

Andreas Witt (Mannheim/Heidelberg)

Leibniz-WissenschaftsCampus „Empirical Linguistics & Computational Language Modeling (LiMo)“

Im Leibniz-WissenschaftsCampus LiMo arbeiten Forscherinnen und Forscher an der Schnittstelle von empirischer germanistischer Linguistik und Computerlinguistik. LiMo wird von der Leibniz-Gemeinschaft, dem baden-württembergischen Landesministerium MWK, der Universität Heidelberg und dem IDS finanziert und wurde im September 2015 offiziell gegründet.

Ein Leibniz-WissenschaftsCampus ist ein Modell, das von der Leibniz-Gemeinschaft initiiert wurde, um insbesondere den Austausch zu forschungsspezifischen Fragen zwischen Leibniz-Instituten und Universitäten zu stärken. Mit diesem Förderinstrument sollen Aktivitäten angestoßen werden, die sonst so nicht möglich wären. Das IDS bewarb sich in diesem Rahmen im Jahr 2014 erfolgreich darum, ein Projekt zu beginnen, da der Bereich Computerlinguistik am Institut für Deutsche Sprache nicht zu den institutionalisierten Arbeitsbereichen gehört, obwohl insbesondere die Anwendung computerlinguistischer Methoden auch im IDS zunehmend relevant wird. An der Universität Heidelberg ist mit dem Institut für Computerlinguistik (ICL) ein Institut etabliert, das – auch international – zu den besten computerlinguistischen Forschungseinrichtungen gehört. Aufgrund der räumlichen Nähe von Heidelberg und Mannheim bestand die Möglichkeit, das Instrument Leibniz-WissenschaftsCampus für die Kooperation von IDS und ICL zu nutzen.

Da – wie in der gesamten international sichtbaren, computerlinguistischen Forschung – auch am ICL englischsprachige Daten im Zentrum der Arbeiten stehen, besteht eine der Herausforderungen von LiMo darin, neue Methoden, Modelle und Werkzeuge zur Erfassung und Analyse von deutschsprachigen digitalen Sprachbeständen zu entwickeln und die Nutzbarkeit existierender computerlinguistischer Verfahren für das Deutsche zu betrachten. LiMo konzentriert sich auf bestimmte maschinelle Lernverfahren, die in der Computerlinguistik derzeit sehr verbreitet sind.

Ein weiteres Ziel dieser Kooperation ist es, computerlinguistische Analyseergebnisse in Form von Annotation in die Textkorpora des IDS zu integrieren, sodass sie dann, z.B. über die Korpusanalyseplattform KorAP (Bański et al. 2013, Diewald et al. 2016), den empirisch arbeitenden Germanistinnen und Germanisten für ihre Arbeit zur Verfügung gestellt werden können.

Literatur

- Bański, Piotr et al. (2013): KorAP: the new corpus analysis platform at IDS Mannheim. In: Vetulani, Zygmunt/Uszkoreit, Hans (Hg.): Human language technologies as a challenge for Computer Science and Linguistics. Proceedings of the 6th Language and Technology Conference. Poznań: Fundacja Uniwersytetu im. A., S. 586–587.
- Diewald, Nils et al. (2016): KorAP architecture – diving in the deep sea of corpus data. In: Calzolari, Nicoletta et al. (Hg.): Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia. Paris: European Language Resources Association (ELRA), S. 3586–3591.