Reykjavík (Iceland), 26th May 2014

Session 2

15:00-16:00

Chairperson: Eveline Wandl-Vogt

15:00-15:30

The LREMap for Under-Resourced Languages

Riccardo Del Gratta, Francesca Frontini, Anas Fahad Khan, Joseph Mariani, Claudia Soria

A complete picture of currently available language resources and technologies for the under-resourced languages of Europe is still lacking. Yet this would help policy makers, researchers and developers enormously in planning a roadmap for providing all languages with the necessary instruments to act as fully equipped languages in the digital era. In this paper we introduce the LRE Map and show its utility for documenting available language resources and technologies for under-resourced languages. The importance of the serialization of the LREMap into (L)LOD along with the possibility of its connection to a wider world is also introduced.

15:30-16:00

Using GrAF for Advanced Convertibility of IGT data

Dorothee Beermann, Peter Bouda

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.