SMILE Swiss German Sign Language Dataset

Sarah Ebling*, Necati Cihan Camgöz†, Penny Boyes Braem*, Katja Tissi*, Sandra Sidler-Miserez*, Stephanie Stoll†, Simon Hadfield†, Tobias Haug*, Richard Bowden†, Sandrine Tornay‡, Marzieh Razavi‡, Mathew Magimai-Doss‡,

*University of Applied Sciences of Special Needs Education (HfH), Schaffhauserstrasse 239, 8050 Zurich, Switzerland
{sarah.ebling, katja.tissi, tobias.haug}@hfh.ch, boyesbraem@fzgresearch.org, sandysidler@gmail.com
†CVSSP, University of Surrey, GU2 7XR, Guildford, UK
{n.camgoz, s.m.stoll, s.hadfield, r.bowden}@surrey.ac.uk
‡Idiap Research Institute, 1920 Martigny, Switzerland
{sandrine.tornay, marzieh.razavi, mathew}@idiap.ch

Abstract

Sign language recognition (SLR) involves identifying the form and meaning of isolated signs or sequences of signs. To our knowledge, the combination of SLR and sign language assessment is novel. The goal of an ongoing three-year project in Switzerland is to pioneer an assessment system for lexical signs of Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*, DSGS) that relies on SLR. The assessment system aims to give adult L2 learners of DSGS feedback on the correctness of the manual parameters (handshape, hand position, location, and movement) of isolated signs they produce. In its initial version, the system will include automatic feedback for a subset of a DSGS vocabulary production test consisting of 100 lexical items. To provide the SLR component of the assessment system with sufficient training samples, a large-scale dataset containing videotaped repeated productions of the 100 items of the vocabulary test with associated transcriptions and annotations was created, consisting of data from 11 adult L1 signers and 19 adult L2 learners of DSGS. This paper introduces the dataset, which will be made available to the research community.

Keywords: Swiss German Sign Language (DSGS), automatic sign language assessment, sign language testing, sign language recognition and analysis, Microsoft Kinect v2, GoPro, L2 acquisition

1 Introduction

Swiss German Sign Language (*Deutschschweizerische Gebärdensprache*, DSGS) has approximately 5,500 Deaf[1] L1 users. In addition, an estimated 13,000 hearing persons use DSGS. Among them are *children of deaf adults* (CODAs), sign language interpreters, teachers, social workers, and persons otherwise interested in the language (Boyes Braem, 2012). With the exception of CODAs, they are often adult L2 learners of DSGS.

DSGS is composed of five dialects that originated in former schools for the Deaf. The differences between the dialects are primarily lexical and pertain, e.g., to semantic fields such as food (distinct signs for regional food items, such as specific breads) and date specifications (distinct signs for weekdays and months) (Boyes Braem, 1983).

The goal of the ongoing three-year SMILE *(Scalable Multimodal Sign Language Technology for Sign Language Learning and Assessment)* project in Switzerland is to pioneer an assessment system for lexical signs of DSGS that relies on sign language recognition (SLR) technology. SLR involves identifying the form and meaning of isolated signs or sequences of signs. While SLR has been applied to sign language learning (Spaai et al., 2005; Huenerfauth et al., 2017), to our knowledge, the combination of SLR and sign language assessment is novel.

The assessment system that is being developed as part of the SMILE project aims to give adult L2 learners of DSGS feedback on the correctness of the manual parameters (i.e., handshape, hand position, location, and movement) of isolated signs they produce. In its initial version, the system will include automatic feedback for a subset of a DSGS vocabulary production test consisting of 100 lexical items. The testing scenario in the project is as follows: Learners are prompted with a DSGS gloss[2] of the sign on a monitor in front of them. They then produce the sign while their production is recorded by a video camera in front of them. Following this, they receive feedback from the automatic assessment system.

State-of-the-art SLR approaches (Camgöz et al., 2017) are based on deep learning (Goodfellow et al., 2016) methods that require vast amounts of data. Therefore, to provide the SLR component of the assessment system with sufficient training samples, a large-scale dataset containing videotaped repeated productions of the 100 items of the vocabulary test with associated transcriptions and annotations was created, the SMILE Swiss German Sign Language Dataset, which consists of data from 11 adult L1 signers and 19 adult L2 signers of DSGS. This is the first DSGS dataset of its kind. The paper at hand introduces the dataset, which will be made available to the research community.

The remainder of this paper is organized as follows: Section 2 introduces existing sign language datasets and corpora. Section 3 describes the process of creating the DSGS dataset: selecting items for the vocabulary production test (Section 3.1), developing the recording software (Section 3.2), carrying out the recordings (Section 3.3),

[1] It is a widely recognized convention to use the upper-cased word *Deaf* for describing members of the linguistic community of sign language users and, in contrast, to use lower-cased *deaf* when describing the audiological state of a hearing loss (Morgan and Woll, 2002).

[2] Sign language glosses are spoken language words used as labels for semantic aspects of signs. Glosses are typically written in upper-case letters.

post-processing, transcribing, and annotating the data (Section 3.4), and distributing the resulting dataset (Section 3.5). Section 4 offers a conclusion and outlook.

2 Related Work

In the context of language, a corpus denotes a "finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration" (McEnery and Wilson, 2001, p. 32), where *text* may refer to original written text, transcriptions of speech, and transcriptions of sign language. The units of interest in the assessment system in our project (cf. Section 1) are not continuous utterances but isolated signs. Transcribed recordings of repeated productions of these signs form a dataset.

Several sign language corpora and datasets exist, some created for the purpose of conducting linguistic analyses, and some to serve as training data for sign language technology systems, e.g., SLR systems. Table 1 provides an overview of different sign language corpora and datasets. Depending on the field of study, researchers prioritized different aspects of data collection. Linguists mainly focused on having large vocabularies to be able to understand and extract underlying rules of sign languages. On the other hand, SLR researchers concentrated on having multiple repetitions of sign samples from different signers to learn distinctive signer-independent representations using statistical machine learning algorithms.

Most SLR methods begin with extracting the upper body pose information, which is a challenging task due to the color ambiguity between the signers and the background (Cooper et al., 2011). With the availability of consumer depth cameras, such as Microsoft Kinect (Zhang, 2012), and real-time pose estimation algorithms (Shotton et al., 2013; Cao et al., 2017), SLR researchers created datasets containing human pose information, which accelerated the field.

Due to the articulated nature of sign languages, datasets which are collected using generic video cameras suffer from motion blur. This limits both the linguistic analysis and SLR algorithms, which try to investigate and learn the manual attributes of signs, respectively. In addition, the estimated pose becomes noisy where the performed signs contain rapid upper body motion. To address this limitation, we used a diverse set of visual sensors including high speed and high resolution GoPro video cameras, and a Microsoft Kinect v2 depth sensor to collect the SMILE dataset.

3 Compilation of the SMILE Swiss German Sign Language Dataset

3.1 Selection of Test Items

As described in Section 1, the assessment system that includes an SLR component in our project is based on a DSGS vocabulary production test consisting of 100 individual signs. In addition, the test features five practice items that are excluded from subsequent processing. The test is aimed at beginning adult L2 learners of DSGS, targeting level A1 of the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2009). Learning materials for some parts of level A1 have been developed for DSGS (Boyes Braem, 2004a; Boyes Braem, 2004b; Boyes Braem, 2005a; Boyes Braem, 2005b). The basis of the development of the DSGS vocabulary production test was a list of glosses of 3,800 DSGS signs taken from these materials.

The work of arriving at a set of 105 test items (100 main items plus five practice items) was carried out by a team of Deaf and hearing sign language researchers and involved both excluding certain categories of signs, similar to what had previously been done for a lexical comparison study involving DSGS (Ebling et al., 2015), and prioritizing specific signs. In particular, signs denoting persons (e.g., CHARLIE-CHAPLIN), organizations (e.g., GALLAUDET), places (e.g., AUSTRALIEN 'AUSTRALIA'), and languages (e.g., DEUTSCH 'GERMAN') were excluded. This was because many of these signs are borrowed from other sign languages, and some are initialized signs, i.e., signs in which the handshape corresponding to the first letter of the spoken language word in the DSGS manual alphabet is used. For example, the sign ASIEN ('ASIA') is produced by combining the letter A from the DSGS manual alphabet with a circular movement.

Body-part signs (e.g., NASE 'NOSE') as well as pronouns (e.g., DU 'YOU [sg.]') were also discarded, as they mostly correspond to indexical (pointing) signs in DSGS. Number signs were removed since they tend to have many variants, particularly numbers greater than ten. For example, there are three variants for the number sign ELF ('ELEVEN') in DSGS. Primarily fingerspelled components were also removed from the list, e.g., signs for the months of the year (such as JANUAR 'JANUARY' consisting of the sign J from the DSGS manual alphabet), as assessing fingerspelling production was not among the core aims of the final test. Signs composed of multiple successive segments were also eliminated; this was because the segments they consisted of were often also contained in the list of 3,800 signs as individual lexemes. For example, the list contained the sign ABENDESSEN ('DINNER') as well as the signs ABEND ('EVENING') and ESSEN ('MEAL'). Signs marked as being old variants were also ignored (e.g., an earlier form of the sign BAUERNHAUS 'FARMHOUSE'), as current-day DSGS learners could not be expected to know them. Like Vinson et al. (2008) and Mayberry et al. (2013), who compiled lists of signs to be used in acceptance/familiarity studies, we excluded productive forms from our list. However, unlike in these studies, our reason for exclusion was that we anticipated it to be hard to elicit the correct forms for productive signs using glosses as prompts. For example, a test taker might not know which form to sign from a gloss like GEHEN-FUSS ('GO-FOOT').

We further removed signs that appeared in less than four of the five DSGS dialects from the list of item candidates to ensure high familiarity of the learners with the signs. Since the items to be selected formed part of a sign production test, our goal was to test production of as many different sign forms as possible. We therefore reduced homonymy in the following way: We identified groups of form-identical signs and for each group gave preference to the sign whose

Study	Language	Research Field	# Items	# Samples	# Signers	Acquisition Tool
The NGT Corpus (Crasborn and Zwitserlood, 2008)	SL of the Netherlands	Linguistic	N/A	15 Hours	92	Video Camera
ATIS (Bungeroth et al., 2008)	Multilingual	Linguistic	292	595 Sentences	N/A	Video Camera
DGS Corpus (Prillwitz et al., 2008)	German SL	Linguistic	N/A	2.25 million Tokens	328	Video Camera
BSL Corpus (Schembri et al., 2013)	British SL	Linguistic	N/A	40000 Lexical Items	249	Video Camera
LSE-SIGN (Gutierrez-Sigut et al., 2015)	Spanish SL	Linguistic	2400	2400 Samples	2	Video Camera
AUSLAN (Johnston, 2010)	Australian SL	Linguistic	N/A	1100 Videos	100	Video Camera
RWTH-BOSTON (Dreuw et al., 2008)	American SL	Linguistic, SLR	483	843 Sentences	4	Video Camera
ASSLVD (Athitsos et al., 2008)	American SL	Linguistic, SLR	3000	12000 Samples	4	Video Camera
Dicta-Sign (Matthes et al., 2012)	Multilingual	Linguistic, SLR	N/A	6-8 Hours (/Participant)	16-18 (/Language)	Video Camera
SIGNUM (von Agris and Kraiss, 2010)	German SL	SLR	450	33210 Sequences	25	Video Camera
CopyCat (Zafrulla et al., 2010)	American SL	SLR	22	420 Phrases	5	Accelerometer & VC
RWTH-PHOENIX-Weather (Forster et al., 2014)	German SL	SLR	1231	6931 Sentences	9	Video Camera
DEVISIGN (Chai et al., 2015)	Chinese SL	SLR	2000	24000 Samples	8	Kinect v1 Sensor
BosphorusSign (Camgöz et al., 2016)	Turkish SL	SLR	636	24161 Samples	6	Kinect v2 Sensor

Table 1: Existing sign language corpora and datasets

Removed:
Name signs: persons, organizations, places, languages
Body-part signs
Pronouns
Number signs
Primarily fingerspelled components
Signs composed of multiple successive segments
Old signs
Productive signs
Signs appearing in less than four of the five DSGS dialects
Homonyms
Signs overlapping with co-speech gestures
Signs with ambiguous German glosses
Signs with occurrence <3 in DSGS corpora
Prioritized:
Signs with concepts in Efthimiou et al. (2009)
Signs for concepts included in all of the following studies: Vinson et al. (2008), Mayberry et al. (2013), and Efthimiou et al. (2009)

Table 2: Item selection for the DSGS vocabulary production test

meaning was contained in a list of 1,000 common sign language concepts (Efthimiou et al., 2009). In cases where several homonyms were contained in this list, we gave preference to the one with the highest overall token count in the small DSGS corpora currently available. We also eliminated signs that overlapped with co-speech gestures, such as SUPER corresponding to a thumbs-up gesture. Chen Pichler (2009) was among the first to point out that gestures represent a "source for phonological transfer" in L2 sign acquisition (p. 39). In this sense, excluding signs that resembled co-speech gestures represented another step towards ensuring that what was being tested was sign language as opposed to spoken language production. Glosses whose underlying German words were semantically ambiguous (e.g., AUFNEHMEN can have the meaning of both *recording* and *including*, LEICHT can denote the concepts *lightweight* and *easy*) were also discarded. We thus tried to ensure that glosses alone would be sufficient as prompts in the test setting. Lastly, we removed signs that occurred fewer than three times in the DSGS corpora available.

From the resulting set, we gave direct preference to signs whose meanings appeared in the list of 1,000 common sign language concepts (Efthimiou et al., 2009) and well as preference to signs that appeared in all three sign/concept lists mentioned previously (Vinson et al., 2008; Mayberry et al., 2013; Efthimiou et al., 2009). Table 2 summarizes the item selection process.

3.2 Recording Software and Setup

To obtain high quality sign samples, we used a diverse set of visual sensors: a Microsoft Kinect v2 depth sensor, two GoPro Hero 4 Black video cameras (one in high speed mode and the other in high resolution mode), and three webcams. The GoPro cameras and the Microsoft Kinect sensor were fitted on a rigid mount. The mount was placed in front of the signer facing the signing space, and three webcams were placed to the left, the right, and the top of the signer to capture the signs performed from different angles. Sample recording output from all of the sensors and our recording setting can be seen in Figure 1 and Figure 2, respectively.

We modified the publicly available BosphorusSign Recording Software due to its user-friendly interface and color-coded multi-view signer-operator interaction scheme, which are described in detail in Camgöz et al. (2016). To synchronize the capture from multiple sensors, we first developed a recording driver to control webcams using EmguCV (Emgu, 2013). In addition, we developed an API for GoPro cameras in C#, which allows the recording software to have access to all of the functionality of the cameras. The interface was modified to give the operator control over the GoPros. The modified BosphorusSign Recording Software interface can be seen in Figure 3.

The recording software allows for capturing video streams from all the sensors simultaneously. Given a recording script, which contains a set of items and their corresponding prompts, the operator starts a recording session by us-

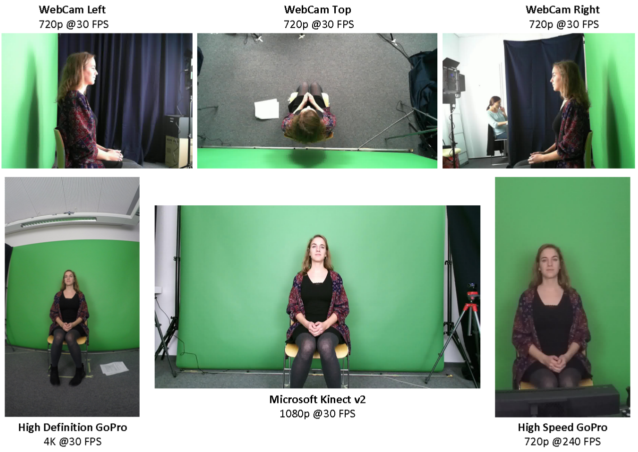


Figure 1: Sample recording output from the video cameras.

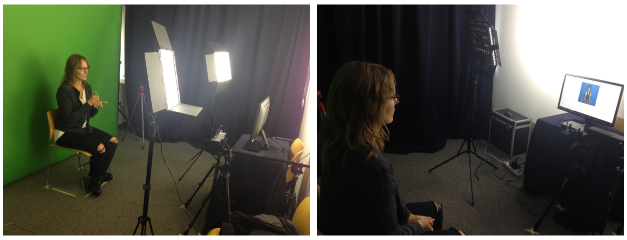
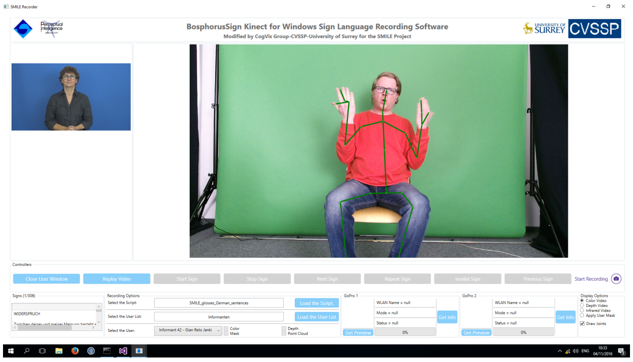


Figure 2: Recording setting.



Signs (1/306)
WIDERSPRUCH
Zwischen deiner und meiner Meinung herrscht

Recording Options
Select the Script: SMILE_glosses_German_sentences
Select the User List: Informanten
Select the User: Informant 42 - Gian Reto Janki
□ Color □ Mask □ Depth □ Point Cloud

GoPro 1
WLAN Name = null
Mode = null
Status = null
0%

GoPro 2
WLAN Name = null
Mode = null
Status = null
0%

Display Options
● Color Video
○ Depth Video
○ Infrared Video
○ Apply User Mask
☑ Draw Joints

ENG 10:33 04/11/2016

Figure 3: Recording interface.

ing the *Start Recording* button (cf. Figure 3). By clicking the *Start Sign* button, the operator indicates the signer to start performing a sign while annotating the beginning of the sample over all streams. When the sign is performed, the operator clicks the *Stop Sign* button to annotate the end of the sample. The operator can then choose to proceed to the next item on the list by using the *Next Sign* button or can request the repetition of the sign by clicking the *Repeat Sign* button. The operator can also use the *Invalid Sign* button to annotate a sign sample as invalid. Once the recording session is finished, the operator presses the *Stop Recording* button and stops the capture on all of the sensors.

3.3 Recording Procedure

The focus of the data collection described in this paper was on obtaining training data for the SLR system. Therefore, in an attempt to reduce the number of instances in which no sign was produced at all, the participating signers were provided with the test items prior to the recordings in the form of a list of glosses with accompanying German example sentences.[3] Table 3 shows a selection of glosses along with context examples. The sentences had been gathered from a DSGS online lexicon[4] and, where necessary, shortened and modified. The rationale behind providing German example sentences in addition to DSGS glosses was to further reduce any semantic ambiguity remaining even after clearly ambiguous glosses had been eliminated in the item selection process (cf. Section 3.1).

Upon recording, participants were asked to perform each sign in three separate passes. The glosses with German example sentences served as prompts for the first two passes, while the prompt for the third pass was a video of a signer performing the sign. The video corresponded to the base form of the sign in a DSGS lexicon (Boyes Braem, 2001).

While the DSGS vocabulary production test is ultimately aimed for use by L2 learners, the goal of the recordings described here was to obtain both L1 and L2 data for training the recognition system. In total 40, 20 L1 and L2 signers each participated in the recordings (due to technical problems, not all recordings were used for the dataset; cf. Section 3.4). The L1 participants were recruited by the Deaf members of our project team; they were native DSGS signers and/or trained DSGS instructors.[5] To recruit L2 participants, a call for participation was released via various channels, such as e-mail, social media, and personal contacts. L2 participants had to have completed four courses in the course framework of the Swiss Deaf Association corresponding to parts of CEFR level A1. Both L1 and L2 participants were asked to complete a background questionnaire prior to the recordings. The background questionnaire was a modified version of a questionnaire developed in the DGS Corpus Project (Hanke, 2017). Participants gave their informed consent for the video recordings and collection of background information as well as usage thereof in the SMILE project.[6] In addition, they were offered the option of giving informed consent for passing the data on to other researchers and to the public via a secure web interface. All but two participants gave their consent for these latter options as well.

L1 participants were paid by the hour, L2 participants were given the choice between getting paid and receiving detailed video feedback on their sign productions from the Deaf members of our project team, who are also trained

[3] In a recent test of the assessment scenario of the project, no sign was produced for 20.56% of all prompts using a nearly identical item set (Haug, 2017).

[4] `https://signsuisse.sgb-fss.ch/` (last accessed September 7, 2017)

[5] Limiting the L1 subject pool to native signers was not an option for DSGS due to the small population of signers upon which one could draw.

[6] A DSGS version of the informed consent had been made available beforehand.

Gloss	Example sentence
ANGESTELLT_1A (‘EMPLOYED_1A’)	Sie ist in einer grossen Firma angestellt. (‘She is employed by a large corporation.’)
THEATER_1A (‘THEATRE_1A’)	Das Theater findet in Basel statt. (‘The theatre play takes place in Basel.’)
WARTEN_1A (‘WAIT_1A’)	Ich warte, bis der Arzt kommt. (‘I am waiting for the doctor to come.’)

Table 3: Glosses and example sentences

DSGS instructors. Since participants were expected to perform 300 signs, it was decided that they should sit on a chair rather than remain standing while signing. In the introductory message signed by a Deaf member of our team and supplemented with German subtitles, participants were told that the goal of the study was to obtain information about natural variation in the production of isolated signs and that following five practice items, they were asked to sign 100 signs three times, the first and second time with glosses as prompts, the third time with a model video of a signer performing the sign. For the third pass, participants were asked to mirror the sign they saw in the video, not repeat a potential dialect variant that they might have produced in the previous two passes. They were told that the order of the signs in the three passes was different and were asked to return to a neutral position after each sign. They were not required to look into a particular camera but rather direct their eye gaze towards the general area of the cameras. Participants were specifically instructed to sign the base forms of the lexical items, not modified versions based on the context evoked in the example sentences. Recordings lasted between 30 and 45 minutes.

3.4 Transcription and Annotation

In the context of sign languages, *transcription* usually refers to the process of providing a written version of signing recorded on video, while *annotation* describes the enhancement of the primary data with additional information, e.g., of linguistic nature. Both steps, transcription and annotation, provide valuable information for an SLR system.

To perform transcription and annotation on the videos obtained through the procedure outlined in Sections 3.2 and 3.3, the videos were postprocessed and imported into iLex, a software tool for creating and analyzing sign language lexicons and corpora (Hanke and Storz, 2008). In iLex, all occurrences of a sign in a transcript (sign tokens) are linked back to their sign type in the lexicon, and changes of the sign type affect all sign tokens in all transcripts.

For each recording, three videos corresponding to three of the six camera perspectives (cf. Figure 1 for an example of all perspectives) were imported and synchronized based on information on the starting and stopping times of the cameras (cf. Section 3.2). Participant and movie metadata were also automatically imported into iLex. One transcript was created for each recording. Based on information on the starting and stopping times of the individual signs (cf. Section 3.2), a tier holding the targeted signs as tags and another tier recording for each tag the pass it belonged to were introduced. The team of Deaf and hearing sign language researchers then manually postcorrected the sign tag boundaries where necessary.

Table 4 shows the transcription/annotation scheme. The scheme consists of twelve tiers. As detailed above, information for the first two tiers, “Pass” and “Target sign”, was automatically imported and manually postcorrected. The team manually annotated information for the remaining tiers for the second pass. If a sign was produced multiple times in this pass (recall from Section 3.2 that self-correction was permitted during the recordings), only the last attempt was considered. A four-eyes principle was observed, i.e., each annotation produced by one annotator was checked by another. In addition, cases for which the annotators were not certain were discussed in weekly group meetings.

The “Sign produced” tier (cf. Table 4) records the glosses of the signs actually performed. “Category of sign produced” is a classification of the productions in this tier into one of six categories:

1. **Same lexeme as target sign:** same meaning, same form
2. **Same lexeme as target sign:** same meaning, slightly different form
3. **Same lexeme as target sign:** same meaning, different form
4. **Same lexeme as target sign:** slightly different meaning, slightly different form
5. **Different lexeme than target sign:** same meaning, different form
6. **Different lexeme than target sign:** different meaning, different form

Instances of Category 1 are sign productions that are identical to the target sign, i.e., to the base form as produced in the model video (cf. Section 3.3). Sign productions assigned to Category 2 have the same meaning as the target sign and a slightly different but acceptable form.[7] For example, the sign SPRACHE_1A (‘LANGUAGE_1A’) might be produced in a slightly different location, resulting in a sign denoted by the *qualified gloss*[8] SPRACHE_1A’loc in the “Sign produced” tier. Members of Category 3 were judged by the annotators to differ clearly and significantly from acceptable variant forms (cf. below for the link between categories and test decisions, i.e., decisions

[7] These instances are sometimes called *allophonic variants*.

[8] Cf. Konrad et al. (2012) for an introduction to qualifiers and qualified glosses.

No.	Tier name	Description
1	Pass	“first”, “second”, or “third”
2	Target sign	Which sign was to be produced?
3	Sign produced	Which sign was actually produced?
4	Category of sign produced	One of six categories
5	Confidence	Confidence of assignment in Tier 4
6	Parameter(s) different	Deviating manual parameter(s)
7	Handedness different	Deviating handedness
8	Hand configuration different	Deviating hand configuration
9	Comment parameter	(free text)
10	Comment style	(free text)
11	HamNoSys (Prillwitz et al., 1989) of target sign	automatically inserted from iLex lexicon
12	HamNoSys of sign produced	HamNoSys notation of sign produced in Tier 3

Table 4: SMILE transcription/annotation scheme

regarding the correctness of the productions). For example, if SPRACHE_1A, which has an open handshape, were to be produced with a closed handshape, this occurrence would be labeled with Category 3 and notated as SPRACHE_1A’hdf in the “Sign produced” tier. Instances of Category 4 are morphophonemic/semantic variants, e.g., modifying SPRACHE_1A from singular to plural, resulting in a slightly different form and slightly different meaning. Sign productions that represent dialect variants are assigned to Category 5, indicating identical meanings but different forms.[9] Sign productions with both an entirely different meaning and form, e.g., productions of the sign BAUM_1A (‘TREE_1A’) for the prompt SPRACHE_1A, are assigned to Category 6.

Table 5 shows the mapping of category assignments to test decisions: Members of Categories 1, 2, 4, and 5 are rated as correct, while members of Categories 3 and 6 are considered incorrect.

A “Confidence” tier (cf. Table 4) records the annotators’ joint confidence of the assignment of Categories 2 and 3 in the “Category of sign produced” tier, with “certain” and “uncertain” as possible values. Our analysis showed that the distinction between permissible variants (Category 2) and incorrect productions (Category 3) of a sign was in some cases especially challenging. Therefore, cases for which the team was uncertain were extracted for presentation to a group of seven outside sign language experts.

For cases in which the sign form produced does not coincide with the target form, a “Parameter different” tier (cf. Table 4) records the deviating parameters, with all cross-combinations of parameters as possible values (“handshape”; “handshape and hand position”; “handshape, hand position, and location”; etc.).

If the number of hands of the target and produced sign differ, this is notated by indicating the handedness of the produced sign as either “one-handed”, “two-handed symmetrical”, or “two-handed dominance”. Similarly, differing hand configuration is recorded along the following values: “one hand next to the other”, “dominant hand on top of non-dominant”, “non-dominant hand on top of dominant”, “dominant hand closer to body”, “dominant hand further away from body”, “one hand crossing the other”, “hands interlocked”, “hands without contact with each other”, and “hands without contact with the body”.

Two tiers allow for comments pertaining to the articulation of the parameters (“Comment parameter”) and to signing style (“Comment style”).

Finally, the last two tiers contain Hamburg Notation System for Sign Languages (HamNoSys) (Prillwitz et al., 1989) notations of the target sign (“HamNoSys target sign”, inserted directly from the lexicon) and the sign produced (“HamNoSys sign produced”). HamNoSys consists of approximately 200 symbols describing the manual parameters hand shape, hand position (with finger direction and palm orientation as sub-parameters), location, and movement. The symbols together constitute a Unicode font.

The second pass of the recordings was completely annotated for 30 transcripts, of which 11 are L1 transcripts and 19 are L2 transcripts. Technical issues were the reason why not all 40 recordings were transcribed/annotated. Figure 4 shows a sample iLex transcript.

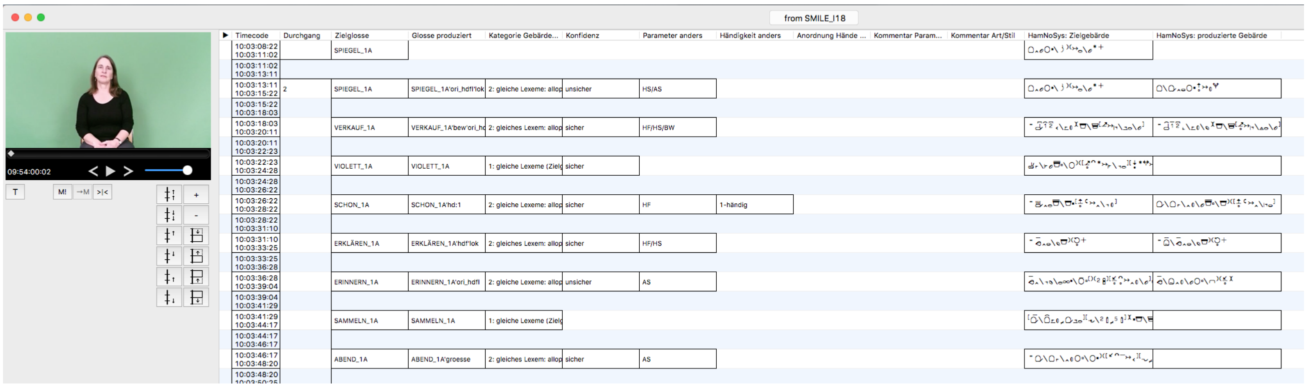
3.5 Distribution

The SMILE Swiss German Sign Language Dataset will be publicly available for academic purposes upon signing an end user license agreement. We will share all of the modalities that were collected using the Microsoft Kinect v2 sensor, namely color videos, depth maps, user masks, and 3D pose information. Other color video streams such as High Definition (4K Resolution) and High Speed (240 frames per second) GoPro and Webcam streams will also be made available. Furthermore, to encourage and to expedite sign language recognition research on our dataset, we will distribute body pose, facial landmarks, and hand pose information extracted using the state-of-the-art deep-learning-based key point detection library OpenPose (Cao et al., 2017). For linguistic research purposes, we will release all of our iLex annotations including sign form and category annotations, which were mentioned in Section 3.4. The contents of the dataset that will be released can be seen in Table 6. The dataset will be available at `https://www.idiap.ch/project/smile`.

[9] Recall from Section 1 that DSGS is composed of five dialects and that the items of the DSGS vocabulary production test at hand are known to appear in at least four of these five dialects.

Category	Same lexeme as target sign?	Same meaning?	Same form?	Test decision
1	yes	yes	yes	correct
2	yes	yes	slightly different	correct
3	yes	yes	no	incorrect
4	yes	slightly different	slightly different	correct
5	no	yes	no	correct
6	no	no	no	incorrect

Table 5: Link between category assignments and test decisions



Timecode	Durchgang	Zielglosse	Glosse produziert	Kategorie Gebärde...	Konfidenz	Parameter anders	Händigkeit anders	Anordnung Hände ...	Kommentar Param...	Kommentar Art/Stil	HamNoSys: Zielgebärde	HamNoSys: produzierte Gebärde
10:03:08:22 10:03:11:02		SPIEGEL_1A										
10:03:11:02 10:03:13:11												
10:03:13:11 10:03:15:22	2	SPIEGEL_1A	SPIEGEL_1A'ori_hdfl'lok	2: gleiche Lexeme: allop	unsicher	HS/AS						
10:03:15:22 10:03:18:03												
10:03:18:03 10:03:20:11		VERKAUF_1A	VERKAUF_1A'bew'ori_hc	2: gleiches Lexem: allop	sicher	HF/HS/BW						
10:03:20:11 10:03:22:23												
10:03:22:23 10:03:24:28		VIOLETT_1A	VIOLETT_1A	1: gleiche Lexeme (Zielg	sicher							
10:03:24:28 10:03:26:22												
10:03:26:22 10:03:28:22		SCHON_1A	SCHON_1A'hd:1	2: gleiche Lexeme: allop	sicher	HF	1-händig					
10:03:28:22 10:03:31:10												
10:03:31:10 10:03:33:25		ERKLÄREN_1A	ERKLÄREN_1A'hdf'lok	2: gleiches Lexem: allop	sicher	HF/HS						
10:03:33:25 10:03:36:28												
10:03:36:28 10:03:39:04		ERINNERN_1A	ERINNERN_1A'ori_hdfl	2: gleiche Lexeme: allop	unsicher	AS						
10:03:39:04 10:03:41:29												
10:03:41:29 10:03:44:17		SAMMELN_1A	SAMMELN_1A	1: gleiche Lexeme (Zielg								
10:03:44:17 10:03:46:17												
10:03:46:17 10:03:48:20		ABEND_1A	ABEND_1A'groesse	2: gleiches Lexem: allop	sicher	AS						
10:03:48:20 10:03:50:25												

Figure 4: Sample transcript in iLex

Modality	File Type	Resolution	Content
Kinect Color Video	.MP4 Video File	1920x1080 Pixels @ 30 FPS	24bpp Image Sequence
GoPro Color Video [HD]	.MP4 Video File	3840x2160 Pixels @ 30 FPS	24bpp Image Sequence
GoPro Color Video [HS]	.MP4 Video File	1280x720 Pixels @ 240 FPS	24bpp Image Sequence
Webcam Color Videos	.MP4 Video File	1280x720 Pixels @ 30 FPS	24bpp Image Sequence
Depth Map	.RAR Binary File	512x424 Pixels @ 30 FPS	16bpp Image Sequence
User Mask	.RAR Binary File	512x424 Pixels @ 30 FPS	8bpp Binary Image Sequence
Kinect Pose Information	.CSV File	25 Joints	3D Joint Coordinates and Angles
Body Pose Information	.JSON File	18 Joints	2D Joint Coordinates and Confidences
Facial Landmarks	.JSON File	70 Joints	2D Joint Coordinates and Confidences
Hand Pose Information	.JSON File	2x21 Joints	2D Joint Coordinates and Confidences
iLex Annotations	.XML File	(not applicable)	Linguistic Annotations

Table 6: Contents of the SMILE Swiss German Sign Language Dataset [HS: High Speed, HD: High Definition]

4 Summary and Future Directions

This paper has introduced the SMILE Swiss German Sign Language Dataset, a large-scale dataset containing videotaped repeated productions of the 100 items of a DSGS vocabulary production test with associated transcriptions and annotations. The dataset was created for use in a project whose goal is to pioneer an assessment system for lexical signs of DSGS that relies on sign language recognition. In its initial version, the system includes automatic feedback for a subset of the items of the vocabulary test. A prototype of the system is currently under development. Following this, the system will be extended to provide feedback for the complete set of items of the vocabulary test.

5 Acknowledgements

This work was funded by the SNSF Sinergia project "Scalable Multi-modal Sign Language Technology for Sign Language Learning and Assessment (SMILE)", grant agreement number CRSII2_160811. We would like to thank the three students of the DSGS interpreting training program who helped collect the data as well as the external DSGS experts who took part in our online and focus group studies. We are greatly indebted to Thomas Hanke for continuing to share his vast expertise in sign language technology. We would also like to thank NVIDIA Corporation for their GPU grant.

6 Bibliographical References

Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., and Thangali, A. (2008). The American Sign Language Lexicon Video Dataset. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–8.

Boyes Braem, P. (1983). Studying Swiss German Sign Language dialects. In *Proceedings of the 3rd International Symposium on Sign Language Research (SLR)*, pages 247–253, Rome, Italy.

Boyes Braem, P. (2001). A multimedia bilingual database for the lexicon of Swiss German Sign Language. *Sign Language & Linguistics*, 4(1/2):133–143.

Boyes Braem, P. (2004a). *Gebärdensprachkurs Deutschschweiz, Stufe 1: Linguistischer Kommentar*. SGB-FSS, Zurich, Switzerland. CD-ROM.

Boyes Braem, P. (2004b). *Gebärdensprachkurs Deutschschweiz, Stufe 2: Linguistischer Kommentar*. SGB-FSS, Zurich, Switzerland. CD-ROM.

Boyes Braem, P. (2005a). *Gebärdensprachkurs Deutschschweiz, Stufe 3: Linguistischer Kommentar*. SGB-FSS, Zurich, Switzerland. CD-ROM.

Boyes Braem, P. (2005b). *Gebärdensprachkurs Deutschschweiz, Stufe 4: Linguistischer Kommentar*. SGB-FSS, Zurich, Switzerland. CD-ROM.

Boyes Braem, P. (2012). Overview of research on signed languages of the Deaf. Lecture held at the University of Basel. Retrieved from `http://www.signlangcourse.org` (last accessed November 13, 2015).

Bungeroth, J., Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A., and van Zijl, L. (2008). The ATIS Sign Language Corpus. In *International Conference on Language Resources and Evaluation (LREC)*.

Camgöz, N. C., Kindiroglu, A. A., Karabuklu, S., Kelepir, M., Ozsoy, A. S., and Akarun, L. (2016). Bosphorus-Sign: A Turkish Sign Language Recognition Corpus in Health and Finance Domains. In *International Conference on Language Resources and Evaluation (LREC)*.

Camgöz, N. C., Hadfield, S., Koller, O., and Bowden, R. (2017). Subunets: End-to-end hand shape and continuous sign language recognition. *IEEE International Conference on Computer Vision (ICCV)*.

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*.

Chai, X., Wanga, H., Zhoub, M., Wub, G., Lic, H., and Chena, X. (2015). DEVISIGN: Dataset and Evaluation for 3D Sign Language Recognition. Technical report, Beijing, Technical Report.

Chen Pichler, D. (2009). Sign production in first-time hearing signers: A closer look at handshape accuracy. *Cadernos de Saúde, Número especial, Línguas gestuais*, 2:37–50.

Cooper, H., Holt, B., and Bowden, R. (2011). Sign language recognition. In *Visual Analysis of Humans*, pages 539–562. Springer.

Council of Europe. (2009). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge University Press, Cambridge.

Crasborn, O. A. and Zwitserlood, I. (2008). The Corpus NGT: An Online Corpus for Professionals and Laymen. In *3rd Workshop on the Representation and Processing of Sign Languages (LREC)*, pages 44–49.

Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., and Ney, H. (2008). Benchmark Databases for Video-Based Automatic Sign Language Recognition. In *International Conference on Language Resources and Evaluation (LREC)*, pages 1–6.

Ebling, S., Konrad, R., Boyes Braem, P., and Langer, G. (2015). Factors to consider when making lexical comparisons of sign languages: Notes from an ongoing study comparing German Sign Language and Swiss German Sign Language. *Sign Language Studies*, 16(1):30–56.

Efthimiou, E., Fotinea, S., Vogler, C., Hanke, T., Glauert, J., Bowden, R., Braffort, A., Collect, C., Maragos, P., and Segouat, J. (2009). Sign language recognition, generation, and modelling: A research effort with applications in Deaf communication. In C. Stephanidis, editor, *Universal Access in Human-Computer Interaction*, pages 21–30. Springer, Berlin, Germany.

Emgu, C. (2013). Emgu cv: Opencv in .net (c#, vb, c++ and more). *Online: http://www. emgu. com.*

Forster, J., Schmidt, C., Koller, O., Bellgardt, M., and Ney, H. (2014). Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather. In *International Conference on Language Resources and Evaluation (LREC)*.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT press.

Gutierrez-Sigut, E., Costello, B., Baus, C., and Carreiras, M. (2015). LSE-Sign: A Lexical Database for Spanish Sign Language. *Behavior Research Methods*, pages 1–15.

Hanke, T. and Storz, J. (2008). iLex: A database tool for integrating sign language corpus linguistics and sign language lexicography. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)*, pages 64–67, Marrakech, Morocco.

Hanke, T. (2017). Wörterbuch ohne Wörter? Zum Entstehen eines Wörterbuches der Deutschen Gebördensprache. In Heidelberger Akademie der Wissenschaften, editor, *Jahrbuch der Heidelberger Akademie der Wissenschaften für 2016*, pages 84–88. Universitätsverlag Winter, Heidelberg.

Haug, T. (2017). Development and Evaluation of Two Vocabulary Tests for Swiss German Sign Language. Master's thesis, University of Lancaster. Submitted.

Huenerfauth, M., Gale, E., Penly, B., Pillutla, S., Willard, M., and Hariharan, D. (2017). Evaluation of Language Feedback Methods for Student Videos of American Sign Language. *ACM Transactions on Accessible Computing*, 10(1).

Johnston, T. (2010). From Archive to Corpus: Transcription and Annotation in the Creation of Signed Language Corpora. *International Journal of Corpus Linguistics*, 15(1):106–131.

Konrad, R., Hanke, T., König, S., Langer, G., Matthes, S., Nishio, R., and Regen, A. (2012). From form to function: A database approach to handle lexicon building and spotting token forms in sign languages. In *Proceedings of the 5th LREC Workshop on the Representation and Processing of Sign Languages*, pages 87–94, Istanbul, Turkey.

Matthes, S., Hanke, T., Regen, A., Storz, J., Worseck, S., Efthimiou, E., Dimou, A.-L., Braffort, A., Glauert, J., and Safar, E. (2012). Dicta-Sign: Building a Multilingual Sign Language Corpus. In *5th Workshop on the*

Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon.

Mayberry, R., Hall, M., and Zvaigzne, M. (2013). Subjective frequency ratings for 432 ASL signs. *Behavior Research Methods*, 46(2):526–39.

McEnery, T. and Wilson, A. (2001). *Corpus Linguistics*. Edinburgh University Press, Edinburgh, Scotland, 2nd edition.

Morgan, G. and Woll, B. (2002). The development of complex sentences in British Sign Language. In Gary Morgan et al., editors, *Directions in Sign Language Acquisition: Trends in Language Acquisition Research*, pages 255–276. John Benjamins, Amsterdam, Netherlands.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T., and Henning, J. (1989). *HamNoSys: Version 2.0: An Introductory Guide*. Signum, Hamburg, Germany.

Prillwitz, S., Hanke, T., König, S., Konrad, R., Langer, G., and Schwarz, A. (2008). DGS Corpus Project–Development of a Corpus Based Electronic Dictionary German Sign Language/German. In *3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, page 159.

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S., and Cormier, K. (2013). Building the British Sign Language Corpus.

Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., and Moore, R. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124.

Spaai, G., Lichtenauer, J., Hendriks, E., de Ridder, H., Arendsen, J., de Ridder, H., Fortgens, C., Bruins, M., and Elzenaar, M. (2005). Elo: An electronic learning environment for practising sign vocabulary by young deaf children. In *Proceedings of the International Congress for Education of the Deaf (ICED)*.

Vinson, D., Cormier, K., Denmark, T., Schembri, A., and Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, 40(2):1079–1087.

von Agris, U. and Kraiss, K.-F. (2010). SIGNUM Database: Video Corpus for Signer-Independent Continuous Sign Language Recognition. In *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 243–246.

Zafrulla, Z., Brashear, H., Hamilton, H., and Starner, T. (2010). A Novel Approach to American Sign Language (ASL) Phrase Verification using Reversed Signing. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 48–55.

Zhang, Z. (2012). Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10.