International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 4431-4440

https://doi.org/10.70517/ijhsa463379

Syntactic Optimization and Semantic Structure Modification for English Translation Based on Semantic Enhancement Modeling

Yongjia Huang[1,*]
[1] School of Foreign Languages, Zhengzhou Shengda University, Zhengzhou, Henan, 451191, China
Corresponding authors: (e-mail: huang0105chen@163.com).

Abstract Improving the ability to deal with complex syntax and semantics is the key for English translation systems to move towards intelligence. In this paper, we incorporate a multimodal parallel fusion architecture into the design of the translation system, combining visual theme enhancement coding with detail fusion decoding to construct a cross-language-cross-modal semantic space. Semantic pre-tuning order training strategy and tree model syntactic encoding method are introduced to optimize the translation quality from source language to English. Experiments show that the BLEU values of this paper's method on four datasets significantly outperform mainstream models. In the translation of long sentences with (35,45] and (45,80] word counts, the BLEU enhancement values are up to 2.51 and 2.67. The range of BLEU values of this paper's method is enhanced to the range of 40-43 in the translation of complex sentences with syntactic and semantic structural adjustments.

Index Terms semantic enhancement, multimodal fusion, English translation, cross-modal semantic space, tree model syntactic encoding

I. Introduction

Machine translation uses high-performance computers as the computing core to realize the conversion between different natural languages, which occupies a large proportion in the field of artificial intelligence and machine learning [1]. The online translation service provided by a large number of machine learning models, although it has a wider use value in the scenarios with lower translation quality requirements, the degree of accuracy of machine translation is still low, and it still can't replace professional translators [2]-[4]. Especially in translating some long utterances, it is difficult to accurately characterize the differences in word order between the source and target languages [5].

However, as English sentences usually have more complex long difficult sentences such as main and subordinate complex sentences, determinative clauses, homonymous clauses, gerund clauses, etc., how to correctly analyze and understand the antecedents of the subordinate clauses as well as the demonstrative pronouns etc. is of great significance for the whole sentence [6]-[8]. At the same time, most of the English sentences, in order to achieve the purpose of strict accuracy and objectivity, often appear as complex and lengthy structure, mostly using passive voice and nominalized structure [9], [10]. Based on this, some scholars have implanted the source language syntactic information into the translation model, which effectively improves the accuracy of the description of long-distance sequencing, and at the same time, combined with the semantic structure analysis method, it effectively improves the translation accuracy in the process of foreign language long sentence translation [11]-[13].

In this paper, we design an English translation system containing multiple modules. The multimodal parallel fusion architecture is selected in the translation module, separating the basic translation module and the scalable visual enhancement module to reduce noise interference. The cross-lingual-cross-modal semantic space is constructed based on the cooperative attention mechanism to enhance syntactic and thematic consistency. Combine the tree LSTM model to encode the syntax and optimize the semantic order alignment by semantic pre-tuning order training. Utilize multiple datasets for performance testing of the models in this paper.

II. Analysis of English Translation Optimization Process Based on Semantic Enhancement

This chapter combines the steps of translation system design, translation semantic pre-tuning order, and syntactic coding to analyze in detail how to enhance the translation of source language-English through semantic enhancement modeling.

II. A.Architecture Design of English Translation System Based on Semantic Enhancement

II. A. 1) General system architecture

Combining multimodal translation modeling and requirement analysis, the English translation system architecture is divided into four major parts, including text processing module, image processing module, translation module, and interaction module. The translation module is the core part of the system, which consists of the basic translation structure and the multimodal fusion architecture, both of which remain architecturally independent in a parallel relationship. In the architecture, the multimode fusion architecture is designed to be pluggable, and the operation of the translation module is not dependent on the fusion architecture. In addition, the fusion architecture can be extended according to the actual needs and is not dependent on the specific multimodal learning form. In this paper, the multimodal fusion architecture is materialized as an encoding architecture for visual topic enhancement and a decoding architecture for cross-modal fusion of visual details. Figure 1 shows the overall architecture of the translation system for modal semantic parallel fusion.

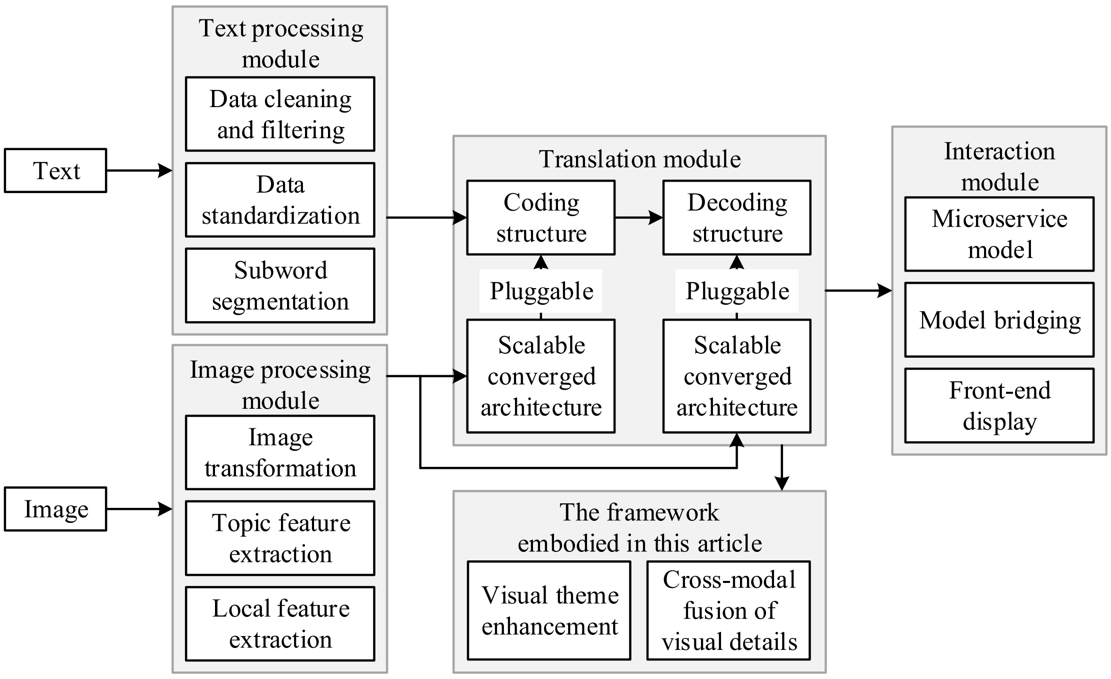


Figure 1: Translation system with parallel fusion of modal semantics

Among them:

(1) The text processing module accepts source language sentences in the training phase or the speculation phase, cleans the data through the built-in translation data preprocessing process, symbols standardization and other processes to get a number of input units, and uses BPE subword slicing technology to narrow the word list and solve the OOV problem, and finally generates the input form acceptable to the translation model.

(2) The image processing module accepts the image and returns the features needed by the translation model, including two major steps of image transformation and feature extraction, the image transformation includes image scaling, channel normalization and so on. The feature extraction returns visual information of corresponding granularity according to the demand using the visual feature extraction model of pain.

(3) Translation module is the core of the translation system, which accepts source language sentences and optionally inputs auxiliary images to generate target language translations. The translation module consists of a standard translation model and a parallel multimodal fusion architecture. According to the modal-semantic relations in the actual environment, the parallel fusion architecture at the codec end is materialized into specific modal learning modes and enhancement architectures, so as to assist the translation model in outputting the visually-semantically-enhanced translation results. In this paper, two semantic enhancement modes are set up, the parallel multimodal fusion architecture at the encoding end is designed for visual theme enhancement, and the parallel fusion architecture at the decoding end is for cross-modal fusion of visual details, thus realizing syntactic optimization and semantic structure modification of English translation.

(4) The interaction module is used to interact with the user, and the client in the text is realized using the Web to provide the user with a simple and easy-to-operate clean surface. Which when the user chooses to translate between low-resource languages, the interaction module will use a bridging approach to utilize the two translation models on the English side to achieve the effect of mutual translation.

II. A. 2) Cross-language-cross-modal semantic space construction tasks

The semantic space task constructs a cross-linguistic-cross-modal semantic space by receiving source language implicit features, target language thematic implicit features, and image thematic features, so that the encoder pays attention to both syntactic structure and semantic features when performing translation, thus enhancing the integration of semantic information. Therefore, the visual information from outside does not directly intervene in the translation model in the theme-enhanced translation model, which reduces the influence of visual noise, while the architecture is dismantled in the speculative phase without relying on visual input.

Specifically, the source language features H_{sou} generated by the encoder and the target language first draft topic features H_{tar} obtained from the one-stage decoding, as well as the image topic features H_{vis} are projected into the same semantic space. Where $H_{sou} \in R^{N\times d_{sou}}, H_{tar} \in R^{M\times d_{tar}}, H_{vis} \in R^{K\times d_{vis}}$, $d_{sou}, d_{tar}, d_{vis}$ denote the dimensions denoting the source language implicit features, the target language topic, and the image topic features, respectively, and N, M, K denote the source language sentences, the length of the primitive drafts, and the number of the images in the input image set, and since this paper uses probabilistic mapping to generate the primitive drafts, here $M = N$. The instances of each modality in the semantic space are represented as individual vectors, thus a collaborative attention needs to be utilized to compute the text-image correlation, taking the source language as an example, firstly, a text-visual affinity matrix is computed, denoted as Eq. (1):

$$A = \tanh(H^T_{sou} W_A H_{vis}) \tag{1}$$

In Eq. (1) $W_A \in R^{d_{sou}\times d_{vis}}$ represents the weight parameter, and $A \in R^N \times K$ is the affinity matrix representing the relevance of the words in the sentence to the pictures in the image group. Subsequently A is incorporated as a feature into a parallel attentional mechanism to learn the attentional mapping between different modalities. For the source language hidden features, their vector representations in the semantic space are obtained using the following equation, Eqs. (2)-(4) demonstrate the computational process:

$$C_{sou} = \tanh(W^C_{sou} H_{sou} + A(W^C_{vis} H_{vis})) \tag{2}$$

$$a_{sou} = softmax(W^a_{sou} C_{sou}) \tag{3}$$

$$E_{sou} = W^E_{sou} \sum_n^N a_{sou_n} H_{sou_n} \tag{4}$$

where $W^C_{sou} \in R^{d_{sou}\times d_{spa}}, W^C_{vis} \in R^{d_{vis}\times d_{spa}}, W^a_{sou} \in R^{d_{spa}}$ is the training parameter in synergetic attention, and d_{spa} denotes the dimensionality of semantic spatial features. $W^E_{sou} \in R^{d_{sou}\times d_{spa}}$ maps the weighted summed source language hidden features to the public semantic space, denoted as E_{sou} . For the input image group, computing its vector representation in the semantic space is a mirroring process as shown in Eqs. (5)-(7):

$$C_{vis} = \tanh(W^C_{vis} H_{vis} + A(W^C_{sou} H_{sou})) \tag{5}$$

$$a_{vis} = softmax(W^a_{vis} C_{vis}) \tag{6}$$

$$E_{vis} = W^E_{vis} \sum_k^K a_{vis_k} H_{vis_k} \tag{7}$$

For the target language first draft feature H_{tar} , the same co-attention computation with the visual feature H_{vis} is performed to obtain the target language representation E_{tar} in the semantic space. Since the visual features are involved in the computation of both source and target languages, thus the representations of the two visual features in the semantic space E^{sou}_{vis} and E^{tar}_{vis} are obtained, and the final vector representation of the visual features in the semantic space is the summed average of the two as in Eq. (8):

$$E_{vis} = \frac{1}{2} E^{sou}_{vis} + \frac{1}{2} E^{tar}_{vis} \tag{8}$$

The ordering loss is utilized to find the pairwise relationship between $E_{sou}, E_{tar}, E_{vis}$, thus constructing the semantic space, and Eq. (9) demonstrates how the ordering loss is expressed:

$$L = L_{sou-vis} + L_{tar-vis} + \beta L_{sou-tar} \tag{9}$$

The $L_{sou-vis}$ and $L_{tar-vis}$ in Eq. (9) is the textual visual inter-modal ranking loss, associated to source language-visual and target language-visual. This loss is used to learn the cross-modal representation, in the case of source-language-visual, assuming that the set of source-language features in the semantic space is Sou, and the set of visual features is Vis, which is represented as Eq. (10):

$$\begin{aligned} L_{sou-vis} = & \sum_{i}\sum_{j}\max\{0,\gamma_1-\cos(Vis_i,Sou_i)+\cos(Vis_i,Sou_{jeG_i})\} \\ & +\sum_{j}\sum_{i}\max\{0,\gamma_1-\cos(Sou_j,Vis_j)+\cos(Sou_j,Vis_{ieG_j})\} \end{aligned} \tag{10}$$

In Eq. (10) γ is the boundary parameter to control the relevance threshold. cos represents the cosine similarity, and i,j is the index of the instances in the set. There will be no loss between visually shared instances, $i \notin G_j, j \notin G_i$ indicates that there is no visual sharing between instances indexed i,j, i.e., the topics are not relevant.

The ranking loss within textual modality is more stringent compared to the cross-modal ranking loss, the index inconsistency between the source language and the target language i.e., the loss occurs, let the set of features of the target language in the semantic space be Tar, denoted as Eq. (11):

$$\begin{aligned} L_{sou-tar} = & \sum_{i}\sum_{j}\max\{0,\gamma_2-\cos(Tar_i,Sou_i)+\cos(Tar_i,Sou_{j\neq i})\} \\ & +\sum_{j}\sum_{i}\max\{0,\gamma_2-\cos(Sou_j,Tar_j)+\cos(Sou_j,Tar_{i\neq j})\} \end{aligned} \tag{11}$$

When the loss decreases, the cosine similarity between topic-independent image-text pairs will decrease, and the cosine similarity between indexed differently (i.e., unpaired in the dataset) source-target language sentences will decrease, with a tighter semantic structure.

II. B. Translation semantic pre-conditioning training

II. B. 1) Pre-conditioning model

In this paper, we establish a pre-tuning order model based on neural networks, decompose the vocabulary tuning problem into a two-by-two ordering problem, and score the ordering through a multi-layer neural network. Specifically, for sentence $src=\{w_1,w_2,\cdots,w_n\}$, the scoring of the ordering result is given as:

$$\begin{aligned} & s(\pi,src)=\sum_{i<j}s_{NN}(i,j)[order(\pi(i),\pi(j))]+ \\ & s_{sparse}(i,j,\pi(i),\pi(j),src)oder(\pi(i),\pi(j))=\begin{cases}0,\pi(i)<\pi(j) \\ 1,else\end{cases} \end{aligned} \tag{12}$$

where s_{NN} and s_{sparse} are neural network and sparse feature computation scoring values, respectively, the inputs are the i,j th contextualized vocabulary, and the outputs are 2D vectors. The output optimal solution for determining the model tuning order is:

$$\pi^* = \arg\max_{\pi} s(\pi) \tag{13}$$

II. B. 2) Sequencing training data

Sequencing data training mainly obtains training data from bilingual parallel precisions and learns the parameters of the model based on the training data. For a bilingual sentence pair (e,f,α) with word alignment information, where e is the source language sentence, f is the target sentence, and a is the word alignment relation between the two. To obtain the source language sentence e reordering π^* so that it is similar to the target language sentence f order, the cross-linking number is used to evaluate the result of the ordering, and the word alignment connection is denoted by the number pair (i,j), i.e., the interconnection relation is established between the i th word of the source language and the j th word of the target language, and then the two words are said to be linked (i_1,j_1) and (i_2,j_2) are intersection relations if the relation in equation (14) is satisfied:

$$(i_1-i_2)*(j_1-j_2)<0 \tag{14}$$

Define at this point:

$$c(i_1,j_1,i_2,j_2)=\begin{cases}1 & (i_1-i_2)*(j_1-j_2)<1 \\ 0 & \text{else}\end{cases} \tag{15}$$

Then the source language is a kind of reordering π the corresponding formula for the number of cross-connections is:

$$crosslink(\pi,e,f,a)=\sum_{\substack{alinkpairs\\(\pi(i_1),\alpha(i_1),\pi(i_2),\pi(i_2))}} c(\pi(i_1),a(i_1),\pi(i_2),a(i_2)) \tag{16}$$

The minimum reordering π^* of the reordering in the source language is obtained by computing the number of cross-connections:

$$\pi^* = \arg\min_{\pi} \mathrm{crosslink}(\pi,e,f,a) \tag{17}$$

When the source and target language orderings are identical, this number of cross-links is 0. The greater the ordering variability, the greater the number of cross-links. Since the relationship between the number of reorderings and the sentence length presents a

$$L(\theta)=\sum_{(e,f,a)} \max(0.1+s(\pi^-)-s(\pi^*)) \tag{18}$$

where π^- is the reordering with the highest score in the ranking. For all sentence pairs in the bilingual corpus, one sentence pair is randomly selected from it and CKY decoded with the current parameter values to obtain π^- and compared with π^*, and if the loss of the comparison is not 0, then the parameters are updated according to the minimum loss gradient obtained:

$$\theta \leftarrow \theta - \gamma\nabla L(\theta) \tag{19}$$

where γ is the learning rate and $\nabla L(\theta)$ is the gradient of the sparse feature weight parameter, corresponding to the relation:

$$\nabla L(\theta) = f_{sparse}(\pi-) - f_{sparse}(\pi^*) \tag{20}$$

f_{sparse} is the system feature vector, which can be obtained by back propagation algorithm calculation. When the model parameters are initialized, the learned vocabulary vector is used as the initial value of the lookup table parameters, the parameters of the two linear layers of the neural network are initialized to a small interval, and the initial value of the sparse feature weights is set to zero.

II. C.Tree model-based syntactic coding approach

Intuitively, a tree-based model is used to encode syntactic information since it has standard tree-structured features. A common tree-based neural network model is the recurrent neural network model (RNN). It is worth noting that the tree-based recurrent neural network model originated from the most basic recurrent neural network model.

Currently, recurrent neural network models are the most commonly used models in computational linguistics, and their temporal information processing features are making them suitable for encoding texts with temporal features. One of the most representative variants of recurrent neural network models is the LSTM-based recurrent neural network model, which, by designing several kinds of gate mechanisms, allows the recurrent neural network to avoid the problem of gradient vanishing or gradient exploding that arises in long text input. Specifically, the tensor flow and computation inside the LSTM tuple at moment t can be represented as the following steps:

$$\begin{aligned}
i_t &= \sigma(W^i[x_t;h_{t-1}],b_i)\\
f_t &= \sigma(W^f[x_t;h_{t-1}],b_f)\\
o_t &= \sigma(W^o[x_t;h_{t-1}],b_o)\\
\hat{h}_t &= \mathrm{Tanh}(W^c[x_t;h_{t-1}],b_c)\\
c_t &= f_t \odot c_{t-1} + i_t \odot \hat{h}_t\\
h_t &= o_t \odot \mathrm{Tanh}(c_t)
\end{aligned} \tag{21}$$

where σ is a Sigmoid function with an output value between 0 and 1 to model the gating mechanism. Specifically, i_t, f_t and o_t are input gates, forgetting gates and output gates, respectively. c_t is a memory gate that controls what proportion of the history information is injected into the cell receiving the current step.Tanh is a nonlinear activation function. $\odot$ is the element-level vector multiplication.

Next, consider transforming the LSTM-based recurrent neural network model into a recurrent neural network model that can handle non-temporal, tree-structured information. Generally, if the basic unit in the recurrent neural network model is set as an LSTM tuple, a tree LSTM model (TreeLSTM) is obtained. Considering that compared with recurrent neural network model tree LSTM model has better tree structure coding effect and wider application, here we take tree LSTM model as the core to introduce the content related to syntactic optimization for English translation. We know that the temporal-based LSTM model processes the current t moment while also considering the information from the previous temporal sequence $t-1$; while for the tree LSTM model, the LSTM tuple at the position of the current t node in the tree structure will consider the information from all its children nodes at the same time. Figure 2 illustrates the internal computational mechanism of an LSTM tuple at the current t moment.

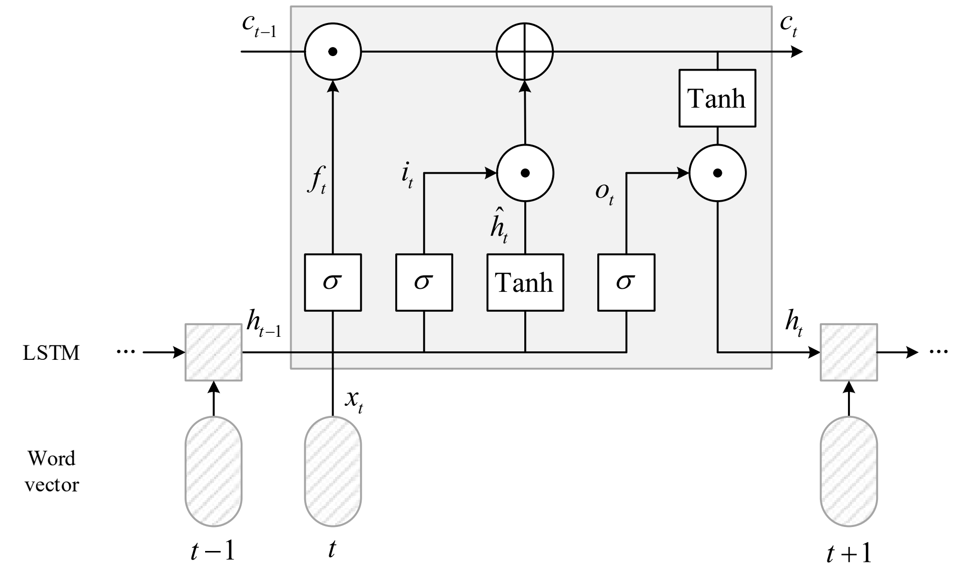


Figure 2: Internal computing mechanism of the LSTM cell at time t

Existing work categorizes tree LSTM models into Child-Sum tree LSTM models and N-ary tree LSTM models based on the difference in the way the information about the child nodes in the tree structure is handled. These two models are suitable for encoding structural information in dependent and phrase-component syntax, respectively. Intuitively, Dependency Syntax trees are suitable for Child-Sum Tree LSTM model because of the complexity of branching and the variable number of children of each node as well as the unordered nature of the tree. In the phrase constituent tree, the order of the children of a phrase constituent node is also more important, and the state information of the children is more fine-grained, for example, the information of the noun phrase constituent from the right branch node at the current node may be more important than that from the verb phrase with the left branch node, so it is suitable to use the N-ary tree LSTM model.

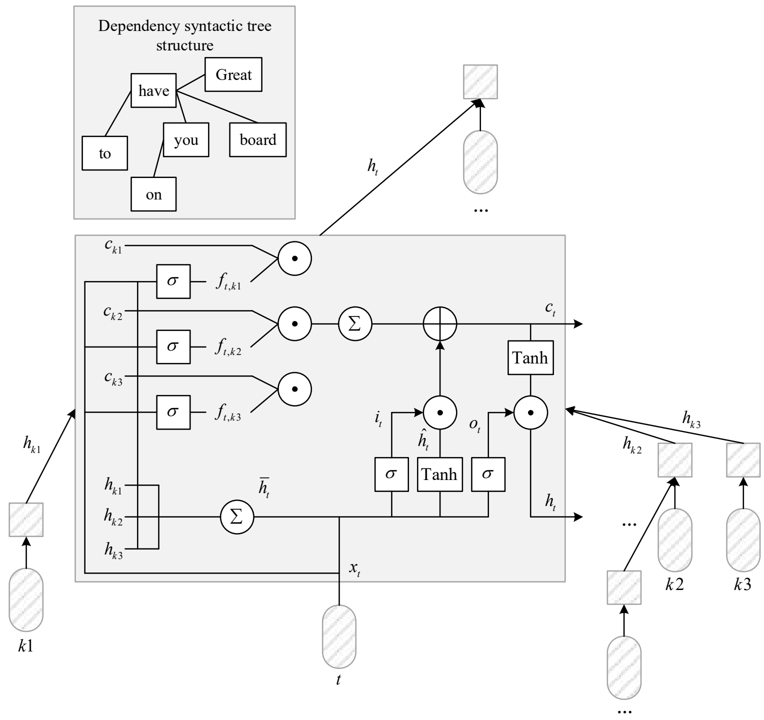


Figure 3: Internal computing process

Specifically, the LSTM tuple computation for the current position t in the Child-Sum tree LSTM model has:

$$
\begin{aligned}
\bar{h}_t &= \sum_{k \in C(t)} h_k \\
i_t &= \sigma(W^i[x_t;\bar{h}_t]],b_i) \\
f_{t,k} &= \sigma(W^f[x_t;h_k],b_f) \\
o_t &= \sigma(W^o[x_t;\bar{h}_t],b_o) \\
\hat{h}_t &= \mathrm{Tanh}(W^u[x_t;\bar{h}_t],b_u) \\
c_t &= \sum_{k \in C(t)} f_{t,k} \odot c_k + i_t \odot \hat{h}_t \\
h_t &= o_t \odot \mathrm{Tanh}(c_t)
\end{aligned}
\tag{22}
$$

where W,U as well as b are parameters and $C(t)$ refers to the children of node t. i_t,o_t and c_t are the individual gated switches, and h_t is the hidden state output representation of the tree LSTM tuple. $f_{t,k}$ is the oblivious gating switch specific to the k th child node of the current t -node. Compared to the temporal LSTM, the Child-Sum Tree LSTM model utilizes information from all child nodes and each child node is equipped with its own forgetting gate, which allows for more flexible control of the information from the child nodes. The specific internal computational flow is given in Fig. 3.

As for the phrase-component syntax, in order to realize a finer-grained control of the information about the children of the current node, the metacells of the N-ary tree LSTM take into account the influence of the state of the children in the input information of each gating:

$$
\begin{aligned}
i_t &= \sigma(W^i x_t + \sum_{k \in C(t)} U_k^i h_k + b) \\
f_{t,k} &= \sigma(W^f x_t + \sum_{q \in C(t)} U_{k,q}^f h_q + b) \\
o_t &= \sigma(W^o x_t + \sum_{k \in C(t)} U_k^o h_k + b) \\
\hat{h}_t &= \mathrm{Tanh}(W^u x_t + \sum_{k \in C(t)} U_k^u h_k + b) \\
c_t &= \sum_{k \in C(t)} f_{t,k} \odot c_k + i_t \odot \hat{h}_t \\
h_t &= o_t \odot \mathrm{Tanh}(c_t)
\end{aligned}
\tag{23}
$$

where the branching children of the first t nodes form an ordered set $C(t)$ and k is the k th branching child of the current node. U_k^* is a separate parameter for the child nodes under gating*. Unlike Child-Sum tree LSTMs, N-ary tree LSTMs have their own exclusive parameter U_k for each child node to control this fine-grained information difference and help capture the order change among the resident child nodes by summing them up individually.

It is worth noting that for tree-based syntactic encoders, there are other schemes for encoding syntactic information other than encoding a complete syntactic tree using the above. For example, different traversal directions such as top-to-bottom as well as bottom-to-top of the tree information are considered, resulting in different encoding schemes. For dependency syntactic trees, it is possible to encode only the substructures related to the target, e.g., consider 1) encoding only the information between a number of neighboring nodes of a certain node of the local second or third order, etc., and 2) encoding only the substructures, or shortest dependency paths, between a certain two target words. For phrase constituent syntactic trees, it can be considered not to encode the label node information of lexical annotation in the first level. Syntactic optimization and semantic structure construction when translating from source language to English is achieved by encoding syntactic information.

III. English Translation Optimization Practice Based on Semantic Enhancement Modeling

In this chapter, four datasets are selected as research data to perform multi-model performance comparison and sentence length experiments, and combined with syntactic and semantic adjustment experiments to test the actual translation effect of this paper's method.

III. A. Validation of model performance benefits

III. A. 1) Experimental data

In order to test the effect of syntactic optimization and semantic structure transformation of English translation based on the semantic enhancement model proposed in this paper, 4 datasets are selected as the base data for source language-English translation experiments.

Table 1 shows the data statistics of the 4 datasets. The datasets used include IWSLT15 Vietnamese→English (Vi→En), NC11 German→English (De→En), IWSLT14 French→English (Fr→En), and WMT18 Turkish→English (Tr→En). 1) The IWSLT15 Vietnamese→English (Vi→En) dataset contains a total of 134K bilingual sentence pairs, and the calibration set and test set contains 1557 sentences and 1266 sentences respectively. 2) NC11 German→English (De→En) dataset totals 239K bilingual pairs, and the calibration and test sets contain 2168 sentences and 2998 sentences respectively. 3) IWSLT14 French→English (Fr→En) dataset totals 162K bilingual pairs, and the calibration and test sets contain 7285 sentences and 6751 sentences. 4) The IWSLT14 French→English (Fr→En) dataset totals 209K bilingual pairs, and the calibration and test sets contain 3005 sentences and 3006 sentences, respectively. The dataset has more data and can reflect the performance level of the model more comprehensively.

Table 1: Data statistics of four datasets

Data set	Data type	Number of sentences
IWSLT15 Vi→En	Training set	134318
	Verification set	1557
	Test set	1266
NC11 De→En	Training set	239281
	Verification set	2168
	Test set	2998
IWSLT14 Fr→En	Training set	162239
	Verification set	7285
	Test set	6751
WMT18 Tr→En	Training set	209071
	Verification set	3005
	Test set	3006

III. A. 2) Translation performance comparison

The method of this paper is compared with Multi-Task method based on shared machine translation and syntactic parsing tasks, MixedEnc, PASCAL, the best performing syntactic application method at present, Mult-Task with parameter optimization, LISA, a method that pools syntactic information and attention, and S&H, a method that incorporates syntactic information into an encoder. In addition to this, comparisons are made with other machine translation models such as ELMo, CVTU, SAWR, DynamicConv, Tied-Transformerl, Macaron, C-MLMU131 which fuses pre-trained models, and BERT-fused on the IWSLT14 and IWSLT15 translation tasks.

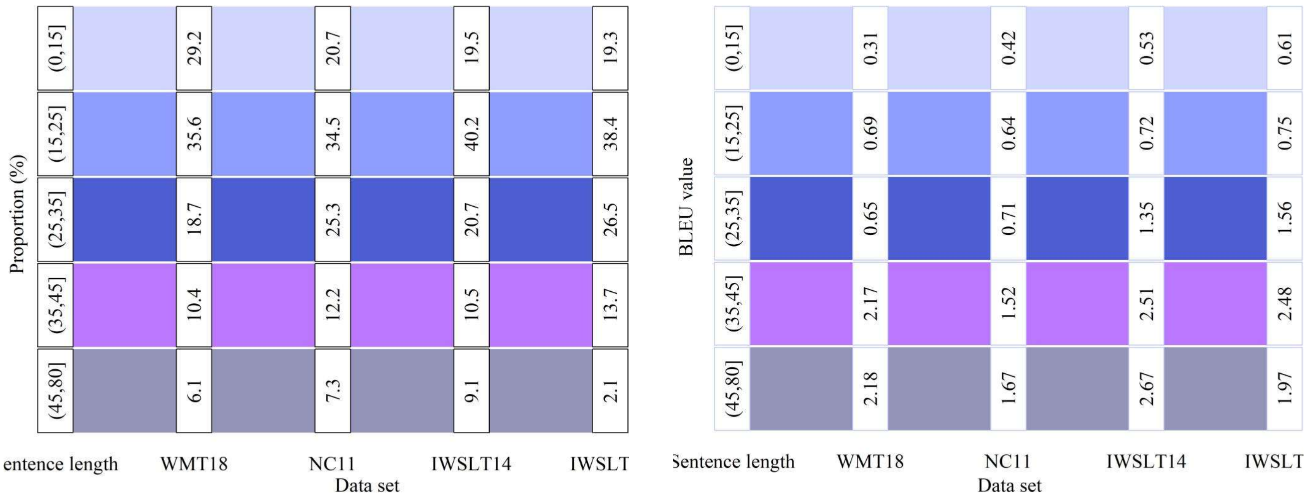
Table 2 shows the performance of the different methods in each English translation task using BLEU as a measure. Due to the different structure of each model, the missing values in the table indicate that the model cannot be applied to the corresponding dataset for training and testing. Comparing the size of BLEU values of different methods in English translation tasks, it is found that the BLEU values of this paper's English translation optimization method based on semantic enhancement model in the four datasets reach WMT18 Test set (14.99), NC11 Test set (25.98), IWSLT14 Verification set (37.88), IWSLT14 Test set (36.17), IWSLT15 Verification set (28.99), and IWSLT15 Test set (29.37), which are higher than the comparison methods. The data comparison results show that this paper's method has more obvious performance advantages in source language-English translation.

Table 2: Evaluation Results

Medol	WMT18	NC11	Medol	IWSLT14		IWSLT15	
	Test set	Test set		Verification set	Test set	Verification set	Test set
-	-	-	ELMo	-	-	-	29.31
-	-	-	SWAR	-	-	-	29.08
Mixed Enc	9.61	-	CVT	-	-	-	29.63
Multi -Task	10.64	-	C-MLM	36.90	35.64	27.86	31.50
Transformer	13.14	25.01	Transformer	35.28	34.40	27.43	30.75
+Multi -Task	14.02	24.83	Tied-Transformer	-	35.52	-	-
+S&H	13.05	25.55	Dynamic Conv	-	35.21	-	-
+LISA	13.66	25.32	Macaron	-	35.44	-	-
+PASCAL	14.04	25.91	BERT -fused	-	36.10	-	31.54
Article method	14.99	25.98	Article method	37.88	36.17	28.99	29.37

III. A. 3) Sentence length experiment

The English translation performance enhancement of this paper's method on sentences of different lengths is further tested. Figure 4 shows the sentence length statistics and translation performance improvement. Figure 4(a) statistics the distribution of source language sentence length in the datasets. It can be seen that the sentence lengths of the four datasets are mainly distributed within 35 words (all more than 50%), e.g., 80.4% of the sentences on the IWSLT14 dataset are less than 35 words long. Figure 4(b) demonstrates the translation performance improvement of each dataset in different sentence length intervals. It can be seen that this paper's method can improve the translation performance more stably on all length intervals of the test set sentences, especially for long sentences. In (0,15] sentence length, the highest BLEU value improvement is 0.61; in (15,25] the highest improvement is 0.75; in (25,35] it reaches 1.56; and in (35,45] and (45,80], the highest performance improvement can reach 2.51 and 2.67. This phenomenon is in line with the expectation, because in general the longer the sentence is, the more complex the syntactic and semantic structure is, and performing translation semantic preordering and tree model syntactic encoding can guide the model to pay more attention to the syntactic relationship between sentences and semantic structure, which in turn improves the quality of translation.



(a) Sentence length statistics

(b) Changes in translation performance

Figure 4: Sentence length statistics and changes in translation performance

III. B. Comparison of the effects of models based on syntactic and semantic adjustments

III. B. 1) Comparison of experimental effects of syntactic and semantic adjustments

In order to investigate whether this paper's English translation method based on semantic enhancement model still has excellent translation performance under complex syntactic and semantic structures, this section sets up tuning experiments to syntactically and semantically adjust the sentences in the four datasets and compares the translation effects of this paper's method with the baseline model before and after the tuning experiments. Table 3 shows the comparison of the effect of this paper's method without adding the tree model and the baseline model in the syntactic and semantic adjustment experiments. The BLEU values of the baseline model in the four datasets are 33.71, 32.55, 32.64, and 33.46, ranging from 32-34, while the range of the BLEU values of this paper's method before the syntactic and semantic structural adjustment of the sentences has reached between 32-36, and the range of the BLEU values is further increased to between 38-40 after the syntactic and semantic structural adjustment of the sentences. It shows that the method in this paper has better English translation performance compared to the baseline model without adding the tree model and with more complex syntactic semantics.

Table 3: The adjustment experiment results without adding the tree model

Data set	Baseline	Original		Syntactic and semantic adjustment	
		Primitive syntax	Original semantics	Syntactic adjustment	Semantic structure transformation
WMT18	33.71	35.08	35.04	38.28	38.27
NC11	32.55	32.72	32.77	38.55	38.58
IWSLT14	32.64	32.89	32.93	39.37	39.49
IWSLT15	33.46	33.75	33.86	39.86	39.93

III. B. 2) Comparison of the experimental effects of syntactic and semantic adjustments added to tree models

Table 4 shows the effect of this paper's method in the syntactic and semantic adjustment experiments after adding the tree model. After adding the tree model, the BLEU value of this paper's method is further improved from the range of 38-40 to 40-43, which also verifies that after combining the tree model with syntactic coding information analysis, this paper's method can realize semantic enhancement in the process of English translation, so that the syntax of the final output translated sentences is more in line with the habits of English usage and the semantic structure is more complete.

Table 4: BLEU value of the adjustment experiment after adding the tree model

Data set	Syntactic and semantic adjustment	Syntactic and semantic adjustment + LSTM
WMT18	38.28	40.73
NC11	38.57	40.98
IWSLT14	39.43	41.65
IWSLT15	39.90	42.51

IV. Conclusion

In this paper, we propose an English translation optimization method based on semantic enhancement model to improve the English sentence translation quality of the system from both syntactic and semantic structures. In the performance test of four datasets, the BLEU value of this paper's method reaches a maximum of 37.88. The performance of long sentence translation is improved by a maximum of 2.67 compared with the benchmark model. The BLEU value of complex sentence translation ranges from [40,43]. The effectiveness of semantic enhancement and syntactic synergy is verified. The adaptive multimodal selection mechanism will be explored in the future to improve translation intelligence.

References

[1] Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., ... & Joulin, A. (2021). Beyond english-centric multilingual machine translation. Journal of Machine Learning Research, 22(107), 1-48.

[2] Lin, L., Liu, J., Zhang, X., & Liang, X. (2021). Automatic translation of spoken English based on improved machine learning algorithm. Journal of Intelligent & Fuzzy Systems, 40(2), 2385-2395.

[3] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., ... & Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics, 5, 339-351.

[4] Bowker, L. (2020). Machine translation literacy instruction for international business students and business English instructors. Journal of Business & Finance Librarianship, 25(1-2), 25-43.

[5] He, H. (2023). An intelligent algorithm for fast machine translation of long English sentences. Journal of Intelligent Systems, 32(1), 20220257.

[6] Zhang, T. (2020). Recognition and Segmentation of English Long and Short Sentences Based on Machine Translation. International Journal of Emerging Technologies in Learning, 15(1).

[7] Bi, S. (2020). Intelligent system for English translation using automated knowledge base. Journal of Intelligent & Fuzzy Systems, 39(4), 5057-5066.

[8] Currey, A., & Heafield, K. (2019, August). Incorporating source syntax into transformer-based neural machine translation. In ACL 2019 Fourth Conference on Machine Translation (pp. 24-33). Association for Computational Linguistics.

[9] Sun, C. A., Mu, J., Xiao, M., Liu, H., & He, P. (2025). Semantic Structure Invariance-Based Metamorphic Testing for Machine Translation Systems. IEEE Transactions on Reliability.

[10] Song, L., Gildea, D., Zhang, Y., Wang, Z., & Su, J. (2019). Semantic neural machine translation using AMR. Transactions of the Association for Computational Linguistics, 7, 19-31.

[11] Liu, X., He, J., Liu, M., Yin, Z., Yin, L., & Zheng, W. (2023). A scenario-generic neural machine translation data augmentation method. Electronics, 12(10), 2320.

[12] Yang, H. (2024). Optimized English Translation System Using Multi-Level Semantic Extraction and Text Matching. IEEE Access.

[13] Wang, Y., Li, D., Shen, J., Xu, Y., Xu, M., Funakoshi, K., & Okumura, M. (2024, November). LAMBDA: Large Language Model-Based Data Augmentation for Multi-Modal Machine Translation. In Findings of the Association for Computational Linguistics: EMNLP 2024 (pp. 15240-15253).