The talk presents the ongoing DFG-funded research project on "Sustainability of Linguistic Data". The project was initiated by the DFG, as a cooperation between three collaborative research centers in Germany - the SFB 441 "Linguistic Data Structures" in Tübingen, the SFB 538 "Multilingualism" in Hamburg, and the SFB 632 "Information Structure" in Potsdam/Berlin. The DFG was concerned about the long-term availability of the large collection of high-quality linguistic resources amassed within these projects. The major aim of this project is therefore to analyze resources acquired in projects from SFBs 441, 538, and 632 and to ensure their interoperability and accessibility since the first of these SFBs will end in 2008. The project also deals with more general aspects of sustainability of linguistic data, e.g. the use of terminologies or ontologies, approaches to meta-data, and legal issues. (Schmidt et. al. 2006)

In the presentation the individual data handling solutions developed and applied at the three centers will be described and compared. An architecture for the integration of the differently annotated language corpora is introduced. (Wörner et. al. 2006) Beside these technical aspects of the representation of linguistic data, the talk also addresses legal issues of linguistic corpora. (Rehm & Witt 2007)

The talk ends with a short advertisement for the Biennial GLDV Conference 2007 "Data Structures for Linguistic Resources and Applications" which will take place in April 2007 in Tübingen. (Rehm et. al. 2007)

Refs.:


1) Lehmberg, T. & F. Zimmermann. Language Corpora: The Legal Point of View,


