abstract 167

**Typology with graphs and matrices**
*Oral presentation*
*Theme session:* Linked Data in Linguistic Typology

In this talk we demonstrate how to leverage Semantic Web technologies to transform data in any number of typological databases, e.g. WALS (Haspelmath et al, 2008), Autotyp (Bickel & Nichols, 2002), PHOIBLE (Moran, 2012), ODIN (Lewis, 2006) or individual databases -- along with metadata from Ethnologue (Lewis, 2009), LLMAP (ILIT), Multitree (ILIT, 2009) and Glottolog (Nordhoff et al, 2012) -- into Linked Data. This is the vision of the Linguistic Linked Open Data Cloud (LLOD; Chiarcos et al, 2012).

Once data from these databases are converted into a homogeneous format, i.e. RDF graph data structures, the contents of these disparate datasets can be merged into one large graph, which allows for their data to be queried in a federate search fashion, in line with how we currently search the content of the Web through popular search engines.

We will illustrate how a user can query and retrieve information about a particular language, from multiple databases, via a language's ISO 639-3 code. For example, a user might be interested in accessing all typological variables described by various databases for a particular language, e.g., word order data from WALS, genealogical information and phonological word domains from Autotyp, and phoneme inventory data from PHOIBLE, etc.

We show how the results of such queries can be combined and output into a matrix format that mirrors recent work in multivariate typology (cf. Witzlack-Makarevich, 2011; Bickel, 2011). By outputting the results of users' queries across different databases into table-based matrix formats, the results can be directly loaded into statistical packages for statistical analyses, and published algorithms can be directly applied to them and tested, e.g. statistical sampling procedures (cf. Cysouw, 2005) and statistical approaches to determine universal preferences, e.g. Family Bias Theory (Bickel, 2011). Furthermore, when typological data are output into tables, state-of-the-art approaches using linear algebra to transform matrices into new datasets can be applied (Mayer & Cysouw, 2011).