

**From edition to basic historical corpus: elements of a simple workflow**

In my talk, I present two projects that involve the re-use and partial enhancement of edition data for electronic corpora: first, the electronic corpus accompanying the Freiburg edition of the Velikie Minei Čet'i (Waldenfels & Rabus 2015) and the lemmatization and translation mapping for the Edition of Dečani 88 (Waldenfels and Taseva 2012).

The first project involves a quick and simple corpus based directly on the electronic version of the book project. XSLT scripts were used to extract the formatted text and reformatted for input in CWB, a standard corpus manager. The corpus is then queried through a web interface that was designed specifically to be useful for both experienced and novice users of corpus linguistic software; see <http://www.vmc.uni-freiburg.de>.

The second project concerns a different aspect of the transformation of an edition to a corpus, namely annotation. For the publication of the South Slavic manuscript Dečani 88 (Kakridis and Taseva 2014), all word forms in the edition were manually lemmatized and aligned to their hypothetical Greek sources. To this end, the edition text was transferred to a spread sheet using XSLT scripts, manually annotated there, and extracted again to build word lists. A side effect of these procedures was the data for a lemmatized and translation source annotated electronic corpus, which has, however, not been published in this case.

Both these projects involve different elements of a chain of procedures that can be used to derive an annotated electronic corpus from a traditional edition. Such procedures are straightforward using existing tools and programming resources which should, in my view, be exchanged between projects; it has to be kept in mind, though, that the most important outcome of such efforts are standard-compliant electronic versions of the edited texts in TEI or other standard formats, since tools and procedures will come and go, while the edition data are unique and will not, as a rule, be compiled a second time.

**References**

- Kakridis, Taseva 2014: Kakridis, Y. and L. Taseva. „Gegen die Lateiner. Traktate von Gregorios Palamas und Barlaam von Kalabrien in kirchenslavischer Übersetzung.“ Freiburg i. Br., 2014.
- Waldenfels, Taseva 2012: Waldenfels, R. v. and L. Taseva. „An der Schnittstelle von Korpuslinguistik und Paläoslavistik: Wörterverzeichnisse zu einer mittelalterlichen Handschrift als Keimzelle eines annotierten digitalen Korpus.“ In *Diachrone Aspekte slavischer Sprachen*, ed. by B. Hansen. München – Berlin, 2012, 243–259 (Slavolinguistica 16).
- Waldenfels, Rabus 2015: Waldenfels, R. v. and A. Rabus. „Recycling the Metropolitan: Building an Electronic Corpus on the Basis of the Edition of the Velikie Minei Čet'i“ *Scripta & e-Scripta* 14/15, 27-38.