IA for Monograms: Interpreting Byzantine Monograms thanks to Artificial Intelligence Approaches

1 Context

The taste for monograms is a feature of the Mediterranean material heritage in the Roman and Byzantine eras. Monograms take less space than a complete name and can be found on diverse objects or architectural elements. They are like puzzles that elegantly combine letters with design and aesthetics. Monograms are mostly seals and markers (signacula) but also on terracotta pottery, amphora caps, signet rings, and architectural elements such as capitals (see Figure 1). Monograms combine letters of a name or a function, or both a name and a title. Reading monograms requires several rules: every letter of a name must be included; some may be used twice or combined with other letters. Therefore, there are different reading possibilities.

Figure 1. Byzantine monograms in different supports.

The most accessible and numerous documents bearing monograms are Byzantine seals (see Figure 2). They enclose essential knowledge about the Byzantine administration, aristocracy, and religion. Their importance derives from the scarcity of surviving Byzantine documents and the large number of existing seals. However, since Byzantine seals are mostly made of lead, they suffer from corrosion and are often damaged. The historians' interpretation work is challenging because some seals have been crushed or shattered, making their inscriptions difficult or impossible to read.

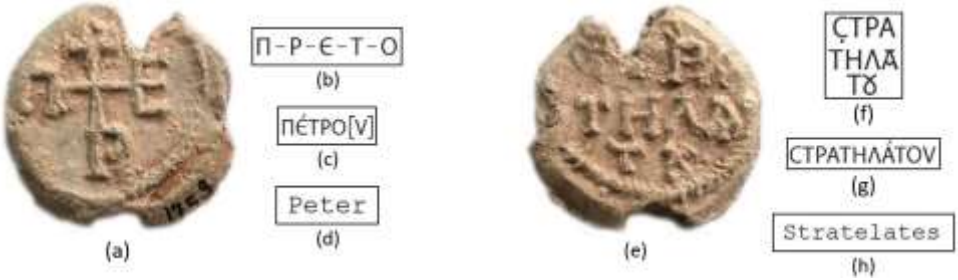

Figure 2. Obverse (a) and reverse (b) of a Byzantine Seal with their plain transcription in Byzantine Greek (b and f), the restored text (c and g) and the English transcription (d and h) [Cheynet 2019, *Peter stratèlates*, Tatish 1753, 5.116, p. 269]

2 Scientific objective

This Digital Humanities project aims to study Byzantine monograms using an innovative IA approach. This project will use new textual and iconographic sources previously unexplored and never linked or compared to read monograms. We propose to combine knowledge engineering techniques, combinatorial optimization, natural language processing (NLP), and deep learning:

1. First, we will **recognize the characters** in the monograms (for example, in Figure 1.b, to detect the characters "ΠΡΕΤΟ") using deep learning and make a ***linguistic study of onomastics*** (compilation of lists of names from published material and prosopographic directories);
2. Secondly, we will develop an algorithm that **suggests combinations** of those characters to **obtain possible names** (for example, in Figure 1.c, identify "Πέτρο[υ]" as a possible name), using AI and NLP techniques to process the data from epigraphic analysis.

The first challenge is to incorporate different data sources lacking explicit contextual linking. Because of the complexity of the subject — the variety of sources, languages, and artistic contexts — we would like to implement knowledge graphs to model Byzantine sigillography. ***Knowledge graphs*** significantly preserve cultural heritage and model human expert knowledge. Since knowledge graph embedding techniques have not been exploited in the sigillography domain, we will explore different approaches using purely knowledge graph embedding, visual embedding, or combined embedding to calculate the similarity between items. We will use ***Natural Language Processing*** (NLP) techniques to extract references to onomastic elements in texts: multilingual for medieval languages and multiple versions of descriptions are a novel challenge.

Deep learning approaches are widely used for many computer vision tasks, such as object recognition, semantic segmentation, or instance segmentation. Results are impressive and often rely on large training databases. However, instead of training a neural network from scratch, the best practice now is to use a pre-trained model and only retrain the last layers. This thesis will benefit from the first results obtained in the BHAI project concerning character recognition in Byzantine seals.

3 Justification of the scientific approach

Monograms are often difficult to interpret because there seems to be no rule for the composition and arrangement of the letters. Furthermore, their interpretation is even more challenging due to the deteriorating condition of Byzantine seals. Hence, the decryption of monograms could greatly benefit from the computing possibilities. From a computational point of view, monograms constitute an exciting object of study to analyze because they include a certain degree of combinations that must be compared with onomastic lists. Thus, the probability must be considered, leading to an analysis of the linguistic characteristics of onomastics (regions, periods).

We will work on photographic collections, which represent thousands of objects. Volume II of the Byzantine Lead Seals, by G. Zacos and A. Veglery, published in 1971 alone, contains 1278 monogram seals. It is possible to use images of seals from the books published by Jean-Claude Cheynet. For the markers, the photographic collection of B. Caseau will be used. She has worked on markers for many years. The number of signacula amounts to several thousand. However, not all of them have monograms. As for the hallmarks on the silver, the monograms listed are around 200. There are enough objects to test and validate hypotheses and computer tools. The use of different types of monograms inscribed in different materials (from lead in seals to stone in capitals) will allow us to have a more complete overview of the uses and motivations behind these enigmatic graphical compositions.

4 Project's Suitability to the Initiative

This project offers new perspectives on material heritage recently reevaluated by historical research. A multidisciplinary approach is essential for the study of Byzantine monograms. An inclusive approach, merging Byzantine expert knowledge and powerful artificial intelligence techniques, can read and interpret the still unresolved challenges of Byzantine monograms. This thesis has the support of Byzantine experts of the ANR BHAI project, who have transcribed and translated most of the seals that will be studied into French. However, due to their complexity and diversity, the monograms are not studied in the ANR BHAI project. The ultimate purpose of this new project is to create a lasting synergy to establish lasting multidisciplinary cooperation around Byzantine seals.

5 Scientific team

The scientific team comprises four researchers: two historians and two computer scientists. Isabelle Bloch, Victoria Eyharabide, and Beatrice Caseau will supervise the future PhD student. As an artificial intelligence expert, I. Bloch specializes in hybrid IA methods. V. Eyharabide works on digital humanities and IA-based approaches for cultural heritage preservation. As an expert in History, B. Caseau specializes in late antiquity and the Byzantine world. Jean-Claude Cheynet is a world-renowned expert in Byzantine sigillography.