

---

Natalia Mull<sup>1</sup>, Sonja Linde<sup>2</sup>, Roland Mittmann<sup>3</sup>, Julia Richling<sup>2</sup>

<sup>1</sup>Schiller-Universität Jena <sup>2</sup>Humboldt-Universität zu Berlin <sup>3</sup>Goethe-Universität Frankfurt a.M.

**REFERENZKORPUS ALTDEUTSCH:  
KOMPLEXE SPRACHHISTORISCHE RECHERCHEN MIT DER DATENBANK ANNIS**

Poster + Demo

---

Das DFG-geförderte Projekt „Referenzkorpus Altdeutsch“ erfasst sämtliche überlieferten Texte des ältesten Deutsch (Althochdeutsch und Altsächsisch, um 750 – etwa 1050 u.Z.) mit ca. 650 000 Wortformen. Die digitalisierten Daten werden linguistisch annotiert und mit zusätzlichen Meta-Daten versehen.

Die entsprechend aufbereiteten Daten werden sukzessive in die im Rahmen des SFBs 632 (Berlin/Potsdam) entwickelte Datenbank ANNIS importiert.

Für das Projekt „Referenzkorpus Altdeutsch“ ist es notwendig, ANNIS in Bezug auf die spezifischen Anforderungen des historischen Textmaterials und den Bedürfnissen einer sprachhistorisch motivierten Recherche zu modifizieren.

Die äußerst vielschichtige Annotation und die durch ANNIS bereitgestellten Werkzeuge ermöglichen umfangreiche Suchabfragen.

So können die altdeutschen Daten nach Lemmata, nach Wortarten und nach flexionsmorphologischen Merkmalen recherchiert werden. Ebenfalls liegt eine einfache syntaktische Kennzeichnung vor. Zudem ist es mit den durch das Referenzkorpus Altdeutsch bereitgestellten Daten erstmalig möglich, statistische Aussagen über linguistische Phänomene des ältesten Deutsch zu erlangen.

Das Poster stellt die Aufbereitung der Daten für die Datenbank und die Möglichkeiten komplexer linguistischer Recherchen in ANNIS in den bis dato öffentlich zugänglichen Teilkorpora des Altdeutschen vor.

Zusätzlich wird eine Präsentation am Laptop angeboten und die KonferenzteilnehmerInnen erhalten die Möglichkeit, mit Unterstützung der AutorInnen selber Suchabfragen vorzunehmen.