

# DGS-BIGEKO: A Dataset for Hypothetical Emergency Scenarios in German Sign Language

Cristina Luna-Jiménez\*<sup>ID</sup>, Lennart Eing\*<sup>ID</sup>, Daksitha Withanage Don\*<sup>ID</sup>,  
Marco González†<sup>ID</sup>, Fabrizio Nunnari‡<sup>ID</sup>, Pamela Perniss†<sup>ID</sup>, Patrick Gebhard‡<sup>ID</sup>,  
Elisabeth André\*<sup>ID</sup>

\*University of Augsburg

{cristina.luna.jimenez, lennart.eing, daksitha.withanage.don, elisabeth.andre}@uni-a.de

†University of Cologne

{m.gonzalezathenas, pperniss}@uni-koeln.de

‡German Research Center for Artificial Intelligence (DFKI)

{fabrizio.nunnari, patrick.gebhard}@dfki.de

## Abstract

In this article, we describe DGS-BIGEKO, a sign language dataset containing a conversation in a crisis scenario signed by a professional deaf interpreter in German Sign Language (DGS). The dataset comprises 14 sentences with common questions and answers from protocols occurring in emergency call scenarios translated into DGS. Additionally, the dataset contains signs for an additional 108 concepts that are relevant to emergency call scenarios. The dataset is intended to support research in sign language linguistics and sign language machine translation by providing resources in a very specific domain, where no previous resources are available in DGS. The dataset is freely available for research purposes at the following address: <https://doi.org/10.5281/zenodo.18458557>.

**Keywords:** corpus, sign language, emergency scenarios

## 1. Introduction

Since the public release of the first Large Language Models (Brown et al., 2020), Machine Translation (MT) has become a reality for many languages. However, for small and minoritized languages, including sign languages, there is a lack of resources available for their use in Machine Translation (De Coster et al., 2023). For sign languages, documentation and the existence of language resources are linked to the availability of technological resources, in particular, video recording technology and archiving facilities. For example, the ASLLRP Continuous Signing Corpora started in the mid-1990s (MacLaughlin et al., 1996) with the collection of a database for studying American Sign Language (ASL). Similarly, a few years later in 2009, the DGS Corpus Project began the data collection and annotation of one of the largest corpora of German Sign Language (DGS) available nowadays (Hanke et al., 2020). However, resources for sign languages remain relatively limited compared to (large) spoken languages, a gap that is even more evident when focusing on specific sign languages and scenarios. For example, datasets such as RWTH-PHOENIX (Camgoz et al., 2018), aimed at studying sign language recognition in DGS, only contain samples related to weather news. Similarly, How2Sign (Duarte et al., 2021) contains instructional videos in ASL. Although relevant, they highlight the need for more resources with different sign languages and scenarios.

With the aim of contributing to research in sign

language processing and machine translation, the focus of the present article is the DGS-BIGEKO (Deutsche Gebärdensprache- Bidirektionale Gebärdensprach Kommunikation, German Sign Language - Bidirectional sign language communication) dataset. The dataset was designed to capture a realistic bidirectional dialogue scenario centered on emergencies. Its content focuses on question–answer exchanges that are likely to occur during such critical events, reflecting the communicative needs of real-world emergency interactions. The primary objective of the dataset is to support research on communication in DGS in human–computer interactions. Although all the data were recorded by a deaf interpreter, the dialogue structure is intended to reflect a target application scenario in which an animated avatar produces signed questions, while a human participant responds with signed answers. This design enables the dataset to serve as a resource for sign language recognition and generation in the context of human–avatar interaction, while maintaining natural human signing in the recorded data.

The remainder of the paper follows: Section 2 describes the related work, summarizing the existing datasets in DGS, and those related to emergency scenarios. Section 3 introduces the dataset acquisition and recording procedure. The features extracted from the dataset are described in Section 4. Afterwards, Section 5 presents the statistics of the dataset from the annotations in DGS and in spoken German. Finally, Section 6 summarizes the conclusions and future lines.

## 2. Related Work

### 2.1. Corpora and Datasets in DGS

When focusing on specific sign languages and themes, DGS contains several open datasets that have allowed the first studies in MT.

The *RWTH-PHOENIX-Weather 2014* (Koller et al., 2015) dataset focuses on the theme of daily news and weather forecast airings on German public TV. It consists of 386 videos transcribed using gloss notations, and has been widely used as a benchmark for continuous sign language recognition (Koller, 2020). Lately, to support end-to-end sign language translation (Núñez-Marcos et al., 2023), it was extended to include transcriptions of the original German speech in *RWTH-PHOENIX-Weather 2014 T* (Camgoz et al., 2018). While both datasets are extensively used by the computer science community, the linguistic community has also criticized their limited variety in topics and interpretation quality (Müller et al., 2023).

More focused on linguistic research, the *DGS Public Corpus* (Hanke et al., 2020) is one of the largest datasets available in DGS, created with the purpose of documenting DGS in everyday life to enable linguistic research and serve as a historical document. The dataset includes approximately 50 hours of video material, annotated using an annotation scheme of types and glosses/lexemes (Konrad et al., 2020, 2022). The annotation scheme (Konrad et al., 2022) of the DGS Corpus offers an important lexeme-based ID-glossing system (Johnston, 2001), and has become a valuable resource for linguists. The Public Corpus contains one conversation referred to as an emergency scenario; generally, the dialogues center on discussing their experiences and personal narratives rather than on simulating or enacting emergency calls or protocols.

In an effort to bridge the linguistic and computer science communities, a recent addition for DGS is *DGS-Fabeln-1* (Nunnari et al., 2024), a parallel corpus of German text and videos containing German fairy tales interpreted by a deaf DGS signer, totaling approximately one and a half hours of video material recorded from different camera perspectives. However, as mentioned, the dataset and its vocabulary are exclusively related to fairy tales.

Recently, Avramidis et al. (2025) collected more than 525 hours of interpreted DGS from hearing interpreters for the multilingual TUB corpus. However, none of the sources were specialized in emergencies.

Although the previously mentioned datasets are valuable for research in linguistics and computer science, they make clear the limited availability of resources for specific domains, and the lack of resources focusing on emergency scenarios.

### 2.2. Sign Language Datasets for Emergency Scenarios

Although no widely available DGS dialogue resources exist for emergency scenarios, a limited number of sign language corpora and automatic recognition models have been developed for the emergency domain in other sign languages.

Initial datasets considered a limited list of glosses to develop prototypes of technical frameworks. For example, in the work of Sarkar et al. (2026), they collected six signs representing six types of emergencies. This small-scale dataset was used to evaluate a recognizer that received signals from flex sensors and a gyroscope as inputs. Similarly, Adithya and Rajesh (2020) collected a small dataset of videos containing eight different signs in Indian Sign Language. During the collection process of this dataset, the body and face of the signer are purposefully blocked from view, removing an essential part of the signs. Although interesting for research, these datasets are limited in their applicability to pursue further investigations on them. In another order of magnitude in terms of number of samples is the *Myanmar Sign language Corpus*. This corpus contains a subset dedicated to emergencies (MSL4Emergency) (Nwe et al., 2017), containing 558 recordings with transcriptions in Myanmar glosses and translated Myanmar written text. Another resource of note for the emergency scenario is the GLex (König et al., 2008) dictionary in DGS. This dictionary is composed of technical terms from the fields of health and nursing care, closely related to terms and vocabulary that may arise in medical emergency scenarios.

For machine translation purposes, one of the largest sign language datasets for emergency domains is the KETI (Ko et al., 2019) dataset. The corpus contains 14,672 videos for emergencies, signed in Korean Sign Language. It includes 105 sentences produced by 14 deaf signers following an initial agreed sign variation and annotated with glosses. This dataset was employed to train sign language recognition and translation models in the reduced sets of recorded sentences.

## 3. Dataset Creation

The new DGS dataset with dialogues in emergency scenarios was collected following the protocols described in this section.

### 3.1. Scenario and Dialogue Structure

The main aim of establishing the dataset was to research technology for application in bidirectional communication in a hypothetical emergency scenario. In the scenario, it is assumed that a deaf

DGS signer would witness an accident on a German highway and would contact emergency services as the first bystander.

One of the motivations for investigating this scenario is that, currently, there are no available systems that would allow natural and real-time communication in German Sign Language (DGS). The available systems in Germany, e.g. the NORA app (<https://www.nora-notruf.de/de-gs>) instead rely on reading and writing German text. However, German is a second language to many signers, impacting the rapidness and accuracy with which conversations and gathering of information may happen in these critical situations. Additionally, in these situations, it is highly likely that sign language interpreters would not be available to facilitate communication due to the prompt succession of events in an accident.

In order to understand which questions and answers are posed and received in real emergency scenarios via voice calls, we contacted several emergency boards, including firefighters and police officers, who shared anonymous conversations and their experiences during these emergency calls. After examining these conversations, we discarded those that were not related to the first version of the scenario (e.g., a fire accident in a house). Additionally, we searched for protocols and guidelines normally followed in critical and emergency scenarios, and we used synthetic conversations from the LeanScope AI (<https://www.leanscope.ai/en/>) based on Large Language Models to supplement the dataset and extract the most relevant questions and answers expected in the scenario.

From the analysis of the guidelines and conversations, we derived a series of questions and answers that may occur during the call. This baseline dialogue contained questions, aimed to be produced by an avatar, and short answers that the user would likely sign as a response to the system. The structure of the dialogue is designed to be followed by a rule-based dialogue system, in which each question has a limited set of possible answer options. Depending on the selected option, the dialogue system should act in accordance by asking for more specific questions, or requesting a repetition of the last answer of the user. The baseline dialogue workflow is included as supplementary material.

In the dialogue structure, most questions (intended to be produced by the avatar) terminate with an explicit list of possible answers that can be accepted by the system. This structure is meant to be beneficial in time-critical scenarios, such as emergencies, to reduce the cognitive load of the witness and to most efficiently assess information about the accident for planning the required emergency response services.

Finally, in addition to the main sentences for im-

plementing the dialogue system, signs that may arise under emergency scenarios were recorded to complement the dataset of emergency scenarios and to serve as a starting point for future research in the computer science field.

### 3.2. Recording set-up and procedure

Once the initial version of the dialogue was created in German, the dialogue was sent to professional deaf German Sign Language (DGS) interpreters, who agreed on the most appropriate signs and translations to perform for the videos. To record the signs in high resolution (1920x1080) and 50 fps, the recording process was carried out in the studios of the deaf interpreters, using frontal cameras and a green background.

Two types of videos were recorded: 14 *sentence videos* and 14 *vocabulary videos*. Each sentence video consists of two utterances of the dialogue, involving the question, plus the answers associated with that question. However, two videos contained only the beginning of the call and the closing of the call. Vocabulary videos contain between 1 and 30 (7 on average) signs, grouped by topic, for a total of 108 signs.

Posteriorly, all videos were annotated by a hearing DGS interpreter, who labeled the beginning and end points of the signs as well as provided glosses according to the DGS Corpus annotation guidelines (Konrad et al., 2022). Additionally, for sentence videos, the original German text was provided, and subtitles were created in .srt format.

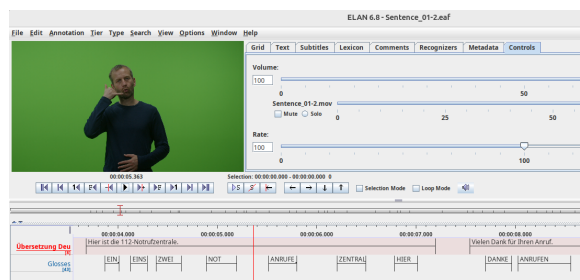
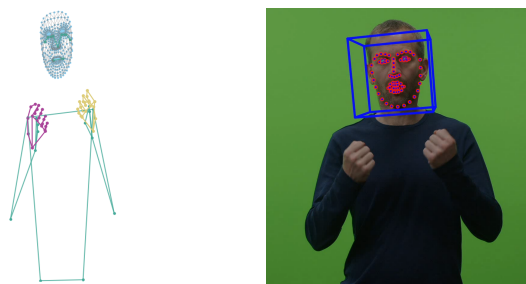


Figure 1: Original videos of one of the annotated sentences in ELAN (Wittenburg et al., 2006).

## 4. Feature Extraction

In this section, the extracted features from the original videos are described. As DGS is rich in facial expressions and mouthing, features related to facial expressions, pose and hand position were provided to allow further research and analysis for the linguistic and computer science communities. Additionally, for completeness, deep vectors were also extracted.

**MediaPipe Landmarks and Blendshapes** The MediaPipe (MP) Holistic framework (Lugaresi et al., 2019) provides a representation of landmarks, each parameterized by normalized spatial coordinates (x,y,z). The pose component comprises 33 landmarks describing major skeletal joints, including the head, shoulders, elbows, wrists, hips, knees, and ankles. The hand component includes 21 landmarks per hand, capturing the wrist and finger joints. Finally, the facial component is represented by a dense mesh of landmarks that delineate facial contours and key features such as the eyes, eyebrows, nose, and lips. In total, the set is composed of 553 landmarks that were extracted employing the Holistic version with the highest complexity level of 2 to ensure the highest possible quality. Figure 2a shows a sample of these landmarks. Additionally, the set of features was complemented with the 52 blendshapes extracted with the FaceLandmarker model after detecting the bounding boxes of the face with the Haar Cascades detector. These features have been employed successfully for Sign Language Recognition and they provide a resource to propose baselines for the DGS-BIGEKO dataset (Luna-Jiménez et al., 2025).



(a) MP landmarks.

(b) OpenFace features.

Figure 2: Features extracted from one of the frames of the video from the vocabulary videos for the DGS sign for 'car' (AUTO).

**OpenFace Facial Features** OpenFace is an open-source toolkit for facial behavior analysis (Baltrusaitis et al., 2018) that allows the extraction of a set of low and high-level features describing facial geometry, appearance, and expression from monocular video. The extracted features include facial landmark locations, head pose estimations, eye gaze direction, and facial action unit (AU) activations and intensities, as well as a histogram of oriented gradients (HOG). Specifically, OpenFace detects a set of facial landmarks that capture the geometric configuration of key facial components such as the eyes, eyebrows, nose, mouth, and jawline, enabling the computation of rigid and non-rigid facial shape parameters. Head pose is represented by three-dimensional translation and rotation parameters (pitch, yaw, and roll), providing

an estimate of the subject's orientation relative to the camera. Eye gaze features are modeled as gaze direction vectors and eye location estimates for each eye. Facial expressions are characterized using the Facial Action Coding System (FACS) (Ekman and Friesen, 1978), with OpenFace estimating both the presence and intensity of a predefined set of action units corresponding to underlying facial muscle activations. Together, these features provide a set of 710 features providing interpretable representations of facial behavior, plus the HOG features. These feature sets have been commonly used in emotion recognition but they contain valuable cues that can apply for sign language recognition given the relevance that non-manuals for disentangle linguistic from emotional cues in the language (Kulshreshtha, 2024).

**DINOv3 Features** DINOv3 (Siméoni et al., 2025) is a vision foundation model trained in a self-supervised paradigm on approximately 1.7 billion images. In addition to tasks such as image classification and semantic segmentation, DINO-type models have been used in the past in sign-language related approaches (Gueuwou et al., 2025b,a; Wong et al., 2024; Liang et al., 2024). For this reason, we provide features from the "ViT-S" type DINOv3 model, distilled from the original DINOv3 model of 7B parameters.

Videos are resized to  $768 \times 1360$  pixels to calculate frame-wise features from their frames. After that, we applied a PCA to compress the number of dimensions per feature token from 384 to 10 dimensions. We provide more information on the structure of the provided files and the code used to generate them in our supplementary materials.

**SAM Synthesized Body Mesh Features** To complement the landmark and image-based representations, we also provide a per-frame *full-body 3D mesh* reconstruction for each video. These features would be specially valuable for avatars or sign language generation studies. Meshes are synthesized from monocular RGB frames using *SAM 3D Body (3DB)* (Yang et al., 2026), as illustrated in Figure 3, a promptable human mesh recovery model designed for robust full-body reconstruction under in-the-wild conditions (e.g., occlusions and uncommon poses). The model estimates parameters of the *Momentum Human Rig (MHR)* (Ferguson et al., 2025), a parametric human body model that represents the body surface together with an explicit skeleton/rig, enabling a consistent kinematic structure across frames.

For each frame, a signer's bounding box is obtained using a person detector and used to crop and resize the input prior to mesh inference. For every frame we store the (i) reconstructed MHR mesh (vertices and faces), (ii) estimated camera parameters, and (iii) underlying MHR parameters

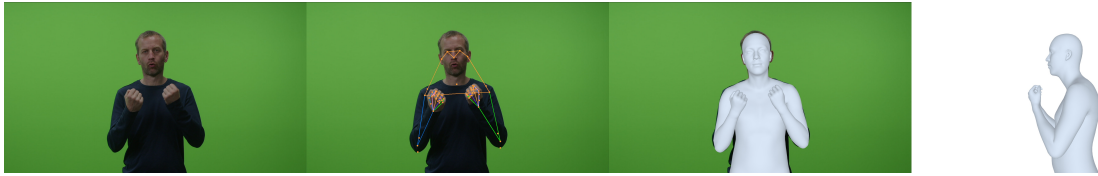


Figure 3: Synthesized mesh from the sign for 'car' (AUTO) from the DGS-BIGEKO dataset.

(pose and shape), enabling both direct rendering and reproducible reconstruction; an example overlay is shown in Figure 3. Finally, we apply temporal smoothing to the recovered pose parameters to reduce jitter while preserving fast articulations relevant for sign language, and we flag frames where mesh estimation fails (e.g., severe truncation or missed detections) with missing values. Accompanying metadata indicates whether results were produced with keypoint prompting or unprompted inference, yielding a compact and physically interpretable representation of full-body motion that complements 2D landmarks and appearance embeddings (e.g., DINOv3, Sec. 4).

## 5. Corpus Analysis

In this section, an analysis of the glosses and the German text is provided.

### 5.1. German Sign Language

The German Sign Language videos (single signs and sentences) were annotated with glosses according to the conventions of the DGS Corpus (Konrad et al., 2022) with the annotation tool of Wittenburg et al. (2006) (ELAN). The ID-glossing conventions established by the DGS Corpus were followed where possible. That is, for signs in the dataset that appear in the same form in the DGS Corpus, the gloss used in the DGS Corpus was used. Due to the special topic of the dataset, focusing on emergency scenarios, not all signs, and thus not all glosses, could be found in the DGS Corpus. The decision to follow the DGS Corpus glossing conventions allows comparability and use across datasets. In addition to the glosses, a translation into German was included on an additional tier in ELAN, as can be seen in Figure 1. In this version of the dataset, annotations included only glosses and translations; non-manuals and features of grammar (e.g., referential locations in space) remain for future work.

In total, the videos of the sentences contain 111 different sign types with an average repetition of  $\mu = 1.95 \pm 1.93$ , resulting in a total of 217 sign tokens annotated.

### 5.2. German

Considering the German sentences from the original dialogue, the sentences contain an average of 4 to 5 words. This reflects the thematic focus and the nature of the task, where signers were instructed to be clear and explicit. In total, the sentences consist of a vocabulary of 269 word types and 370 tokens, employing the Moses tokenizer (Koehn et al., 2007). Despite the limited number of sentences, consisting of 14 utterances in the dialogue, the Type-Token Ratio (TTR) of 56.13 indicates a relatively large linguistic diversity in the words contained in the dataset. The percentage of short words (35.68% of words with less than 4 characters) in the sentences indicates intended direct interactions in the dialogue. Similarly, the Flesh Reading Ease of 81.80 indicates clear and easy-to-read sentences, suitable for emergency scenarios. The complete list of statistics is shown in Table 1, following similar metrics employed in previous corpus (Nunnari et al., 2024).

Metrics for Sentences	Value
# Annotations	60
# Unique Annotations	53
# words	269
# tokens	370
# characters	1,493
Avg. words in sentence	4.20
% hapax legomena	41.64
TTR	56.13
% Short words (<4 characters)	35.69
Avg. Flesh Reading Ease	81.80
Avg. Wiener Sachtextformel	3.48

Table 1: Statistics for the German texts.

The videos of single vocabulary items include signs related to the topics: means of transport in an accident, cities and places, emergencies, number of people, clarification statements, plus additional questions and answers that may arise in these scenarios. Moreover, generic signs that may occur in any situation were also recorded, such as the finger alphabet, cardinal and ordinal numbers from 0 to 10, and personal pronouns. Overall, this vocabulary set consists of 108 additional signs.

## 6. Conclusion

Sign Language Translation has become an open and active research field. However, from a technological perspective, a main limitation is the lack of high-quality data in different contexts, scenarios, and languages. To enrich the available data in DGS, we propose the DGS-BIGEKO dataset. In addition to the video and feature files, we provide guidelines for dialogues that may occur in an emergency scenario, and propose a dialogue of possible questions and answers. The dialogue was translated into German Sign Language (DGS) by an expert deaf interpreter, deciding which signs would be most appropriate and understandable for the situation and recording them in HD videos. In total, the dialogue contained 14 utterances of questions and answers. Additionally, a list of 108 signs was recorded as part of the vocabulary to complement the case study of the emergency scenario. From the videos, we performed feature extraction employing different models, from established libraries such as MediaPipe for landmarks extraction, to deep features extracted from the DINOv3 models. This dataset can serve as the basis for further developments on this scenario from a computer science perspective, to enhance the sign generation and production, as well as the communication protocols to create dialogue systems based on avatars. For future research, we plan to acquire a larger number of subjects to have more variability in terms of dialects and signing conventions in DGS, as well as to enlarge the vocabulary and sentences that may arise in other emergencies. Additionally, we will provide baselines on Isolated Sign Language Recognition.

## 7. Limitations

While the dataset provides a valuable initial resource for translation-based communication in German Sign Language (DGS) within emergency scenarios, it has certain limitations. The current version contains a limited number of samples and is recorded from a single signer, which restricts variability in signing style, speed, and expressiveness. In addition, the degree to which some of the DGS sentences reflect structures associated with LBG (Lautsprachbegleitende Gebärden; “speech-accompanying signs”) needs to be checked; the original translation task (from German to DGS) may have influenced this. LBG reflects the structure of German and is distinct from DGS. These sentences and any videos used in Sign Language Translation research must be checked and reviewed by more deaf DGS signers in addition to our deaf interpreters before their possible use in commercial applications.

Despite this, the recorded sequences consistently include the expected keywords and key phrases that models should be able to detect and recognize, ensuring that the dataset remains useful for initial model training and evaluation in hypothetical emergency scenarios.

## 8. Ethical Considerations

The people recorded have given informed consent for the use of their data for the purposes of the dataset creation. Crucially, this dataset was recorded for research purposes, and it is not intended to train models to directly translate in emergency scenarios, but to provide guidelines for the protocol followed during the data acquisition to continue future research in hypothetical emergency scenarios. Given the critical and sensitive domain of emergency scenarios, it is important to highlight that more research is still required to have these systems in production covering more sentences, scenarios, and data acquisition since only high accurate systems could substitute currently existing text and button-based solutions. In the framework of the project, we explored a first version of what it is required and the limitations and issues that should be considered from all the parts of the systems, being the recognition of signs one of the most critical and data-dependent.

## 9. Acknowledgments

This research is funded by the German Ministry for Education and Research (BMBF) through the BIGEKO project, grant number 16SV9094. Additionally, we would like to acknowledge and thank the contribution of yomma GmbH ([www.yomma.de](http://www.yomma.de)) employees for helping with the recordings and translations. We also thank Centigrade GmbH (<https://www.centigrade.de/en/>) and Charamel GmbH ([www.charamel.de](http://www.charamel.de)) for their support in the guidelines and dialogue flow creation.

## 10. Bibliographical References

- V. Adithya and R. Rajesh. 2020. Hand Gestures for Emergency Situations: A Video Dataset based on words from Indian Sign Language. *Data in Brief*, 31:106016. doi:10.1016/j.dib.2020.106016.
- Eleftherios Avramidis, Vera Czehmann, Fabian Deckert, Lorenz Hufe, Aljoscha Lipski, Yuni Amaloea Quintero Villalobos, Tae Kwon Rhee, Mengqian Shi, Lennart Stölting, Fabrizio Nunnari, and Sebastian Möller. 2025. The

- TUB Sign Language Corpus Collection. In *Adjunct Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents*, pages 1–7, Berlin, Germany. ACM. doi:10.1145/3742886.3756709.
- Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, page 59–66, Xi'an, China. IEEE Press. doi:10.1109/FG.2018.00019.
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20*, Online. Association for Computing Machinery. URL: <https://dl.acm.org/doi/abs/10.5555/3495724.3495883>.
- Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural Sign Language Translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7784–7793, Salt Lake City, UT, USA. IEEE Computer Society. URL: [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Camgoz\\_Neural\\_Sign\\_Language\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Camgoz_Neural_Sign_Language_CVPR_2018_paper.html).
- Mathieu De Coster, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. 2023. Machine Translation from Signed to Spoken Languages: State of the art and Challenges. *Universal Access in the Information Society*, 23(3):1305–1331. doi:10.1007/s10209-023-00992-1.
- Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i Nieto. 2021. How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2734–2743, Online. IEEE Computer Society. URL: <https://doi.ieeecomputersociety.org/10.1109/CVPR46437.2021.00276>.
- Paul Ekman and Wallace V. Friesen. 1978. Facial Action Coding System. Institution: American Psychological Association. doi:10.1037/t27734-000.
- Aaron Ferguson, Ahmed A. A. Osman, Berta Bescos, Carsten Stoll, Chris Twigg, Christoph Lassner, David Otte, Eric Vignola, Fabian Prada, Federica Bogo, Igor Santesteban, Javier Romero, Jenna Zarate, Jeongseok Lee, Jinyoung Park, Jinlong Yang, John Doublestein, Kishore Venkateshan, Kris Kitani, Ladislav Kavan, Marco Dal Farra, Matthew Hu, Matthew Cioffi, Michael Fabris, Michael Ranieri, Mohammad Modarres, Petr Kadlec, Rawal Khirodkar, Rinat Abdrashitov, Romain Prévost, Roman Rajbhandari, Ronald Mallet, Russell Pearsall, Sandy Kao, Sanjeev Kumar, Scott Parrish, Shouo-I Yu, Shunsuke Saito, Takaaki Shiratori, Te-Li Wang, Tony Tung, Yichen Xu, Yuan Dong, Yuhua Chen, Yuanlu Xu, Yuting Ye, and Zhongshi Jiang. 2025. MHR: Momentum Human Rig. Preprint. arXiv:2511.15586.
- Shester Gueuwou, Xiaodan Du, Greg Shakhnarovich, and Karen Livescu. 2025a. SignMusketeers: An Efficient Multi-Stream Approach for Sign Language Translation at Scale. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 22506–22521, Vienna, Austria. Association for Computational Linguistics. doi:10.18653/v1/2025.findings-acl.1157.
- Shester Gueuwou, Xiaodan Du, Greg Shakhnarovich, Karen Livescu, and Alexander H. Liu. 2025b. SHuBERT: Self-Supervised Sign Language Representation Learning via Multi-Stream Cluster Prediction. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 28792–28810, Vienna, Austria. Association for Computational Linguistics. doi:10.18653/v1/2025.acl-long.1397.
- Thomas Hanke, Marc Schuler, Reiner Konrad, and Elena Jahn. 2020. Extending the Public DGS Corpus in Size and Depth. In *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*, pages 75–82, Marseille, France. European Language Resources Association (ELRA). URL: <https://aclanthology.org/2020.signlang-1.12/>.
- Trevor Johnston. 2001. "The Lexical Database of Auslan (Australian Sign Language)". *Sign Language & Linguistics*, 4(1-2):145–169. doi:10.1075/sll.4.12.11joh.

- Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2019. Neural Sign Language Translation Based on Human Keypoint Estimation. *Applied Sciences*, 9(13):2683. doi:10.3390/app9132683.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, page 177–180, Prague, Czech Republic. Association for Computational Linguistics. URL: <https://aclanthology.org/P07-2045/>.
- Oscar Koller. 2020. Quantitative Survey of the State of the Art in Sign Language Recognition. arXiv Preprint. arXiv:2008.09918.
- Oscar Koller, Jens Forster, and Hermann Ney. 2015. Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems handling Multiple Signers. *Computer Vision and Image Understanding*, 141:108–125. Pose & Gesture. doi:10.1016/j.cviu.2015.09.013.
- Lutz König, Susanne König, Reiner Konrad, and Gabriele Langer. 2008. Corpus-based Sign Dictionaries of Technical Terms – Dictionary Projects at the IDGS in Hamburg. In *Proceedings of the LREC 2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pages 94–100, Marrakech, Morocco. European Language Resources Association (ELRA). URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/08017.pdf>.
- Reiner Konrad, Thomas Hanke, Gabriele Langer, Dolly Blanck, Julian Bleicken, Ilona Hofmann, Olga Jeziorski, Lutz König, Susanne König, Rie Nishio, Anja Regen, Uta Salden, Sven Wagner, Satu Worseck, Oliver Böse, Elena Jahn, and Marc Schulder. 2020. MEINE DGS – annotiert. Öffentliches Korpus der Deutschen Gebärdensprache, 3. Release / MY DGS – annotated. Public Corpus of German Sign Language, 3rd release. Universität Hamburg. Dataset. doi:10.25592/dgs.corpus-3.0.
- Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. Public DGS Corpus: Annotation Conventions / Öffentliches DGS-Korpus: Annotationskonventionen. doi:10.25592/uhhfdm.10251.
- Neha Kulshreshtha. 2024. The Effect of Emotional State on the Position of Wh-signs in Indian Sign Language. In *First Workshop on Expressing Emotions in Sign Languages (ExEmSiLa-2024)*, Universität Hamburg. URL: <https://www.idgs.uni-hamburg.de/forschung/tagungen/expressing-emotions-in-sign-languages/programm/programm-pdfs/kulshreshtha---the-effect-of-emotional-state-on-the-position-of-wh.pdf>.
- Han Liang, Chengyu Huang, Yuecheng Xu, Cheng Tang, Weicai Ye, Juzhe Zhang, Xin Chen, Jingyi Yu, and Lan Xu. 2024. LLaVA-SLT: Visual Language Tuning for Sign Language Translation. Preprint. arXiv:2412.16524.
- Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Perceiving and Processing Reality. In *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019*, pages 3746–3753, Long Beach, CA. Cornell University. URL: [https://mixedreality.cs.cornell.edu/s/NewTitle\\_May1\\_MediaPipe\\_CVPR\\_CV4ARVR\\_Works\\_hop\\_2019.pdf](https://mixedreality.cs.cornell.edu/s/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Works_hop_2019.pdf).
- Cristina Luna-Jiménez, Lennart Eing, Annalena Bea Aicher, Fabrizio Nunnari, and Elisabeth André. 2025. Lightweight Transformers for Isolated Sign Language Recognition. In *Proceedings of the 27th International Conference on Multimodal Interaction, ICMI '25*, page 155–163, Canberra, Australia. Association for Computing Machinery. doi:10.1145/3716553.3750772.
- D MacLaughlin, C Neidle, and RG Lee. 1996. Design Specifications for SignStream™, a Multimedia Database Tool for Language Research. Boston University, Boston. American Sign Language Linguistic Research Project. doi:10.3758/bf03195384.
- Mathias Müller, Zifan Jiang, Amit Moryossef, Annette Rios, and Sarah Ebling. 2023. Considerations for meaningful sign language machine translation based on glosses. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 682–693, Toronto, Canada. Association for Computational Linguistics. doi:10.18653/v1/2023.acl-short.60.
- Fabrizio Nunnari, Eleftherios Avramidis, Cristina España-Bonet, Marco González, Anna Hennes, and Patrick Gebhard. 2024. DGS-Fabeln-1: A Multi-Angle Parallel Corpus of Fairy Tales between German Sign Language and German

- Text. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4847–4857, Torino, Italia. ELRA and ICCL. URL: <https://aclanthology.org/2024.lrec-main.434/>.
- Hlaing Myat Nwe, Hnin Aye Thant, Hnin Wai Wai Hlaing, Nandar Win Min, Ni Htwe Aung, Swe Zin Moe, and Ye Kyaw Thu. 2017. MSL4Emergency. GitHub. URL: <https://github.com/ye-kyaw-thu/MSL4Emergency>.
- Adrián Núñez-Marcos, Olatz Perez de Viñaspre, and Gorka Labaka. 2023. A Survey on Sign Language Machine Translation. *Expert Systems with Applications*, 213:118993. doi:10.1016/j.eswa.2022.118993.
- Saugata Sarkar, Sayan Datta, Soumodeep Banerjee, Dhritiman Mukherjee, and Moumita Basu. 2026. EmSLR Framework: Enhancing Communication for Deaf Individuals During Mass Emergency. In *Human-centered Data Analytics: Technology for Sustainable Development*, pages 89–103, Kolkata, India. Springer Nature Switzerland. doi:10.1007/978-3-032-09153-6\_6.
- Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Santana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. 2025. DINOv3. Preprint. arXiv:2508.10104.
- Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA). URL: <https://aclanthology.org/L06-1082/>.
- Ryan Wong, Necati Cihan Camgoz, and Richard Bowden. 2024. Sign2GPT: Leveraging Large Language Models for Gloss-Free Sign Language Translation. Preprint. arXiv:2405.04164.
- Xitong Yang, Devansh Kukreja, Don Pinkus, Anushka Sagar, Taosha Fan, Jinhyung Park, Soyong Shin, Jinkun Cao, Jiawei Liu, Nicolas Ugrinovic, Matt Feiszli, Jitendra Malik, Piotr Dollar, and Kris Kitani. 2026. SAM 3D Body: Robust Full-Body Human Mesh Recovery. Preprint. arXiv:2602.15989.