

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

Deutsche Gebärdensprache zugänglich und nutzbar machen: Das Öffentliche DGS-Korpus und das Digitale Wörterbuch der Deutschen Gebärdensprache

Reiner Konrad, Julian Bleicken, Calvin Khan, Amy Isard, Gabriele Langer, Anke Müller & Marc Schulder (Universität Hamburg, Institut für Deutsche Gebärdensprache und Kommunikation Gehörloser)

Das DGS-Korpus-Projekt, ein Langzeitprojekt der Akademie der Wissenschaften in Hamburg, das am Institut für Deutsche Gebärdensprache und Kommunikation Gehörloser der Universität Hamburg durchgeführt wird, hat seit September 2019 ca. 50 Stunden annotierter Gespräche online gestellt. Das Referenzkorpus der DGS, für das 560 Std. Videomaterial zur Verfügung stehen, wird zum Projektende 2023 fertiggestellt, Vorabinträge des korpusbasierten **Digitalen Wörterbuchs der Deutschen Gebärdensprache (DW-DGS)** sind bereits veröffentlicht. Im Öffentlichen DGS-Korpus wird das Videomaterial in drei verschiedenen Portalen präsentiert, um den unterschiedlichen Bedarfen und Interessen der Nutzer:innen zu entsprechen (Jahn et al. 2018, Hanke et al. 2020): **MEINE DGS** für die Sprachgemeinschaft, die sich für die Kultur Gehörloser und deren Sprache interessiert, **MEINE DGS – annotiert** für alle, die ein bestimmtes Forschungsinteresse oder bereits Erfahrung mit annotierten Gebärdensprachdaten haben, und ab Release 4 **MEINE DGS – ANNIS**. Damit steht auch für die DGS ein in der Korpuslinguistik bekanntes Korpuswerkzeug zur Verfügung (Krause 2019, Krause et al. 2016). Bei der Auswahl der Gespräche lag der Schwerpunkt auf Inhalten, die die sozialen und kulturellen Aspekte der Sprachgemeinschaft widerspiegeln sowie die Erfahrungen Gehörloser als sprachliche Minderheit in einer hörenden Mehrheitsgesellschaft.

MEINE DGS präsentiert die Videos mit deutschen Untertiteln sowie einigen Filtern (nach Region, Altersgruppe und Gesprächsformaten) und einer Suchmöglichkeit (Schlagwörter). MEINE DGS – annotiert enthält Online-Transkripte mit Annotationen (Übersetzung, Glossen zur Lemmatisierung der darin enthaltenen Gebärden (Types) und Mundbilder bzw. Mundgestik), eine Types-Liste mit Tokens als KWIC-Konkordanz sowie einen Sachindex der Schlagwörter. Zu jedem Transkript können sowohl die Videos als auch die Annotations- und Metadaten heruntergeladen werden inkl. OpenPose-Daten für computergestützte automatisierte Auswertung (Schulder/Hanke 2020). Das Datenset als solches sowie alle darin enthaltenen Transkripte und Types-Einträge sind über DOIs eindeutig adressierbar. Mit MEINE DGS – ANNIS kann nach Gebärden (Glossen), deren Form (HamNoSys, Hanke 2004) und Kontext sowie nach Text (Übersetzung, Mundbild) gesucht werden. Die verschiedenen Suchkriterien können sowohl untereinander als auch mit Metadaten kombiniert werden.

Im DW-DGS werden die Gebärden als Videos und Micons (moving icons) repräsentiert. Die Datengrundlage der lexikographischen Analyse ist das gesamte DGS-Korpus (Langer et al. 2018a/b). Die Wörterbucheinträge fassen die Ergebnisse zusammen und geben den Nutzer:innen einen strukturierten Überblick u.a. zu Lesarten, Kollokationen und semantischen

Präferenzen sowie Formvarianten und Variation nach Region und Altersgruppen. Hinzu kommen authentische Verwendungsbeispiele aus dem Öffentlichen DGS-Korpus, die die Gebärde im Kontext zeigen.

DW-DGS und Öffentliches DGS-Korpus sind in mehrfacher Weise miteinander verlinkt. Die Nutzer:innen können von den Informationen im Wörterbuch zu den Videos und annotierten Daten im Öffentlichen DGS-Korpus gelangen als auch von den Types des Öffentlichen DGS-Korpus zu den bereits vorhandenen Wörterbuch-Einträgen (Müller et al. 2020).

In der Präsentation zeigen wir die vielfältigen Zugriffsmöglichkeiten auf das Wörterbuch sowie die Inhalte von Release 4 des Öffentlichen DGS-Korpus, indem wir auf verschiedene Nutzungsszenarien eingehen, vom interessierten Laien zur DGS-Lehre, vom Studium zur Forschung, von DGS-Muttersprachler:innen zu DGS-Lernenden.

Links:

ANNIS: <https://corpus-tools.org/annis/>

DGS-Korpus-Projekt: <http://dgs-korpus.de>

Digitales Wörterbuch DGS – Deutsch <http://dw-dgs.de/>

MEINE DGS: <http://meine-dgs.de>

MEINE DGS – annotiert: <http://ling.meine-dgs.de>

Literatur:

Hanke, Thomas (2004): HamNoSys. Representing sign language data in language resources and language processing contexts. In: Streiter, Oliver / Vettori, Chiara (eds.): Proceedings of the LREC 2004 Workshop on the Representation and Processing of Sign Languages: From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication. Paris: ELRA, S. 1-6.

URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/04001.pdf> (letzter Zugriff: 2022-02-18).

Hanke, Thomas / Konrad, Reiner / Jahn, Elena / Schulder, Marc (2020): Extending the Public DGS Corpus in Size and Depth. In: Efthimiou, Eleni / Fotinea, Stavroula-Evita / Hanke, Thomas / Hochgesang, Julie A. / Kristoffersen, Jette / Mesch, Johanna (eds.): Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. Language Resources and Evaluation Conference (LREC 2020), Marseille, 11-16 May 2020, S. 75-82. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20016.pdf> (letzter Zugriff: 2022-02-18).

Jahn, Elena / Konrad, Reiner / Langer, Gabriele / Wagner, Sven / Hanke, Thomas (2018): Publishing DGS corpus data: Different Formats for Different Needs. In: Bono, Mayumi / Efthimiou, Eleni / Fotinea, Stavroula-Evita / Hanke, Thomas / Hochgesang, Julie / Kristoffersen, Jette / Mesch, Johanna / Osugi, Yutaka (eds.): Workshop Proceedings. 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. Language Resources and Evaluation Conference (LREC), Miyazaki, Japan, 12 May 2018. Paris: ELRA, S. 83-90. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/18018.pdf> (letzter Zugriff: 2022-02-18).

Krause, Thomas (2019): ANNIS: A graph-based query system for deeply annotated text corpora. Berlin (Diss., HU zu Berlin). DOI: [10.18452/19659](https://doi.org/10.18452/19659).

Krause, Thomas / Zeldes, Amir (2016): ANNIS3: A new architecture for generic corpus query and visualization. Digital Scholarship in the Humanities, 31(1). pp. 118-139. DOI: [10.1093/llc/fqu057](https://doi.org/10.1093/llc/fqu057).

Langer, Gabriele / Müller, Anke / Wähl, Sabrina (2018a): Queries and Views in iLex to Support Corpus-based Lexicographic Work on German Sign Language (DGS). In Bono, Mayumi / Efthimiou, Eleni / Fotinea, Stavroula-Evita / Hanke, Thomas / Hochgesang,

- Julie / Kristoffersen, Jette / Mesch, Johanna / Osugi, Yutaka (eds.): Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. 11th International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 12.5.2018. Paris, Frankreich: European Language Resources Association (ELRA), S. 107-114. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/18026.pdf> (letzter Zugriff: 2022-02-18).
- Langer, Gabriele / Müller, Anke / Wähl, Sabrina / Bleicken, Julian (2018b): Authentic Examples in a Corpus-Based Sign Language Dictionary – Why and How. In: Čibej, Jaka et al. (Hrsg.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts (EURALEX 2018), Ljubljana, Slowenien, 17.-21.7.2018. Ljubljana, Slowenien: Ljubljana University Press, S. 483-497. URL: <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/2933-1> (letzter Zugriff: 2022-02-18).
- Müller, Anke / Hanke, Thomas / Konrad, Reiner / Langer, Gabriele / Wähl, Sabrina (2020): From Dictionary to Corpus and Back Again – Linking Heterogeneous Language Resources for DGS. In Efthimiou, Eleni / Fotinea, Stavroula-Evita / Hanke, Thomas / Hochgesang, Julie A. / Kristoffersen, Jette / Mesch, Johanna (eds.): Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. 12th International Conference on Language Resources and Evaluation (LREC 2020), Marseille, Frankreich, 16.5.2020. Paris, Frankreich: European Language Resources Association (ELRA), S. 157-164. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20025.pdf> (letzter Zugriff: 2022-02-18).
- Schulder, Marc, & Hanke, Thomas. (2020, May 11). OpenPose in the Public DGS Corpus (Version 2). DOI: [10.25592/uhhfdm.1866](https://doi.org/10.25592/uhhfdm.1866).