Journal of Integrated Information Management electronic journal

Computer-Assisted Translation of Egyptian-Coptic into Greek

Argyro Kontogianni[1], Theodoros Ganetsos[1], Panagiotis Kousoulis[2], Evangelos C. Papakitsos[1]
[1]University of West Attica, Greece, [2]University of the Aegean, Greece
akontogianni@uniwa.gr, ganetsos@uniwa.gr, kousoulis@rhodes.aegean.gr, papakitsev@uniwa.gr

Article Info

Article history:
Received 06 March 2021
Received in revised form 15 March 2021
Accepted 18 March 2021

https://doi.org/10.26265/jiim.v5i2.4470

Abstract:

Purpose *- The Coptic language is the final evolutionary stage of Ancient Egyptian, firstly attested in the 1st century CE (Old Coptic) until the 16th century CE, which survives nowadays as the liturgical language of the Coptic Orthodox Church. Egyptian-Coptic had been transliterated through the alphabetization of written Egyptian, based on the Greek script. In this work, the designing of a software tool is presented, for the computer-assisted learning and translation of Egyptian-Coptic into Greek.*

Design/methodology/approach *- The software design herein emphasizes the ability of processing inscriptions on artifacts through a simple interactive interface for its usage by Greek-speaking scholars of Humanities, having just essential familiarization with computers. For this purpose, a model of processing has been selected, based on an existing software tool for other ancient languages, which is being currently modified for processing Coptic scripts.*

Findings *- Especially in Greece, several museums and institutions preserve known manuscripts and artifacts of this language and art. Besides the large number of Coptic corpora (gospels and scriptures), a significant and historically important part of texts is found on Coptic artifacts, being difficult to be automatically processed, especially in a Greek context. Considering the above issues, it had been decided in the research project to implement a semi-automated approach, which will allow the user to choose the processed unit, whether this will be smaller, equal or larger than word, following a word-by-word computer-assisted translation into Greek.*

Originality/value *- Considering the cultural importance of Coptic, they create a continuous of almost 5000 years of written sources in Egyptian language, with an undisputable value in the diachronic history of languages, and prove the unsurpassed cultural relationship between the Ancient Greek and Egyptian civilizations. Therefore, the development of the presented software application aims at further facilitating the Greek-speaking researchers and scholars of Coptic to study this language in depth.*

Index Terms **-** computer-assisted translation, Coptic, Egyptian, cultural heritage digitization, cultural informatics applications, software for ancient language translation.

I. Introduction

The Egyptian-Coptic script 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 **[1]**. These evolutionary phases had been **[2, 3]**:

- The Old or Archaic Egyptian (2600 - 2100 BCE), found on the monumental inscriptions of the Ancient Kingdom (e.g., at the Pyramids), and having two writing systems with a parallel function and appearance developed: the hieroglyphic and the hieratic script; the latter being a simpler form of the former.
- The Middle or Classical Egyptian (ca. 2100 - 1300 BCE} that mark the classical phase of the language, during the Middle Kingdom.
- The Late Egyptian (ca. 1300 - 600 BCE) that differed particularly in grammar, also having many linguistic loans of Semitic, Hittite and African origin.
- The Demotic Egyptian (ca. 650 BCE - 300 CE) of the Late and Greco-Roman periods, when significant changes happened in the structural function of the language, due to foreign linguistic influences; the corresponding writing system that appeared around 650 BCE is the evolution of Hieratic, it is read from right to left and lasted for about a thousand years **[4]**; therefore, it coexisted with Coptic during the last three centuries of its existence; Tomas Young **[5]** firstly concluded that this script is a mixture of alphabetic and hieroglyphic signs, in his attempt to decipher hieroglyphics.
- The Coptic (300 BCE - 1500 CE), which is the last phase of the evolution of the vernacular language, with the simultaneous use of the Greek alphabet; the grammatical rules of the Late Egyptian had been generally retained along with an important part of the Egyptian vocabulary, containing Greek and other words of Pharaonic origin; nowadays it survives as the liturgical language of the Coptic Orthodox Church, written in an alphabetical script.

A. The Coptic Language

Since 332 BCE, Egypt was conquered by Alexander the

Great, and consequently the language of the conqueror (i.e., Greek) became the standard one for the documents of public administration, instead of Egyptian. Since the end of the 1st century CE, when the Christian religion had started spreading in the Eastern Mediterranean, those Egyptians who followed Christianity seemed reluctant to use the Demotic Egyptian script for writing their sacred texts, because it had a direct connection to the former pagan religion. Therefore, they preferred to transliterate the Egyptian language by using the letters of the Greek alphabet, thus creating the Coptic script. Coptic became the dominant way of writing for secular and sacred texts, by the 5th century CE.

The influence of Greek in the creation of the Coptic language was decisive, being evidenced by the number of Greek loans. For example, in the database of Freie Universität Berlin there are 8,000 Egyptian-Coptic entries and 3,250 Egyptian-Greek entries **[6]**. A wide range of documentary texts from the later Roman, Byzantine, and early Islamic periods are included in the Coptic corpora. They constitute a rich, extensive and original translated Christian literature of particular importance for the early history of Christian monasticism, being considered excellent witnesses of great Gnostic, Manichaean and Hermetic texts **[7]**.

Six distinct dialects and various sub-dialects appear in Coptic. These are **[1]**: the Sahidic dialect (also known as the "dialect of Upper Egypt" or former Thebaic), the Bohairic dialect (also known as the "dialect of Lower Egypt" or Memphitic), the Fayyumic dialect (or Bashmuric), the Akhmimic dialect, the Lycopolitan dialect {alias Subakhmimic or Assiutic), and the Oxyrhynchite dialect (or Mesokemic). Each dialect comprises a corpus with separate linguistic features and occasionally separate thematic features. Although eventually Coptic was replaced by Arabic, as the language of daily life in Egypt, it remains today as a functional language in the Christian communities of Egypt and the expatriate Coptic communities around the world **[1]**.

B. Artifacts and Objectives

The Coptic artifacts represent 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. Museums and institutions preserve evidence of this language and art. 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. Therefore, the development of the presented herein software application aims at facilitating the researchers and scholars of Coptic to study this language in depth, by making it easier for them to learn and study Coptic, without being exceptionally familiarized with computers.

II. Materials And Methods

The methodology of developing a software tool for the study of Coptic by scholars of the Humanities can be dictated by three factors: the nature of the artifacts to be studied, the features of the particular script and the existing software applications for this purpose. As mentioned previously, the nature of the artifacts varies, having though as a common characteristic that the ability of automated processing of text is rare. The usual scenario is to have a photo of the text on an artifact and then to use a program of pattern recognition, if any, as a first step for its translation. Another way is to have a software tool that may facilitate 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. For this purpose, the features of the script have to be studied and the relevant software applications that follow.

A. The Coptic Script

A writing system or script reflects the phonological and phonetic system of the corresponding language. In this respect, although the Coptic script marks the final stage of the ancient Egyptian language, it is not exactly an evolution or a simplification of the previous writing systems (hieroglyphic, hieratic and demotic), since the Egyptians adopted the Greek alphabet, at this stage of linguistic depiction. After the conquest of Egypt by Alexander the Great, in 332 BCE, the influence of Greek vocabulary in Coptic led to the adoption of Greek terminology in administrative affairs and to the progressive integration of Greek words into the Demotic Egyptian (Old Coptic). The spoken Egyptian language was written in Greek characters; therefore, the Coptic alphabet initially reflects the phonological values of the Greek prototype **[1]**. In addition though, the Coptic alphabet is enriched with eight more signs for representing the consonants that does not exist in the Greek language. Consequently, a script of 32 signs is formed, corresponding to 26 distinctive sounds.

B. Relevant Software Tools

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. For example, there is a unique application developed by Snyder et al. **[8]**, with the aim of automatically deciphering the Ugaritic language, a lost but well-known language classified as Western Semitic of the 14th century BCE, written in a cuneiform alphabet of consonants [9]. Snyder et al. [8] developed here a statistical model of unsupervised machine translation (i.e., without

human intervention), which uses a dictionary of the Hebrew language, due to the affinity of Ugaritic with Hebrew. The results have remarkable decryption accuracy (90.53%), but show that no matter how accurate an unsupervised system is, it still requires some human intervention for the rest of the text **[10]**. The system of Snyder et al. **[8]** though is a machine-translation one for a language previously unknown; this is not at all the case herein.

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 **[11]** that indicate the international interest on digital Coptic. These resources of KELLIA Project comprise an online dictionary that includes a virtual keyboard option for entering Coptic characters **[12]**, a treebank with full syntactic annotations, a search engine **[11]**, an optical character recognition (OCR) tool and various NLP tools and services **[13]** that are being developed for part-of-speech tagging, lemmatization and language-of-origin tagging [14]. Desirable software tools are also a converter of Coptic characters into the Unicode Coptic character set and a tokenizer **[14]**.

C. Processing Issues

Despite the plethora of software resources, presented in the previous subsection (*II.B. Relevant Software Tools*), several processing issues emerge, regarding Coptic. The need for tokenization has been already mentioned because of the agglutinative morphology of Coptic [14], while another related issue is that Coptic was originally written in manuscripts without spaces between words (*scriptio continua*) **[12]**. These two particular features seriously complicate automated processing, along with the frequent existence of diacritics (such as supralinear strokes or circumflexes), punctuation and abbreviations that require a *normalizer* for further machine-enabled processing **[14]**. In overall, there are multi-word entries from corpora and units smaller than words for productive derivations and incorporation **[12]**. Finally, 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 **[15]**.

Considering the above issues, it had been decided in the herein research project to implement a semi-automated approach, which will allow the user to choose the processed unit, whether this will be smaller, equal or larger than word, following a word-by-word computer-assisted translation into Greek. The software design 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. For this purpose, another model of processing has been selected **[16]**, based on an existing software tool for ancient languages **[17]**, which has been successfully used for processing Linear-B script **[10, 18]** and is being also modified for processing Linear-A script **[19]**.

III. RESULTS

According to the previous decision, the developed software tool consists of three main modules: the Coptic-Greek digital dictionary, the search-engine and the interface. From these three modules only two are visible to the user: the dictionary and the interface; they will be briefly described in the respective sub-sections, below. 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.

A. The Digital Dictionary

The Coptic-Greek digital dictionary is a lexical database being implemented as a single spreadsheet file (Figure 1). 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 **[20]** and online **[21-22]**. 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* **[1]**.

B. The Interface

The application's interface is being designed as a window-screen very simple to use (Figure 2). The entire user-guide is accessible on screen by activating the Help button, first on the right side of the window-screen. The Coptic characters are placed on a virtual keyboard on the left side of the window-screen, in alphabetical order. The users can select the characters of a word by activating the corresponding icons on the virtual keyboard. The sequence of the selected characters is displayed on the central upper text-box (*Coptic Word*).

	A	B	C
1	**Coptic**	**Greek**	**Comments**
2	ⲁⲓⲁⲓ	αυξάνω, μεγαλώνω, μεγεθύνω	
3	ⲁⲓⲟⲩ	ταξιδεύω, πηγαίνω	(πιθανόν), το νόημα είναι άγνωστο
4	ⲁⲟⲩⲱ	παύω	
5	ⲁⲃⲏⲱ	__	άγνωστο νόημα, επίθετο που χρησιμοποιείται για κάποια εδώδιμα
6	ⲁⲗⲟⲩ	παιδί, νέος	
7	ⲁⲓⲡⲉ	αρχιμανδρίτης	
8	ⲁⲱⲁⲓ	πληθαίνω, πολλαπλασιάζω	
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			

1letter | 2letters | 3letters | **4letters**

Figure 1. A sample of the digital dictionary.

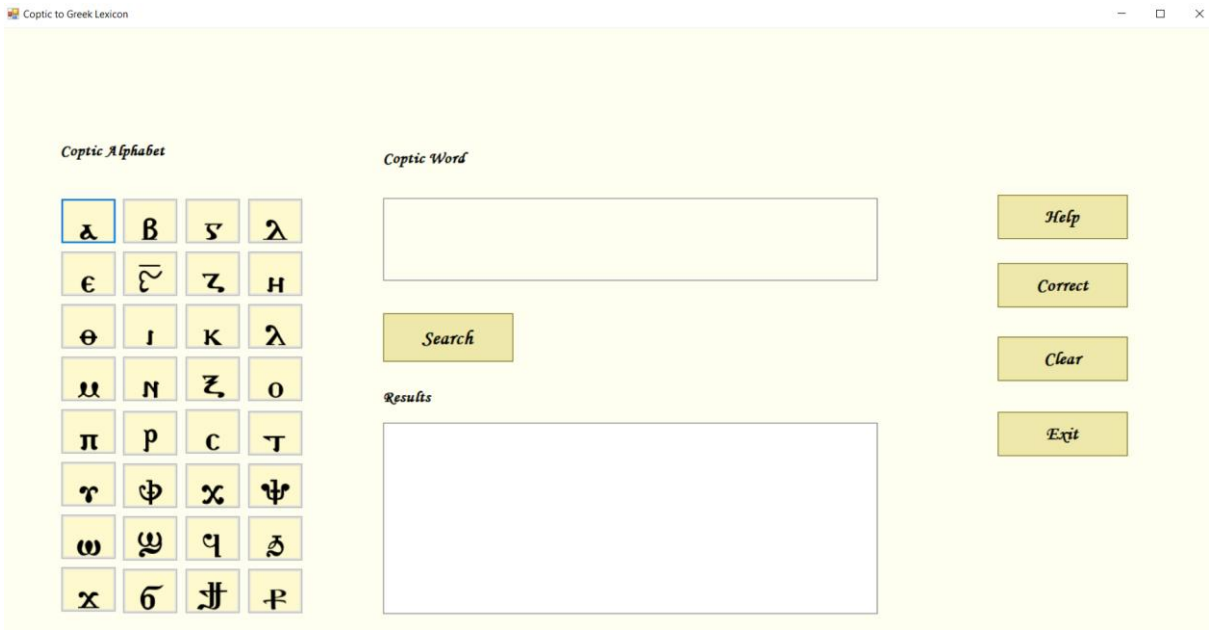

Figure 2. The interface of the software application.

If an icon is selected incorrectly, it can be deleted by activating the Correct button, second on the right side of the window-screen. The search-engine is activated through the Search button, in the center of the window-screen. 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.

The users can then activate the Clear button, third on the right side of the window-screen, to repeat the search with a new word. Various messages that may appear in pop-up windows inform the user of incorrect/failed actions. By activating the Exit button, last on the right column of the window-screen, the users may close the session and the application; alternatively, the standard "x" button can be used, on the upper right corner of the window-screen. Finally, each translation session is simultaneously recorded and thus saved in a simple text-file, which can be printed afterwards. 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.

IV. Discussion

Regarding the results of linguistic software applications, as presented in the previous section, the more complex the software is, the less supervised the application may be. Systems that translate text from one language (or script) to another (here from Coptic into Greek) can be also considered as computer-assisted translation systems, rather than a standalone application. Here, the human participation is still necessary, in various parts and tasks of the software tool, because language rules or phenomena cannot always be formulated in a computationally

unsupervised manner. The unsupervised processing may work well mainly on manuscripts, yet, it cannot perform adequately in the cases of artifacts, corrupted or fragmentary text that may also lack syntactic cohesion (see subsection II.C. *Processing Issues*). On the contrary, a supervised software tool, like the one presented herein, may also facilitate the teaching/learning of the target-language (i.e., Coptic), as demonstrated in a similar case [18].

In this respect, the current software development process modifies an existing digital platform for translating Coptic into Greek. The development of similar software systems could greatly aid the painstaking efforts of the researchers involved and also contribute to the rescue of the studied artifacts from possible physical damage, through digitization. In addition, the translation, study and interpretation of existing texts, but also the facilitation of reading new ones that may be discovered, are enhanced, since such applications can be supported by rich, recent and explanatory databases, with the ability of continuous enrichment.

V. Conclusion

Although Coptic is considered a dead language, its study and digitizing efforts remain undiminished. In Greece, there is a plethora of artifacts, which Coptic script is found on, but the theoretical background for Coptic is poor and the software tools related to this language for Greek are absent. Under the auspices of the University of West Attica, it was deemed necessary to develop a similar software tool, according to the existing methodology related to other non-spoken languages (i.e., Linear-A and Linear-B). This software development is being adapted for computer-assisted translation of Coptic into Greek, in such a way as to meet the complex features of the Coptic language. This software system introduces a digital platform for deciphering, translating and interpreting existing ancient texts by digitizing them, which 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.

REFERENCES

[1] A. Kontogianni, T. Ganetsos, N. Zacharis, P. Kousoulis, and E. C. Papakitsos, “A Detailed Study about Egyptian-Coptic and Software Engineering,” Archaeology, vol. 9, no. 1, pp. 28-33, 2021 (Proceedings of the 7th Balkan Symposium on Archaeometry, 22-25 September 2020, University of West Attica, Athens, Greece).

[2] P. J. Allen, “Middle Egyptian: An Introduction to the Language and Culture of Hieroglyphs,” Cambridge University Press, Cambridge, 2000.

[3] P. Kousoulis, “Introduction to Ancient Egyptian Language and Scripture” [in Greek], Association of Greek Academic Libraries, Athens, 2015.

[4] P. J. Allen, “The Ancient Egyptian Language, An Historical Study,” Cambridge University Press, 2013.

[5] A. Robinshon, “Writing and Script: A Very Short Introduction,” Oxford University Press, 2009.

[6] Database and Dictionary of Greek Loanwords in Coptic (DDGLC), available from: https://www.geschkult.fu-berlin.de/en/e/ddglc/index.html

[7] T. G. Wilfong, “Coptic language,” Oxford Classical Dictionary, 2018.

[8] B. Snyder, R. Barzilay, and K. Knight, “A Statistical Model for Lost Language Decipherment,” Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, pp. 1048-1057, 2010 (available from: https://dl.acm.org/citation.cfm?id=1858788).

[9] D. Pardee, “Ugaritic,” In: R. D. Woodard (editor), “The Ancient Languages of Syria-Palestine and Arabia,” Cambridge University Press, New York (NY), pp. 5-6, 2008.

[10] A. Kontogianni, T. Gkanetsos, and E. C. Papakitsos, “Digitization of Cultural Heritage Works: The Case of Linear B” [in Greek], 3rd Panhellenic Conference on Digitization of Cultural Heritage (EuroMed 2019), University of West Attica, 25-27 September, Egaleo, Greece, 2019.

[11] C. T. Schroeder, H. Behlmer, E. Platte, U. Schmid, A. Zeldes, S. Miyagawa, and F. Feder, “KELLIA Final White Paper,” National Endowment for the Humanities (Grant HG-229371), Deutsche Forschungsgemeinschaft (BE 4172/1-1), 2019 (available from: https://kellia.uni-goettingen.de/downloads/KELLIA-white-paper.pdf).

[12] F. Feder, M. Kupreyev, E. Manning, C.T. Schroeder, and A. Zeldes, “A Linked Coptic Dictionary Online,” Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature. Association for Computational Linguistics (Anthology ID: W18-4502), Santa Fe, New Mexico, pp. 12–21, 2018 (available from: https://www.aclweb.org/anthology/W18-4502.pdf).

[13] A. Zeldes, and C. T. Schroeder, “An NLP Pipeline for Coptic,” Proceedings of the 10th ACL SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH2016), Berlin, 2016 (available from: https://doi.org/10.18653/v1/W16-2119).

[14] C. T. Schroeder, and A. Zeldes, “Raiders of the Lost Corpus,” Digital Humanities Quarterly, vol. 10, no. 2, pp. 1-14, 2016 (available from: http://digitalhumanities.org/dhq/vol/10/2/000247/000247.html).

[15] I. Fourtounas, “The Greekness of the Coptic language; Pantainos, the Greek creator of Coptic” [in Greek], Published by the author, Athens, 2012.

[16] E. C. Papakitsos, A. Kontogianni, C. Papamichail, and I. K. Kenanidis, “An Application of Software Engineering for Reading Linear-B Script,” International Journal of Applied Science, Vol. 1, no. 2, pp. 58-67, 2018.

[17] A. Kontogianni, “Research Software: Development of a Linear B Lexicon” [in Greek], Interdisciplinary Joint Postgraduate Course “Technoglossia” (MSc), National & Kapodistrian University of Athens and National Technical University of Athens, 2014.

[18] A. Kontogianni, C. Papamichail, and E. C. Papakitsos, “Educational Software for Learning Linear B” [in Greek], Proceedings of the 9th Conference on Informatics in Education (CIE2017), pp. 423-433, University of Piraeus, Greece, 13th-15th of October, 2017.

[19] A. Mavridaki, E. Galiotou, and E. C. Papakitsos, “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. https://doi.org/10.1145/3437120.3437299

[20] W. E. Crum, “A Coptic Dictionary,” Wipf and Stock, Eugene, OR, 2005.
[21] “Coptic Dictionary Online,” ed. by the Koptische/Coptic Electronic Language and Literature International Alliance (KELLIA), available from: https://coptic-dictionary.org/ (accessed 2021-03-17).
[22] “Coptic Dictionary,” CopticBook: Books of the Coptic Orthodox Church, available from: http://e.copticbook.net/copticdictionary.php (accessed 2021-03-17).