

Prof. Dr. Erhard Hinrichs  
Text+  
Leibniz-Institut für Deutsche Sprache  
R-5, 6–13  
68161 Mannheim

**GSCL**

Gesellschaft für  
Sprachtechnologie  
& Computerlinguistik

**Erster Vorsitzender / President**

Prof. Dr.-Ing. Torsten Zesch  
Sprachtechnologie / Language  
Technology  
Universität Duisburg-Essen  
Forsthausweg 2  
47057 Duisburg

*Telefon:* +49-331-977-2691

*Telefax:* +49-331-977-2761

*E-Mail:* torsten.zesch@uni-  
due.de

25. September 2020

**National Research Data Infrastructure: Support for Consortium Text+**

Dear Professor Hinrichs,

The German Society for Computational Linguistics and Language Technology (*Gesellschaft für Sprachtechnologie und Computerlinguistik*, GSCL) sees an urgent need for a national research data infrastructure in the field of text- and language-based data. The GSCL supports the plans by the Text+ consortium, which make provisions for researchers not only from the humanities, but also from the more technical fields such as computer science. The implementation of a national infrastructure in general and of the Text+ in particular would be highly beneficial for improving Germany's position in the field of speech and language processing.

The GSCL is the association for research, teaching and professional work in natural language processing and computational linguistics in the German-speaking area. Historically, the GSCL has positioned itself between the fields of humanities and the field of computer science. The interests of our current members range from foundational aspects of linguistics to building practical natural language and speech processing tools, with a focus on the latter.

All our members make use of research data in the form of language- and text-based resources. For modern language and speech processing methods, which are based on statistics, machine learning and artificial intelligence, collections of linguistic material are needed as training material. The automatic processing of text needs open standards and interfaces that allow access to resources across institutional boundaries. A more open and integrated infrastructure is a possibility for researchers to build excellent software with "big data" without relying on proprietary resources provided by external and commercial entities. The CLARIN and DARIAH infrastructures have begun to integrate these aspects. Text+ plans to build on these achievements by providing, among others, integrated workspaces and an authorization infrastructure that facilitates extended usage scenarios. For instance, it will enable model development based on resources which cannot be made publicly available in their entirety, e.g., due to privacy issues or other sensitive data. Text+ will nevertheless enable scientific benchmarking of methods. Another important aspect is the integration of heuristic tools and lexical resources, as those can be used to enhance existing tool chains. Overall, the reuse of components will improve the free flow of information and ideas. This will be a great incentive for the development of new approaches and the improvement of existing ones.

The GSCL strongly welcomes the plans of the consortium, which take into account several relevant interests, and which are open both to scholars in the humanities and to researchers in language and speech processing. The GSCL sees great potential for cooperation in building and integrating language- and text-based resources. Resources and tools developed by the scientific community can be integrated easily into the Text+ infrastructure due to adherence to open standards and interfaces. Integration would allow that new resources and tools are tested more easily, and by increasing interoperability (standards), accessibility (single entry point) and findability (central registries), it would increase reuse of data and tools greatly.

In the past, the GSCL has supported several shared tasks, which included providing training and gold standard evaluation materials. The use of a unified infrastructure would have greatly facilitated the construction of these materials, but also the competition process as all participants could have operated on the same interface.

The GSCL hereby states its willingness to support Text+ in dissemination and will actively encourage its members to participate in Text+ activities and development.

The GSCL has always mediated between scholars in the humanities and computer scientists, and it will continue in this direction. It would welcome the Text+ consortium as a strong partner in this enterprise and expresses its hope that the application will be successful.

Kind Regards.

On behalf of the GSCL



Prof. Dr.-Ing. Torsten Zesch  
President, GSCL



Dr. Annemarie Friedrich  
Vice-President

GSCL Letter of Support Text+ NFDI Consortium



