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 (29 June 2020): Original publication.

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.

Contents

1 Introduction

This document is a data statement for the Public DGS Corpus. To provide the reader with the required context, this introduction provides a brief explanation of what a data statement is (Section 1.1) as well as descriptions of the dataset (Section 1.2) and the annotation process (Section 1.3). The remaining sections

cover the topics specified by Bender and Friedman (2018) for inclusion in long form data statements.[1]

1.1 What is a Data Statement?

Data statements were introduced by Bender and Friedman (2018) as a new professional practice for the description of language datasets. Data statements focus on describing those aspects of a dataset that are relevant for judging the linguistic nature of the data, such as which language varieties it contains, how it was curated and the demographic parameters of both participants and annotators. Their purpose is to enable researchers to better judge the linguistic nature of datasets, helping them to gauge how well information gained from the dataset might generalise and what potential biases might be inherent in the data.

1.2 The Dataset

The Public DGS Corpus is an annotated corpus of German Sign Language (DGS). It is a subset of the DGS Corpus by the DGS-Korpus project[2], chosen for public release with high-quality annotations. 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.

The Public DGS Corpus is made available in two formats, each of which can be accessed through a website. *MY DGS*[3] 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*[4] 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.[5]

[1] Some section titles were adjusted to better match sign language terminology.

[2] `www.dgs-korpus.de`

[3] `http://meine-dgs.de`

[4] `http://ling.meine-dgs.de`

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

The Public DGS Corpus was originally released in 2018 (see Jahn et al., 2018) and has received updates in 2019 and 2020, adding additional recordings as well as extensions to its annotation (see Hanke et al., 2020).

1.3 The 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. English translations of the German translations were added to provide access for an international audience. They are rather free translations and therefore 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 (specified by a citation form) and subtypes (representing a conventionalised form-meaning relation). 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 oriented 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)).

2 Curation Rationale

2.1 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.

2.1.1 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. 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., 2010).[6] 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).

2.1.2 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

[6] 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.

studio that would be set up in different locations around Germany. Similarly, all persons present during the recording (i. e. informants 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 informants.

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 informants 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.

2.1.3 Informant Distribution

Due to the lack of census data on the Deaf population in Germany, the target number of informants 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 informants per region (to cover four age groups times two sexes with at least two informants each), this resulted in a target number of 328 participants. In total, 330 informants 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.

2.1.4 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 informants, arranging a studio location, being a host and moderator during recordings and being a point of contact for informants. Their duties are described

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

in detail in project notes AP02-2009-01 and AP02-2009-02 (König et al., 2011; Langer et al., 2010, respectively).

2.2 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 curated to

- be balanced for region, sex, and age,
- include all elicitation tasks (with the exception of "Sign names" (for reasons of data protection) and "Isolated items"),
- cover a variety of topics,
- cover different styles of signing, and
- include each of the 330 informants at least once (apart from 3 informants who limited their consent to project-internal use of their recordings).

For further details on the curation of the public subset, see AP06-2013-01 (Salden and Konrad, 2014). For details on the quality assurance steps taken prior to its publication, see AP05-2017-01 (Konrad and Salden, 2018).

3 Language Variety

The corpus consists of video recordings of natural utterances in German Sign Language (common shortform: DGS; ISO 639-3 tag: gsg). The annotated data contains translations and sign glosses[7] in German and English. The English versions are based on the German ones. The language use mainly represents free informal signing (see Section 6 for more details).

[7]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 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.

4 Participant Demographic

All 330 informants 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 informants with early language acquisition. During the selection of informants the aim was to include a wide variety of occupational and educational backgrounds. None of the informants have pathological conditions that would result in disordered language use.

4.1 Collection of Demographic Information

Demographic information was collected through two questionnaires, which can be found in AP02-2020-01 (Langer et al., 2020). The first questionnaire was filled out by potential informants to judge their suitability for the project. The questionnaire covered, among other things, the informant's age, sex[8], 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 informants were recruited by contact persons (see Section 2.1.4). A separate committee of project members chose which candidates should be included as informants, 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.

The second questionnaire, which informants 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 informant's own degree of deafness, whether they use hearing aids, and their handedness.

The information from the questionnaires was also used when selecting informant pairs for the recordings (see Section 6 for details).

[8]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. None of the informants commented on this issue. While it is possible that there were informants who identified as non-binary at the time but chose not to divulge this, we assume that all informants identified as male or female at the time.

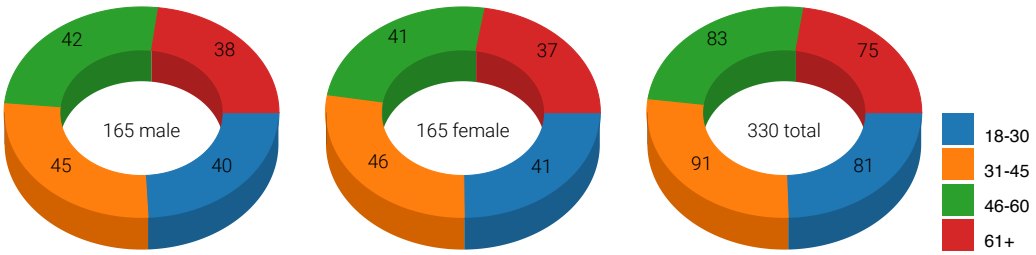

Figure 2: *Distribution of age groups among male, female and all informants of the DGS Corpus.*

4.2 Public Demographic Information

The Public DGS Corpus contains 227 informants. For each informant it identifies their sex, age group[9] and the region they live in (see Section 2.1.2). All other information is not released publicly to protect the privacy of the informants.

Across regions, the age groups are fairly well-balanced with respect to age groups, and perfectly with respect to sex (see Figure 2).[10] The map in Figure 3 shows the distribution of informants across the different regions.

5 Annotator Demographic

The annotation of the DGS Corpus, including the Public DGS Corpus, involved teams of translators (Section 5.1) and annotators (Section 5.2). Their workflow was based on the annotation guidelines that were developed by members of the project (Section 5.3). We also provide information on the studio crew (moderators and technicians) that facilitated the recordings (Section 5.4).

Moderators signed the same informed consent forms as the participants. For all other groups described in this section, no consent for the publication of personal information was collected, so the amount of information that can be released publicly is limited. The majority of information is therefore 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 workflow see the annotation guidelines, published as AP03-2018-01 (Konrad et al., 2020b). For information on the studio setup, see Hanke et al. (2010).

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

[10]These statistics include the three informants who were omitted from the Public DGS Corpus.

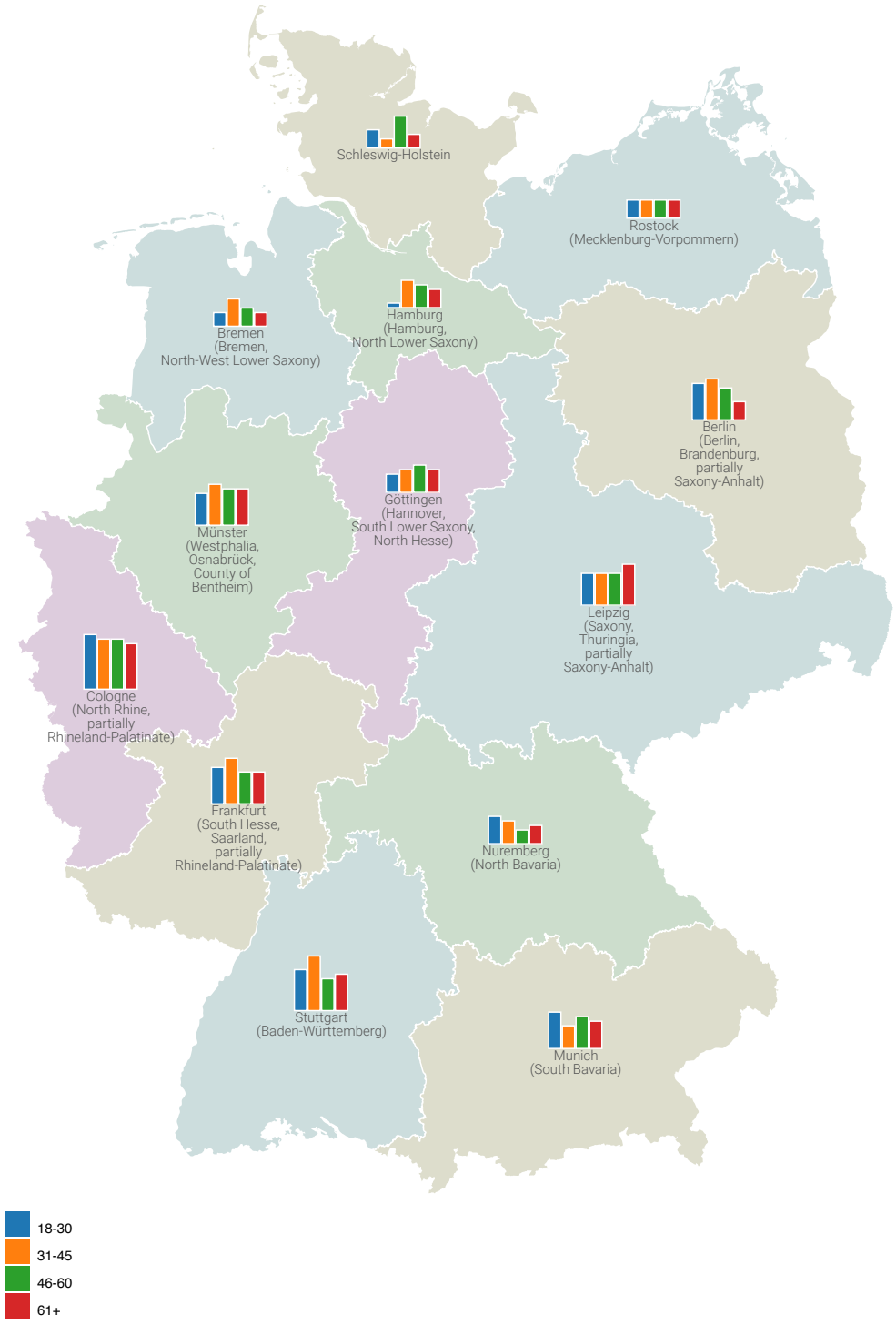

Figure 3: *The distribution of informants across the 13 collection regions, separated by age group.*

5.1 Translators

The DGS-to-German translations for the DGS Corpus were made by external contractors. The contractors were professional translators and interpreters for that language pair.

The English translations were not made by the contractors. Rather, annotators with good English skills translated the German translations to English. These 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.

5.2 Annotators

All annotators were required to have good language skills in DGS (DGS III proficiency or better) and to have English language skills. There were also required to have basic knowledge of sign language linguistics, as provided by introductory university courses.

Most annotators were 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 were female.

While the majority of annotators were L2 signers of DGS, several others were 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 was more mixed.

5.3 Annotation Guideline Developers

The annotation guidelines were developed by Konrad et al. (2020b). All developers have experience in sign language linguistics and strong competence in DGS. One developer is a Deaf signer of DGS.

5.4 Studio Crew

As described by Hanke et al. (2010), 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 native or near-native signers of DGS. The moderator always came from the same geographic region as the participants (see Section 2.1.4). No hearing people were present during recordings.

6 Language Production Situation

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. (2010).

Rather than inviting informants to Hamburg, the studio was set up in the geographic region of the informants in question (see Section 2.1.2). The exact location of each studio was chosen based on regional familiarity for the informants, 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 informants would adjust their language use for them (see Section 5.4).

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

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

The moderator explained elicitation tasks for the informants 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., 2010).

To have an overview of the session plan, keep track of time and to present elicitation material to informants, the moderator had use of the software tool *SessionDirector*, which was developed for this purpose by the DGS-Korpus project. *SessionDirector* is described in detail in AP04-2011-01 (Hanke, 2011).

Basically all elicitation tasks involve spontaneous dialogues between two people or monologues, bounded more or less strongly by the task. A limited amount of interaction occurred with the moderator. For more information on the elicitation tasks, see Section 2.1.1.

The intended audience was, in the immediate sense, the other informant with whom the dialogue happened. The moderator was an additional participant who was present but passive. Informants also knew that a technician was monitoring the recording from another room. At a more abstract level, the informants 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, informants tended to forget about the recording situation and signed freely with the other informant.

7 Text Characteristics

Genre and topic depended on the elicitation task at hand (see Section 2.1.1). In the Public DGS Corpus dataset, the metadata of each recording specifies the elicitation task and which topics were discussed. Conversations were mainly informal, despite the studio environment.

8 Recording 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 screen. 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. (2010).

9 Other

9.1 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.

9.2 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.

When referring to the *MY DGS – annotated* dataset of the Public DGS Corpus 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)

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. (2010)
- **Elicitation tasks:** Nishio et al. (2010)
- **Segmentation** Hanke et al. (2012)
- **Annotation of mouth activities:** Hanke (2014)
- **Anonymisation:** Bleicken et al. (2016)

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 notes cited in this work are:

- **Contact person manual**
  - **Part 1 (project, promotion, recruitment, studio search):** AP02-2009-01 (König et al., 2011)
  - **Part 2 (elicitation tasks, data collection, consent):** AP02-2009-02 (Langer et al., 2010)
- **Informant Questionnaires:** AP02-2020-01 (Langer et al., 2020)
- **Session Director:** 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)

For a list of all published project notes, see https://www.sign-lang.uni-hamburg.de/dgs-korpus/arbeitspapiere/.

10 Provenance Appendix

The Public DGS Corpus is the publicly available part of the DGS Corpus. 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.

The contents of the Public DGS Corpus were selected to present a balanced and representative sample of the full corpus. However, the amount of content for individual tasks is balanced differently for the public corpus. This is described in AP06-2013-01 (Salden and Konrad, 2014). 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. Relative to the remainder of the DGS Corpus, complete annotation and quality assurance of the Public DGS Corpus was prioritised.

Acknowledgements

This document was originally drafted in the course of the LREC 2020 Workshop on Data Statements in NLP.[11] 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.

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.

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.

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

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: *Proceedings of the 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, Asunción Moreno, Jan Odijk, and Stelios Piperidis. Paris, France: European Language Resources Association (ELRA), pp. 3303–3306. ISBN: 978-2-9517408-9-1. URL: https://www.aclweb.org/anthology/L16-1526.

Hanke, Thomas (2014). "Annotation of mouth activities with iLex". In: *Beyond the Manual Channel. Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages. 9th International Conference on Language Resources and Evaluation (LREC 2014)* (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.

Hanke, Thomas, Lutz König, Sven Wagner, and Silke Matthes (2010). "DGS Corpus & Dicta-Sign: The Hamburg Studio Setup". In: *Corpora and Sign Language Technologies. Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. 7th International Conference on Language Resources and Evaluation (LREC 2010)* (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.

Hanke, Thomas, Silke Matthes, Anja Regen, and Satu Worseck (2012). "Where Does a Sign Start and End? Segmentation of Continuous Signing". In: *Interactions between Corpus and Lexicon. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. 8th International Conference on Language Resources and Evaluation (LREC 2012)* (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.

Hanke, Thomas, Marc Schulder, Reiner Konrad, and Elena Jahn (2020). "Extending the Public DGS Corpus in Size and Depth". In: *Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages. 12th International Conference on Language Resources and Evaluation (LREC 2020)* (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.

Jahn, Elena, Reiner Konrad, Gabriele Langer, Sven Wagner, and Thomas Hanke (2018). "Publishing DGS Corpus Data: Different Formats for Different Needs". In: *Involving the Language Community. Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages. 11th International Conference on Language Resources and Evaluation (LREC 2018)* (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.

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: *Corpora and Sign Language Technologies. Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. 7th International Conference on Language Resources and Evaluation (LREC 2010)* (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. 178–181.

Project Notes

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

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 (2011). *Handbuch für Kontaktpersonen Teil I: Projekt, Werbung, Informantensuche, Raumsuche*. Project Note AP02-2009-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University.

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, and Reiner Konrad (2020). *Informant Questionnaires*. Project Note AP02-2020-01. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. To be published.

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 (2010). *Handbuch für Kontaktpersonen Teil II: Erhebung, Einverständniserklärung*. Project Note AP02-2009-02. Version 3. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University.

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.

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, 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.