

58. Jahrestagung des Leibniz-Instituts für Deutsche Sprache vom 15. bis 17. März 2022
(als Online-Konferenz)

KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

METHODENMESSE

Mittwoch, 16. März 2022, 15:45 Uhr bis 17:45 Uhr

Swiss-AL: Korpus und Workbench für mehrsprachige digitale Diskurse

Julia Krasselt, Philipp Dreesen, Matthias Fluor, Klaus Rothenhäusler, Maren Runte, Cerstin Mahlow, Anna Ulasik

Swiss-AL ist ein mehrsprachiges Korpus zur Analyse digitaler Diskurse in der Schweiz (Krasselt et al. 2020). Es wird seit 2016 am Departement für Angewandte Linguistik der Zürcher Hochschule für Angewandte Wissenschaften (ZHAW) entwickelt. Das Korpus enthält wichtige Akteure der Schweizer öffentlichen Kommunikationen aus den Bereichen Politik (z.B. Bundesämter und Parteien), Wirtschaft (z.B. Verbände und Spitäler), Wissenschaft (z.B. Hochschulen), Medien (Print und Online) und Zivilgesellschaft (z.B. NGOs). Die Medientexte stammen aus Schweizer Tages- und Wochenzeitungen sowie von Online-Newsportalen und werden systematisch über die Schweizer Mediendatenbank bezogen. Die Texte der anderen vier Bereiche werden mit Hilfe von Webcrawling kontinuierlich gesammelt. Dafür wird systematisch auf diejenigen Teile der Webseite eines Akteurs zugegriffen, die News, Medienmitteilungen und Blogeinträge enthalten. Die Erstellung des Korpus wird mit einer eigenen Verarbeitungspipeline prozessiert, mit je eigenständigen Komponenten für Datenverwaltung (mit Elastic Search), Preprocessing (insbesondere mit Komponenten zu Deduplikation und Scraping) und linguistische Verarbeitung (Wortarten, Dependenz-Parsing, Named Entity Recognition, semantische Annotationen). Das Korpus wird außerdem ergänzt durch Daten aus sozialen Medien (Twitter), Gesundheitsforen sowie Blogs (Fokus Rechtspopulismus).

Swiss-AL steht der breiten Öffentlichkeit zur Verfügung und wird in anwendungsorientierten Forschungsprojekten sowie in der linguistischen und kommunikationswissenschaftlichen Lehre genutzt. Dafür werden basierend auf spezifischen Modellierungskriterien Teilkorpora erstellt (Dreesen & Stücheli-Herlach 2019). Das Korpus wurde bisher für mehrsprachige Diskursanalysen eingesetzt, beispielsweise Diskurse zu erneuerbaren Energien, Covid-19, Antibiotikaresistenzen und Sozialhilfe. Die in Swiss-AL enthaltenen Medientexte bilden außerdem die Grundlage für die Wahl des Schweizer Wort des Jahres in den Sprachen Deutsch, Französisch, Italienisch und Rätoromanisch (Liste Lamas 2018).

Das Korpus kann über die an der ZHAW entwickelte Swiss-AL-Workbench genutzt werden (Krasselt et al. 2021). Die Workbench zeichnet sich durch eine leicht zugängliche und einfach gestaltete Benutzeroberfläche aus, bietet eine breite Palette von Analysemöglichkeiten und visualisiert Ergebnisse in unterschiedlicher Form. Die Workbench ermöglicht klassische korpuslinguistische Analysen, die für Diskursanalysen inzwischen etabliert sind (Konkordanzen, Kollokationen, Keywords). Darüber hinaus können computerlinguistische Methoden genutzt werden, die sich insbesondere für vertiefte Analysen an der Schnittstelle von Semantik und Pragmatik eignen (Topic Modeling, Word Embeddings). Die Swiss-AL-Workbench basiert auf RShiny und erstellt Abfragen in Echtzeit mit Hilfe der IMS Corpus Workbench.

Auf der Methodenmesse der IDS-Jahrestagung 2022 sollen sowohl das Korpus als auch die verfügbare Workbench vorgestellt werden.

Link: <https://swiss-al.linguistik.zhaw.ch/rsconnect/workbench/>

Literatur:

Dreesen, P., & Stücheli-Herlach, P. (2019). Diskurslinguistik in Anwendung. Ein transdisziplinäres Forschungsdesign für korpuszentrierte Analysen zu öffentlicher Kommunikation. *Zeitschrift für Diskursforschung*, 7(2), 123–162.

Krasselt, J., Fluor, M., Rothenhäusler, K., & Dreesen, P. (2021). A workbench for corpus linguistic discourse analysis. In D. Gromann, G. Sérasset, T. Declerck, J. P. McCrae, J. Gracia, J. Bosque-Gil, F. Bobillo, & B. Heinisch (Hrsg.), *3rd conference on language, data and knowledge (LDK 2021)* (Bd. 93, S. 26:1-26:9). Schloss Dagstuhl – Leibniz-Zentrum für Informatik. <https://doi.org/10.4230/OASlcs.LDK.2021.26>

Krasselt, J., Dreesen, P., Fluor, M., Mahlow, C., Rothenhäusler, K., & Runte, M. (2020). Swiss-AL: A Multilingual Swiss Web Corpus for Applied Linguistics. *Proceedings of the 12th Language Resources and Evaluation Conference*, 4138–4144. <https://www.aclweb.org/anthology/2020.lrec-1.510/>

Liste Lamas, E. (2018). "Wort des Jahres Schweiz" als Ergebnis systematischer Diskursmessung im mehrsprachigen Kontext. *Text und Diskurs*. 2018(11), S. 503-516. <https://doi.org/10.7311/tid.11.2018.24>