

## WORKSHOP IN COMPUTATIONAL LINGUISTICS AND LATIN PHILOLOGY

Place: University of Innsbruck, 15. International Colloquium on Latin Linguistics

Date: April 6, 2009

Workshop organizers: David Bamman (Perseus Project, Tufts University), Dag Haug (University of Oslo), Marco Passarotti (Catholic University of Milan)

Invited speaker: Roberto Busa, S.J.

Classical Studies has long had a history of driving pioneering research in linguistics and literary studies. The great Classical philologists and lexicographers of the 19<sup>th</sup> century are arguably some of the world's earliest and finest corpus linguists – but we find ourselves now lagging behind the achievements of other languages due in large part to the absence of structured digital resources on which to base our research. While the TLG and the Packard Humanities Institute each released their respective Greek and Latin corpus in the 1970s (only shortly after the release of the Brown Corpus of English in 1967), they remain today – almost 40 years later – two of our most widely used electronic resources. Those ensuing 40 years have seen the rise and widespread development of structured knowledge bases, such as huge treebanks to encode syntactic information in English, Czech, Arabic and over twenty other languages, lexical ontologies such as WordNet, and new corpora being annotated not just with their semantics and syntax disambiguated, but their named entities and propositional data made explicit as well.

We are, however, now beginning to see these same resources being developed for Latin, along with the automatic tools that can exploit them (such as automatic syntactic parsers and morphological taggers) and a new interest in quantitative research that can only exist as a result. As we enter this new era, we must take care to work together as a community going forward – the three organizers, for instance, are each leading the development of independent treebank projects for different eras of Latin (Classical, Biblical and Thomistic) and we recognize that the value of each project is exponentially greater when compatible with the others. This workshop aims to bring together scholars working in the field – both those developing such resources and those conducting linguistic research using them – to share such work and experience.

We invite presentations including the following:

- \* Electronic resources for Latin in development
- \* Corpus linguistic research
- \* Application and evaluation of NLP tools on Latin texts
- \* Development of corpus-driven lexica
- \* Standards and standardization of annotation styles on different linguistic layers (e.g., morphological, syntactic, semantic, propositional)

Please submit abstracts of up to two a4-pages to Dag Haug at [daghaug@ifikk.uio.no](mailto:daghaug@ifikk.uio.no) before December 1, 2008. Notifications will be sent before January 1, 2009.