

Donnerstag
09.03.2017
13:45 – 14:15
B3 1, 0.12

Annotating a multiregional diachronic corpus of Early New High German handwritten texts

Lisa Dücker
Universität Hamburg
lisa.duecker@uni-hamburg.de

Stefan Hartmann
Universität Hamburg
stefan.hartmann@uni-hamburg.de

Renata Szczepaniak
Universität Hamburg
renata.szczepaniak@uni-hamburg.de

This talk discusses the methodologies and annotation guidelines employed in encoding a corpus of 57 Early New High German protocols of witch trials. To our knowledge, this corpus represents the first attempt to make spontaneously produced, handwritten German texts available as a linguistic corpus. The corpus was compiled in a project on the interplay of semantic and syntactic factors in the evolution of sentence-internal capitalization (*Satzinterne Großschreibung*, SiGS for short) in German.

Handwritten texts entail specific challenges for the linguistic encoding of the data, such as defining sentence or word boundaries (Szczepaniak & Barteld 2016). In this talk, we will give a brief overview of how these challenges were approached. However, our main focus will be on the semantic annotation of the data. As animacy plays a major role in the development of noun capitalization (cf. Bergmann & Nerius 1998), all nouns were encoded for animacy. Semantic role annotation is another major step in encoding the corpus data. In a first step, we use a binary coding scheme (“agent” vs. “non-agent”), which will be further extended in a subsequent phase of the project. However, the SiGS corpus is not only a valuable resource for studying graphemic change. The data can also be used to pursue, for instance, morphological or syntactic research questions. Therefore, the data will eventually be made available via ANNIS. In sum, we believe that the SiGS corpus is an important addition to the German corpus landscape and that the decisions made during the encoding process can also prove insightful for other researchers dealing with historical language data.

References: • Bergmann, R. & D. Nerius (1998): *Die Entwicklung der Großschreibung im Deutschen von 1500–1700*. Heidelberg. • Szczepaniak, R. & F. Barteld (2016): Hexenverhörprotokolle als sprach-historisches Korpus. In S. Kwekkeboom & S. Waldenberger (eds.), *PerspektivWechsel* 43–70. Berlin: Schmidt.