

## **Interim Report**

National Endowment for the Humanities, Office of Digital Humanities HG-229371

Koptische/Coptic Electronic Language and Literature International Alliance (KELLIA)  
Project

Principal Investigators:

Caroline T. Schroeder, University of the Pacific

Amir Zeldes, Georgetown University

Institution: University of the Pacific

May, 2018

Elizabeth Platte

Caroline T. Schroeder

Amir Zeldes

## **Introduction**

This interim report documents the activities of the American participants in the joint NEH-DFG supported KELLIA project from December 2017, to May 2018.

In addition to the progress described on the original grant outcomes (below), key project participants held weekly virtual meetings; we consulted with our Advisory Board members on an individual basis; we prepared grant applications for further funding for the project.

### **Outcome 1 - Milestones for Data Standards**

This outcome is generally supported by the German partners. These partners began writing a draft of the data standards white paper.

### **Outcome 2 - Server-based batch conversion tools**

A TEI export module has been integrated into our online transcription and multi-layer annotation tool (GitDox). Corpora are now published via GitDox and openly released and versioned in TEI XML on GitHub.

### **Outcome 3 - Integration of linguistic tools and methods to produce collaborative digital editions**

These tools were developed in earlier phases. We have updated the natural language processing pipeline tools with new data, substantially improving morphological segmentation and syntactic parsing accuracy.

### **Outcome 4 - development of a web-based, multi-layer annotation tool for collaborative text annotations and stand-off markup**

This tool (GitDox) was developed in an earlier phase; during this period, in addition to the export facilities noted above, the data validation mechanisms were updated by Emma Manning (Ph.D. student, Georgetown University).

### **Outcome 5 - Sharing, linked data, and textual re-use**

Geographic data in our corpora is now discoverable in Pelagios and Pleiades due to linking our data with Pelagios. We continued ongoing conversations with Pleiades about direct integrations.

Version 2.2 of the Coptic treebank was released as part of the Universal Dependencies framework for cross-linguistic research (<http://universaldependencies.org/cop/>). Contributors are Amir Zeldes, Elizabeth Davidson, Mitchell Abrams.

The online dictionary updated some entries based on new decisions regarding standardizations of word segmentations and lemmatizations.

## **Publications and Presentations resulting from this project phase**

Miyagawa, So and Zeldes, Amir (2018) "A Semantic Map of the Coptic Complementizer če-based on Corpus Analysis: Grammaticalization and Areal Typology in Africa". *International Workshop on Semantic maps: Where do we stand and where are we going?* Liège, Belgium.

Schroeder, Caroline T. "Annotating Heresies (Keynote)." *Future Philologies*. Institute for the Study of the Ancient World. New York University. April 2018.

Our spring corpus release includes more textual data from our German partners Alin Suciu and Diliana Atanassova (texts by pseudo-Theophilus and Apa Johannes) as well as from American contributors (more sayings in the Apophthegmata Patrum). See <https://github.com/CopticScriptorium/corpora/releases/tag/v2.5.0>. We also released new versions of pre-digitized Old Testament and New Testament corpora machine annotated for syntactic parsing.

The Universal Dependencies 2.2 release is at

<https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1983> and

<http://universaldependencies.org/cop/> .