
Sprachgebrauch süddeutscher Klosterfrauen des 17. Jahrhunderts: Linguistische Aufbereitung und Annotation eines historischen Spezialkorpus

Jessica Nowak¹, Stefan Hartmann²

¹University of Mainz, ²University of Bamberg

nowakj@uni-mainz.de, stefan1.hartmann@uni-bamberg.de

In historischen Referenzkorpora des Deutschen sind Texte, die von Frauen verfasst wurden, nach wie vor stark unterrepräsentiert. Gerade aus kultur-, aber auch aus sprachhistorischer Perspektive kann sich jedoch der Vergleich von Texten, die von Schreiberinnen stammen, mit solchen, die von Männern verfasst wurden, als aufschlussreich erweisen, wie etwa die Befunde zeigen, die auf Grundlage des Fürstinnenkorrespondenz-Korpus von Lühr (2016) erzielt wurden. In diesem Poster stellen wir ein Projekt vor, das zur digitalen Verfügbarkeit von Texten frühneuzeitlicher Schreiberinnen in linguistisch annotierter Form beiträgt und damit wichtige Grundlagenarbeit für weiterführende Analysen etwa aus kulturgeschichtlicher oder genderlinguistischer Sicht leistet. Ziel des Projekts ist die linguistische Aufbereitung und öffentliche Zugänglichmachung der von Nolting (2010) edierten Briefe und chronikalischen Texte süddeutscher Klosterfrauen des 17. Jahrhunderts. Mithilfe von CAB („Cascaded Analysis Broker“, Jurish 2012) werden die edierten Texte graphematisch zum Neuhochdeutschen hin normalisiert und auf dieser Basis POS-getaggt und lemmatisiert. In einem weiteren Schritt erfolgt eine manuelle Korrektur der Annotation. Weiterhin werden nach dem Vorbild des Fürstinnenkorrespondenz-Korpus grammatische, diskurs- und textpragmatische Annotationen vorgenommen, soweit diese auf die Klosterfrauen-Texte übertragbar sind. Hierzu gehören beispielsweise die Annotation syntaktischer Funktionen oder für die Brieftexte die Auszeichnung einzelner Kategorien des Briefstils (bspw. Grußformeln). Zusammengenommen können diese Annotationen die Grundlage einer umfassenden grammatischen wie auch pragmatischen Analyse bilden. Für die manuelle Annotation wird dabei die Software GATE verwendet (Cunningham et al. 2002). Für die Veröffentlichung werden die Daten für das browserbasierte Korpusabfragesystem ANNIS (Krause & Zeldes 2016) aufbereitet und können voraussichtlich Ende 2019 öffentlich zugänglich gemacht werden.

References: Cunningham, Hamish, Diana Maynard, Kalina Bontcheva & Valentin Tablan (2002): GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. *Proceedings of the 40th Anniversary Meeting of ACL*. • Jurish, Bryan (2012): Finite-state Canonicalization Techniques for Historical German. PhD Thesis, University of Potsdam. [urn:nbn:de:kobv:517-opus-55789](http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-55789). • Krause, Thomas & Amir Zeldes (2016): ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1), 118–139. • Lühr, Rosemarie (2016): Satzkomplexität in fürstlichen Korrespondenzen der frühen Neuzeit. In Sarah Kwekkeboom & Sandra Waldenberger (eds.), *PerspektivWechsel, Bd. 1*, 455–478. Berlin: Erich Schmidt Verlag.