

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

The Multilingual GRUG Parallel Treebank – Syntactic Annotation for Under-Resourced Languages

Author

Oleg Kapanadze

Abstract

In this paper, we describe outcomes of an undertaking on building Treebanks for underresourced languages Georgian, Russian, Ukrainian, and German - one of the “major” languages in the NLT world (Hence, the treebank ’s name – GRUG). The monolingual parallel sentences in four languages were syntactically annotated manually using the Synpathy tool. The tagsets follow an adapted version of the German TIGER guidelines with necessary changes relevant for the Georgian, the Russian and the Ukrainian languages grammar formal description. An output of the monolingual syntactic annotation is in the TIGER-XML format. Alignment of monolingual repository into the bilingual Treebanks was done by the Stockholm TreeAligner software. The parallel treebank resources developed in the GRUG project can be viewed at the URL of Saarland and Bergen Universities: <http://fedora.clarin-d.uni-saarland.de/grug/> , <http://clarino.uib.no/iness>.