

Digitization of Coptic Heritage

Argyro Kontogianni¹, Theodoros Ganetsos², Nikolaos Zacharis³, Evangelos C. Papakitsos⁴

¹ Department of Industrial Design & Production Engineering, University of West Attica, Egaleo, Greece, email: akontogianni@uniwa.gr

² Department of Industrial Design & Production Engineering, University of West Attica, Egaleo, Greece, email: ganetsos@uniwa.gr

³ Department of Informatics & Computer Engineering, University of West Attica, Egaleo, Greece, email: nzach@uniwa.gr

³ Department of Industrial Design & Production Engineering, University of West Attica, Egaleo, Greece, email: papakitsev@uniwa.gr

Abstract. The Coptic language began in the 1st century AD with the old Coptic and extends to the 16th century AD, creating a continuum of written sources in the Egyptian language of almost 5000 years. Especially in Greece, several museums and institutions preserve known manuscripts and artifacts of this language and art. Besides the large number of Coptic corpora, the lack of interactive computational tools in Greek makes it difficult for Greek scholars to study the language in depth and allows various misinterpretations. In this present work, we will present a software tool for semi-automated Coptic to Greek translation, which will facilitate the scholars and will and will overcome the obstacles of that someone may meet, not only at corpora but also at artifacts. We will also examine the necessity of a double-based data sources for accelerating the results of our search engine and extracts conclusions about the better way that familiar tools should be structured.

Keywords: *Coptic, computer-assisted translation, cultural heritage digitization.*

1. Introduction

1.1 Coptic Language

The Egyptian-Coptic is the final phase of the long-standing evolution of the Ancient Egyptian language, one of the oldest written languages of the world, which appeared since the 33rd century BCE, belonging to the Afro-Asiatic (alias Hamito-Semitic) family of languages. Specifically, the Coptic language is spoken around 300 BC until 1500 AD and is the last phase of the Egyptian language. It is the evolution of the vernacular Egyptian language but written in the Greek alphabet. Coptic displays six distinct dialects and many more secondary sub-dialects. Each dialect consists of a set of texts with separate linguistic features, but also in some cases separate thematic features. Nevertheless, there are common features that place a dialect in specific geographical areas, but not with precise boundaries for each.

In summary, the 6 distinct dialects are the following: i. Sahidic, ii. Bohairic, iii. Fayyumic, iv. Akhmimic, v. Lycopolitan, vi. Oxyrhynchite.

1.2 Coptic Script

The influence of Greek vocabulary in Coptic began with the conquest of Egypt in 332 BC by Alexander the Great and led to the integration of Greek terminology in administrative affairs. So, Coptic in essence is not exactly an evolution or simplification of the previous Egyptian script (hieratic and hieroglyphic), as the Egyptians -at this stage of linguistic depiction- adopted the Greek alphabet. Gradually, colloquial Egyptian begins to be written in Greek characters, and until the 3rd century BC the language is formed more strictly and consolidates. As long as the writing system reflects the phonological and phonetic system of a language, in Coptic with its six basic dialects and numerous sub-dialects we were obliged to identify a common substrate to represent its writing system. Obviously, it is a reduction to a common alphabet of all dialects, a “theoretical” construction but at the same time so useful for scholars of this language, that it could be characterized as “Common Coptic”. Thus, Coptic script is a writing system of 32 characters that respond to 26 distinct sounds.

The oldest Coptic texts date to about the 1st to 2nd century AD, while the Coptic alphabet that reflects the phonological values of the Greek prototype belongs at the latest to the 3rd century AD (Allen, 2013) revealing that the tradition of writing the Egyptian language in alphabetical

characters is at least three centuries older than the first confirmed old Coptic texts.

2. The interaction of Coptic with Greek

Greek language does not influenced Coptic only at alphabetical level. It should be noted that from the earliest years Coptic was associated with Christianity in Egypt, Coptic texts contain a wide range of documents from the later Roman Byzantine and early Islamic periods that constitute an extensive and rich body of original and translated Christian literature of great importance for the early history of Roman Christian monasticism. They are also considered excellent witnesses of great Gnostic, Manichaean, and Hermetic texts (Wilfong, 2018). Coptic is the language in which the Christian religious texts were written by the political monks, it is the language of the gospels, the language of the letters but also the language of all the financial archives and administrative documents. Nowadays, it survives as a liturgical language of the Coptic Orthodox Church. It is therefore a huge cultural heritage that must be preserved and can be studied with greater comfort.

2.1 Coptic Evidence in Greece

Coptic element also exists in modern Greece. Museums and institutions manage to preserve some elements of this more "dead" language and art, that represents an amazing patchwork of cultures, constituting a unique transition from paganism to Christianity and finally to Islam, bringing together the old with the new, so creating an amazing, distinct, and holistic identity. Especially in Greece, known artifacts and manuscripts can be found on display or at archives in four museums (the Byzantine and Christian Museum of Athens, the Benaki Museum at Athens, the Museum of Modern Greek Culture at Athens and the Peloponnesian Folklore Foundation “V. Papantoniou” at Nafplion), the Holy Monastery of Iveron on Mount Athos and the National Library of Greece at Athens. Coptic scripts can be found in various materials of different durability, such as rocks (limestone), ivory, wood, clay, fabric (silk, linen, wool), and papyri, being difficult to be automatically processed because of their nature. Especially for Greek scholars, digital resources are totally absent. (Kontogianni et al., 2021: 33-34).

3. Digitization of Coptic Heritage

3.1 Theoretical Background and Existing Software

In order to develop a software tool for the study of Coptic by scholars of the Humanities three factors were considered: i) the nature of the artifacts to be studied; ii) the features of the particular script; iii) the existing software applications for this purpose. The diversity of nature of the artifacts leads us to focus on a computer-assisted translation, word-by-word, i.e., by inserting a word and getting its translation back in a given language. The translation of single words, one at a time, is often useful for texts that can be inscriptions on artifacts, without syntactic cohesion, fragmentary or corrupted (Kontogianni et al., 2020: 28).

The computational tools for deciphering or translating texts of ancient languages are based on the design of processing software, with the existence and assistance of corresponding lexical databases. Regarding Coptic, there are several digital resources, mainly databases but also Natural Language Processing (NLP) tools, some notable ones developed by the Koptische/Coptic Electronic Language and Literature International Alliance (KELLIA) Project (Schroeder et al., 2019) that indicate the international interest on digital Coptic.

Coptic was originally written in manuscripts without spaces between words (*scriptio continua*). This combined with the agglutinative morphology of Coptic and the frequent existence of diacritics (such as supralinear strokes or circumflexes), punctuation and abbreviations seriously complicates the automated processing. Furthermore, there are multi-word entries from corpora and units smaller than words for productive derivations and incorporation. The existing software tools are focused on corpora, without a particular concern of inscriptions on artifacts that may lack syntactic cohesion, can be fragmentary or corrupted, and without providing a translation into Greek, in spite of their immense cultural and linguistic affinity (Kontogianni et al., 2020: 27-28).

Considering the above issues our tool is based in semi-automated translation and emphasizes the ability of processing inscriptions on artifacts through a simple interactive interface for its usage by scholars of Humanities, having just essential familiarization with computers. It is structured on an existing software tool for ancient languages (Kontogianni, 2014), which has been successfully used for processing Linear-B script (Kontogianni et al., 2017; 2019) and is being also modified for processing Linear-A script (Mavridaki et al., 2020; 2021).

3.2 The Software

The developed software tool consists of three main modules, the Coptic-Greek digital dictionary and the interface (which are visible to the user) and the search-engine.

3.2.1 The Search Engine

The search engine has been implemented in Visual C# computer programming language. It receives a sequence of Coptic characters as input and it returns a matching translation into Greek, along with an accompanied commentary, if any found in the digital dictionary. Otherwise, an appropriate message is displayed.

3.2.2 The Lexical Database

The Coptic-Greek digital dictionary is a lexical database being implemented as a single spreadsheet file. There, the Coptic words are sorted into lists firstly by size, according to the number of their characters, and then alphabetically in each separate list. This particular setting makes it easier for the search-engine to achieve a faster retrieval. The structure of each list includes three columns: the first one contains a transliteration of the words in Coptic; the second one has the corresponding translation into Greek; the third one contains a commentary on the corresponding Coptic word (e.g., original source, dialect, part-of-speech, etc.). The information contained in the dictionary can be modified or enriched easily, through the spreadsheet application (in this case, the Microsoft Office Excel). The data sources for this lexical database include Coptic dictionaries available both in printed form (Crum, 2005) and online (Coptic Dictionary Online and CopticBook, n.d.). The commentary of their entries is being translated into Greek (by the authors). The initial dataset for testing/validation includes scripts on artifacts and manuscripts exhibited at the Byzantine and Christian Museum of Athens, the Benaki Museum (Athens) and the National Library of Greece (Kontogianni et al., 2021: 32).

3.2.3 Searching Process and Corpora

The search process is done in two ways. First, there is a Cartesian dictionary in which the words were arranged in size and alphabetically doing a linear search. Secondly, another base will be created in order to be used for a weighted linear search, a process based on Zipf's law (Papakitsos, 2013). It is also sequential, like the previous one, but it is

executed in tables with the data in terms of their probability of occurrence, in descending order (preloading) (Papakitsos, 2014).

For creating the frequency tables Scriptorium corpora will be used. Scriptorium is a collaborative, digital project created by Caroline T. Schroeder (University of Oklahoma) and Amir Zeldes (Georgetown University) (copticcriptorium.org). They provide us with 39 separate texts for reading, analysis and complex searches. Their subject matter is quite long like magic papyri, the Book of Ruth, manuscripts, the Gospel of Mark, the Assumption of John, various biographies, etc.

The benefits of dual search are great. It will achieve faster and more accurate search results while at the same time will provide the necessary conclusions on how to rank databases for faster search with the most reliable results.

3.2.4 Interface

The application's interface is being designed as a window-screen very simple to use. It depicts the Coptic alphabet at left side and a range of buttons in order to facilitate the user (help, correct, clear, exit). At the center of the window-screen the Coptic word is displayed based on the letters selected by the user, while the search-engine is activated through the Search button. If the formed word exists in the dictionary, then the Greek translation and the corresponding comments appear in the text-box below the Search button (Results), otherwise a failure message is displayed there. Finally, the software provides the potential to save in a simple text-file, the complete searching process of a session. In this way, the entire inscription of an artifact can be printed in a translated form, being commented as well, for facilitating both the synchronous and the asynchronous study of the artifact.

4. Conclusions

Coptic has given important findings in our country, however, the tools that will help the scholars to process and study it, are poor. This tool it works on the one hand like a digital platform for deciphering, translating, and interpreting existing ancient texts by digitizing them, and on the other can be also used for teaching the rendered languages. Therefore, it can become another promising tool for the digitization of cultural heritage, a model of computer-assisted processing of ancient languages, and an aid for the scholar of the Coptic language. Using additional double search will be able to give not only reliable results in the tool itself but also to draw general

conclusions about how similar software should be structured, thus NLP processing and computer-assisted translation usually are formed upon databases.

References

- Allen, P. J. (2013). *The Ancient Egyptian Language: An Historical Study*. Cambridge: Cambridge University Press.
- Coptic Dictionary Online (n.d.). Ed. by the Koptische/Coptic Electronic Language and Literature International Alliance (KELLIA), available from: <https://coptic-dictionary.org/> (accessed 2021-03-17).
- Coptic Dictionary (n.d.). CopticBook: Books of the Coptic Orthodox Church, available from: <http://e.copticbook.net/copticdictionary.php> (accessed 2021-03-17).
- Crum, W. E. (2005). *A Coptic Dictionary*. Eugene, OR: Wipf and Stock.
- Kontogianni, 2014 [Κοντογιάννη, Α. (2014). Ερευνητικό Λογισμικό: Ανάπτυξη Ηλεκτρονικού Λεξικού Γραμμικής Β΄. Διπλωματική Εργασία: Διεπιστημονικό Διαπανεπιστημιακό ΠΜΣ «Τεχνολογία» (ΕΚΠΑ & ΕΜΠ)].
- Kontogianni et al., 2017 [Κοντογιάννη, Α., Παπαμιχαήλ, Χ., & Παπακίτσος, Ε.Χ. (2017). Εκπαιδευτικό Λογισμικό για την Εκμάθηση της Γραμμικής Β΄. Πρακτικά 9ου συνεδρίου «Η Πληροφορική στην Εκπαίδευση» (CIE2017, σ. 423-433), Πανεπιστήμιο Πειραιώς, 13-15 Οκτωβρίου 2017].
- Kontogianni, A., Ganetsos, T., Kousoulis, P., and Papakitsos, E. C., (2020). Computer-Assisted Translation of Egyptian-Coptic into Greek. *Journal of Integrated Information Management*, 5(2), 26-31.
- Kontogianni, A., Ganetsos, T., Zacharis, N., Kousoulis, P., and Papakitsos, E. C., (2021). A Detailed Study about Egyptian-Coptic and Software Engineering. *Archaeology*, 9(1), 28-33.
- Kontogianni et al., 2019 [Κοντογιάννη, Α., Γκανέτσος, Θ., & Παπακίτσος, Ε.Χ. (2019). Ψηφιοποίηση Έργων Πολιτιστικής Κληρονομιάς: Η Περίπτωση της Γραμμικής Β΄. Εισήγηση στο 3ο Πανελλήνιο Συνέδριο Ψηφιοποίησης Πολιτιστικής Κληρονομιάς: EuroMed 2019, Συνεδριακό Κέντρο Πανεπιστημίου Δυτικής Αττικής, Αιγάλεω, 25-27 Σεπτεμβρίου 2019].
- Mavridaki, A., Galiotou, E, and Papakitsos, E. C., (2020). Designing a Software Application for the Multilingual Processing of the Linear A Script. *Proceedings of the 24th Pan-Hellenic Conference on Informatics (PCI 2020)*, Athens, Greece, November 20–22, 2020. ACM, New York, NY, USA, 3 pages.

- Mavridaki, A., Galiotou, E., and Papakitsos, E.C. (2021). Developing a Software Application for the Study and Learning of Linear A Script. *Review of Computer Engineering Research*, 8(1), 8-13.
- Papakitsos, 2013 [Παπακίτσος, Ε.Χ. (2013). Φυσική Γλώσσα & Υπολογιστικά Μαθηματικά. Αθήνα: Διεπιστημονικό Διαπανεπιστημιακό ΠΜΣ «Τεχνογλωσσία» (ΕΚΠΑ & ΕΜΠ)].
- Papakitsos, 2014 [Παπακίτσος, Ε.Χ. (2014). Γλωσσική Τεχνολογία Λογισμικού: Π. Πραγμάτωση. Αθήνα: ISBN 978-960-93-6235-1].
- Schroeder, C., and Zeldes, A. (n.d.). University of Oklahoma & Georgetown University, available from: <https://copticcriptorium.org/> (accessed 2021-06-20).
- Schroeder, C. T., Behlmer, H., Platte, E., Schmid, U., Zeldes, A., Miyagawa, S., and Feder, F. (2019). KELLIA Final White Paper. National Endowment for the Humanities (Grant HG-229371), Deutsche Forschungsgemeinschaft (BE 4172/1-1) (available from: <https://kellia.uni-goettingen.de/downloads/KELLIA-white-paper.pdf>).
- Wilfong, T. G. (2018). Coptic language. *Oxford Classical Dictionary*.