

Linguistische Basisannotation für älteres Deutsch: Tokens, Tags und Tools in der Praxis

Claudia Resch (Österreichische Akademie der Wissenschaften)

Vorliegender Beitrag berichtet aus der Praxis der Erstellung und Annotation historischer Spezialkorpora. Da die Mehrzahl dieser Projekte zeitlich und personell limitiert ist, bleibt oftmals kaum Zeit zur vergleichenden Evaluierung und Erprobung unterschiedlicher Tagsets oder Tools. Weil beste Ergebnisse in Form von verlässlich abfragbaren Interfaces erwartet werden, besteht die Herausforderung darin, im Annotationsprozess nachvollziehbare und auch nachhaltige Entscheidungen zu treffen, die möglichst alle vorausgegangenen relevanten Beobachtungen berücksichtigen.

Nicht ohne auf Ambiguitäten und hartnäckige Zweifelsfälle hinzuweisen, soll anhand von Beispielen aus dem „Austrian Baroque Corpus“ und aus einem kleineren Korpus frühneu- zeitlicher Flugblätter gezeigt und diskutiert werden, auf welche bewährten Tagsets und Tools bei der automatischen linguistischen Basisannotation von älterem Neuhochdeutsch und deren manueller Nachbearbeitung derzeit in der Praxis gesetzt werden kann.

Quellen

Dipper, Stefanie et al. 2013. HiTS: ein Tagset für historische Sprachstufen des Deutschen. In *Journal for Language Technology and Computational Linguistics* 28 (1), 85-137. http://www.jlcl.org/2013_Heft1/5Dipper.pdf

Hinrichs, Erhard, und Thomas Zastrow. 2012. Linguistic Annotations for a Diachronic Corpus of German. In *Linguistic Issues in Language Technology* 7, 1-16.

Mittmann, Roland. 2015. Automated quality control for the morphological annotation of the Old High German text corpus. Checking the manually adapted data using standardized inflectional forms. In *Historical Corpora. Challenges and Perspectives* (= CLIP 5), S. 65-76.

Resch, Claudia, und Ulrike Czeitschner. 2017. Morphosyntaktische Annotation historischer deutscher Texte: Das Austrian Baroque Corpus. In *Digitale Methoden der Korpusforschung in Österreich* (= Veröffentlichungen zur Linguistik und Kommunikationsforschung Nr. 30). Wien: Verlag der ÖAW, 39-62.

Resch, Claudia. 2017. »Etwas für alle« – Ausgewählte Texte von und mit Abraham a Sancta Clara digital. *Zeitschrift für digitale Geisteswissenschaften*. http://www.zfdg.de/2016_005