

Korpuskonzeptionelle Überlegungen zu einem syntaktischen Referenzkorpus des Neuhochdeutschen

Volker Emmrich / Mathilde Hennig (Universität Gießen)

Im Vortrag sollen sowohl das Vorhaben des DFG-Projekts „Syntaktische Grundstrukturen des Neuhochdeutschen. Zur Fundierung eines Referenzkorpus Neuhochdeutsch“ vorgestellt sowie insbesondere korpuskonzeptionelle Überlegungen der syntaktischen Annotation beleuchtet werden. Ein zentrales Ziel des Projekts ist es, die theoretischen und empirischen Grundlagen für eine Syntax des Neuhochdeutschen zu legen. Dazu werden Texte aus den Bereichen Wissenschaft, Alltag, Gebrauchsliteratur und Belletristik aus dem 17. bis 19. Jahrhundert syntaktisch erschlossen.

Um die syntaktischen Strukturen des Neuhochdeutschen umfassend darstellen zu können und um eine breite Nutzbarkeit in verschiedenen Forschungskontexten zu ermöglichen, wird im Korpus eine Mehrebenenannotation verfolgt. Zentrale Ebenen sind dabei satzübergreifende textgrammatische Phänomene, satzsyntaktische Grundstrukturen sowie Wort- und Ausdrucksarten. Mit der Vergabe von atomaren Werten und modularen sowie funktionalen Annotationsebenen soll zum einen Anschlussfähigkeit an bestehende historische Korpora gewährleistet werden. Zum anderen soll Nutzern die Möglichkeit eröffnet werden, ihre theoretischen Annahmen durch die Vielzahl von Annotationsebenen – und vor allem ihre Kombination – in komplexen Suchanfragen realisieren zu können.

Das Herzstück der satzsyntaktischen Korpuskonzeption bildet eine eigens für das Projekt entwickelte Baumbankstruktur. Im Gegensatz zu bisherigen Baumbankansätzen ist diese auf der Basis der Mehrebenenrepräsentation in ANNIS mit weiteren Analyseebenen (bspw. textgrammatischer und wortartmorphologischer Art) kombinierbar; der Mehrwert besteht vor allem aber in der Ausdehnung der Erfassung hierarchischer Form-Funktions-Beziehungen auf den Bereich der internen Struktur von Wortgruppen. Das Baumbankkonzept wird im Vortrag am Beispiel des für die sprachhistorische Forschung einschlägigen Phänomenbereichs vorangestellter Adverbialsätze vorgeführt.