Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 2541–2547
Marseille, 20-25 June 2022
© European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0

Crowdsourcing Kazakh-Russian Sign Language: FluentSigners-50

M. Mukushev[1], A. Kydyrbekova[1], A. Imashev[1], V. Kimmelman[2], A. Sandygulova[1]
[1]Department of Robotics and Mechatronics, School of Engineering and Digital Sciences
Nazarbayev University, Nur-Sultan, Kazakhstan
[2]Department of Linguistic, Literary, and Aesthetic Studies
University of Bergen, Bergen, Norway
{mmukushev, aigerim.kydyrbekova, alfarabi.imashev, anara.sandygulova}@nu.edu.kz
vadim.kimmelman@uib.no

Abstract

This paper presents the methodology we used to crowdsource a data collection of a new large-scale signer independent dataset for Kazakh-Russian Sign Language (KRSL) created for Sign Language Processing. By involving the Deaf community throughout the research process, we firstly designed a research protocol and then performed an efficient crowdsourcing campaign that resulted in a new FluentSigners-50 dataset. The FluentSigners-50 dataset consists of 173 sentences performed by 50 KRSL signers for 43,250 video samples. Dataset contributors recorded videos in real-life settings on various backgrounds using various devices such as smartphones and web cameras. Therefore, each dataset contribution has a varying distance to the camera, camera angles and aspect ratio, video quality, and frame rates. Additionally, the proposed dataset contains a high degree of linguistic and inter-signer variability and thus is a better training set for recognizing a real-life signed speech. FluentSigners-50 is publicly available at https://krslproject.github.io/fluentsigners-50/

Keywords: dataset, sign language processing, deaf, crowdsourcing

1. Introduction

Natural languages utilized largely by deaf populations across the world are known as sign languages. In sign languages, meaning is expressed by a series of motions involving the hands, torso, arms, head, and face. Sign languages, like spoken languages, contain several levels of linguistic structure, such as phonology, morphology, syntax, semantics, and pragmatics. Over 300 different sign languages have been recognized thus far (Koller, 2020).

One of the main challenges, as Bragg et al. (2019) correctly point out in their review of research in the field of Sign Language Processing (SLP), is related to significant limitations of public sign language datasets that limit the power and generalizability of recognition systems trained on them. SLP is a research area focusing on sign language recognition, generation, and translation. Apart from obvious dataset limitations such as the size of the vocabulary (due to expensive recording and annotation processes), most datasets only contain isolated signs (such as MS-ASL (Joze and Koller, 2018) and Devisign (Chai et al., 2014)), which are insufficient for most real-world use cases that require natural signing (continuous and spontaneous) and training on complete sentences and longer utterances (Bragg et al., 2019).

Another disadvantage of most presently used datasets is the lack of environmental heterogeneity since they are often recorded in the same setting(s) and have just one vocabulary domain, resulting in overfitting when applied to models that are architecturally more complex (Koller et al., 2016). As a result, several Continuous Sign Language Recognition (CSLR) techniques concentrate solely on cropped hands to make the task easier (Bragg et al., 2019). It loses vital verbal and grammatical information provided by body movements, facial expressions, and mouthing. Furthermore, many sign language datasets include inexperienced or non-native contributors (i.e., students), who sign slower and simplify the style and vocabulary to make the computer vision problem simpler, although of no real utility (Bragg et al., 2019).

This paper describes our methodology utilized to crowdsource a new large-scale Kazakh-Russian Sign Language dataset (FluentSigners-50) with the help of the Deaf community in Kazakhstan. By following rigorous ethical standards, we carefully designed a research protocol in close consultation and collaboration with the representatives of the Deaf community. This collaboration resulted in an efficient crowdsourcing campaign in which 50 dataset contributors understood the task and independently performed data collection. This paper proposes a crowdsourcing methodology for dataset collection and details our experience and key takeaways.

The objective of FluentSigners-50 is to address three shortcomings of commonly used datasets identified by Bragg et al. (2019): *continuous signing*, *signer variety*, and *native signers*. FluentSigners-50's main advantage is in its large signer variety: age (ranging from 8 to 57 years old), gender (18 male and 32 female), clothing, skin tone, body proportions, disability (deaf or hard of hearing), and fluency. Additionally, as the dataset was crowdsourced: the participants were using a variety of their own recording devices (such as smartphones and web cameras), it resulted in a large variety of backgrounds, lighting conditions, camera quality, frame rates, camera aspect ratios, and angles. Fi-

Datasets	Language	Signers	Deaf	Vocabulary	Samples	In the wild
The SIGNUM (Von Agris and Kraiss, 2007)	DGS	25	Yes	780	780	No
The RWTH-BOSTON-400 (Dreuw et al., 2008)	ASL	4	Yes	483	843	No
The RWTH-PHOENIX-Weather 2014T (Camgoz et al., 2018)	DGS	9	No	2887	8257	No
Video-Based CSL (Huang et al., 2018)	CSL	50	No	178	25000	No
The BSL-1K (Albanie et al., 2020)	BSL	40	Yes	1064	-	No
The How2Sign (Duarte et al., 2020)	ASL	11	Yes	16000	35000	No
FluentSigners-50	**KRSL**	**50**	**Yes**	**278**	**43250**	**Yes**

Table 1: Datasets used for Continuous Sign Language Recognition. This list excludes datasets of isolated signs. *Deaf* column indicates if deaf signers contributed to the dataset. *In the wild* column indicates if recording settings varied. *No* means that the settings were the same for all samples.

nally, FluentSigners-50 contains recordings of 50 contributors that use sign language on a daily basis: either deaf, hard of hearing, hearing CODA (Child of Deaf Adults), and hearing SODA (Sibling of a Deaf Adult). As a result, the dataset contains a certain degree of linguistic variability, including phonetic, phonological, lexical, and syntactic variations. Figure 1 demonstrates ten participants showcasing signer variety as well as video-related differences.

The remainder of this paper discusses Related Work, followed by descriptions of our crowdsourcing methodology for the data collection. We then briefly introduce the data itself. The paper concludes with guidelines about how future studies could perform crowdsourcing of sign language datasets.

2. Sign Language Datasets

Sign language datasets are of great importance in order to advance the tasks of SLP. For example, datasets recorded using standard cameras have direct utility in real-life situations. Such datasets contain videos of either isolated signs or continuous signing. Table 1 presents an overview of the most commonly used sign language datasets that are appropriate for the problem of CSLR with the inclusion of FluentSigners-50.

The high performance of deep learning methods for sign language recognition and translation tasks requires thousands of samples of data for training machine learning methods. Bragg et al. (2019) highlight that only a few publicly available and large-scale sign language corpora exist. Furthermore, they specify the main concerns of existing datasets: a relatively small vocabulary size, absence of spontaneous (real-life) signing, novice signers and interpreters (e.g., students), and lack of signers' variety. Because of the importance of fluency and the naturalness of signing, we should distinguish between datasets containing contributors whose experience in sign language is unknown (e.g., learned a few gestures for the sake of dataset collection) and signers who use sign language as their first language. Many datasets record professional interpreters who are often not native signers (i.e., CODA) (Mukushev et al., 2020). While being professional and fluent, the act of interpretation changes the execution (e.g., use of a calque or loan translation, i.e., a literal word-for-word translation). Additionally, datasets should differentiate between desired content and "real-life" signs (i.e., self-generated rather than prompted) (Bragg et al., 2019) and datasets collected *in the wild* (i.e., varying recording settings and devices).

RWTH-Phoenix-Weather-2014 (Koller et al., 2015) is a German Sign Language (DGS) dataset used as a benchmark for the most recent works in SLP. It features nine signers who performed sign language translations of the weather forecast on TV broadcasts. RWTH-Boston-400 (Dreuw et al., 2008) is one of the first CSLR benchmark datasets for American Sign Language (ASL). Nevertheless, it has only four signers present in the videos. In contrast, Video-Based CSL (Chinese Sign Language) (Huang et al., 2018) provides a large number of participants (n=50) involved in collecting the dataset. At the same time, they are all recorded in the same recording settings, and most participants seem to be unfamiliar with sign language as they sign in slow and artificial ways without involving any facial expressions. SIGNUM (Von Agris and Kraiss, 2007) is a signer-independent CSLR dataset of DGS with all participants being fluent in DGS and are either deaf or hard of hearing. However, all videos were shot with a single RGB camera in a supervised condition with the same lighting and uniform blue background. These concerns of existing datasets limit the accuracy and robustness of the models developed for SLR and their contribution to the challenges of real-world signing. More recent datasets aim to address most challenges of the previous datasets: BSL-1K (Albanie et al., 2020) provides the largest number of annotated sign data while How2Sign (Duarte et al., 2020) provides the largest vocabulary size. Similar to older datasets, they were either recorded in a controlled lab environment or extracted from the TV broadcast. From this perspective, FluentSigners-50 is the first sign language dataset that includes 1) a large signer variety recorded in various environmental conditions and 2) fluent sign language contributors (deaf, hard of hearing, CODA, or SODA). Future SLR and SLT models can now be benchmarked on more than one dataset, which will help build more reliable and applicable solutions.

Figure 1: Signers showing the sign HELLO

3. Deaf-friendly Methodology

3.1. Ethics

The Ethics Committee of Nazarbayev University approved this research. Written informed consent forms were accompanied with a video an experienced KRSL interpreter translating the written text into sign language. It included detailed explanations of the purpose of the research and descriptions of how the data would be stored, used, and shared. In particular, it was explained that the objective is to collect a publicly available dataset for the research community to download and use for either linguistics or machine learning purposes. Monetary compensation was equivalent to one full day's salary as data collection involved at least three hours (a maximum of six hours) per participant.

3.2. Research protocol

As advised by Singleton et al. (2015), there is a need for researchers to understand the implications of their research protocol from the Deaf community's perspective and to be aware that they are ethically accountable for fully debriefing the Deaf participants and for sharing with the Deaf community the findings of their research. To this end, to evaluate our proposed data collection protocol, we invited six professional sign language interpreters who work at the national television. They are native to KRSL since they were born and grew up in families with at least one deaf parent. At first, we planned to distribute the written sentences to participants for them to interpret them to KRSL with complete freedom for the selection of signs and their order. Since one of the main objectives for the dataset was to be appropriate for both linguists and machine learning researchers, the protocol was to obtain people's natural responses to common questions. For example, in response to the question "what is the weather like in winter?" people would respond similarly by saying it is cold and windy. However, it was decided against this idea. It would not be possible to achieve a well-performing machine learning architecture given the limited amount of data collected and its high risk of receiving different responses.

Therefore, it was decided to brainstorm and compose a set of phrases and sentences for the dataset that are commonly used in the Deaf community on various topics (e.g., greetings, introductions, family, profession, hobby, food, habits, and others) for a total of 173 sentences and provide their written translations in Russian and Kazakh languages for people to come up with their interpretations of these sentences without restricting them what signs or sign order to use. Again, the invited interpreters warned us that the linguistic variability would be too diverse if people had complete freedom of interpretation. Such diverse data and its limited size would not be appropriate for machine learning. Currently, machine learning architectures are not good enough to handle wide-range data and provide substantial performance only on limited-vocabulary datasets. Therefore, it was decided to record exemplary performances of the 173 translations in KRSL in addition to the welcome message, explanation of the task, and instructions for the recording settings. We recorded several interpreters and their way of translating the written sentences into KRSL to allow for linguistic variability. In addition, each person repeated each sentence five times to have more data for machine learning purposes. We used the Logitech C920 Pro web camera and a dark background (see Figure 1's top-left signer). We later distributed these videos to all other contributors of the dataset. Thus, for all remaining contributors (44 people), the task was to repeat one of the variations of the KRSL sentences they saw in the exemplary recordings

3.3. The Data

The FluentSigners-50 dataset consists of everyday conversational phrases and sentences in KRSL, the sign language used in the Republic of Kazakhstan. The summary of FluentSigners-50 dataset is presented in

Table 2. KRSL is closely related to Russian Sign Language (RSL) and some other sign languages of the ex-Soviet Union (Imashev et al., 2020). While no official research comparing KRSL with RSL exists, our observations based on our experience researching both languages are that they show a substantial lexical overlap and are entirely mutually intelligible (Kimmelman et al.,). The sentences and phrases of FluentSigners-50 represent the following sentence types: statements, polar questions, wh-questions, and requests. Table 3 presents a subset of sentences (translated to English).

Video resolution	Range
Number of Signers	50
Repetitions	5
Number of sentences	173
Video duration	2∼11 seconds
Body joints	Upper-body involved
Mean number of signs per sentence/phrase	4
Vocabulary size	278
Total number of videos	43250
Total number of hours	43.9 (∼150 raw)

Table 2: Statistics of the FluentSigners-50 dataset

ID	English translation
S000	Hello
S001	Hi
S002	How are you?
S003	How is your job?
S004	I am doing great
S005	I am all good
S006	I am fine
S007	I am doing terribly
S008	I am very bad
S009	What are you doing?
...	...
S166	What a wonderful day
S167	Today is so hot
S168	What is the weather right now?
S169	Is it snowing outside?
S170	Good weather
S171	It is very cold outside
S172	I love when it rains
S173	I do not like the heat
S174	I like the wind
S175	There is a very strong wind outside

Table 3: Subset of sentences used in the study (translated to English). Full list of sentences can be downloaded from the dataset's website

3.4. FluentSigners-50 Contributors

Given the importance of signer independence and signer variety, we involved the local Deaf community in FluentSigners-50 data collection.

The contributors to the dataset were recruited via word of mouth, and they were friends, relatives, or colleagues of the initial six interpreters. All contributors participated in the data collection voluntarily and signed an informed consent form accompanied by the video with KRSL translation to enable full accessibility. For underage participants consent was collected from the parents, who also were dataset contributors. All contributors received monetary compensation for their participation and agreed to have their data shared as a dataset. All FluentSigners-50 contributors use sign language on a daily basis as they are either deaf (N=32), hard of hearing (N=6), hearing SODA (N=3), or hearing CODA (N=9). All our signers are fluent in KRSL but might not be considered "native" since early acquisition may be necessary for developing native language abilities. Other factors, particularly the quality of language input, may play a role (Lu et al., 2016). According to this distinction, FluentSigners-50 has 30 CODA contributors (including nine hearing signers) and 20 who are not CODA (16 deaf, one hard of hearing, and three hearing SODA). We decided to name our dataset "FluentSigners-50" because all of our contributors use sign language daily, and it is their primary language of communication. They all came from various regions of Kazakhstan and are of different age and gender groups. Figure 2 shows the age distribution of participants, with ages ranging from 8 to 57 years old.

3.5. Crowdsourcing the data

The participants were asked to watch the pre-recorded sentences one by one and record themselves repeating each sentence five times. Such a data collection process did not require the presence of a researcher or an interpreter. Even though the signers were asked to repeat the pre-recorded KRSL sentences, many of them added their minor corrections. They performed the sentences in their way since they relied on their own communication experience, method of interpretation, etc. All collected videos have different quality and resolutions since they used their mobile phones or web cameras with varying backgrounds, lighting and illumination conditions, quality of the videos, camera aspect ratios and angles, distance to the camera, and frame rates making the FluentSigners-50 dataset diverse and realistic compared to other CSLR datasets. The filming process of each contributor took about 3.5 hours. Recorded videos were then shared with one of the researchers via Whatsapp, Google Drive, Mail.Ru cloud, or similar file-sharing solutions. The duration of all raw videos was more than 150 hours. Each video was carefully validated and annotated by one of the researchers, resulting in 43 hours of labeled trimmed materials. When some translations were missing or did not have five repetitions, the researcher contacted that contributor via Whatsapp messenger.

Signer independence is one of the main challenges

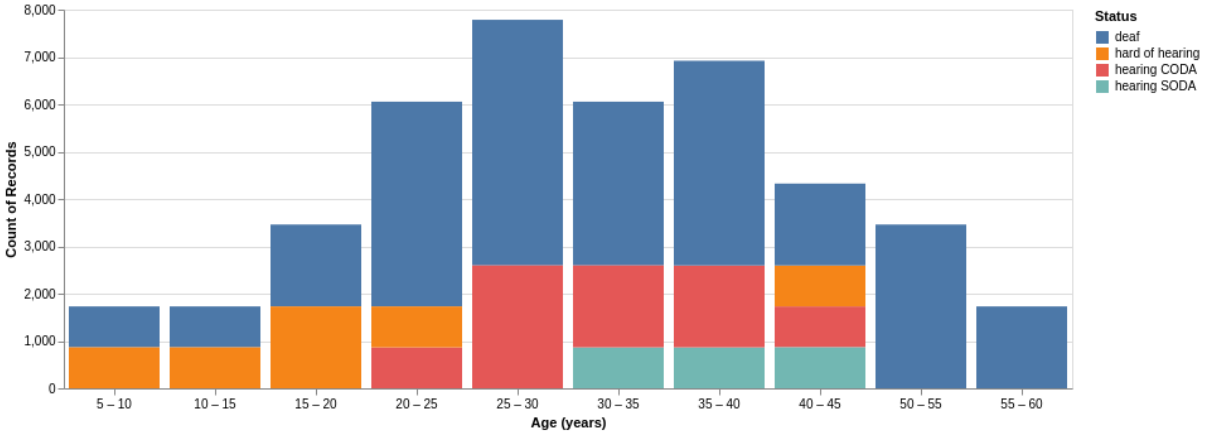

Figure 2: Number of videos per age distribution of participants

that must be addressed for the real-life value of machine learning models. Our dataset can help address this challenge as it provides both visual-related differences of each signer, such as variability in postures, distance to the camera, lighting, backgrounds, camera aspect ratios and angles, frame rates, quality, as well as linguistic-related differences of each signer, such as phonetic, phonological, lexical and syntactic variations.

4. Linguistic properties of the data

While a complete linguistic analysis of the data set is yet to be conducted, we can already observe a large amount of variation at different levels, as would also be expected in naturalistic sign language production (Schembri and Johnston, 2012). Phonetic and phonological variations are observed in many signs. For example, the sign HELLO is produced with 2, 3, or 4 movement repetitions by different signers, which is most likely phonetic variation. Lexical variation is also found in the dataset, where different lexical signs for the same concept are chosen by different signers. For example, in sentence 109, different lexical signs for 'adore' are used by different signers. Syntactic variation can be observed as well. Word order varies between signers: one pattern concerns the position of wh-signs; for instance, in sentence 65 'Where were you born?', the wh-sign WHERE occurs either in the initial or medial position, as also described for some other sign languages (Cecchetto, 2012).

5. Guidelines for future research

Reflecting on our overall experience, there emerge some guidelines about how future studies could conduct crowdsourcing involving Deaf community:

- Deaf community input early in the research stage helps identify potential issues and shape data collection methods. The initial consultation with the Deaf community allows for the design of a better research protocol.
- Deaf participants should be fully aware of the research purposes, procedure, and plans for sharing the data. In particular, a signed version of an informed consent form has to be provided. An interpreter has to be available for questions, etc. Researchers have to follow rigorous ethical standards to minimize potential ethical issues. Researchers are advised to perform debriefing and share their findings with the Deaf community.
- For people with low technological literacy, researchers or interpreters could be present or be available online to help with data collection.
- There might still be some contributors who need to be contacted to resend or recollect unclear or missing data. The process requires manual checking for quality.
- In some cases, contributors might need at least one exemplary video to minimize linguistic variability in signing. In other cases, contributors might be free to sign their interpretations or even answer open-ended questions to allow for the collection of natural and spontaneous signing.
- In collaboration with interpreters, researchers need to develop instructions, guidelines, or tips for participants to use during data collection. Some technical skills will be helpful both within and outside data collection.
- Although the Deaf community is eager to be involved in research, it is advised to compensate the contributors for their time.

6. Conclusion

This paper details our methodology used for data collection of the FluentSigners-50 dataset, a new large-scale Kazakh-Russian Sign Language dataset that aims to contribute to the development of continuous sign language recognition by introducing a new large-scale multi-signer benchmark. The main difference with

other sign language datasets is its large number of sign language contributors who are deaf, hard of hearing, and hearing CODA or SODA. Every video was recorded in different settings with varying backgrounds and lighting using their web or mobile phone cameras, which resulted in considerable variability in videos' resolution and frame rate. Additionally, the FluentSigners-50 dataset contains a high degree of linguistic and inter-signer variability and thus is a better training set for recognition of a real-life signed speech. The dataset is fully open and is available online at https://krslproject.github.io/fluentsigners-50.

7. Acknowledgements

We would like to thank the dataset contributors for agreeing to participate in data collection. This work was supported by the Nazarbayev University Faculty Development Competitive Research Grant Program 2019-2021 "Kazakh Sign Language Automatic Recognition System (K-SLARS)". Award number is 110119FD4545.

8. Bibliographical References

Albanie, S., Varol, G., Momeni, L., Afouras, T., Chung, J. S., Fox, N., and Zisserman, A. (2020). BSL-1K: Scaling up co-articulated sign language recognition using mouthing cues. In *European Conference on Computer Vision*.

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., and Ringel Morris, M. (2019). Sign language recognition, generation, and translation: An interdisciplinary perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, pages 16–31, New York, NY, USA. ACM.

Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., and Bowden, R. (2018). Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7784–7793.

Cecchetto, C. (2012). Sentence types. In Roland Pfau, et al., editors, *Sign language: An international handbook*, pages 292–315. De Gruyter Mouton.

Chai, X., Wang, H., and Chen, X. (2014). The devisign large vocabulary of chinese sign language database and baseline evaluations. *Technical report VIPL-TR-14-SLR-001. Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS*.

Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., and Ney, H. (2008). Benchmark Databases for Video-Based Automatic Sign Language Recognition. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., Torres, J., and Giro-i Nieto, X. (2020). How2sign: a large-scale multimodal dataset for continuous american sign language. *arXiv preprint arXiv:2008.08143*.

Huang, J., Zhou, W., Zhang, Q., Li, H., and Li, W. (2018). Video-based sign language recognition without temporal segmentation. *arXiv preprint arXiv:1801.10111*.

Imashev, A., Mukushev, M., Kimmelman, V., and Sandygulova, A. (2020). A dataset for linguistic understanding, visual evaluation, and recognition of sign languages: The k-rsl. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 631–640.

Joze, H. R. V. and Koller, O. (2018). MS-ASL: A Large-Scale Data Set and Benchmark for Understanding American Sign Language. *arXiv preprint arXiv:1812.01053*.

Kimmelman, V., Imashev, A., Mukushev, M., and Sandygulova, A.). eyebrow position in grammatical and emotional expressions in kazakh-russian sign language: A quantitative study. *PLOS ONE*.

Koller, O., Forster, J., and Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125.

Koller, O., Ney, H., and Bowden, R. (2016). Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3793–3802.

Koller, O. (2020). Quantitative survey of the state of the art in sign language recognition. *arXiv preprint arXiv:2008.09918*.

Lu, J., Jones, A., and Morgan, G. (2016). The impact of input quality on early sign development in native and non-native language learners. *Journal of Child Language*, 43(3):537–552.

Mukushev, M., Sabyrov, A., Imashev, A., Koishibay, K., Kimmelman, V., and Sandygulova, A. (2020). Evaluation of manual and non-manual components for sign language recognition. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association (ELRA).

Schembri, A. and Johnston, T. (2012). Sociolinguistic aspects of variation and change. In Roland Pfau, et al., editors, *Sign language: An international handbook*, pages 788–816. Mouton de Gruyter.

Singleton, J. L., Martin, A. J., and Morgan, G., (2015). *Ethics, Deaf-Friendly Research, and Good Practice When Studying Sign Languages*, chapter 1, pages 5–20. John Wiley Sons, Ltd.

Von Agris, U. and Kraiss, K.-F. (2007). Towards a video corpus for signer-independent continuous sign language recognition. *Gesture in Human-Computer*

Interaction and Simulation, Lisbon, Portugal, May, 11.