DGS-Fabeln-1: A Multi-Angle Parallel Corpus of Fairy Tales between German Sign Language and German Text

Fabrizio Nunnari[1], Eleftherios Avramidis[1], Cristina España-Bonet[1], Marco González[2], Anna Hennes[1], Patrick Gebhard[1]
[1]German Research Center for Artificial Intelligence (DFKI)
[2]University of Köln
{fabrizio.nunnari, eleftherios.avramidis, cristinae, patrick.gebhard}@dfki.de
m.gonzalezathenas@uni-koeln.de, annahennes@hotmail.de

Abstract

We present the acquisition process and the data of *DGS-Fabeln-1*, a parallel corpus of German text and videos containing German fairy tales interpreted into the German Sign Language (DGS) by a native DGS signer. The corpus contains 573 segments of videos with a total duration of 1 hour and 32 minutes, corresponding with 1428 written sentences. It is the first corpus of semi-naturally expressed DGS that has been filmed from 7 angles, and one of the few sign language (SL) corpora globally which have been filmed from more than 3 angles and where the listener has been simultaneously filmed. The corpus aims at aiding research at SL linguistics, SL machine translation and affective computing, and is freely available for research purposes at the following address: https://doi.org/10.5281/zenodo.10822097.

Keywords: sign language, corpus, fairy tales, machine translation, sign language translation

1. Introduction

In the realm of machine translation (MT), sign language (SL) translation is recognized as a "difficult task", much harder than text-to-text translation. Since SLT still represents a low-resource scenario (De Coster et al., 2023), the translation quality in generic domains is very low (Müller et al., 2022, 2023; Zhu et al., 2023). Despite the attempt to involve the sign language communities in the collection of data, as for example of the American Sign Language (Joze and Koller, 2019; Desai et al., 2023), parallel corpora of sign languages and text/speech are still scarce (Kopf et al., 2022) as compared to the ones between spoken languages.

Parallel corpora are of interest for computer scientists to train and evaluate machine translation systems, as well as for linguists, to deepen the understanding of a relatively new language structure: two communities that have been criticized for a lack of collaboration (Börstell, 2023; Fox et al., 2023).

The disjoint research effort between sign-language linguists and machine-learning scientists reflects in the nature of the few corpora available for DGS (Deutsche Gebärdensprache; German Sign Language). Corpora appreciated by the former scientist group, are not appreciated by the latter, and vice versa. For example, the most used corpus for training machine translation systems between German and DGS, the RWTH-PHOENIX corpus (Camgöz et al., 2018), is a ready-to-use collection of short sentences of a single domain (weather forecast). However, linguists criticize it for the unnatural signing style (communication to a camera) and its questionable interpretation quality (Müller et al., 2022). Vice versa, the most popular corpus for DGS, the Public DGS corpus (Hanke et al., 2020), has been conceived for natural conversation, on different topics, with a variety of elicitation tasks. However, computer scientists have many difficulties in using it for training MT engines (e.g. Zhu et al., 2023), because of the highly informal way of conversing, wide variance in dialects, and high variety of conversation topics, which leads to a large vocabulary with few occurrences.

This paper focuses on DGS and presents the acquisition procedure and the data of *DGS-Fabeln-1*; a parallel corpus of videos narrating German fairy tales interpreted into DGS. The corpus, apart from obvious pedagogical uses related to the fairy tales, has been created with the aim to fulfil the requirements of both research communities, according to the following criteria:

First, the recorded **signer is signing frontally while standing**, directed to a "virtual camera". This is needed by computer scientists because it is implicitly desired when building a system aiming at creating virtual signers for very specific, highly controlled circumstances, like public announcements or news, where signers sign to a camera.

Second, despite the signing being monodirectional, the signer is **directed to a human** who is known to understand the SL. This increases the naturalness of the signing at the small cost

LREC-COLING 2024, pages 4847–4857

of creating a slight offset between the signer eye-gaze direction with respect to the camera.

Third, the whole dataset is performed by **a single signer**. This can be seen as a limitation, both in terms of variety of signing style (for linguists) and variance to support generalization (for computer scientists). However, for the latter category, having long footage with a singe signer allows building a generation system with coherent aesthetics (e.g., when using GAN-based approaches; Saunders et al., 2021, 2022) and behaviour (for motion synthesis; Ventura et al., 2020; Brock et al., 2020; Nguyen et al., 2021).

Fourth, we use a novel capture setup consisting of **eight synchronized videos**: seven different angles for the signer plus one framing the listener. This fulfils the requests of researchers who ask for side-view recordings in order to better analyse hand and body movement relatively to their position. The multiple points of view allow computer vision and computer graphics experts to experiment with video interpolation techniques for transferring the recorded motion from the video signal to high-precision 3D avatar animation data. For example, see (Michoud et al., 2007; Alldieck et al., 2018).

Finally, for affective analysis, the corpus provides a **backchannel video** framing the listener of the tales. This gives the opportunity for further studies analysing the way SL enacts communicative emotions.

Due to its limited size, this corpus is not sufficient as unique source for training MT engines. However, we envision several possible usage scenarios:

- high-quality *test* set for engines trained on other big-sized DGS corpora;
- train or test data for the reconstruction of skeletal motion (specifically for the challenging case of sign language);
- analysis of the correlations between the story plot and the back-channeling behaviour (e.g., emotion conveyed by the story vs. recipient's facial expression).

In any case, DGS-Fabeln-1 represents the first of a series and meanwhile it aims at defining a new methodology for the creation of future corpora featuring the needs of both linguists and computer scientists.

2. Related Work

Despite the numerous and often very elaborate efforts to create datasets for SLs, these are by far inferior to the number of SLs, and the overall coverage of domains is scarce. They have nevertheless been valuable, as they have allowed for initiating research in several directions. In this section, we attempt to describe similar efforts and compare and contrast DGS-Fabeln-1 with prior work.

2.1. Dataset categorization

A SL dataset can be categorized based on several technical or content-related characteristics. A first categorization refers to the **continuity** of the content, distinguishing among the fingerspelling datasets (Dreuw et al., 2006; Shi et al., 2019), the ones with isolated signs (e.g. Athitsos et al., 2008; Mesch and Wallin, 2012) and the ones with continuous SL (e.g. Albanie et al., 2020, 2021), where our corpus is included.

Secondly, a basic categorization of continuous SL corpora is drawn based on **how the data has been collected**. Here, we find the corpora whose content has been expressed:

1. first in a SL and then translated into text (e.g. Hanke et al., 2020),
2. first in a spoken language and then interpreted into a SL for the corpus creation purposes (e.g. Sisto et al., 2023) and
3. first in a spoken language with simultaneous SL interpretation for third use (e.g., parliament talks or TV broadcasts), and then utilized as a corpus (e.g. McGill and Saggion, 2023; Camgöz et al., 2021).

This categorization also orders the corpora with regard to their value from a linguistic perspective and the quality of parallel data. Whereas corpora of the first category are of a higher quality, their collection is more demanding, and therefore it is harder to collect big quantities. On the contrary, the latter category may result into larger corpora which are suitable for data-demanding computational processes but whose linguistic validity may be very low. Our corpus falls in category 2, containing semi-natural SL expression, while interpreting known content which had been previously documented in written form of the spoken language.

The Institute of German SL and Communication of the Deaf in Hamburg has recently curated a compendium of available SL corpora and resources (Kopf et al., 2022)[1] and set the minimum requirements for including a corpus there, specifying that it should contain (semi-)spontaneous signing by native (L1) speakers, provide transcriptions or translations for at least some of its content and contain at least 10 hours of SL recordings. We provide the original texts and intend to expand our corpus in future iterations to fulfil the last requirement. Regarding spontaneity, we provide a semi-spontaneous dataset where, even if the signing was driven by the text, the signer was asked to follow the guidelines described later in section 3.3.

[1] https://www.sign-lang.uni-hamburg.de/lr/compendium/

2.2. Multiple recording angles

Despite the importance of multiple recording angles, only a small amount of SLs corpora have accounted for this (Crasborn and Zwitserlood, 2008; Dreuw et al., 2008; Duarte et al., 2021; Ko et al., 2019; Matthes et al., 2012; Neidle and Vogler, 2012). Our corpus is one of the few offering more than 3 angles, along with *How2Sign* for ASL (Duarte et al., 2021), and the first corpus of DGS which is recorded from 7 different angles.

2.3. Corpora for the DGS

Only a handful of prior corpora have included content in DGS: the Vidi Sign Space (Aslı Özyürek, UZ et al., 2009), the ECHO (Nonhebel et al., 2004), the DictaSign (Matthes et al., 2012), the RWTH-PHOENIX-Weather 2014T (further mentioned as PHOENIX; Camgöz et al., 2018) and the Public DGS Corpus (Hanke et al., 2020). We provide details on the two latter, since they have been the most influential in SL research.

The RWTH-PHOENIX-Weather 2014T dataset (Camgöz et al., 2018) contains about 11 hours of videos of German TV weather broadcasts interpreted into DGS, aligned with transcriptions of spoken German into German text and DGS signs to glosses. Despite the limitations mentioned in Section 1, its test set has become a *de facto* evaluation for MT systems translating between DGS and German text. The Public DGS Corpus (*DGS* in our tables) (Hanke et al., 2020) contains a collection of dialogues between native signers of German Sign Language (∼50 hours). It includes the translation into German, lemmatisation, annotation of mouthing and mouth gestures and a detailed gloss annotation scheme. Also, relevant are the two corpora of Swiss German Sign Language DSGS, parallelized with German text, released by the Shared Task of Sign Language Translation of the Seventh and Eighth Conference of Machine Translation (Müller et al., 2022, 2023). The characteristics of the above corpora which have a transcription in German will be compared with ours in Section 4.3.

Finally, previous linguistic research on the linguistic phenomena of *perspective taking* and *role shift* of DGS has been conducted on Aesop fables (Herrmann and Pendzich, 2018), which highlights the special linguistic characteristics of this genre.

3. Corpus creation

3.1. Personnel

The recording procedure involved three persons: a Signer, a Director, and a software Operator.

The Signer was selected following three criteria. First, a deaf person fully immersed in the deaf culture. Second, someone with Sign Language as first native language, meaning that they grew up talking German sign language with their parents. Third, an expressive signer: needed for our goal of capturing emotional expressiveness in German sign language. After a consultation with a sign language teacher working at the Saarland University (language centre), we contacted and hired a native speaker, deaf in her third generation, with a strong interest in signing fairy tales for children.

The Director was a MSc degree student of Psychology with a B1 qualification in DGS. She could communicate with the Signer and act as translator between other researchers and the Operator. The Operator was not qualified for DGS, but was the main developer of the recording software, thus able to monitor for malfunctions and improve the software during the recording period according to the emerging needs.

3.2. Material preparation

The topic of fairy tales is selected as a reasonable candidate to form a corpus with relatively simple vocabulary, enough emotional content for an expressive SL performance and as a source of learning material for deaf children.

The selection of the material was conducted after surveying both existing scientific corpora and other free-of-use material available online.

Following these considerations, we opt for the collection of fairy tales available online at the NDR German national television broadcaster[2]. The right for using such material for non-profit scientific research was ensured after directly contacting a NDR representative. This repository features fairy tales that, in addition to the DGS video recordings, come together with pre-formatted text and audio narrative, thus giving potentially the possibility to create a three-way parallel corpus between text, video, and audio (after proper segmentation).

Initially, the idea was to use the existing videos as reference for the interpretation, and re-enact the existing videos with our video capturing system. However, after a first consultation with the Signer, it was clear that this would have been impossible without sacrificing the naturalness of the signing. At the end, the existing videos were used as reference only when the Signer and the Director were doubtful about some interpretation choices. This resulting in effectively having two alternative frontal videos of the DGS interpretation for the same source text.

The source text was divided in "segments" by the Director, prior to a recording. The definition of a

[2] https://www.ndr.de/fernsehen/barrierefreie_angebote/gebaerdensprache/Maerchen-in-Leichter-Sprache-und-Gebaerdensprache,maerchengebaerden100.html

segment, which can correspond to more than one written sentence, followed two principles. First, limit the need to refer to entities of previous segments (i.e., pronouns should refer only to entities in the same segment). Second, keep the segment short enough to be memorized and interpreted with reasonable low effort. However, on-the-fly-adjustments were made during the recording, e.g., when a segment resulted being too long to be recorded with a single take without mistakes.

3.3. Interpretation guidelines

The Signer and the Director, also with the advice of MT experts, worked together to establish a set of guidelines driving the interpretation process. Four major guidelines were identified:

1. Similarly to how German spoken language is increasingly incorporating English words, international signs are increasingly found in DGS, mainly used by young people. For this corpus, we decided to conform to DGS vocabulary.

2. As the Signer grew up learning two different DGS variants, she was asked to be consistent and try to use only one of them, preferably the most popular in Germany.

3. When more than one signs were good candidates to translate a German word, the Signer limited herself to one possible translation for these words and uses the same sign consistently throughout all fairy tales.

4. Finally, there was an effort to find a balance between incorporating signs and preforming a word-for-word translation that stays close to the original text. For example, DGS often uses *incorporation* to compress signs and shorten the translation. Another example is that signers often use INDEX signs to refer to previously mentioned people, objects or places. This concept can be compared to replacing nouns with pronouns in spoken languages. The Signer often did this instinctively, even if the original noun was used repeatedly in the text. In this case, the Signer was asked to stick to the text by signing again and again the original noun. The goal being to keep each segment as much as possible decontextualized (i.e., independent of entities introduced in previous segments), thus helping the SLT translation systems, which are at the moment unable to model broad SL context, in building an association between words and their most naturally corresponding signs, rather than to the INDEX signs. Nevertheless, the Signer had to make sure for every situation that she could follow the guidelines without damaging the naturalness and integrity of DGS.

3.4. Recording setup

The recording devices are eight Samsung S9 tablets. As shown in figures 1 and 2, one tablet frames the Signer frontally. Six other tablets frame the Signer at 30, 60 and 90 degrees on both sides. Finally, an eighth tablet is positioned over the head of the Signer to frame the Director, who is back-channelling while watching the sign performance. Every device was placed at a distance of 2 meters from the Signer. All devices recorded at Full HD resolution (1920x1080) and a frequency of 30 Hz.

Figure 1: The recording setup from the point of view of the Director.

The video recordings have been conducted using a customized version of the RecSync software (Akhmetyanov et al., 2021), which allows synchronizing the camera of multiple Android devices connected only through a Wi-Fi network and record parallel videos with a shooting discrepancy below 1 millisecond: enough precision to catch, simultaneously, the moment of a photo flash. In so doing, the several recordings are synchronized at frame-level, even when it concerns the fastest hand movements or eye blinks. As an example, figure 3 shows frame number 199 of Segment 1 from the tale *Frau Holle*.

The setup is completed by: i) one laptop used by the Operator to coordinate the camera devices, ii) a second laptop used to show to the Signer the text of the segment to be enacted (and optionally the existing SL video from NDR), and iii) a big screen to show to the Signer a preview of the last captured video.

The recording software, called RecSyncNG, composed by an Android application and a desktop application acting as remote controller, is available as open source project[3].

3.5. Recording procedure

The creation of the corpus is a rather exhausting task. Hence, recording days were organized in two sessions: 2 hours in the morning and 2 more hours in the afternoon, with about 1 hour break in between. Each session of 2 hours was meant to be split in 45 minutes work and 15 minutes pause, but exceptions were often applied according to the

[3] `https://github.com/DFKI-SignLanguage/RecSyncNG`

Figure 2: A 180 degrees panoramic view of the recording setup from the point of view from the Signer. The Director sits just behind the frontal camera. The Operator sits at the table next to the Director.

Figure 3: Left: a still collage from the 7 videos, framing the Signer. Right: the corresponding frame of the back-channel video.

tiredness of the Signer or the "flow" of the recording: it was often preferred to finish a narrative section before having a break.

The Director, the Signer, and the Operator coordinated each other following a methodology inspired from movie production, adapted also to allow for the coordination between the non-signing Operator and the deaf Signer:

1. Director: prepares a new segment on screen.
2. Operator: sets the segment number and the take number on the recording software.
3. Director: nods to the Signer to get ready.
4. Signer: rehearses and confirms to be ready.
5. Director: says "Camera" + nods to the Operator (asks for recording to start).
6. Operator: says "Yes" (confirms recording has started).
7. Director: says "Action!" + nods and points to the Signer.
8. Signer: performs the segment. Might interrupt the interpretation, communicating that the performance was not good.
9. Director: says "Stop!", followed by "good" or by "again" (i.e., another take is needed).
10. Operator: stops recording. If the Director said "again", increase the take number and go back to step 3. Otherwise continue.
11. Operator: downloads the videos from the recording devices and shows the Front video to the Signer.
12. Signer and Director discuss the quality of the performance.
13. Signer: confirms that take was good (signs "OK"), or ask for a new take (signs "again").
14. Director: speaks out the decision of the Signer (to double-check with the Operator).
15. Repeat from step 1 (new segment) or 2 (new take).

If any technical problem occurs, the Operator asks to interrupt the recordings until it is fixed.

The seven tales of DGS-Fabeln-1 were recorded in 8 days of work, recording every tale in one single day, except for the longest one, *Scnheewittchen* (=Snow White) that spanned over two days.

The duration of a take cycle was varying a lot according to the time needed by the Director and the Signer to discuss the strategy of interpretation and some choices about the sign to use when more synonyms were available (steps 4 and 12). The discussion time could range from a few seconds to a few minutes when, for example, the Signer decided to watch the reference NDR videos before taking a final choice.

3.6. Content structure

An overview of the corpus statistics is shown in Table 1. The corpus consists of 7 fairy tales, each

ID	Full name	Date	Segments	Frames	Minutes	Dropped Frames	Missing Videos
1	Der Hase und der Igel	23.05.23	37	11365	6	6,83%	50
2	Frau Holle	26.05.23	77	26100	15	6,73%	16
3	Der Wolf und die 7 Geißlein	31.05.23	73	22091	12	6,71%	2
4	Schneewitchen	06.06.23	150	42025	23	6,61%	13
5	Hänsel und Gretel	15.06.23	92	26482	15	7,23%	0
6	Dornröschen	04.07.23	63	16787	9	5,09%	8
7	Die Bremer Stadtmusikanten	06.07.23	81	20824	12	5,03%	2
			573	165674	92	6,40%	91

Table 1: Summary of the DGS-Fabeln-1 corpus content.

divided in segments (between 37 and 149). The whole corpus features 573 segments, for a total of about 92 minutes of recording, hence averaging 9.6 seconds per segment.

Due to technical failures, 92 out of 4584 videos (573 segments per 8 cameras) could not be properly saved. Only two segments miss all video material, due to human errors.

As Android camera devices typically perform *frame dropping* (when unable to shoot or store a frame correctly), all synchronized videos present intermittent, irregular black frames. For this reason, each video is accompanied by a CSV file with two columns: i) a timestamp, in nanoseconds, when the frame was shot, and ii) a flag indicating if the frame is taken from the `Original` recorded video or a `Generated` black frame. On average, we measured a 6.40% of frame dropping over the whole corpus. All videos contain the same number of frames and are verified to be synchronized with <10 ms precision, allowing for full body skeletal motion reconstruction via dedicated software (e.g., `http://freemocap.org`).

Finally, an additional frontal video, as originally recorded with the Android device without the synchronization, is included as per request by linguists, because the flashing black frames present in `Front.mp4` disturb the evaluation and annotation processes.

Each fairy tale is distributed as a main folder containing all the segments. Each segment is contained in a folder with all the videos and CSV files. For each tale, an additional CSV file contains the list of recorded segments and the original text used by the Signer.

3.7. Productivity

With a total of 573 segments, 9598 words, and 8595 glosses, the average recording productivity was almost 18 segments/hour, 300 words/hour and 270 glosses/hour. Considering the lower pace at the beginning of the recordings, because of software tuning, hardware checks, and coordination warm-up, such values can be considered as a lower bound.

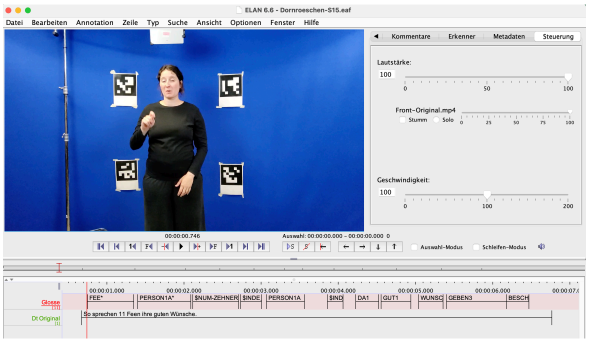

Figure 4: Example of annotation with ELAN.

4. Corpus Analysis

4.1. German Sign Language

The recorded DGS-Signed Videos were annotated with glosses using the annotation tool ELAN (`archive.mpi.nl/tla`). Figure 4 shows a screenshot.

For glossing the manual parts of the sign language, we used the conventions for annotation of the DGS-Korpus (Konrad et al., 2022). We also use the same glosses for the same signs, as long as they have the same signs (e.g. "FEE*" (fairy) does not appear in the open DGS Corpus). The decision was made to have later the possibility to compare both corpora.

Each sequence was separately annotated. The annotated DGS sequences are unrelated to the spoken or written segments. That means, sometimes the DGS sequences have three sentences, even if the written text shows just one or two.

For this first annotation, we only annotated the manual SL elements after the signs. Annotation of non-manuals is reserved for future work.

4.2. German Spoken Language

The natural language part of the corpus is aligned at segment level with the videos. Each segment

	Fab1	Fab2	Fab3	Fab4	Fab5	Fab6	Fab7	all
# segments	37	77	73	150	92	63	81	573
# sentences	126	266	207	333	196	130	179	1437
# tokens	1011	2059	1532	2751	1682	1023	1421	11479
# words (w/o punct&num)	848	1746	1265	2304	1442	847	1198	9650
# types	231	329	282	422	325	244	283	1045
sentence length (words)	8.0	7.7	7.4	8.3	8.6	7.9	7.9	8.0
word length (characters)	4.3	4.5	4.9	5.1	4.6	4.9	4.4	4.7
% function words	49.9	47.6	46.3	46.2	46.1	46.8	50.7	47.4
% hapax legomena	13.3	7.1	9.9	7.7	10.5	14.0	10.6	4.0
TTR	27.2	18.8	22.3	18.3	22.5	28.8	23.6	10.8
Entropy (words)	6.7	7.1	6.8	7.2	7.0	6.9	7.0	8.1
Yule K (words)	220.7	172.8	214.0	185.2	198.9	160.4	177.9	121.4
% short words ($<$4chars)	40.0	36.6	39.4	37.5	35.6	37.5	40.7	37.9
% words $>$ 2 syllables	4.4	4.8	11.1	14.0	3.9	12.5	4.3	8.3
Flesch Reading Ease	94.1	89.9	85.3	82.5	85.8	84.7	93.5	87.4
Wiener Sachtextformel	1.3	1.6	3.4	5.4	3.5	4.6	1.6	3.3

Table 2: Statistics and complexity of the textual part of the 7 fairly tales of *DGS-Fabeln-1*.

contains one or more sentences, with an average of 2.5 sentences per segment and 8 words per sentence. The complete statistics per fairly tale are summarized in Table 2. The first block of the table shows the basic statistics,[4] the second and third ones report on metrics related to text complexity: richness and readability respectively.

Large type-token ratios (TTR) and the abundance of hapax legomena (words that only appear once in a text) indicate rich texts where the author uses a varied vocabulary. But large values can be an artefact of short texts, as seen with the fact that both of them diminish significantly when we consider the full set of tales. Shannon's entropy (Shannon, 1948) and specially Yule's K (Yule, 1944) are close to constant measures that converge after a certain amount of text. Low entropy and high Ks as seen in the table imply low complexity. The same trend is observed with the readability measures Flesch reading ease (Kincaid et al., 1975) with its adaptation to German (Amstad, 1978) and the first Wiener Sachtextformel (Schulz, 1985) specifically designed for German. The first one, FRE, scales between 0 and 100 going from difficult to easy. The second one, WSTF, with no hard limits, usually goes between 4 and 15 from easy to difficult, roughly reproducing the Austrian/German grade levels (a text with WSTF=x would be adequate for the xth grade). According to all the metrics, *Schneewitchen* (=Snow White, Fab4) is the most complex fairy tale (also the longest one) and *Der Hase und der Igel* (=The Hare and the Hedgehog, Fab1) the easiest (also the shortest one).

[4]Sentence splitting has been done manually, and we used Moses tokenizer (Koehn et al., 2007) for further processing.

4.3. Comparison to other parallel corpora

Several corpora are used for research in DGS (Section 2). Well studied are the *Phoenix* [5] and the Public DGS Corpus (*DGS*). More recently, two editions of the shared task in SL translation have released training and testing data (Swiss German Sign Language corresponding with German text), *SLT22* and *SLT23*. In Table 3, we show the comparison of all the test sets from these corpora with DGS-Fabeln-1 (*Fabeln*), and a comparison test set used to evaluate (textual) MT into German, the *WMT22* test set. DGS-Fabeln-1 has a similar length to Phoenix, SLT22 and SLT23 in number of segments, but doubles or triples the other ones when the number of sentences are considered. Both DGS and WMT22 are significantly larger. Notice that DGS-Fabeln-1 is similar to the DGS corpus regarding sentence and word lengths. However, Yule-K, FRE and WSTF show that the text in DGS-Fabeln-1 is much more simple. SLT22, SLT23 and WMT22 are the most complex datasets, and this can explain in part the low performance of SLT in the SL shared tasks performed at WMT. The status of current SLT technology is not ready to tackle translation tasks at the same level as text-to-text translation.

5. Conclusion

In this paper, we described the motivation and the procedure used for creating DGS-Fabeln-1, a par-

[5]Notice that the Phoenix corpus is available in its lowercased version. In order to restore the original capitalisation, we trained a recaser using the Moses scripts (Koehn et al., 2007) on the German Wikipedia dump from May 2020 extracted with Wikitailor (España-Bonet et al., 2023).

	Fabeln	DGS	Phoenix	SLT22	SLT23	WMT22
# segments	573	5113	642	488	496	19856
# sentences	1437	5113	642	488	496	19856
# tokens	11479	53537	8458	6712	5992	304004
# words (w/o punct&num)	9650	44609	7816	5731	5149	259585
# types	1045	4894	1001	2040	2080	46058
sentence length (words)	8.0	10.5	13.2	13.8	12.1	15.3
word length (characters)	4.7	4.8	5.3	5.7	5.8	5.9
% function words	47.4	57.8	50.1	45.9	47.2	44.3
% hapax legomena	4.0	5.8	5.3	24.0	29.9	11.8
TTR	10.8	11.0	12.8	35.6	40.4	17.7
Entropy (words)	8.1	9.1	8.2	9.5	9.5	11.4
Yule K (words)	121.4	78.1	87.4	56.4	54.1	47.3
% short words (<4chars)	37.9	40.4	34.2	34.0	33.3	34.2
% words > 2 syllables	8.3	12.7	15.9	23.1	22.8	24.7
Flesch Reading Ease	87.3	79.2	69.0	58.5	58.3	52.1
Wiener Sachtextformel	3.3	4.9	4.9	9.3	8.9	10.1

Table 3: Stylistic comparison of *DGS-Fabeln-1* (Fabeln) with respect to other parallel corpora in German.

allel corpus between text and SL videos. We claim that this corpus comes with all the features needed to be "appreciated" by both linguists and computational linguists. For linguists, the corpus features a native signer that is motivated by an active listener. For computer vision scientists, the corpus features a well-structured format, accompanied by annotations, and offers the possibility to test SL recognition from multiple points of view. The low complexity of the content, as well the intentional reduction of broad SL context anaphora, makes it more friendly to the infant algorithms for MT of SLs. An additional back-channeling video offers researchers of affective computing the possibility to explore for correlations between the semantic content of the tales and the facial reactions of the listener, thus supporting a deeper understanding on the communication of emotions in SL.

For the future, we plan to employ the same procedure, capturing hardware, and software, to re-record blurry videos and for a new set of fairy tales and other domains.

6. Limitations

The corpus has been recorded at Full HD resolution and 30 Hz frequency. Modern devices would allow recording at least at 4K format and 60 Hz, but our devices were limited at 30 Hz, and we tested that recording at 4K resolution would have increased the frame drop up to 20%.

Due to the low maturity of the software and the procedure, about 2% of the videos were not correctly recorded, especially in the first tales. However, both the software settings and the recording procedure improved from tale to tale, resulting in no videos missing for tale 5 and only 10 videos missing for the last two tales.

Another issue present in the first two tales is the camera sensor aperture and sensitivity, which were left at their defaults (1/32th sec aperture and sensitivity 800). This led to blurry videos that, when the signer's hands are moving at high speed, makes it virtually impossible to automatically recognize hand fingers. For later tales, the settings were changed to 1/250th aperture and sensitivity 2000, leading to sharper frames.

The first tales present a Charuco marker at the back of the signer: the intent being to give the possibility for relative camera localization during video post-processing and skeletal reconstruction. Unfortunately, we could test the efficacy of the marker only after the finalization of the first tales, and we recognized problems due to excessive occlusion and light reflections. The last tales present a more robust setting, featuring 4 Aruco markers positioned at the sides of the signer (see figure 4).

Finally, the most evident artefact is a continuous autofocus of all the devices, which adjusts the focal length while the signer signs. This is due to a combination of the light conditions and the high contrast between the light skin of the signer and her dark dresses. Unfortunately, this behaviour could not be controlled before the end of the recordings, but the software includes now a manual focus control mechanism that will be employed in future sessions.

7. Ethics

The people recorded have given informed consent for the usage of their footage for the corpus purposes. The corpus contains no references to real persons, so anonymization is not needed. Our institution works closely with local SL communities and does their best to ensure the applicability of the FAIR and CARE principles.

8. Acknowledgements

This work was partially funded by the German ministry for education and research (BMBF) through projects BIGEKO (grant number 16SV9093) and SOCIALWEAR (Socially Interactive Smart Fashion, grant number 01IW20002)). Our gratitude to Deborah Monzel (the signer). Thanks also to Saurabh Pandey for his contribution in the development of the RecSyncNG application.

9. Bibliographical References

Azat Akhmetyanov, Anastasiia Kornilova, Marsel Faizullin, David Pozo, and Gonzalo Ferrer. 2021. Sub-millisecond Video Synchronization of Multiple Android Smartphones. In *2021 IEEE Sensors*, pages 1–4, Sydney, Australia. IEEE.

Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. 2018. Video Based Reconstruction of 3D People Models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Toni Amstad. 1978. *Wie Verständlich Sind Unsere Zeitungen? Universität Zürich*. Ph.D. thesis, Philosophische Fakultät I. Universität Zürich.

Heike Brock, Felix Law, Kazuhiro Nakadai, and Yuji Nagashima. 2020. Learning three-dimensional skeleton data from sign language video. *ACM Trans. Intell. Syst. Technol.*, 11(3).

Carl Börstell. 2023. Ableist Language Teching over Sign Language Research. In *Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2023)*, pages 1–10, Tórshavn, the Faroe Islands. Association for Computational Linguistics.

Mathieu De Coster, Dimitar Shterionov, Mieke Van Herreweghe, and Joni Dambre. 2023. Machine translation from signed to spoken languages: State of the art and challenges. *Universal Access in the Information Society*.

Cristina España-Bonet, Alberto Barrón-Cedeño, and Lluís Màrquez. 2023. Tailoring and evaluating the wikipedia for in-domain comparable corpora extraction. *Knowledge and Information Systems*, 65(3):1365–1397.

Neil Fox, Bencie Woll, and Kearsy Cormier. 2023. Best practices for sign language technology research. *Universal Access in the Information Society*.

Annika Herrmann and Nina-Kristin Pendzich. 2018. Between narrator and protagonist in fables of German Sign Language. pages 275–308.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, and Anja Regen. 2022. Public DGS Corpus: Annotation Conventions / Öffentliches DGS-Korpus: Annotationskonventionen. Publisher: Universität Hamburg Version Number: 4.1.

Maria Kopf, Marc Schulder, and Thomas Hanke. 2022. The sign language dataset compendium: Creating an overview of digital linguistic resources. In *Proceedings of the LREC2022 10th Workshop on the Representation and Processing of Sign Languages: Multilingual Sign Language Resources*, pages 102–109, Marseille, France. European Language Resources Association.

Brice Michoud, Erwan Guillou, Hector Briceno, and Saida Bouakaz. 2007. Real-Time Marker-free Motion Capture from multiple cameras. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–7, Rio de Janeiro, Brazil. IEEE.

Mathias Müller, Zifan Jiang, Amit Moryossef, Annette Rios, and Sarah Ebling. 2022. Considerations for meaningful sign language machine translation based on glosses.

Lan Thao Nguyen, Florian Schicktanz, Aeneas Stankowski, and Eleftherios Avramidis. 2021. Automatic generation of a 3D sign language avatar on AR glasses given 2D videos of human signers. In *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 71–81, Virtual. Association for Machine Translation in the Americas.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2021. Anonysign: Novel human appearance synthesis for sign language video anonymisation. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8.

Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. 2022. Signing at Scale: Learning to Co-Articulate Signs for Large-Scale Photo-Realistic Sign Language Production. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5141–5151.

Renate A. Schulz. 1985. Lesen-verstehen-lernen-schreiben: Die schwierigkeitsstufen von texten in deutscher sprache. *Die Unterrichtspraxis / Teaching German*, 18(2):366–369.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Lucas Ventura, Amanda Duarte, and Xavier Giro-i-Nieto. 2020. Can Everybody Sign Now? Exploring Sign Language Video Generation from 2D Poses.

C Udny Yule. 1944. *The statistical study of literary vocabulary*. Cambridge University Press.

Dele Zhu, Vera Czehmann, and Eleftherios Avramidis. 2023. Neural machine translation methods for translating text to sign language glosses. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12523–12541, Toronto, Canada. Association for Computational Linguistics.

10. Language Resource References

Samuel Albanie, Gül Varol, Liliane Momeni, Triantafyllos Afouras, Joon Son Chung, Neil Fox, and Andrew Zisserman. 2020. BSL-1K: Scaling Up Co-articulated Sign Language Recognition Using Mouthing Cues. In *Computer Vision–ECCV 2020: 16th European Conference*, volume 12356, pages 35–53, Cham. Springer International Publishing.

Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, and Andrew McParland. 2021. BBC-Oxford british sign language dataset. *arXiv preprint arXiv:2111.03635*.

Aslı Özyürek, UZ, Pamela Perniss, Engin Arik, Deniz Ilkbasaran, and Daniela Happ. 2009. Collection "VIDI Sign space project".

Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Quan Yuan, and Ashwin Thangali. 2008. American sign language Lexicon video dataset (ASLLVD). In *Proceedings of the Workshop on Human Communicative Behaviour Analysis*.

Necati Cihan Camgöz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7784–7793.

Necati Cihan Camgöz, Ben Saunders, Guillaume Rochette, Marco Giovanelli, Giacomo Inches, Robin Nachtrab-Ribback, and Richard Bowden. 2021. Content4all: Open research sign language translation datasets. In *Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–5. IEEE.

Onno Crasborn and Inge Zwitserlood. 2008. The Corpus NGT: An online corpus for professionals and laymen. In *Sign-Lang@ LREC 2008*, pages 44–49. European Language Resources Association (ELRA).

Aashaka Desai, Lauren Berger, Fyodor O Minakov, Vanessa Milan, Chinmay Singh, Kriston Pumphrey, Richard E Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, and Danielle Bragg. 2023. ASL Citizen: A Community-Sourced Dataset for Advancing Isolated Sign Language Recognition. *arXiv preprint arXiv:2304.05934*.

Philippe Dreuw, Thomas Deselaers, Daniel Keysers, and Hermann Ney. 2006. Modeling image variability in appearance-based gesture recognition. In *ECCV Workshop on Statistical Methods in Multi-Image and Video Processing*, pages 7–18.

Philippe Dreuw, Carol Neidle, Vassilis Athitsos, Stan Sclaroff, and Hermann Ney. 2008. Benchmark Databases for Video-Based Automatic Sign Language Recognition. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*.

Amanda Duarte, Shruti Palaskar, Lucas Ventura, Deepti Ghadiyaram, Kenneth DeHaan, Florian Metze, Jordi Torres, and Xavier Giro-i-Nieto. 2021. How2sign: A large-scale multimodal dataset for continuous american sign language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2735–2744.

Thomas Hanke, Marc Schulder, Reiner Konrad, and Elena Jahn. 2020. Extending the Public DGS Corpus in size and depth. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 75–82, Marseille, France. European Language Resources Association (ELRA).

Hamid Reza Vaezi Joze and Oscar Koller. 2019. MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language. ArXiv:1812.01053 [cs].

Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2019. Neural sign language translation based on human keypoint estimation. *Applied sciences*, 9(13):2683.

Silke Matthes, Thomas Hanke, Anja Regen, Jakob Storz, Satu Worseck, Eleni Efthimiou, Athanasia-Lida Dimou, Annelies Braffort, John Glauert, and Eva Safar. 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. Satellite Workshop to the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*.

Euan McGill and Horacio Saggion. 2023. BSL-Hansard: A parallel, multimodal corpus of English and interpreted British Sign Language data from parliamentary proceedings. In *Proceedings of the Second International Workshop on Automatic Translation for Signed and Spoken Languages*, pages 38–43, Tampere, Finland. European Association for Machine Translation.

Johanna Mesch and Lars Wallin. 2012. From meaning to signs and back: Lexicography and the Swedish Sign Language Corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 123–126.

Mathias Müller, Malihe Alikhani, Eleftherios Avramidis, Richard Bowden, Annelies Braffort, Necati Cihan Camgöz, Sarah Ebling, Cristina España-Bonet, Anne Göhring, Roman Grundkiewicz, Mert Inan, Zifan Jiang, Oscar Koller, Amit Moryossef, Annette Rios, Dimitar Shterionov, Sandra Sidler-Miserez, Katja Tissi, and Davy Van Landuyt. 2023. Findings of the second WMT shared task on sign language translation (WMT-SLT23). In *Proceedings of the Eighth Conference on Machine Translation*, pages 68–94, Singapore. Association for Computational Linguistics.

Mathias Müller, Sarah Ebling, Eleftherios Avramidis, Alessia Battisti, Michèle Berger, Richard Bowden, Annelies Braffort, Necati Cihan Camgöz, Cristina España-bonet, Roman Grundkiewicz, Zifan Jiang, Oscar Koller, Amit Moryossef, Regula Perrollaz, Sabine Reinhard, Annette Rios, Dimitar Shterionov, Sandra Sidler-miserez, and Katja Tissi. 2022. Findings of the first WMT shared task on sign language translation (WMT-SLT22). In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 744–772, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Carol Neidle and Christian Vogler. 2012. A new web interface to facilitate access to corpora: Development of the ASLLRP data access interface (DAI). In *Proc. 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC*, volume 3. Citeseer.

Annika Nonhebel, Onno Crasborn, and Els van der Kooij. 2004. Sign language transcription conventions for the ECHO project. In *Proceedings of the Annotation Convention*, volume 9, page 20.

Bowen Shi, Aurora Martinez Del Rio, Jonathan Keane, Diane Brentari, Greg Shakhnarovich, and Karen Livescu. 2019. Fingerspelling recognition in the wild with iterative visual attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5400–5409.

Mirella De Sisto, Vincent Vandeghinste, and Dimitar Shterionov. 2023. A New English-Dutch-NGT Corpus for the Hospitality Domain. In *Proceedings of the Second International Workshop on Automatic Translation for Signed and Spoken Languages*, pages 34–37, Tampere, Finland. European Association for Machine Translation.