

## Gesprochene Fremdsprache Deutsch: Instrumente für Forschung und Vermittlung

Simon Sauer (Humboldt-Universität zu Berlin)

BeMaTaC: Ein tief annotiertes multimodales Map-Task-Korpus gesprochener Lerner- und Muttersprache

Korpora bieten die Möglichkeit, den tatsächlichen Sprachgebrauch von DaF-LernerInnen quantitativ und qualitativ auszuwerten. Insbesondere bei gesprochener Sprache ist jedoch oftmals gar nicht klar, wie die Zielsprache aussehen sollte, da Grammatikwerke und linguistische Theorien (wenn überhaupt empirisch fundiert) nahezu ausschließlich auf geschriebener (Zeitungs-)Sprache basieren.

Das Berlin Map Task Corpus (BeMaTaC) ist ein frei verfügbares und erweiterbares Korpus gesprochener Sprache. Es besteht sowohl aus taskorientierten Dialogen fortgeschrittener DaF-LernerInnen, als auch auf unter identischen Bedingungen erhobenen Aufnahmen mit deutschen MuttersprachlerInnen. Hierbei erklärt eine Person einer anderen eine Route auf einer Karte mit Landmarken. Dieses Setting erlaubt sowohl spontane konversationale Sprache, als auch einen thematisch und pragmatisch kontrollierten Kontext, um die beiden Gruppen systematisch vergleichen zu können.

Die Transkription wurde separat normalisiert und mit STTS-Wortarttags und Lemmatisierungen versehen. Weiterhin sind verschiedene, für gesprochene Sprache typische, Phänomene annotiert: Disfluencies wie z.B. Filler (äh, ähm), Wortdehnungen und Wiederholungen, sowie Äußerungsspannen, Reparaturen und Backchanneling. Das Audiosignal, sowie eine Videoaufnahme der die Route nachzeichnenden Hand, sind mit allen Daten aligniert und verknüpft. Anhand von umfangreich erhobenen Metadaten können eine detaillierte Datenauswahl sowie Vergleiche zwischen beliebig ad hoc festgelegten Gruppen erfolgen.

Neben dem Korpus an sich, soll der Vortrag anhand von Beispielen mögliche Untersuchungsszenarien und Anwendungsmöglichkeiten für Deutsch als Fremdsprache vorstellen.

Zugang zum Korpus, weitere Informationen und Literatur: <http://u.hu-berlin.de/bematac>