Digital Discoveries and Collaborative Tool Development in the Classics

In recent years, classicists have frequently turned to computer software in order to drive their research and improve their teaching. Usually, this technology is designed for some kind of general purpose unconnected to our field. Many examples come immediately to mind: statistical software, algorithms for network analysis, programs for drafting or mapping, and packages for measuring, displaying and even printing artifacts. These products have been amazingly useful in the hands of skillful classicists, but we must take them as we find them and often must engage in a frustrating struggle to adapt them to our needs. But there is also another paradigm, one in which developers and classicist end-users work together for the building and improvement of software designed with *our* goals in mind. It is the purpose of this panel to illustrate the advantages and difficulties of this process by focusing on the development and use of the software tools associated with the Alpheios Project and the Perseids Platform. This panel will serve as a demonstration of the great benefits (and occasional pains) from such a collaboration, as well as an invitation for all who may be interested in joining the work.

The initial two papers describe research projects which would have been impossible without the tools produced through the efforts of Alpheios and Perseids. The first, "Authorship Identification of Short Texts Using Only Syntactic Features," describes the use of information generated by the treebanking annotation modules in order to execute exciting new advances in one of the oldest sub-fields practiced by computer-literate humanists: stylometry. Stylometry has the goal of identifying and quantifying the features of style that typify an author, genre, etc. There have been some striking successes in this research since the computer became ubiquitous, but one problem that has stubbornly remained difficult is the application of accepted stylometric methods to small texts, those that classicists would call "fragments." This paper represents a first

step toward that goal. It demonstrates how data from dependency treebanks yield results in stylometric tests that are markedly superior to vocabulary-based studies (99% accuracy in texts of 1000 words and over 94% accuracy in texts of 100 words).

Computers can notice patterns in style too small to be visible to the human eye. The second paper, "A Wealth of Variables: Using Syntactic Stylometry to Distinguish Signature Constructions in Herodotus and Thucydides," describes how syntactic data from the ancient Greek treebanks illuminate salient differences between these two influential authors. Unlike vocabulary, many aspects of an author's "syntactic fingerprint" are difficult to notice and describe. They are often mundane, significant only in the aggregate. The computer allows us to "lift" these fingerprints from a text by identifying the precise syntactic features that contribute significantly (in a statistical sense) to its particularity. Once these structures are known, examples may be gathered and analyzed along traditional lines, and the results used to draw "syntactic portraits" of texts and authors.

The next two papers will be presented by computer programmers involved in the development and maintenance of specialty classics software. The third paper, "Student, Scholar, Teacher, Software Developer: Working Together to Build Software for Engaging with Classics," describes the collaborative process ("wires exposed") employed in designing and implementing software for learning, research, and pedagogy in Classics. In this model, scholars and students fill the roles of product designers, testers, and developers, providing feedback, testing early releases of the software, identifying bugs, and verifying bug fixes. The resulting tools are versatile and robust enough to support a wide variety of pedagogical and research needs, but also highly tuned to the requirements of the community of participants.

The fourth paper, "Maintaining, Preserving, and Distributing Open-Source Software for the Classics," discusses the problem of future-proofing open-source software by writing it defensively and by making the project welcoming so that, both technically and non-technically, developers outside of academia can be encouraged to get involved. It will discuss how such a development model can benefit a project, and how these issues are being addressed within Perseids.

The final paper is titled, "From Reading to Research: How the Study of the Classics with Contemporary Resources Uncovers Opportunities with Broad Implications for the Humanities." It will discuss the conceptual adjustments entailed by the desire to produce an optimal environment for reading classical literature in the Alpheios Project. It describes a journey that begins with the modest desire to develop specialty software in Classic and progresses to encompass an exploration of the problems of computer-assisted language learning. Ultimately, the scope of the project has transformed from simple (!) digital tool development to incorporate cognitive science and explore the idea of "neurophilology."

The panel will conclude with comments by a respondent.