LREC 2014 Workshop

CCURL 2014
Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era

Reykjavík (Iceland), 26th May 2014

Oral Presentation

Title

Using GrAF for Advanced Convertibility of IGT data

Authors

Dorothee Beermann, Peter Bouda

Abstract

With the growing availability of multi-lingual, multi-media and multi-layered corpora also for lesser-documented languages, and with a growing number of tools that allow their exploitation, working with corpora has attracted the interest of researchers from the theoretical as well as the applied fields. But always when information from different sources is combined, the pertaining lack of interoperability represents a problem. This is in particular a challenge for corpora from endangered and lesser described languages since they often originate from work by individual researchers and small projects, using different methodologies and different tools. Before this material can become a true resource, well-known differences in the physical as well as the conceptual data structure must be leveraged against ways of future data use and exploitation. Working with Interlinear Glossed Text (IGT), which is a common annotation format for linguistic data from lesser described languages, we will use GrAF to achieve Advanced Convertibility. Our goal is to build data bridges between a number of linguistic tools. As a result data will become mobile across applications.