**Excellent**

**Schroeder**

**Coptic SCRIPTORIUM: Corpus, Tools, and Methods for Corpus Linguistics**

This project comprises a highly qualified and diversified team with needed skills in Coptic, computational linguistics (corpus linguistics, morphological parsing, annotation), and digital humanities (e.g., XML markup through TEI) to carry out an ambitious and well conceived project. Tool development will include a Part of Speech tagger, mechanisms for scholars to contribute texts and annotate others already in the corpus, and database search and visualization capabilities. The PI makes a strong case for this project is based on (1) the importance of Coptic for studying ancient history; (2) the lack any comparable resource on this language; (3) the significance of the potential deliverables. In regard to the latter, the PI proposes (a) TEI markup using EpiDoc; (b) development of computational tools to analyze, process and visualize this language; (c) a database created from the texts; (d) a collaborative platform for corpus enrichment and analysis. An impressive proof-of-concept video (which I have viewed) is available online. Although this is considered a pilot project the Schroeder and Zeldes have accomplished a level of preliminary work that makes this not only the best pilot I have ever seen but an outstanding proposal that will establish and strong foundation for all future work on Coptic. The proof of concept suggests that this project will be successful and have a significant impact on the study of ancient texts and religion.

The tools themselves (character encoding converter, tokenizer, lemmatizer, PoS tagger; see Appendix D: Coptic Scriptorium Tool List) are precisely what is needed for textual analysis including semantic and morphosyntactic research on Coptic. Unfortunately the morphological complexity of Coptic is not discussed beyond a brief mention that the language is agglutinative. One of the major roadblocks to lemmatization and PoS tagging is ambiguity of the input and the need for humans to disambuate in parses and glosses when needed. Some mention of this would have been helpful but the project team inspires confidence that the best available options will be used to parse the Coptic texts.

The PI has taken care to present the project, which is a complicated initiative, is a clear and concise manner. She has thought of many ways to ensure both that there is a strong foundation and that subsequent growth is facilitated (by online access as well as mechanisms by which other scholars can upload additional texts or annotate those already archived.

I strongly hope that this project be funded.

**Overview**

Intellectual significance: This is a highly significant project. It makes Coptic texts available in digital format (marked up with EpiDoc); it creates the infrastructure for advanced analysis of Coptic texts; it develops a mechanism by which other scholars may contribute to the corpus and database.

Impact on research and technology: This project will have a great impact on both substantive research on ancient texts and on the development of software technology to process these texts.

Innovation: The project is highly innovative not simply in the sense of creating new tools for analyzing Coptic texts but in the tightly integrated manner in which the project has been conceived and presented.

Proposal development: The PI is requesting a relatively small portion of funding given the extensive work that has already gone into digitizing Coptic texts and creating other preliminary software that will be used in the present effort.

Feasibility: The proposal inspires confidence that all goals will be met and that any difficulties that do emerge will be resolved.

Qualifications: The team is extremely well qualified and the collaboration between Schroeder, Zeldes and Almas ensure that all project needs will be administered to by highly competent experts.