

Linguistic Linked Open Data: What's in for Machine Translation?

Christian Chiarcos

Angewandte Computerlinguistik

Fachbereich für Informatik und Mathematik

Goethe-Universität Frankfurt am Main

Robert-Mayer-Str. 10

D-60325 Frankfurt am Main

chiarcos@informatik.uni-frankfurt.de

During the past years, the notion of Linked (Open) Data has gained considerable reception in different communities working with language resources, ranging from academic and applied linguistics over lexicography to natural language processing and information technology. In this context, the Open Linguistics Working Group of the Open Knowledge Foundation (OWLG, <http://linguistics.okfn.org/>), founded in 2010 in Berlin, Germany, is playing an important integrative role, by reaching out to a broad band-width of disciplines, by facilitating interdisciplinary information exchange through meetings, workshops, datathons and joint publications, but most notably by introducing and maintaining the Linguistic Linked Open Data (LLOD) cloud diagram. Being deeply involved in this emerging community at the intersection between the different disciplines mentioned above, I will introduce the basic concepts of Linked Open Data for linguistics/NLP, summarize motivations and history of Linguistic Linked Open Data so far. Since creating the first instantiation of the LLOD cloud diagram in 2012, LLOD has attracted a lot of activity, we have reached an agreement on vocabularies for many aspects of language resources and the number of resources included is continuously on the rise. This growth is documented, for example, by declaring LLOD “the new hot topic in our (= language resource) community” (Nicoletta Calzolari, LREC-2014 closing session). But with substantial amounts of data being available, the focus of activity in the LLOD community is slowly shifting from resource creation to applications of Linguistic Linked Open Data. The primary promise of providing open, but heterogeneously structured and scattered language resources in a more interoperable way has been fulfilled, and it facilitates using and re-using existing language resources in novel contexts. Beyond this, innovative LLOD-based applications for common problems in Natural Language Processing, Digital Humanities and linguistics are on the horizon. The second part of the talk will give a glimpse on these prospects by discussing use cases and potential applications of LLOD for (Deep) Machine Translation.