

## **KORPUSBASIERTE METHODEN ZUR ERSTELLUNG DER LEGEDE-RESSOURCE**

**Meike Meliss / Christine Möhrs / LeGeDe-Team (IDS)**

Am IDS Mannheim wurde im September 2016 die Arbeit an dem Forschungsprojekt „Lexik des gesprochenen Deutsch“ (LeGeDe) aufgenommen, welches die Erstellung einer korpusbasierten elektronischen Ressource des gesprochenen Deutsch auf der Grundlage von Untersuchungen der Besonderheiten von mündlichem vs. schriftlichem Sprachgebrauch im Bereich des Lexikons anvisiert.

Die Untersuchungen zum Forschungsgegenstand dieses Projektes, der gesprochenen Sprache des Deutschen, erfolgen auf der Datengrundlage des „Forschungs- und Lehrkorpus gesprochenes Deutsch“ (FOLK, vgl. Schmidt 2014a, 2016; Kupietz/Schmidt 2015), das als erstes großes Gesprächskorpus des Deutschen am IDS aufgebaut wird. Die „Datenbank gesprochenes Deutsch“ (DGD: vgl. Schmidt 2014b), über die FOLK recherchierbar ist, ist mit innovativen korpus-technologischen Funktionalitäten ausgestattet und beinhaltet vielfältige Optionen der Erschließung mündlicher Daten nach linguistischen und interaktionalen Merkmalen.

Im Rahmen der Methodenmesse sollen unterschiedliche empirische Methoden vorgestellt werden: Einerseits soll der Gegenstandsbereich – typische lexikalische Einheiten der gesprochenen vs. geschriebenen Sprache – identifiziert und erfasst werden. Andererseits sollen besondere Eigenschaften der gesprochenen Lexik auf formaler, inhaltlicher und kommunikativ-funktionaler Ebene im Vergleich zu den Einheiten des geschriebenen Deutsch beschrieben werden. Frequenzgesteuerte Daten zu Lemmata und Wortformen in Verbindung mit ihrer Kombinatorik (Bi- und Trigramme) und ihren Kookkurrenzprofilen sowie die detaillierte, semiautomatische Kodierung von zufallsgenerierten Stichproben zu ausgewählten Lemmata sind einige der methodologisch relevanten Verfahren (Perkuhn et al. 2012; Seretan 2010). Sie werden sowohl für die Erstellung der Stichwortliste als auch für die Makro-, Mikro- und Mediostruktur der geplanten LeGeDe-Ressource eingesetzt.

### **Literatur:**

- Kupietz, Marc/Schmidt, Thomas (2015): Schriftliche und mündliche Korpora am IDS als Grundlage für die empirische Forschung. In: Eichinger, Ludwig (Hg.): Sprachwissenschaft im Fokus. (= Jahrbuch des Instituts für Deutsche Sprache 2015). Berlin/Boston: de Gruyter, S. 297-322.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc (2012): Korpuslinguistik. Paderborn: Fink.
- Schmid, Hans-Jörg (2015): A blueprint of the Entrenchment-and-Conventionalization Model. Yearbook of the German Cognitive Linguistics Association 3, 1-27.

Schmidt, Thomas (2014a): The Research and Teaching Corpus of Spoken German – FOLK. In: Proceedings of LREC'14, Reykjavik, Iceland: ELRA.

Schmidt, Thomas (2014b): The Database for Spoken German – DGD2. In: Proceedings of LREC'14, Reykjavik, Iceland: ELRA.

Schmidt, Thomas (2016): Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken German. In: Kirk, John M./Andersen, Gisle (Hg.): Compilation, transcription, markup and annotation of spoken corpora. Special Issue of the International Journal of Corpus Linguistics [IJCL 21:3], 396-418.

Seretan, Violeta (2010): Syntax-Based Collocation Extraction. Text, Speech and Language Technology. Dordrecht.

**Ressourcen:**

Datenbank gesprochenes Deutsch (= DGD): [http://dgd.ids-mannheim.de/dgd/pragdb.dgd\\_extern.welcome](http://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome).

FOLK: <http://agd.ids-mannheim.de/folk.shtml>.