

Thomas Schmidt und Dolores Batinić

Das Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK) als Quelle für lexikographische Arbeit

Das Forschungs- und Lehrkorpus Gesprochenes Deutsch (<http://agd.ids-mannheim.de/folk.shtml>) ist ein großes aktuelles Gesprächskorpus des Deutschen. Es umfasst aktuell Audioaufnahmen von 219 verschiedenen Interaktionen im Umfang von knapp 170 Stunden (entspricht ca. 1.6 Millionen Tokens), die literarisch transkribiert, standardorthographisch normalisiert sowie mit Lemma- und Part-Of-Speech-Tags annotiert werden. FOLK zielt auf eine möglichst große Variationsbreite in den Interaktionstypen. Aktuell enthält es verschiedenste Gesprächstypen aus dem privaten Alltagsbereich (z.B. Tischgespräche, private Telefongespräche), aus beruflicher und institutioneller Kommunikation (z.B. Schulunterricht, Verkaufsgespräche, berufliche Meetings) und aus öffentlicher Kommunikation (z.B. Podiumsdiskussionen).

Die Arbeit im Projekt „Lexik des Gesprochenen Deutsch“ stützt sich wesentlich auf dieses Korpus, das sowohl als primäre Analysegrundlage als auch als Quelle für lexikographische Beispiele dient. In unserem Beitrag stellen wir das Korpus vor und gehen auf einige speziell für die lexikographische Arbeit relevanten Aspekte ein, die wir in der bisherigen Projektarbeit identifiziert haben.