Neural Machine Translation between Myanmar and Korean Languages

HNIN NANDAR ZAW

M.C.Sc. **DECEMBER 2022**

Neural Machine Translation between Myanmar and Korean Languages

By

HNIN NANDAR ZAW

B.C.Sc.

A Dissertation Submitted in Partial Fulfillment of the

Requirement for the Degree of

Master of Computer Science

M.C.Sc.

University of Computer Studies, Yangon

DECEMBER 2022

ACKNOWLEDGEMENTS

To complete this thesis, many things are needed like my hard work and the supporting of many people who gave a lot of idea to me during my study period.

First of all, I would like to thank the Union Minister, the Ministry of Science and Technology, for allowing me the advanced study and providing full facilities during the master degree course at the University of Computer Studies, Yangon.

I would like to express very special thanks to **Dr. Mie Mie Khin**, Rector of the University of Computer Studies, Yangon, for her kind permission to submit this dissertation.

I would like to express very special thanks to **Dr. Yuzana,** Pro-Rector of the University of Computer Studies, Pyay, for her kind permission to submit this dissertation.

I specially thank the external examiner, **Dr. Aung Nway Oo**, Professor and Head of Faculty of Computer Science, University of Information Technology, for his patience in critical reading, valuable suggestions and comments in the preparation of thesis.

I would like to express my deepest gratitude to my thesis supervisor, **Dr. Khin Mar Soe**, Professor and Head of Faculty of Computer Science, University of Computer Studies, Yangon, for her excellent guidance, valuable suggestions, patience, and providing me with excellent ideas for this thesis.

I would especially want to thank **Dr. Si Si Mar Win**, Professor at the University of Computer Studies in Yangon, and **Dr. Tin Zar Thaw**, Professor, University of Computer Studies, Yangon for their superior suggestions and administrative supports during my academic study.

I would like to express my sincere gratitude to my thesis co-supervisor, **Dr. Yi Mon Shwe Sin**, Lecturer, University of Computer Studies, Yangon, for her close supervision, proper guidance, valuable suggestions, advice and encouragement during the course of this work.

I would like to sincerely thank **Daw Aye Aye Khine,** Associate Professor, Department of English, for her valuable supports from the language point of view and pointed out the correct usage in my dissertation.

I also like to acknowledge my thanks to **all my teachers** who taught me throughout the bachelor degree and master's degree courses and **my friends** for providing support and friendship that I needed. And I wish to express my gratitude to **my beloved parents, my elder sister and my younger sister** for their endless love, invaluable support and encouragement to fulfill my wish.

If not for the above-mentioned people, my thesis would never have been completed in such a successfully manner. I once again extend my sincere thanks to all of them.

ABSTRACT

No matter where they are in the world, people or individuals may now easily and cheaply communicate with one another because of the Internet. In addition to maintaining social networks and relationships, people can discover and exchange ideas. In addition, Natural Language Processing (NLP), a branch of Artificial Intelligence and Linguistics, investigates issues related to the automated generation and comprehension of natural human languages. It also aims to provide computers with the ability to comprehend instructions given to them in commonly used human languages. And then, NLP is an effort to allow users to communicate with computer in a natural language.

However, communication with people from other countries continues to be restricted by language barriers, which are still a significant barrier. Consequently, there are increasingly more difficulties in translating from one language to another. The researchers are investigating machine translation to help users quickly translate from one language to several languages in order to solve these challenges. As a result, machine translation is growing in popularity among researchers and helps to reduce linguistic barriers. Artificial intelligence (AI) is employed in machine translation to automatically translate text between languages without the assistance of human interpreters.

Even though there have only been a few studies on machine translation systems for translating from Myanmar to another language, there are still some difficulties in the early stages due to a lack of resources and a small number of publicly available data corpus.

The aim of this paper is to develop a neural machine translation system implementation for the languages of Myanmar and Korean. There

are two primary sections to this performance. The creation of a new Myanmar-Korean parallel corpus is the early part. The attention-based neural machine translation system for the Myanmar-Korean language pair is introduced in the second. The baseline system for the proposed model's experiments is a word-based neural machine translation model. BLEU (Bilingual Evaluation Understudy) score is used to evaluate on the Myanmar-Korean translation results.

TABLE OF CONTENTS

Pages

LIST OF FIGURES

Pages

LIST OF TABLES

Pages

LIST OF EQATIONS

Pages

CHAPTER 1
INTRODUCTION

Language translation is becoming more widespread and diverse as globalization and the advance in information technology. One of the key functions of a machine translation system in natural language processing (NLP) is to translate between languages. The field of NLP research has a particular focus in creating high-quality machine translation systems. Artificial intelligence is used in machine translation to translate text from one language to another without the need for a human translator. In machine translation, words or phrases from one language are substituted for words or phrases in another language using a machine translation engine. There are currently some machine translation approaches on the market. The most widely used being approaches are Statistical Machine Translation (SMT), Rule-Based Machine Translation (RBMT), Hybrid Systems, which combine RBMT and SMT and Neural Machine Translation (NMT). Automating translation may be accomplished using neural machine translation (NMT), a type of end-to-end learning. Recently, neural machine translation (NMT) has gained popularity as a successful end-to-end statistical machine translation strategy.

In contrast to traditional machine translation, which uses a set of predefined rules from the beginning, neural machine translation uses the neural network of the program to encode and decode the source text. The application of machines to translate natural languages has gained popularity in recent years. Neural Machine Translation (NMT) makes an effort to create and train a single, sizable neural network that can read an input sentence and produce a correctly translated sentence. The translation quality of several language pairings has significantly improved using NMT and the attention-based encoder-decoder system. Both neural machine translation (NMT) and statistical machine translation (SMT) systems rely on sizable parallel data corpora for model training. Additionally, the size of the parallel data corpus has a significant impact on the effectiveness of neural machine translation systems.

On the other hand, Myanmar language is one of the low resource languages and Myanmar-Korean parallel corpus is rare. Therefore, in this system, a parallel corpus for Myanmar and Korean is firstly constructed, and then a neural machine translation system between the two languages is suggested. In order to create the Myanmar-Korean

parallel corpus, Myanmar sentences from the UCSY-corpus [23], which comprises of the Myanmar-English language pair, are used. These Myanmar phrases are then manually translated into Korean. Additionally, this corpus includes parallel sentences from the spoken and written text books for school in both languages. More than 37K parallel sentences can be found in the Myanmar-Korean parallel corpus.

This paper intends to develop the implementation for Neural Machine Translation System between Myanmar and Korean language.

1.1 Neural Machine Translation

Because of its effectiveness in localization and language translation, neural machine translation (NMT) has attracted a lot of attention recently. The neural machine translation method, which uses a single neural network, has gained popularity and is showing promising results in various languages. A huge neural network is trained end-to-end for neural machine translation, which translates one language into another. Neural machine translation is the process of translating new text using data extracted from existing corpora. Nowadays, neural machine translation systems have succeeded in almost all language pairs, and the field is developing quickly. Neural networks, which can handle very huge datasets and require little supervision, are used in neural machine translation to convert source text to target text. Encoder and decoder networks are the two basic components of neural machine translation systems. They are both neural networks.

1.2 Motivation of the Thesis

The number of languages used worldwide in the twenty-first century ranges from 6000 to 8000. This happens as a result of changes in the number of speakers, how frequently they appear in various media, migration, and linguistic mixing with other languages. Learning a foreign language is a difficult task to grasp and difficult to learn. In order to address this issue, machine translation of natural languages has gained popularity in the real world from the late twentieth century. Myanmar (Burmese) is generally regarded as one of the low-resource languages. Additionally, it is rare to find a parallel corpus for Myanmar-Korean translation. As a result, it is necessary to build a new parallel corpus for Myanmar and Korean language. Additionally, the system

described in the paper uses the constructed corpus to translate sentences between Myanmar and Korean employing neural machine translation and an attention-based model.

1.3 Objectives of the Thesis

The proposed Myanmar-Korean machine translation system is capable of translating texts in either way at the request of the user. The main objectives of the thesis are as follows:

(i) To create a Myanmar-Korean parallel corpus

(ii) To have clear understanding of how to work neural networks for language translation

(iii) To study attention-based neural machine translation in language translation

(iv) To develop Myanmar-Korean machine translation model in both directions

(v) To evaluate the accuracy of Myanmar-Korean machine translation in BLEU Score

1.4 Contributions of the Thesis

The system develops the implementation for Attention-based Neural Machine Translation System between Myanmar and Korean language. The suggested approach is quite helpful for translating between Myanmar and other languages. The contributions of the thesis are as follows:

(i) New Myanmar-Korean corpus is constructed.

(ii) Neural Machine Translation system with attention model for Myanmar-Korean language pair is proposed.

1.5 Organization of the Thesis

This thesis is organized into five chapters. In chapter one, a neural machine translation system with attention model for Myanmar and Korean language is

introduced. The motivation, contributions, and the aim of the research work are also discussed in this chapter.

Chapter two discusses background theory and related research. The third chapter introduces the Korean language and Myanmar language in detail. In chapter four, the proposed system's design and implementation, evaluation the NMT models and deployment of the proposed system are explained.

The conclusion of the research work is presented in chapter five. Further extensions that suggest some potential improvements are offered in this chapter. This chapter also discusses the advantages and limitations of the system.

CHAPTER 2
BACKGROUND THEORY

This chapter provides an introduction to natural language processing, the development and application of machine translation systems, and the various types of neural machine translation systems. This chapter begins with outlining the characteristics of natural language processing. The overview of machine translation is covered in the second section of this chapter. Then, many types and models of machine translation are presented. The Attention-based Neural Machine Translation system is explained in detail in the end.

2.1 Natural Language Processing

Natural language processing (also known as NLP), which combines computer science, information engineering, and artificial intelligence (AI), aims to improve the machines' capacity to understand language and comprehend messages. NLP also equips computers with the capacity to understand natural languages in the same way that people do. Regardless of whether the language is spoken or written, natural language processing employs artificial intelligence to take actual information, process it, and make sense of it in a way that a computer can understand. People regularly communicate with one another using natural or human languages. While individuals communicate through words, computers only speak a language of numbers. Nevertheless, these numbers can act as a connection between the numerous languages used around the globe. We can create a translation system that will allow us to communicate openly and effectively by using NLP to solve this problem. Computers are increasingly able to understand and interpret human language thanks to a process called natural language processing. Computational linguistics, which employs rule-based modeling of human language, is combined with statistical, machine learning, and deep learning models in NLP. NLP supports all computer programs that translate text between languages, respond to spoken commands, and summarize enormous volumes of text fast, even in real time.

Participants in a two-way communication process generate and share meaning while also exchanging information, ideas, and emotions. Language is the earliest form of communication. Language is necessary for every aspect of daily life since it allows

individuals to communicate and share ideas. Additionally, language is the most efficient means for people to communicate with one another since it allows us to express our thoughts, feelings, desires, and more. Receptive and expressive language fall into these two groups. Receptive language is how we understand a language and is often achieved through listening or reading, whereas expressive language is how we use language and is typically accomplished by speaking and writing.

2.2 Machine Translation Overview

Machine translation (MT), usually referred to as automated translation, is the process by which computer software translates a text from one language into its equivalent in another. Machine translation, often known as MT or robotized interpretation, is the process of using artificial intelligence (AI) to mechanically translate contents from one natural language (the source) to another (the target) language without the assistance of a human. A large number of source and target languages are compared and matched when using a machine translation engine. Early in the 1950s, mainly in the United States, one of the earliest ideas of using computers to automatically translate human languages emerged. Beginning in the 1970s, research was being done to develop automatic translation. Over the years, researchers have tried a number of different approaches to solve the problems with machine translation. The following four techniques of machine translation are typical:

(i) Rule-based Machine Translation (RBMT)
(ii) Statistical Machine Translation (SMT)
(iii) Hybrid Machine Translation (HMT)
(iv) Neural Machine Translation (NMT)

2.2.1 Rule-based Machine Translation (RBMT)

Rule-based translation (RBMT) systems, which gather their linguistic data from dictionaries of the source and destination languages as well as grammar and apply rules developed by linguists, were the first commercial machine translation systems. To create the translated sentence, RBMT performs a grammatical analysis of the source and destination languages. A rule-based machine translation system is made up of a set of grammatical rules as well as bilingual or multilingual lexicon rules. Early in the

1970s, the first RBMT systems were developed. To map input words to output words, RBMT systems need monolingual and bilingual dictionaries that require human input. Rule-based machine translation systems can be categorized using one of three main methods. They are transfer-based machine translation, dictionary-based machine translation and interlingual machine translation.

2.2.2 Statistical Machine Translation (SMT)

The statistical machine translation system is inherited from the empirical machine translation (EMT) systems. These systems learn how to translate by analyzing a huge number of previously translated texts by humans, despite the fact that they rely on a great number of parallel aligned corpora rather than linguistic concepts or words. A statistical machine translation system is a framework for text translation from one natural language to another. These systems are based on knowledge models and statistics drawn from parallel corpora. For statistical machine translation to be effective, sentences in the source language and the target language or languages must be bilingual or multilingual. An algorithm for statistical machine learning is utilized to build the statistics tables. The training procedure is what is referred to, and the statistical tables include the statistical data. This statistical data is used to identify the best outcome during the decoding phase. In machine translation, there are three different statistical methods. They are hierarchical phrase-based model, phrase-based translation, and word-based translation.

2.2.3 Hybrid Machine Translation (HMT)

The employment of many machine translation techniques inside a single machine translation system characterizes hybrid machine translation as a method of machine translation. Combining the best aspects of two or more MT approaches is the main goal of hybrid methods. Rule-based machine translation (MT), statistical MT, and example-based MT are the three main components of hybrid machine translation (HMT), which fills in the gaps left by separate MT methods. Based on a statistical transfer technique using linguistic and statistical aspects, a system and method for hybrid machine translation has been developed. One can translate from one language into another using the system and method. A statistical translation module, a rule-based

translation module, and a hybrid machine translation engine may all be present in the system. Source, target, rule-based, and statistical language models are all stored in the database(s). Based on rule-based language models, the rule-based translation module converts source text. Based on statistical language models, the statistical translation module converts source text. In order to translate source text into target text using the rule-based and statistical language models, a hybrid machine translation engine with a maximum entropy algorithm is connected to the rule-based translation module and the statistical translation module.

2.2.4 Neural Machine Translation (NMT)

A large neural network called Neural Machine Translation (NMT) is trained end-to-end to translate one language into another. Neural Machine Translation (NMT) is an algorithm that is also used to translate words from one language to another. Deep neural networks and artificial intelligence are used in NMT to train neural models, which is a fundamentally different approach to the problem of language translation and localization. In just three years, there has been a significant shift from SMT to NMT, making NMT the main machine translation methodology. When compared to statistical machine translation methods, neural machine translation often delivers translations of a significantly higher quality, with greater fluency and appropriateness. High quality NMT can identify the context of the translation and apply models to provide a more accurate translation. Recently, the idea of neural machine translation has been put out. Neural machine translation aims to establish a single neural network that can be collaboratively modified to maximize translation performance, in contrast to standard statistical machine translation. The most current models for neural machine translation frequently come from the family of encoder-decoders and convert a source sentence into a fixed-length vector, from which a decoder produces a target sentence. The models for neural machine translation that have been recently developed frequently belong to the encoder-decoder family and encode the source sentence into a fixed-length vector, from which a decoder produces the translation. Using a single, enormous neural network, NMT trains itself to translate. With improved outcomes with language pairs, this approach is gaining popularity. NMT has been developed recently and has high levels in languages with complete resources. The encoder and decoder frameworks are the foundation for several neural machine translation models. Recurrent neural

networks serve as both the encoder and decoder in this architecture (RNNs). RNNs are a subclass of neural networks that support hidden states and allow preceding outputs to be utilized as inputs.

2.3 Attention-based Neural Machine Translation

An attention-based NMT (Bahdanau et al., 2014) is an encoder-decoder network. Traditional encoder-decoder neural network models consist of two parts: encoder and decoder. Additionally, there are sequence-to-sequence encoder-decoder models based on recurrent neural networks (RNNs). The encoder reads the full input sequence before encoding it into a context vector, which is a vector with a set length. Additionally, the decoder creates the output sequence using the encoded representation as one of its inputs. An RNN decoder produces data for another sequence whereas an RNN encoder accepts input for one sequence.

Bi-directional recurrent neural network (BiRNN), which performs better for long sentences, is used particularly by the encoder in attention-based encoder-decoder architecture. Each source word's annotation is encoded by the encoder in order to obtain the word that comes before it and the word that comes after it. A BiRNN is made up of forward RNN and forward backward RNN. The input word sequence is given to a forward RNN ($\overrightarrow{f}$) in the direction of left to right. And then it calculates a sequence as forward hidden states ($\overrightarrow{h_1}, \ldots\ldots, \overrightarrow{h_{T_x}}$). A backward recurrent neural network($\overleftarrow{f}$) also takes the sequence in the other direction, which is from right to left. It actually means from the beginning of the sentence to the end. And it results in a sequence of backward hidden states ($\overleftarrow{h_1}, \ldots\ldots\ldots, \overleftarrow{h_{T_x}}$). An annotation for each word x_j (an input sequence x like [x_1,.....,x_{T_x}]) is obtained by joining the forward hidden state $\overrightarrow{h_j^T}$ and the backward hidden state $\overleftarrow{h_j^T}$, i.e., $h_j = \left[\overrightarrow{h_j^T}, \overleftarrow{h_j^T}\right]^T$.

It is suggested that the use of attention can align and translate. The only issue with machine translation is alignment. While translation is the act of using this information to choose the proper output, alignment is the identification of the relationships between the input and output words. This alignment is known as "attention" in the field of neural machine translation, and encoder-decoder models with attention are now frequently utilized. As a result, in addition to encoders and decoders,

attention mechanisms are included in attention-based models. The decoder and alignment model compute the context vector using the sequence of annotations. The decoder accepts as inputs a representation of the context of the input, the previous hidden state, and the word prediction for the output. And following that, it generates a new hidden decoder state and a new output word prediction is:

$$p\ (y_i \mid y_1, \ldots\ldots, y_{i-1}, x) = g\ (y_{i-1}, s_i, c_i)$$

Equation 2.1

A sequence of hidden states si which are computed from the previous hidden state s_{i-1}, the embedding of the previous output word y_{i-1}, and the input context c_i as follows:

$$s_i = f\ (s_{i-1}, y_{i-1}, c_i)$$

Equation 2.2

To compute the context state c_i, the decoder gave the output as a sequence of word representations $h_j = (\overrightarrow{h_j}, \overleftarrow{h_j})$.

The attention mechanism is informed by all input word representations $h_j = (\overrightarrow{h_j}, \overleftarrow{h_j})$ and the previous hidden state of the decoder s_{i-1} and it produces a context state c_i and are computed as follows:

$$\alpha_{ij} = \frac{\exp\ (a(si-1, h_j))}{\sum_k \exp(a(si-1), h_k))}$$

Equation 2.3

Finally, the normalized attention is value use to weigh the contribution of the input word representation h_j to the context vector c_i and are done.

$$c_i = \sum_j \alpha_{ij} h_j$$

Equation 2.4

2.4 Related Work

The previous translations of the Myanmar language are discussed in this section, along with articles relevant to neural machine translation (NMT).

The authors first presented the Attention mechanism in [4], which creates an alignment model between the source and target characters or words. The parallel corpora used to evaluate the system's performance on the English-to-French translation tasks totaled 850M words. The RNN Encoder-Decoder (RNNencdec) and the suggested model are trained by the system using the equivalent settings (RNNsearch). The RNNencdec's encoder and decoder each has 1000 hidden units. One thousand hidden units each make up the forward and backward recurrent neural networks (RNN) that make up the encoder for the RNNsearch. Its decoder contains 1000 hidden units. Each model is trained by the authors using the minibatch stochastic gradient descent (SGD) technique with Adadelta (Zeiler, 2012). Each SGD update direction is computed using a minibatch of 80 sentences. The authors trained each model for approximately 5 days. On longer sentences, the presented models perform well, and the proposed RNNsearch greatly beats the traditional encoder-decoder model (RNNencdec).

An enhanced RNN neural network translation model is put out by the author in [22]. This article also compared the BLEU results of the Chinese-Korean corpus I, II, and III with the results of the conventional translation model. According to the results, the three corpora examined by this model have BLEU scores of about 45 points, compared to only 30 points for the traditional ones. The translation model in this paper's BLEU score has raised by roughly 15 points, which shows that the translation quality has greatly improved.

According to the author in [23], attention-based neural machine translation models are introduced based on word-to-word, character-to-word, and syllable-to-word levels and a parallel corpus for the Myanmar-English language pair is formed. The author used the default settings of the PyTorch OpenNMT [18]. Moreover, to decrease the low resource problem, source side monolingual data are also used. The experimental results show that syllable (Myanmar) to word (English) level neural machine translation model obtains an improvement over the other systems.

Recurrent neural networks (RNN), transformer, and convolutional neural networks (CNN) were researched by the author in [20] and tested on a parallel text corpus in Myanmar and Rakhine. Additionally, word embeddings use the word byte

pair encoding (Word-BPE) and syllable byte pair encoding (SyllableBPE) segmentation techniques. Rakhine's word segmentation was carried out manually, and the total number of words is 123,018. For training, 2,485 sentences were used for development, while 1,812 sentences were used for evaluation. According to experimental findings, Syllable-BPE segmentation produces the best NMT and SMT performances for both types of translation.

CHAPTER 3
MYANMAR AND KOREAN LANGUAGES

The introduction to Myanmar and the Korean language is described in this chapter. This chapter also presents the facts about the Korean language and its sentence structure as well as to the nature of the Myanmar language and its grammar.

3.1 Introduction to Myanmar language

About a hundred different languages are spoken in Myanmar (also known as Burma). The primary language of the Republic of the Union of Myanmar is Myanmar language, which is spoken in that country. It is also referred to as Burmese language. Myanmar belongs to the Tibeto-Burman ethnic group. About 34.5 million people speak Myanmar as their first language, while another 10 million people speak it as a second language. Additionally, Myanmar is a language that is spoken in a few regions of the United States as well as in nearby nations like Bangladesh, Malaysia, and Thailand. Despite speaking their own native tongues, ethnic groups also speak Myanmar as a second language. The syllable-based Myanmar language has its own script. Even after English was declared the official language during the colonial era, Myanmar remained the dominant language in all other scenarios.

3.2 Nature of Myanmar Language

The Myanmar script was initially derived from the Mon script, according to the history. Pali, an ancient Indian language used in the writing of Theravada Buddhism, is the foundation of the Mon script.

Myanmar script consists of (33) consonants, along with Independent vowels, Dependent consonant signs (also called Medials), Dependent vowel signs, Dependent varied signs (also called Pali Words), Punctuation, and Digits. They are depicted in Figure (3.1). The writing system used in Myanmar is left to right. The spoken style and the written style are the two types of language [23].

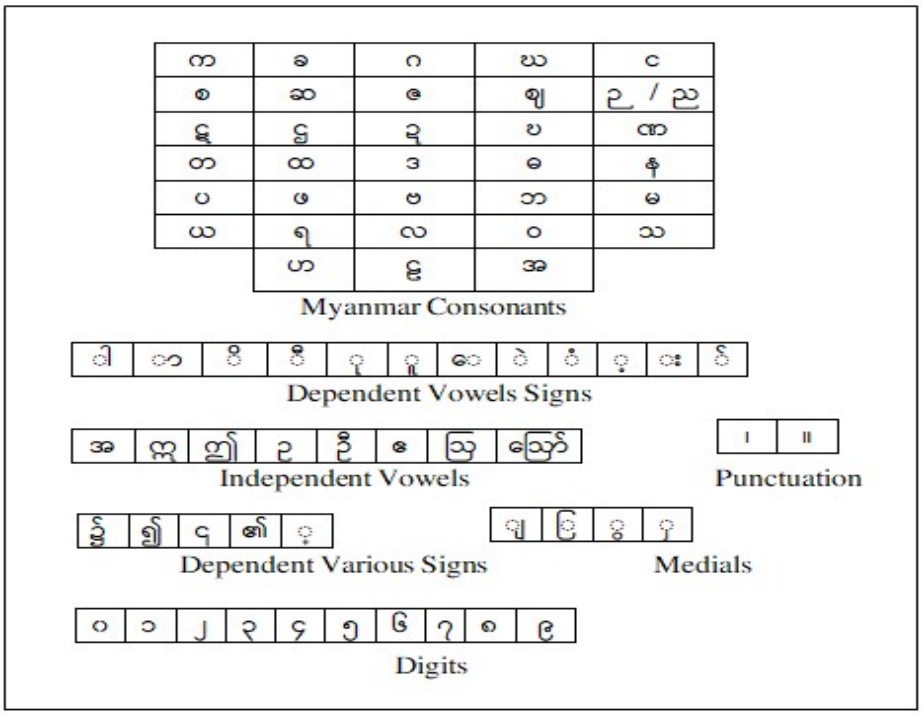

Figure 3.1 Myanmar Character Patterns

Additionally, the sentences construction is subject-object-verb (SOV). In Table 3.1, the way sentences are formed in Myanmar is illustrated.

Table 3.1 Formation of Myanmar Sentence

English Sentence	I give you this dress.										
Myanmar Sentence	ဒီဝတ်စုံကမင်းကိုငါပေးတာပါ။										
Myanmar Phrases or clauses	Noun Phrase				Noun Phrase		Noun Phrase	Verb Phrase			Punctuation
	ဒီဝတ်စုံက				မင်းကို		ငါ	ပေးတာပါ			။
Myanmar Word	ဒီ	ဝတ်စုံက			မင်းကို		ငါ	ပေးတာပါ			။
Myanmar Syllables	ဒီ	ဝတ်	စုံ	က	မင်း	ကို	ငါ	ပေး	တာ	ပါ	။
Myanmar Characters	ဒ ဝ စ က မ က င ပ တ ပ ီ ် ံ ု ် း ိ ု ါ ေ ာ ။										

The spacing between words in the Myanmar language, like other languages of Southeast Asia, is not specified. Usually, there is no space between sentences. It is occasionally written with spaces between phrases. Sentences can be easily determined with sentence boundary maker "။" which is called ပုဒ်ဖော and pronounced as "Pou ma". However, there is no set guideline on how to write in Myanmar.

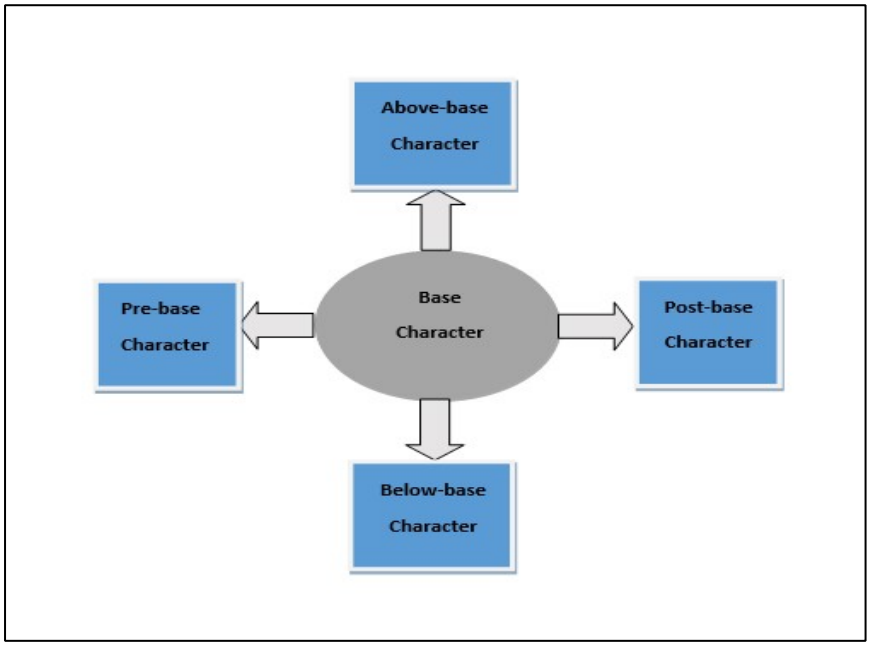

Figure 3.2 Positioning of Characters in a Myanmar Syllable

The sentence is made up of one or more words or phrases, following Myanmar sentence structure. There are one or more syllables in each word. And one or more characters make up a syllable. However, a word might sometimes just have consonants and no vowels. Figure 3.2 illustrates the placement of characters within a Myanmar syllable.

3.3 Myanmar Grammar

Myanmar words can be merged morphologically to create new words. As a result, the language of Myanmar may be inflective and agglutinative. Morphemes can be joined in any order, much like in Chinese. However, because Myanmar is primarily a head-final language, it shares many syntactic similarities with Japanese and Korean languages.

In terms of grammar, the Myanmar language has nine primary parts of speech, which are represented in Figure 3.3. They consist of the following: Noun, Pronoun, Adjective, Verb, Adverb, Particle, Post-positional, Conjunction, and Interjection.

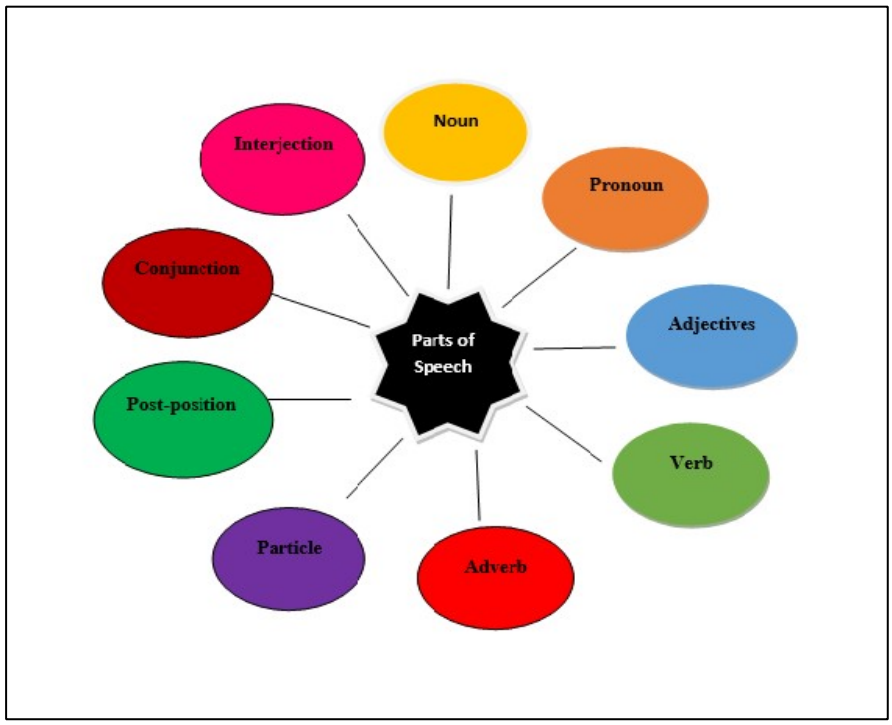

Figure 3.3 Nine main parts of speech in Myanmar Grammar

3.3.1 Nouns (နာမ်များ)

Nouns typically refer to a person, an item, or intangible concepts. Nouns can be single words or word compounds. In Myanmar, there are two categories of nouns. The four different structures and the four different meanings or representations are these. Proper nouns, abstract nouns, common nouns, and collective nouns are the four different categories of meaning or representation. Indivisible noun, compound noun, verb modification noun, and qualitative noun are the four different types of construction.

In Myanmar language, some nouns are usually started with အ (အချစ် = 사랑 = love) and ended with မှု (ပျော်ရွှင်မှု = 행복 = happiness), ခြင်း (ကြွယ်ဝခြင်း = 부유한 = wealthy), ရေး (ကျန်းမာရေး = 건강 = health), ချက် (အားနည်းချက် = 약함 =

weakness). In Myanmar language, plural nouns are formed by suffixing the particle "များ၊ တို့၊ တွင်" into singular noun. "များ" is used in writing. "တို့၊ တွင်" is used in spoken form. The Myanmar language simply recognizes the distinction between the sexes, i.e., the masculine and feminine, and does not recognize any artificial or grammatical gender. Gender-indicating particles come before nouns. "ဦး၊ ကို၊ မောင်၊ ထီး" is used to describe male gender, and "ဒေါ်၊ မ" is employed for female gender. Particles, often called measure words, are suffixes that are added to nouns to denote their type. For example, "ယောကျ်၊ ဦး" is employed to count persons. "ကောင်" is used to count animals. "ခု" is applied to general classifiers. "လုံး" is employed for spherical things. "ပြား" is applied for flat objects and "စု" is employed for group objects. Additionally, some nouns are distinguished by a concatenation of particles attached to a verb or an adjective. Post-positional makers (PPM), which are particles, are also suffixed to the end of the word. The postposition provides an explanation.

3.3.2 Pronouns (နာမ်စားများ)

When referring to persons or things, pronouns are needed. There are four types of pronouns in the Myanmar language: the personal pronoun, the referential pronoun, the question pronoun, and the mathematical pronoun.

In place of a person, a personal pronoun is used. "저,나 = ကျွန်တော်၊ ကျွန်မ = I", "당신, 너 = သင်၊ မင်း၊ ခင်ဗျား၊ နှင် = You", "그 = သူ = He", "그녀 = သူမ = She" are the personal pronouns used in Myanmar language. Referential pronouns are used to indicate that something or someone is being pointed to such as, "이 = ဤ = this", "그 = ထို = that". And then, Question pronouns are words that are similar to English terms like "what," "who," and "where." For example, in the Myanmar question: "What does she like?", "무엇 = ဘာ" that refers to "the thing that she likes" (noun), and it is called a question pronoun. By the same logic, "where" in the question: "Where did he go?",

"어디 = ဘယ်" that refers to "the place that he went" (noun), and it is also a question pronoun. Mathematical expressions for "one person," "two cups," "three groups," "five items," "some," "few," "a few," "all," "half," etc.

3.3.3 Adjectives (နာမဝိသေသနများ)

A word called an adjective is employed to change the noun. Adjectives in Myanmar typically terminate with the particles "သော". Adjectives in Myanmar are divided into three levels. In normal adjectives is concluded with "သော". In comparative adjectives, the particles "ပို၍၊ သာ၍" is prefixed to the adjectives. In superlatives adjectives, the particles "ဆုံး" is suffixed to the adjectives.

In Myanmar, there are two categories of adjectives. These are the two different kinds of constructions and the four different kinds of meaning or representation. Qualitative adjectives, referential adjectives, mathematical adjectives, and questionnaires of adjectives are the four categories of meaning or representation. The two types of construction are indivisible adjectives (느린 = နှေးသော = slow) and compound adjectives (깨끗한 = သန့်ရှင်းသော = clean).

Adjectives that modify a noun's quality by describing how something or someone is are known as qualitative adjectives. For example ("부유한 = ချမ်းသာကြွယ်ဝသော = rich"). Adjectives with references to someone or something are called referential adjectives. ဤ (이), သည် (이), ထို (그), အခြားသော (다른 = other), etc., are referential adjectives. Mathematical adjectives are words used to describe "how many" of something or someone, "what position" in an ordered list of the something, and unspecified numbers are in this category. Furthermore, it is classified into three types. They are quantitative, ordinal numbers and unspecified numbers. Quantitative adjectives are the words that described numbers followed by measure words. For example: (개 두 마리 = ခွေးနှစ်ကောင် = two dogs).

Ordinal numbers are the words that show position in the ordered list of numbers such as "first, second and third." For example: (21 번째 생일 = ၂၁ကြိမ်မြောက်မွေးနေ့), in there, “မြောက်” is an ordinal number of adjectives. Unspecified number of adjectives is the words that are used as quantifiers without numbers. အားလုံး (모두, 다), အချို့သော (조금), etc., are unspecified number of adjectives. Questionnaires of adjectives are မည်မျှ (몇), မည်သို့သော (어느, 어떤), etc.,

3.3.4 Verbs (ကြိယာများ)

The word "verb" represents an event, an action, and a situation. Usually, the root word, prefix, and suffix of a Myanmar verb can be used to identify it. In the Myanmar language, one or more particles are always added after or before the verb roots. This particle communicates information about the present tense, intentions, manners, emotion, etc. The suffix "သည်၊ ၏၊ ပြီ" can be used as a marker making the present tense statements and also as a verb post-positional marker. The suffix "ခဲ့သည်၊ ခဲ့၏၊ ခဲ့ပြီ" can be used as a marker making the past tense statements and as a verb post-positional marker. The suffix "နေသည်" can be used to describe an action in progression of happening and equivalent to the Korean verb ‘-고 있어요’ form.

The suffix "မည်၊ လိမ့်မည်၊ လတ္တံ့၊ အံ့" can be used as a marker making the future tense statements and as a verb post-positional marker. Myanmar verbs are negated by the particle "မ", and "မ" is prefixed or infixed to the verb. Usually the marker "ပါ၊ ဘူး" are used with negative verbs. For example: (စား = 먹어요, မစားပါ = 안 먹어요), (စာဖတ်သည် = 읽어요, စာမဖတ်ဘူး = 안 읽어요).

3.3.5 Adverbs (ကြိယာဝိသေသနများ)

Adverbs are words that are used to change verb tenses. Myanmar adverbs are usually ended with the particles "စွာ". There are five categories of adverbs in Myanmar.

စောစော (이른), ကြာမြင့်စွာ (너무 오래), မကြာခဏ (가끔), ချက်ချင်း (당장), ယခင် (전에) are identified as the time indicators of adverbs. ရိုသေစွာ (정중하게), လျှင်မြန်စွာ (빨리, 빠르게) are the manner indicator of adverbs. And, ဧကန် (확실히, 분명히), စင်စစ် (반드시), သည်းထန်စွာ (세게) are the situational indicator of adverbs. အလွန် (매우, 아주, 정말), နည်းနည်း (조금), လုံးဝ (전혀, 조금도) are the quantity indicator of adverbs. Moreover, မည်မျှ (몇), မည်သို့သော (어느, 어떤), ဘယ်လို (어떻게) are the questionnaires indicator of adverbs. Additionally, certain adverbs are created by integrating words that are positive and negative in opposition to one another. For example, ကောင်းမကောင်း (좋게든 안 좋게든).

3.3.6 Particles

A word that qualifies on a noun, pronoun, adjective, verb, and adverb is called a particle. Defining a particle is adding it to the end of a noun, pronoun, verb, adjective, or adverb. A word that cannot be translated is a particle. Some Particles are များ၊ တို့၊ သော၊ သည့်၊ မည့်၊ သာ၊ သင့်၊ ပင်၊ ဝုံ၊ ရက်၊ ရှာ၊ တော့၊ နှင့်၊ ပါ and etc.

3.3.7 Post-positional (PPM)

A post-positional word is one that comes after or is added to a noun, pronoun, or verb. Pronouns and nouns marked with PPM denote the subject and object, respectively.

PPM verbs indicate the time and the mood. Despite the fact that particles cannot be translated, several post-positional makers can, with the exception of subject maker and object maker. The noun post-positional makers are shown in Table 3.2 below.

Table 3.2 Noun PPM

No.	Makers	PPM	Examples
1.	Subject Makers	သည်၊ က၊ မှာ	ကျွန်မသည် (I)
2.	Object Makers	ကို	စာအုပ်ကို (book)
3.	Receiver Makers	အား	ဆရာမအား (teacher)
4.	Place Makers (Location)	၌၊ မှာ၊ တွင်၊ ဝယ်၊ က	at, on, in
5.	Place Makers (Departure)	မှ၊ က	from
6.	Place Makers (Destination)	သို့၊ ကို	to
7.	Place Makers (Direction)	သို့၊ ကို ၊ ဆီသို့	to, towards
8.	Place Makers (Continuation of place)	တိုင်တိုင်၊ အထိ	until, till
9.	Time Makers	မှာ၊ တွင်	at, on, in
10.	Continuous of Time	တိုင်တိုင်၊ အထိ	up to, till
11.	Instrumentality Makers	ဖြင့်၊ နှင့်အတူ	by, with
12.	Cause Makers	ကြောင့်၊ သဖြင့်	because, because of
13.	Possessive Makers	၏၊ ရဲ့	's
14.	Accordance Makers	အလိုက်၊ အရ	as, according to
15.	Accompaniment Makers	နှင့်၊ နှင့်အတူ၊ နှင့်အညီ	and, with
16.	Choice Makers	တွင်၊ အနက်၊ မှ၊ ထဲမှ	between, among
17.	Purpose Makers	ရန်၊ ဖို့၊ အတွက်	to, for

Verb PPM comes in two types. There are four different categories and three different tenses. In Table 3.3, verb post-positional makers are presented.

Table 3.3 Verb PPM

No.	Makers	PPM
	Three types of tense	
1.	Present Tense	သည်၊ ၏၊ ပြီ
2.	Past Tense	ခဲ့သည်၊ ခဲ့၏၊ ခဲ့ပြီ
3.	Future Tense	မည်၊ လိမ့်မည်၊ လတ္တံ့
	Four types of tense	
1.	command [literary] Maker	လော့
2.	Consensus Maker	စို့၊ ရအောင်
3.	sympathy and mercy Maker	ပါရစေ
4.	Judgment Maker	စေ

3.3.8 Conjunctions (စကားဆက်များ)

A conjunction joins and holds words, phrases, clauses and sentences. Conjunctions are used for joining the similar things or items. For example, "하고, (이)랑 = နှင့်", "(이)랑 같이, (이)랑 함께 = နှင့်တကွ", "에 따라서 = နှင့်အညီ", "그밖에 = ထို့အပြင်" and etc. To describe the contrasts, conjunctions can also be used such as "그러나 = သို့သော်", "하지만 = သို့ပေမယ့်" and "그런데도 = သို့ပါသော်လည်း". In addition, conjunctions are also used to represent the constraints and challenges, for example, "는데도 = သော်လည်း". Conjunctions are also used to show "or else" choice, logical consequence such as "그래서 = ထို့ကြောင့်", and "결과적으로 = ရလဒ်အနေဖြင့်", or to describe the desired end result such as "하기 위해, 하려고 = ဖြစ်စေရန် " and "하도록 = စေရန်".

3.3.9 Interjections (အာမေဍိတ်များ)

Words or phrases used as interjections to convey emotion. Interjections are not subject to any rules pertaining to a sentence's grammatical function. Additionally, there is no connection between these words and the rest of the statement. This sentence still makes sense even though an interjection word is missing from it. An interjection can also be used on its own. Myanmar people frequently utilize interjections to express their emotions. For example: (အမလေး၊ ဘုရား၊ အမွေ့၊ ဟယ်၊ ဩော်၊ ဟွဲ).

Interjections come in seven different varieties. They are displayed in Table 3.4 below.

Table 3.4 Seven types of Interjection

Interjections for greeting and farewell	နှုတ်ဆက်ခြင်းအတွက် အာမေဍိတ်များ
Interjections for joy	ပျော်ရွှင်ခြင်းအတွက် အာမေဍိတ်များ
Interjections for attention	အာရုံစိုက်ခြင်းအတွက် အာမေဍိတ်များ
Interjections for approval or praise	သဘောတူညီချက် (သို့) ချီးမွမ်းခြင်းအတွက် အာမေဍိတ်များ
Interjections for surprise	အံ့အားသင့်ခြင်းအတွက် အာမေဍိတ်များ
Interjections for sorrow or pain	ဝမ်းနည်းခြင်း (သို့) နာကျင်ခြင်းအတွက် အာမေဍိတ်များ
Interjections for expressing doubt or hesitation	သံသယကို ဖော်ပြခြင်း (သို့) တုံ့ဆိုင်းခြင်းအတွက် အာမေဍိတ်များ

3.4 Korean Language

Both North and South Koreans speak Korean as their official and native tongue. However, the two Koreas have acquired certain distinct vocabularies throughout the previous 74 years of political division. There are about 80 million Korean speakers in the world. The term "Korean" can refer to a language, a group of people, or a feature of

a culture. Although they speak in separate dialects—the South Korean dialect and the North Korean dialect—both languages share the same basic components. South Korea is known as ROK (Republic of Korea), and North Korea is known as DPKR (Democratic People's Republic of Korea) [5][9]. Additionally, Korean's syntax and sentence structure are comparable to those of Myanmar. The alphabet and writing system used for the Korean language are unique. The Korean Peninsula, which is occupied by South Korea and North Korea, has two varieties of the Korean language.

The majority of Korean language students are learning the South Korean language, known as 한국어 (hangugeo). The other language spoken on the Korean Peninsula is referred to as North Korean language 문화어 (munhwaeo).

King Sejong Daewang (세종 대왕), the fourth king of Korea's Joseon dynasty, created the Korean alphabet in 1443, also known as Hangul or Hangeul (한글) in South Korea and Chosongul (조선글) in North Korea. After World War II and the Korean War, it rose to prominence as the most significant writing system in both North and South Korea. Before the creation of Hangul, Koreans used different native phonetic writing systems together with Classical Chinese characters (also known as Hanja, 한자) to express themselves in writing. However, because Korean and Chinese are fundamentally different languages, a lot of lower-class individuals lack literacy. Therefore, King Sejong personally invented and established a new alphabet: the Korean alphabet—in an effort to aid in the literacy of more common people. People with little no formal education can simply learn to read and write thanks to the new writing system. In 1896, Koreans started writing sentences with gaps between words. In the writing system known as Hangul (sometimes written "Hangeul"), there are 14 consonants and 10 vowels. 14 consonants and 10 vowels make up the 24 basic letters of the Korean alphabet, known as Hangul. There are also 19 complex letters made up of 11 complex vowels and 5 tense consonants that are created by joining the basic letters. They are described in Figure 3.4.

(i) Original consonants + additional 5 consonants

(ii) Original vowels + additional 11 vowels

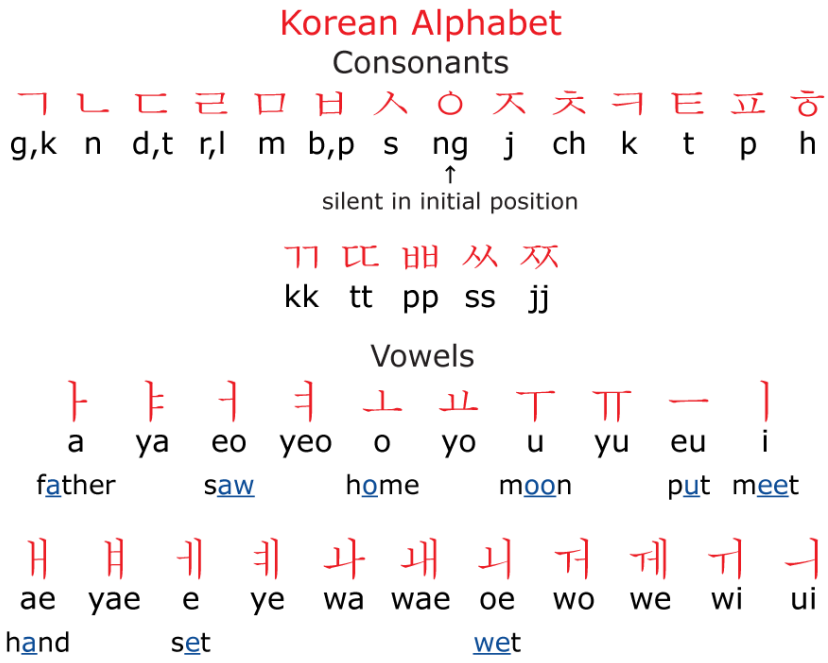

Figure 3.4 Korean Alphabet

3.4.1 Korean Consonants and Vowels

Hangeul (한글) is a combination of the Korean words han (한), which means "great," and geul (글), which means "script." The name can also be translated to "Korean script" because the word han (한) is also frequently used to refer to Korea as a whole [7][14]. The Korean alphabet is written in a system known as ganada (가나다순), which separates consonants from vowels. Consonants are listed first, followed by vowels. With representing mother and indicating son, the syllabic Korean script can be broken down into vowels (모음, mo-eum) and consonants (자음, ja-eum). Korean consonants are listed in the Tables 3.5. The romanized spelling of each individual consonant is listed next to it. Whether the consonants are at the beginning or the end of the syllable affects how the word is spelled. Only the spelling of the Korean word in English letters uses the romanization. Korean vowels appear on the second Table 3.6. The romanized spelling for every vowel is located next to it. The vowels are spelled consistently and without variation.

Table 3.5 Korean Consonants

Korean Consonant	Name of the consonant	Romanized Spelling
ㄱ	기역	giyeok
ㄴ	니은	nieun
ㄷ	디귿	digeut
ㄹ	리을	rieul
ㅁ	미음	mieum
ㅂ	비읍	bieup
ㅅ	시옷	siot
ㄲ	쌍기역	ssangiyeok
ㄸ	쌍디귿	ssangdigeut
ㅃ	쌍비읍	ssangbieup
ㅆ	쌍시옷	ssangsiot
ㅉ	쌍지읒	ssangjieut
ㅇ	이응	ieung
ㅈ	지읒	jieut
ㅊ	치읓	chieut
ㅋ	키읔	kieuk
ㅌ	티읕	tieut
ㅍ	피읖	pieup
ㅎ	히읗	hieut

Table 3.6 Korean Vowels

Vowel/ Name of the Vowel	Romanized Spelling
ㅏ	a
ㅐ	ae
ㅑ	ya
ㅒ	yae
ㅓ	eo
ㅔ	e
ㅕ	yeo
ㅖ	ye
ㅗ	o
ㅘ	wa
ㅙ	wae
ㅚ	oe
ㅛ	yo
ㅜ	u
ㅝ	wo
ㅞ	we
ㅟ	wi
ㅠ	yu
ㅡ	eu
ㅢ	ui
ㅣ	i

3.4.2 Sentence Structure of Korean

The subject-object-verb structure is the foundation of basic Korean sentences (SOV). The sentence structure of the Korean language is the same as that of the Myanmar language.

ကျွန်မသည်	ရေကူးခြင်းကို	ကြိုက်သည်။
저는	수영하기를	좋아합니다.

In this Korean sentence, “는” is subject particle, “를” is object particle and “-ㅂ니다” is declarative ending (polite form). There are three important Korean sentence patterns. They are:

(i) Subject-Object-Verb (SOV)

(ii) Subject-Verb (SV)

(iii) Subject-Adjective (SA)

The most fundamental sentence structure in Korean is the first pattern, SVO. In this pattern, the subject and object are introduced in the first half of the phrase, and the verb that occurs between them is described in the second. A SOV phrase pattern is essential in constructing the structure. Additionally, the Myanmar sentence structure also follows the same pattern. For example: “저는 한국어를 공부해요.” = “ကျွန်မသည် ကိုရီးယားစာကို လေ့လာသည်။” In this example sentence, subject is “저 = ကျွန်မ”, “한국어 = ကိုရီးယားစာ” is the object and the last part is verb “공부해요 = လေ့လာသည်”. Some example sentences of SOV pattern are described in Table 3.7.

Table 3.7 Example sentences of SOV pattern

그녀**는** 문**을** 닫았어요.	သူမသည် တံခါးကို ပိတ်ခဲ့သည်။
저**는** 책**을** 읽고 있어요.	သူသည် စာအုပ်ကို ဖတ်နေသည်။
그**는** 경기**를** 볼 거예요.	သူသည် ပြိုင်ပွဲကို ကြည့်လိမ့်မည်။

SV is the second pattern. Sometimes the meaning of the sentence can be understood without an object. Just the subject and the verb can be used. This is the simplest sentence pattern in Korean. Equivalent sentence structures can be found in English. So, utilizing this specific sentence structure, it is not difficult to construct a sentence. Table 3.8 displays a few SV pattern example sentences.

Table 3.8 Example sentences of SV pattern

할아버지**가** 오셨어요.	အဖိုးက လာခဲ့သည်။
엄마**는** 울었어요.	အမေက ငိုခဲ့သည်။
저**는** 노래해요.	ကျွန်မ သီချင်းဆိုသည်။
그녀**는** 들었어요.	သူမက ကြားခဲ့သည်။
그**는** 말했어요.	သူက ပြောခဲ့သည်။
나**는** 실패 할 거예요.	ငါ ကျရှုံးလိမ့်မယ်။

Subject-Adjective (SA) pattern is the final pattern. A sentence may usually contain just a subject and an adjective. It is a typical sentence structure in Korean. The S+A sentence structure is also common in English. Examples of SA-structured sentences are shown in Table 3.9.

Table 3.9 Example sentences of SA pattern

저**는** 바빠요.	ကျွန်မ အလုပ်ရှုပ်နေတယ်။
날씨**가** 덥다.	ရာသီဥတုက ပူလိုက်တာ။
영화**는** 길었다.	ရုပ်ရှင်က ရှည်တယ်။
학생들**은** 긴장했어요.	ကျောင်းသားတွေက စိတ်လှုပ်ရှားနေတယ်။
내 개**는** 게을러요.	ငါ့ ခွေးကတော့ ပျင်းလိုက်တာ။

3.4.3 Korean Particles: Markers and Indicators

Despite the fact that particles are tiny, the subject of Korean particles is actually quite large. This is due to the fact that "Korean particles" serves as a general word. There are characters that appear right after the nouns in Korean sentences. These characters follow nouns like the tail of a dog. In a Korean sentence, a particle is usually attached to the majority of the words. They are 은, 는, 이, 가, 을 and 를. And they show up in Korean sentences over and over again. Indicator or marker words are particles in the Korean language. These particles identify each word's function in a sentence, i.e., which word is the subject or object. There are quite a lot of Korean particles. However, approximately 20 are frequently employed.

The following three elements are crucial to Korean sentence structure:

1. Topic Markers: 은 and 는

This is used to denote that a word is the subject of a sentence by placing it after the word. “은” and “는” are essentially the same. “은” is used for nouns that end with a consonant, while “는” is used for nouns that end with a vowel. For example: “저 = 저는”, “집 = 집은”.

2. Subject Markers: 이 and 가

To introduce a NEW subject, this participle is utilized. “이” is used for nouns that end with a consonant, and “가” is for nouns that end with a vowel. For example: “석양이 아름답다 = နေဝင်ချိန်က လှတယ်”, “개가 나를 물었다 = ခွေးက ငါ့ကို ကိုက်တယ်”.

3. Object Markers: 을 and 를

This is used to denote that the word is the sentence's object by placing it after the word. It is applied to all Objects. In grammar, the object being acted upon is known as the sentence's object. The verb is applied to the object. “을” is used when the preceding noun ends with a consonant, and “를” follows a noun that ends with a vowel. For

example: “영화를 = ရုပ်ရှင်ကို”, “책을 = စာအုပ်ကို”. The object comes before the verb in an SOV sentence form.

4. Linking Particles: 와, 과, 하고 and (이)랑

The next group of particles is equivalent to the word "and" in English and “နှင့်” in Myanmar language. They serve to denote the pairing or grouping of nouns. For example: “ပန်းသီးနှင့်လိမ္မော်သီး = 사과하고 오렌지”, “ခွေးနှင့်ကြောင် = 개와 고양이”. There are numerous particles that can accomplish this. There is 와, 과, 랑, 이랑 and 하고. 와 and 과 are more suited for speeches, presentations and written formats while 랑, 이랑 and 하고 are utilized in everyday conversation. When the previous noun ends with a vowel, the symbol 와 is used. When the previous noun ends with a consonant, the symbol 과 is used. In the case of the other pair, 랑 is used when the noun before it ends in a vowel and 이랑 is used when it finishes in a consonant. 하고 is a free-form word that can be used with vowels and consonants.

5. Plural Particle: 들

To the end of nouns in English, we usually add a "s" or "es". In Korean, the noun is followed by the word "들(deul)". Similar to this, “များ၊ တွေ” is employed in Myanmar for plural nouns. However, unlike in English and Myanmar, Korean rarely pluralizes nouns. There is no significant distinction between singular and plural nouns in Korean. Native speakers have little trouble with this because context frequently suffices to notify the listener whether the word is single or plural. As a result, a sentence like “나는 펜을 샀다 (Na-neun pen-eul sa-dda)” can mean “I bought a pen” or “I bought pens”.

6. Possessive Particle: 의

The last one expresses ownership or possession and is the equivalent of the English apostrophe + s and “၏၊ ရဲ့” in Myanmar meaning. When two nouns are

encountered together, the particle “의 (ui)” moderates their relationship. The nouns' order is important in this sentence. The owner will be the first noun, and the thing owned will be the second noun, the one that comes after. For example: 오늘**의** 게임 = today’s game, ဒီနေ့ရဲ့ ကစားပွဲ. Frequently, in speech, “의 (ui)” is pronounced as “에 (e)”. When pronouns such as “I (저, 나)” and “you” (너), 의 or 에 is added to obtain the possessive forms “my” and “your”, the result is a contraction: 나의 becomes **내** *(nae)* — my, 저의 becomes **제** *(je)* — my, 너의 becomes **네** *(ne)* — your

Sometimes in Korean language, the subject can be dropped. This indicates that there is no subject in some sentences. For example: a Korean older brother tells the younger, “문 닫아!” (Shut the door). In this, there is no explicit subject in the sentence, but the context tells everything the younger brother needed.

CHAPTER 4
SYSTEM DESIGN AND IMPLEMENTATION

This chapter presents the precise neural machine translation implementation between Myanmar and Korean language. This chapter also provides a description of the proposed system design. The graphical user interface of the system is shown at the end, along with illustrations that clarify each step of the process and the experimental findings.

4.1 Design of Myanmar-Korean Translation

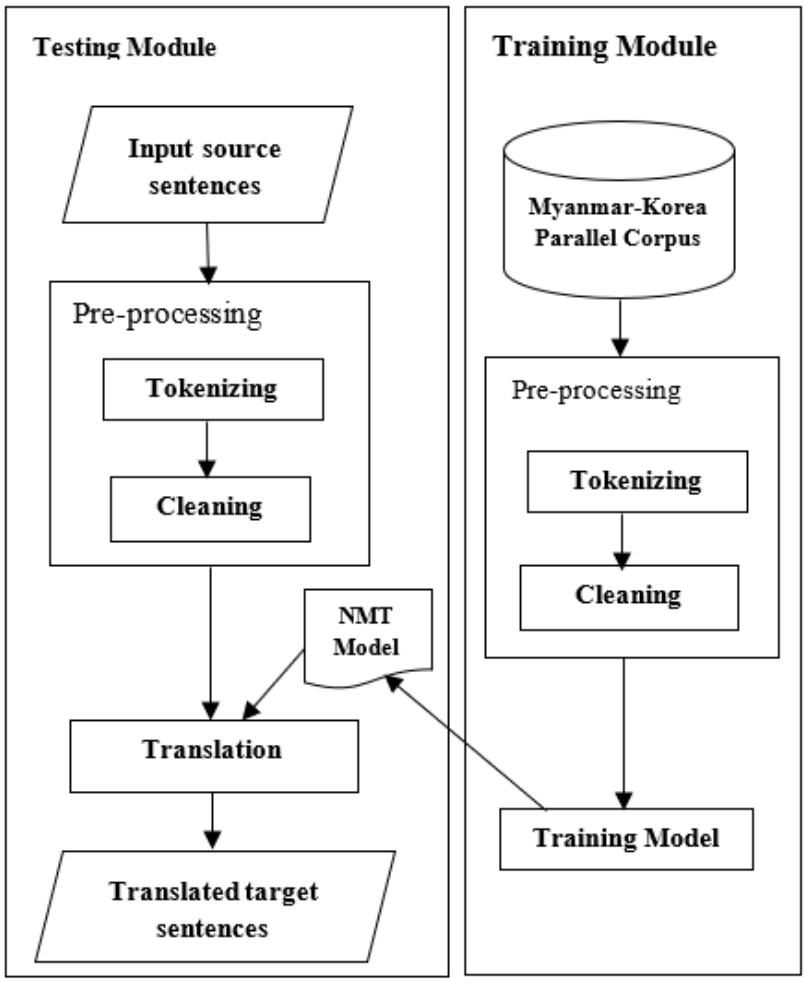

Figure 4.1 Design of Myanmar-Korean Translation

In Figure 4.1, the overall design diagram of the system is displayed. There are two primary modules in the proposed system. Training Module is the first, while Testing Module is the second. First, a parallel corpus for Myanmar and Korean needs to be built for the training module. It is necessary to tokenize and clean the corpus data

in preparation for data pre-processing. The NMT models are then trained using the PyTorch OpenNMT toolkit [18]. It is necessary to pre-process the input source sentences in the testing module. After that, a trained NMT model is used to translate sentences and produce the translated sentences.

4.2 Implementation

The implementation of the Myanmar-Korean Neural Machine Translation models in both directions using the PyTorch OpenNMT toolkit [18], which is accessible on GitHub. We will demonstrate how to implement the Myanmar-Korean neural machine translation models in this section, including the experimental setting for the system, the data preprocessing step, the training models, and the translation processes.

4.2.1 Dataset and Preprocessing Tools

One of the low resource languages is Myanmar. There are not many parallel corpora between Myanmar and Korean at the present. For this system, the author must therefore construct a parallel Myanmar-Korean corpus. In order to create a new Myanmar-Korean parallel corpus, Myanmar sentences from the UCSY Myanmar-English Corpus [23] are collected and manually translated into Korean language. About parallel sentences from local news, travel-related articles, school textbooks, and spoken textbooks in both languages are included in the corpus. More than 37K parallel sentences may be found in this parallel corpus. These parallel sentences are divided up at the word level. UCSY NLP Word Segmenter tool [21] was used for the word segmentation task, and Korean sentences were segmented manually according to Korean grammar rules. For cleaning the corpus, Moses's clean scripts [16].

The Myanmar-Korean parallel corpus is randomly divided into three division files as shown in the following table 4.1 in order to train the Myanmar-Korean NMT models:

Table 4.1 Statistics of Korean-Myanmar parallel corpus

Files	**No. of sentences**
Training File	33925
Validation File	3017
Testing File	700
Total Sentences	37642

4.2.2 Neural Machine Translation Model

These days, nearly all language pairs may be successfully translated via neural machine translation, and the field is expanding rapidly. Furthermore, there are several toolkits available for the study, creation, and use of neural machine translation systems. Various NMT implementations are currently in use. The PyTorch OpenNMT tool [18], which is accessible on GitHub, was used to create the Myanmar-Korean neural machine translation models. A 2-layer long short-term memory with 500 hidden units on both the encoder and decoder is employed for the translation system. Drop-out was set to 0.3, and each direction's computations were done with 64 batches of data and a learning rate of 1.0. The Korean language has 25,282 terms in its vocabulary, compared to 15,188 words in the Myanmar language.

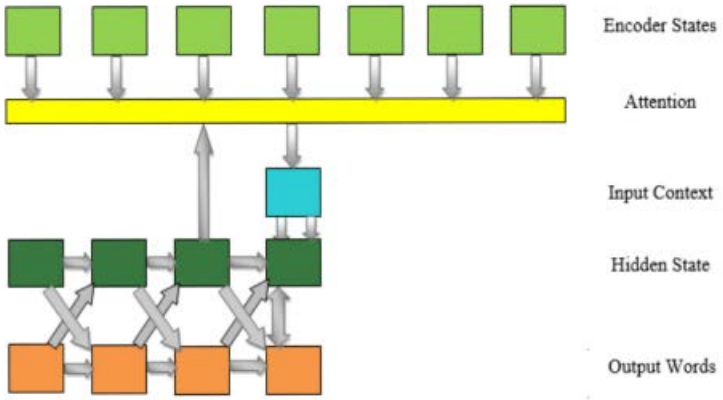

Figure 4.2 Attention Model Architecture

Figure 4.2 describes how does the Attention-based model architecture perform in machine translation.

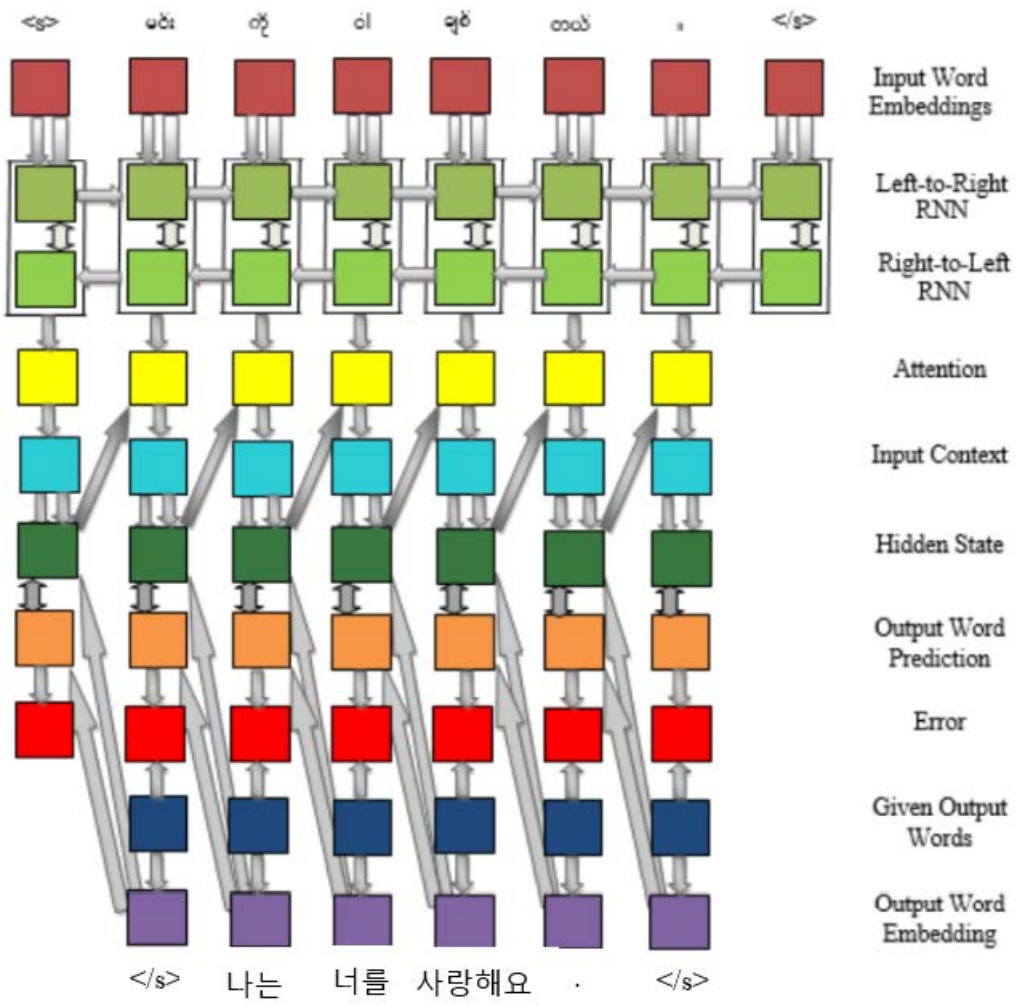

Figure 4.3 Fully computed training sample graph with 7 source words and 5 result words.

4.2.3 Evaluation

Bilingual Evaluation Understudy (BLEU), one of the de facto recognized automatic evaluation metrics, is utilized to evaluate the translation results. Additionally, BLEU is an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another. BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a set of high-quality reference translations. A value of 0 means that the machine-translated output has no overlap with the reference translation (low quality) while a value of 1 means there is perfect overlap with the reference translations (high quality) like the following equation:

$$\text{BLEU} = \underbrace{\min\Big(1, \exp\big(1 - \frac{\text{reference-length}}{\text{output-length}}\big)\Big)}_{\text{brevity penalty}} \underbrace{\Big(\prod_{i=1}^{4} precision_i\Big)^{1/4}}_{\text{n-gram overlap}}$$

$$precision_i = \frac{\sum_{\text{snt} \in \text{Cand-Corpus}} \sum_{i \in \text{snt}} \min(m^i_{cand}, m^i_{ref})}{w^i_t = \sum_{\text{snt'} \in \text{Cand-Corpus}} \sum_{i' \in \text{snt'}} m^{i'}_{cand}}$$

Equation 4.1

This BLEU metrics is employed to examine the experiments of Myanmar-Korean Neural Machine Translation models. These experiments' BLEU scores are displayed in Table 4.2.

Table 4.2 Evaluation result of Korean-Myanmar NMT models

NMT Model	**BLEU**
Myanmar-Korean	12.58
Korean-Myanmar	20.32

The experimental results show the performance of Korean-to-Myanmar model is better than the Myanmar-to-Korean model. In translation results, it is found that some Foreign Names are unable to translate, and some Korean sentences are not using the subject word to translate. For the future experiments, the existing Myanmar-Korean parallel corpus cannot be sufficient to train the translation models since the quality of Neural Machine Translation (NMT) systems is strongly depended on the size of parallel corpus. Therefore, the collection of more data and more work of other neural models require to increase the translation performance.

4.3 Deployment of the System

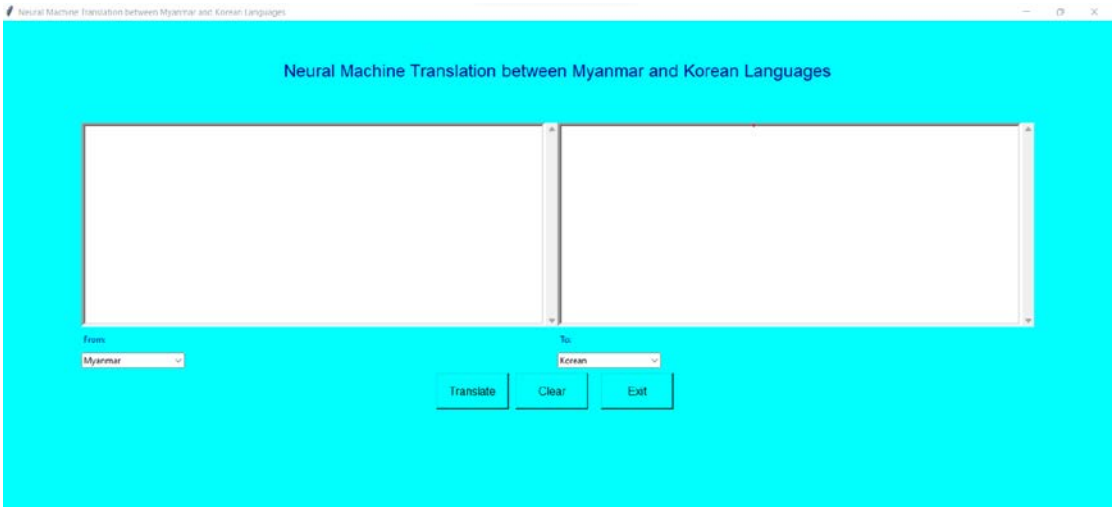

Figure 4.4 Graphical User Interface Design of the System

The graphical user interface for the neural machine translation between Korean and Myanmar is shown in Figure 4.4. This system has two text boxes: one for the input source sentence and another for the translated sentence. Two combo boxes with options for source and target languages are set up in the left and right panels, respectively. This system also has three different kinds of buttons: Translate, Clear, and Exit. The Translate button allows you to convert sentences from the source language into the preferred language. Similarly, the sentences that were entered are eliminated by using the Clear button. The final Exit button is then activated to shut off the system.

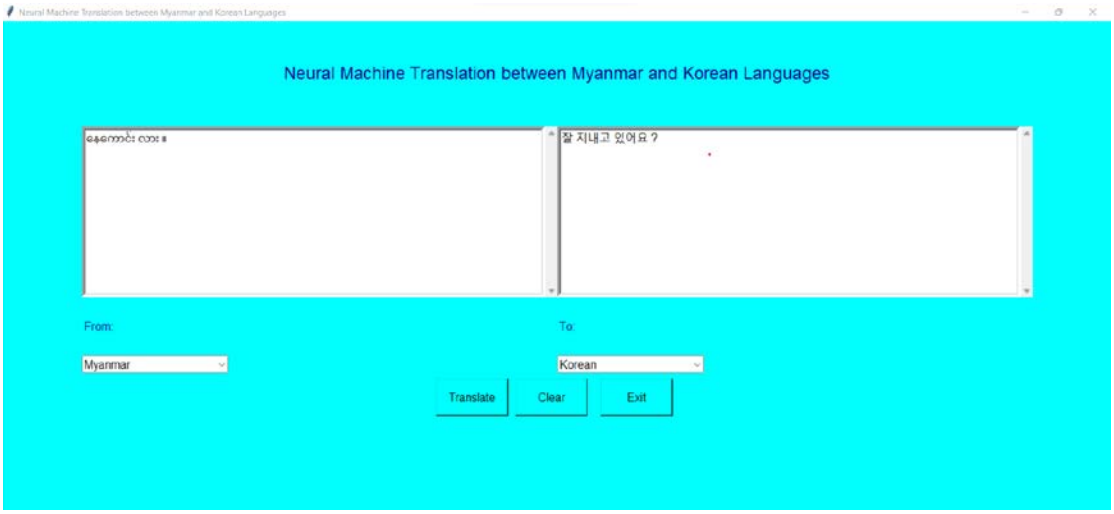

Figure 4.5 An accurate sentence translation from Myanmar to Korean language

The example sentence translation from Myanmar to Korean is shown in Figure 4.5. Translating Myanmar greeting word “နေကောင်း လား ။” outputs the Korean word “잘 지내고 있어요 ?”. The accurate translation of the input sentence can be obtained from this experiment.

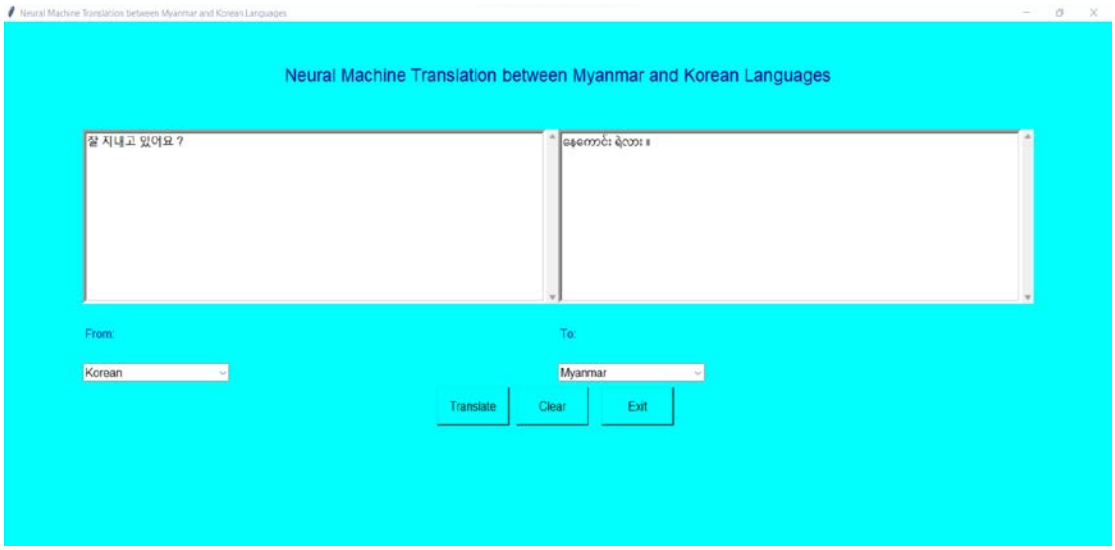

Figure 4.6 An accurate sentence translation from Korean to Myanmar language

Figure 4.6 displays an example of a sentence being translated from Korean to Myanmar. The Myanmar welcome phrase "နေကောင်း ရဲ့လား ။" is produced by translating the Korean word "잘 지내고 있어요 ?" in the opposite direction. This experiment provides a correct translation of the provided sentence.

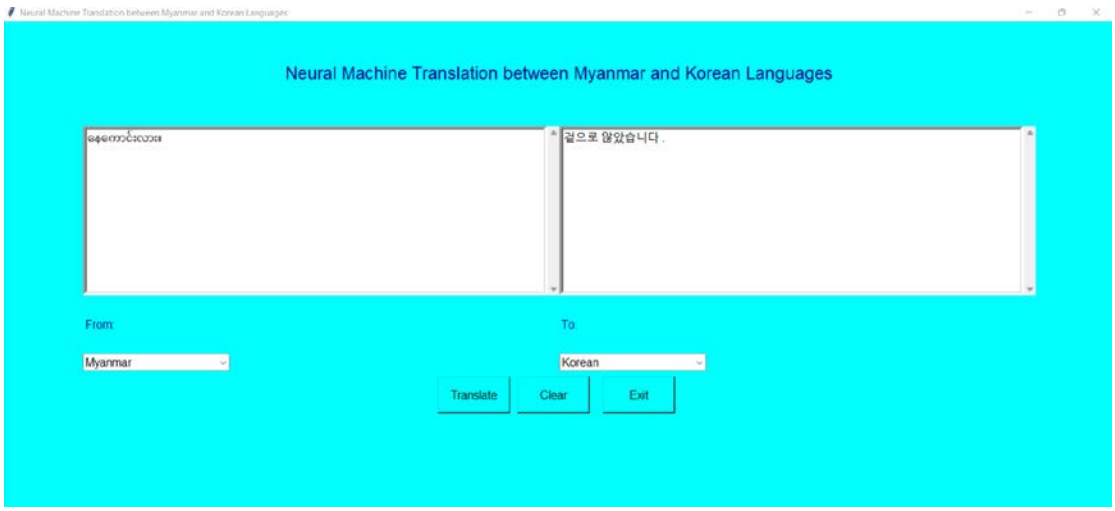

Figure 4.7 False translation result of Myanmar sentence into Korean language

Due to incorrect segmentation of the input Myanmar sentence, Figure 4.7 shows an incorrect translation of the input sentence into Korean. The segmented sentences in both languages make up the training corpus. As a result, the trained model is unable to translate the input sequence accurately.

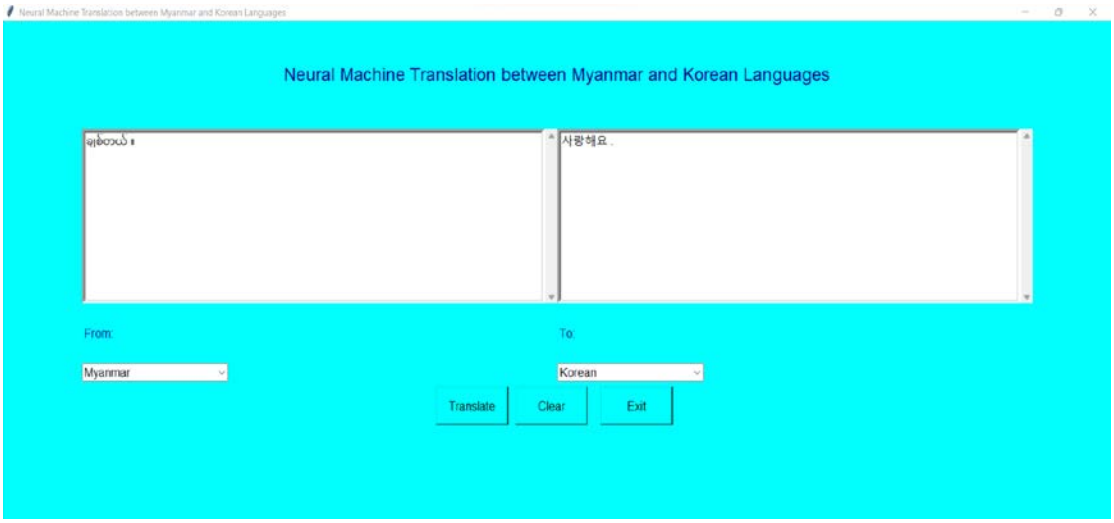

Figure 4.8 A correct sentence translation from Myanmar to Korean language

Figure 4.8 represents the translation of a correct sentence from Myanmar to Korean language. “ချစ်တယ် ။” in Myanmar language are translated into “사랑해요 .” in Korean language.

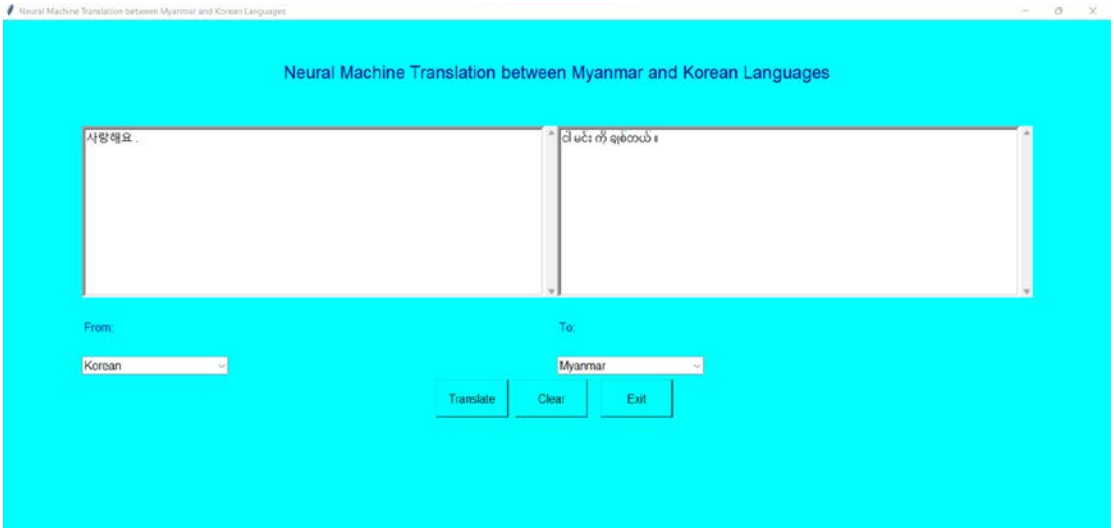

Figure 4.9 A correct sentence translation from Korean to Myanmar language

The correct translation of the Korean sentence, which omits the subject, is shown in Figure 4.9. Typically, Koreans omit the subjects of their sentences when speaking in casual conversation.

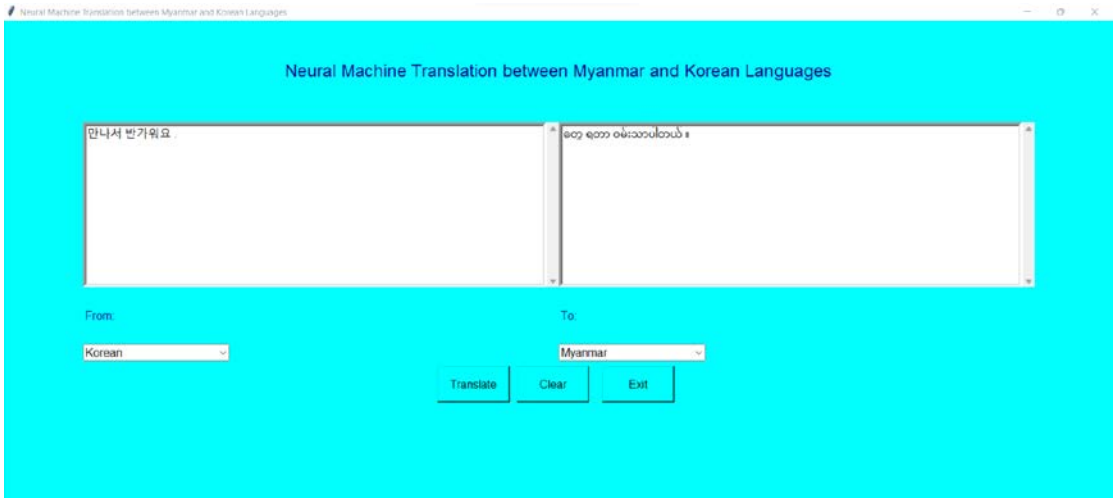

Figure 4.10 A correct sentence translation from Korean to Myanmar language

The precise sentence translations between Korean and Myanmar are shown in Figure 4.10.

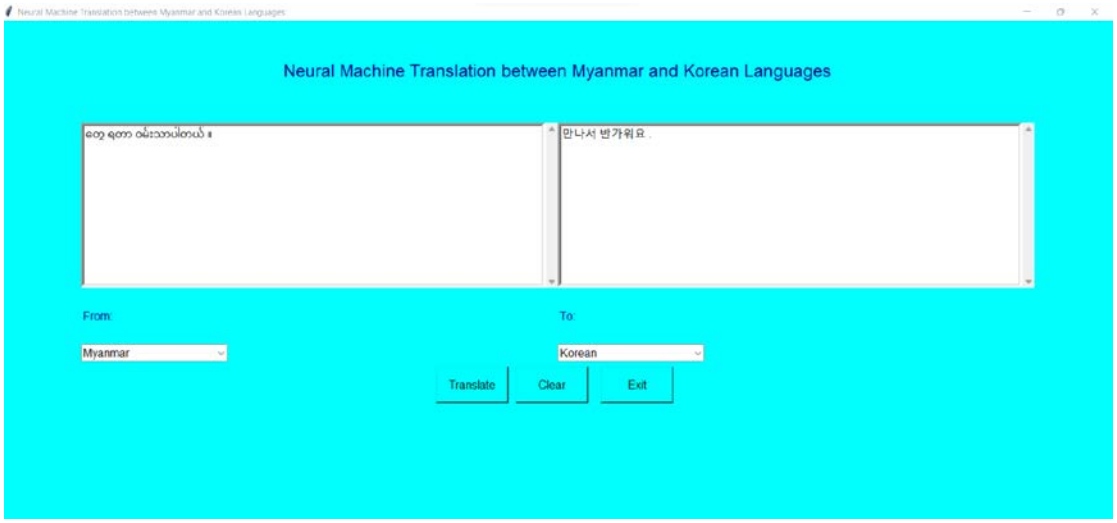

Figure 4.11 A correct sentence translation from Myanmar to Korean language

Similar to Figure 4.10, Figure 4.11 shows the exact sentence translations between Korean and Myanmar.

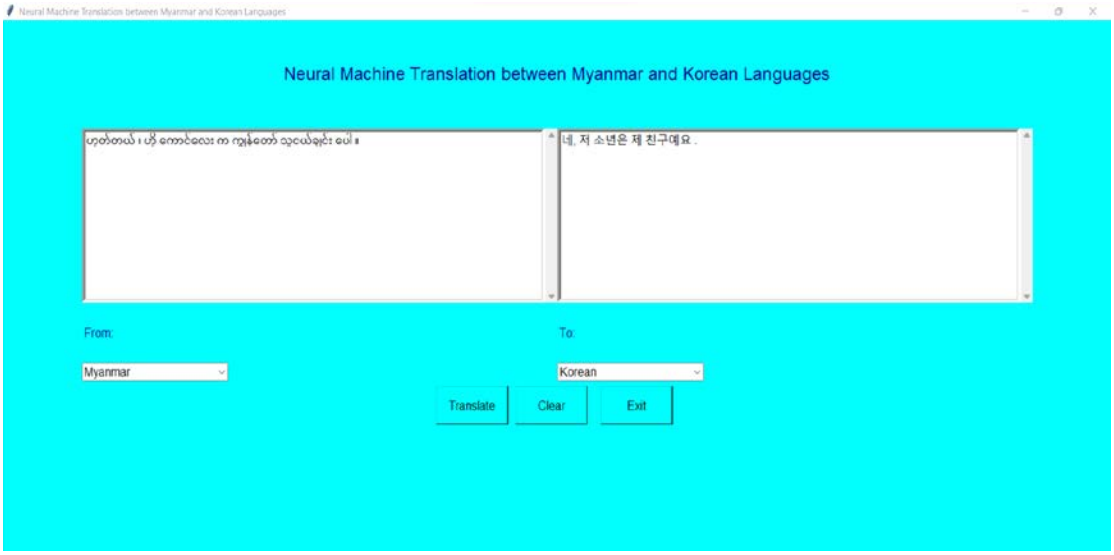

Figure 4.12 A correct long sentence translation from Myanmar to Korean language

Additionally, this approach can translate lengthy sentences in both languages that are used in daily communication in addition to everyday speech. This is depicted in Figure 4.12 and Figure 4.13.

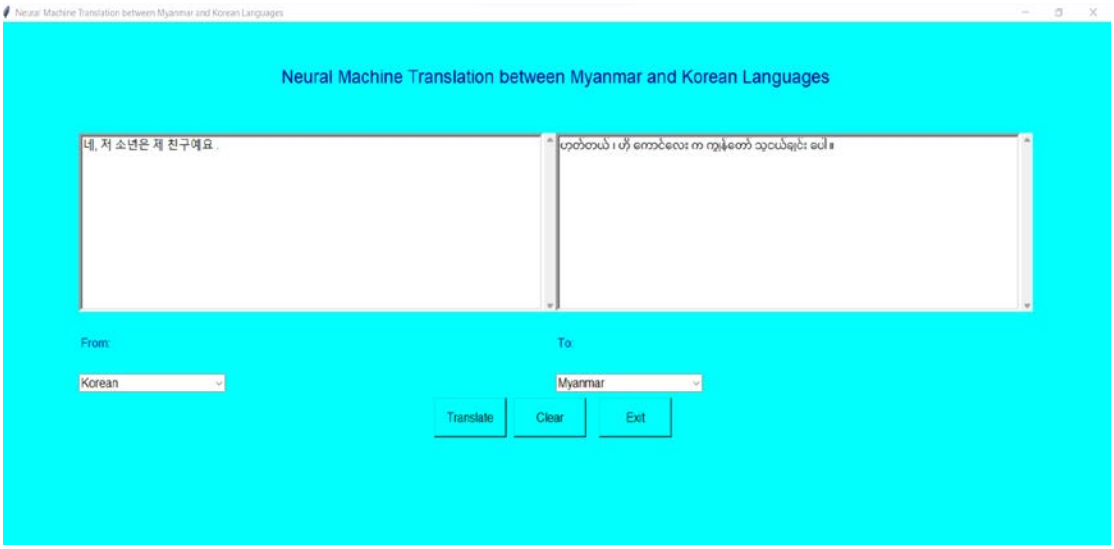

Figure 4.13 A correct translation of a long sentence in Korean to Myanmar language

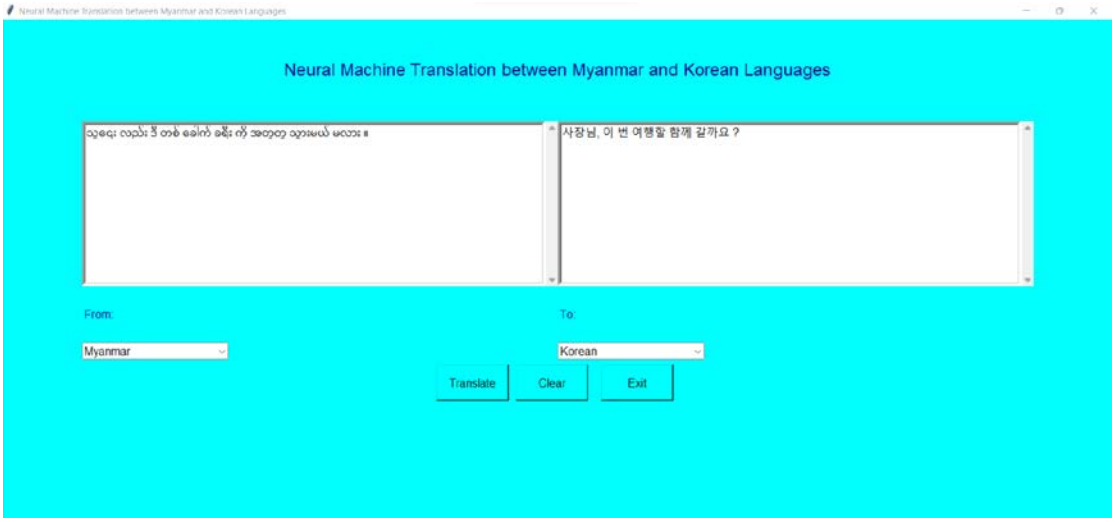

Figure 4.14 A good translation with a small error in the final product

The proposed model is able to translate the appropriate statement in the mentioned Figure 4.14 with a tiny mistake. In actuality, “သူငယ်ချင်း လည်း” and “ခရီး” must be “사장님도” and “여 행”. As a result, the author can spot two errors in this translated output.

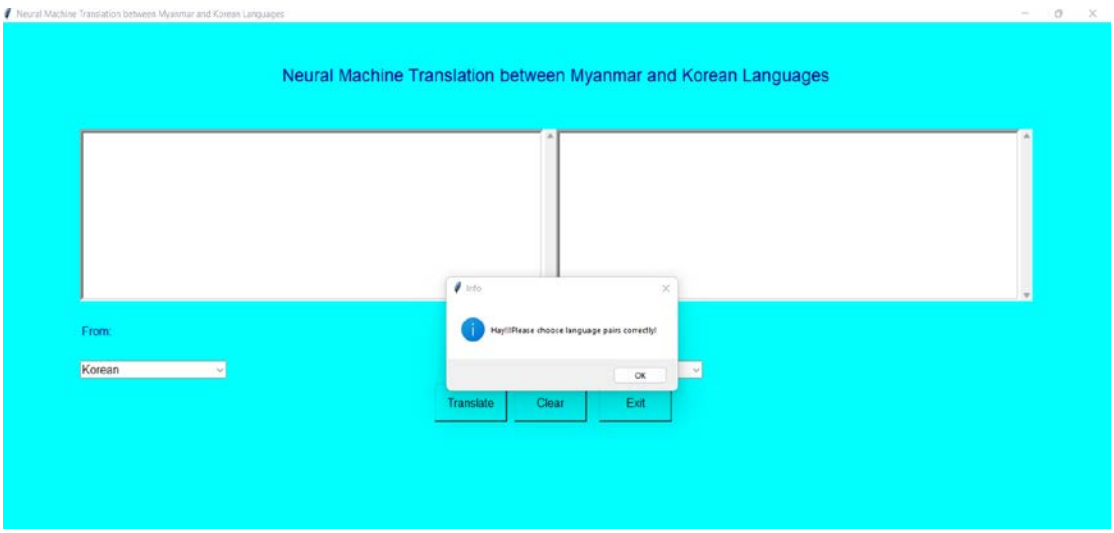

Figure 4.15 Advisory regarding appropriate language selection

If it does not correctly identify the source and target languages, the system will display a warning notice and prompt us to do so immediately, according to Figure 4.15.

CHAPTER 5
CONCLUSION AND FURTHER EXTENSIONS

This paper aims to create a neural machine translation system for the Myanmar-Korean language pair using an attention model based on a newly established parallel corpus in both languages. A new Myanmar-Korean parallel corpus is created for the system because there is a limited number of publicly available Myanmar-Korean parallel corpora. The key points of the paper are summarized, the benefits and drawbacks of the system are discussed, and ideas for further paper are offered in this chapter.

The proposed system implements neural machine translation between the languages of Myanmar and Korean. This paper proposes a mechanism that users can employ to translate sentences from Myanmar into Korean as well as Korean into Myanmar. One neural network is used in the neural machine translation (NMT) system, which has gained popularity in recent years and is showing promising results in various languages. Neural Machine Translation (NMT) is the process of translating new text using data extracted from existing corpora. Parallel corpus is the primary prerequisite for machine translation. Therefore, a new Myanmar-Korean parallel corpus is built for this system. Myanmar sentences are taken from the UCSY Myanmar-English Corpus [23] and manually translated into Korean to build a parallel corpus in Myanmar-Korean for this system. The constructed corpus contains parallel sentences from a school text book, a tour guide, and spoken text. With the assistance of instructors from Korean language classes, manual translation into Korean was completed.

5.2 Advantages and Limitations

For students and teenagers in Myanmar who are learning Korean language, the proposed system is user-friendly and helpful. The new Myanmar-Korean parallel corpus can be used by researchers for future Myanmar-Korean machine translation tasks. To build neural machine translation models and get the best performance in language translations, however, the size of this parallel corpus is insufficient. As a consequence of the translation efforts, the author has discovered that some foreign names cannot be translated, and that the subject word cannot be translated in some

Korean sentences. Additionally, if the source sentence is not accurately segmented before being input, the translation cannot produce an appropriate result. This is due to the fact that the proposed training model is based on a word-level segmentation scheme linked to both languages.

5.3 Further Extensions

The proposed method is created by utilizing a small parallel corpus of Korean and Myanmar. The current Myanmar-Korean parallel corpus will no longer be sufficient to train the translation models in the future. Since the size of the parallel corpus has a significant impact on the accuracy of Neural Machine Translation (NMT) systems, the current Myanmar-Korean parallel corpus will not be adequate for training translation models in the upcoming experiments. In order to improve translation performance, more data must be collected and other neural models must be trained hard work.

AUTHOR'S PUBLICATION

[1] Hnin Nandar Zaw, Yi Mon Shwe Sin, Khin Mar Soe "Neural Machine Translation between Myanmar and Korean Languages, National Journal of Parallel and Soft Computing, Yangon, Myanmar, 2022

REFERENCES

[1] Attention Model Intuition. https://www.youtube.com/watch?v=SysgYptB198

[2] BASIC KOREAN: A GRAMMAR AND WORKBOOK

[3] CHENCHEN DING, YE KYAW THU, MASAO UTIYAMA, and EIICHIRO SUMITA, National Institute of Information and Communications Technology, ACM Trans. Asian Low-Resource Lang. Inf. Process., Vol. 15, No. 4, Article 22 [Word Segmentation for Myanmar (Myanmar)]

[4] Dzmitry Bahdanau, Jacobs University Bremen, Germany and KyungHyun Cho, Yoshua Bengio, Universite de Montreal, published as a conference paper at ICLR 2015 [Neural Machine Translation By Jointly Learning To Align And Translate]

[5] https://en.wikipedia.org/wiki/Korean_language

[6] https://kikaben.com/neural-machine-translation-with-attention-mechanism/

[7] https://linguapsych.com/korean-sentence-structure/

[8] https://phrase.com/blog/posts/neural-machine-translation/

[9] https://www.90daykorean.com/how-to-learn-the-korean-alphabet/

[10] https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/

[11] https://www.fluentu.com/blog/korean/korean-sentence-structure/

[12] https://www.memoq.com/tools/what-is-machine-translation

[13] Diana Lăpușneanu [A Quick Guide to Hangul, the Korean Alphabet-Pronunciation and Rules]

[14] Korean Words Common Vocabulary Used Most Often

[15] Minh-Thang Luong, Hieu Pham and Christopher D. Manning [Effective Approaches to Attention-based Neural Machine Translation]

[16] Moses Toolkit: http://www2.statmt.org/mo ses/

[17] Naver Dictionary. https://en.dict.naver.com/#/main?sLn=en

[18] PyTorch-OpenNMT. https://github.com/OpenNMT/OpenNMT-py

[19] Sameul E. Martin. https://www.britannica.com/topic/Korean-language

[20] Thazin Myint Oo, Ye Kyaw Thu and Khin Mar Soe, UCSY NLP Lab, Myanmar Language and Semantic Technology Research Team (LST), NECTEC, Thailand Language and Speech Science Research Lab, Waseda University,

Japan [Neural Machine Translation Between Myanmar (Myanmar) and Rakhine (Arakanese)]

[21] Win Pa Pa, Nilar. Thein, "Myanmar Word Segmentation using Hybrid Approach", Proceedings of 6th International Conference on Computer Applications, 2008, Yangon, pp-166-170
http://www.nlpresearch-ucsy.edu.mm/wordsegmentation.html

[22] Yang Dong, Nanyang Institute of Technology, Nanyang, Henan 473000, China [RNN Neural Network Model for Chinese-Korean Translation Learning]

[23] Yi Mon Shwe Sin, Khin Mar Soe, International Journal on Natural Language Computing (IJNLC) Vol.8, No.2, April 2019 [Attention-based syllable level neural machine translation system for Myanmar to English Language Pair]

[24] Yi Mon Shwe Sin, Thazin Myint Oo, Hsu Myat Mo, Win Pa Pa, Khin Mar Soe and Ye Kyaw Thu [UCSYNLP-Lab Machine Translation Systems for WAT 2018]

[25] Yongkeun Hwang, Yanghoon Kim (Department of Electrical and Computer Engineering, Seoul National University) and Kyomin Jung (Automation and Systems Research Institute, Seoul National University) [Context-Aware Neural Machine Translation for Korean Honorific Expressions]