Computational Approaches to the Study of Ancient Greek Conflict

Natalie Ayers
Department of Government, Harvard University | natalieayers@g.harvard.edu

Transformer Models: Conflict Words

Application of transformer language model to Ancient Greek texts referencing conflict

- Fine-tune Ancient Greek BERT-based model on conflict-related lines from canonical Greek literature to obtain **contextualized word embeddings**.[1,2] Implement in Python.
  - Previous uses of transformers with Ancient Greek text have focused on Part-of-Speech tagging; this project assesses the viability of transformers for Ancient Greek semantic analysis
- Visualize embeddings of words in UMAP, highlighting areas of divergence between usage of the same or similar words for further study[3]

Text to Words - Example

οἵ μοι ἐφώρμησαν πόλεμον πολύδακρυν Ἀχαιῶν → [ἐφωρμάω, πόλεμος, πολύδακρος, Ἀχαιός]

"..who roused against me the tearful war of the Achaeans" (Homer Il. 3.146)

Words longer than 3 letters converted to standard form (nouns: nom. sg.; verbs: 1st sg. pres.), then tokenized

Introduction

This work brings together three strands of existing scholarship:

I. Computational and quantitative methods, which are continually improving in capability, accessibility, and applicability
II. Modern conflict research, which has thought extensively about the nature of warfare in the post-industrial world
III. Ancient Greek studies, concerning one of the most widely (qualitatively) studied Western civilizations

Two main areas of focus for the initial phase of this research:

1. Comparison of ancient and modern warfare using the tools of modern conflict scholarship, including computational methods
   - Ancient warfare should be an additional source of data for the overall study of human conflict
2. Consideration of a "democratic peace" in Ancient Greece, and what similarities and differences this holds to our modern system

Here, I present two methodological approaches applied to Ancient Greek data:

Analysis of contextual word embeddings from transformer model applied to Ancient Greek conflict-related texts

Analysis of Ancient Greek city-state conflict behavior using Bayesian network model

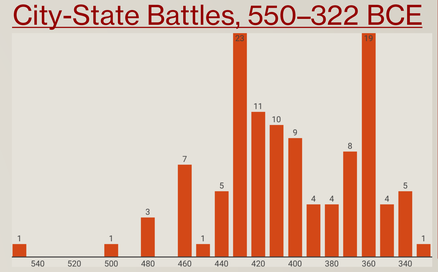
Locations of city-state battles from 550 – 322 BCE (red X) and locations of city-states which participated in these battles (blue temple).

Network Analysis: City-State Battles

Application of Mixed-Membership Stochastic Blockmodel (MMSBM) Regression to City-State Conflicts

- Bayesian modeling approach for network of city-states in conflict, treating **city-states** as **nodes** in a graph with presence / absence of **conflict** in a year as binary **edges**. Implement in R with NetMix.[4,5]
- Analyze conflicts between Greek city-states from 550 – 322 BCE, including city-state Fame, Size, Walls, Democracy (Regime), and more[6]

City-State Battles, 550–322 BCE



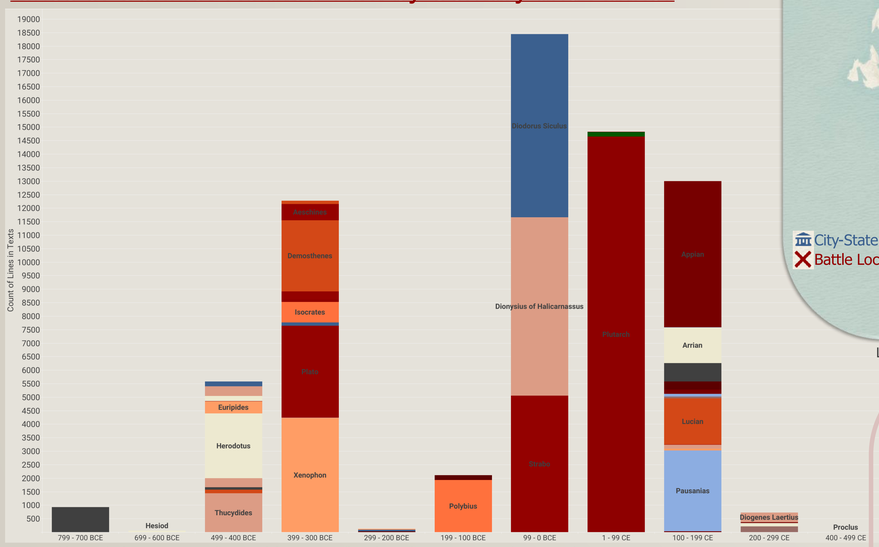
# Battles	104
# Participating City-States	106
# City-States with some democracy	59
# City-States with some fortification	95

Top 5 City-State Members per Group

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Naupaktos (0.253)*	Delphi (0.831)	Thisbae (0.19)*	Abdera (0.61)*	Euboia (0.38)*	Ialysos (0.43)
Argos – Akarnania (0.251)*	Athens (0.625)*	Pisa (0.18)	Samos (0.61)*	Opous (0.29)	Tenedos (0.39)
Halieis (0.249)*	Sparta (0.597)	Orchemenos (0.17)*	Kyzikos (0.61)*	Galeria (0.26)	Thera (0.38)
Oropos (0.246)*	Agyrion (0.575)*	Thespiae (0.17)*	Ephesos (0.60)*	Pharsalos (0.26)*	Pisa (0.36)
Anaktorion (0.246)*	Mantinea (0.517)*	Tegea (0.17)*	Argos - Argolid (0.60)*	Spartolos (0.17)*	Lebedeia (0.30)

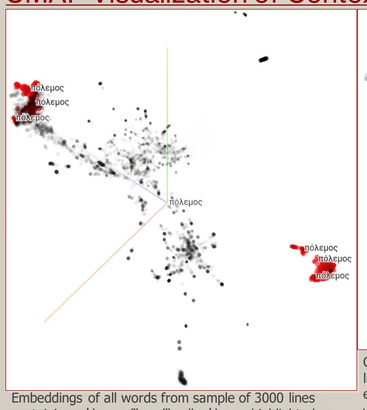
Blue: evidence of a democratic regime; Red: no evidence of democracy; Grey: regime uncertain.
* indicates presence of Walls.

Lines of Conflict-Related Text by Century and Author

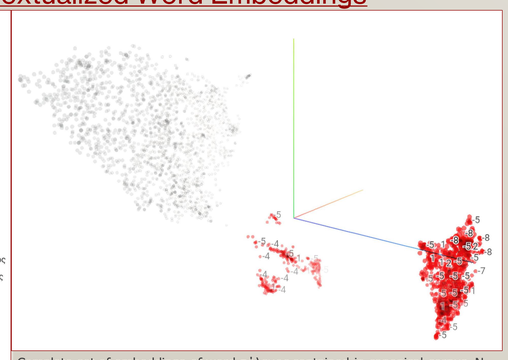


Count of lines which contain a conflict word (101 manually selected) in Ancient Greek canonical corpus. N authors: 59; N texts: 522.

UMAP Visualization of Contextualized Word Embeddings



Embeddings of all words from sample of 3000 lines containing πόλεμος ("war"); all πόλεμος highlighted.



Complete set of embeddings of word πόλεμος contained in canonical corpus, N lines of text: 6,017. Highlighted labels indicate century (-5 is 5 BCE, 1 is 1 CE, etc).

Visualized with Tensorflow's Embedding Projector: https://projector.tensorflow.org/

Results and Discussion

Text Analysis of Conflict Words:

The word πόλεμος ("war, battle") was selected as a test case for analysis of contextualized word embeddings.

- Majority of πόλεμος embeddings clustered in single space, but 2 smaller clusters indicate potentially different semantic usage (left UMAP)
- When πόλεμος embeddings viewed by century (right UMAP), the clusters indicate πόλεμος meaning differences may relate to certain time periods

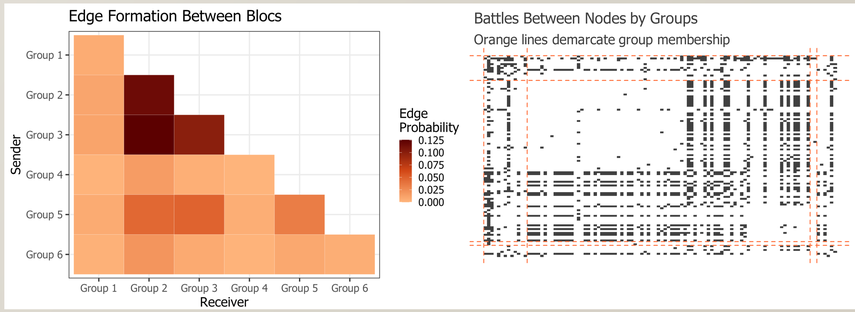
Network Analysis of City-State Conflicts:

A 6-group model offered the strongest analytic potential, producing the results on the right.

- Top members of Group 4, the largest by membership and least likely to experience conflict, all had some form of democracy
- Standard errors on coefficients are too high to make any claims, but Walls and Size are generally associated with less conflict, while Fame and some democracy are associated with increased conflict

Future Directions:

- Continued analysis of divergent πόλεμος usages identified with embeddings, as well as similar analysis for other conflict terms
- Enhance city-state dataset to incorporate Walls, Size, and Regime change over time, where possible, to use temporally dynamic version of MMSBM[5]



Estimated Monadic Predictor Coefficients on Conflict

Predictor	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Intercept	11.3957 (49.2)	11.1652 (49.2)	10.1994 (49.2)	10.7486 (49.2)	10.4603 (49.2)	11.4165 (49.2)
Fame	0.9822 (49.3)	1.2400 (49.3)	0.9504 (49.3)	1.0312 (49.3)	1.0492 (49.3)	1.0341 (49.3)
Size	-1.2522 (49.3)	-1.6139 (49.3)	-1.0370 (49.3)	-1.1976 (49.3)	-1.4230 (49.3)	-1.1711 (49.3)
Walls	-3.2491 (41.2)	-3.6826 (41.2)	-3.5754 (41.2)	-2.5185 (41.2)	-2.7600 (41.2)	-4.3491 (41.2)
Regime	2.0932 (49.3)	2.1691 (49.3)	2.1366 (49.3)	2.7186 (49.3)	2.3460 (49.3)	2.5089 (49.3)

N nodes: 106; N dyad-year pairs; 2,548,770; N years: 319; N dyad-years with conflict: 2,172; Lower bound at convergence: -228,761.7

1. Crane, Gregory R., ed. (2023). *Perseus Digital Library*
2. https://huggingface.co/Sonnenblume/bert-base-uncased-ancient-greek-v4
3. McInnes, Leland, John Healy, and James Melville (Sept. 17, 2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.*
4. Olivella, Santiago, Adeline Lo, et al. (Nov. 16, 2022). *NetMix: Dynamic Mixed-Membership Network Regression Model. V.* 0.2.0.1.
5. Olivella, Santiago, Tyler Pratt, and Kosuke Imai (July 3, 2022). "Dynamic Stochastic Blockmodel Regression for Network Data: Application to International Militarized Conflicts". *Journal of the American Statistical Association.*
6. City-state data from POLIS v5, Josiah Ober et al. (2017). Battle data from Jordan Adamson (unpublished), supplemented with Montagu (2000).