Dr. Oliver Hellwig
Institut für Sprachen und Kulturen Südasiens
Freie Universität Berlin
Germany

A computational framework for linguistic research in post-Vedic Sanskrit

While Vedic Sanskrit has been studied intensively in Indological research, the linguistic developments in the post-Vedic language have received much less attention. This is mainly due to two facts. First, post-Vedic Sanskrit is assumed to be a static idiom regulated to a large extent by the rules of Pāṇini's grammar, whose – unquestionable – influence is taken to have prevented major linguistic changes in post-Vedic Sanskrit. Second, the extent of the post-Vedic literature has made it impossible to study the (temporal) distribution of high-frequency linguistic phenomena systematically, as long as traditional philological methods are applied. However, advances in statistics, informatics, and corpus linguistics have created a "toolbox" of statistical and computational methods that can be used to (re-)examine some long-standing problems in the linguistic history of post-Vedic Sanskrit and its literature. The presentation does not address a specific problem in Sanskrit linguistics, but describes a generic computational framework that can guide research about high-frequency linguistic phenomena. First, the presentation sketches how Sanskrit texts have to be prepared for a computational analysis, how a corpus of Sanskrit texts should be structured, and which computer programs are available for building such a corpus. Next, it describes which classes of philological questions can be solved when existing computational and statistical procedures are applied to such a corpus. At this point, the presentation focusses on questions of text segmentation and automatic dating, which belong to the most complex problems in Indology. After a short theoretical discussion of the statistical and computational procedures that can be used to tackle these problems the results of some recent case studies from Sanskrit lexicography and stilometry are reported. Finally, the presentation sketches the direction of future research in computational Indology, which may yield promising results for the chronology of Sanskrit literature.