SwissSLi: the Multi-parallel Sign Language Corpus for Switzerland

Zifan Jiang, Anne Göhring, Amit Moryossef, Rico Sennrich, Sarah Ebling
University of Zurich
jiang@cl.uzh.ch

Abstract

In this work, we introduce SwissSLi, the first sign language corpus that contains parallel data of all three Swiss sign languages, namely Swiss German Sign Language (DSGS), French Sign Language of Switzerland (LSF-CH), and Italian Sign Language of Switzerland (LIS-CH). The data underlying this corpus originates from television programs in three spoken languages: German, French, and Italian. The programs have for the most part been translated into sign language by deaf translators, resulting in a unique, up to six-way multi-parallel dataset between spoken and sign languages. We describe and release the sign language videos and spoken language subtitles as well as the overall statistics and some derivatives of the raw material. These derived components include cropped videos, pose estimation, phrase/sign-segmented videos, and sentence-segmented subtitles, all of which facilitate downstream tasks such as sign language transcription (glossing) and machine translation. The corpus is publicly available on the SWISSUbase data platform for research purposes only under a CC BY-NC-SA 4.0 license.

Keywords: sign language, sign language processing, multi-parallel corpus

1. Introduction

Sign languages are the primary means of communication for up to 70 million deaf persons (World Health Organization, 2021; World Federation of the Deaf, 2022). They use the visual-gestural modality to convey meaning through manual articulations in combination with non-manual elements such as the face and body (Sandler and Lillo-Martin, 2006).

This work introduces SwissSLi, the first sign language corpus that covers all three sign languages in Switzerland and their corresponding spoken languages. The corpus is valuable in that: (1) it contains parallel sign language videos and spoken language subtitles; (2) it consists primarily of translation as opposed to interpretation data, the former produced without time pressure; (3) it contains most data from deaf as opposed to hearing signers; (4) it is partly multi-parallel, which is distinctly rare in the domain of sign language data.

The corpus serves linguistic research and development of computational models in sign language processing (SLP) (Bragg et al., 2019; Yin et al., 2021). It is publicly available on the SWISSUbase data platform[1] under a CC BY-NC-SA 4.0 license.

2. Background

Data is essential to solve natural language processing tasks especially when deep learning (Goodfellow et al., 2016) is employed. Current efforts in SLP are often limited by a severe lack of training material. In this section, we categorize current resources and summarize related works in Table 1.

[1]https://www.swissubase.ch/en/catalogue/studies/20709/19836/overview

2.1. Sign Language Dictionaries

Bilingual dictionaries for sign languages such as the one provided by Sehyr et al. (2021) map a spoken language word or short phrase to a sign language video. Dictionaries may help create lexical mappings between languages or a crude sign-by-sign translation system (Moryossef et al., 2023b). However, they do not take account of the differences in grammar between sign languages and spoken languages and the usage of signs in context.

SpreadTheSign[2] is a notable multilingual dictionary containing around 23,000 words with up to 41 different spoken-sign language pairs and more than 600,000 videos in total. From it, Yin et al. (2022) collected ten thousand simple sentences in ten spoken languages that correspond to 14 hours of signing in ten sign languages to build a multilingual sign language translation dataset called *SP-10*.

2.2. Continuous Sign Corpora

Continuous sign corpora (summarized in Table 1) typically contain sequences of signs aligned with spoken language sentences. Available corpora are limited in orders of magnitude fewer sentence pairs than similar corpora for spoken languages. For example, the training data for the first WMT shared task on sign language translation, 17,000 parallel examples for DSGS-German (Müller et al., 2022), is negligible compared to the data for the general machine translation shared task of the same year, with 296 million parallel examples for English-German (Kocmi et al., 2022). Consequently, current sign language translation systems are barely usable at the moment (Müller et al., 2022, 2023).

[2]https://www.spreadthesign.com/

LREC-COLING 2024, pages 15448–15456
20-25 May, 2024.

Dataset	Languages	Signing mode	Signers	Domain	Duration (h)
KETI (Ko et al., 2019)	KVK/Korean	translation	deaf	emergency situations	28
CSL-Daily (Zhou et al., 2021)	CSL/Chinese	translation	deaf	daily lives	23
PHOENIX (Forster et al., 2014)	DGS/German	interpretation	hearing	weather forecasts	11
Public DGS Corpus (Hanke et al., 2020)	DGS-German	source	deaf	elicited dialogues	50
How2Sign (Duarte et al., 2021)	ASL/English	translation	mixed	instructional monologues	79
OpenASL (Shi et al., 2022)	ASL/English	mixed	mixed	news, vlogs on YouTube	288
YouTube-ASL (Uthus et al., 2024)	ASL/English	mixed	mixed	mixed on YouTube	984
BOBSL (Albanie et al., 2021)	BSL/English	interpretation	hearing	mixed TV programs	1,467
SP-10 (Yin et al., 2022)	10 sign/10 spoken	translation	deaf	dictionary entries	14
AfriSign (Gueuwou et al., 2023b)	6 sign/English	translation	deaf	Bible verses	152
JWSign (Gueuwou et al., 2023a)	98 sign/51 spoken	translation	deaf	Bible verses	2,530

Table 1: Well-known continuous sign language datasets, separated into two classes: bilingual (top) and multilingual (multi-parallel) (bottom). *KVK = Korean Sign Language, CSL = Chinese Sign Language, DGS = German Sign Language, ASL = American Sign Language, BSL = British Sign Language*. *SP-10*, *AfriSign* and *JWSign* contain multiple sign and spoken languages. *AfriSign* is a subset of *JWSign*.

Signing Mode and Data Quality We break down the corpora in Table 1 into different levels of data quality based on the signing mode and the hearing status of the signers that produce the content.

Firstly, we define data where sign language represents the *source*, i.e., data produced without undergoing any kind of translation and free from the effect of translationese (Graham et al., 2020), to be of superior quality. However, this kind of data is scarce[3]. The Public DGS Corpus is a notable resource where sign language serves as the *source*, though elicited in a studio recording environment.

On the other hand, signed *interpretation* of spoken language content as introduced in PHOENIX and BOBSL are more readily found. While larger in quantity, these datasets bear three disadvantages: (1) sign language interpreters to date are often hearing persons, who are frequently second-language learners of sign language rather than first-language signers; (2) under time pressure, they are especially prone to following the grammar of the spoken language they are more familiar with rather than that of sign language; (3) again due to time restrictions, they tend to skip certain pieces of information to make up for the time lag in interpreting.

Sign language *translation* data such as KETI, CSL-Daily, How2Sign, and the corpus introduced in this paper represent a middle ground, which is usually transferred from spoken language to sign language in a non-live setting, i.e., without time pressure. They are commonly recorded in a studio.

Data Quality, Quantity, and Licensing There is tension between high-quality sign language data preferred by the deaf community and linguists and the great quantity of data required by data-hungry algorithms/models such as Transformers (Vaswani et al., 2017) for training SLP systems.

Similar to what happened for text corpora (Bañón et al., 2020), large sign language corpora of mixed content from the web have emerged. OpenASL and YouTube-ASL curate videos from YouTube, and AfriSign contains translations of English Bible verses into African sign languages from the Jehovah's Witnesses website. However, concerns may be raised regarding the licensing of video content in the public domain, especially relevant to people's faces and identities. Anonymizing data by blurring or blackening signers' faces is unfeasible, as the face carries linguistic information (see §1).

2.3. Sign Languages in Switzerland

The focus of this work is the three Swiss sign languages. DSGS has approximately 5,500 deaf L1 users and an estimated 13,000 hearing users (Ebling et al., 2018). LSF-CH and LIS-CH are varieties of the same languages that are used in France and Italy, with an estimated 1700 and 300 deaf users, respectively (Boyes Braem et al., 2012).

Known lexical resources for these three sign languages are DSGS iLex (Ebling and Boyes Braem, 2016) and Signsuisse (DSGS, LSF-CH, LIS-CH) (Jiang et al., 2023a). In contrast to the bilingual sign/spoken language landscape of most countries, the multilingual environment of Switzerland leads to the possibility of a multi-way parallel corpus.

3. Data Overview

Our corpus contains the sign language version of two sets of television programs in three spoken/sign languages released between 2020 and 2023 fall[4]. They were created by the Swiss Broadcasting Corporation (SRG) in spoken language and then translated into sign language.

[3] For example, on Swiss television, only the program *Signes* satisfies this requirement.

[4] The following text in the paper describes the statistics of the programs up to this point and matches the initial data release on SWISSUbase planned in April 2024.

Dataset	Languages	Signing mode	Signers	Domain	Duration (h)
MEDIAPI-SKEL (Bull et al., 2020)	LSF/French	source	deaf	deaf press	27
Content4All Weather (Camgöz et al., 2021)	DSGS/German	interpretation	hearing	weather forecasts	12
Content4All News (Camgöz et al., 2021)	DSGS/German	interpretation	hearing	general news	76
FocusNews WMT22 (Müller et al., 2022a)	DSGS/German	source	deaf	general news	19
Daily News WMT22 (Müller et al., 2022b)	DSGS/German	interpretation	hearing	general news	16
Daily News WMT23 (Jiang et al., 2023b)	DSGS/German	interpretation	hearing	general news	437
Ours	3 sign/3 spoken	translation	mixed	mixed	30

Table 2: Comparison of our corpus to existing corpora of the same languages in Switzerland. Our corpus contains all three sign languages used in Switzerland and their corresponding spoken languages.

Program	Languages	#episodes	#hours	#subtitle files	#subtitles	#sentences	#signers
mitenand	DSGS/German	46/85	5.85	84	5,257	4,287	2/9
Ensemble	LSF-CH/French	105/105	6.72	82	5,076	3,265	3/3
Insieme*	LIS-CH/Italian	35/98	6.70	86	5,167	3,161	2/6
Helveticus (DSGS)	DSGS/German	52/52	3.56	52	2,672	2,196	1/1
Helveticus (LSF)	LSF-CH/French	50/50	3.42	49	2,418	1,716	4/4
Helveticus (LIS)	LIS-CH/Italian	52/52	3.64	52	2,316	2,049	1/1
Total	-	340/442	29.89	405	22,906	16,674	13/24

Table 3: Statistics of the television programs until fall 2023. *program* includes the link to the sign language version of the program. *#episodes* denotes the number of single broadcast videos (usually one on a single day) and *#hours* denotes the total duration in hours. *#subtitle files* denotes the number of available associated subtitle files. *#subtitles* denotes the number of subtitle units and *#sentences* denotes the number of well-formed spoken language sentences in the subtitles. *#signers* denotes the number of signers with informed consent so far/in total, the same applies to the *#episodes* they appear in. *Insieme* is marked with * because its distribution agreement is still in progress as of this writing.

These programs cover a broad domain. *mitenand* (German/DSGS), *Ensemble* (French/LSF-CH), and *Insieme* (Italian/LIS-CH) report the work of charity organizations, are parallel and produced weekly; *Helveticus* in three languages covers historical content with animations and is produced irregularly. Most episodes originate from German and are then translated into other languages. We show parallel examples in Appendix A.

As shown in Figure 1, the raw data includes signing and subtitles linked to the original audio track. It is available online on the SRG websites and we re-distribute on the SWISSUbase platform the episodes that are translated by deaf signers from whom we have obtained informed consent.

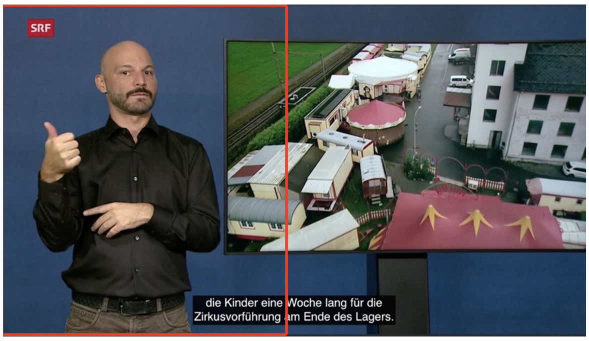

Figure 1: The *mitenand* program, translated by a deaf signer, comprised of DSGS signing (within the red bounding box) and German subtitles.

3.1. Comparison to Existing Corpora

We perform a comparison with other continuous signing corpora of the same languages in Table 2:

(1) Most existing corpora, including Content4All and Daily News, stem from the interpretation of Swiss television footage by hearing interpreters and are DSGS-German bilingual. The original subtitles are partially re-aligned and corrected by deaf signers to better accommodate the signing.

(2) FocusNews from the former deaf online channel *FocusFive* and MEDIAPI-SKEL from the deaf media company *Média-Pi* provide small but valuable deaf signing data with well-aligned subtitles.

(3) Our corpus fills the gap by adding translation data produced mostly by deaf signers. Despite not being manually re-aligned, the quality of signing and the default alignment between signing and subtitles is considered better[5] than interpreted data for the reasons (hearing interpreters, time pressure,

[5] To verify this intuition empirically, we analyzed the average offset between the subtitles and the signing in some episodes. The offset values (0.1 to 0.3 seconds) are substantially smaller than those in interpreted data such as Daily News WMT (around 1 second).

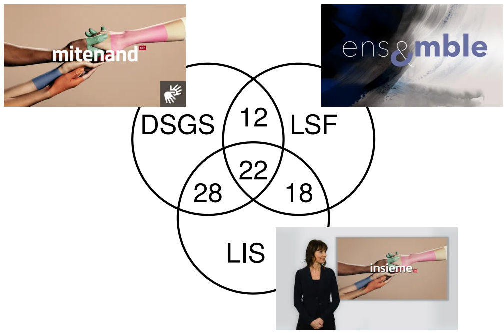

Figure 2: Venn diagram of the episode-level alignment between *mitenand*, *Ensemble*, and *Insieme*. The overlaps represent the number of parallel episodes between the respective languages collected until fall 2023.

and information loss) discussed in §2.2. In addition, our corpus contains all three Swiss sign languages.

3.2. Data Statistics

We present the detailed statistics of the six programs in Table 3[6]. The duration of each episode of all the programs is about four minutes. *#sentences* is derived from the subtitle files; the sentence segmentation process is discussed in §4.3.

Episode-level Alignment Due to inconsistencies in the title and release dates of parallel programs in different languages, we manually determined which episodes are aligned with each other. As illustrated by Figure 2, the alignment between *mitenand*, *Ensemble*, and *Insieme* is partial, while for *Helveticus*, it is almost complete, i.e., all episodes are aligned except two missing LSF ones (see Table 3).

4. Data Processing

This section describes the workflow to construct the corpus. It is highly automatic and reproducible to be continuously rerun for future data collection. Nevertheless, manual checks are involved in spotting and fixing incorrect and missing data.

4.1. Metadata Collection

The first step in our workflow is to collect metadata from the program websites. We use the *Selenium* library to automatically extract a list of episodes. Each entry has a title, a date, a description, a link to the web page, a link to the video, and a link to the subtitle, as well as a unique *id*. These metadata are stored in a master CSV file for each program and are included as part of the dataset release.

4.2. Video Processing

We use the *FFmpeg* software to download the videos and subtitles. The videos are encoded in H.264/*.mp4* format regardless of their original encoding; frame rate (FPS = 25.0) and video resolutions (1920x1080 or 1280x720) are inherited.

As shown by the red bounding box in Figure 1, the raw videos are then cropped into the left half region where the signer stands[7]. The step eliminates any potential interference of the program's original content with the subsequent pose estimation step.

Pose Estimation We use Mediapipe Holistic (Grishchenko and Bazarevsky, 2020) to estimate the signers' poses, i.e., the location of body keypoints, from the videos. The results are stored in a binary *.pose* format using the *pose* library (Moryossef et al., 2021). Each frame consists of 576 keypoints (33 pose landmarks, 468 face landmarks, 21 hand landmarks per hand, and 33 world landmarks).

The pose estimation step can be seen as an interpretable way of spatial downsampling or feature extraction from the dense video dimensionality of RGB frames that leads to more efficient downstream computation. One additional usage of pose estimation is that it potentially conceals the identity of the signers[8]. We refer to Isard (2020) for more about anonymizing sign language data.

Video Segmentation Based on the pose estimation, we run the sign language video segmentation model proposed by Moryossef et al. (2023a) to identify individual signs and phrases appearing in the videos. The segmentation model was originally trained and tested on the Public DGS Corpus (Hanke et al., 2020)[9]. The results are stored in *.eaf* ELAN (Wittenburg et al., 2006) files with two tiers for signs and phrases, respectively. This is a novel contribution since most current sign language corpora either do not have a segment-level annotation or have it done through human experts.

For our data, we expect the automatic segmentation quality to be the highest for DSGS data due

[6]The current data agreement with SRG excludes *Insieme* and only allows re-distributing episodes until May 2023. We will upload the missing part to SWISSUbase once we extend the agreement. Also, some *mitenand* episodes and *Insieme* before January 2023 are translated by hearing interpreters. For the initial release, we focus on deaf signer data and exclude them.

[7]This is congruent with our legal agreement with SRG.

[8]The extent to which poses are anonymous representations of signers is still under empirical investigation.

[9]The reported model performance in terms of *intersection over union* is 0.69 for signs and 0.85 for phrases, tested on a separate test set from Public DGS Corpus. The authors also tested the DGS-trained model on LSF data and it generalizes well under a zero-shot setting.

to the similarity of DGS and DSGS. We acknowledge that differences in signing mode (Public DGS Corpus: sign language as the source; ours: sign language as the translation target) may affect segmentation performance as well. That said, we are performing human correction to better understand the segmentation process and will include the latest results of both automatic and human segmentation in the data release.

4.3. Subtitle Processing

On the subtitle side, after normalizing the format into plain text and adjusting the wrong timecodes, we follow Müller et al. (2022) to convert the raw subtitle units into well-formed spoken language sentences and store them as "pseudo" *.srt* files where each subtitle unit is essentially a sentence.

5. Outlook

The corpus with the above-mentioned derivatives facilitates downstream tasks. On the sign level, the predicted sign candidates can be enriched by further gloss annotation (Johnston and De Beuzeville, 2016), either by hand or automatically by isolated sign language recognition (Adaloglou et al., 2021).

On the phrase level, the candidates can be used for refining the timings of the spoken language sentences in subtitle units (Renz et al., 2021), to create better-aligned parallel video-text example pairs.

Finally, the multi-parallel nature of the corpus offers the possibility for comparative linguistic study and specially designed multilingual sign language translation systems such as Yin et al. (2022).

6. Conclusion

We introduced a multi-parallel sign language corpus of three sign languages and three spoken languages in Switzerland. We believe that this corpus is a valuable asset for researchers within the fields of computer vision, natural language processing, and sign language linguistics.

7. Ethics Statement

All the signing videos that we re-distribute on the SWISSUbase data platform are for research purposes only under a CC BY-NC-SA 4.0 license and are shared with the informed consent of the signers involved. In addition, the videos and the subtitles are the subject of a data-sharing agreement between the SRG and the Department of Computational Linguistics, University of Zurich.

8. Acknowledgements

This work was funded by the Swiss Innovation Agency (Innosuisse) flagship IICT (PFFS-21-47). We also thank colleagues from SWISS TXT for their support in data collection.

9. Bibliographical References

Nikolas Adaloglou, Theocharis Chatzis, Ilias Papastratis, Andreas Stergioulas, Georgios Th Papadopoulos, Vassia Zacharopoulou, George J Xydopoulos, Klimnis Atzakas, Dimitris Papazachariou, and Petros Daras. 2021. A comprehensive study on deep learning-based methods for sign language recognition. *IEEE Transactions on Multimedia*, 24:1750–1762.

Samuel Albanie, Gül Varol, Liliane Momeni, Hannah Bull, Triantafyllos Afouras, Himel Chowdhury, Neil Fox, Bencie Woll, Rob Cooper, Andrew McParland, and Andrew Zisserman. 2021. BOBSL: BBC-Oxford British Sign Language Dataset.

Marta Bañón, Pinzhen Chen, Barry Haddow, Kenneth Heafield, Hieu Hoang, Miquel Esplà-Gomis, Mikel L. Forcada, Amir Kamran, Faheem Kirefu, Philipp Koehn, Sergio Ortiz Rojas, Leopoldo Pla Sempere, Gema Ramírez-Sánchez, Elsa Sarrías, Marek Strelec, Brian Thompson, William Waites, Dion Wiggins, and Jaume Zaragoza. 2020. ParaCrawl: Web-scale acquisition of parallel corpora. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4555–4567, Online. Association for Computational Linguistics.

Penny Boyes Braem, Tobias Haug, and Patty Shores. 2012. Gebärdenspracharbeit in der Schweiz: Rückblick und Ausblick. *Das Zeichen*, 90:58–74.

Danielle Bragg, Oscar Koller, Mary Bellard, Larwan Berke, Patrick Boudreault, Annelies Braffort, Naomi Caselli, Matt Huenerfauth, Hernisa Kacorri, Tessa Verhoef, et al. 2019. Sign language recognition, generation, and translation: An interdisciplinary perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 16–31.

Hannah Bull, Annelies Braffort, and Michèle Gouiffès. 2020. MEDIAPI-SKEL - a 2D-skeleton video database of French Sign Language with aligned French subtitles. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6063–6068, Marseille, France. European Language Resources Association.

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 *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–5. IEEE.

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.

Sarah Ebling and Penny Boyes Braem. 2016. Linking a web lexicon of DSGS technical signs to iLex. In *Proceedings of the LREC2016 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, pages 59–62, Portorož, Slovenia. European Language Resources Association (ELRA).

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, and Mathew Magimai-Doss. 2018. SMILE Swiss German sign language dataset. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. 2014. Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-weather. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1911–1916, Reykjavik, Iceland. European Language Resources Association (ELRA).

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Yvette Graham, Barry Haddow, and Philipp Koehn. 2020. Statistical power and translationese in machine translation evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 72–81, Online. Association for Computational Linguistics.

Ivan Grishchenko and Valentin Bazarevsky. 2020. Mediapipe holistic.

Shester Gueuwou, Sophie Siake, Colin Leong, and Mathias Müller. 2023a. JWSign: A highly multilingual corpus of Bible translations for more diversity in sign language processing. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 9907–9927, Singapore. Association for Computational Linguistics.

Shester Gueuwou, Kate Takyi, Mathias Müller, Marco Stanley Nyarko, Richard Adade, and Rose-Mary Owusuaa Mensah Gyening. 2023b. Afrisign: Machine Translation for African Sign Languages. In *4th Workshop on African Natural Language Processing*.

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).

Amy Isard. 2020. Approaches to the anonymisation of sign language corpora. 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 95–100.

Trevor Johnston and Louise De Beuzeville. 2016. Auslan corpus annotation guidelines. *Auslan Corpus*.

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.

Tom Kocmi, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Thamme Gowda, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Rebecca Knowles, Philipp Koehn, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Michal Novák, Martin Popel, and Maja Popović. 2022. Findings of the 2022 conference on machine translation (WMT22). In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 1–45, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Amit Moryossef, Zifan Jiang, Mathias Müller, Sarah Ebling, and Yoav Goldberg. 2023a. Linguistically motivated sign language segmentation. In *Findings of the Association for Computational*

Linguistics: EMNLP 2023, pages 12703–12724, Singapore. Association for Computational Linguistics.

Amit Moryossef, Mathias Müller, and Rebecka Fahrni. 2021. pose-format: Library for viewing, augmenting, and handling .pose files. https://github.com/sign-language-processing/pose.

Amit Moryossef, Mathias Müller, Anne Göhring, Zifan Jiang, Yoav Goldberg, and Sarah Ebling. 2023b. An open-source gloss-based baseline for spoken to signed language translation. In *Proceedings of the Second International Workshop on Automatic Translation for Signed and Spoken Languages*, pages 22–33, Tampere, Finland. European Association for Machine Translation.

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.

Katrin Renz, Nicolaj C Stache, Neil Fox, Gul Varol, and Samuel Albanie. 2021. Sign segmentation with changepoint-modulated pseudo-labelling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3403–3412.

Wendy Sandler and Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge University Press.

Zed Sevcikova Sehyr, Naomi Caselli, Ariel M Cohen-Goldberg, and Karen Emmorey. 2021. The ASL-LEX 2.0 Project: A Database of Lexical and Phonological Properties for 2,723 Signs in American Sign Language. *The Journal of Deaf Studies and Deaf Education*, 26(2):263–277.

Bowen Shi, Diane Brentari, Gregory Shakhnarovich, and Karen Livescu. 2022. Open-domain sign language translation learned from online video. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6365–6379, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Dave Uthus, Garrett Tanzer, and Manfred Georg. 2024. Youtube-asl: A large-scale, open-domain american sign language-english parallel corpus. *Advances in Neural Information Processing Systems*, 36.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. Elan: a professional framework for multimodality research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559.

World Federation of the Deaf. 2022. World federation of the deaf - our work. https://wfdeaf.org/our-work/.

World Health Organization. 2021. Deafness and hearing loss. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

Aoxiong Yin, Zhou Zhao, Weike Jin, Meng Zhang, Xingshan Zeng, and Xiaofei He. 2022. Mlslt: Towards multilingual sign language translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5109–5119.

Kayo Yin, Amit Moryossef, Julie Hochgesang, Yoav Goldberg, and Malihe Alikhani. 2021. Including signed languages in natural language processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7347–7360, Online. Association for Computational Linguistics.

Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, and Houqiang Li. 2021. Improving sign language translation with monolingual data by sign back-translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1316–1325.

10. Language Resource References

Jiang, Zifan and Moryossef, Amit and Müller, Mathias and Ebling, Sarah. 2023a. *Signsuisse DSGS/LSF/LIS lexicon*. LaRS - Language Repository of Switzerland.

Jiang, Zifan and Moryossef, Amit and Müller, Mathias and Ebling, Sarah. 2023b. *SRF DSGS Daily news broadcast: video and original subtitle data*. LaRS - Language Repository of Switzerland.

Müller, Mathias and Ebling, Sarah and Camgöz, Necati Cihan and Jiang, Zifan and Battisti, Alessia and Moryossef, Amit and Rios, Annette and Bowden, Richard and Wong, Ryan. 2022a. *WMT-SLT FocusNews: Training data for the WMT shared task on sign language translation*. Zenodo.

Müller, Mathias and Ebling, Sarah and Camgöz, Necati Cihan and Jiang, Zifan and Battisti, Alessia and Tissi, Katja and Sidler-Miserez, Sandra and Perrollaz, Regula and Berger, Michèle and Reinhard, Sabine and Moryossef, Amit and Rios, Annette and Bowden, Richard and Wong, Ryan and Ribback, Robin and Schori, Severine. 2022b. *WMT-SLT SRF: Training data for the WMT shared task on sign language translation (videos, subtitles)*. Zenodo.

A. Parallel Examples

In this section, we show the first few sentences of a randomly selected *Helveticus* episode in three languages as an example of the parallel sentences in the corpora[10]. The spoken language and sign language versions of the episodes are translations of a single source, the German episode.

Language	Sentences
German	Unsere Geschichte spielt um 1230 in den Regionen Uri und Tessin.
	Wir sind mitten in den Alpen unterhalb des Gotthards.
	Um auf die andere Seite eines Berges zu kommen, mussten damals sehr lange Wege zurückgelegt werden.
	Um gemeinsame Geschäfte abschliessen zu können, hatten sich die Leute vor und hinter dem Gotthard schon lange einen direkteren Weg gewünscht.
French	Notre histoire se passe vers 1230 dans les régions de Uri et du Tessin.
	Nous voici au beau milieu des Alpes, au pied du Gothard.
	A cette époque, pour aller de l'autre côté de la montagne, il faut la contourner par des chemins très longs.
	Pour faire du commerce ensemble, les habitants des deux côtés aimeraient une route plus directe.
Italian	La nostra storia si svolge nel 1230 in Uri e in Ticino.
	Siamo ai piedi del Gottardo.
	Per andare dall'altra parte della montagna si devono fare lunghe trade.
	Gli abitanti vorrebbero una via più diretta.
English*	Our story takes place around 1230 in the regions of Uri and Ticino.
	We are in the middle of the Alps below the Gotthard.
	To get to the other side of a mountain in those days, very long distances had to be covered.
	In order to be able to do business together, the people in front of and behind the Gotthard had long wished for a more direct route.

Table 4: Parallel examples in three languages, as well as an English translation from German by DeepL.

[10]We refer to the spoken languages, German, Italian, and French, as the varieties of these languages as used in Switzerland. For the German part, the audio content might contain both the Swiss German dialect and Swiss Standard German, and the subtitle text is always in Swiss Standard German, which is close to Standard German of Germany with some lexical differences. Swiss French and Swiss Italian are much closer to the standard varieties used in France and Italy, respectively, with even fewer lexical differences. Our corpus focuses on the subtitles (which are always in standard language) and the signing and does not include the audio.