

## **SEMATIA: Linguistic Annotation of the Greek Documentary Papyri – Detecting and Determining Contact-Induced, Dialectal and Stylistic Variation**

This paper presents preliminaries of a project-to-be, the goal of which is to establish a linguistic annotation system for the corpus of Greek documentary papyri. This corpus illuminates both the vernacular and the administrative language and the registers in between in different time periods. Therefore, it is worthwhile to consider how to improve linguistic research possibilities on this material.

Digital research possibilities in the field of papyrology are advanced and being continuously developed (see <http://papyri.info/>). The digital papyrological corpus, the Duke Databank of Documentary Papyri (DDbDP), comprises of ca. 80 000 Greek documents of various size. Currently, it enables lemmatized and substring searches of the texts. These search possibilities are valuable for many types of research and when editing new texts. They are, however, difficult or even impossible to use in identifying linguistic features and variation.

How do we find variation, then? We can not annotate features as "substandard" vs. standard. Rather the answer lies in detailed annotation on several linguistic levels: phonological, morphological and syntactic.

Phonological variation and change can be studied from the papyrological material mainly through orthography. At the moment anomalous spellings can only be searched as guessing possible wrong spellings. It would be useful that the instances where the editor has suggested a regularization (to follow "standard spelling") could be searched as well. This is not an impossible scenario even now, since the regularizations are tagged in the XML in DDbDP, but we would still always need to search for certain words with known orthographic problems; we could not search single graphemes interchanging with each other. Thus, a more specific annotation is needed, so that all variant spellings could be found.

Morphological forms can, to some extent, be searched in the DDbDP. This requires, however, that the searched form does not seriously overlap with some other string of letters with a different meaning. Moreover, the historical changes in pronunciation caused overlapping spellings in, e.g., different case endings and some verbal forms which at the moment cannot be deciphered from the electronic corpus. Morphological annotation would make it possible for the first time to differentiate the intended morphological forms from the phonologically spelled ones and to study, for example, how the orthographic changes affected the case system and the loss of certain grammatical cases. In general, using part-of-speech-tags, as used in the Ancient Greek and Latin Dependency Treebanks (Perseus) would significantly improve the morphosyntactic analysis on papyrological texts.

At the syntactic level, the linguistically annotated database would enable study currently impossible. In DDbDP, we can not search for the subject or

the object of the clause, or their dependants, in any meaningful manner. The same concerns the relationship of a relative pronoun and its antecedent, infinitival constructions, and a countless number of other syntactic and pragmatic structures.

In this paper I will discuss what sort of annotation system would be most useful when considering 1) the fragmentary state of the texts, and 2) the aim of detecting variation. The system in Ancient Greek and Latin Dependency Treebanks (Perseus), do not take into account orthography, nor possible differences between functional vs. formal analysis (e.g. if a nominative case is used where an accusative would be expected). Thus, the customization needed for SEMATIA would concern both orthography and morphosyntactic variation.

I will present some of these issues by way of demonstrating the DDbDP available online and via examples from two different (sub)corpora: 1) papyri written by bilingual notaries in Hellenistic Upper Egypt, 2) papyri from a family archive from Early Byzantine Petra, Jordan.

Marja Vierros, PhD  
Department of World Cultures  
University of Helsinki  
marja.vierros@helsinki.fi