
Nina Bartsch^{*}, Stefanie Dipper^{*}, Lars Eschke[‡], Birgit Herbers^{*}, Thomas Klein[‡],
Sarah Kwekkeboom^{*}, Elke Weber[‡], Klaus-Peter Wegera^{*}
^{*} Ruhr-Universität Bochum, [‡] Universität Bonn

ANNOTIERTES REFERENZKORPUS MITTELHOCHDEUTSCH (1050-1350)

Poster + Demo

Das DFG-geförderte Projekt 'Annotiertes Referenzkorpus Mittelhochdeutsch (1050-1350)' hat zum Ziel, die hochdeutschen Sprachdenkmäler von ca. 1050 bis 1200 weitestgehend vollständig und von ca. 1200 bis 1350 in strukturierter Auswahl für eine breite Nutzergemeinde für historiolinguistische und mediävistische Forschung verfügbar zu machen. Die ausgewählten Texte, bei denen es sich ausschließlich um Primärquellen handelt, werden diplomatisch (d.h. handschriftengetreu) digital erfasst und in weiteren Arbeitsschritten z.T. semiautomatisch lemmatisiert und morphologisch sowie nach Wortarten annotiert. Für die Transkription der Texte wird das im Projekt entwickelte Tool OTTO eingesetzt. Die Annotation geschieht mit den schon für das Mittelhochdeutsche Grammatik-Korpus verwendeten Bonner Tools, die für die Arbeit am Referenzkorpus angepasst und verbessert worden sind. Soweit möglich werden überdies auch Annotate aus retrodigitalisierten Textglossaren extrahiert. Das Annotationsschema orientiert sich am Stuttgart-Tübingen Tagset (STTS). Insgesamt wird das Referenzkorpus Mittelhochdeutsch etwa 325 mhd. Texte im Gesamtumfang von knapp 2 Mio. Wortformen enthalten.

Die transkribierten Texte und ihre Annotationen werden im Austauschformat PAULA gespeichert und in das Korpustool ANNIS importiert. ANNIS unterstützt die Suche nach Textfragmenten kombiniert mit Beschränkungen über Annotationen. Damit steht das Korpus für Recherchen nach lexikalischen und grammatischen Phänomenen zur Verfügung. Von ANNIS aus soll auch ein Zugriff auf Abbildungen der Handschriften möglich sein.

Referenzen

Korpus der Mittelhochdeutschen Grammatik:

<http://www.degruyter.de/files/pdf/9783484110038Quellenkorpus%20Uebersicht.pdf>

OTTO: <http://www.linguistics.ruhr-uni-bochum.de/otto>

ANNIS: <http://www.sfb632.uni-potsdam.de/annis/>

PAULA: <http://www.sfb632.uni-potsdam.de/~d1/paula/doc/>