# The PROIEL corpus's architecture and annotation schemes. How can we make the most out of scanty data? /  Àðõèòåêòóðà êîðïóñà PROIEL è ñõåìû àííîòèðîâàíèÿ. Êàê ìû ìîæ

Àâòîð Hanne Martine Eckhoff

04.10.2009 ã.
Ïîñëåäíåå îáíîâëåíèå  28.10.2009 ã.

Ìàòåðèàëû ê ëåêöèè (ïðåçåíòàöèÿ)

In this lecture we explain the architecture of the PROIEL parallel corpus of New Testament texts. The corpus consists of the original Greek NT and its earliest translations into Indo-European languages: Latin, Gothic, Classical Armenian and Old Church Slavic (OCS). We will use examples mostly from OCS. We will look at file formatting for import, text processing and the many layers of annotation employed in the corpus: for morphology, syntax and information structure. We will also describe our system of free tags that in principle can be used by any scholar for any purpose: we will exemplify this by discussing our animacy tagging.

Practice session. In this practice session we import two chapters from the Gospel of Matthew from the Codex Zographensis and let the participants try out the PROIEL application for annotation in the various layers.