

Marius L. Jøhndal, Hanne M. Eckhoff, Dag Haug (University of Oslo)

## The PROIEL parallel corpus of old Indo-European translations of the New Testament

In this talk we report on the ongoing creation of the Slavic and Greek parts of the PROIEL corpus of old Indo-European New Testament translations and demonstrate some applications in historical linguistics.

The texts are first annotated morphologically and syntactically (dependency-based). Syntactic units are then automatically aligned using Gale and Church's algorithm, which performs well on these literal translations. Finally, syntactic constituents are linked manually. These links are modeled as relations between nodes in aligned syntactic graphs or zero-terminated links when correspondences are lacking.

For a study of the rendering of the Greek genitive, for example, a link between the heads of `о приставлении домовънѣмь` and `τῆς οἰκονομίας σου` (Lk 16:2) gives a correct representation of one equivalence, but implies another incorrect equivalence between `σου` and some element of the Slavic phrase.

By adding links to subgraphs, the equivalences can be made progressively more precise. We thus sacrifice query precision for recall rate, but make the data useful to scholars immediately, as they can combine querying with refinement of the data.