MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SIENCE AND TECHNOLOGY

NGUYEN THI BICH DIEP

RESEARCH AND DEVELOPMENT OF APPROACHES BASED ON STRUCTURE AND STATISTICS IN AUTOMATIC TRANSLATION OF VIETNAMESE SIGN LANGUAGE

SUMMARY OF DISSERTATION ON COMPUTER SCIENCE

Code: 9 48 01 01

Ha Noi – 2023

The dissertation is completed at: Graduate University of Science and Technology, Vietnam Academy Science and Technology

Supervisors:
Supervisor 1: Dr. Vu Tat Thang, Institute of Information Technology, Vietnam Academy Science and Technology
Supervisor 2: Prof. Dr. Phung Trung Nghia, Thai Nguyen University of Information and Communication Technology

Referee 1:

Referee 2:

Referee 3:

The dissertation will be examined by Examination Board of Graduate University of Science and Technology, Vietnam Academy of Science and Technology, at ………..……………………….

This dissertation could be found at:
1. The National Library of Vietnam
2. The Library of Graduate University of Science and Technology

INTRODUCTION

1. The importance of the dissertation

Sign Language (SL) is the official language of the Vietnamese deaf community. Translation of sign language involves the process of converting SL to and from regular language. Specifically, translating from regular language to SL is a crucial task with significant implications for conveying information and providing social knowledge to the deaf community.

The process of translating regular language into SL consists of three steps:

Step 1: Speech recognition into text

Step 2: Processing from regular text to correct syntax in sign language

Step 3: Simulating from correctly formatted text in sign language to visual representations

Step 2 is the most critical in this process as it determines the conveyed message. However, it is also the most significant challenge due to the limited vocabulary of sign language compared to spoken/written language. If machine translation is not executed well, information may not be successfully conveyed, or in some cases, the transmitted message may differ in meaning from the original.

Global research on sign language translation commonly employs both traditional and modern methods. Traditional methods utilize grammatical rules to convert from regular language to SL. Modern methods, on the other hand, rely on Deep Learning to automatically learn the features of sign language from input data.

The dissertation, "Translation of Vietnamese Sign Language," focuses on the problem of translating regular text language into correctly structured syntax in Vietnamese Sign Language (text-to-text). The dissertation proposes the deployment, suggestion, and improvement of machine translation models for translating regular text language into correctly structured syntax in Vietnamese Sign Language. Simultaneously, the dissertation also constructs vocabularies, VSL dictionaries, and parallel datasets through testing and developing methods to enrich data for the problem.

Sign language translation is a challenging task but holds significant importance for the deaf community. The dissertation, "Translation of Vietnamese Sign Language," is a crucial study contributing to enhancing the quality of Vietnamese Sign Language translation, enabling the deaf to access information and social knowledge comprehensively and accurately.

2. Objectives of the dissertation

The primary aim of this dissertation is to address a problem of both scientific and practical significance. This entails proposing effective models and methods in the field of machine translation to tackle the task of translating Vietnamese into properly structured sign language text in VSL. Following this, the emphasis lies on the experimental process, analysis, and evaluation of the results of the translation task in comparison with the proposed methods, considering sign languages worldwide and specifically VSL.

3. Contributions

The dissertation explores the machine translation problem from Vietnamese sentences into properly structured syntactic sign language in Vietnamese Sign Language. The main contributions of the dissertation include:

1) The dissertation proposes a simple and effective translation approach for the problem using a rule-based translation model. Although a classical method, it proves to be suitable for the given task. This contribution has been published in the works [CT1], [CT2], [CT3].
2) Introduces a data enrichment method based on word embeddings for bilingual data of Vietnamese sentences and syntactically correct sign language sentences in VSL. the dissertation has developed datasets: VSL-Lexicon for Vietnamese Sign Language, and bilingual data sets Vie-VSL10k, Vie-VSL60k, released for the research community to use. This contribution has been published in the work [CT5].
3) Improves a basic statistical translation model and several neural network-based modern translation models for the problem. This contribution has been published in the works [CT4], [CT6].

4. Scope of the dissertation

The scope of this dissertation is specifically focused on machine translation methods addressing the translation challenge from Vietnamese sentences into appropriately structured syntactic sign language in Vietnamese Sign Language. The dissertation does not extend its discussion to include 3D visualization models or other ultimate outputs associated with Vietnamese Sign Language in this contex.

5. Dissertation Structure

The main content of the dissertation is organized into an introduction and four chapters, structured as follows:

Introduction Section: This section provides an overview of the problem of translating symbolic language, with a focus on machine translation methods for translating regular Vietnamese text into syntactically correct symbolic language. The content discusses the significance and necessity of the dissertation, providing an overview of the research context.

Chapter 1: This chapter introduces an overview of the research problem in the dissertation, presenting and analyzing existing issues in domestic and international studies related to the problem of translating symbolic language.

Chapter 2: This chapter introduces some foundational knowledge related to the research content of the dissertation.

Chapter 3: Researches the approach based on structure in automatic translation of Vietnamese sign language, conducts experiments, and evaluates the results of this method.

Chapter 4: Presents a method for enriching data based on network words for the problem at hand.

Chapter 5: Explores some classical statistical machine translation models and modern neural network-based machine translation models in the automatic translation of Vietnamese sign language. It includes experiments and evaluations of the results obtained using these methods.

Finally, the dissertation concludes with a summary of the achievements, highlighting strengths and weaknesses, and outlining future directions for development.

CHAPTER 1
INTRODUCTION TO THE VIETNAMESE SIGN LANGUAGE TRANSLATION

1.1. Overview of sign language

Sign language has evolved early in conjunction with the development of spoken language. The deaf community created sign language as a distinct form of communication and knowledge acquisition for humanity. Unlike spoken language expressed through sound and speech, sign language involves the movement of hands and arms, coupled with facial expressions. Thus, in linguistic terms, it falls under the category of natural languages. However, it is not a body language a non-linguistic form of communication. Sign language exhibits distinctive syntactic features, characterized by brevity, emphasis on focal points, and alterations in syntax order compared to spoken languages.

1.2. Relevant studies

The global issues related to sign language translation are categorized into two problem classes. One is the translation from regular language to sign language, and the other is the reverse translation, meaning from sign language to written or spoken language in the conventional sense.

However, this dissertation specifically examines studies related to the translation of text/speech to sign language. This is because it is a problem of significant importance for conveying information, providing social knowledge to the deaf community. In this problem, many studies focus on the translation of regular text into syntactically correct Sign Language (SL) text.

In recent years, structure-based translation has continued to find application in some sign language translation problems. Statistical methods are also frequently applied for text-to-text translation in sign language translation tasks, especially when dealing with a limited amount of data. Recent studies leverage advancements in Natural Language Processing (NLP), Deep Neural Networks (DNN), and Machine Translation (MT) to develop systems capable of translating between sign language and spoken language. The goal is to bridge the communication gap between the sign language-speaking community and the community using spoken language.

In summary, one of the significant drawbacks of many aforementioned projects is the limited emphasis on sign language syntax, with insufficient consideration for the unique characteristics of each independent sign language. This results in issues related to language comprehension. Additionally, there is a challenge associated with insufficiently large databases, and notably, evaluations within the deaf community are often overlooked.

1.3. The Vietnamese sign language translation problem

In the realm of VSL translation, a noticeable void exists as there is currently no publicly disseminated database available for the research community. Consequently, this dissertation directs its attention toward a pivotal objective—establishing a robust database for VSL machine translation. The initial aspiration involves crafting a comprehensive VSL lexicon (VSL-lexicon) enriched with annotations, each vocabulary entry intricately linked to a 3D illustrative model. Concurrently, the creation of a "bilingual data" set ensues, comprising pairs of Vietnamese sentences juxtaposed with their syntactically sound equivalents in VSL.

Thus, this dissertation delves into the intricacies of the VSL machine translation problem, accentuating specific concerns related to classical and contemporary machine translation methodologies (structure-based, statistical, and neural network-based machine translation), alongside the meticulous construction of data pertinent to the task.

1.4. Conclusion of this chapter

This chapter has presented an overview of sign language in general and the distinctive syntactic characteristics of Vietnamese sign language. The imperative nature of the sign language translation problem is illuminated through the analysis and evaluation of various research endeavors on sign language translation globally. From these observations, three primary issues are identified for the VSL machine translation problem.

Firstly, there is a consideration of applying purportedly classical machine translation methods, deemed effective and suitable for the VSL translation task. Secondly, the implementation of data enrichment methods is emphasized, a focal point for evaluating and testing translation models. Lastly, a proposition is made for modern statistical machine translation models that align with the VSL translation problem..

CHAPTER 2
FUNDAMENTAL KNOWLEDGE

2.1. Fundamental knowledge on machine translation

Machine Translation (MT), also known as automatic translation, is the process where computer software translates text from a source language into a text in a target language.

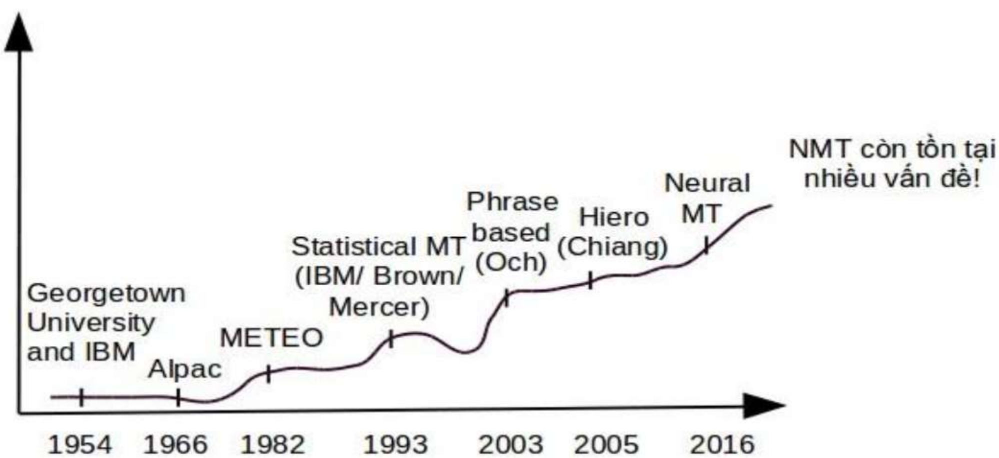


Figure 2.1. Evolution of Machine Translation Process

2.2. Rule-Based Machine Translation

Rule-Based Machine Translation (RBMT) utilizes a set of rules concerning morphology, syntax, and semantics between pairs of source and target languages [33]. Vietnamese and VSL are closely related in terms of syntax. Therefore, translation can be achieved through syntax analysis and various techniques. Figure 2.2 illustrates a rule-based translation system.

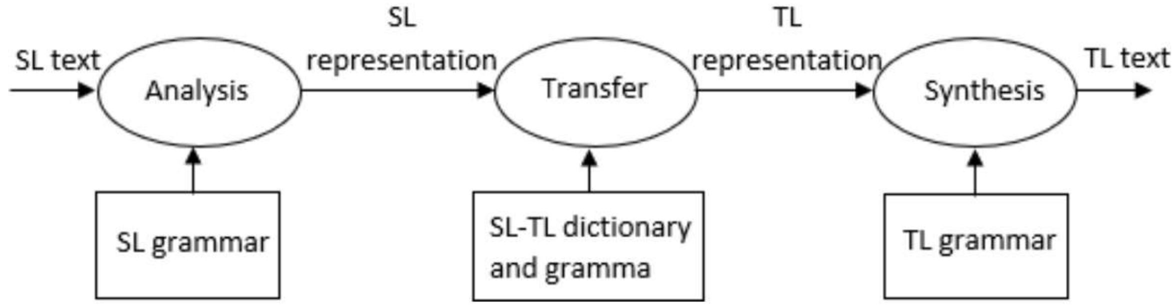


Figure 2.2. Rule-Based Machine Translation Diagram.

2.2. Statistical Machine Translation

The statistical machine translation method was first proposed by Brown in 1993 using the noisy channel model. The problem is formulated as follows:

For a given sentence f in the source language $f \in f^J = \{f_1, f_2, \ldots, f_J\}$, the system needs to translate it into a sentence e in the target language $e \in e^I = \{e_1, e_2, \ldots, e_I\}$. The translation system will select a sentence e with the highest probability among many possible translation options.

$$e* = argmax_e p(e)p(f|e) \tag{2.3}$$

With the formula 2.3, the SMT model is decomposed into two sub-models: the language model $p(e)$ and the translation model $p(f|e)$.

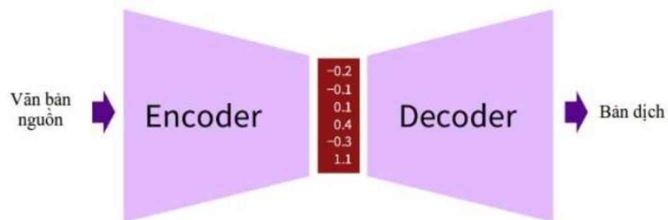
he translation model is the central problem of SMT. In the translation model, the focal point of modeling the translation probability $p(f/e)$ is determining the correspondence between words in the source sentence and words in the target sentence. There are several different methods to model the translation process, categorized into three main approaches: word-based translation, phrase-based translation, and syntax-based.

2.3. Neural Network-Based Machine Translation

Recurrent Neural Networks (RNN), proposed by Elman in 1990, is an architecture that allows processing input sequences and computing outputs through hidden states. RNNs have been successfully applied to language modeling in recent studies by Mikolov and colleagues [38]. In machine translation, RNNs take a sequence of input vectors, updating their memory and generating hidden states at each time step through a recursive expression.

2.3.1. Mô hình Sequence to Sequence

With the successful application of RNNs to language modeling, researchers have proposed the sequence-to-sequence model (commonly referred to as seq2seq), based on the encoder-decoder architecture with RNNs as the central components. The encoder-decoder architecture is illustrated in Figure 2.6



:

Figure 2.6. Encoder-Decoder Architecture Using RNNs

In the seq2seq model, the RNNs of the Encoder and Decoder operate concurrently during the training process.

2.3.2. Transformer Model

The Transformer model is an artificial neural network architecture introduced in the paper "Attention Is All You Need" by Vaswani and colleagues in 2017. This model has become one of the most significant architectures in the field of natural language processing.

The Transformer model employs an encoder-decoder structure to perform natural language processing tasks. It consists of two main parts: the encoder and the decoder, each comprising multiple identical layers.

Each layer of the Transformer model consists of two main components: a self-attention layer and a feed-forward layer. The self-attention layer allows the model to attend to different parts of the input sentence during encoding and decoding. The feed-forward layer then applies a non-linear function to compute the output. The overall architecture of the Transformer model is formed by the connection of multiple encoder and decoder layers. During the encoding process, each layer of the encoder takes an input sequence of words and produces a sequence of hidden states. During decoding, each layer of the decoder takes an input sequence in the target language and the hidden state computed by the preceding layer of the decoder

2.4. Evaluation Metric for Machine Translation

The most common automatic evaluation method is BLEU (Bilingual Evaluation Understudy). BLEU calculates the similarity between machine-generated translations and the reference translation by comparing n-grams (subsequences of n items, usually words) in the two texts. Similarity is evaluated by calculating the ratio of the number of matching n-grams in the translation to the reference.

$$score = exp\left\{\sum_{i=1}^{N} w_i \, log(p_i) - \max\left(\frac{L_{\text{ref}}}{L_{\text{tra}}} - 1,0\right)\right\}$$

- $P_i = \frac{\sum_j NR_j}{\sum_j NT_j}$
- NR_j: is the number of n-grams in the reference segment j
- jNT_j: is the number of n-grams in the translation segment j
- $w_i = N^{-1}$
- L_{ref}: is the number of words in the reference translation, and its length is typically close to the length of the machine-generated translation..
- L_{tra}: is the number of words in the machine-generated translation.

One advantage of BLEU is its simplicity and fast computation, allowing for a quick evaluation of machine translation system quality. Therefore, the dissertation has chosen the BLEU evaluation method for the translation task.

2.5. Conclusion of the Chapter

Chapter 2 presents the foundational knowledge used in this dissertation. The content includes fundamental concepts of machine translation, classical and modern machine translation models for SIGN LANGUAGE translation. This encompasses rule-based machine translation, the IBM statistical machine translation model, and neural network-based translation models (Seq2Seq and Transformer). Foundational knowledge about evaluating machine translations is also covered, with a specific focus on detailing the calculation formula for BLEU score – the quality evaluation metric used in this dissertation.

CHAPTER 3
STRUCTURE-BASED APPROACH IN AUTOMATIC SIGN LANGUAGE TRANSLATION FOR VIETNAMESE SIGN LANGUAGE

3.1. Database Construction

3.1.1. VSL-Lexicon

In the VSL-Lexicon dataset, lexical units are stored along with accompanying information such as part of speech, annotation codes, synonyms, and corresponding illustrative models. Table 3.1 describes the structure of the VSL-Lexicon dataset.

Table 3.1. Description Table of VSL-Lexicon Lexical Dataset

No.	Lexical Unit	Lexical category	Synonyms	Tag code	Corresponding 3D animation model
1	a	Alphabet		VSL0001	M3D0001.FBX
2	ă	Alphabet		VSL0002	M3D0002.FBX
153	tôi	Pronoun (P)	tao, tớ	VSL0153	M3D0153.FBX
154	họ	Pronoun (P)		VSL0154	M3D0154.FBX
296	chết	Verb (V)	hi sinh,..	VSL0296	M3D0296.FBX
3035	trường học	Noun (N)		VSL3035	M3D3035.FBX
3036	nhà	Noun (N)		VSL3036	M3D3036.FBX
6176	xương rồng	Alphabet		VSL6176	Not available in the database

3.1.2. The bilingual dataset Vie-VSL10k

The constructed dataset is named Vie-VSL10k. This dataset was semi-manually built with 10,000 sentence pairs in everyday communication domain. The collected data underwent partial automatic processing through various text summarization algorithms and rough syntax conversion. Subsequently, it was re-evaluated by language experts. The final dataset for the dissertation comprises 10,000 bilingual sentence pairs in Vietnamese and

VSL (Vietnamese Sign Language), serving the development of a rule-based translation system with a vocabulary of 4,626 lexical units. Statistics regarding the Vie-VSL-10k database are presented in Table 3.2.

Table 3.2. Statistical Figures for Vietnamese Sentences in Vie-VSL-10k Dataset

No.	Lexical category	Sign	Quantity in Vietnamese sentence	Quantity in VSL sentence
1	Noun	N	16182	16182
2	Proper noun	Np	7030	7030
3	Common noun	Nc	1069	1069
4	Countable noun	Nu	172	172
5	Verb	V	15528	13559
6	Adjective	A	4241	4241
7	Pronoun	P	3424	3424
8	Determiner	L	537	0
9	Numeral	M	1560	1560
10	Adverb	R	8477	4689
11	Preposition	E	4471	2910
12	Conjunction	C	1480	0
13	Interjection	I	175	0
14	Auxiliary, modal, or mood word	T	878	0
15	Word formation element	S	10	0
16	Unclassifiable words	X	322	0

3.2. The problem of synthesizing rules

With the characteristic syntax features presented in VSL, some simplified syntax features and conversions of sentences in VSL are synthesized for the structure-based translation problem. The dissertation demonstrates the application of parsing tools to the 10,000 bilingual sentence pairs in Vie-VSL for rule extraction. As a result, 8,025 rules are derived from the bilingual data. The table 3.13 describes some of the extracted rules.

From the 8,025 rules extracted from the dataset of 10,000 bilingual sentence pairs, we proceed to construct a rule-based machine translation system. The effectiveness of this translation method is analyzed and evaluated in the following section. Refer to the 8,025 rules at https://github.com/BichDiep/rules-VSL.git

Bảng 3.13. Một số luật trích rút cho hệ thống dịch Rule-based

No.	Vietnamese sentence parsed	Grammar rules	Sign language sentence parsed	Extraction rules
1	SQ (NP (N Bạn) (N tên)) (VP (V là) (WHNP (P gì))) (? ?)	1	SQ (NP (N Bạn) (N tên) (P gì)) (? ?)	SQ (NP (N) (N)) (VP (V) (WHNP (P)) (? ?) → SQ (NP (N) (N) (P)) (? ?)
2	S (NP (P Tôi)) (NP (N tên)) (VP (V là) (NP (Np Hiếu))) (..)	1	S (NP (P Tôi)) (NP (N tên) (Np Hiếu)) (..)	S (NP (P)) (NP (N)) (VP (V) (NP (Np)) (..) → S (NP (P)) (NP (N) (Np)) (..)
3	S (NP (N Khế)) (C thì) (AP (A chua)) (..)	1	S (NP (N Khế)) (AP (A chua)) (..)	S (NP (N)) (C) (AP (A)) (..) → S (NP (N)) (AP (A)) (..)
4	S (NP (N Mít)) (C thì) (AP (A ngọt)) (..)	1	S (NP (N Mít)) (AP (A ngọt)) (..)	
5	S (NP (P Tôi)) (NP (M 19) (N tuổi)) (..)	2	S (NP (P Tôi)) (NP (N tuổi) (M 19)) (..)	S (NP (P)) (NP (M) (N))) (..) → S (NP (P)) (NP (N) (M)) (..)
..	…	..	…	…

3.3. Building a Rule-Based Translation System

The algorithm for rule syndissertation and the rule-based machine translation system is described based on pseudocode as follows:

Algorithm: Rule-based-MT-VSL

Input: Sentence S in Vietnamese,
Output: Sentence S' in the syntax of VSL.

```
1. R is set of syntax conversion rules
2. WD = Ø; (WD: Waiting Dataset)
3. SYN is Synonyms files with n line: SYN[n,1] in VSL dictiary;
   SYN[n,i] is a synonym of SYN[n,1]; (i=1:m).
4. Si ⇐ Tokenization(S)
5. While ∃ Si in SYN:
          Si = SYN[n,1]
6. (TS, PS)⇐ Parsing (S)
7.  If (Find Ps in R)
            ST= Transform(TS)
      Else
            Add S to WD
8. S'= Shorten(ST)
9. Return S'
```

3.4. Experiments and Evaluation of the Rule-Based Translation System

To evaluate the effectiveness of the rule-based translation method for the Vie-VSL translation task, the dissertation conducts assessments on prepared test sets. The BLEU score evaluates the translation quality in automatic Vie-VSL translation for datasets as follows.

Table 3.16. Aggregate BLEU scores for the Rule-Based translation system on selected test sets

Dataset	BLEU Score
Data set 1: The domain of sentences in communication	81.15
Data set 2: The domain of sentences in literature	48.68
Data set 3: The domain of sentences in technical context	64.13
Data set 4: The domain of sentences in medical context	55.72
Average	62.55

Overall, the BLEU scores on the test sets are superior compared to BLEU scores for some other languages. This is because, in the dissertation problem, the translation model remains almost unchanged for the majority of language units between the two languages. Only a few words not present in sign language are replaced with synonyms. Regarding sentence structure, VSL mostly consists of simple sentence patterns, lacking the diversity observed in other language pairs.

3.5. Conclusion of this Chapter

In this chapter, the dissertation presents a method to address the translation problem using a rule-based model. To implement this model, it is essential to have resources such as the VSL dictionary and grammar rule data. The machine translation model will utilize grammar rules to analyze and translate sentences. These rules are predefined and structured according to Vie-VSL conversion rules. The achieved results in this section include: the VSL-Lexicon dictionary database with components and features distinct from conventional dictionaries, the Vie-VSL-10k bilingual database consisting of 10,000 pairs of Vietnamese sentences and syntactically correct VSL rule sentences for constructing the syntax rules of the rule-based translation model.

Finally, a simple and effective rule-based translation model for the problem is presented. The BLEU evaluation score for translation quality reaches 62.55 with the analyzed features. Relevant publications related to this section include [CT1], [CT2], [CT3].

CHAPTER 4
AUGMENTING DATA FOR THE AUTOMATIC TRANSLATION OF VIETNAMESE SIGN LANGUAGE

4.1. Data augmentation background

Utilizing the semantic relationships between words in Wordnet, the idea behind data augmentation involves replacing words in a sentence to generate new data. The newly generated sentence maintains its syntactic structure and logical semantic coherence. To translate it into Vietnamese Sign Language (VSL), the conversion rules are preserved. This ensures that the translation is accurate and maintains semantic fidelity, as evaluated by similarity metrics in the experimental section.

In our problem, we use 3 criteria:

Sibling criterion:

That is: $SV = \{S_i^{jk} / S_i^{jk} \in S_i^j (\forall j: 0 \le j \le n_i^j): \exists S_p: (S_P\ is\_hyper\ S_i^{jk})\}$

Parent-child criterion:

$$SV = \{S_i^{jk} / \exists S_p \in S_i^h\ (h \in [1 \dots n_i^j])\ S_i^{jk} \in S_i^h (\forall j: 0 \le j \le n_i^j, \quad j \ne h): (S_P\ is\_hyper\ S_i^{jk})\}$$

Grandparent- grandchildren criterion:

$$SV = \{S_i^{jk} / \exists S_g \in S_i^h\ (h \in [1 \dots n_i^j]), S_i^{jk} \in S_i^j (\forall j: 0 \le j \le n_i^j, j \ne h)(S_g\ is\_dist\_hyper\ S_i^{jk})\}$$

Thus, when the word W appears in a phrase, W can be replaced with W' if W and W' satisfy the sibling, parent-child, and grandparent-grandchild criteria. Therefore, depending on the structure of the hypernyms and hyponyms and other characteristics of wordnet, we may construct fuzzy data by changing words in previous phrases according to predetermined criteria. The data augmentation algorithm is described by pseudocode as follows:

Algorithm: Data-Augment-VSL
Input: Sentences S ***Output***: Set of sentences S′ are generated based on S.
1: Split W word ∈ S 2: X ⇐ W. hypernyms() n = len(X); 3: For i=1,n do Xi⇐X.hyponyms() Add Xi to set T 4: While !∃ Xi.hyponyms: Yi ⇐ Xi.hyponyms() Add Yi to set T 5: S′= Replace(W,Ti)

4.3. Evaluation of results.

To evaluate the experiments, the dissertation relies on criteria related to the degree of data enrichment and data similarity for analysis. Firstly, regarding the degree of data enrichment, the dissertation addresses the issue of the number of sets T constructed from the data augmentation algorithm and examines certain semantic aspects of the newly generated sentences from the original data. For the group of vocabulary consisting of verbs, these experiments yield unreasonable results in terms of semantics in Vietnamese sentences.

After experimenting with some data, it was observed that for verbs when using the method of searching for words with hyponyms based on sibling, parent-child, and grandparent-grandchild standards, the results were not appropriate in terms of semantics. Therefore, the examination focuses only on the unit word groups, including pronouns, nouns, and adjectives.

Table 4.1 presents several sets T and summarizes the number of sentences enriched from the proposed algorithm (where T is the set of words with the same hyponym as the applied standards for each word type group, WS is the number of original data sentences containing 1 word from the examined word type group, W'S is the number of sentences enriched from all original sentences containing 1 word from the examined word type group).

Table 4.1. Results of the data augmentation algorithm from Vie-VSL10k

Lexical category	Group	Example	T	W_S	W'_S
Noun	Plant 1 (fruits)	Bưởi, cam, nho, táo,.. (Pomelo, orange, grape, apple, etc.)	92	35	3220
	Plant 2 (flowers)	Hoa cúc, hoa hồng, hoa ly,… (Chrysanthemum, rose, lily, etc.)	183	5	915
	Plant 3 (general)	Cây, hoa, cỏ, lá, rau,… (Tree, flower, grass, leaf, vegetable, etc.)	438	10	2628
	Food	Bánh, kẹo, bia, thịt, rau… (Cake, candy, beer, meat, vegetable, etc.)	471	3	1413
	Animal 1 (pets)	chó, chó con, chó xù, gà, mèo,… (Dog, puppy, poodle, chicken, cat, etc.)	25	5	125
	Animal 2 (others)	Báo, hổ, hươu,.. (Tiger, lion, giraffe, etc.)	708	3	2124
	Object 1 (household items)	Bàn, ghế, tủ,.. (Table, chair, cabinet, etc.)	257	11	2827
	Object 2 (tools)	Buá, kéo, máy,.. (Hammer, scissors, machine, etc.)	1564	4	5056
	Object 3 (vehicles)	Xe máy, ô tô, xe chở hàng, .. (Motorcycle, car, truck, etc.)	78	7	546
	Weather	Nắng, mưa, gió,.. (Sun, rain, wind, etc.)	63	5	315
	Occupation	Giáo viên, công nhân,… (Teacher, worker, etc.)	21	8	168
	Body parts	Chân, tay, tóc, má, môi,… (Leg, arm, hair, cheek, lips, etc.)	231	4	924
	Geometric shapes	Tam giác, hình tròn, hình vuông,… (Triangle, circle, square, etc.)	134	3	402
Adjective	Color	Đỏ, xanh, vàng, tím,… (Red, green, yellow, purple, etc.)	12	36	432
	Material property	Nặng, nhẹ, Cứng, mềm,… (Heavy, light, hard, soft, etc.)	45	2	90
	Size	To, rộng, dài, ngắn,…	15	4	60

		(Big, wide, long, short, etc.)			
	Emotions	vui, buồn, lo lắng,… (Happy, sad, worried, etc.)	279	7	1953
	Personality	hài hước, cục cằn, dễ thương… (Funny, grumpy, adorable, etc.)	23	4	92
Pronoun		Tôi, họ, chúng ta, .. (I, they, we, etc.)	12	3424	41088
Total:					64378

The similarity of the language dataset before and after augmentation can be evaluated based on the perplexity of each type. Perplexity is a measure used in probability and statistics to assess the effectiveness of a language model. In the n-gram language model, perplexity measures the model's predictive ability on a new piece of text based on the probability of n-gram sequences in the model. Table 4.3 presents the perplexity indices for the language datasets constructed with a 3-gram language model

Table 4.3. Perplexity scores for the constructed language corpora

Data corpus	**The average perplexity index within the range**
Vie10k	P_1= 420
VSL10k	P_2= 300
Vie60K	P_1’= 520
VSL60K	P_2’= 450

Thus, the size is more than 6 times larger than the original data, but the perplexity score is not more than 1.5 times higher. This indicates that the language dataset with a 3-gram model performs well. There is high similarity between the original sentences and the newly generated ones as the syntactic structure remains unchanged. In terms of semantics, the similarity is ensured by the nature of words sharing the same hyponyms based on the applied standards. The BLEU score is also a criterion for comparing the effectiveness of translation models, which will be presented in Chapter 5.

4.4. Conclusion of the Chapter

In this chapter, the dissertation presents the construction of two datasets, Vie-VSL-10k and Vie-VSL-60k, consisting of bilingual pairs of Vietnamese sentences and their syntactically correct counterparts in VSL. The Vie-VSL-60k dataset is the result of a data augmentation method applied to the base dataset Vie-VSL-10k. The proposed idea for data augmentation relies on the hypernym and hyponym structure of the WordNet and utilizes the Vietnamese WordNet database. The data augmentation algorithm helps generate bilingual Vie-VSL sentence pairs based on the original dataset comprising 10,000 Vie-VSL bilingual sentences.

From the analysis and evaluation of the data augmentation algorithm's experimental process, it is observed that the augmented dataset exhibits high similarity to the original data, as it preserves the syntactic structure of the original sentences. Moreover, the newly generated sentences, produced by replacing words, adhere to standards based on the properties of WordNet, ensuring semantic relevance. The Vie-VSL-60k dataset is used for the dissertation's experimental evaluations with various statistical machine translation models and neural machine translation models in the following chapters.

CHAPTER 5

APPROACH BASED ON STATISTICS AND NEURAL NETWORKS IN AUTOMATIC TRANSLATION OF VIETNAMESE SIGN LANGUAGE

5.1. Improving the IBM translation model for the vie-vsl translation task

In this section, the dissertation introduces a simple statistical translation model for sign language machine translation based on word translation. This method requires a dictionary mapping words from the source language to the target language. In the Vie-VSL translation task, this word mapping dictionary is much simpler than in translation tasks between other languages such as English-Vietnamese, Vietnamese-Chinese, or Vietnamese-Japanese. This simplicity arises because all words are mapped one-to-one. The dissertation addresses the use of statistical data based on the number of words in the text corpus or bilingual texts. It is necessary to estimate the probability distribution of word translation. This function will return a probability for each VSL translation option, indicating the likelihood of that translation

$$P_f = e \rightarrow P_f(e)$$

By leveraging the probability distribution for word translation, we can proceed to the first model using only the probability of word translation. We represent the probability of translating a Vietnamese word f to a VSL word e with the conditional probability function *t(e|f)*. The association between input words and output words can be illustrated by the diagram:

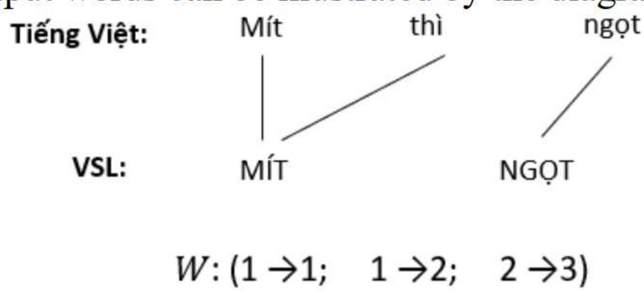


Figure 5.1. Linkages between input words and output words in Vie-VSL sentence translation

The mapping function in the example above, each output word in VSL at position i corresponds to a Vietnamese input word at position j: w: ($j \rightarrow i$)

This is a very straightforward alignment method since the corresponding Vietnamese words and VSL words do not follow the exact same order. This means that the words must be rearranged during the translation process, as illustrated in the following example:

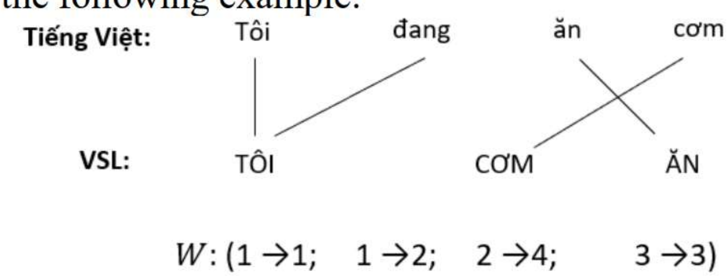


Figure 5.2. An example illustrating the rearrangement of words in Vie-VSL sentence translation.

With a word-based alignment model, each output can be associated with one or multiple input words, as determined by the alignment function. The IBM model can be implemented to align words based on the probability of word translation. There are 3 IBM models to map words from the source language Vie to the target language VSL, with an improved algorithm based on phrase alignment for the Vie-VSL translation task.

IBM1 model defines the translation probability for a Vietnamese sentence $f = (f_1, \dots, f_{l_f})$ to a VSL sentence $e = (e_1, \dots, e_{l_e})$ of length l_e with the alignment of each word e_j to an English word f_i according to the alignment function w: $(j \rightarrow i)$ as follows:

$$p(e, w|f) = \frac{\varepsilon}{(l_f + 1)^{l_e}} \prod_{j=1}^{l_e} t(\, e_j |\, f_{w(j)})$$

Considering the algorithm on a small portion of the Vie-VSL-10k dataset with 3 Vietnamese words as input: "tôi" (I), "ăn" (eat), "cơm" (rice), and 3 words in VSL as output: "TÔI" (I), "ĂN" (eat), "CƠM" (rice).

EM Optimization in the IBM-1 Model.

From the formula: $$p(w|e, f) = \frac{p(e, w|f)}{p(e|f)}$$

We can improve the result by adding d_w between *e* and *f*, resulting in:

$$p(w|e, f) = \frac{\alpha . p(e, w|f) + (1 - \alpha) . d_w(e, f)}{p(e|f)}$$

Where α is the similarity coefficient between two words e and f. The standard value for α used in the experiments is 0.5. Table 5.4 presents the comparison results applied to a small dataset consisting of two corresponding sentence pairs.

Table 5.4. Translation probability results with optimized IBM-1 model.

e	f	Initial	Iteration 3	Iteration 3 and phrase alignment
TÔI	tôi	0.33	0.75	0.96
TÔI	ăn	0.33	0.21	0.03
TÔI	cơm	0.33	0.21	0.02
ĂN	tôi	0.33	0.04	0.01
ĂN	ăn	0.33	0.42	0.77
ĂN	cơm	0.33	0.42	0.32

CƠM	tôi	0.33	0.04	0.01
CƠM	ăn	0.33	0.42	0.28
CƠM	cơm	0.33	0.77	0.95

Therefore, for the statistical model in this problem, the phrase alignment technique and phrase-based translation are suitable. The dissertation utilizes the Vie-VSL-10K and Vie-VSL-60K language datasets for experimenting with the proposed statistical translation model.

5.2. Sequence to Sequence Model

5.2.1. Model of Encoder and Decoder

The constants for the model are: embedding_dim = 256; units = 1024. We begin by constructing the encoder.

Encoder:

Retrieve a list of token ID codes (from input_text_processor).

Look up an Embedding vector for each token code (Using an Embedding technique).

Process the "embeddings" into a new sequence.

Result:

- The processed sequence - will be passed to the attention head.
- Internal state - will be used to initialize the decoder.

We use the Vie-VSL-10K and Vie-VSL-60k data for the translation model with the specified settings and processes. Then, the experimental evaluation figures are analyzed and compared in Section 4.4. The Seq2Seq model for VSL translation is published on GitHub athttps://github.com/BichDiep/Seq2seq-VSL.

5.3. Mô hình Transformer cho bài toán dịch

- Hyperparameter settings: The base model used is:

num_layers = 6, d_model = 512, dff = 2048.

- Optimizer: Use the Adam optimizer with a customized learning rate schedule (The Adam optimization algorithm is an extension of stochastic gradient descent that has recently been widely applied to deep learning applications in computer vision and natural language processing).

- Training and Testing:

After each training step, checkpoints are saved by creating a checkpoint path and using a checkpoint manager.

The input for the problem is a regular Vietnamese sentence, and the correct syntax in Vietnamese sign language is the output. To perform inference, we need to follow these steps:

Step 1: The encoder processes the input sentences in standard Vietnamese and uses the Vietnamese encoding tokenizer (tokenizers.Vie). The encoding outputs are utilized as inputs.

Step 2: Subsequently, these values are initialized into the token (START).

Step 3 involves the computation of padding masks and look ahead masks.

Step 4: The decoder produces predictions based on examining the output of the encoder and its own output (self-attention mechanism).

Concatenate the predicted token with the input of the decoder and feed it back into the decoder. In this approach, the decoder predicts the next token based on its previously predicted tokens.

5.4. Experiments and Evaluation Results

The evaluation of the proposed approaches is based on BLEU scores, comparing the enriched dataset with the original dataset across various machine translation models. In these experiments, translation performance is assessed using BLEU scores.

Table 5.5. Comparison of BLEU scores across several translation models between the original dataset and the enriched dataset.

No.	Translation model	Original data	Augmented data
1	Rule-based translation	68.02	68.02
2	IBM model	42.31	60.32
3	The improved IBM model	48.75	76.25
4	Seq2Seq	58.53	81.44
4	Transformer	65.22	89.23

Through the experimentation with various models as described above, we observe that with a training dataset of 10,000 sentence pairs, rule-based translation yields higher BLEU scores compared to statistical models. However, with a larger dataset, statistical models show superior and increasing results. Among the statistical models used in our research, the Transformer model currently provides the best results.

Referring to the achieved results of the dissertation in comparison with some sign language translation studies in other languages, we observe that BLEU scores in the applied models for the Vie-VSL problem are significantly superior to machine translation models for other language pairs. Therefore, we find that the Transformer model provides excellent translation results for Vietnamese sign language within the scope of this problem. The BLEU score, which evaluates the translation quality, is very high for the reasons analyzed. Specifically, this is due to the convergence of the language model; the translation model remains almost unchanged for most language units that are the same in both languages.

5.4. Conclusion of this Chapter

Chapter 5 has presented several statistical models and improvements applied to the translation problem in the field of information technology. Specifically, the IBM translation model with enhancements based on phrase-based translation and the addition of a scaling factor along with sequence alignment techniques were discussed. Through experiments conducted on a small dataset and the entire dataset, the proposed translation model demonstrated significant improvements compared to the baseline. Additionally, the enriched data from the algorithm presented in Chapter 3 was utilized as test data for some modern neural machine translation models: Seq2Seq and Transformer. Finally, an analysis and evaluation of the datasets were performed using the proposed translation models. With the models proposed for the translation problem, it is observed that the Transformer model yields the best translation results for translating Vietnamese symbolic language within the scope of this study.

CONCLUSION AND RESEARCH DIRECTIONS

Automatic translation of Vietnamese Sign Language (VSL) is a challenging problem for researchers and developers in the field of natural language processing. Vietnamese Sign Language is a distinct language with its own syntax structure compared to spoken/written language.

This dissertation focuses on the translation problem from Vietnamese to VSL. The process of translating Vietnamese text into grammatically correct sentences in VSL is one of the most crucial aspects of this task.

The dissertation has achieved several key results:

- Proposed a simple and effective translation approach for the problem using rule-based translation models.
- Proposed a data enrichment method based on word embeddings for Vietnamese-VSL bilingual sentence data. This led to the creation of datasets: VSL-Lexicon dictionary, bilingual data sets "Vie-VSL10k," "Vie-VSL60k."
- Improved a basic statistical translation model and some modern neural network-based models for the task.

These results have contributed to enhancing the quality of automatic Vietnamese Sign Language translation, enabling the deaf community to access information and social knowledge more comprehensively and accurately. In the future, further research will focus on proposing new models and methods to continue improving automatic sign language translation. Additionally, there is a need to develop more optimized models for machine translation tasks, especially for languages with limited resources. These goals will contribute to building more comprehensive translation systems, helping the deaf community interact and integrate effectively within society.

LIST OF THE PUBLICATIONS RELATED TO THE DISSERTATION

[CT1]. Diep Nguyen Thi Bich, Trung-Nghia Phung, Thang Vu Tat and Lam Phi Tung, "*Special Characters of Vietnamese Sign Language Recognition System Based on Virtual Reality Glove*", the International Conference on Advances in Information and Communication Technology – ICTA, 2016.

[CT2]. Thi Bich Diep Nguyen and Trung-Nghia Phung, "*Some issues on syntax transformation in Vietnamese sign language translation". Sign Language Studies*. IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.5, pp 292-297, 2017.

[CT3]. Thi Bich Diep Nguyen, Trung-Nghia Phung, Tat-Thang Vu , "*A rule-based method for text shortening in Vietnamese sign language translation".* Springer AISC, Vol. 672, Proc. of INDIA-2017, Vietnam, 2017.

[CT4]. Nguyen Thi Bich Diep, "*Using the Transformer Machine Translation Model for Automatic Vietnamese Sign Language Translation*", Proceedings of the 24th National Conference on Information and Communication Technology, Vietnam, 2021.

[CT5]. Diep Nguyen Thi Bich, Tuyen Ho Thi, "*Data Augmentation Techniques in Automatic Translation of Vietnamese Sign Language for the Deaf*", International Conference on the Development of Biomedical Engineering - BME9, 2022.

[CT6]. Thi Bich Diep Nguyen, Trung-Nghia Phung, Tat-Thang Vu, *A Study of Data Augmentation and Accuracy Improvement in Machine translation for Vietnamese sign language,* Journal of Computer Science and Cybernetics, Vol 39, N2, pp 143-158, 2023.