

Raising Awareness for a Korean Sign Language Corpus among the Deaf Community

Sung-Eun Hong¹, Seongok Won¹, Il Heo¹, Hyunhwa Lee²

¹Korea National University of Welfare, Department Sign Language Interpreting, Korea

²National Institute of Korean Language, Promotion Division of Special Languages, Korea
Sungeunhong2001@gmail.com, woonsok@knuw.ac.kr, heo1@knuw.ac.kr, lhh1127@korea.kr

Abstract

This paper contains strategies that need to be implemented before the sign language community can be involved in corpus work to raise awareness for the need of corpus work. The Korean Sign Language (KSL) Corpus Project began in order to create a linguistic corpus with 60 deaf native and near-native signers from the area of Seoul. In the process of building the KSL Corpus by collecting sign language data and annotating it the project was faced with the challenge that the concept of corpus was completely new to the Korean deaf community. The KSL Corpus Project developed three strategies in order to inform and explain what the KSL Corpus is about. First, the research project produced numerous KSL videos and posted them on social networking websites in a weekly rhythm. Second, the project organized a workshop, where only deaf people were invited to participate. Third, the KSL Corpus project selected prominent deaf people who were schooled and provided with corpus materials in order to inform others about KSL Corpus by connecting to their friends and families. The experiences and outcomes of the above strategies are of special importance since the data collection of the KSL Corpus is still in process.

Keywords: Korean Sign Language, corpus linguistics, involvement of deaf community, deaf workshop

1. KSL Corpus

The Korean Sign Language (KSL) is the sign language of the deaf people in South Korea. Due to the Korean Sign Language Act, which was enacted in 2016, KSL has now gained legal recognition and is the official language of deaf people in South Korea with its own rights and an equal status to the spoken Korean language (Hong et al., in press). One of the aims of the KSL Act is to protect the linguistic rights of deaf communities and KSL users through a variety of KSL-related research projects (Jung, 2016). One of these projects is the KSL Corpus Project which is funded by the National Institute of Korean Language and carried out by researchers from the Korea National University of Welfare.

The KSL Project has collected sign language data from 60 deaf signers in the area of Seoul. The spontaneous and elicited data (90 hours of raw data) has been tagged and annotated using ELAN (Wittenburg et al., 2006). ELAN is a professional tool used to create complex annotations on video and audio resources. It was developed by the Max-Planck-Institute. So far only the sign language data of informants from Seoul have been collected. It is the intention of the project to collect more KSL data of other areas in the future.

1.1 Data Collection

1.1.1 Informants

60 deaf native and near-native signers from the area of Seoul were invited in pairs. The informants have been recruited with the help of the Korea Association of the Deaf and its 25 offices in Seoul. In the process of selection we realized how important it was to recruit the informants with the help of the Deaf Association. The Korea Association of the Deaf has an important and central role within the deaf community and was able to reach out to the deaf people. In addition, the informants participated in the research project just because it was supported by the Deaf Association.

All informants were at least 19 years old. Their most used and most comfortable language is KSL. They have graduated from a deaf school (exceptions were made, when the informants are a child of deaf parents). They have either lived in Seoul for at least 10 years or they have lived close to Seoul and work in Seoul for at least 10 years. Lastly, they meet deaf people at least three times a week. Informants have been prioritized in the process of selection when they had deaf parents, siblings or partners and if they had acquired KSL before they entered school.

1.1.2 Elicitation Materials

The elicitation materials used in the KSL Corpus contain pictures, photographs, movie clips, animations, topics for an open conversation, signed videos and a combination of pictures and written words. These types of stimuli are processed in 13 tasks which are mainly based on the elicitation material of the DGS Corpus (Nishio et al., 2010). Most of the tasks from the DGS Corpus Project have been adapted to the Korean deaf culture. The elicitation materials have been tested twice before the data collection took place. Each task is introduced and explained in sign language. In order to ensure that all informants receive the same input, the instructions were presented in a video.

1.2 Annotation and Translation

Each session of this naturalistic, controlled and elicited signed language sample has a length of about three hours. This means that the complete recordings contain about 90 hours of sign language data. All the video material has been cut into the length of each task, converted, compressed into MPEG format and synchronized in order to use the data with ELAN. About two thirds of the data was translated into Korean by competent KSL-interpreters. And almost 17 hours of the KSL data has been annotated in ELAN. The annotation of the KSL Corpus Project can be seen as the first attempt in Korea to transcribe and annotate KSL data in a systematic and scientific way. The

KSL Corpus Project recruited numerous hearing and deaf transcribers. But perhaps different than sign language annotators in countries like Germany for example, the Korean annotators work at home and participate in a weekly annotation meeting where annotation problems are discussed and clarified.

2. Raising Awareness of the Need for a KSL Corpus among Deaf Community

When the data collection of the KSL Corpus Project started in 2015, we assumed that corpus work was unknown to deaf people just like the majority of hearing people. We explained to the informants what a corpus is and why a KSL Corpus is of great importance to the sign language community. However, we had the desire to reach more people than just the informants. Our goal was to raise awareness for a KSL Corpus and to let the deaf community know that the KSL Corpus is not just a research object of (hearing) linguists. But it is a language resource which belongs to the deaf community and deaf people should be proud of it, because KSL is unique and an independent full-fledged language with its own structure and grammar. The building of the KSL Corpus comes together with the KSL Act which was enacted in 2016 (Hong et al., in press). Because of this legislation, the KSL Corpus has received much attention from the deaf community. The strategies developed by the KSL project contain three different approaches, which all have the aim to inform and draw interest to the KSL Corpus.

2.1 KSL Videos

The KSL Project produced 16 KSL videos in order to reach out to the deaf community. First of all, a deaf member and a CODA member of our research team created six short KSL videos in which both were shown in a conversation about the following issues: What is a KSL corpus? What is annotation about? What does a data-collection session look like? Why is a corpus important? What are the advantages of having a KSL corpus? How can a corpus be used? Why is a KSL corpus an important contribution to deaf Culture? Why is it important that deaf people get involved in corpus research? The video clips each 5-6 min long were not recorded in a studio but were recorded in a coffee shop to keep a relaxed atmosphere.

	A: CODA researcher B: Deaf researcher
A: Hi!	
B: Hi!	
A: What are we doing here?	
B: We are sitting in front of the camera because there is a research project from the NIKL and the Korean University of Welfare. It's called KSL Corpus Project and we are here to talk about it and this video will be posted on Facebook. We hope that our video can inform about the project and communicate with people who are curious. It is possible to post questions under the video and we will answer the questions.	
A: Oh, I see! Well, there is not much I know... what could we talk about?	
(...) How did you sign corpus?	
B: I signed it this way. There is no KSL sign for corpus yet, but I use this sign. In the Netherlands they sign it like this and in Germany they sign it like this. I will use this sign.	
A: Ok (continues to repeat the sign for corpus). I understand. So, what is a corpus? I have heard about it recently, but to be honest I don't have a clue what it means.	
B: Corpus is a collection of a huge bunch of language data. But you cannot just go and collect it the way you want. There have to be certain criteria and you also have to transcribe them.	
A: Stop! What you mean by transcribe?	
B: To transcribe means in an easy way to write things down somehow. You could write down a sign by using a pen and sheet of paper, but if you do it that way, it's not really searchable. So the writing should be done on a computer.	
A: Wait, lots of my friends have recorded KSL data with their cameras. Do they all have a corpus then?	
B: No. Having KSL videos doesn't mean you have a corpus. A Corpus means...	

Figure 1 : Part of the English translation of a KSL video

The deaf colleague explained the above points in KSL and the CODA colleague asked questions (also in KSL) and stopped the deaf colleagues when things got too complicated. The language of both was kept easy and casual and technical terms were avoided as much as possible (see fig. 1).

These videos were uploaded on our Facebook page in a weekly rhythm (see fig. 2). Facebook is the most used social media within the deaf community in Korea and the numbers of people who saw, commented and shared the KSL videos were very satisfying to us. The viewers were asked to formulate questions about the content and the research team answered them on Facebook openly. The research project posted numerous other KSL videos in order to announce events (e.g., deaf workshop), give updated information about the research project or to find deaf informants or deaf annotators for example.

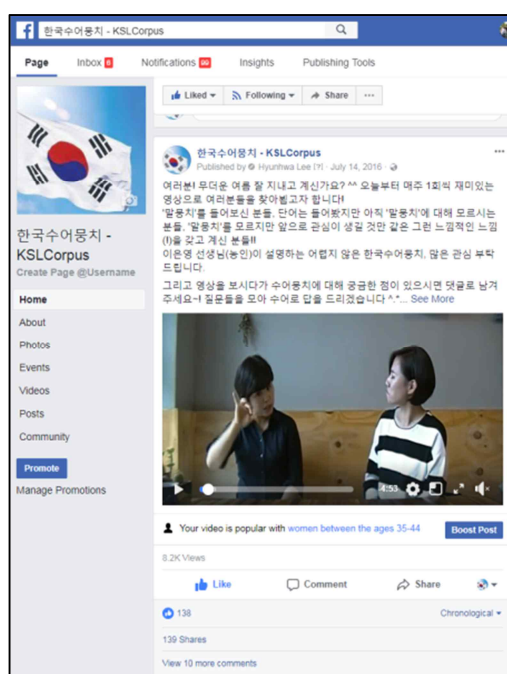


Figure 2: KSL video posted on the project Facebook page

2.2 Deaf Workshop

The second strategy was to organize a workshop in which only deaf people were invited to attend (fig. 3). This workshop contained three parts: A KSL talk on the importance of a KSL corpus, small group work on different aspects of a KSL corpus and a poster session. The KSL talk explained the importance of a KSL corpus in detail. It was possible to explain things in context and show examples by presenting visual material which aided in the understanding of the concepts being presented. While the KSL videos were kept very casual, the talk on the deaf workshop had an academic character. After the talk, the workshop participants were asked to join small groups (fig. 3) which discussed issues such as: Do we need standardization in KSL? What is the role of the KSL corpus concerning standardization? Does KSL corpus succeed in language documentation as well as documenting the culture of the deaf? What happens if “wrong” signs are included in the KSL corpus? These topics were chosen because we had the feeling that these things were currently discussed in the deaf community

which could be also seen by the comments on Facebook. Especially standardization is always an issue among deaf people in Korea, because various sign language dictionaries in the past claimed to present standardized signs, but couldn't support it with language data (Lee, 2017).



Figure 3: Official poster of the Deaf Workshop

All small groups (fig. 4) were moderated by a deaf or CODA project member, who received special training. The small groups were equipped with marker pens and white posters and were asked to draw the results of their discussion. This method seemed more deaf-friendly than using written language. After the discussion time a participant of each small group presented the results by showing and commenting on the small group poster to all workshop participants.



Figure 4: Small group session at the Deaf Workshop

The breaks in the workshop were used as a poster session. The research group prepared several posters with different topics. The issues were presented simply and visually. Additionally, deaf project members explained the posters.

Figure 5 shows one of the workshop posters, which explains in what ways a KSL corpus benefits the deaf community. Issues like regional or age variation of signs, corpus-based sign language research, preservation of KSL, corpus as a language resource, and other relevant concerns are presented visually on the poster. Figure 6 shows the studio setup of the data collection.

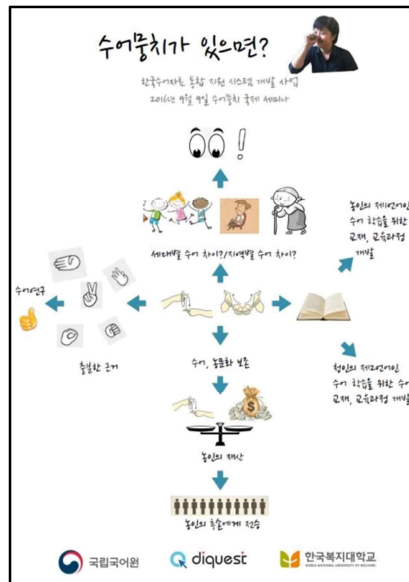


Figure 5 : Poster at the Deaf Workshop

The deaf workshop received much attention and positive feedback in the deaf community¹. The decision to exclude hearing people (which was taken seriously, the hearing researchers were also not allowed to participate actively at the workshop) helped many deaf people to express themselves curiously, critically and openly.

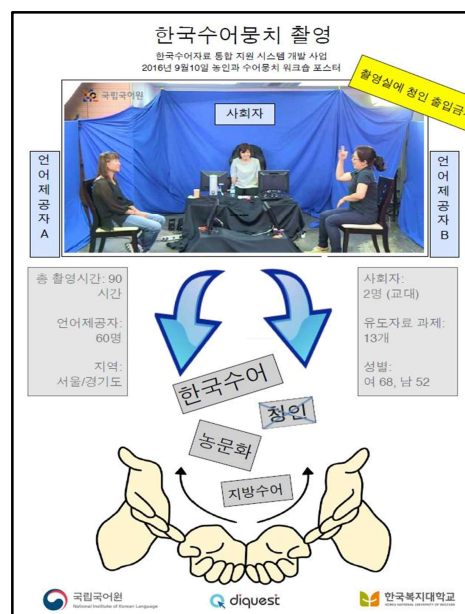


Figure 6 : Poster at the Deaf Workshop

¹ Deaf participants were asked if they liked and what they liked about the workshop. Their answers were filmed and edited to a video which was posted on Facebook.

2.3 Schooling of Prominent Deaf People

The third strategy had the aim to educate prominent deaf people and to send them to their deaf communities to inform other deaf people about sign language corpus work. We chose deaf people who had leadership roles and/or a significant influence in the deaf community. These chosen prominent deaf members participated at the deaf workshop and were additionally educated by the research team. We provided them with teaching material such as KSL videos, posters, pictures and PowerPoint presentations and asked them to go to their circle of friends and acquaintances and to meet 3-5 deaf people for about an hour to inform them about what a sign language corpus is about and what it can accomplish for the deaf community. Each educated deaf member was asked to arrange several of these sessions in places of their choice (e.g., at home, at church, in coffee shops) and to record one minute of each session on video (see fig. 7). This strategy was developed for two reasons: first, this way of information also reaches out to deaf people in rural areas so long as the prominent deaf person comes from a rural area. Since our deaf workshop took place only in Seoul, there was no possibility for us to reach deaf people in the provinces. Second, this method reached people who weren't always well connected to the mainstream deaf community. Through this method, people got to know about the KSL corpus in a very private and intimate atmosphere from a person they trusted.



Figure 7: Deaf person telling his friends about the KSL Corpus

3. Conclusion

All three strategies had the aim to raise awareness of a sign language corpus in the deaf community. The KSL video method reached mostly young people because the videos were spread through an internet medium. This also had the advantage of being locally unbounded. The deaf workshop was a great opportunity to assemble numerous deaf people in order to discuss corpus-related matters. It also turned out to have a great impact on the deaf community because it was the first event in the framework of sign language research excluding hearing people and showed that a sign language corpus is about deaf people (and not about hearing people doing sign language research). Schooling prominent deaf people was an approach to meet more specific deaf communities in a private surrounding. Depending on the person who was educated, the groups of people he or she would meet would be specific to their community. An elder deaf person would probably meet elder people and a deaf

person from the country would be meeting people from the country. The feedback of the deaf participants concerning the above methods was highly positive. Many deaf people were thrilled to experience sign language research in a visual and deaf-friendly way. The main point of the introduced strategies is that all these things cannot be done without very experienced deaf researchers. Although the concept of the strategies was developed together in the project, the three strategies couldn't have been carried out without the professional role of the deaf researchers. It is not a new finding that promoting and training of deaf researchers is beneficial. But this is especially true when the aim is to involve the deaf community in corpus work. The KSL Corpus Project aims to train more deaf persons and to find the deaf leaders who can actively involve the deaf community to the KSL Corpus.

4. Acknowledgements

This work is supported by the National Institute of Korean Language. We acknowledge our deaf colleagues Eun-Young Lee and Kang-Suk Byun, whose role in this part of research was essential.

5. Bibliographical References

- Hong, S.-E., Lee, H., Lee, M.-H., Byun, S.-I. (In Press). The Korean Sign Language Act. In Maartje De Meulder, Joseph J. Murray and Rachel McKee, editors, *Recognizing Sign Languages: An International Overview of National Campaigns for Sign Language Legislation and their Outcomes*. Multilingual Matters.
- Jung, H. (2016). Korean Sign Language Policies: Trends & Outlook. International Conference "Sign Language and Social Communications", National Institute of Korean Language and Korea Deaf Association, Seoul.153-167.
- Lee, H. 이현화 (2017). 한국수어 정비 사업: 한국수어사전(한국수어의 기록). Korean Sign Language Inspection Project: Korean Sign Language Dictionary(Documentation of Korean Sign Language). 새국어생활, *New language life*. 27-2, 국립국어원, National Institute of Korean Language, pp. 67-84.
- Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T., Rathmann, C. (2010). Elicitation methods in the DGS (German Sign Language) corpus project. In *Proceedings of the 4th workshop on representation and processing of sign languages: corpora and sign language technologies*. European Language Resources Association, pp. 178-185.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pp. 1556-1559.