

# Non-Standard Text Explorer - Ein Suchtool für historische und andere nicht-standardisierte Texte

Fabian Barteld (Universität Hamburg)

Das korpusbasierte Arbeiten mit historischen Texten wirft Probleme auf, die bei standardisierten gegenwartssprachlichen Texten nicht auftreten. Ein prominentes Beispiel hierfür ist die Schreibvariation (vgl. z.B. Baron et al. 2009): Bereits für das einfache Durchsuchen der Texte nach Wortformen müssen Verfahren angewendet werden um möglichst auch Schreibvarianten zu finden (vgl. u.a. Jurish et al. 2014). Während sich für die Nutzung von annotierten Korpora ANNIS (Krause/Zeldes 2016) als Standard etabliert hat, existiert keine Entsprechung für die Suche in nicht-annotierten historischen Texten. Diese Lücke soll mit dem im Vortrag vorgestellten „Non-Standard Text-Explorer“ geschlossen werden. Das Programm schlägt mithilfe von Verfahren zur automatischen Schreibvariantenidentifikation (vgl. u.a. Barteld 2017) Schreibvarianten für eine gegebene Suchanfrage vor und erlaubt es diese Vorschläge anzupassen, d.h. zu entfernen aber auch hinzuzufügen. Dies vereinfacht das Auffinden von relevanten Belegstellen in Texten mit Schreibvariation, ohne eine Annotation der Texte vorauszusetzen.

## Quellen

Baron, Alistair/Rayson, Paul/Archer, Dawn. „Word frequency and key word statistics in historical corpus linguistics“. In: *Anglistik: International Journal of English Studies* 20, Nr. 1, 2009, S. 41–67.

Barteld, Fabian. „Detecting spelling variants in non-standard texts“. In: *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 2017, S. 11–22.

Jurish, Bryan/Thomas, Christian/Wiegand, Frank. „Querying the Deutsches Textarchiv.“ In: *MindTheGap@ iConference*, 2014, S. 25–30.

Krause, Thomas/Zeldes, Amir: ANNIS3: A new architecture for generic corpus query and visualization. In: *Digital Scholarship in the Humanities* 31 (2016), S. 118-139.