Automatic Gloss-level Data Augmentation for Sign Language Translation

Jin Yea Jang[⋆,§]**,Hanmu Park**[§]**,Saim Shin**[§]**,Suna Shin**[†]**,Byungcheon Yoon**[‡]**,Gahgene Gweon**[⋆]

[⋆]Department of Intelligence and Information, Seoul National University
[§]AIRC, Korea Electronics Technology Institute
[†]Sign Language Linguistics Laboratory, Korea Nazarene University
[‡]Department of Sign Language Interpreting, Korea Nazarene University
[⋆]1 Gwanak-ro, Gwanak-gu, Seoul , Republic of Korea
[§]22 Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea
[†,‡]48 Wolbong-ro, Seobuk-gu, Cheonan-si, Chungcheongnam-do, Republic of Korea
{jinyea.jang, ggweon}@snu.ac.kr,{hanmu,sishin}@keti.re.kr,ksl_power@daum.net,ybch5778@kornu.ac.kr

Abstract

Securing sufficient data to enable automatic sign language translation modeling is challenging. The data insufficiency issue exists in both video and text modalities; however, fewer studies have been performed on text data augmentation compared to video data. In this study, we present three methods of augmenting sign language text modality data, comprising 3,052 Gloss-level Korean Sign Language (GKSL) and Word-level Korean Language (WKL) sentence pairs. Using each of the three methods, the following number of sentence pairs were created: blank replacement 10,654, sentence paraphrasing 1,494, and synonym replacement 899. Translation experiment results using the augmented data showed that when translating from GKSL to WKL and from WKL to GKSL, Bi-Lingual Evaluation Understudy (BLEU) scores improved by 0.204 and 0.170 respectively, compared to when only the original data was used. The three contributions of this study are as follows. First, we demonstrated that three different augmentation techniques used in existing Natural Language Processing (NLP) can be applied to sign language. Second, we propose an automatic data augmentation method which generates quality data by utilizing the Korean sign language gloss dictionary. Lastly, we publish the Gloss-level Korean Sign Language 13k dataset (GKSL13k), which has verified data quality through expert reviews.

Keywords: sign language translation, sign language gloss, data augmentation

1. Introduction

Research on automatic sign language translation technologies to improve accessibility for the hearing impaired is being actively conducted (Camgoz et al., 2018; Ko et al., 2019; Miyazaki et al., 2020). A two-stage model, which is shown in Figure 1, is a popular translation model. The first stage recognizes sign language gloss (Ormel et al., 2010) units (Sincan and Keles, 2020; Imashev et al., 2020), and the second stage translates gloss sequences into target-language sentences. The two-stage model has two main advantages. First, such a model can alleviate the out-of-vocabulary problem when using a separate sign language gloss dictionary. Second, a two-stage model is flexible, in that various NLP techniques can be applied independent of the first stage.

To ensure adequate performance of the two-stage model, sufficient data, especially gloss-level text data is required. However, accumulating sufficient sign language data is challenging due to the lack of expert-annotated datasets. Herein, we present new methods for automatically augmenting translation data on Gloss-level Korean Sign Language (GKSL) and Word-level Korean Language (WKL) sentence pairs.

Gloss-level data augmentation methods are rare. A related study (Moryossef et al., 2021) exists; however, only rule-based methods were applied. A rule-based method requires extensive prior linguistic knowledge
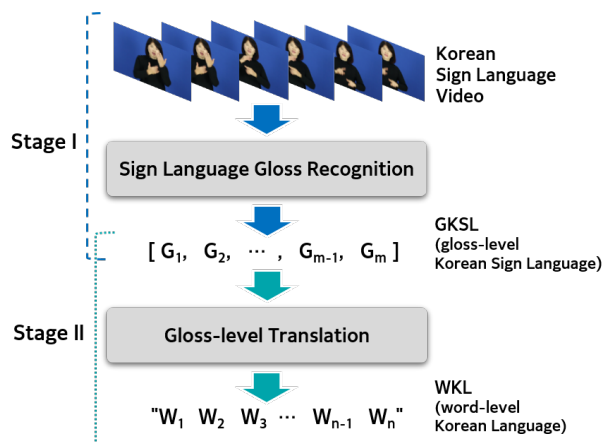

Figure 1: The two-stage sign language translation model.

of the target language. However, rule-based methods have difficulty handling irregular rules. If we utilize target-language linguistic knowledge via a large-scale pre-trained language model, the a priori burden can be reduced to facilitate processing, particularly when applying modern NLP techniques (Yin et al., 2021).

In this study, we applied three methods of gloss-level text data augmentation: blank replacement (BR), sentence paraphrasing (SP), and synonym replacement
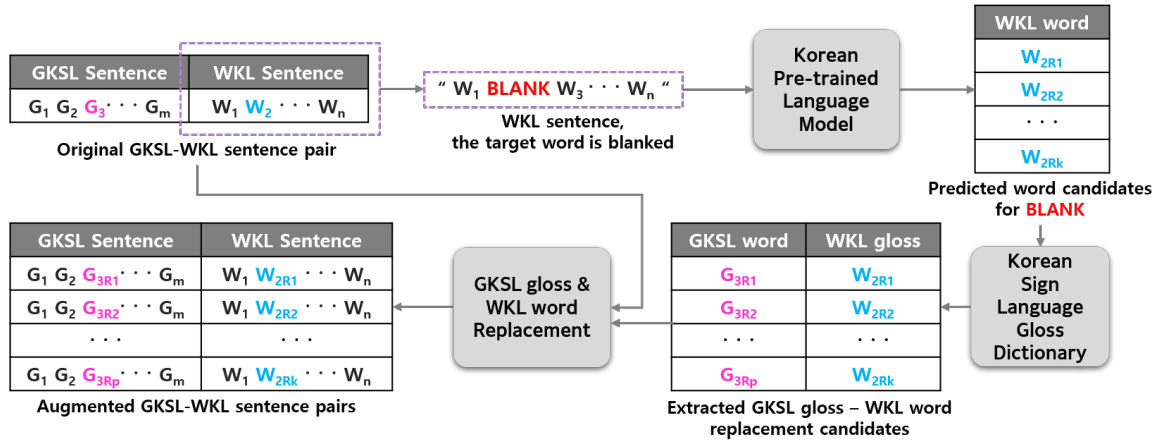
Figure 2: The procedure of Blank Replacement (BR) method.

(SR). Among the three methods, BR produces the largest amount of automatically augmented data using linguistic knowledge from a pre-trained language model. Quantitative and qualitative evaluations are used to evaluate the generated augmented data.

The three contributions of this study are as follows. First, we demonstrate that three different augmentation techniques used in existing NLP methods can be applied to sign languages. Second, an automatic data augmentation method is proposed that generates quality data by utilizing the Korean sign language gloss dictionary. Lastly, the Gloss-level Korean Sign Language 13k data (GKSL13k) is published, verified by experts for data quality.

2. Automatic Augmentation Methods

The sign language translation data in this study were augmented using BR, SP, and SR. Each of these is discussed in the following subsections.

2.1. Blank Replacement (BR)

BR leverages the language knowledge of a large-scale language model (Devlin et al., 2019; Liu et al., 2019) with Transformer (Vaswani et al., 2017) structure as the backbone. It learns linguistic knowledge in units of sentences with a word object that matches a masked word (i.e., a blank) in the input sentence during training. This model extracts candidate words to fit the blank such that the sentences are made plausible. Currently, language models for WKL exist; however, those for GKSL do not. Therefore, the blanks for data augmentation were determined based on WKL.

Because BR produces a candidate word to fill a blank in the sentence, the new sentence may be semantically different from the original. If the meaning of the WKL sentence changes, the GKSL version of the sentence must be accordingly modified. To this end, we used the Korean sign language gloss dictionary to determine the gloss corresponding to the replacement word of the WKL sentence, and we modified the GKSL sentence

```
function GetReplacePair(W_candidate):
    if Lemma(W_candidate) in GKSLdict.keys() then
        W += W_candidate
        G += KSLdict(W_candidate)
    end
    return (W,G)
for each (GKSL,WKL) pair do
    if POS of W_n is in [NNG, NNP, NP, VV] then
        WT_candidate += W_n
    end
    (GT,WT) ← GetReplacePair(WT_candidate)
    for each (GT,WT) do
        if GT is not null then
            WR_candidate ← KorLM(WKL with WT blanked)
            (GR,WR) ← GetReplacePair(WR_candidate)
            if GR is not null then
                NewWKL ← WKL with WT replaced by WR
                NewGKSL ← GKSL with GT replaced by GR
                DataAugmented += (NewGKSL,NewWKL)
            end
        end
    end
end
```

Figure 3: Blank Replacement (BR) algorithm.

as well. We used our own Korean sign language gloss dictionary, which refers to the Korean Sign Language Dictionary [1] of the National Institute of Korean Language.

The overall procedure of the BR method is shown in Figure 2. First, one word of the WKL sentence is set to a blank and input into the Korean pre-trained language model. The model then outputs a list of k probabilistic candidates for filling the blank position such that the sentence is intelligible. If the Korean sign language gloss dictionary is searched using the candidate words, the corresponding p gloss lists are extracted, where $p \leq k$. If there is no corresponding gloss in the search result, the word outputted is not used for replacement. When a matching pair of a WKL word and GKSL gloss is found, a new translation pair sentence is created by filling the blank with the new word.

Data name	Domain	Number of GKSL-WKL pair
KETI-Emergency	emergency	105
NIA-2020	transportation	2,000
KETI-Airport	airport	397
KETI-Daily	daily	550
Total	-	3,052

Table 1: Statistics of the original data.

Augmentation method	Number of GKSL-WKL pair
Blank Replacement (BR)	10,654
Sentence Paraphrasing (SP)	1,494
Synonym Replacement (SR)	899
Total	13,047

Table 2: Statistics of the GKSL13k data.

More details of the BR algorithm are shown in Figure 3. `Lemma` is a function that lemmatizes words; GKSLdict represents the dictionary of Korean sign language gloss and Korean word pairs; POS means part-of-speech; `KorLM` is a pre-trained Korean language model; GKSL is a list of glosses, G_n and WKL is a list of words, W_n. To reduce the chance of generating poor-grammar sentences, the POS of the word-replacement target in the WKL sentence is limited to general nouns, proper nouns, pronouns, and verbs. We developed the BR augmentation model using KaKao Brain's PORORO (Heo et al., 2021) to leverage a Korean pre-trained language model and determine word candidates[2].

2.2. Sentence Paraphrasing (SP)

Translation data augmentation using paraphrasing is frequently applied to NLP (Sugiyama and Yoshinaga, 2019). Because one sentence can be expressed as several different sentences with the same meaning, their number can be increased to increase the size of the dataset and improve training. Back translation (Prabhu-moye et al., 2018) is often used to acquire paraphrased entries; in our case, we did so by translating a WKL sentence into English and translating the result back into WKL using the Google Cloud Platform translation API[3]. Respective translation pair data were created by matching paraphrased sentences with GKSL sentences.

2.3. Synonym Replacement (SR)

A thesaurus is another useful data augmentation tool (Wei and Zou, 2019). By replacing a word in a sentence with a synonym, the original meaning is maintained, but a new sentence is created. The number of additional sentences is determined based on the number of words targeted for replacement and the number of synonyms corresponding to each word.

In SR, augmentation is first attempted using WKL sentences from the paired translation data. When performing one SR, the number of words targeted for replacement in one sentence is limited to one; thus, the meaning of the newly created sentence is likely to be the same as that of the original. Consider, for example, the sentence, "The student is smart." The number of

words that can be replaced is four. A new WKL sentence is created by finding a synonym for each target word one by one and replacing them in turn with synonyms. To guarantee the quality of the newly generated WKL sentences and reduce the probability of generating ungrammatical sentences, the part-of-speech of the replacement target word is limited to general nouns, proper nouns, and pronouns. Owing to the nature of the agglutinative Korean language, the role of a word and its form change in a sentence, depending on the postposition or ending attached to the word. Thus, the careful application of SR is crucial.

To maintain consistency with GKSL sentences, replacements were performed with synonyms mapped to existing gloss matches in the Korean sign language gloss dictionary. New translation pair data were then generated by matching the generated WKL sentences with the GKSL sentences paired with existing ones.

3. Gloss-level Korean Sign Language 13k Data (GKSL13k)

The data used for augmentation comprised 3,052 pairs of GKSL and WKL sentences. Details of the data are summarized in Table 1. Based on the domain, the data were divided into emergency situations, using public transportation, using airports, and everyday situations. KETI-Emergency[4] and NIA-2020[5] sources contain sign language video clips. However, because this study only dealt with gloss-level text, video data were excluded.

Table 2 lists the statistics of the data generated using the three augmentation methods. In total, 13,047 new translation pairs[6] were produced automatically. The BR method, using GKSL and WKL sentences, generated 10,654 new translation pairs, providing the largest increase in quantity compared to those provided by SP and SR. For the latter two, fewer quantitative results were obtained compared with the size of the original data. SP, which utilizes back translation via WKL sentences, was expected to increase the quantity as a multiple of the number of pairs in the original data. However, the back translation results often

---

[2]https://kakaobrain.github.io/pororo/tagging/fill.html
[3]https://cloud.google.com/translate

[4]https://aihub.or.kr/opendata/keti-data/recognition-laguage/KETI-02-003
[5]https://aihub.or.kr/aidata/7965
[6]https://github.com/AIRC-KETI/GKSL-dataset

	GKSL→WKL	WKL→GKSL
Original	0.246	0.137
Original+SP	0.373	0.238
Original+SR	0.380	0.258
Original+BR	0.436	0.299
Original+All	**0.450**	**0.307**

Table 3: Sign language translation performance results: the numbers indicate the BLEU score.

Evaluation Label	Count
Correct	298
Incorrect	173
Borderline	29
Total	500

Table 4: Experts review results of 500 GKSL-WKL pair samples.

differed from the meanings of the original sentences, which necessitated manual filtering. Hence, 1,494 new pairs were obtained. For SR, even fewer augmentation results were obtained, probably because we limited replacement candidates to WKL synonyms using the same gloss from the Korean sign language gloss dictionary for quality.

4. Augmented Data Quality Evaluation

We used quantitative and qualitative methods to evaluate automatically augmented sign language translation pair data. Quantitatively, we compared model translation performance by training an automatic version with data before and after augmentation. Qualitatively, three sign language experts reviewed the augmented data. Each method is introduced in detail in the following sections.

4.1. Translation Model Performance

To evaluate the quality of the augmented data, the performance of the gloss-level sign language translation models trained using data before and after augmentation was compared. We also trained models according to the types of augmentation methods to determine which method contributed the most. Both the GKSL-to-WKL (which takes a GKSL sentence and produces a WKL sentence) and the WKL-to-GKSL (which receives a WKL sentence and generates a GKSL sentence) translation models were trained. The models were implemented using the small-size model of KE-T5 (Kim et al., 2021), a Korean pre-trained language model implemented using Google's T5 algorithm (Raffel et al., 2019), as the backbone. The model has an encoder-decoder structure, wherein the encoder generates a hidden representation vector for the input sentence and feeds it to the decoder; thereafter, the decoder generates the next token using the output of the encoder and the token information generated in the previous step. All ten models were trained through 50 epochs, and the performance of the validation set was measured at each epoch. The model the parameters showing the best performance were stored, and the performance of the test set was measured. The performance results of the translation model summarized in Table 3, reflecting the BLEU score (Papineni et al., 2002).

Regardless of the type of augmentation, the models trained with augmented data yielded better per-

formances compared with those trained with original data. Among the augmentation methods, BR exhibited the best performance improvements of about 19% and about 16% over GKSL-to-WKL model and the WKL-to-GKSL models, respectively. Apart from the size of the augmented data, BR was the only method that increased the vocabulary size of the GKSL gloss. A rich vocabulary is key to training adequate translation models (Gowda and May, 2020).

4.2. Experts Review

Three sign language experts reviewed the results of the BR method, wherein GKSL sentence modifications were performed. The first expert sign language reviewer is a professor in the Department of Sign Language Interpretation and has been teaching Korean sign language linguistics since 2005. The second expert has worked as a Korean sign language interpreter since being licensed in 2009. This person's spouse is Deaf[7]. The third expert is a Deaf person who graduated from schools for the Deaf. This person has worked for the Korea Deaf Association for 15 years.

They sampled 500 random BR results and labeled them "correct" if the automatically generated GKSL sentence was well structured according to the the WKL sentence meaning; otherwise, it was labeled "incorrect". If the sentence was difficult to judge, it was labeled as "borderline."

Table 4 summarizes the results of the expert review. Among the 500 samples, 60% of the GKSL sentences were judged to be correct, and approximately 65%, including borderlines, were judged plausible. Approximately 35% of the sentences were judged incorrect. According to the error analysis, the use of incorrect glosses; processing errors for homonyms; awkwardness caused by literal translation; loss of additional information (e.g., units after a fingerspelling gloss); and grammar errors of WKL sentences were the most common causes of incorrect.

Examples of each error or cause presented in Table 5. The first example uses the wrong gloss. In the first row of Table 5, instead of "무엇 [what]" in the GKSL sentence, "방법 [method]" should have been generated. In the second row, the homonyms error occurred because the Korean word "성" has several different mean-

---

[7] A capital "D" is used to designate a person whose original language is sign language

Type	GKSL sentence	WKL sentence
Use of wrong gloss	서울 가다 무엇 Seoul go what	서울로 가려면 어떻게 가나요? How do I get to Seoul?
Processing errors for homonyms	사람 각각 성 다르다 person each gender/castle/surname different	사람의 성은 각각 다르다. Each person's gender/castle/surname is different.
Missing additional information	아이 나이 지금 5 되다 child age now 5 to-be	아이의 나이가 다섯살이 되었다. The child is five years old.
Ungrammatical WKL sentence	계절 덥다 끝 봄 다가오다 season hot end spring to-come-closer	여름이 끝나면 봄이 온다. When summer ends, spring comes.
Literal translation	물 울다 이용 꽃 만들다 water cry use flower to-make	눈물로 꽃을 만들다. To make flowers out of tears

Table 5: Augmented data error and cause examples (blue text indicates English translation.)

ings: "gender", "castle", and "surname". Korean sign language expresses these differently; however, its gloss could not be identified. In the third row, "Missing additional information," the corresponding GKSL sentence required the gloss "살 [years]" after "5," indicating a unit of age. The example in the fourth row is a case wherein the translation of a GKSL sentence to a WKL sentence is literally correct. However, there are meaning errors in the resultant WKL sentence. The fifth row shows an example that requires paraphrasing instead of providing a literal translation. The expression "making flowers with tear" can also be understood as a literary expression of enduring pain and achieving good results. The literal translations of such expressions may be difficult to understand in the target language.

Among these errors, incorrect gloss usage and homonym errors can be improved with the Korean sign language gloss dictionary update. Missing unit information can be remediated through a post-verification process. Grammar and meaning errors in WKL sentences may also be automatically eliminated via language model validation. Currently, such automatic data augmentation methods have significant room for improvement.

5. Conclusion

In this study, the quantity of sign language data at the gloss level was augmented using a well-known augmentation technique that is stably used with NLP techniques. The results showed that the performance of the gloss-level sign language translation model improved when augmented data were used. Moreover, we demonstrated the plausibility of our augmentation approach based on the expert evaluation of our automatically generated GKSL and WKL sentences quality. In the future, we will update the Korean sign language gloss dictionary using homonyms, synonyms, parts of speech, and head-word arrangements to automatically build quality into the augmented data. We also intend to reduce the time and human costs required for augmentation by applying a simplified method of filtering ungrammatical sentences among the automatically augmented ones using the linguistic knowledge of a large-scale language model.

7. Bibliographical References

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.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Gowda, T. and May, J. (2020). Finding the optimal vocabulary size for neural machine translation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*.

Heo, H., Ko, H., Kim, S., Han, G., Park, J., and Park, K. (2021). Pororo: Platform of neural models for natural language processing. https://github.com/kakaobrain/pororo.

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.

Kim, S., Jang, J. Y., Jung, M., and Shin, S. (2021). A model of cross-lingual knowledge-grounded response generation for open-domain dialogue systems. In *Findings of the Association for Computational Linguistics: EMNLP 2021*.

Ko, S.-K., Kim, C. J., Jung, H., and Cho, C. (2019). Neural sign language translation based on human keypoint estimation. *Applied Sciences*, 9(13):2683.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Miyazaki, T., Morita, Y., and Sano, M. (2020). Machine translation from spoken language to sign language using pre-trained language model as encoder. 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 139–144.

Moryossef, A., Yin, K., Neubig, G., and Goldberg, Y. (2021). Data augmentation for sign language gloss translation. In *Proceedings of the 18th Biennial Machine Translation Summit, 1st International Workshop on Automatic Translation for Signed and Spoken Languages*.

Ormel, E., Crasborn, O., van der Kooij, E., van Dijken, L., Nauta, E. Y., Forster, J., and Stein, D. (2010). Glossing a multi-purpose sign language corpus. In *sign-lang@ LREC 2010*, pages 186–191. European Language Resources Association (ELRA).

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Prabhumoye, S., Tsvetkov, Y., Salakhutdinov, R., and Black, A. W. (2018). Style transfer through back-translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.

Sincan, O. M. and Keles, H. Y. (2020). Autsl: A large scale multi-modal turkish sign language dataset and baseline methods. *IEEE Access*, 8:181340–181355.

Sugiyama, A. and Yoshinaga, N. (2019). Data augmentation using back-translation for context-aware neural machine translation. In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 35–44.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Wei, J. and Zou, K. (2019). EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6382–6388. Association for Computational Linguistics.

Yin, K., Moryossef, A., Hochgesang, J., Goldberg, Y., and Alikhani, M. (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*.