

## **Philipp ROELLI and Jan CTIBOR**

*Corpus Corporum*: current state and planned further development

In our talk we will explain the status quo of the *Corpus Corporum* and some of the planned future improvement. Today, the *Corpus Corporum* is the largest Latin full-text collection in existence. It is free and open and depends for its texts on the work of many individuals and initiatives at European and North American universities. It is organised in individual corpora usually stemming from one source (hence the name). The main idea was and is to present digitised Latin texts in a manner usable for linguistic studies and for online perusal in a sustainable way. At present we have about 160 Million Latin words in some 8000 texts. All texts are encoded in standard TEI XML and can usually be downloaded by users. We currently use a rather small subset of TEI's possibilities when TEI files are loaded into the MySQL database. The resource automatically lemmatises all words (using TreeTagger). This enables the reader to call up dictionary content about them simply by clicking inflected words, which makes it possible to read texts quickly. At present we use 12 Latin and Greek dictionaries, among them the standard Latin dictionaries translating into German, French, and English. The most recent acquisition is the *Latinitatis medii aevi lexicon Bohemorum* (thanks to our Czech colleagues, especially Pavel Nývlt). Besides, several types of complex queries can be performed (including lemmatised and time-dependent), and information about texts and authors is provided. Recently we have started to intensify our collaboration with SISMELE Firenze to link texts and authors to their resource *mirabile*.

Currently we are working on a new, much faster and more user-friendly interface, including an API allowing better linkage with other Latin resources. Planned new features include: searches flexible for mediaeval spellings, cross-links between several editions of one work (we have some 200 works in more than one edition), linking texts to external translations and graphical resources, displaying a wider subset of TEI's possibilities (like tables, or a more extended display of critical apparatuses), and the possibility to search only in one (the most recent) edition of each work. In the long run, when things become more stable we plan to release our software on internet platforms, enabling people to set up their own *Corpus Corporum*, online or offline. Without colleagues providing texts and dictionaries such a resource would be unthinkable. So we would be happy to find new collaborators and to contribute in other free and open initiatives working with digital Latin texts.