DGS-Korpus Project Note AP06-2020-01

Data Statement for the Public DGS Corpus

Authors: Marc Schulder, Dolly Blanck, Thomas Hanke, Ilona Hofmann, Sung-Eun Hong, Olga Jeziorski, Lutz König, Susanne König, Reiner Konrad, Gabriele Langer, Rie Nishio, Christian Rathmann

DOI (latest version): 10.25592/uhhfdm.1745

Releases:

Version 1	2020-06-29	Original publication.
Version 2	2021-11-11	Update to Data Statement Version 2 schema.

Abstract

This data statement of the Public DGS Corpus provides information relevant to judging the nature of the language content of the corpus. It covers how the corpus was curated, specifies the language varieties it covers, and provides demographic information for participants and annotators. It also describes the technical and sociological conditions under which the language data was recorded as well as its topical characteristics. The data statement provides a general overview, supported by references to a variety of publications that cover individual topics in more detail.

Zusammenfassung (German Abstract)

Dieses Data Statement zum Öffentlichen DGS-Korpus stellt Informationen bereit, die dazu dienen, den sprachlichen Inhalt des Korpus einzuschätzen. Es umfasst, wie das Korpus kuratiert wurde, spezifiziert die Sprachvarianten, die es abdeckt, und bietet demographische Angaben zu den Teilnehmenden und den Annotierenden. Desweiteren beschreibt es die technischen und soziologischen Bedingungen, unter welchen die Sprachdaten erhoben wurden, sowie deren thematische Charakteristika. Das Data Statement bietet einen allgemeinen Überblick und unterstützt durch Referenzen zu einer breiten Zahl an Veröffentlichungen, die individuelle Themen in größerer Tiefe behandeln.

Contents

1 Header

Dataset Title: MY DGS – annotated. Public Corpus of German Sign Language

Dataset Curators: 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, Marc Schulder. *Institute of German Sign Language and Communication of the Deaf, Universität Hamburg.*

Dataset Version: 3

Dataset Citation: Hanke et al. (2020), Konrad et al. (2020a)

Data Statement Authors: Marc Schulder, Dolly Blanck, Thomas Hanke, Ilona Hofmann, Sung-Eun Hong, Olga Jeziorski, Lutz König, Susanne König, Reiner Konrad, Gabriele Langer, Rie Nishio, Christian Rathmann. *Institute of German Sign Language and Communication of the Deaf, Universität Hamburg.*

Data Statement Version: 2

Data Statement Citation: Marc Schulder, Dolly Blanck, Thomas Hanke, Ilona Hofmann, Sung-Eun Hong, Olga Jeziorski, Lutz König, Susanne König, Reiner Konrad, Gabriele Langer, Rie Nishio and Christian Rathmann (2021). *Data Statement for the Public DGS Corpus*. Project Note AP06-2020-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.1745

2 Executive Summary

The Public DGS Corpus is an annotated corpus of German Sign Language (DGS) that was designed to be both a linguistic resource and a historic account of the experiences of signers of DGS. It consists of 50 hours of video recordings of native and near-native signers, which are lemmatised, annotated for mouthings/mouth gestures and translated to German and English. The recordings cover a range of elicitation tasks, most of which are free-form dialogues on given topics, while a few tasks focus on a single participant.

3 Curation Rationale

This section provides the curation rationale for the complete DGS Corpus and its publicly available subset, the Public DGS Corpus.

For the DGS Corpus (Section 3.1) we describe the choice of elicitation formats (Section 3.1.1), how the data collection regions were defined (Section 3.1.2), the distribution of participants across these regions (Section 3.1.3) and the role of the contact persons (Section 3.1.4).

For the Public DGS Corpus (Section 3.2) we describe how content was selected from the full corpus (Section 3.2.1), in what formats it is provided to the public (Section 3.2.2) and how its annotation was performed (Section 3.2.3).

3.1 DGS Corpus

The aim of the DGS-Korpus project is to create a reference corpus of German Sign Language (DGS) and to compile a corpus-based dictionary DGS – German. The reference corpus (called simply *DGS Corpus*) was designed to be both a linguistic resource and a historic account of the experiences of signers of DGS. Special attention was paid to actively involving the deaf community and to creating resources that they can actively profit from.[1]

3.1.1 Data Elicitation Formats

To address the various needs of linguistic research, Nishio et al. (2010) chose a total of 20 elicitation tasks, such as discussions of a given topic, free conversation and the retelling of stories originally presented in sign, picture or movie format. These tasks cover new formats as well as some that have been established by other corpus projects. The sequence of tasks for individual recording sessions was carefully planned to make the mix of formats diverting enough for participants to keep them engaged. For a detailed description of the elicitation tasks, see the original work.

As the article was published while data collection was still ongoing, certain details were omitted to prevent spoiling the elicitation material. These details can be found in project note AP02-2009-02 (Langer et al., 2020).[2] For information on the selection of topics for the elicitation tasks, see AP02-2010-02, AP02-2011-01 and AP02-2012-01 (Konrad et al., 2011; Konrad and Wagner, 2012a; Konrad and Wagner, 2012b, respectively).

[1] Regarding our decision to write *deaf* in lowercase throughout instead of differentiating between *deaf* and *Deaf*, see Section 14.1.2.

[2] At the time of writing, certain project notes are only available in German. English translations may be released in the future. DOIs of project notes in the list of references lead to the latest version of each project note and provide access to all available language versions.

3.1.2 Data Collection Regions

Experiences in previous projects have shown that having participants travel to a recording studio in a different region of the country can have an impact on language use during recording (e. g. changes in sign choice). To preserve as much of the "local spirit" of language use as possible, it was decided to use a mobile studio that would be set up in different locations around Germany. Similarly, all persons present during the recording (i. e. participants and moderators) had to be from that region. A person was considered to be from a region if they had lived there for at least the last ten years.

A total of 13 data collection regions was defined, taking into account

- the catchment areas of current and former schools for the deaf,
- state (Bundesland) borders, on account of their influence of the educational setting,
- the former border between East and West Germany,
- suspected dialectal borders, and
- practical considerations such as the required travel time for participants.

The resulting regions are shown in Figure 1. Each region had one studio location. The regions were further subdivided into up to five sub-regions to allow a balanced selection of participants from different parts of the region, separating large metropolitan areas from rural and mixed ones. In the case of the Berlin region, the selection of sub-regions also took into account the historical separation of West Berlin from the remainder of the region.

3.1.3 Participant Distribution

Due to the lack of census data on the deaf population in Germany, the target number of participants per region was based on the distribution of the general population, with a weight of 2 for larger cities to reflect the common (though unconfirmed) observation that Deaf people often prefer to live in larger cities. Together with a fixed minimum of 16 participants per region (to cover four age groups times two genders with at least two participants each), this resulted in a target number of 328 participants. In total, 330 participants were recorded. 327 of these gave permission for their contributions to be released to the public. These 327 are all represented in the Public DGS Corpus. For more information on the demographic distribution and categorisation of participants, see Section 6.

3.1.4 Contact Persons

To facilitate the corpus recordings, a team of 22 contact persons was assembled. These were members of the deaf community located in the various data collection regions. They fulfilled a variety of tasks, such as advertising the project, recruiting participants, arranging a studio location, being a host and moderator during recordings and being a point of contact for participants. Their duties are described in detail in project notes AP02-2009-01 and AP02-2009-02 (S. König et al., 2020; Langer et al., 2020, respectively).

Figure 1: *Map of the 13 data collection regions.*

3.2 Public DGS Corpus

The Public DGS Corpus was created as a representative subset of the DGS Corpus with high quality annotations fit for public release (Jahn et al., 2018). It was originally released in 2018 and has received updates in 2019 and 2020, adding additional recordings, extensions to its annotation and automatically computed pose information (see Hanke et al., 2020). Further updates are planned for the remainder of the project duration.

3.2.1 Content Curation

The Public DGS Corpus was curated to

- be balanced for region, gender, and age,
- include all elicitation tasks, with the exception of "Sign names" (names of individuals, excluded for reasons of data protection) and "Isolated items" (elicitation of individual signs, excluded because it does not involve full utterances and thus does not fit the focus of the public corpus),
- cover a variety of topics,
- cover different styles of signing, and
- include each of the 330 participants at least once (apart from 3 participants who limited their consent to project-internal use of their recordings).

While contents for the Public DGS Corpus were selected to present a balanced and representative sample of the full corpus, the amount of content for individual tasks was balanced differently for the public corpus. Due to its dual purpose of research resource and historic account, a quantitative focus was given to conversational data, such as discussions and narratives of deaf life experiences and culture. More restricted and constructed tasks, such as retellings of presented media, were mainly included to demonstrate the variety of tasks in the full corpus. AP06-2013-01 (Salden and Konrad, 2014) provides further details on how tasks are balanced for the public corpus.

Relative to the remainder of the DGS Corpus, complete annotation and quality assurance of the Public DGS Corpus was prioritised. Details on the quality assurance steps taken prior to its publication are provided in AP05-2017-01 (Konrad and Salden, 2018).

3.2.2 Content Portals

The Public DGS Corpus is made available in two formats, each of which can be accessed through a website.

MY DGS (`www.meine-dgs.de`) is the community portal, intended for the deaf community and those interested in it. It provides a web viewer for watching the recordings of the corpus with optional German subtitles. Recordings of purely research-oriented elicitation tasks are omitted.

MY DGS – annotated (`http://ling.meine-dgs.de`) is the research portal, intended for an international audience of linguistics researchers. It provides access to the fully annotated version of the public corpus (including research-oriented elicitation tasks) both via a web viewer and as downloadable files. All content of the research portal is available in both German and English.[3]

[3]Two exceptions exist regarding the bilingual coverage of the annotation. Mouthings in the English annotation

3.2.3 Annotation

The annotation of the Public DGS Corpus covers translation, lemmatisation and annotation of mouthings and mouth gestures. This section provides a brief summary of the annotation process. For a detailed description of the annotation conventions, see project note AP03-2018-01 (Konrad et al., 2020b).

As a first step, German translations of the recordings were created by contracted sign language translators and interpreters. German translations were kept as close to the DGS utterances as possible to allow their use as an aid to annotators. For more details on the translations, see AP05-2014-01 (Salden and Konrad, 2015).

The remaining steps processing the translations were performed by student co-workers under the guidance of deaf and hearing project members. Translations were split into sentence- or utterance-segments and time aligned to the signed utterances. The students also proofread the translations with the support of deaf project members. For transcripts included in the Public DGS Corpus, English translations were also added. These are rather free translations of the German translations (as opposed to being direct translations of the source DGS utterance) intended to provide access for an international audience. As such they may be more concise than the German ones.

Sign segmentation was performed to identify the exact start and end points of individual signs. These signs were then lemmatised using a double glossing type hierarchy that differentiates between types and their subtypes.[4] German and English glosses were created. Annotators relied on the German glosses, while the English glosses (like the English translations) where added later to facilitate use of the corpus by an international audience.

Mouthings and mouth gestures are labeled separately from signs and contribute to their contextual meaning. Their timespan can cover one or multiple signs. Start and end points of mouthing/mouth gesture labels are anchored on those of the signs they cover, rather than being an exact indicator of the articulation span. Mouthings are annotated as fully realised target words, rather than exactly representing the articulated form. These words follow German spelling conventions, but are differentiated from regular German words by always being written in lowercase (German nouns are capitalised) and in certain cases by being written as only the word stem (with suffixes provided in brackets purely for readability). This is done because word endings are often either not mouthed or not clearly identifiable. Using word stems in such cases avoids making judgements about the intended part of speech when the mouthing in fact provides no such indicators. Mouth gestures are primarily given the label *'[MG]'*, although a small number of other labels exists also (see AP03-2018-01 (Konrad et al., 2020b)).

4 Documentation for Source Datasets

The Public DGS Corpus is the publicly available part of the DGS Corpus. As of release 3 (Konrad et al., 2020a) it covers 50 of the 560 hours of DGS conversations found in the DGS Corpus. All parts of this data statement apply to both the full and public corpus, unless noted otherwise.

are still given in German, as they are linked directly to the visible articulation and translating them would make little sense. The other exception is that an editorial decision was made not to translate or otherwise annotate recordings from the *'Joke'* elicitation task for either portal.

[4]Types are specified by the citation form of a sign (see Section 15). Subtypes represent conventionalised form-meaning relations, e. g. the combination of a polysemous sign type with a disambiguating mouthing. For a more detailed definition, see the section *'Type hierarchy (double glossing)'* of AP03-2018-01 (Konrad et al., 2020b).

5 Language Varieties

The corpus consists of video recordings of natural utterances in German Sign Language (common shortform: DGS; ISO 639-3/BCP-47 tag: gsg), the most commonly used signed language in Germany. Each transcript is furthermore identified by its collection region (see Section 3.1.2). The language use mainly represents free informal signing (see Section 8 for more details).

The annotated data contains translations and sign glosses in German as spoken in Germany (BCP-47: de) and English (BCP-47: en). The English versions are based on the German ones and most translators were L1 German, L2 English users (see Section 7 for details).

6 Participant Demographic

All 330 participants of the DGS Corpus identify as part of the deaf community and use DGS as their main language of communication. The age of language acquisition varies. Preference was given to participants with early language acquisition. During the selection of participants the aim was to include a wide variety of occupational and educational backgrounds. None of the participants have conditions that would result in disordered language use. All demographic information is self-reported.

Section 6.1 describes the collection process for demographic information, while Section 6.2 outlines the demographic distribution of the corpus, based on the aspects that are made publicly available. Section 6.3 provides information on the studio crew (moderators and technicians) that facilitated the recordings.

6.1 Collection of Demographic Information

Demographic information was collected through two questionnaires. The original version of the first questionnaire can be found in AP02-2009-01 (S. König et al., 2020). The second questionnaire and English translations of both questionnaires are due to be released in the near future. The first questionnaire was filled out by potential participants to judge their suitability for the project. The questionnaire covered, among other things, the participant's age, gender[5], occupation, education level, current and previous places of residence as well as contact information. It also contained questions about their language experience, such as when they first acquired DGS language skills, whether there are other deaf people in their immediate environment (parents, partners, friends), whether they have experience teaching DGS and whether they use DGS for performing art.

Potential participants were recruited by contact persons (see Section 3.1.4). A separate committee of project members chose which candidates should be included as participants, based entirely on information from the first questionnaire (excluding the contact information). Members of the selection committee did not have direct contact with the candidates until after their selection.

[5]The questionnaires do not take into account the possibility of non-binary gender identities, as they were designed prior to the widespread public debate of the matter in Germany. None of the participants commented on this issue. While it is possible that there were participants who identified as non-binary at the time but chose not to divulge this, we assume that all participants identified as male or female at the time. It is also worth noting that the questionnaires use the German term *Geschlecht*, which can mean both gender or sex, so participants were free to interpret it either way. See Section 14.1.3 for information on how *Geschlecht* was translated for English texts.

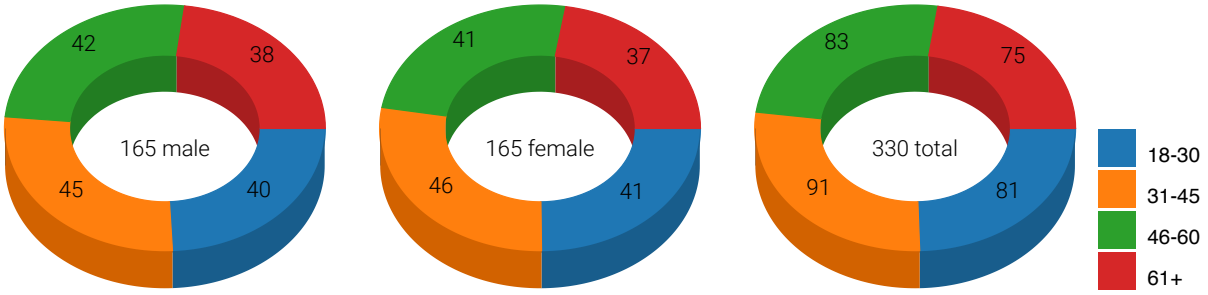

Figure 2: *Distribution of age groups among male, female and all participants of the DGS Corpus.*

The second questionnaire, which participants filled out after they had been selected for the project, contained more detailed versions of the questions from the first questionnaire, as well as additional ones, such as the participant's own degree of deafness, whether they use hearing aids, and their handedness.

The questionnaires were filled out between November 2009 and January 2012 in preparation for the recordings in their respective regions (see Section 3.1.2). The information from the questionnaires was also used when selecting participant pairs for the recordings (see Section 8 for details).

6.2 Public Demographic Information

The Public DGS Corpus provides information on the gender, age group[6] and collection region (see Section 3.1.2) of individual participants. All other information is not released publicly to protect the privacy of the participants.

Overall, the DGS Corpus is fairly well-balanced with respect to age groups, and perfectly with respect to gender (see Figure 2). The distribution of participants across the different collection regions is shown on the map in Figure 3.

6.3 Studio Crew

As described by Hanke et al. (2010b), two studio crew members were present for recordings: a moderator and a technician. During recordings, the moderator was in the studio with the participants (except during the *'free conversation'* elicitation task), while the technician was in an adjacent room from where they monitored the recording.

Both crew members were deaf signers of DGS. Moderators signed the same informed consent forms as the participants and always came from the same geographic region as the participants (see Section 3.1.4). No hearing crew were present during recordings.

7 Annotator Demographic

The annotation of the DGS Corpus involves teams of translators (Section 7.1) and annotators (Section 7.2). Their workflow is based on the annotation guidelines that were developed by members of the project (Section 7.3). The annotation of the DGS Corpus began in 2011 and

[6]There are four age groups: 18–30, 31–45, 46–60 and 61+.

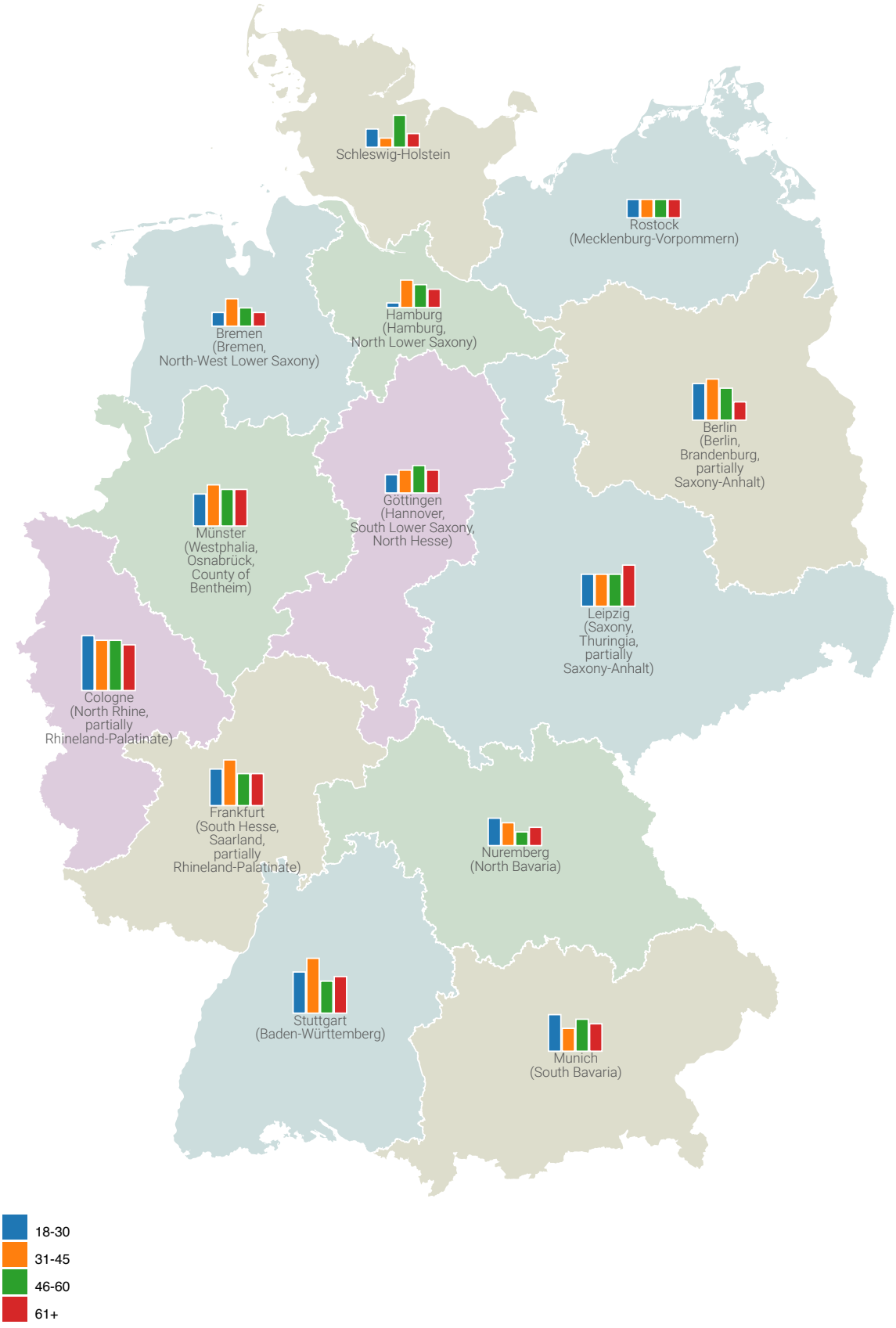

Figure 3: *The distribution of participants across the 13 collection regions, separated by age group.*

is still ongoing. Transcripts published as part of the Public DGS Corpus were completely transcribed and translated before their release.

No consent for the publication of personal information was collected from annotators or translators, so the amount of information that can be released publicly is limited and mainly based on the official job requirements for the respective tasks. The cultural background in all groups is predominantly that of Germany.

For more information on the annotation process, see Section 3.2.3 and the annotation guidelines, published as AP03-2018-01 (Konrad et al., 2020b).

7.1 Translators

The DGS-to-German translations for the DGS Corpus were primarily done by external contractors. The contractors were professional translators and interpreters for that language pair. While the majority of them were hearing, several were deaf or hard-of-hearing. Contracts were publicly advertised and awarded to the lowest bidder among qualified candidates. Over 30 different translators were involved. Translations were done between 2013 and 2020.

For transcripts to be included in the Public DGS Corpus, additional English translations were created before their release. These were not made by contractors, but rather by annotators with good English skills who translated the German translations to English. Most spoke English as a second language. They were primarily students enrolled in English language degrees. While a few of these annotators lacked language skills in DGS (and were therefore employed solely for German-to-English translation) experience showed that the quality of English translations was considerably better when the annotator was able to access the original utterance in addition to the German translation. Annotators with good language skills in both DGS and English were therefore strongly preferred. All additional demographic factors match those of the other annotators. They are described in the next section.

7.2 Annotators

Annotation of the DGS Corpus was started in 2011 and is still ongoing. Over the years it has involved over a hundred different annotators. Annotators are trained and supervised by deaf and hearing project members.

All annotators are required to have good language skills in DGS (DGS III proficiency or better) and to have English language skills. They are also required to have basic knowledge of sign language linguistics, as provided by introductory university courses. They are paid hourly wages at the standard Hamburg University student employment rate.

Most annotators are students of the sign language linguistics and sign language interpreting degrees at Hamburg University in Germany. Due to the gender distribution in those degrees and the resulting candidate pool, most annotators are female. While the majority of annotators are L2 signers of DGS, several others are native or near-native signers. The same applies regarding hearing and deaf cultural backgrounds. Among the (near-)native signers the choice of degree and university affiliation is more mixed.

7.3 Annotation Guideline Developers

The annotation guidelines were developed by the authors of Konrad et al. (2020b). The initial 2011 guidelines were based on others previously developed for a number of corpus-based sign

language dictionaries (cf. L. König et al., 2008). They have been revised throughout the course of the project.

All guideline developers have experience in sign language linguistics and strong competence in DGS. All were involved in the creation of the aforementioned other datasets as developers, annotators or both. One developer is a deaf signer of DGS.

8 Language Production Situation and Text Characteristics

All recordings were made between January 2010 and March 2012. Recordings were made in a mobile studio environment. The studio setup is described in detail in Hanke et al. (2010b).

Rather than inviting participants to Hamburg, the studio was set up in the geographic region of the participants in question (see Section 3.1.2). The exact location of each studio was chosen based on regional familiarity for the participants, good transport links and technical requirements.

Great care was taken to not have any hearing people in the studio during recording, to avoid situations where participants would adjust their language use for them (see Section 6.3).

Recording sessions lasted for 7 hours, including 2 hours of breaks. Two participants sat, facing each other, with a moderator sitting in the background between them. Several cameras were aimed at each participant, but all of them outside the immediate line of sight between the two participants.

Participants were matched to be of a similar age, background and sub-region where possible. Both pairs of the same and of opposite genders were selected. Information on how well participants knew each other was also recorded, although it is not part of the publicly available data.

The moderator explained elicitation tasks for the participants and structured the activity where required. Moderators were trained to facilitate the conversation but not participate in it. For the free conversation task the moderator would leave the room. For more details on the behavioural instructions for moderators, see AP02-2009-02 (Langer et al., 2020).

To have an overview of the session plan, to keep track of time and to present elicitation material to participants, the moderator had use of the software tool *SessionDirector*, which was developed for this purpose by the DGS-Korpus project. A detailed description of *SessionDirector* is available in German as part of AP02-2009-03 (Hong et al., 2020) and in English in AP04-2011-01 (Hanke, 2011).

Genre and topic depended on the elicitation task at hand (see Section 3.1.1). In the Public DGS Corpus dataset, the metadata of each recording specifies the elicitation task and which topics were discussed. Basically all elicitation tasks involve spontaneous dialogues between two people or monologues, bounded more or less strongly by the task. Conversations were mainly informal, despite the studio environment. A limited amount of interaction occurred with the moderator.

The intended audience was, in the immediate sense, the dialogue partner. The moderator was an additional participant who was present but passive. Participants also knew that a technician was monitoring the recording from another room. At a more abstract level, the participants were fully aware they were performing these tasks for the corpus project, whose aims they had been informed of. They were also aware that they would be able to view all their recordings and veto the use or publication of any part. In practice, participants tended to forget about the recording situation and sign freely with their dialogue partner.

9 Preprocessing and Data Formatting

9.1 Anonymisation

All data of the Public DGS Corpus has been anonymised to protect the identities of participants and third parties (Bleicken et al., 2016). This includes the censoring of person names as well as place names, dates or other information that would make individuals uniquely identifiable.

For participants, identifying information was anonymised when it exceeded the amount covered by the consent form. Identifying information about third parties was always anonymised except for individuals of public life who are commonly known in the general German population or the German deaf community.

Information was anonymised in videos by covering it with black rectangles, in annotations and translations by replacing glosses and words with categorial placeholders (e. g. *Name#1*) and in pose information by removing model outputs for the relevant frames.

See Bleicken et al. (2016) for more information on the rationale and approach to video and annotation anonymisation and AP06-2019-01 (Schulder and Hanke, 2020) for information on how this was extended to pose information.

9.2 Pose Information

Pose information was generated automatically for all corpus recordings using OpenPose (Cao et al., 2019). We use the precomputed `BODY_25` body model by Cao et al. (2019) and the hand and face models by Simon et al. (2017). These models were designed for general purpose pose recognition without a special focus on sign language.

OpenPose data for the Public DGS Corpus was published through release 2 (frontal views) and release 3 (side view). Preprocessing steps for this data involved anonymisation (see Section 9.1), limited error correction and introducing a single-file wrapper format for OpenPose data. Regarding error correction, while a thorough manual correction was not feasible, automatic scripts were used to detect and repair some errors. These mainly revolve around the supposed recognition of more people in a frame than were in fact present in the given environment. Such errors were either due to phantom recognitions or fragmentation of one actual person into multiple recognised entities. All preprocessing steps are described in AP06-2019-01 (Schulder and Hanke, 2020). Error correction code has been published alongside it.[7]

10 Capture Quality

Data was recorded from multiple angles using HD cameras. Recordings were made with resolutions of 1080i25 and 720p50. No sound was recorded. The published versions of the recordings are downscaled to 640x360 pixels at 50 frames per second. The original resolution of each recording is identified in its metadata. Participants were recorded sitting in front of a blue background. Participants were asked to wear single colour clothes that provide high contrast to skin colour. For further details on the studio setup, including the camera equipment, see Hanke et al. (2010b).

[7] https://doi.org/10.25592/uhhfdm.8239

11 Limitations

While the German translations of the corpus were created with the aim of being as close as possible to the original DGS utterances, the goal of the English translations was to make the data accessible for an international audience. They are freer and more concise re-translations of the German translations (i. e. not direct translations of the DGS utterance) and as such further removed from the DGS source material.

Sign transcriptions in the Public DGS Corpus identify the citation form of individual signs and mark individual tokens with a * when they differ from this form. The exact nature of the difference is not indicated.

The pose information for the corpus was generated automatically using OpenPose. Keypoint locations are given in a 2-dimensional space and lack depth information. While some recognition errors were corrected before release using automatic post-processing steps (see Section 9), no manual correction was performed.

12 Metadata

12.1 License Conditions

The *MY DGS – annotated* dataset of the Public DGS Corpus is intended for linguistic research. You may download and use the material for this sole purpose. If you publish your research based on this material, please cite the corresponding publications by the DGS-Korpus project. The full license conditions of the dataset can be found at `http://ling.meine-dgs.de/license_en.html`.

12.2 How to Cite

This data statement is primarily a summary of previously published information. As such, when citing this work, please also cite the original work(s) relevant to your text. Where citation space is limited, preference should always be given to the primary resource(s). An overview of relevant works is given in this section.

12.2.1 The Dataset

The research dataset representation of the Public DGS Corpus is released under the name *MY DGS – annotated*. When referring to the dataset in general, please cite both the dataset and its associated publication.

- **Release 1**
  - **Dataset:** Konrad et al. (2018)
  - **Publication:** Jahn et al. (2018)
- **Release 2**
  - **Dataset:** Konrad et al. (2019)
  - **Publication:** Hanke et al. (2020)

- **Release 3**
  - **Dataset:** Konrad et al. (2020a)
  - **Publication:** Hanke et al. (2020)

12.2.2 Peer-reviewed Publications

When referring to specific details of the corpus design, please cite the appropriate publication. These include, but are not limited to:

- **Studio setup:** Hanke et al. (2010b)
- **Elicitation tasks:** Nishio et al. (2010)
- **Segmentation:** Hanke et al. (2012)
- **Annotation of mouth activities:** Hanke (2014)
- **Lemma revision:** Konrad and Langer (2009)
- **Anonymisation:** Bleicken et al. (2016)

12.2.3 Project Notes

In addition to peer-reviewed publications, a lot of information is also published in the form of project notes. (This data statement is such a project note itself.) Project notes should be cited when no peer-reviewed publication covers the information in question.
The collection of all published project notes of the DGS-Korpus project can be found at https://doi.org/10.25592/dgs.korpus.aps. The notes cited in this work are:

- **Contact person manual**
  - **Part 1 (project, promotion, recruitment, studio search):** AP02-2009-01 (S. König et al., 2020)
  - **Part 2 (elicitation tasks, data collection, consent):** AP02-2009-02 (Langer et al., 2020)
  - **Part 3 (moderator software):** AP02-2009-03 (Hong et al., 2020)
- **Session Director moderator software:** AP04-2011-01 (Hanke, 2011)
- **Topics for *'Subject Area'* elicitation task**
  - **Original selection:** AP02-2010-02 (Konrad et al., 2011)
  - **Evaluation and adjustments (2011):** AP02-2011-01 (Konrad and Wagner, 2012a)
  - **Evaluation and adjustments (2012):** AP02-2012-01 (Konrad and Wagner, 2012b)
- **Annotation Conventions:** AP03-2018-01 (Konrad et al., 2020b)
- **Translation:** AP05-2014-01 (Salden and Konrad, 2015)
- **Selection of recordings for public corpus:** AP06-2013-01 (Salden and Konrad, 2014)

- **Formal and content validation steps for corpus release:** AP05-2017-01 (Konrad and Salden, 2018)
- **OpenPose in the Public Corpus:** AP06-2019-01 (Schulder and Hanke, 2020)

13 Disclosures and Ethical Review

13.1 Consent

Informed consent was received from all participants and moderators for the use of their recordings. The consent and licensing procedure is outlined by Hanke et al. (2010a). At the start of their recording sessions, participants were provided with information in DGS and German regarding the purpose of the research, the data collection process, use of data within the project and their options regarding data sharing approval. They were also given the opportunity to ask questions. Participants gave their consent by putting their signature on the German consent form and then signing their consent in DGS. The entire process was recorded.

Regarding their options for data sharing approval, participants could choose whether to approve the sharing of their data for non-commercial educational purposes, other non-commercial purposes and whether their contact details could be shared with other researchers. In each case they could choose to opt out in general, choose on a case by case basis or to delegate the decision to the project.

After data collection all participants were sent a copy of their recording and the opportunity to indicate sections that they would like to be excluded from publication.

13.2 Funding

This document has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the Academies of Sciences and Humanities.

14 Other

14.1 Terminology changes over time

The DGS-Korpus project is a long-term project that began in 2009 and is still ongoing at the time of writing in 2021. Terminology used by the project was always carefully selected to be inclusive and appropriate. However, terminology evolves and in 2021 we revisited our choices on a number of terms, deciding to change them for future corpus and document releases. These changes are described in the following subsections.

14.1.1 *informant* and *participant*

Originally, persons who took part in our recordings (other than moderators) were referred to as *informants* (German: *Informanten*) to indicate that they were the ones contributing linguistic and cultural information.

In 2021 we decided to change our terminology to the more neutral *participants* (German: *Teilnehmende*).[8] This was to avoid association with past colonialist attitudes in linguistics which treated indigenous and minority language users as mere sources of information rather than as equitable partners. It also helps to avoid associations with the espionage and state policing senses of *informant*.

14.1.2 *deaf* and *Deaf*

The following matter only applies to English. As such it is of relevance to our documentation efforts, but did not affect the data collection process, which involved only DGS and German.

For written English, a common practice in the deaf community and sign languages research community is to differentiate between the audiological attribute of deafness and the identity of belonging to the community of deaf people by writing the former as lowercase *deaf* and the latter as capitalised *Deaf* (Pudans-Smith et al., 2019). We originally followed this practice for English texts related to the DGS corpus.

We changed our approach in 2021 to always using lowercase *deaf*. This change was made in light of evolving views among the deaf community and recent research practices by Fisher et al. (2018), Loos et al. (2020) and others. The intention behind this change is 'to avoid identity politics and unintended marginalisation of deaf people who might not fit neatly within arbitrary boundaries regarding those with a certain audiological status and those with a certain cultural status. This choice reflects not a minimisation of culture, but an attempt to be inclusive of the various and individual ways of being deaf.' (Loos et al., 2020, p. 74)

14.1.3 *sex* and *gender*

The following matter only applies to English. As such it is of relevance to our documentation efforts, but did not affect the data collection process, which involved only DGS and German.

The German word *Geschlecht* can refer to both *sex* and *gender*. As part of the participant questionnaires (see Section 6.1) participants were asked to specify their *Geschlecht*, leaving it up to them to interpret the term in their preferred way.

For the creation of English documentation and publications, the project consulted a fellow linguist who is an L1 user of English for recommendations on how to best translate *Geschlecht*. At the time the recommendation was to use *sex* in the given context. In 2021 this choice was revisited and the decision was made to use *gender* instead from now on, as we consider it a better representation of how the questionnaire question can be understood.

15 Glossary

Contact person: A member of the deaf community that supported the data collection process through a number of tasks. See Section 3.1.4 for details.

Citation form: Also referred to as lemma, canonical form or dictionary form. A minimally marked form of a sign or word, as it would be provided in a dictionary entry. For example, in English the citation form of *books* is *book* and that of *annotated* and *annotating* is *annotate*. In the case of sign languages, citation forms usually specify manual and non-manual components, but not mouthings.

[8] Equivalent terminology changes are made to new contributions in DGS and International Sign.

DGS: A commonly used acronym for German Sign Language (ISO 639-3: gsg), based on its German name, *Deutsche Gebärdensprache*.

Gloss Like basically all sign languages, DGS has no commonly used written form. To enable annotation, we follow the common practice in sign language linguistics of using ID glosses. These consist of a gloss name, a number and possibly further markers. Gloss names are expressions in another (written) language that represents approximate lexical translations of the dominant sense of a sign. The index is used to differentiate distinct signs that share the same gloss name. Glosses are aids for linguistic research and should not be mistaken for context-appropriate translations. Sign language ID glosses should not be confused with other linguistic uses of the term *gloss*, such as for brief lexicographic word sense definitions or interlinear glosses.

L1/L2: First language (L1) and second language (L2) refer to the point of language acquisition. Commonly, L1 is used to refer to language proficiency acquired from birth or early in life, while L2 refers to proficiency in language skills acquired later. However, there is no generally accepted definition of the boundary between the two categories, particularly in the area of sign language acquisition.

Pose information: Information on the location of human bodies in an image or video frame, down to individual joints (shoulder, wrist, knee, etc.) and other points of interest like the nose and eyes. Each such point is referred to as a *keypoint*. Pose information is usually gathered through automatic image recognition (as is the case in this dataset) or with the help of motion capture suits. Depending on the software, recognition model or motion capture setup used, aspects of the information can differ, such as the selection of keypoints and whether coordinates are in two- or three-dimensional space.

Token: A token is the specific realisation of a sign or word in an utterance, as opposed to its general concept (cf. **Type**). For example, the English phrase 'word for word' contains three tokens: the first occurrence of *word*, *for* and the second occurrence of *word*.

Type: A type is the general concept of a particular sign or word, as opposed to a specific instance of it being used (cf. **Token**). Commonly (and in this particular dataset), a type is represented as its citation form and is seen to represent all token realisations of the sign or word, irrespective of their particular morphological inflection or other variations. For example, in the English utterance 'one word, two words, all the wooooords', the tokens *word*, *words* and *wooooords* are all instances of the type *word*.

About this document

A data statement is a characterisation of a dataset that provides context to allow developers and users to better understand how experimental results might generalise, how software might be appropriately deployed, and what biases might be reflected in systems built on the software (Bender and Friedman, 2018).

This data statement was written based on the template for the Data Statements Version 2 Schema (Bender et al., 2021). The template was prepared by Angelina McMillan-Major, Emily M. Bender, and Batya Friedman and can be found at http://techpolicylab.uw.edu/data-statements.

The following changes were made to the official format of the schema:

- The start of the document follows the format for project notes of the DGS-Korpus project, adding author list, DOI, release log, abstract and table of contents.
- The name of Section 6 was changed from *"Speaker Demographic"* to *"Participant Demographic"*
- The name of Section 8 was changed from *"Speech Situation and Text Characteristics"* to *"Language Production Situation and Text Characteristics"*

Acknowledgements

An initial draft of this document was created in the course of the LREC 2020 Workshop on Data Statements in NLP.[9] We would like to thank the workshop organisers, Emily M. Bender, Batya Friedman and Angelina McMillan-Major, for providing us with that opportunity. We would also like to thank Zeerak Waseem, Jiagi Li and Stacey Taylor for their feedback during the workshop drafting sessions.

References

References have been grouped by whether they are publications, project notes or datasets.

Publications

Bender, Emily M. and Batya Friedman (2018). 'Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science'. In: *Transactions of the Association for Computational Linguistics* 6. Ed. by Mark Johnson, Ani Nenkova and Brian Roark. Action ed. by Yuji Matsumoto, pp. 587–604. DOI: 10.1162/tacl_a_00041.

Bender, Emily M., Batya Friedman and Angelina McMillan-Major (2021). *A Guide for Writing Data Statements for Natural Language Processing*. University of Washington. URL: https://techpolicylab.uw.edu/wp-content/uploads/2021/10/Data_Statements_Guide_V2.pdf (visited on 11/11/2021).

Bleicken, Julian, Thomas Hanke, Uta Salden and Sven Wagner (2016). 'Using a Language Technology Infrastructure for German in order to Anonymize German Sign Language Corpus Data'. In: *10th International Conference on Language Resources and Evaluation (LREC 2016)* (Portorož, Slovenia). Ed. by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk and Stelios Piperidis. Paris, France: European Language Resources Association (ELRA), pp. 3303–3306. ISBN: 978-2-9517408-9-1. URL: https://aclanthology.org/L16-1526.

Cao, Zhe, Gines Hidalgo Martinez, Tomas Simon, Shih-En Wei and Yaser A. Sheikh (2019). 'OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields'. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.1, pp. 172–186. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2019.2929257.

[9] https://sites.google.com/uw.edu/data-statements-for-nlp/

Fisher, Jami N., Gene Mirus and Donna Jo Napoli (2018). 'STICKY: Taboo Topics in Deaf Communities'. In: *The Oxford Handbook of Taboo Words and Language*. Ed. by Keith Allan. Oxford, UK: Oxford University Press. Chap. 8, pp. 140–159. ISBN: 978-0-19-880819-0. DOI: 10.1093/oxfordhb/9780198808190.013.8.

Hanke, Thomas (2014). 'Annotation of mouth activities with iLex'. In: *9th International Conference on Language Resources and Evaluation (LREC 2014). Proceedings of the LREC2014 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel* (Reykjavik, Iceland). Ed. by Onno Crasborn, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Jette Kristoffersen and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 67–70. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/14029.pdf.

Hanke, Thomas, Sung-Eun Hong, Susanne König, Gabriele Langer, Rie Nishio and Christian Rathmann (2010a). 'Towards Fair Licences for Data from the DGS Corpus Project (SLCN 4)'. Poster presented at the 4th Workshop of the Sign Language Corpus Network (Berlin, Germany, 3rd–4th Dec. 2010). URL: http://dgs-korpus.de/files/inhalt_pdf/Poster_Berlin_Consent.pdf.

Hanke, Thomas, Lutz König, Sven Wagner and Silke Matthes (2010b). 'DGS Corpus & Dicta-Sign: The Hamburg Studio Setup'. In: *7th International Conference on Language Resources and Evaluation (LREC 2010). Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies* (Valletta, Malta). Ed. by Philippe Dreuw, Eleni Efthimiou, Thomas Hanke, Trevor Johnston, Gregorio Martínez Ruiz and Adam Schembri. Paris, France: European Language Resources Association (ELRA), pp. 106–109. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/10047.pdf.

Hanke, Thomas, Silke Matthes, Anja Regen and Satu Worseck (2012). 'Where Does a Sign Start and End? Segmentation of Continuous Signing'. In: *8th International Conference on Language Resources and Evaluation (LREC 2012). Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon* (Istanbul, Turkey). Ed. by Onno Crasborn, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Jette Kristoffersen and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 69–74. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/12028.pdf.

Hanke, Thomas, Marc Schulder, Reiner Konrad and Elena Jahn (2020). 'Extending the Public DGS Corpus in Size and Depth'. In: *12th International Conference on Language Resources and Evaluation (LREC 2020). 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* (Marseille, France). Ed. by Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen and Johanna Mesch. Paris, France: European Language Resources Association (ELRA), pp. 75–82. ISBN: 979-10-95546-54-2. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/20016.pdf.

Jahn, Elena, Reiner Konrad, Gabriele Langer, Sven Wagner and Thomas Hanke (2018). 'Publishing DGS Corpus Data: Different Formats for Different Needs'. In: *11th International Conference on Language Resources and Evaluation (LREC 2018). Proceedings of the LREC2018 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community* (Miyazaki, Japan). Ed. by Mayumi Bono, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie A. Hochgesang, Jette Kristoffersen, Johanna Mesch and Yutaka Osugi. Paris, France: European Language Resources Association

(ELRA), pp. 83–90. ISBN: 979-10-95546-01-6. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/18018.pdf.

König, Lutz, Susanne König, Reiner Konrad and Gabriele Langer (2008). 'Corpus-based Sign Dictionaries of Technical Terms – Dictionary Projects at the IDGS in Hamburg'. In: *6th International Conference on Language Resources and Evaluation (LREC 2008). Proceedings of the LREC2008 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (Marrakech, Morocco). Ed. by Onno Crasborn, Eleni Efthimiou, Thomas Hanke, Ernst D. Thoutenhoofd and Inge Zwitserlood. Paris, France: European Language Resources Association (ELRA), pp. 94–100. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/08017.pdf.

Konrad, Reiner and Gabriele Langer (2009). 'Synergies between transcription and lexical database building: The case of German Sign Language (DGS)'. In: *Proceedings of the Corpus Linguistics Conference (CL 2009)* (Liverpool, United Kingdom). Ed. by Michaela Mahlberg, Victorina González-Díaz and Catherine Smith. Liverpool, United Kingdom: University of Liverpool. URL: http://ucrel.lancs.ac.uk/publications/cl2009/346_FullPaper.doc.

Loos, Cornelia, Jens-Michael Cramer and Donna Jo Napoli (2020). 'The linguistic sources of offense of taboo terms in German Sign Language'. In: *Cognitive Linguistics* 31.1, pp. 73–112. ISSN: 1613-3641. DOI: 10.1515/cog-2018-0077.

Nishio, Rie, Sung-Eun Hong, Susanne König, Reiner Konrad, Gabriele Langer, Thomas Hanke and Christian Rathmann (2010). 'Elicitation methods in the DGS (German Sign Language) Corpus Project'. In: *7th International Conference on Language Resources and Evaluation (LREC 2010). Proceedings of the LREC2010 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies* (Valletta, Malta). Ed. by Philippe Dreuw, Eleni Efthimiou, Thomas Hanke, Trevor Johnston, Gregorio Martínez Ruiz and Adam Schembri. Paris, France: European Language Resources Association (ELRA), pp. 178–185. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/10026.pdf.

Pudans-Smith, Kimberly K., Katrina R. Cue, Ju-Lee A. Wolsey and M. Diane Clark (2019). 'To Deaf or not to deaf: That is the Question'. In: *Psychology* 10.15 (15), pp. 2091–2114. DOI: 10.4236/psych.2019.1015135.

Simon, Tomas, Hanbyul Joo, Iain Matthews and Yaser Sheikh (2017). 'Hand Keypoint Detection in Single Images Using Multiview Bootstrapping'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, Hawaii, USA, pp. 4645–4653. DOI: 10.1109/CVPR.2017.494.

Project Notes

Hanke, Thomas (2011). *SessionDirector*. Project Note AP04-2011-01. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.825.

Hong, Sung-Eun, Sven Wagner, Silke Matthes and Thomas Hanke (2020). *Handbuch für Kontaktpersonen Teil III: SessionDirector*. Project Note AP02-2009-03. Version 4. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.1896.

König, Susanne, Gabriele Langer, Thomas Hanke, Reiner Konrad, Dolly Blanck, Stefan Goldschmidt, Ilona Hofmann, Sung-Eun Hong, Olga Jeziorski, Lutz König, Rie Nishio, Christian Rathmann, Silke Matthes and Satu Worseck (2020). *Handbuch für Kontaktpersonen Teil I: Projekt, Werbung, Informantensuche, Raumsuche*. Project Note AP02-2009-01. Version 3.

Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.1892.

Konrad, Reiner, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio and Anja Regen (2020b). *Öffentliches DGS-Korpus: Annotationskonventionen / Public DGS Corpus: Annotation Conventions*. Project Note AP03-2018-01. Version 3. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.822.

Konrad, Reiner and Uta Salden (2018). *Formale und inhaltliche Prüfschritte zur Korpusveröffentlichung*. Project Note AP05-2017-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.838.

Konrad, Reiner and Sven Wagner (2012a). *Auswertung und Anpassung der Elizitationsmaterialien: Sachthemen*. Project Note AP02-2011-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.811.

— (2012b). *Auswertung und Anpassung der Elizitationsmaterialien: Sachthemen*. Project Note AP02-2012-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.813.

Konrad, Reiner, Sven Wagner and Gabriele Langer (2011). *Erstellung der Elizitationsmaterialien: Sachthemen*. Project Note AP02-2010-02. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.805.

Langer, Gabriele, Susanne König, Thomas Hanke, Reiner Konrad, Dolly Blanck, Stefan Goldschmidt, Ilona Hofmann, Sung-Eun Hong, Olga Jeziorski, Lutz König, Rie Nishio, Christian Rathmann, Silke Matthes and Satu Worseck (2020). *Handbuch für Kontaktpersonen Teil II: Erhebung, Einverständniserklärung*. Project Note AP02-2009-02. Version 4. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.1894.

Salden, Uta and Reiner Konrad (2014). *Auswahl von Aufnahmen für das Teilkorpus*. Project Note AP06-2013-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.840.

— (2015). *Übersetzung*. Project Note AP05-2014-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.836.

Schulder, Marc and Thomas Hanke (2020). *OpenPose in the Public DGS Corpus*. Project Note AP06-2019-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI (latest version): 10.25592/uhhfdm.842.

Datasets

Konrad, Reiner, Thomas Hanke, Gabriele Langer, Dolly Blanck, Julian Bleicken, Ilona Hofmann, Olga Jeziorski, Lutz König, Susanne König, Rie Nishio, Anja Regen, Uta Salden, Sven Wagner and Satu Worseck (2018). *MY DGS – annotated. Public Corpus of German Sign Language, 1st release*. Dataset. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI: 10.25592/dgs.corpus-1.0.

— (2019). *MY DGS – annotated. Public Corpus of German Sign Language, 2nd release*. Dataset. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI: 10.25592/dgs.corpus-2.0.

Konrad, Reiner, 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 (2020a). *MY DGS – annotated. Public Corpus of German Sign Language, 3rd release*. Dataset. Hamburg,

Germany: DGS-Korpus project, IDGS, Hamburg University. DOI: 10.25592/dgs.corpus-3.0.