Journal of Theoretical and Applied Information Technology
15th February 2025. Vol.103. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TRANSFORMING MACHINE TRANSLATION FOR ISOLATING LANGUAGES WITH MULTI-SOURCE NEURAL MODEL

NGUYEN NGOC LAN[1], TRINH BAO NGOC[2*], LE PHUONG THAO[3], NGUYEN THANH CONG[4], LE MANH TOAN[5], , TRAN DINH DIEN[6], BUI PHAN TUE ANH[7], VUONG THI VAN[8], NGUYEN NHAT TRANG[9], NGUYEN TIEN THANH[10]

[1]Department of Scientific Research, Hanoi University, Hanoi 100000, Viet Nam

[2,3,5,9,10] Faculty of Information Technology, Hanoi University, Hanoi 100000, Viet Nam

[4,7] Faculty of Chinese, Hanoi University, Hanoi 100000, Viet Nam

[6] Faculty of Information Technology, Fisheries and Technical, Economics College, Bac Ninh 16000, Viet Nam

[8] Faculty of Chinese, University of Industry, Hanoi 100000, Viet Nam

E-mail: [1]lannn@hanu.edu.vn, [2]ngoctb@hanu.edu.vn, [3]lephuongthao6872@gmail.com, [4]congnt@hanu.edu.vn, [5]toanlemanh2003@gmail.com, [6] tddien@cdts.edu.vn, [7]anhbpt@haui.edu.vn, [8]van.vuong @haui.edu.vn, [9]nguyentrangrmg@gmail.com, [10]ntienthanhn@gmail.com

ABSTRACT

Recent advancements in Artificial intelligence (AI) and Deep learning have facilitated the rapid development of machine translation technologies, among them, Neural machine translation (NMT) models have demonstrated impressive performance, especially in handling multiple language pairs. However, due to their complexity and lack of appropriate data, contemporary NMT models still have a lot of challenges when applied to isolated languages, despite their great accomplishments. This paper proposes a multi-source neural model that employs two different encoders to process both the source word sequence and the linguistic feature sequences of isolating languages. Unlike traditional NMT models, this approach improves the encoders' input embeddings by incorporating a second encoder that integrates the linguistic elements, including part-of-speech (POS) tags and lemma. To enhance the source sentence's context representation, this article combines the encoders' conditional data with the outputs of the decoders using a serial combination technique. In this way, different metrics such as METEOR and BLEU are examined to assess the suggested model's precision of translation. Experimental results indicate that our methodology works efficiently for isolating language translation, as evidenced by the improvement of +3.9 BLEU and +3.2 METEOR scores on translation tasks conventional NMT models perform. This highlights a significant advancement in integrating linguistic features to enhance translation accuracy for isolating languages.

Keywords: *Artificial Intelligence, Neural Machine Translation, Linguistic Features, Isolated Language, BLEU, METEOR*

1. INTRODUCTION

Artificial intelligence (AI) is increasingly transforming fields like healthcare, finance, transportation, and natural language processing (NLP) [1]. One notable AI application in NLP is neural machine translation (NMT), which leverages deep learning models to facilitate language translation. NMT creates smoother translations than statistical machine translations (SMT) by combining all aspects of the translation process into a single model. Moreover, NMT employs an architecture of encoders and decoders where the target sentence is produced by the decoder after the encoder has processed input embeddings. Additionally, NMT models have gained power with advancements like gated recurrent units [2], transformers, and long short-term memory [3], which enhance translation accuracy and handle complex contexts more effectively [4]. By minimizing semantic discrepancies, this technology fosters improved communication across different cultural backgrounds [5].

NMT models have demonstrated notable efficacy in a variety of language pairings, including both commonly used languages like French and under-resourced ones like Tagalog and Khmer [6], [7]. However, integrating linguistic features into NMT systems poses challenges [8], especially for isolating languages like Vietnamese and Chinese. First, the diversity and complexity of these languages complicate the identification and encoding of relevant linguistic attributes. Each language possesses unique structural and expressive characteristics, necessitating a flexible methodological approach to processing. Additionally, obtaining vast and diverse datasets to represent linguistic traits is extremely challenging and requires a combination of advanced technology and linguistic expertise [9]. Finally, optimizing hyperparameters to enhance linguistic feature representation without compromising overall performance is complex task [10].

Table 1: Some challenges for NMT [9] and examples.

Challenge	Example
Data Training Size	A Vietnamese - English model may misinterpret "đá" (stone or to kick) due to a lack of quality training data that covers its different meanings and contexts.
Long Sentence	"The little boy kicked the ball into the garden" could be incorrectly translated into Chinese as "小男孩把球踢进了商店," which means "The little boy kicked the ball into the store" instead of "the garden."
Word Alignment	"She sells seashells by the seashore" could be mistranslated as "She sells the seashore"
Beam Search	The model might choose "She is happy" instead of "She is very happy" if the beam size is too small.

The effectiveness of machine translation systems is largely influenced by the parallel corpus utilized during training, especially regarding its quantity and quality [11]. Nevertheless, creating a parallel, high-quality corpus is difficult and costly, especially for a particular domain-parallel corpus. Additionally, the commonly used recurrent neural network (RNN) model struggled to fully convey the significance of longer documents, leading to the low-quality translation [12]. Due to limitations in data augmentation and base RNN models for large documents, few researchers have focused on integrating linguistic features into NMT models. This integration can significantly enhance translation accuracy, particularly for languages with flexible grammatical rules and patterns [13]. Zhang applied a CAEncoder that enhances NMT models by learning both historical and future contexts, improving upon traditional bidirectional encoders [12]. Studies on translation tasks from Chinese to English revealed that the suggested model performed better than the standard RNN system. Sennrich et al also demonstrated that the NMT model could effectively integrate linguistic features into the attentional encoder-decoder architecture [13]. Their approach incorporated morphological traits, POS tags, lemmas, and dependency labels in experiments on English - German and English - Romanian translations. The improvements in these tasks highlight the benefits of leveraging linguistic properties.

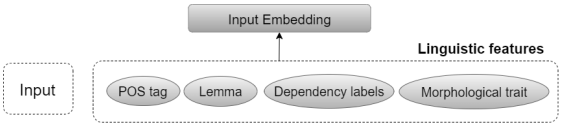

Figure 1: Knowledge-Based Encoder's Input Embedding Layer [13]

The above studies, however, do not take into account the issue of isolating language translation without inflectional morphology and the heavy reliance on context cues. Instead, their primary focus is on high-resource machine translation issues that need sufficient semantic analysis tools and vast amounts of parallel data. This article provided a multi-source neural model to extend NMT model's capacity to represent the source-side sentence's linguistic information. In this literature, our contributions include:

- For the purpose of capturing underlying context, our multi-source NMT utilizes the inherited linguistic traits of isolating languages into a knowledge-based encoder.

- For the linguistic representation, the knowledge-based encoder's input embedding layer is altered to simultaneously encode each word's linguistic attributes.

- Conditional information of the source sequence is combined hierarchically from the encoders, enabling NMT to acquire improved context representation on the source-side sequences. Initially, a compressed vector is produced by fusing the source-side linguistic information with the target sequence representation. This vector then integrates with the source sentence's representation to produce

a context vector. Lastly, each target-side word is predicted orderly using the context vector.

2. RELATED WORKS

Recent developments in machine translation emphasize multi-source methods to boost efficiency, especially for isolated and low-resource languages. In this context, Cao et al. explored language-specific latent spaces for fine-tuning multilingual NMT models [14], while Brazier and Rouas integrated emotional context into large models [15]. In 2024, Zeng used generative adversarial networks to augment data [16], and Honda developed context-aware NMT for business dialogue [17]. Furthermore, additional research includes Nzeyimana, who applied advanced modeling techniques to NMT for low-resource languages [18], and Her who assessed the performance of NMT models for these languages [19], focusing on Bavarian as a case study. Besides, Sennrich et al proposed using sub-word units to manage rare words [20], aiding models in handling previously unseen terms in isolating languages. In addition, Li applied multi-source techniques to enrich semantic and contextual understanding [21], and Zoph et al demonstrated that multi-source methods and transfer learning can greatly enhance the caliber of translations for languages with limited resources [22].

The original multi-source NMT system for crosslingual translation was developed by Zoph and Knight from the foundation of single-source NMT [23]. By assigning each source language to a synthetic four-layer encoder, it encodes its own secret si and cell state cj to an intermediate unit called combiner blocks. Under the hood, this block combines states from encoders into single output s and c states without the modified dimension, then passes them to only a four-layer decoder for target language prediction.

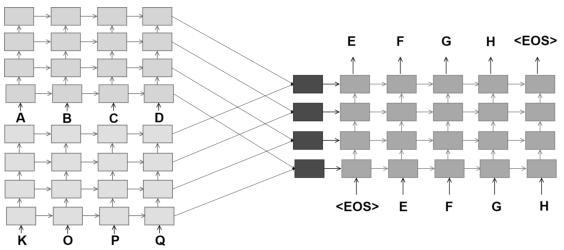

Figure 2: Model of Multi-Source Encoder-Decoder For MT [23]

In Figure 2, two source sentences (A B C D and K O P Q) in different languages. Each language is processed by its own encoder, which outputs final secret states and cell states. These states are then passed to a collection of combiners (illustrated in blue) that merge them into a unified representation. This novel architecture has been consecutively improved to facilitate low resource language and corpus scarcity by data augmentation [24], [25] while several researchers raised the translation accuracy by injecting linguistic representations as main contributions [12], [26], [27].

Supervised learning in NLP relies on large labeled datasets, but limited labeled data requires leveraging additional resources. Pitler et al reviewed linguistic quality metrics in text summarization, achieving 90% accuracy in system comparison and 70% in summary ranking [26]. Hashimoto et al. (2017) introduced a multi-task model [29], such as POS tagging, NER, and syntactic analysis, optimizing weights and achieving strong results. Gururangan et al. (2018) found large NLI datasets contain clues for identifying labels, with a simple model achieving 67% accuracy on SNLI and 53% on MultiNLI [30].

3. MATERIAL AND METHOD

3.1 Standard NMT Model

This article used the Transformer architecture-based NMT model that was first put out by Vaswani et al [31]. This model followed a typical encoder–decoder structure operating on single-source translation. We utilized the basic Transformer model as the foundation to enhance machine translation performance. For comprehension, Figure 3 summarizes several main points of this.

Figure 3: Base Architecture of Transformer Model With Single Source: Encoder-Decoder Framework [32]

The inputs to both the encoder and decoder use the same embedded logic. The inputs to both the encoder and decoder use the same embedded logic. Let $x = (x_1, x_{2,..}, x_m)$ represent the source sequence and $y = (y_1, y_{2,..}, y_n)$ denote the target sequence. As input embeddings, each source word x_i is represented by a corresponding vector while output embedding is responsible for vector conversion from each target word y_i orderly. Generally, a sequence is mapped into an embedding matrix $e = (e_1, e_{2,..}, e_l)$ by the input as well as the output embedding layers. Each embedding e_i calculated as follows:

$$e_i = a_i \cdot E_a \quad (1)$$

with:

- $a_i \in R^{Ka}$: the one-hot vector representing the i^{th} word.

Journal of Theoretical and Applied Information Technology
15th February 2025. Vol.103. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- Ea $\in R^{K_a \cdot d_a}$: matrix of word embedding, d_a the embedding size and K_a is the vocabulary size.

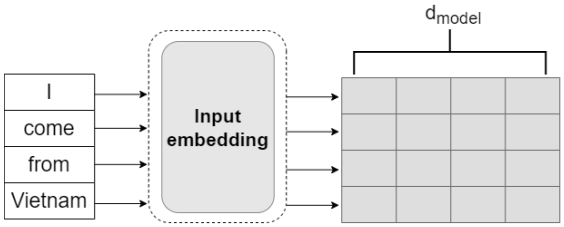

Figure 4: Embedding Process For The Sequence 'I come from Vietnam'

Following the embedding process, the Transformer model applies positional encoding (PE) to capture each word's position in the input sequence. A PE matrix with the same dimension as the sequence matrix is created by encoding these positions. PE uses sine functions for even elements and cosine functions for odd ones [33], allowing the model to assign positions and determine distances between elements (Figure 5).

$$PE(pos,2i) = \sin\left(\frac{pos}{10{,}000^{\frac{2i}{d_{model}}}}\right), \quad (2)$$

$$PE(pos,2i+1) = \cos\left(\frac{pos}{10{,}000^{\frac{2i}{d_{model}}}}\right) \quad (3)$$

with:

- i refers to the dimension within the positional encoding vector.
- pos is the word's position in the sequence (e.g., the 1^{st}, 2^{nd} word, etc.).
- d_{model} is the dimension index for input embedding.

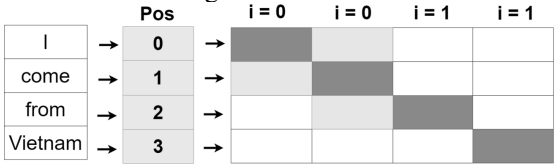

Figure 5: PE Matrix For The Sequence 'I come from Vietnam'

Once applying positional encoding to establish word positions, the positional input embedding X' is formed by combining the input embedding with the positional encoding matrix (Figure 6). The Encoder then uses X' in self-attention to learn contextual relationships among words, evaluating how each word relates to others regardless of distance. Each word is represented by three vectors: Query (Q), Key (K), and Value (V), as illustrated in Figure 7.

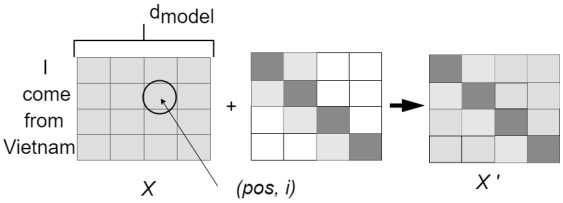

Figure 6: Positional Input Embedding X' For The Sequence 'I come from Vietnam'

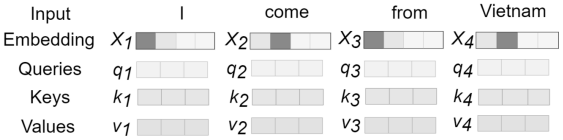

Figure 7: Queries, Keys and Values For The Sequence 'I come from Vietnam'

The attention score is calculated by comparing a word's Query with the Keys of other words and applying softmax normalization. This converts raw scores into a probability distribution that sums to one, enabling the model to weigh the Value vectors effectively for the final output [31], [34]. The softmax function and core attention function are defined as:.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{i=1}^{n} e^{z_i}},$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

with:

- z_i represents the attention scores.
- d_k is the key vectors' dimension.

In the context of Multi-Head Self-Attention, each encoder layer identifies relationships among words in a sentence (Figure 8) by mapping the Q, K, and V sets into an attention matrix [31]. Multi-head attention enhances single attention by using multiple heads for Q, K, and V. Each head computes attention values by multiplying Q, K, and V with their respective weight matrices W_i^Q,W_i^K, and W_i^V from all heads are concatenated using the Concat function, then multiplied by the embedding matrix WE to produce the final attention matrix.

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V),$$

$$\text{MultiHead}(Q,K,V) = \text{Concat}(\text{head}_1, .., \text{head}_h)W_E \quad (5)$$

with:

- W_i^Q, W_i^K, W_i^V being the weight matrices for each attention head.
- Q being the query set, K the key set, and V the value set.

Journal of Theoretical and Applied Information Technology
15th February 2025. Vol.103. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 8: Multi-Head Self-Attention Approach For The Sentence 'I come from Vietnam'

After the multi-head attention value has been determined, the output is run through a fully connected feed-forward network consisting of two linear layers with a ReLU activation function. This adds non-linearity to the model by computing position-wise transformations on the attention layer's output X and filtering out negative values. As illustrated in Figure 9

$$FFN(X) = \max(0, XW_1 + b_1)W_2 + b_2 \quad (6)$$

with:

- b_1 ,b_2 are bias terms.
- W_1 ,W_2 weight matrices.

Figure 9: Feed-Forward Network Based On Position In Transformer Model

The decoder has a structure that closely resembles that of the encoder, as it generates a contextual representation from the inputs and processes this representation through a feed-forward network [31]. Ultimately, the target word sequence is predicted using a softmax function and a linear layer, which determines the likelihood of every word in the lexicon and selects the most likely term to be the output. To improve training stability and efficiency, layer normalization and residual connections (Figure 10) are also applied after every sub-layer.

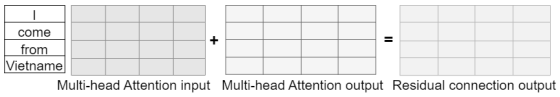

Figure 10: Process Of Residual Connection For Sentence 'I come from Vietnam'

3.2 Multi-Source Neural Model

This article developed the Multi-source model from the standard NMT model to enhance natural language processing capabilities by integrating information from multiple sources. While the standard NMT model uses a word-based encoder to encode word-level source information, the Multi-Source model uses two distinct encoders—a knowledge-based encoder as well as a word-based encoder [35]. This makes it possible to encode source-side linguistic information and source word features efficiently. Figure 11 shows the suggested multi-source neural model's framework. Model sub-layers have an output dimension of of d_x=512, with d_k and d_v defined as d_k=d_v=d_x/h. This paradigm is described in depth in the parts that follow. This paradigm is described in depth in the parts that follow.

Figure 11: The Suggested Multi-Source Neural Model's Architecture [35]

3.2.1 Word-based encoder

The word-based encoder is essential for encoding the source words' features, operating similarly to the encoder in traditional Neural Machine Translation (NMT) models. It processes each word in the source sequence to produce a detailed representation vector H_1 that captures essential contextual information and semantic features [35]. In addition to this representation, the encoder outputs the sets of vectors Q , K , and V, which facilitate the attention mechanisms. By leveraging these outputs, the NMT model effectively encodes and interprets the nuances of the source language, enabling more accurate and contextually appropriate translations.

3.2.2 Knowledge-based encoder

Unlike the encoder of the regular NMT model, the knowledge-based encoder's input embedding layer is enhanced to incorporate linguistic features such as lemma and POS tags for each source word. The model's capacity to convey meaning and context is enhanced by this integration, as shown in Figure 12, and is particularly useful for languages with complex structures.

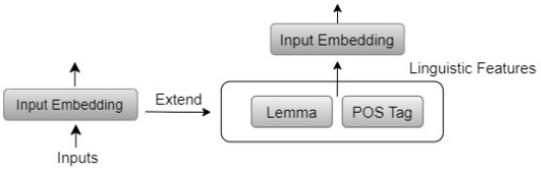

Figure 12: Input Embedding Layer With The Knowledge-Based Encoder's Linguistic Features

Based on the following calculation, linguistic annotation sequences $k_1 = (k_{11}, .., k_{1m})$ and $k_F = (k_{F1}, .., k_{Fm})$ are transformed into an embedding matrix $e^* = (e_1^*, .., e_m^*)$ via the embedding layer [35].

$$e_1^* = \sum_{t=1}^{F} k_{ti} \cdot E_t \quad (7)$$

with:

- ∪ being the operator for vector concatenation.
- $k_{ti} \in R^{1 \times K_t}$ being the one-hot vector.
- K_t being the vocabulary size of the t-th feature.
- $E_t \in R^{K_t \times d_t}$ is the the embedding matrix of the t-th feature.
- d_t being the embedding size of the t-th feature such that $\sum_{t=1}^{F} d_t = d_x$.

Together with the Q* ,K*, and V* vectors utilized in the attention mechanism, the knowledge-based encoder also produces the H_2 representation vector for linguistic feature.

3.2.3 Serial combination approach

To effectively merge the encoders' outputs, we utilize a serial combination method. This approach

concatenates the representation vectors H_1, H_2 into a single vector. The tanh activation function transforms the output to a range of -1 to 1, ensuring non-linearity and enhancing learning, followed by linear and non-linear transformations applied to the resulting vector:

$$H = [H_1 ; H_2], h = \tanh(W_\theta H) \quad (8)$$

with:

- W_θ the trainable weight matrix that determines the importance of the inputs.
- H is the input vector containing information from both encoders.

In order to merge conditional data generated by encoders with the decoder's output, a serial combination approach is used [34]. Furthermore, a second multi-head attention layer is added to the decoder to carry out the attention function [35].

Initially, knowledge-based encoder vector sets K*, V*, and Q from the previous sub-layer of the decoder are mapped by the multi-head attention to produce an attention matrix:

$$Att_F(Q,K^*,V^*) = Concat(head_1,.., head_h)W_F \quad (9)$$

Next, the multi-head attention layer that follows converts the word-based encoder's sets of vectors K and V as well as the attention matrix into a context representation:

$$Att_C(Att_F,K,V) = Concat(head_1,.., head_h)W_C \quad (10)$$

Subsequently, the context representation is subjected to a fully linked feed-forward network in the manner described below:

$$FFN(Att_C) = \max(0,Att_C W_1 H, +b_1 H)W_2 H + b_2 H \quad (11)$$

Finally, the result of the feed-forward network is subjected to a linear layer and a softmax layer in order to produce the desired word sequence.

3.3 Neural Model Comparisons

In this section, the evaluation compared single-source neural models to the suggested multi-source neural model.

- Single-source neural model: An improved NMT model that provides a better representation of source words by integrating linguistic elements into the input-embedding layer [11].
- Multi-source neural model: A model with two encoder blocks—one for source word features and another for linguistic features—designed to improve context and translation accuracy.

4. EXPERIMENT

This section outlines the tests intended to assess the neural machine translation model that has been suggested for language isolation. The experiments focus on the integration of linguistic features and the effectiveness of the training datasets. The complete code for the experiments is available and can be accessed at the https://github.com/multi_source/corpus_sentence_extraction_script.

4.1 Linguistic Features

Three linguistic characteristics of isolating languages are used in this article. Lemma is a trait that is frequently utilized in information retrieval. Normalization allows different forms of a word to share a common representation, thereby improving the efficiency of information retrieval [37]. The second feature is POS tags, which provide information about syntactic roles within context. According to Clark, assigning POS tags to words in Chinese helps identify grammatical functions, supporting syntactic analysis and clarifying the structure of the language [38].

4.2 Experimental Data

In the context of isolating languages, we chose three medium-sized corpora derived from Asia. Leveraging Opus, an open-access platform for parallel corpora storage, we accessed diverse datasets. For Vietnamese - English, a data collection of translated movie subtitles: OpenSubtitles [39], QED [40] and wikimedia [41] as a training dataset, and 10% of the training dataset was reused as a validation dataset, while utilizing the ,wiki- media [41], TED2020 [42], NeuLab-TedTalks [41], and QED [40] as four test datasets. For Chinese-English translation, this article utilized wikiMatrix [43] corpus for training, validating and testing. Tables 2 and 3 present the training sample dataset's statistics for the language pairings Chinese-English and Vietnamese - English, as well as samples from the corresponding corpora used in this investigation.

Table 2: Some challenges for NMT [9] and examples.

Language	# Sentences	# Tokens
Vietnamese	5,248,771	77,016,972
English (Vietnamese)	5,248,771	59,578,519
Chinese	786,512	1,516,164
English (Chinese)	786,512	16,750,129

Table 3: Translation corpus sample of Vietnamese to English & Chinese to English.

Language	Source sentence	Target sentence

Vietnamese - English	Điều này có lẽ khác với những gì bạn mong đợi	This might have been different from what you expected.
	Album đó bán được nhiều hơn những album trước đó	The album sold more copies than previous releases of the band.
Chinese - English	正月，朝其主（王)	He is sometimes seen having arguments with his lord.
	我也不能找到任何防止们吃人的法律	Nor can I find any law which prevents us from eating people.

Although these datasets were carefully selected from diverse sources, the exclusion of data from other languages or text genres may limit the model's generalizability. Besides, the model's application in high-complexity domains such as healthcare, education, and others remains unexplored, and its performance with non-isolating or low-resource languages is yet to be tested. These factors could influence the model's generalizability in real-world applications.

4.3 Data Preprocessing

For each corpus, first we perform data filtering following these steps: remove NaN tokens, remove duplicate rows, delete data pairs whose source and target sentences are identical, and filter out html tags. Merging data from three corpora and training the Unigram vocabulary model will help cover all words from all three. For the final procedure of data processing, we tokenize data by subwording method via the Python SentencePiece library to reduce OOV in recognizing unseen words in the follow-up training step.

Exclusively in the multi-source training procedure, we process POS tagging data for the external input. In addition, we utilize specific Python libraries specialized to the linguistic characteristics of each selected language in three corpuses. We employ NLTK for the English dataset. Regarding the Vietnamese data, we use VnCoreNLP, and for the Chinese corpus, we rely on Jieba. These libraries are used for part-of-speech (POS) tagging, as they access dictionaries for each language and correctly identify the grammatical functions of every word in a sentence. Processed POS tag files for each language will then be passed into the external input of multi-source training.

4.4 Model Parameter Setting

In order to enhance the Transformer model's functionality in OpenNMT-tf, several modifications have been made to the configuration, including the incorporation of both token-only input and external sources. Table 4 shows the detailed parameter settings.

Table 4: Some challenges for NMT [9] and examples.

Property	Specification
Toolkit	OpenNMT-tf Transformer model configuration with external input and token-only input.
Oversampling Method	Weighted oversampling with ratios 1:10:20 for OpenSubtitles, QED, and Wikimedia corpora. This ensures equal sampling without significantly increasing the dataset size.
Batch Size	Set to 144, balancing training performance, hardware utilization, and model generalization.
Training Step	Set to 100,000 steps, equivalent to 3.6 epochs, sufficient for model data coverage as per OpenNMT default.

5. RESULTS AND DISCUSSION

Tables 5 and 6 summarize the experimental outcomes for Vietnamese to English and Chinese to English translation tasks. The outcomes presented in Table 5 indicate a quantitative comparison of BLEU and METEOR scores obtained through rigorous evaluation using SacreBLEU [44] and PyMeteor [45]. SacreBLEU is a standard tool known for its consistent and reproducible BLEU score calculations, while PyMeteor is a Python implementation of the METEOR metric, which emphasizes semantic similarity in translation assessment. Both tools are widely recognized for their accuracy in assessing machine translation performance. Specifically, the BLEU score for the multi-source model reached 25.95, signifying a 12.22% improvement over the single-source model's score of 23.125. Similarly, the METEOR score showed an increase of 4.02%, with the multi-source model achieving a score of 39.55 compared to 38.02

for the single-source model. These results, validated by our team, underscore the multi-source model's superior ability to generate translations that are more relevant for the context and more accurate. The improvements highlighted by these metrics suggest significant advancements in translation quality, particularly for challenging language pairs like Vietnamese - English.

Table 5: The outcomes of experiments on machine translation tasks between Vietnamese - English.

	BLEU scores		METEOR scores	
Dataset	Single source	Multi source	Single source	Multi source
wikimedia	23.1	25.7	38.1	39.4
TED2020	22.9	25.5	37.6	39.2
NeuLab-TedTalks	23.8	26.2	38.5	40.1
QED	23.7	26.4	37.9	25.6

Likewise, Table 6 elucidates that the multi-source model has yielded notable advancements in performance within the Chinese-English translation task. Although the BLEU score increases modestly by 0.5 points (from 17.91 to 18.4), the METEOR score reflects a significant increase to 36.44, suggesting that higher-quality translations are produced using the multi-source paradigm.

Table 6: Some challenges for NMT [9] and examples.

Model	BLEU scores	METEOR scores
Single-source model	17.91	35.73
Multi-source model	18.4	36.44

Table 7 indicates that the Multi-source NMT model yields results that are more aligned with the target sentences than those of the Single-source model. In the first example, the Multi-source NMT delivers a more accurate translation by maintaining the structure and meaning of the original sentence, especially in emphasizing skills and time pressure. Similarly, in the second example, the Multi-source translation aligns more closely with the target sentence by choosing a more natural expression rather than using an academic term as seen in the Single-source version. This demonstrates that the Multi-source NMT has better accuracy and contextual relevance, resulting in higher-quality translation as opposed to the single-source NMT.

Table 7: Sample translation results for the pair Vietnamese-English and Chinese-English: Single-source and Multi-source translation tasks.

Source sentence 1	Điều này có lẽ khác với những gì bạn mong đợi.
Single-source model	This could be different from your expectations.
Multi-source model	This might be different from what you expected.
Target sentence	This might have been different from what you expected
Source sentence 2	我也不能找到任何防止我们吃人的法律
Single-source model	I also cannot find any laws that prevent us from cannibalism
Multi-source model	I cannot find any laws that stop us from eating people either.
Target sentence	Nor can I find any law which prevents us from eating people.

6. CONCLUSION

The heavy reliance on translation context in existing high-resource machine translation models limits translation performance for isolating languages. This study proposed a multi-source neural model to address this issue by enhancing the representation capability of NMT systems. Two independent encoders were utilized to process lexical and linguistic features, optimizing language representation from multiple sources. Experimental results indicated significant improvements, with a METEOR score increase of +1.5 and a BLEU enhancement of +2.4. These improvements demonstrate a more accurate translation output, particularly in maintaining semantic coherence and lexical consistency, which are critical for applications requiring precise cross-language communication in fields such as business documentation, academic content, and multilingual education. Future research will explore incorporating advanced linguistic features, attention mechanisms, and broader language pairs to build on these improvements while assessing real-world applications in healthcare and education.

REFERENCES:

[1] Z. Tan, S. Wang, Z. Yang, G. Chen, X. Huang, M. Sun, and Y. Liu, "Neural machine translation: A review of methods, resources, and tools," AI

Open, Dec. 2020, doi: 10.48550/arXiv.2012.15515.

[2] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in Proc. 2014 Conf. Empirical Methods Natural Language Process. (EMNLP), Doha, Qatar, Oct. 2014, pp. 1724–1734, doi: 10.3115/v1/D14-1179.

[3] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS), Montreal, QC, Canada, 2014, pp. 3104–3112, doi: 10.48550/arXiv.1409.3215.

[4] W. Khan, A. Daud, K. Khan, S. Muhammad, and R. Haq, "Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends," NLP, vol. 2023, Art. no. 100026, doi: 10.1016/j.nlp.2023.100026.

[5] C. Zhai, S. Wibowo, and L. D. Li, "Evaluating the AI dialogue system's intercultural, humorous, and empathetic dimensions in English language learning: A case study," Cogn. Artif. Intell. Educ., vol. 2024, Art. no. 100262, doi: 10.1016/j.caeai.2024.100262.

[6] N. Arivazhagan, A. Bapna, O. Firat, D. Lepikhin, M. Johnson, M. Krikun, M. X. Chen, Y. Cao, G. Foster, C. Cherry, W. Macherey, Z. Chen, and Y. Wu, "Massively multilingual neural machine translation in the wild: Findings and challenges," arXiv, vol. 1907, Art. no. 05019, Jul. 2019, doi: 10.48550/arXiv.1907.05019.

[7] S. Hoshino, A. Kato, S. Murakami, and P. Zhang, "Cross-lingual transfer or machine translation? On data augmentation for monolingual semantic textual similarity," arXiv, vol. 2403, Art. no. 05257, Mar. 2024, doi: 10.48550/arXiv.2403.05257.

[8] K. Alkmim, R. Patel, and D. Dave, "Learning machine translation with linguistic interpretation," Preprints.org, Art. no. PPR774588, Dec. 2023, doi: 10.20944/preprints202312.0921.v1.

[9] P. Koehn and R. Knowles, "Six challenges for neural machine translation," in Proc. 1st Workshop on Neural Machine Translation, Vancouver, Aug. 2017, pp. 28–39, doi: 10.18653/v1/W17-3204.

[10] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in Proc. 6th Int. Conf. on Machine Learning (ICML), 2019, pp. 1–10, doi: 10.48550/arXiv.1905.02249.

[11] N. L. Pham, V. V. Nguyen, and T. V. Pham, "A data augmentation method for English-Vietnamese neural machine translation," IEEE Access, vol. 11, pp. 28034–28044, Mar. 2023, doi: 10.1109/ACCESS.2023.3252898.

[12] B. Zhang, D. Xiong, J. Su, and H. Duan, "A context-aware recurrent encoder for neural machine translation," IEEE/ACM Trans. Audio, Speech, Language Process., vol. 25, no. 12, pp. 2424–2432, Dec. 2017, doi: 10.1109/TASLP.2017.2751420.

[13] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," arXiv:1508.07909 [cs.CL], v5, 10 Jun. 2016, doi: 10.48550/arXiv.1508.07909.

[14] Z. Cao, Z. Qu, H. Kamigaito, và T. Watanabe, "Exploring intrinsic language-specific subspaces in fine-tuning multilingual neural machine translation," arXiv:2409.05224 [cs.CL], 8 Sep. 2024, doi: 10.48550/arXiv.2409.05224.

[15] Brazier, C., & Rouas, J.-L. (2024). "Conditioning LLMs with emotion in neural machine translation," Computational Linguistics, 50(1), 45-63. DOI: 10.48550/arXiv.2408.03150

[16] L. Zeng, "Generative-adversarial networks for low-resource language data augmentation in machine translation," Proc. ICNLP 2024, pp. 1–8, 2024, doi: 10.48550/arXiv.2409.00071.

[17] S. Honda, P. Fernandes, và C. Zerva, "Context-aware neural machine translation for English-Japanese business scene dialogues," Proc. MT Summit 2023, pp. 1–8, 2023, doi: 10.48550/arXiv.2311.11976.

[18] A. Nzeyimana, "Low-resource neural machine translation with morphological modeling," in Proc. NAACL Findings 2024, 2024, doi: 10.48550/arXiv.2404.02392.

[19] W. H. Her and U. Kruschwitz, "Investigating Neural Machine Translation for Low-Resource Languages: Using Bavarian as a Case Study," in Proc. 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages (SIGUL 2024), 2024, doi: 10.48550/arXiv.2404.08259.

[20] R. Sennrich and B. Haddow, "Linguistic Input Features Improve Neural Machine Translation," in Proc. First Conf. on Machine Translation: Volume 1, Research Papers, Berlin, Germany, Aug. 2016, pp. 83–91, doi: 10.18653/v1/W16-2209.

[21] P. Li, Y. Wang, J. Liu, A. Luo, S. Xu, and Z. Zhang, "Enhanced semantic representation model for multisource point of interest attribute alignment," Inf. Fusion, vol. 98, Art. no. 101852, Oct. 2023, doi: 10.1016/j.inffus.2023.101852.

[22] B. Zoph and K. Knight, "Multi-source neural translation," arXiv preprint, arXiv:1601.00710 [cs.CL], Jan. 2016. doi: 10.48550/arXiv.1601.00710.

[23] B. Zoph, D. Yuret, J. May, and K. Knight, "Transfer learning for low-resource neural machine translation," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, 2016, pp. 1568–1575. doi: 10.18653/v1/D16-1163.

[24] Y. Nishimura, K. Sudoh, G. Neubig, and S. Nakamura, "Multi-source neural machine translation with data augmentation," in Proceedings of the 15th International Conference on Spoken Language Translation, Brussels, 2018, pp. 48–53. doi: 10.18653/v1/2018.iwslt-1.7.

[25] G.-H. Choi, J.-H. Shin, and Y.-K. Kim, "Improving a multi-source neural machine translation model with corpus extension for low-resource languages," arXiv preprint arXiv:1709.08898, 2017. doi: 10.48550/arXiv.1709.08898.

[26] Q. Li, D. F. Wong, L. S. Chao, M. Zhu, T. Xiao, and J. Zhu, "Linguistic knowledge-aware neural machine translation," IEEE/ACM Trans. Audio, Speech, Language Process., vol. 26, no. 12, pp. 2341–2354, Dec. 2018, doi: 10.1109/TASLP.2018.2864648.

[27] Z. Z. Hlaing, Y. K. Thu, T. Supnithi, và P. Netisopakul, "Improving neural machine translation with POS-tag features for low-resource language pairs," Heliyon, vol. 8, no. 8, Art. no. e10375, Aug. 2022, doi: 10.1016/j.heliyon.2022.e10375.

[28] E. Pitler, A. Louis, and A. Nenkova, "Automatic evaluation of linguistic quality in multi-document summarization," in Proc. 48th Annu. Meet. Assoc. Comput. Linguistics, Uppsala, Sweden, 2010, pp. 544–554.

[29] K. Hashimoto, C. Xiong, Y. Tsuruoka, và R. Socher, "A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks," in Proc. 2017 Conf. Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017, pp. 1923–1933, doi: 10.18653/v1/D17-1206.

[30] S. Gururangan, S. Swayamdipta, O. Levy, R. Schwartz, S. R. Bowman, và N. A. Smith, "Annotation Artifacts in Natural Language Inference Data," in Proc. NAACL 2018, vol. cs.CL, 2018, pp. 1–6, doi: 10.48550/arXiv.1803.02324.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, và I. Polosukhin, "Attention Is All You Need," in Proc. NIPS 2017, vol. cs.CL, 2017, pp. 1–15. doi: 10.48550/arXiv.1706.03762.

[32] H. Santiago-Benito, D.-M. Córdova-Esparza, N.-A. Castro-Sánchez, T. García-Ramirez, J.-A. Romero-González, và J. Terven, "Automatic Translation between Mixtec to Spanish Languages Using Neural Networks," Appl. Sci., vol. 14, no. 7, Art. no. 2958, Mar. 2024. doi: 10.3390/app14072958.

[33] D. Gizlyk, "Neural networks made easy (Part 10): Multi-Head Attention," MetaTrader 5 — Examples, 4 Mar. 2021, [Online]. [Accessed: 23-Oct-2024].

[34] B. Gao và L. Pavel, "On the Properties of the Softmax Function with Application in Game Theory and Reinforcement Learning," arXiv, vol. 1704.00805, Art. no. 4, Aug. 2018. doi: 10.48550/arXiv.1704.00805.

[35] Y. Pan, X. Li, Y. Yang, và R. Dong, "Multi-Source Neural Model for Machine Translation of Agglutinative Language," Future Internet, vol. 12, no. 6, Art. no. 96, Jun. 2020, doi: 10.3390/fi12060096.

[36] J. Libovický, J. Helcl, và D. Mareček, "Input Combination Strategies for Multi-Source Transformer Decoder," arXiv, vol. 1811.04716, Nov. 2018, doi: 10.48550/arXiv.1811.04716.

[37] S. Bird, E. Klein, và E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly, Jan. 2009, ISBN: 978-0-596-51649-9.

[38] J. Zhao, X. Qiu, S. Zhang, F. Ji, và X. Huang, "Part-of-Speech Tagging for Chinese-English Mixed Texts with Dynamic Features," in EMNLP-CoNLL '12: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea, 2012, pp. 1379–1388.

[39] P. Lison và J. Tiedemann, "OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles," in Proc. Tenth Int. Conf. Lang. Resour. Evaluation (LREC'16), Portorož, Slovenia, May 2016, pp. 923–929.

[40] A. Abdelali, F. Guzman, H. Sajjad, và S. Vogel, "The AMARA Corpus: Building Parallel Language Resources for the Educational

Domain," in Proc. 9th Int. Conf. Lang. Resour. Evaluation (LREC), 2014.

[41] J. Tiedemann, "Parallel Data, Tools and Interfaces in OPUS," in Proc. Eighth Int. Conf. Language Resources and Evaluation (LREC'12), Istanbul, Turkey, May 2012, pp. 2214–2218.

[42] N. Reimers and I. Gurevych, "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation," in Proc. EMNLP 2020, 2020, doi: 10.48550/arXiv.2004.09813.

[43] H. Schwenk, V. Chaudhary, S. Sun, H. Gong, và F. Guzmán, "WikiMatrix: Mining 135M Parallel Sentences in 1620 Language Pairs from Wikipedia," arXiv, vol. 1907, no. 05791, 2019, doi: 10.48550/arXiv.1907.05791.

[44] M. Post, "A Call for Clarity in Reporting BLEU Scores," in Proceedings of the Third Conference on Machine Translation: Research Papers, Brussels, Belgium, Oct. 2018, pp. 186–191. doi: 10.18653/v1/W18-6319.

[45] R. Zembrodt, "PyMETEOR: Python implementation of METEOR," Test PyPI, vol. 0.0.1, Oct. 24, 2018

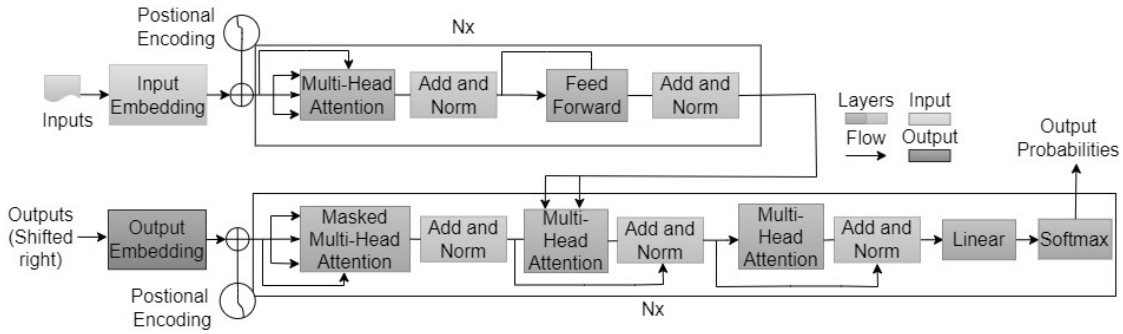

Figure 3: Base Architecture Of Transformer Model With Single Source: Encoder-Decoder Framework [32]

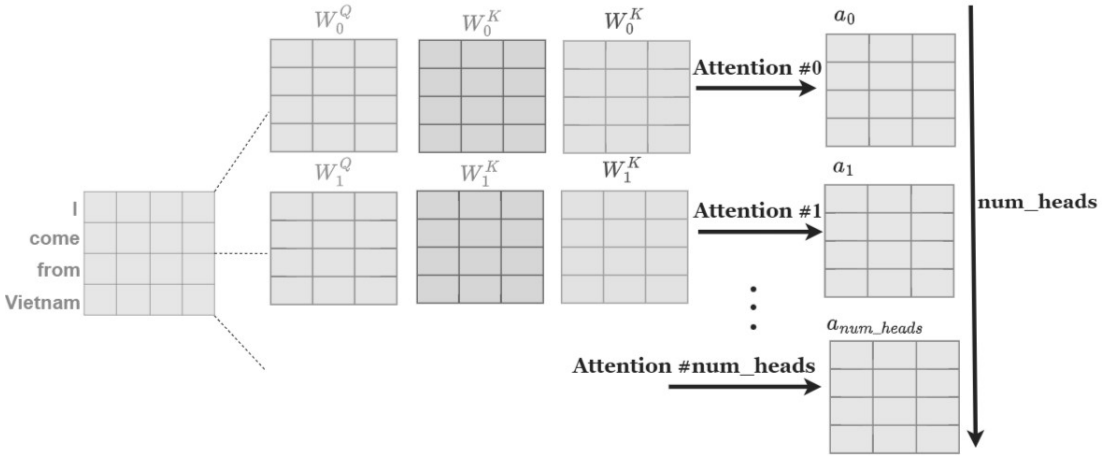

Figure 8: Multi-Head Self-Attention Approach For The Sentence 'I come from Vietnam'

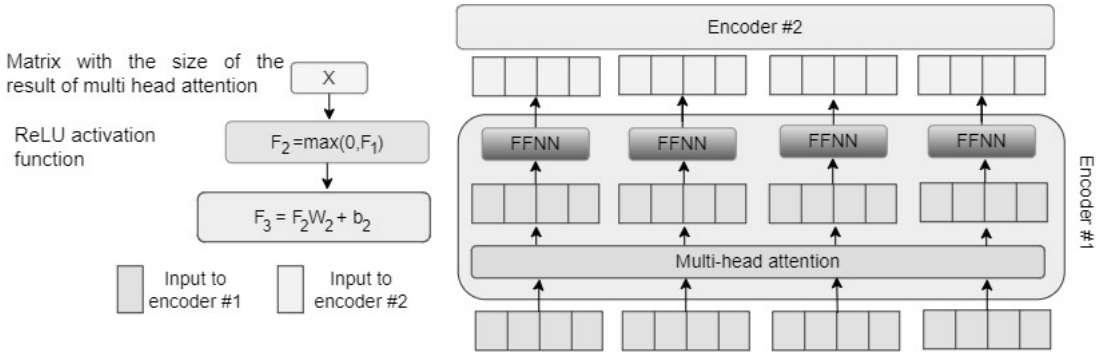

Figure 9: Feed-Forward Network Based On Position In Transformer Model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

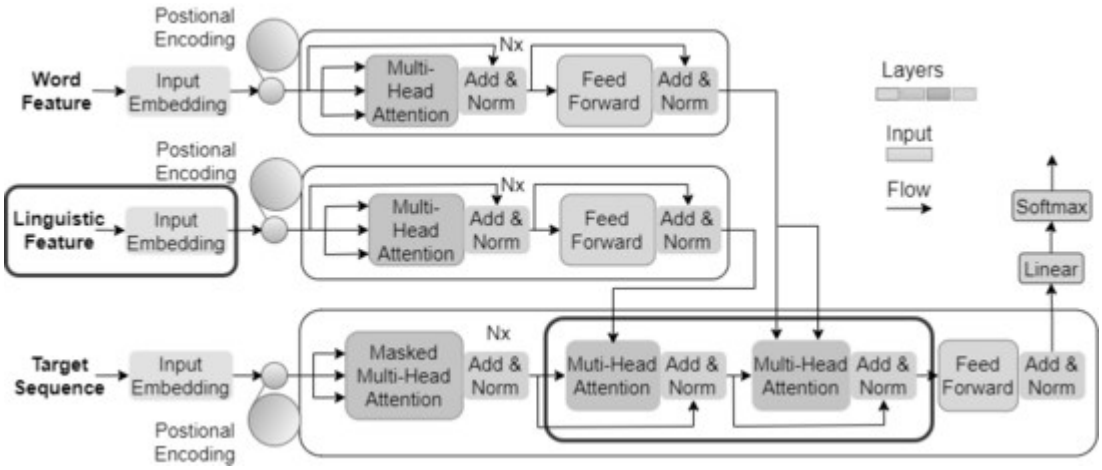

Figure 11: The Suggested Multi-Source Neural Model's Architecture [35]