

Linguistisches Kolloquium

13.12.2016

BERNHARD FISSENI, BERNHARD SCHRÖDER

(DUISBURG-ESSEN)

FnhdC: teilrenoviert und neu vermessen.

Untersuchungen am Bonner Frühneuhochdeutschkorpus

Das *Bonner Frühneuhochdeutschkorpus* (FnhdC) wurde zwischen 1972 und 1974 an der Forschungsstelle *Frühneuhochdeutsch* der Universität Bonn zusammengestellt und stellt einen Ausschnitt eines Gesamtkorpus von 1500 Texten dar, dessen 40 Texte aus 10 der insgesamt 22 Sprachlandschaften kommen und 7 Zeitschnitten à 50 Jahre (1350–1700) zugeordnet sind; das Korpus wurde für die Erstellung der *Grammatik des Frühneuhochdeutschen* entwickelt und für Substantive, Verben und Adjektive morphologisch annotiert. Zu Beginn des Jahrtausends nach XML konvertiert und 2007 um eine Wortform-Suchmaschine und eine HTML-Präsentation der morphologischen Annotation erweitert.

Renovierung

In der Zwischenzeit haben sich einige Fehler in der Kodierung des FnhdC bemerkbar gemacht und es ist gelungen, weitere Annotationen, etwa zur Adjektiv-Syntax und Verb-Morphologie, zu entschlüsseln bzw. besser zu verstehen. Da das Korpus häufig für Untersuchungen zum Frühneuhochdeutschen verwendet wird, scheint es sinnvoll, noch einmal Arbeit in Korrekturen und Erweiterungen zu investieren, auch wenn es sich strenggenommen weniger um Linguistik als um angewandte Computerarchäologie handelt.

Graphematische Vermessung

Als Fallbeispiel für eine Untersuchung am FnhdC wird ein Vergleich der graphematischen Systeme vorgestellt.

Dabei kommen mehrere Systeme zum Einsatz:

- die Clusterung der Texte nach n-Grammen,
- die qualitative und quantitative Betrachtung von Abständen zur neuhochdeutschen Orthographie anhand von Orthographie-Transformations-Regeln und
- ein Versuch, Wörter zwischen verschiedenen Orthographien zu alignieren.

Die für diese Untersuchung entwickelten Werkzeuge können auch bei der Renovierung des Korpus eingesetzt werden.