# Improving Multilingual Sign Language Translation with Automatically Clustered Language Family Information

**Ruiquan Zhang**[1,2,3,4,*], **Cong Hu**[1,2,3,4,*], **Pei Yu**[1,2,3,4], **Yidong Chen**[1,2,3,4,†]

[1]Department of Artificial Intelligence, School of Informatics, Xiamen University, 361005, P.R. China

[2]Key Laboratory of Digital Protection and Intelligent Processing of Intangible Cultural Heritage of Fujian and Taiwan (Xiamen University), Ministry of Culture and Tourism, 361005, P.R. China

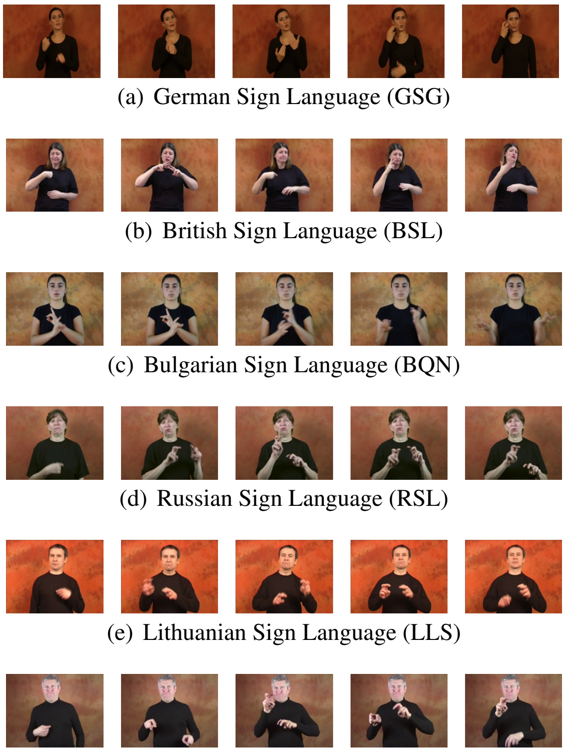
[3]National Language Resources Monitoring and Research Center for Education and Teaching Media, Xiamen University, 361005, P.R. China

[4]Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, 361005, P.R. China

{rqzhang, hucong11235, yupei}@stu.xmu.edu.cn, ydchen@xmu.edu.cn

## Abstract

Sign Language Translation (SLT) bridges the communication gap between deaf and hearing individuals by converting sign language videos into spoken language texts. While most SLT research has focused on bilingual translation models, the recent surge in interest has led to the exploration of Multilingual Sign Language Translation (MSLT). However, MSLT presents unique challenges due to the diversity of sign languages across nations. This diversity can lead to cross-linguistic conflicts and hinder translation accuracy. To use the similarity of actions and semantics between sign languages to alleviate conflict, we propose a novel approach that leverages sign language families to improve MSLT performance. Sign languages were clustered into families automatically based on their Language distribution in the MSLT network. We compare the results of our proposed family clustering method with the analysis conducted by sign language linguists and then train dedicated translation models for each family in the many-to-one translation scenario. Our experiments on the SP-10 dataset demonstrate that our approach can achieve a balance between translation accuracy and computational cost by regulating the number of language families. The source codes and models are available at FamilyCluST.



(a) German Sign Language (GSG)

(b) British Sign Language (BSL)

(c) Bulgarian Sign Language (BQN)

(d) Russian Sign Language (RSL)

(e) Lithuanian Sign Language (LLS)

(f) Ukrainian Sign Language (UKL)

Figure 1: The degree of difference in sign language expressions varies among different sign language pairs. When observing the expressions of ”I’m looking for a job”, we may find that the expressions in (a), (b) and (c) are quite unique, while those in (d), (e) and (f) are more similar to one another.

## 1 Introduction

Sign Language Translation (SLT) is a sophisticated cross-modal task that converts sign language videos into spoken language texts, making it easier for hearing-impaired people to communicate with non-sign language speakers (Camgoz et al., 2018). In recent years, research on SLT has attracted more and more attention from researchers (Camgoz et al., 2020; Fu et al., 2023; Hu et al., 2024; Zhao et al., 2024).

In addition to the research on conventional Bilingual Sign Language Translation (BSLT), such as American Sign Language to English or Chinese Sign Language to Chinese, inspired by the research on Multilingual Neural Machine Translation (MNMT), several recent studies have delved into the Multilingual Sign Language Translation (MSLT) domain. Yin et al. (2022) developed and released the first large-scale parallel multilingual sign language understanding dataset, Spreadthesign-

[*]These authors contributed equally to this work.

[†]Corresponding author

*Proceedings of the 31st International Conference on Computational Linguistics*, pages 3579–3588
January 19–24, 2025. ©2025 Association for Computational Linguistics

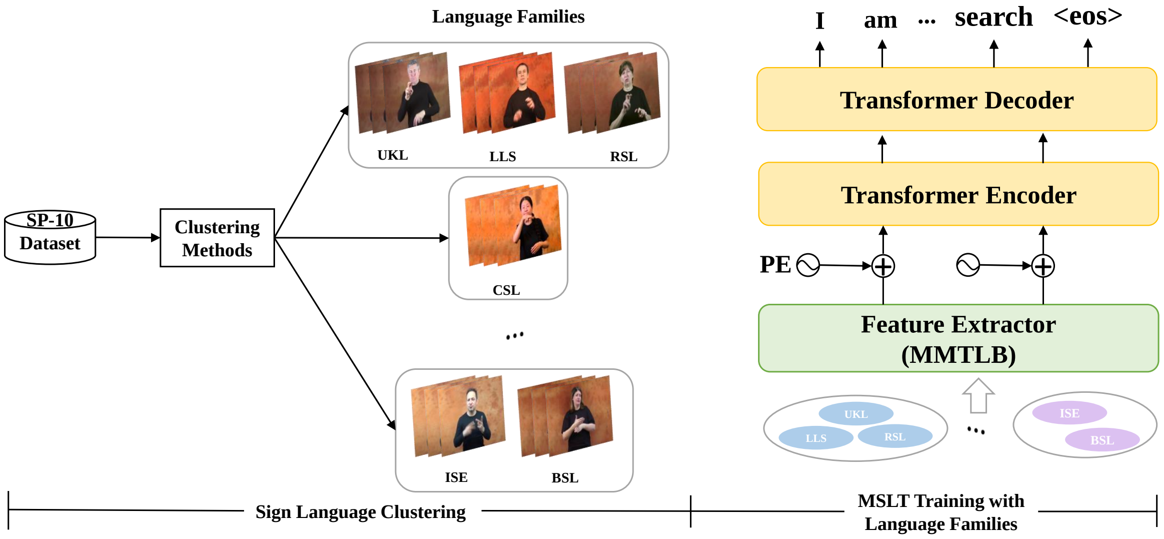


Figure 2: Many-to-one sign language translation model. The architecture consists of a feature extractor (MMLTB (Chen et al., 2022a)) and several transformer modules (Vaswani et al., 2017).

Ten (SP-10), and introduced the first framework for multilingual sign language translation, MLSLT. Gueuwou et al. (2023) presented the JWSign dataset, which includes 98 sign languages. These studies primarily aim to integrate numerous language pairs into a unified translation model.

However, unlike BNMT, which has been extensively studied, the research on MSLT is still in its early stages, with several issues yet to be thoroughly investigated. One key challenge is that when we attempt to train all available sign language data in single model, the diversity of sign languages, which often results in cross-linguistic conflicts (Wei and Chen, 2023; Ye et al., 2024), may harm the performance of the final model.

By observing the MSLT data, such as SP-10, we find that the degree of difference in sign language expressions varies among different sign language pairs and may have something to do with the language families. To specifically demonstrate this problem, Figure 1 illustrates that the phrase "I am looking for a job" across various national sign languages exhibits consistency in some countries' sign vocabularies and even entire sentences. Different nations utilize unique gestures to convey the same meaning, notably in GSG (1(a)), BSL (1(b)), and BQN (1(c)).

Moreover, the role of similarities between sign languages in MSLT remains largely uncharted. Linguistic studies further confirm that historical political and educational influences have led to similarities among national sign languages (Wittmann, 1991; Yu et al., 2018; Abner et al., 2020). Figure 1 reveals that RSL (1(d)), UKL (1(f)), and LLS (1(e)) display highly similar expressions. In contrast, although BQN originates from an Eastern European country, its expressions are significantly different from Russian Sign Language and others.

We posit that employing a single MSLT model for language pairs with significant differences could adversely affect the training process. Conversely, similar language pairs could mutually enhance each other during model training. This observation motivates us to investigate and leverage language families in multilingual translation contexts, aiming to identify languages with fewer conflicts and higher similarities in sign language to enhance translation accuracy and efficiency.

Thus, we initially categorize the sign languages of different countries into language families. Within these language families, the signs from various countries are approximately similar. Subsequently, we train a dedicated translation model for each sign language family in the many-to-one sign language translation scenario. Our experiments on the SP-10 dataset show that by grouping sign languages into families, we reduce memory and storage demands and increase computational efficiency. Training a single model for each family leverages similarities within those groups, enhancing translation accuracy. This approach demonstrates our method's balance between performance and cost, offering a scalable solution for multilingual sign language translation.The contributions of this paper are as follows:

- We are the first to employ language feature distributions within the MSLT network to delve into the underpinnings of sign language families. We compare and analyze the proposed family clustering method with the findings of sign language linguists.

- To underscore the significance of sign language families to MSLT model performance, we have conducted experiments on the SP-10 dataset in the many-to-one context. By modulating the number of families, we achieve a synergy between translation accuracy and resource usage.

- Through ablation studies and a series of comparative trials, we demonstrate the paramount importance of accurate family clustering in enhancing the translation accuracy of multilingual sign language translation.

## 2 MSLT Using Sign Language Families

We constrain the experimental scenario to many-to-one translation to eliminate potential interference arising from the complex task of mapping inputs to multiple target languages, focusing solely on combinations of input sign languages.

As shown in Figure 2, our system is divided into two steps. According to different clustering methods, we first divide the sign language in the data set into different sign language families, and then we train a single translation model for each sign language family. We will introduce various sign language clustering methods and analyze the clustering results in Sec. 3.

Then we employ a basic Transformer encoder-decoder to construct a multilingual many-to-one sign language translation benchmark model. Given a sign language sequence $V = (v_1, \ldots, v_L)$ of length $L$ in a specific language, we input it into the sign language feature extractor, add positional information, and then input it into the Transformer for translation. The final output is a spoken language sequence $Y = (y_1, \ldots, y_T)$ with $T$ words. The objective function is defined as the Negative Log Likelihood (NLL) loss:

$$\mathcal{L}_{\text{NLL}} = -\sum\nolimits_{i=1}^{T} t_i \ln \left( P_{\text{MSLT}}^{i} \mid t_i \right) \qquad (1)$$

where $t_i$ and $T$represent the length of the $i$ output word and the target spoken translation, respectively.

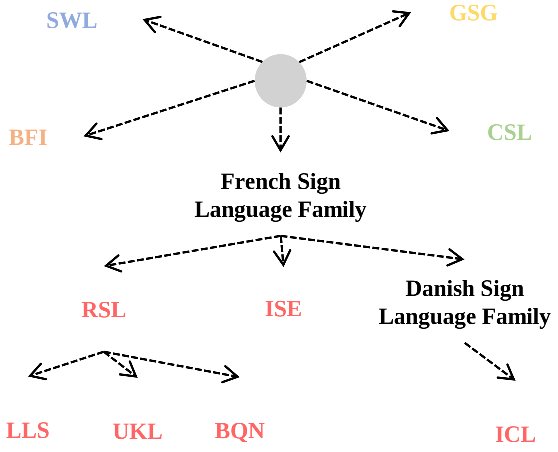


Figure 3: Linguistics-based clustering results of the 10 languages in SP-10 dataset according to language family. There are 5 different language families in this dataset, and different colors represent different language families, among which the French Sign Language Family contains the five sign languages in the dataset. Full name for the abbreviation can be found in Table 1.

## 3 Sign Language Clustering Methods

In this section, we first introduce various methods for sign language clustering. We begin with a Linguistics-based clustering approach and then explore the use of neural networks for automated clustering. Additionally, we present three distribution-based clustering methods. Finally, we compare and analyze the clustering results for the proposed family of methods.

| language | Abbreviation (ISO 639-3) |
|---|---|
| Chinese sign language | CSL |
| Ukrainian sign language | UKL |
| Russian sign language | RSL |
| Bulgarian sign language | BQN |
| Icelandic sign language | ICL |
| German sign language | GSG |
| Italian sign language | ISE |
| Swedish sign language | SWL |
| Lithuanian Sign language | LLS |
| British Sign Language | BFI |

Table 1: The abbreviations of different sign languages in the SP-10 dataset used in this paper. The ISO 639-3 international language code name standard is used for sign language abbreviations.

### 3.1 Linguistics-based Clustering

We delve into the exploration of the evolution of the European Sign Language families (Reagan, 2019;

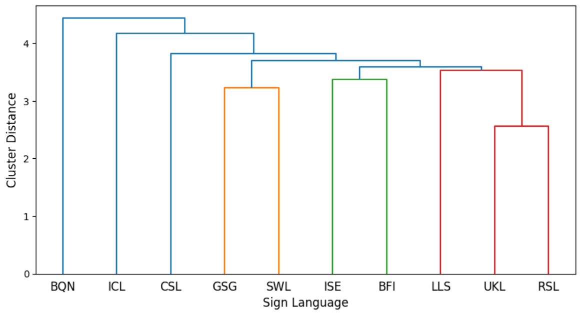


Figure 4: The hierarchical clustering based on encoder output distribution. The Y-axis represents the distance between two languages or families. Languages in the same color are divided into the same family, where blue color agglomerates different clusters together. If a language is marked as blue, then it forms a cluster itself. Family #1: BQN; Family #2: ICL; Family #3: CSL; Family #4: GSG, SWL; Family #5: ISE, BFI; Family #6: LLS, UKL, RSL.

Anderson, 1979; Wittmann, 1991), along with other studies on language independence (Power, 2022; Reagan, 2021; Abner et al., 2020). Based on these studies, we mapped out a sign language family tree for the 10 sign languages included in the SP-10 dataset, as shown in Figure 3. In the figure, the leaf nodes denote the sign languages present in the dataset, with different colors distinguishing different families.

It is noteworthy that the French Sign Language family exhibits a broad coverage in the tree. Particularly, the RSL branch is remarkable as it includes Eastern European sign languages such as LLS, UKL, and BQN. Based on the evolution of the French Sign Language, we grouped these four Eastern European sign languages with ISE and ICL under the same family (Abner et al., 2020). Additionally, BFI, CSL, GSG, and SWL are categorized as independent families due to their linguistic independence (Power, 2022).

These linguistics-based clusterings often derive from lexical analysis and do not capture the holistic sentence-level features of a language. To address the limitation, we propose several clustering methods based on the language distribution in the MSLT model.

### 3.2 Clustering Based on Encoder Output Distribution

In this section, we use a language-agnostic feature extractor to represent each sign language and apply hierarchical clustering to group similar feature vectors, forming related sign language families.

First, we removed the language labels from the many-to-one sign language translation model, making it unable to identify which specific sign language it is processing. However, when sign languages share similar expressions—such as hand movements or facial expressions—the model generates nearly identical feature outputs for them. On a broader scale, each sign language has a distinct linguistic distribution, and those with similar expressions tend to have closely aligned feature distributions.

Building on this idea, we designed a sign language family clustering algorithm based on the distribution of encoder outputs, aiming for the model to discover similarities while embedding semantic information of sign language features.

After multiple rounds of training with the MMLTB encoder (Chen et al., 2022a), a robust SLT baseline model trained via transfer learning, we obtain the feature vector $S = (s_1, \ldots, s_L)$, where $L$ represents the sequence length and $D$ is the network dimension. To acquire a unique representation for each sign language for feature distribution comparison, we average the features across all videos and frames from $N$ sign language videos, condensing them into a single, compact feature vector per language. This is calculated as follows:

$$Lang = \frac{1}{N} \sum\nolimits_{i=1}^{N} \left( \frac{1}{L} \sum\nolimits_{l=1}^{L} s_i \right) \tag{2}$$

We derived language feature vectors for each sign language. We then applied agglomerative hierarchical clustering on these vectors, using average linkage to measure distances between clusters. Figure 4 clearly illustrates the clustering process of sign languages, showing the relationships among their distributions.

Based on the clustering distance threshold and linguistic analyses,the ten languages in SP-10 are clustered into six language families. Considering the clustering results by linguists in the previous section, our observations from the figure are as follows:

- The red line in Figure 4 differentiates LLS, UKL, and RSL, which all belong to the Eastern European Sign Language family, consistent with linguistic consensus. Additionally, CSL and ICL are significantly distinct from mainland European sign languages and are thus separately classified.

- BQN shows minimal association with most other sign languages, contrasting with its previous clustering into the French Sign Language family.

- The language distribution reflects geographical influences. For instance, GSG and SWL are geographically closer to each other.

### 3.3 Clustering Based on Original Input Feature Distribution

In this analysis, we further investigated the original input distribution, specifically focusing on the sign language video features obtained using the MMLTB encoder (Chen et al., 2022a). Employing the pooling strategy outlined in Equation 2, we condensed these feature sequences into a unified vector representation. These original features do not contain semantic information. The t-SNE method was applied to reduce the dimensionality of 500 samples (50 per sign language) in the training set, as shown in Figure 5. The results revealed that sign language features from the same country tend to cluster together, with distinct language distributions. For instance, ICL is far from other sign languages, while RSL and UKL are relatively close.

Building on these observations, we conducted hierarchical clustering on the original sign language input features in the training set. As shown in Figure 6, the sign languages of the four Eastern European countries were grouped together, while GSG and SWL were clustered into one family. However, this clustering method based on original input features also grouped CSL with ICL, and subsequent experiments showed that joint training of Chinese Sign Language with European sign languages led to performance degradation.

### 3.4 Clustering Based on Language Label Distribution

Additionally, we referred to the method of MLSLT (Yin et al., 2022), which incorporates language labels into the model input. Each sign language has a corresponding language parameterized label vector. According to previous research (Tan et al., 2019; Ma et al., 2023), these language label vectors can be considered as comprehensive features of the languages and can be clustered.

As shown in Figure 7, the differences between the label vectors of different sign languages are minimal. Compared to clustering methods based on original input features and encoder outputs, the Euclidean distances between these vectors vary by an order of magnitude. This suggests that the model finds it challenging to capture distinct language features. Using a defined threshold, we categorize the data into four language families, with ICL and UKL grouped into one family independently.

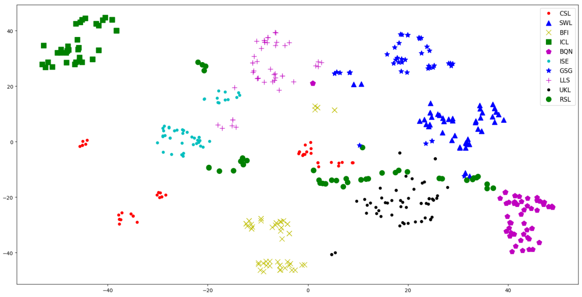


Figure 5: Linguistic phenomena manifested by the dimensionality reduction of the original input sign language features. We apply the t-SNE method to reduce the dimensionality of 500 samples in the training set (50 samples for each sign language).

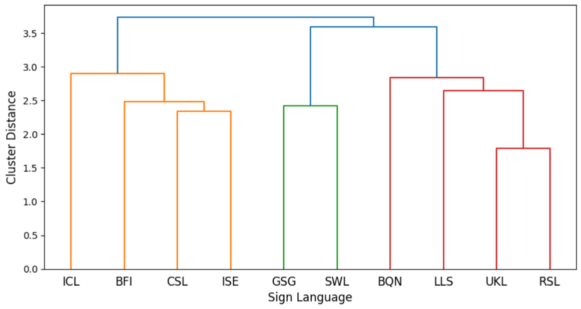


Figure 6: The hierarchical clustering based on original input feature distribution. Family #1: ICL, BFI, CSL, ISE; Family #2: GSG, SWL; Family #3: BQN, LLS, UKL, RSL.

## 4 Experiments

### 4.1 Experimental Setting

**Datasets and Metric.** The SP-10 dataset[1] has collected video materials of 10 different sign languages from the SpreadTheSign platform (Hilzensauer and Krammer, 2015), along with corresponding spoken language translation texts. It includes 830 training samples, 142 development samples, and 214 test samples, each containing ten videos with their respective spoken translations.

A pre-trained multilingual tokenization model, trained on Wikipedia, is employed to tokenize the

[1] https://github.com/MLSLT/SP-10/tree/main

| Metric | Method | CSL | UKL | RSL | BQN | ICL | GSG | ISE | SWL | LLS | BFI | MEAN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BLEU-4 | Individual(10) | 4.90 | 5.91 | 5.45 | **3.62** | 6.68 | 6.03 | 6.41 | 6.24 | 6.21 | 7.96 | 5.94 |
| | Universal(1) | 4.51 | 5.89 | 4.92 | 1.93 | 5.54 | 5.44 | 6.81 | 5.80 | 5.44 | 5.90 | 5.22 |
| | MLSLT(1) (Yin et al., 2022) | **5.16** | 5.42 | 4.95 | 3.28 | 6.76 | 5.18 | 7.05 | 6.33 | 6.08 | 7.03 | 5.72 |
| | Encoder Distribution(6) | 4.90 | **7.54** | **6.91** | **3.62** | 6.68 | **8.61** | **8.35** | **8.56** | **7.40** | **8.02** | **7.06** |
| ROUGE | Individual(10) | 33.25 | 34.23 | 33.65 | **31.50** | 34.28 | 34.08 | 34.67 | 34.49 | 34.41 | 36.56 | 34.11 |
| | Universal(1) | 32.90 | 34.20 | 33.43 | 19.01 | 33.72 | 33.61 | 34.96 | 34.13 | 33.62 | 33.95 | 32.35 |
| | MLSLT(1) (Yin et al., 2022) | **34.59** | 34.04 | 31.62 | 27.98 | 35.29 | 33.50 | 37.96 | 36.02 | 34.48 | **37.25** | 34.27 |
| | Encoder Distribution(6) | 33.25 | **37.54** | **35.20** | **31.50** | 34.28 | **37.61** | **38.35** | **37.56** | 35.76 | 37.02 | **35.81** |
| BLEURT | Individual(10) | **35.55** | 34.93 | 34.80 | **29.76** | **36.35** | **34.58** | 34.87 | 34.70 | 32.65 | 31.36 | 33.96 |
| | Universal(1) | 34.85 | 34.92 | 34.47 | 25.90 | 34.16 | 34.14 | 34.98 | 31.31 | 34.09 | 34.25 | 33.31 |
| | Encoder Distribution(6) | **35.55** | **35.98** | **36.40** | **29.76** | **36.35** | 34.21 | **37.23** | **36.30** | 32.78 | **36.21** | **35.08** |

Table 2: Comparison of the translation results from multiple sign languages to English on the SP-10 validation set.

| Metric | Method | CSL | UKL | RSL | BQN | ICL | GSG | ISE | SWL | LLS | BFI | MEAN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BLEU-4 | Individual(10) | **6.19** | 6.21 | **7.03** | **2.38** | **6.73** | 3.26 | 5.81 | 6.18 | 5.70 | 5.74 | 5.52 |
| | Universal(1) | 4.97 | 3.22 | 4.60 | 1.29 | 4.15 | 3.80 | 3.98 | 5.47 | 2.79 | 4.33 | 3.86 |
| | MLSLT(1) (Yin et al., 2022) | 5.19 | 4.18 | 3.66 | 2.85 | 3.93 | 4.97 | 6.70 | 3.70 | 5.72 | 5.73 | 4.66 |
| | Encoder Distribution(6) | **6.19** | **7.66** | 6.01 | **2.38** | **6.73** | **5.99** | **9.68** | **7.34** | 6.32 | **7.83** | **6.61** |
| ROUGE | Individual(10) | **34.63** | 35.21 | **35.03** | **28.78** | **34.73** | 31.26 | 34.81 | 35.18 | 34.70 | 33.74 | 33.81 |
| | Universal(1) | 30.97 | 29.22 | 30.60 | 15.29 | 30.15 | 30.80 | 30.98 | 34.47 | 28.79 | 30.33 | 29.16 |
| | MLSLT(1) (Yin et al., 2022) | 33.33 | 34.07 | 31.54 | 25.75 | 33.25 | 32.13 | 35.37 | 33.09 | 33.11 | 35.34 | 32.70 |
| | Encoder Distribution(6) | **34.63** | **35.66** | 34.01 | **28.78** | **34.73** | **32.99** | **38.68** | **36.94** | 34.32 | **37.82** | **34.86** |
| BLEURT | Individual(10) | **35.62** | 34.52 | **35.59** | 28.28 | **36.06** | 32.91 | 35.39 | 33.75 | **35.65** | 33.71 | 34.15 |
| | Universal(1) | 31.26 | 29.03 | 31.72 | 25.60 | 31.46 | **33.43** | 32.13 | 32.25 | 30.46 | 31.68 | 30.90 |
| | Encoder Distribution(6) | **35.62** | **36.34** | 33.00 | 28.28 | **36.06** | 30.89 | **35.70** | **35.31** | 35.42 | **36.12** | **34.27** |

Table 3: Comparison of the translation results from multiple sign languages to English on the SP-10 test dataset.

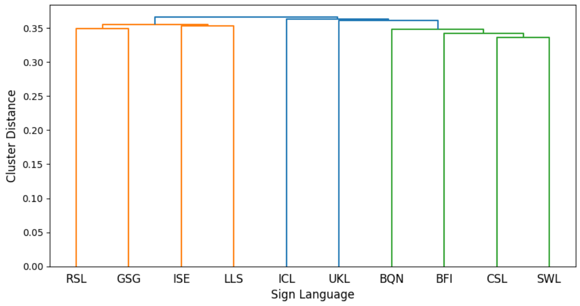


Figure 7: The hierarchical clustering based on language label distribution. Family #1: RSL, GSG, ISE, LLS; Family #2: ICL, UKL; Family #3: BQN, BFI, CSL, SWL.

text data in the SP-10 dataset [2]. The resulting vocabulary size is 10912. Following previous studies (Yin et al., 2022), we use BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004) to measure the translation quality of SLT. In order to fully assess the semantic expressiveness of the model, this section also uses the BLEURT (Sellam et al., 2020) scoring system for additional validation.

**Implementation Details.** We adopt the Transformer model architecture (Vaswani et al., 2017), widely used in multilingual machine translation, as the base model. The hidden size is set to 256, and the feed-forward neural network dimension is set to 1024. Both the encoder and decoder are configured with 3 layers. The training strategy in this section also utilizes the Adam optimizer (Kingma and Ba, 2014) and a learning rate schedule based on the Noam construction. The batch size is set to 8, and the learning rate is set to $1e^{-4}$. The dropout rate and label smoothing (Müller et al., 2019) are set to 0.3 and 0.2, respectively. The evaluation settings include a beam search width of 6 and a penalty factor of 1. The model is trained on four NVIDIA RTX 3090 GPUs with 24 GB of memory each.

### 4.2 Baselines

First, the experimental results table reports the performance of the bilingual (Individual) and multilingual (Universal) baseline models trained using our proposed framework on the current dataset. Consistent with the findings of MLSLT (Yin et al., 2022), the bilingual baseline model trained solely on sign language data outperforms the multilingual baseline model trained on ten languages when trans-

[2] https://github.com/bheinzerling/bpemb

lating into English, as evidenced by the average translation performance on the validation and test sets. Particularly on the test set, the difference in BLEU-4 scores widens to 1.66, mainly due to potential conflicts between sign languages.

### 4.3 Main Results

For comparative analysis with existing research methods, we standardized the target language of many-to-one translation to English. As shown in Tables 2 and 3, we detailed the translation performance under different experimental setups, with the corresponding number of models noted in parentheses.

The experimental results demonstrate the positive impact of our best approach on translation results: the encoder output distribution-based clustering method (Encoder Distribution). On the test set, the Encoder Distribution method produces the best translation results, with a BLEU-4 score of 6.61, a ROUGE score of 34.86, and a semantic evaluation benchmark BLEURT score of 34.27. This suggests that joint training of sign languages within language families effectively reduces conflicts and enables mutual data enhancement, thereby improving overall translation performance.

### 4.4 Influence of Different Language Distributions

Additionally, we compare the translation results obtained from random family clustering methods to verify the importance of accurate family clustering in many-to-one sign language translation. As shown in Table 5, we used the same number of categories as in the encoder output distribution method and ensured that each category contained no more than three languages. We performed five random clusterings and trained translation models based on these random clusterings.

We proceeded to evaluate the translation performance of the aforementioned five clustering methods, as shown in Table 4. The models derived from random family clustering exhibited the poorest translation performance, even underperforming a single model trained in all sign languages. This outcome emphasizes the importance of selecting an appropriate family clustering method to improve both translation efficiency and performance. In general, the encoder output distribution-based method demonstrated the highest translation performance, highlighting the superior sentence-level semantic alignment effect of the Transformer encoder.

### 4.5 Zero-shot Experiments

Furthermore, we investigate the capability of a many-to-one model to discover associations between similar sign language expressions across different sign languages. Specifically, we control for the absence of a particular sign language-English translation pair in the training set during our many-to-one experiments.

Based on the Universal model, we successively excluded the RSL, UKL, and LLS to English translation pairs, and tested the model on the corresponding datasets in the validation set. The results, as presented in the Zero-shot(9x1) row in Table 6, indicate that the model fails to translate unseen sign languages, with a significant drop in ROUGE scores.

However, when we train models using sign languages with strong similarities, the translation of similar sign language expressions improves. For instance, in the zero-shot experiment for UKL-en, we train a model with RSL-en and LLS-en translation pairs. We applied this approach similarly to the other two languages. As shown in the Zero-shot(2x1) results in Table 6, the translation performance surpasses that of the model trained with nine sign languages.

Additionally, we conduct case studies on the models from the Zero-shot(9x1) and Zero-shot(2x1) experiments, as illustrated in Figure 8. In the validation set, the spoken text "Are you looking for anything?" has similar sign language expressions across Eastern European countries. The Zero-shot(9x1) model, trained with nine language pairs excluding UKL-en, failed to understand the Ukrainian Sign Language expression, resulting in the zero-shot translation "How do you sign?", which is entirely unrelated to "Are you looking for anything?". Conversely, the Zero-shot(2x1) model correctly translated the UKL sentence to "What are you looking for?". The translation examples from Zero-shot(2x1) demonstrate that joint training with data from the same language family can enhance data augmentation.

## 5 Related Works

### 5.1 End-to-end Sign Language Translation

End-to-end Sign Language Translation (SLT) aims to directly convert sign language video sequences into spoken text using neural network models, which in recent years have included multi-task learning (Camgoz et al., 2020; Zhou et al., 2021b;

| Metric | Method | CSL | UKL | RSL | BQN | ICL | GSG | ISE | SWL | LLS | BFI | MEAN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BLEU-4 | Random(6) | 3.45 | 4.37 | 5.76 | 2.37 | 4.78 | 5.91 | 5.34 | 5.75 | 5.13 | 6.58 | 4.94 |
| | Linguistics(5) | **4.90** | 4.71 | 5.21 | 1.36 | **7.47** | 6.03 | 6.30 | 6.24 | 6.68 | 7.96 | 5.69 |
| | Input Distribution(3) | 4.10 | 6.10 | 5.71 | 0.46 | 3.86 | **8.61** | 7.15 | **8.56** | 6.51 | 5.10 | 5.62 |
| | Label Distribution(4) | 3.77 | 5.91 | 5.16 | 2.49 | 6.68 | 7.02 | 6.93 | 5.63 | 6.94 | 5.85 | 5.64 |
| | Encoder Distribution(6) | **4.90** | **7.54** | **6.91** | **3.62** | 6.68 | **8.61** | **8.35** | **8.56** | **7.40** | **8.02** | **7.06** |
| ROUGE | Random(6) | 31.95 | 32.90 | 34.30 | 29.50 | 32.70 | 34.05 | 33.60 | 33.90 | 33.30 | 34.40 | 33.06 |
| | Linguistics(5) | **33.25** | 33.15 | 34.12 | 15.85 | **36.78** | 34.08 | 34.56 | 34.49 | **35.82** | 36.56 | 32.87 |
| | Input Distribution(3) | 32.45 | 35.20 | 33.90 | 10.95 | 32.40 | **37.61** | 35.60 | **37.56** | 34.85 | 32.90 | 32.34 |
| | Label Distribution(4) | 32.00 | 34.23 | 34.10 | 29.60 | 34.28 | 35.65 | 34.75 | 33.60 | 34.71 | 35.00 | 33.79 |
| | Encoder Distribution(6) | **33.25** | **37.54** | **35.20** | **31.50** | 34.28 | **37.61** | **38.35** | **37.56** | 35.76 | **37.02** | **35.81** |

Table 4: Comparison of the validity of clustering methods. The translation results on the SP-10 validation set using different families are reported in the table.

| **NO** | **Sign Language Families** |
|---|---|
| 1 | (CSL, UKL) (RSL, BQN) (ICL, GSG) (ISE, SWL) (LLS) (BFI) |
| 2 | (SWL, BFI) (LLS, UKL) (RSL, BQN) (CSL, GSG) (ICL) (ISE) |
| 3 | (LLS, BFI, GSG) (SWL, BQN) (RSL, ICL) (UKL) (ISE) (CSL) |
| 4 | (ISE, RSL) (BFI, CSL) (SWL) (LLS, UKL) (BQN, ICL) (GSL) |
| 5 | (RSL, GSG) (ISE, UKL) (SWL, CSL) (BFI, LLS) (ICL) (BQN) |

Table 5: Random Language families in five experiments.

| Method | RSL-en | | UKL-en | | LLS-en | |
|---|---|---|---|---|---|---|
| | B4 | R | B4 | R | B4 | R |
| Universal | 4.92 | 33.43 | 5.89 | 34.20 | 5.44 | 33.62 |
| Zero-shot(9x1) | 1.40 | 13.49 | 0.59 | 10.62 | 0.90 | 15.91 |
| Zero-shot(2x1) | **2.53** | **25.13** | **1.21** | **19.22** | **2.84** | **26.72** |

Table 6: Zero-shot experiments. We report the translation results on the SP-10 validation set. B represents BLEU-4 and R represents ROUGE.

Zhang et al., 2023; Zhao et al., 2024), data augmentation (Zhou et al., 2021a; Zhang et al., 2023) and the use of pre-trained models (Chen et al., 2022a,b), and many other methods are applied to bilingual SLT. In addition, multilingual sign language translation has gradually attracted the attention of researchers. Yin et al. (2022) successfully constructed the first massively parallel multilingual sign language comprehension dataset, Spreadthesign-Ten (SP-10), and proposed for the first time a multilingual sign language translation framework MLSLT. Gueuwou et al. (2023) introduced the JWSign dataset, encompassing a diverse array of sign languages for multilingual sign language translation and proposed a linguistics clustering method based on lexical analysis. However, this method fails to capture the holistic sentence-level features of the languages.

(a) Russian Sign Language (RSL)

(b) Lithuanian Sign Language (LLS)

(c) Ukrainian Sign Language (UKL)

Figure 8: Expressions of the sentence ”What are you looking for?” in the sign languages for several Eastern European countries.

## 5.2 Sign Language Families

Since the 1970s, Many researchers have attempted to address the issue of historical-comparative linguistics in sign languages and to discern different language families. These research efforts have employed many theories and methods from traditional historical linguistics, including historical homology (Wittmann, 1991) and methods such as lexicostatistics (Woodward, 2011; Abner et al., 2020). It is worth mentioning that Wittmann published a pioneering study on the linguistic families of sign languages in 1991 (Wittmann, 1991), in which he initially identified five sign language families by conducting a comparative study of some 80 sign languages. In recent years, Abner et al. (Abner et al., 2020)also based their analysis on historical homology on the basis of their predecessors. Some researchers have utilized similarity between language families to improve the quality of machine translation (Tan et al., 2019; Ma et al., 2023; Arivazhagan et al., 2019). We innovatively apply Sign Language Families' information to multilingual sign language translation.

## 6 Conclusion

In this study, we investigate the role of sign language families for enhancing multilingual sign language translation. Specifically, we focus on the SP-10 sign language translation dataset and propose the Linguistics-based Clustering approach as well as several Language Distribution-based Clustering methods.

Our approach offers significant advantages over training individual models for each language pair, as it reduces computational costs. Furthermore, compared to training a single model for all language pairs, our method demonstrates superior translation accuracy.

## Limitations

We used the only open-source SP-10 dataset for our experiments, and the number of videos available for training for each sign language is 830, which is an extremely scarce amount of data. Moreover, the language family clustering study in this paper only deals with European sign languages and does not cover the global diversity of language families.

Future research needs to use more comprehensive sign language datasets, such as the soon-to-be open-sourced JWSign dataset [3], which contains 98 different sign languages, to validate the efficacy and wide applicability of the feature distribution-based sign language family clustering method.

## Ethical Considerations

This research aims to improve the accessibility of sign language translation, which directly benefits the hearing-impaired community. However, we fully recognize the critical importance of safeguarding privacy and obtaining informed consent when utilizing sign language data, particularly in datasets that include individual characteristics such as facial expressions, hand gestures, and upper body movements. All datasets used in this study, such as SP-10, were collected with appropriate consent from participants, following ethical guidelines to protect their identity and privacy. Additionally, we are committed to ensuring that our approach does not unintentionally reinforce biases or exclude any sign language groups, fostering inclusivity and fairness in the development of multilingual sign language translation systems.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 62476232 and 62076211 and in part by the University-Industry Cooperation Programs of Fujian Province of China under Grant No. 2023H6001.

## References

Natasha Abner, Carlo Geraci, Shi Yu, Jessica Lettieri, Justine Mertz, and Anah Salgat. 2020. Getting the upper hand on sign language families: Historical analysis and annotation methods. *FEAST. Formal and Experimental Advances in Sign Language Theory*, 3:17–29.

L. Anderson. 1979. A comparison of some american, british, australian and swedish signs: Evidence on historical changes in signs and some family relationships of sign languages. In *First International Symposium on Sign Language*.

Naveen Arivazhagan, Ankur Bapna, Orhan Firat, Dmitry Lepikhin, Melvin Johnson, Maxim Krikun, Mia Xu Chen, Yuan Cao, George Foster, Colin Cherry, et al. 2019. Massively multilingual neural machine translation in the wild: Findings and challenges. *arXiv preprint arXiv:1907.05019*.

Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020. Sign language transformers: Joint end-to-end sign language recognition and translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Yutong Chen, Fangyun Wei, Xiao Sun, Zhirong Wu, and Stephen Lin. 2022a. A simple multi-modality transfer learning baseline for sign language translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Yutong Chen, Ronglai Zuo, Fangyun Wei, Yu Wu, Shujie Liu, and Brian Mak. 2022b. Two-stream network for sign language recognition and translation. *Advances in Neural Information Processing Systems*, 35:17043–17056.

Biao Fu, Peigen Ye, Liang Zhang, Pei Yu, Cong Hu, Xiaodong Shi, and Yidong Chen. 2023. A token-level contrastive framework for sign language translation.

[3] https://github.com/ShesterG/JWSign-Machine-Translation

In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

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

Marlene Hilzensauer and Klaudia Krammer. 2015. A multilingual dictionary for sign languages:"spreadthesign". In *ICERI2015 Proceedings*, pages 7826–7834. IATED.

Cong Hu, Biao Fu, Pei Yu, Liang Zhang, Xiaodong Shi, and Yidong Chen. 2024. An explicit multi-modal fusion method for sign language translation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3860–3864. IEEE.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint*.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*.

Xinyu Ma, Xuebo Liu, and Min Zhang. 2023. Clustering pseudo language family in multilingual translation models with fisher information matrix. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13794–13804, Singapore. Association for Computational Linguistics.

Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. 2019. When does label smoothing help? *Advances in neural information processing systems*, 32.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 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.

Justin M Power. 2022. Historical linguistics of sign languages: Progress and problems. *Frontiers in Psychology*, 13:818753.

T. Reagan. 2019. *Linguistic Legitimacy and Social Justice*. Springer International Publishing.

Timothy Reagan. 2021. Historical linguistics and the case for sign language families. *Sign Language Studies*, 21(4):427–454.

Thibault Sellam, Dipanjan Das, and Ankur Parikh. 2020. BLEURT: Learning robust metrics for text generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7881–7892, Online. Association for Computational Linguistics.

Xu Tan, Jiale Chen, Di He, Yingce Xia, Tao Qin, and Tie-Yan Liu. 2019. Multilingual neural machine translation with language clustering. 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 963–973.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Fangyun Wei and Yutong Chen. 2023. Improving continuous sign language recognition with cross-lingual signs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23612–23621.

Henri Wittmann. 1991. Classification linguistique des langues signÉes non vocalement. *Revue québécoise de linguistique théorique et appliquée*, 10:215–288.

James Woodward. 2011. Some observations on research methodology in lexicostatistical studies of sign languages. *Deaf around the World: The Impact of Language*.

Jinhui Ye, Xing Wang, Wenxiang Jiao, Junwei Liang, and Hui Xiong. 2024. Improving gloss-free sign language translation by reducing representation density. *Preprint*, arXiv:2405.14312.

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

Shi Yu, Carlo Geraci, and Natasha Abner. 2018. Sign languages and the online world online dictionaries & lexicostatistics. In *LREC Proceedings (Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Biao Zhang, Mathias Müller, and Rico Sennrich. 2023. Sltunet: A simple unified model for sign language translation. *arXiv preprint arXiv:2305.01778*.

Rui Zhao, Liang Zhang, Biao Fu, Cong Hu, Jinsong Su, and Yidong Chen. 2024. Conditional variational autoencoder for sign language translation with cross-modal alignment. In *Proceedings of the 38th Annual AAAI Conference on Artificial Intelligence*, pages 19643–19651.

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

Hao Zhou, Wengang Zhou, Yun Zhou, and Houqiang Li. 2021b. Spatial-temporal multi-cue network for sign language recognition and translation. *IEEE TMM*.