

Annotations as category-based interpretations of texts

Carolin Odebrecht, Humboldt-Universität zu Berlin

The terms *assignment*, *encoding*, and *annotation* are notions for a categorical interpretation of texts. The identification and definition of exponents, annotation concepts, tags, assignment guidelines and modes are essential requirements based on and motivated by research contexts and questions for example in literary studies, linguistics, philology and digital humanities. Each of these aspects creates dependencies and consequences with respect to their application. This is evident by various adaptations of best practice tagsets, different implementations in different formats and applications to different text models. Therefore, it is crucial to define analyses goals, the annotation model and its realization. Which categories need to be annotated (and how) in order to solve given research questions? Which annotation concepts are appropriate? Which text concepts are used in the data? Should annotations support text visualisation, close reading, quantitative or/and qualitative, corpus-based or/and corpus-driven analysis methods?

In my talk, I will discuss these questions by using the TEI Guidelines (TEI Consortium 2019) as a framework example that is used in different research contexts for encoding historical texts. As corpus examples, I will use the corpora ELTeC (Schöch 2019) and RIDGES (Lüdeling et al. 2018).

References:

- Lüdeling, Anke, Carolin Odebrecht, Laura Perlitz und Amir Zeldes, Hrsg. (2018). RIDGES Herbolology. Version 8.0. Humboldt-Universität zu Berlin. doi: https://doi.org/10.34644/laudatio-repository-VWhgk2oB6bp_h9NaOboz_1557324977. url: <http://korpling.org/ridges/>.
- Schöch, Christof (2019). COST-ELTeC. Zenodo. doi: <https://zenodo.org/record/3462436>.
- TEI Consortium (2019). TEI P5: Guidelines for Electronic Text Encoding and Interchange. url: <http://www.tei-c.org/Guidelines/P5/> (last accessed 01.11.2019).