabelled 9 and omitted 7 out of 125 manually annotated entities, and GRC_A mislabelled 8 and omitted 2. Overall, GRC_A performs considerably worse in labelling location names.

The label MISC was almost exclusively used to classify people groups and, in particular, ethnonyms. This leads to some issues when trying to create more fine-grained classifications: disambiguation in complex texts like the Histories is not limited to simple entity labelling but is often depending on context (for example, it is customary in Ancient Greek historiography to indicate a location through the name of its inhabitants). Another relevant issue is the lack of performance on multiple-token entities, which depends on the specific training data that does not include such instances. Both these aspects require more context-specific analysis of the data.

Notes

1. https://huggingface.co/UGARIT/flair_grc_multi_ner
2. https://huggingface.co/UGARIT/flair_grc_bert_ner
3. http://www.perseus.tufts.edu/hopper/
4. https://scaife.perseus.org/
5. https://scaife.perseus.org/reader/urn:cts:greekLit:tlg0008.tlg001.perseus-grc4
6. https://www.digitalathenaeus.org/
7. https://inception-project.github.io/
8. We aimed to follow the convention used in the STOA since most pre-trained NER models use four classes (Persons, Locations, Organizations, and Misc). Since Organizations are not part of the training data, we used only three classes. However, the pipeline can be updated for more fine-grained classes, such as Ethnic, since the training dataset has a reasonable amount of examples.
9. https://recogito.pelagios.org/
10. https://huggingface.co/UGARIT/grc-alignment
11. https://huggingface.co/pranaydeeps/Ancient-Greek-BERT
12. https://github.com/flairNLP/

Bibliography

Akbik, Alan / Bergmann, Tanja / Blythe, Duncan / Rasul, Kashif / Schweter, Stefan / Vollgraf, Roland (2019): "FLAIR: An easy-to-use framework for state-of-the-art NLP." In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations), pp. 54-59.

Babeu, Alison / Bamman, David / Crane, Gregory / Kummer, Robert / Weaver, Gabriel (2007): "Named entity identification and cyberinfrastructure." In ECDL, pp. 259-270.

Barker, Elton / Terras, Melissa (2016): "Greek literature, the digital humanities, and the shifting technologies of reading.".

Berti, Monica (2019): “Named entity annotation for ancient greek with inception”. In Kiril Simov and Maria Eskevich, editors, CLARIN Annual Conference Proceedings, pages 1–4, Leipzig,. CLARIN.

Blackwell, Christopher / Crane, Gregory (2009): "Conclusion: Cyberinfrastructure, the Scaife Digital Library and classics in a digital age." Digital Humanities Quarterly 3, no. 1.

Burns, Patrick J. (2019): "Building a text analysis pipeline for classical languages." Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution 10 : 159-176.

Erdmann, Alexander / Brown, Christopher / Joseph, Brian / Janse, Mark / Ajaka, Petra, Elsner, Micha / de Marneffe, Marie-Catherine (2016): "Challenges and solutions for Latin named entity recognition." In Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH), pp. 85-93.

Johnson, Kyle P. / Burns, Patrick J. / Stewart ,John / Cook, Todd / Besnier, Clément / William JB Mattingly (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.

Kemp, Josh (2021): “Beyond translation: Building better greek scholars”, Pelagios Blog.

Palladino, Chiara / Karimi, Farimah / Mathiak, Brigitte (2020): "Ner on ancient greek with minimal annotation." In https:// dh2020. adho. org/. DH2020.

Singh, Pranaydeep / Rutten, Gorik / Lefever, Els (2021): "A pilot study for BERT language modelling and morphological analysis for ancient and medieval Greek." In 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, co-located with EMNLP 2021, pp. 128-137. Association for Computational Linguistics.

Vaswani, Ashish / Shazeer, Noam / Parmar, Niki / Uszkoreit, Jakob / Jones, Llion / Gomez, Aidan N. / Kaiser, Łukasz / Po-losukhin, Illia (2017): "Attention is all you need." Advances in neural information processing systems 30.

Yousef, Tariq / Palladino, Chiara / Shamsian, Farnoosh / d’Orange Ferreira, Anise / Ferreira dos Reis, Michel (2022). "An automatic model and Gold Standard for translation alignment of Ancient Greek." In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 5894-5905.

Yousef, Tariq / Palladino, Chiara / Wright, David J. / Berti, Monica (2022). "Automatic Translation Alignment for Ancient Greek and Latin." In Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages, pp. 101-107.