

Sustainability of language resources: a project report.

Andreas Witt, University of Tübingen

In 2005, the DFG initiated a research project with the goal of providing an accessible dissemination and sustainable storage of corpora developed or annotated within projects from three Collaborative Research Centres, located in Hamburg, Berlin/Potsdam and Tübingen. For this purpose, the resources have been converted into a generalized format that contains all the information included in the original data. This format is an integral part of a software platform currently under development. The platform allows for storing, retrieving, querying, and presenting the linguistic resources. Its architecture, *Generalised Architecture for Sustainability* (GENAU), aims at unifying conceptually different approaches to multi-layered annotations. To achieve this, corpora based on diverse formalisms are converted into multiply annotated XML documents, without losing information. Since the source formats are based on the Annotation Graphs framework, on hierarchically annotated linguistic data, or on stand-off annotated texts all current paradigms for linguistic annotations are covered by this approach.