Translate Arabic Text to Arabic Gloss for Sign Language

Doaa Alghamdi[1*], Mansour Alsulaiman[1], Yousef Alohali[2], Mohamed A. Bencherif[1], Mohammed Algabri[3] and Mohamed A. Mekhtiche[1]

[1]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University
doaaalghmdi@gmail.com, msuliman@ksu.edu.sa, mabencherif@ksu.edu.sa ,
mmekhtiche@ksu.edu.sa
Riyadh 11543, Saudi Arabia.
[2]Department of Computer Science, College of Computer and Information Sciences, King Saud University
yousef@ksu.edu.sa
Riyadh 11495, Saudi Arabia
[3]Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University
mohmahgabri@gmail.com
Riyadh, 13713, Saudi Arabia.

Abstract- Arabic Sign Language (ArSL) is a language used by the deaf community across Arab countries, but the lack of familiarity with ArSL among the hearing population often leads to social isolation for deaf individuals. The structural differences between ArSL and spoken Arabic pose significant challenges for machine translation. In this study, we enhance Arabic to ArSL gloss translation by employing data augmentation techniques, expanding the dataset from 600 to over 23,328 samples using sequence-to-sequence transformer models. Our approach achieved a substantial performance improvement, increasing the BLEU score from 11.1% in the baseline model to 52.72% on original test set. The best model achieved a BLEU score of 85.17% on augmented data test, underscoring the effectiveness of data augmentation in enhancing ArSL translation quality.

1.   Introduction

The global deaf and hearing-impaired community, which constitutes over 5% of the world's population, relies heavily on sign languages for communication [1]. Sign languages are rich, visual-spatial languages that employ a combination of hand gestures, facial expressions, and body movements to convey meaning [2]. ArSL, in particular, serves as the primary mode of communication for the deaf community in Arab countries [3].   Despite its importance, ArSL

remains largely unfamiliar to the hearing population, contributing to the social isolation of deaf individuals. Unlike spoken Arabic, ArSL has its own distinct syntax, grammar, and lexicon, making translation between these two languages a complex challenge.

The development of a semantic rule-based machine translation system for converting Arabic text to ArSL gloss, as demonstrated by [4], has laid an important foundation. However, these approaches have been constrained by the availability of training data and the inherent limitations of rule-based methodologies. The work was based on a relatively small parallel corpus of 600 Arabic sentences translated into ArSL gloss. While useful, this dataset is insufficient to capture the full variability of natural language. The rule-based system achieved BLEU score of 35%, highlighting the challenges in preserving the intended meaning and grammatical structure in translations. These limitations restrict the scalability and adaptability of the translation models, resulting in low accurate translations.

The effectiveness of machine translation systems, particularly those designed for various language pairs such as Arabic language and ArSL, is heavily dependent on the availability of large, high-quality datasets. A more extensive dataset would allow for better training and generalization, leading to more accurate translations [5]. Furthermore, advancements in Natural Language Processing (NLP) and machine learning techniques, such as sequence-to-sequence models, have demonstrated significant potential in improving translation accuracy by learning complex language patterns and relationships directly from data [6].

To address these limitations, our research aims to fill this gap by utilizing data augmentation techniques such as Blank Replacement, Synonym Replacement, and Sentence Paraphrasing to expand the original dataset from 600 to over 23,328 sentences. In addition, we evaluate the generated data by using the advanced Arabic sequence-to- sequence machine translation models and apply different data proportion techniques to examining the impact of dataset size on model performance. This approach makes the data more robust basis for training, capturing a wider range of linguistic diversity.

The contribution of this work is twofold: (1) We explore different data augmentation techniques to enhance the dataset size and quality of ArSL translation. (2) We investigate and compare different sequence-to-sequence machine translation models, by testing their performance on both the original test data and augmented test data.

2.   Related Work

Translating Arabic text into ArSL is essential for integrating deaf individuals into their communities. However, developing effective translation systems faces challenges due to the scarcity of parallel corpora and incomplete documentation of ArSL's grammar and structure. ArSL translation research is still in its early stages compared to other sign languages [7] like American Sign Language (ASL) [8] and British Sign Language (BSL) [9]. Many existing systems rely on

rule-based methods, requiring extensive linguistic knowledge to map spoken or written text to corresponding sign language expressions.

In the context of ArSL, several approaches have been explored. [10] focused on translating prayer-related Arabic sentences into ArSL using Sign Writing, limited by a small corpus and lack of coverage for various sentence structures. [11] used a chunk-based example-based machine translation (EBMT) approach, but their reliance on Google Tashkeel and example similarity led to high error rates. [12] developed a rule-based system that achieved high accuracy at the word level but did not adequately address sentence-level grammatical differences. [13] explored syntax transformations but were limited to specific grammatical structures.

Recent advancements have incorporated data-driven techniques, such as statistical ma- chine translation (SMT) and example-based machine translation (EBMT), which show promise but are limited by the availability of large, high-quality datasets. [4] developed a rule-based system for translating Arabic text into ArSL, utilizing a 600-sentence health domain corpus. While their system achieved over 80% accuracy, its limited dataset size restricted broader language applicability and generalization.

Additionally, efforts have been made to improve the availability and annotation of sign language data. The Jumla Sign Language Annotation Tool, described by [14], provides a web-based solution for annotating Qatari Sign Language (QSL) with written Arabic text, supporting the creation of annotated datasets such as the Jumla Qatari Sign Language Corpus. [15] extended this work with the development of the JUMLA-QSL-22 corpus, containing 6,300 records annotated with glosses, translation, signer identity, and location. These tools and datasets are crucial for advancing sign language processing (SLP) and highlight the ongoing efforts to establish more comprehensive linguistic resources for Arabic-related sign languages.

Machine translation (MT) techniques have evolved significantly, with neural machine translation (NMT) models, such as sequence-to-sequence architectures, emerging as state- of-the-art methods. These models, including Recurrent Neural Networks (RNNs) and Transformer-based architectures, have demonstrated a strong ability to handle complex language pairs by learning from vast amounts of data to capture linguistic patterns and context [16]. For sign language translation, these models have been applied to other sign languages [17] and [18], achieving varying degrees of success. The challenge lies in effectively applying these models to ArSL, where data scarcity and linguistic differences pose significant obstacles. However, utilizing augmented datasets in conjunction with NMT models can enhance translation accuracy and help bridge the gap between spoken Arabic and ArSL.

In our previous work [19], we have shown the performance of the AraT5-V2 model for Arabic gloss machine translation which was evaluated using various data augmentation methods, including BR, SP, and SR. Experimental results shows that the BR method demonstrated superior performance, likely due to its larger dataset size of 22,404 samples. In contrast, the SP and SR methods, which utilized produced smaller datasets, exhibited higher validation losses and significantly lower BLEU scores. Further analysis was conducted by combining all three

augmentation methods and compare with other models, including the original AraT5 Base and mT5 models, the AraT5 V2 model outperformed with test BLEU score of 90.93.

In this work, we extend the investigation by employing data augmentation techniques to expand the original dataset to over 23,000 samples and testing models on the original test set, while we tested the models by the augmented test set in our previous study. Moreover, in this study we apply data proportion techniques to study the impact of dataset size. In all experiment we compare the performance of AraT5 base, Arat5 v2 and mT5 models.

3.   Methodology

In this section, we outline the methodology used to enhance the translation from Arabic text to gloss text. Our approach focuses on significantly expanding the dataset size through various data augmentation techniques and implementing advanced machine translation models. By enriching the dataset and leveraging sequence-to-sequence models, we aim to address the limitations of previous rule-based systems and improve translation accuracy and reliability.

Figure 1 shows an example of translation process from Arabic spoken language to sign language including an intermediate gloss representation, which serves as a crucial step in bridging the gap between the syntax of spoken Arabic and the grammar of ArSL. By translating spoken language into gloss text, the translation text to accurate sign language representations will be more easier for machine translation models.
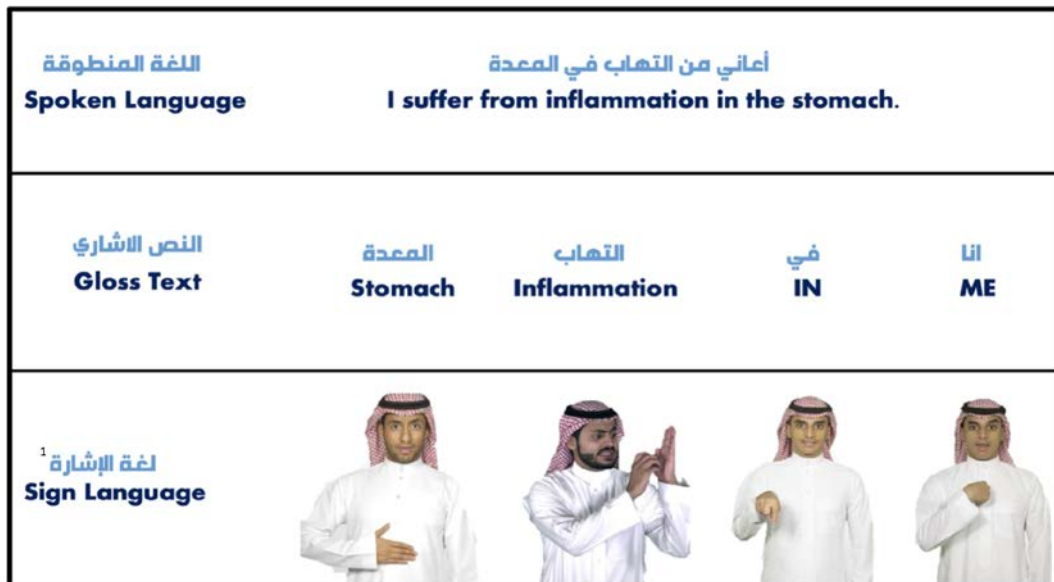

Figure 1. Illustration of the translation process from spoken Arabic to gloss text and then to sign language[1].

---

[1] https://sshi.sa/

3.1 Data Augmentation Techniques

In order to overcome the limitations of the original ArSL dataset, which consisted of only 600 sentences primarily from healthcare settings, we employed data augmentation methods to expand the dataset to 23,328 sentences. The primary augmentation techniques used include BR, SR, and SP. These techniques were selected based on their ability to enhance the variability and robustness of the training data, which is essential for capturing the complex linguistic features of Arabic and ArSL.

A crucial component of our methodology is the development and use of an indexing algorithm, which ensures systematic processing and accurate mapping between Arabic text and its corresponding ArSL gloss. The indexing algorithm assigns indices to each word in both the original Arabic sentences and their gloss translations, maintaining alignment and consistency throughout the data augmentation process. This alignment is essential for preserving the semantic meaning of sentences when applying augmentation techniques, as it ensures that any modifications made to the original text are accurately reflected in the gloss version. The use of the indexing algorithm facilitates seamless integration of new data samples, allowing for scalable and efficient data augmentation, and setting a foundation for creating a high-quality, diverse dataset.

Blank Replacement (BR), is a data augmentation methodology used in NLP to simulate missing or unknown words and enhance the performance of machine learning models. This technique involves masking selected words in a sentence and predicting those words based on surrounding context using the fill mask tool[2], and the AraELECTRA model [20]. By creating masked versions of original sentences and generating new candidate words, BR enables enriching the dataset with up to 21,804 new samples.

Synonyms Replacement (SR), introduces lexical diversity by substituting words with their synonyms from a predefined dictionary, thus increasing vocabulary variety while maintaining the overall meaning of the sentences. This method leverages a custom-built dictionary based on the Saudi Sign Language dictionary and the ArSL dataset, which ensures that synonyms are contextually relevant to ArSL. By exposing the model to different lexical choices that convey similar meanings, SR enhances the model's ability to generalize across various linguistic expressions and improves its adaptability to different word usages, resulting in the generation of 684 new sentences.

Sentence Paraphrasing (SP), is used to provide the training data with some variation and help the model learn various ways to deliver the same information by creating paraphrased versions of sentences through back translation (i.e., translating sentences into English and then back into Arabic). This process generates alternative phrasings that preserve the original meaning but differ in structure. Such variability is crucial for handling the significant structural differences between Arabic and ArSL. By training on paraphrased data, the model becomes more flexible in recognizing and accurately translating a wide range of sentence structures, thereby improving its

---

[2] https://huggingface.co/tasks/fill-mask

capacity to capture meaning and maintain grammatical consistency in translations. SP is resulting in 840 new sentences.

Figure 2 shows the examples of dataset with augmented data that span a wide range of vocabulary and sentence structures, providing the model with the necessary exposure to capture linguistic nuances.
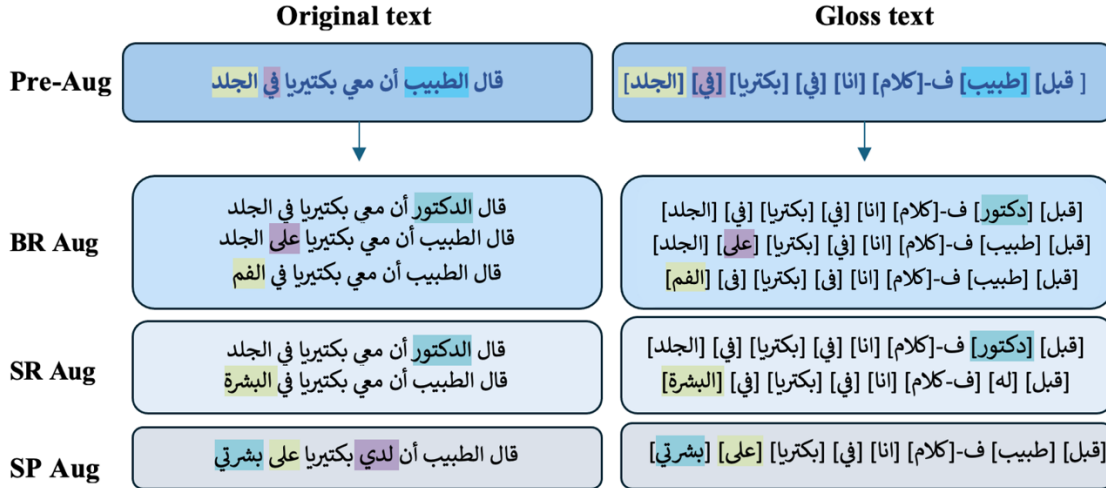

Figure 2. Arabic to Arabic gloss dataset samples

Employing these data augmentation techniques, as detailed in our previous work [19], significantly enhances the dataset's size and diversity. This enriched dataset provides a solid foundation for training effective machine translation models, supporting the development of a robust system capable of accurately translating Arabic text into Arabic Sign Language gloss and improving accessibility and communication for the deaf community.

3.2 Sequence-to-Sequence Machine Translation Model

Sequence-to-sequence transformer models have become a cornerstone in natural language processing, particularly for tasks that involve transforming input sequences into output sequences, such as machine translation. These models are designed to handle input and output sequences of variable lengths, making them appropriate for translating text from one language to another while preserving the meaning of the samples. The T5 (Text-to- Text Transfer Transformer) model is a state-of-the-art Seq2Seq model that unifies various NLP tasks under a single framework by converting them into text-to-text tasks [6]. This architecture is particularly well-suited for translation tasks due to its ability to handle diverse linguistic patterns and context effectively.

In our study, we also utilize AraT5-V2, a variant of the T5 model training specifically for the Arabic language [21]. AraT5-V2 leverages the robust architecture of T5, optimized for handling the intricacies of Arabic syntax and semantics. The model consists of an encoder with multiple

layers that each have a self-attention mechanism and a feed-forward network, followed by a decoder that also includes cross-attention to the encoder's output. This structure allows the model to generate translations that accurately reflect the source language's meaning and align with the target language's grammatical norms. Figure 3 shows the architecture of Arabic gloss machine translation using AraT5 Model.
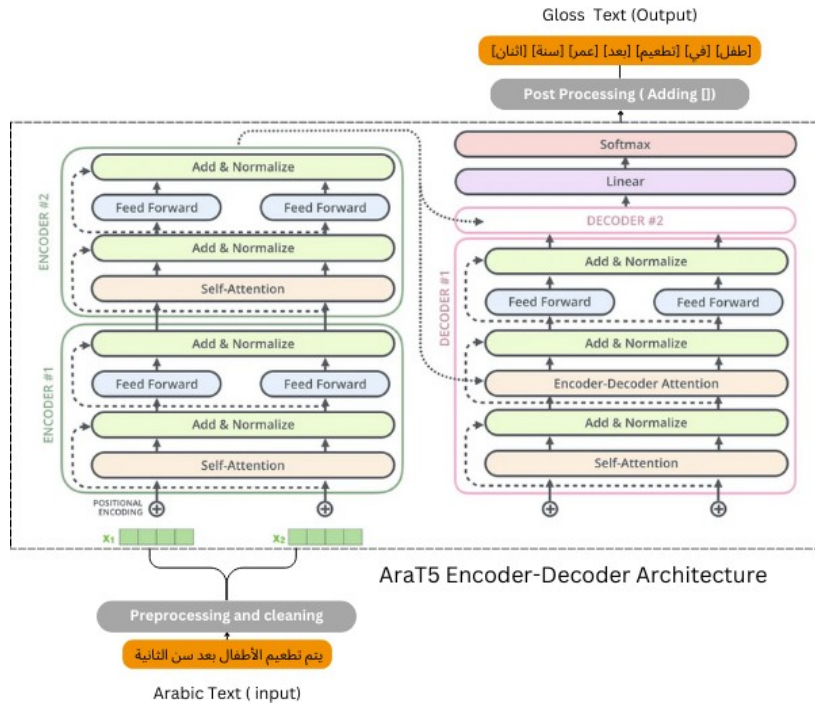

Figure 3. AraT5 Model architecture from gloss machine translation

As shown in Figure 3, the process of training AraT5-V2 for translating Arabic text into Arabic Gloss Language involves a sequence of carefully structured steps to optimize the model's performance. Initially, the pre-trained AraT5-V2 model, which already understands general Arabic language structures, is further training on a parallel dataset of Arabic sentences and their corresponding Arabic gloss translations.

During training, the encoder processes the input, converting it into a series of contextual representations that capture the sentence's meaning. These representations are then passed to the decoder, which generates the gloss output in an autoregressive manner—predicting one token at a time while using previously generated tokens to inform the prediction of the next. The model's parameters are adjusted using backpropagation, where the differences between the predicted gloss sentences and the actual gloss sentences are minimized using optimization algorithms like AdamW. Hyperparameters, including learning rate and batch size, are training to achieve optimal performance, and techniques such as early stopping are employed to prevent overfitting. Once the model demonstrates satisfactory performance on a validation set, it is considered ready for testing.

Training AraT5-V2 model is capable of taking an Arabic sentence as input and generating its

corresponding Arabic gloss translation. The encoder-decoder architecture ensures that the generated output maintains the semantic meaning and follows the syntactic rules of the gloss language, reflecting the knowledge gained during training. The performance of the model is evaluated using metrics such as BLEU scores, which measure the accuracy of the translated sentences against human-annotated references.

4.   Experimental Results

We employed the AraT5 V2[3] model for our experiments, a state-of-the-art neural machine translation model tailored for Arabic text. Additionally, we evaluated two other models; mT5 [22], a multilingual transformer model capable of handling various languages, and AraT5 Base [21] a foundational version of the AraT5 tailored for Arabic but without the enhancements present in the V2 version. The primary evaluation metric used in these experiments is the BLEU score, which is commonly used in machine translation tasks to assess the quality of translations. In addition to BLEU scores, we also measure validation and test losses and compare model predictions against reference test sets. Training was conducted using an Adaptive Learning Rate with the AdamW optimizer, along with a dropout rate of 0.1 to prevent overfitting. We employed a batch size ranging from 8 to 128, adjusted based on the dataset size, and a linear learning rate scheduler. The training ran for 22 epochs, with evaluation and model saving performed every 500 steps to monitor progress and prevent overfitting.

4.1 Performance of Data Augmentation techniques

To evaluate the performance of the trained AraT5-V2 model using different data augmentation methods, we calculated the BLEU scores for each method separately. Each data augmentation method was assessed using its own validation split during training, while the test BLEU scores were calculated using the original ArSL test set 90 samples to determine the overall impact on the translation task. The results of the AraT5-V2 model for each data augmentation method are presented in Table 1, alongside the results from the original dataset before augmentation.

Table 1. Performance Metrics of AraT5-V2 for Different Data Augmentation Methods

Metric	BR	SP	SR	Original ArSL
Val. Loss	0.260	2.173	1.795	3.116
Val. BLEU	**92.778**	25.75	29.30	15.273
Test Loss	2.413	2.396	2.383	2.159
Test BLEU	**52.71**	13.33	12.90	11.069
Dataset Size	22404	1440	1284	600

As shown in Table 1, the BR method consistently demonstrated the best performance, with a validation loss of 0.260 and a validation BLEU score of 92.778. The test BLEU score for BR

---

[3] https://huggingface.co/UBC-NLP/AraT5v2-base-1024

was 52.71, significantly higher than those of the other methods. The large dataset size of 22,404 samples for BR likely contributed to its superior performance, allowing the model to learn more robust translation patterns and effectively generalize to unseen data.

In contrast, SP and SR exhibited poorer performance, with validation losses of 2.173and 1.795, respectively, and evaluation BLEU scores of 25.75 for SP and 29.30for SR. The SR method, which used the smallest dataset size of 1,284 samples, had the lowest test BLEU score of 12.90. This indicates that limited data and vocabulary coverage significantly reduced its effectiveness. However, both SP and SR methods still performed better than the original unaugmented ArSL dataset, which had a test BLEU score of 11.069, demonstrating the value of data augmentation in enhancing translation quality.

Moreover, we combined all three data augmentation methods (BR, SP, and SR), resulting in a dataset of 23,328 samples which is used to train the AraT5 V2 model and compared against the AraT5 Base and mT5 models using the combined data augmented dataset. Note that we used the original test set to evaluate the models.

Table 2 demonstrates the results of the comparison of these models. As shown in Table 2, AraT5 V2 achieved the highest BLEU scores and the lowest validation and test losses, with a validation BLEU score of 86.16 and a test BLEU score of 69.41. The Base model also performed well in terms of validation BLEU score, reaching 35.190, but it had a significantly lower test BLEU score of 33.62. The mT5 model demonstrated moderate performance with a validation BLEU score of 72.380 and a much lower test BLEU score of 15.157.

Table 2. Comparison of Different Machine Translation Models

Model	Val. Loss	Val. BLEU	Test Loss	Test BLEU
AraT5 V2	**0.492**	**86.16**	0.174	**69.41**
Base	1.610	35.190	0.979	33.62
mT5	0.586	72.380	**0.265**	15.157

4.2 Results of Data Augmentation Proportion

To further investigate the impact of data augmentation on the performance of our machine translation model, we conducted a data augmentation proportion experiment. This study examines how varying the proportions of augmented data—specifically 20%, 40%, 80%, and 100%—affects the AraT5 V2 model's translation accuracy. By systematically adjusting the amount of augmented data while maintaining consistent training setups, this experiment aims to identify the optimal balance between data variety and volume for enhancing model performance. The effectiveness of these proportions will be evaluated using BLEU scores, providing insights into how different levels of data augmentation contribute to the robustness and accuracy of the translation models. Table 3 shows the size of data proportion for different data augmentation methods.

Table 3. Data Proportion and Dataset Sizes for Different Data Augmentation Methods

Method	Split	20%	40%	80%	100%
BR	Train. Size	3,908	7,398	14,374	17,863
	Val. Size	526	962	1,834	2,270
	Test Size	526	962	1,834	2,270
	Total Size	4,360	8,722	17,443	21,804
SP	Train. Size	1,985	1,942	1,907	2,173
	Val. Size	18.45	22.45	23.55	25.75
	Test Size	106	123	157	174
	Total Size	168	336	672	840
SR	Train. Size	531	640	859	968
	Val. Size	103	117	144	158
	Test Size	103	117	144	158
	Total Size	137	274	547	684
All	Train. Size	4,995	8,728	16,191	19,923
	Val. Size	735	1,202	2,135	2,602
	Test Size	735	1,202	2,135	2,602
	Total Size	4,665	9,332	18,662	23,328

Table 3 shows the dataset sizes for different data augmentation methods including BR, SP, SR, and the combination of them at proportions of 20%, 40%, 80%, and 100%. The Total Size represents the number of samples corresponding to each proportion from the complete dataset. Validation and test sizes were created by splitting the total dataset and including an additional 90 samples from the original ArSL dataset into the test set.

This approach ensures that each augmentation method's impact can be evaluated under consistent conditions, providing insights into the effectiveness of varying data sizes and combinations on translation accuracy. Table 4 shows the results of test BLEU for each different data augmentation methods and proportions.

Table 4. Test BLEU Scores for Different Data Augmentation Methods and Proportions

Method	20%	40%	80%	100%
BR	42.49	72.54	90.97	**91.57**
SP	12.63	16.52	20.27	21.77
SR	12.01	15.89	15.71	18.49
All	34.73	65.29	82.46	85.17

As shown in Table 4, the BR method outperformed other methods across all proportions, achieving the highest test BLEU score of 91.57 at 100% proportion. This indicates that the BR method provides the most robust training data for the model, likely due to its ability to capture diverse linguistic patterns effectively. Even at lower proportions, the BR method demonstrated

substantial improvements, with a BLEU score of 42.49 at 20% augmentation, highlighting its strong impact even with less data.

In contrast, the SP and SR methods showed relatively lower test BLEU scores across all proportions, with the SP method reaching a maximum BLEU score of 21.77% at 100% augmentation and the SR method peaking at 18.4%. These results suggest that while SP and SR contribute to model performance, their impact is less pronounced compared to BR. The lower effectiveness of SP may be due to the inherent limitations of back- translation, which sometimes produces paraphrases that are too similar to the original or introduces noise that does not enhance the training process. For SR, the relatively poor performance could be attributed to the limited vocabulary coverage and context relevance of the synonym dictionary used, which might have resulted in substitutions that did not significantly vary the training data or, in some cases, distorted the sentence meaning.

The combined data augmentation methods (All) showed a balanced performance, achieving a test BLEU score of 85.17% at 100% augmentation, indicating that a mixture of augmentation techniques can yield high performance but may not surpass the effectiveness of BR alone. Overall, these findings emphasize the importance of selecting appropriate data augmentation methods and optimizing their implementation to enhance machine translation accuracy, as well as the need for more sophisticated approaches to improve SP and SR.

5.   Conclusion

In this study, we enhanced the translation of Arabic text into Arabic gloss text using advanced data augmentation techniques and the AraT5 V2 model. Our results showed that the Blank Replacement method provided the highest translation accuracy, while the combined augmentation method also improved performance but did not surpass BR. However, in this study we used small dataset which was developed in the heath field which does not cover a common of Arabic words. Future work will focus on complete the second phase to translate the gloss to text to Arabic sign language motions, which leads to develop more robust and accurate sign language translation systems to better serve the Arabic-speaking deaf community.

References

[1]   Kushalnagar, R. (2019). Deafness and hearing loss. Web accessibility: A foundation for research, pages 35–47.

[2]   Luqman, H., Mahmoud, S. A., et al. (2017). Transform-based arabic sign language recognition. Procedia Computer Science, 117:2–9.

[3]   Al-Fityani, K. and Padden, C. (2010). Sign language geography in the arab world. Sign languages: A Cambridge survey, 20.

[4] Luqman, H. and Mahmoud, S. A. (2019). Automatic translation of arabic text-to-arabic sign language. Universal Access in the Information Society, 18(4):939–951.

[5] Koehn, P. and Knowles, R. (2017). Six challenges for neural machine translation. arXiv preprint arXiv:1706.03872.

[6] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140):1–67.

[7] Sidig, A. a. I., Luqman, H., and Mahmoud, S. A. (2018). Arabic sign language recog- nition using optical flow-based features and hmm. In Recent Trends in Information and Communication Technology: Proceedings of the 2nd International Conference of Reliable Information and Communication Technology (IRICT 2017), pages 297–305. Springer.

[8] Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., and Palmer, M. (2000). A machine translation system from english to american sign language. In Envisioning Machine Translation in the Information Future: 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000 Cuernavaca, Mexico, October 10–14, 2000 Proceedings 4, pages 54–67. Springer.

[9] Marshall, I. and Sáfár, É. (2003). A prototype text to british sign language (bsl) trans- lation system. In The companion volume to the proceedings of 41st annual meeting of the association for computational linguistics, pages 113–116.

[10] Almasoud, A. M. and Al-Khalifa, H. S. (2012). Semsignwriting: A proposed semantic system for arabic text-to-signwriting translation.

[11] Almohimeed, A., Wald, M., and Damper, R. I. (2011). Arabic text to arabic sign language translation system for the deaf and hearing-impaired community. In Proceedings of the second workshop on speech and language processing for assistive technologies, pages 101–109.

[12] El, A., El, M., and El Atawy, S. (2014). Intelligent arabic text to arabic sign language translation for easy deaf communication. International Journal of Computer Applica- tions, 92(8).

[13] Al-Rikabi, S. and Hafner, V. (2011). A humanoid robot as a translator from text to sign language. In 5th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC 2011), pages 375–379.

[14] Othman, A., Dhouib, A., Chalghoumi, H., Elghoul, O., and Al-Mutawaa, A. (2024). The acceptance of culturally adapted signing avatars among deaf and hard-of-hearing individuals. IEEE Access.

[15] Othman, A., El Ghoul, O., Aziz, M., Chemnad, K., Sedrati, S., and Dhouib, A. (2023). Jumla-qsl-22: Creation and annotation of a qatari sign language corpus for sign lan- guage processing. In Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments, pages 686–692.

[16]   Angelova, G., Avramidis, E., and Möller, S. (2022). Using neural machine translation methods for sign language translation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 273–284.

[17]   Jang, J. Y., Park, H.-M., Shin, S., Shin, S., Yoon, B., and Gweon, G. (2022). Automatic gloss-level data augmentation for sign language translation. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 6808–6813.

[18]   Kayahan, D. and Güngör, T. (2019). A hybrid translation system from turkish spo- ken language to turkish sign language. In 2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), pages 1–6. IEEE.

[19]   Alghamdi, D., Alsulaiman, M., Alohali, Y., Bencherif, M. A., and Algabri, M. (2024). Arabic gloss machine translation through data augmentation. In Proceedings of the Third SmartTech Conference (Manuscript submitted for publication). King Saud Uni- versity.

[20]   Antoun, W., Baly, F., and Hajj, H. (2020). Araelectra: Pre-training text discriminators for arabic language understanding. arXiv preprint arXiv:2012.15516.

[21]   Nagoudi, E. M. B., Elmadany, A., and Abdul-Mageed, M. (2021). Arat5: Text-to-text transformers for arabic language generation. arXiv preprint arXiv:2109.12068.

[22]   Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., and Raffel, C. (2020). mt5: A massively multilingual pre-trained text-to-text transformer. arXiv preprint arXiv:2010.11934.