MYANMAR-WA MACHINE TRANSLATION BASED ON TRANSFER LEARNING

FLORANCE YUNE

UNIVERSITY OF COMPUTER STUDIES, YANGON

June, 2024

Myanmar-Wa Machine Translation based on Transfer Learning

Florance Yune

University of Computer Studies, Yangon

A thesis submitted to the University of Computer Studies, Yangon in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

June, 2024

University of Computer Studies, Yangon

This is to certify that we have examined the thesis entitled "Myanmar-Wa Machine Translation based on Transfer Learning", submitted by MS. FLORANCE YUNE, Roll Number: 13Ph.D. - 19, 2020-2024, in fulfillment of the requirements for the degree of Doctor of Philosophy in Information Technology and we recommend it to the Steering Committee for the Ph.D. Programme for acceptance.

Board of Examiners:

1. Dr. Mie Mie Khin
   Rector ____________________
   University of Computer Studies, Yangon (Chairperson)

2. Dr. Khin Mar Soe
   Professor ____________________
   University of Computer Studies, Yangon (Supervisor)

3. Dr. Win Le' Le' Phyu
   Professor ____________________
   University of Computer Studies, Yangon (Member)

4. Dr. Aye Mya Hlaing
   Associate Professor ____________________
   University of Computer Studies, Yangon (Member)

5. Dr. Thinn Thinn Wai
   Professor ____________________
   University of Information Technology (External Examiner)

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

…..…………………………… ………….......…………………………

Date Florance Yune

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to the Almighty God for His boundless grace and guidance throughout this research endeavor.

I am profoundly indebted to His Excellency, the Minister for Science and Technology, for the unwavering support provided throughout my Ph.D. journey at the University of Computer Studies, Yangon, ensuring access to essential facilities.

Special appreciation is extended to Dr. Mie Mie Khin, the esteemed Rector of UCSY, for her institutional support and invaluable guidance, which laid the foundation for this research.

Heartfelt thanks to Dr. Win Lelt Lelt Phyu, esteemed Professor at UCSY and class coordinator for the 13 Ph.D. - 19 batch, for her academic guidance and unwavering support. Her dedication to her students and commitment to excellence have been a constant source of inspiration.

Continuing on, I am deeply grateful to Dr. Khin Mar Soe, my nurturing mentor and steadfast guide, whose unwavering wisdom and encouragement propelled me forward in my academic pursuits.

I express my sincere appreciation to Dr. Win Pa Pa, Professor at the NLP lab (UCSY). Dr. Win Pa Pa's insightful feedback and dedicated mentorship were instrumental in refining the study's methodologies and enhancing the overall depth and precision of the findings.

I extend my deepest gratitude to Dr. Aye Mya Hlaing for her expert contributions. Her extensive knowledge, guidance, and support were instrumental in shaping the direction and depth of this research work.

I am also thankful to Daw Aye Aye Khine, Professor and Head of English Department, for her invaluable support in refining the language and ensuring correctness throughout the Ph.D. coursework and dissertation.

Additionally, I extend my sincere thanks to Dr. Thinn Thinn Wai, external examiner and Professor at University of Information Technology, for her invaluable feedback and critical insights during the examination of this thesis.

I am also profoundly grateful to Dr. Yi Mon Shwe Sin for her expert contributions. Her thoughtful input and detailed analyses provided significant depth to the research.

Special recognition is extended to Mr. AI MHAWM KHAM for his expertise in language translation, which enhanced the accessibility of this research.

Family support has been my rock throughout this journey, and I am profoundly grateful for their unwavering encouragement. I am deeply appreciative of my sister's consistent emotional and financial backing, as well as the unwavering support of “Lwin Po Po Maung” who is like a sister to me.

My respectful Professors for their invaluable lectures and knowledge shared during the Ph.D. coursework. My friends from Ph.D. 13th Batch for their cooperation and encouragement.

I am grateful for the collaborative contributions of my colleagues at UCSY NLP Lab, whose insights and support enriched this study. I would also like to express my gratitude to all contributors, whose collective efforts have contributed to the success of this research endeavor.

In Memoriam: We mourn the passing of Dr. Mie Mie Thet Thwin, the esteemed Rector of UCSY, whose legacy of leadership and commitment to academic excellence will forever be cherished.

I extend my sincere gratitude to Pastor Rev. AI NAP TAO, whose guidance and mentorship have been invaluable to me. He not only serves as my spiritual guide but also as a father figure, offering unwavering support for my endeavors, both spiritually and financially. Additionally, I appreciate the spiritual encouragement provided by Associate Pastor Rev. AI KHWAT PAN and Rev. KYAW NAING WIN, which has been a source of strength and inspiration for me.

Lastly, I express profound gratitude to U Saw Lu, the esteemed founder of the "Wa" social association, and notably, my grandfather. Although he is no longer with us to witness the outcome of my research, his unwavering support and encouragement have ignited and sustained my scholarly endeavors, instilling in me the motivation and vigor needed to pursue academic inquiry.

ABSTRACT

Artificial Intelligence (AI) is rapidly changing the world, affecting every aspect of human daily lives. Natural Language Processing (NLP) is itself a broad field that lies under AI. NLP depends upon linguistics and is responsible for making computers understand text and spoken words the same way humans do. NLP combines rule-based modeling concepts for human speech language, computational linguistics with some statistics, Machine Learning, and Deep Learning to enable computers to understand human speech language, which can be in the form of text or voice data.

Machine Translation (MT), translates meaningful text from one specific language to another language without human involvement. The research area of NLP in machine translation has been a significant advancement in the research area of AI. NLP allows computers to comprehend, analyze, and generate human language in a way that's more organic and contextual. It involves several subtasks such as part-of-speech tagging, sentiment analysis, named entity recognition, and more. These are applied in various stages of the translation process, augmenting the understanding of the specific source language and the generation of the target language. Incorporating NLP into machine translation has enhanced its capabilities and has led to the creation of more sophisticated translation models. However, it is important to note that NLP-based machine translation is still a developing field. Challenges such as handling low-resource languages, maintaining the source text's style and tone in the translated version, and understanding cultural references and idioms still persist.

Transfer learning, although its exact nature is unknown, enhances the quality of machine translation for low-resource systems. While there are many advantages to transfer learning, the three primary ones are reduced training time, improved neural network performance (for the most part), and reduced data requirements. The main problem of some Machine Translation (MT) systems is the need for a large range of parallel resource data for source-to-target language translation. To overcome this problem, previous research has shown that pivoting, if the pivot language is closely connected to the source and target language pair, produces translations of higher quality. In this exploration, pivot transfer learning-based MT is applied for the translation from Myanmar language to Wa language using English as the pivot language (Myanmar-English and English-Wa).

A critical component of this exploration is the implementation of tokenizer fusion strategies, which combine the strengths of different tokenizers. Specifically, the fusion of the MyanBERTa Tokenizer and the facebook/bart-base Tokenizer plays a pivotal role. Correct tokenization is essential for machine translation as it directly impacts the model's ability to understand and generate language accurately. By effectively merging these tokenizers, the model can better handle linguistic nuances, leading to more precise and contextually appropriate translations. This approach underscores the importance of tokenizer selection and fusion in enhancing the performance and effectiveness of machine translation systems.

The BLEU (BiLingual Evaluation Understudy) score is used to evaluate machine translation. A metric for automatically assessing text translated by machines is called BLEU. The machine translation is rated from 0 to 1, with a higher score indicating better quality. Using the BiLingual Evaluation Understudy (BLEU) scores, this research performed experimental analysis on three major baseline approaches (Transformer-based Neural MT, Traditional Transfer Learning using T5, Pivot-based Transfer Learning(using a bridge language)). The experimental results of the proposed model demonstrate that transfer learning with language-specific tokenizers achieves the best BLEU scores for Myanmar-Wa translation.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The research area of NLP, which studies how well computers comprehend natural language, has advanced quickly in recent years as a result of improved algorithms, more data, and processing capacity. It serves as the foundation for a number of use cases, including conversational bots, Machine Translation and opinion mining. Traditionally, rule-based systems—basically, a collection of carefully designed rules that dictated the system's behavior—were used to accomplish NLP tasks. One example is rule-based Machine Translation, in which linguists create new rules iteratively to improve the accuracy of the translations. NLP models were traditionally trained after the model parameters, also known as weights, were randomly initialized. Deep learning models were able to converge more quickly and with comparatively less fine-tuning data requirements thanks to the process known as transfer learning, in which a neural network is fine-tuned on a particular task after being pre-trained on a generic task. In the past, transfer learning has mostly been related to optimizing deep neural networks for different computer vision tasks that were trained on the ImageNet dataset [49]. However, transfer learning can now be done in this field as well because transfer learning leverages pre-existing knowledge from large-scale datasets and this knowledge can be transferred to specific translation tasks, even when limited data is available.

Transfer learning is a potent machine learning approach that can be model to apply the knowledge they have learnt from one task to the another related activity in order to perform better. Transfer learning has the potential to decrease requirements while also considerably improving the highest-quality translations when used in Machine Translation for highly expensive training on a significant number of concurrent datasets. Parallel corpora were applied to train traditional Machine Translation systems, such as Statistical Machine Translation (SMT) and earlier Neural Machine Translation (NMT) models, to translate language text from one language to another. Aligned sentences in the source language and target languages were gathered using these parallel corpora. Through a sophisticated series of neural or statistical transformations, the models are trained to map text input sentence in source language to output text in the target language. However, there are a number of drawbacks to these earlier conventional Machine Translation methods. For every language pair, they needed enormous volumes of parallel data, which might not be available for all topics or languages.

Furthermore, using the scores from each language pair to build Neural Machine Translation models can be both numerically valuable and taxing. Transfer learning can be used to address these issues by utilizing pre-trained models as the foundation for Machine Translation on a related activity, like building a language model or other translation work. The massive volumes of data have taught the pre-trained models grammar, contextual comprehension, and linguistic traits, making them a useful tool for more modern jobs like translation.

It is impossible to overestimate the crucial role Machine Translation plays in removing language obstacles in the context of the growing globalization period [47]. Machine Translation systems have developed to enable across linguistic communication and cultural barriers, making them vital tools for promoting global collaboration, information sharing, and cross-cultural interactions. The purpose of this work is to examine the evolution of Machine Translation systems historically, paying particular attention to how these systems were adapted to the complex linguistic context of the Myanmar-Wa corpus. The rapid pace of globalization, which is defined by the smooth ideas flow, business and information through the international borders, emphasizes the need for accurate and effective language translation [54]. Driven by advances in computational linguistics and artificial intelligence, Machine Translation has become a key component of the global communication paradigm. As demonstrated by the Wa language in the Myanmar-Wa corpus, it has not only overcome the limitations of traditional translation but also made it possible to bridge languages that were previously thought to be difficult owing to scarce resources.

The techniques that follow are set up to achieve the goals of this investigation. The study begins with a previous review of Machine Translation approaches, covering its origins and early difficulties. It then explores the unique features of the Myanmar-Wa corpus, such as its linguistic features and cultural relevance, as well as the difficulties involved in gathering data. This study then examines the early attempts to apply Machine Translation to the Myanmar-Wa corpus, explaining the shortcomings and constraints of those efforts. Additionally, it looks into cutting-edge methods and technical developments that have improved Machine Translation approach for the particular language pair.

The study of the next portion delves into linguistic analysis, breaking down the distinctive qualities of the Wa language and looking at the metrics and processes used to assess the effectiveness of the Machine Translation system. This analysis highlights the crucial role that better Machine Translation plays in cultural preservation and the greater goal of language revitalization, focusing on the practical applications and implications of enhanced Machine Translation for the Myanmar-Wa language pair. The analysis of the system is concluded with a summary of the major discoveries, a discussion of their implications, and final reflections on the dynamic field of Machine Translation and its critical role in bridging linguistic barriers.

1.1 Machine Translation in Natural Language Processing

Within the field of computational linguistics, Machine Translation (Machine Translation) employs computer algorithms to translate speech or text between languages without the need for human intervention. The objective of Machine Translation is to achieve cost-effectiveness, minimal mistakes, and reasonably high accuracy. Given the vast array of intricate natural languages that exist today, Machine Translation is an extremely significant yet challenging task. The digital revolution has resulted in an explosion of data in the modern world, and as language is the most efficient means of human communication, there is a growing need for NLP tools and other methods of language translation.

In Machine Translation, the language you intend to translate content into is termed the target language, while the original text is designated as the source language. Machine Translation operates through a straightforward two-step process: decoding the meaning of the original text in the source language and encoding that meaning into the target language.

The need for translation services has grown significantly as a result of the exponential development in information sharing across different regions in distinct regional languages. Because there are many words with different meanings, sentences with multiple possible interpretations, and certain grammatical relations in one language could not exist in another, Machine Translation from one language to another using NLP methods is notoriously difficult. The typical flow of the Machine Translation process from Myanmar to Wa is depicted in Figure 1.1.

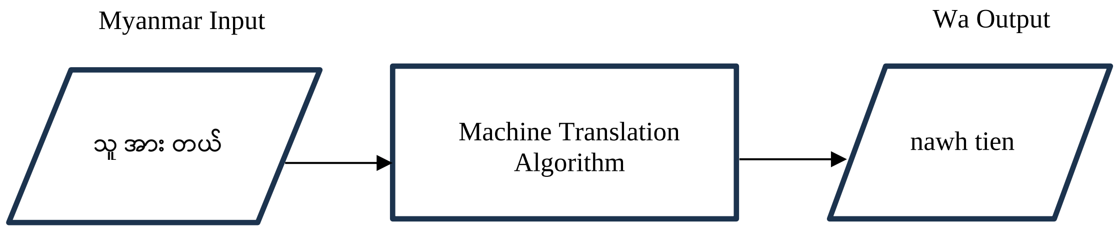


Figure 1.1 Machine Translation for Myanmar-Wa

There are four main types of Machine Translation. They are Statistical Machine Translation (SMT), Rule-based Machine Translation (RBMT), Hybrid Machine Translation (HMT) and Neural Machine Translation (NMT).

- Statistical Machine Translation (SMT)

  SMT operates by making references to statistical models that rely on the analysis of enormous amounts of multilingual material. It seeks to determine whether two terms: one from the target language and one from the source language correspond.

- Rule-based Machine Translation (RBMT)

  The fundamentals of grammar rules are essentially translated by RBMT. To construct the translated sentence, it guides a grammatical analysis of both the target and source languages. Nevertheless, RBMT requires a great deal of editing, and its strong reliance on dictionaries implies that mastery takes some time.

- Hybrid Machine Translation (HMT)

  HMT is clearly more successful in terms of quality because it creates the use of a translation memory. However, even HMT has many disadvantages, the common significant of which is the need for extensive editing and the additional requirement for human translators. HMT is approached in a variety of ways, including confidence-based, generation of statistical rule, multi-engine, and multi-pass.

- Neural Machine Translation (NMT)

  Neural network models, which are modeled after the human brain, are used in Neural Machine Translation (NMT) to create statistical models that are then translated. One of NMT's main advantages is that it offers a single system that is capable of decoding both source and target text. As a result, it is independent of certain mechanisms that are common to SMT and other Machine Translation systems.

1.2 Motivation of the Research

A major issue with some Machine Translation (Machine Translation) systems is that they require a lot of concurrent data resources to translate languages, either source or target. In order to solve this issue, prior studies have demonstrated that, in cases when the bridge language is closely linked to the source and target language pair, rotating the Machine Translation system improves the quality of the translation. Innovations in Machine Translation include cross-lingual pre-training, transfer learning techniques, and low-resource language modifications to pre-existing models. Transfer learning has several benefits, but the three main ones are shorter training times, generally better neural network performance, and less data usage.

The difficulties were made worse by the Wa language's poor resource status, which is marked by odd grammar and a limited vocabulary. Wa's nuances were difficult for conventional Machine Translation models to translate into because they were made for languages with more resources [52]. Additional challenges included linguistic differences between Wa and Myanmar, which included word order, syntax, and vocabulary. These resulted in less than ideal translation quality. The necessity for innovative solutions was highlighted by the fact that early attempts sometimes ignored cultural allusions and nuances, which are essential to accurate translation. In order to get around these obstacles, academics and business experts investigated creative solutions that made use of state-of-the-art NLP techniques and modified pre-existing models to fit the Myanmar-Wa language combination. In this system, tokenizer fusion of pretrained models based Machine Translation are applied for the translation Myanmar Language to Wa Langue.

1.3 Objectives of the Research

A huge parallel corpus is one of the key prerequisites for the neural machine translation systems to operate well. Good parallel corpuses of this size are not available for language pairings that are remote, low resource. This creates a significant obstacle in the development of high-quality Machine Translation for low resource language pairs. Enhancing the quality of Machine Translation for remote, low resource languages is the aim of this work.

1.4 Contribution of the Research

This study delves into the realm of transfer learning, aiming to harness its potential in bolstering the efficacy of Machine Translation (Machine Translation), particularly in contexts involving non-English language pairs such as Myanmar-Wa. While much of Machine Translation research traditionally concentrates on language pairs involving English, owing to the abundance of parallel corpora, our investigation shifts the focus towards languages beyond this predominant domain.

A cornerstone of our methodology lies in the innovative application of tokenizer fusion strategies, which serve as a linchpin in augmenting the quality of translation outputs. By underscoring the pivotal role of tokenizer fusion strategies and leveraging insights from transfer learning, our research makes significant strides in advancing Machine Translation capabilities for language pairs that have been historically underrepresented. This not only facilitates improved translation accuracy but also opens up new avenues for fostering cross-lingual communication and mutual comprehension.

1.5 Organization of the Research

This dissertation is structured into six chapters. The first chapter introduces Machine Translation along with the study's objectives and contributions. Chapter 2 provides a survey of current approaches, discusses relevant literature, and examines various tagging formats. Chapter 3 offers an explanation of the features of the Myanmar language, including the proposed joint word segmentation and Part-of-Speech Tagger, the construction of a corpus with POS tags, and the establishment of morphological rules for post-processing. Chapter 4 outlines the proposed model for translating Myanmar-Wa. In Chapter 5, Machine Translation based on Transfer Learning is expounded upon, along with an assessment of trial outcomes using BLEU score measurements for Myanmar-Wa translation. Finally, Chapter 6 summarizes the findings of this research study and suggests potential directions for further research.

CHAPTER 2

LITERTATURE REVIEW AND RELATED WORK

One of the most challenging and challenging tasks in the literature on Natural Language Processing is Machine Translation. In scenarios where Statistical Machine Translation (SMT) is used, the issue is handled by computing the probability of the translation model for the language pair and the target language model independently. The translation model must also handle issues with alignment and other things. However, using a broad, end-to-end model to do this task is now achievable thanks to Neural Machine Translation (NMT). The review will now focus on certain attention-based neural Machine Translation research methodologies. The relevant earlier efforts on this subject are thoroughly explained in this chapter.

2.1 Languages

In a Machine Translation system, the languages involved play a crucial role. The language being translated into is termed the target language, while the original language or text is referred to as the source language. The process of Machine Translation generally involves two fundamental steps: decoding the meaning of the original text in the source language and encoding this meaning into the target language.

In this proposed model, Myanmar language or text is used as the source language and Wa language (a language spoken mostly in northern Burma as well as in neighboring China and Thailand, and part of the Mon-Khmer family's Palaungic branch) is defined as target language of the proposed model.

2.1.1 Myanmar Language

A Myanmar Language text is a character string without exact boundary markup word, written from left to right sequence without constant spacing for inter word, whereas spacing of inter-phrase may be applied periodically. There are three groups of Myanmar characters: medial, vowels and consonants. The basic Myanmar consonants can be aggregated by medial. Words or Syllables of Myanmar text are organized by combination of consonants and vowels. At the same time, some syllables can be formed by only consonants without using any vowels. The special characters, signs, punctuation and numerals were included as some characters of Myanmar text.

The Union of Myanmar's official language is the Myanmar language, usually referred to as Burmese, and it dates back more than a thousand years. The Burmese script is used to write the tonal and analytical language of Burma. An Indian (Brahmi) prototype served as the basis for this phonologically based script, which was developed from Mon. A Myanmar text is a collection of characters written left to right without clear word boundary marking and without the usual inter-word space, however inter-phrase spacing may occasionally be utilized. Consonants, medial, and vowels are the three categories into which characters from Myanmar may be divided that are shown on the following table 2.1, 2.2, 2.3 and 2.4 respectively.

Table 2.1 Basic Consonants of Myanmar Language

Basic Consonants (ဗျည်းများ)				
က	ခ	ဂ	ဃ	င
စ	ဆ	ဇ	ဈ	ည
ဋ	ဌ	ဍ	ဎ	ဏ
တ	ထ	ဒ	ဓ	န
ပ	ဖ	ဗ	ဘ	မ
ယ	ရ	လ	ဝ	သ
	ဟ	ဠ	အ	

Table 2.2 Vowels of Myanmar Language

Vowels (သရများ)				
ေ◌	◌ာ	◌ု	◌ူ	◌ဲ
	◌ိ	◌ီ		

Table 2.3 Medial of Myanmar Language

Medials (ဗျည်းတွဲများ)			
◌ျ	ြ◌	◌ွ	◌ှ

Table 2.4 Special Characters of Myanmar Language

Special Characters				
၌	၍	၏	ဪ	ဤ

2.1.2 Wa Language

Wa, one language from the Palaungic branch of the Mon-Khmer family, which is spoken by around 950,000 people, predominantly in northern Myanmar and adjacent China and Thailand. Wa comes in three different forms: Parauk, Vo, and Awa, each of which has a variety of dialects and is occasionally thought of as a separate language.

The Parauk language, also known as Baroag, Phalok, Praok, Standard Wa, or Wa, is spoken by about 400,000 people in Burma. Most of these individuals' hail from the Southeast, East, and Northeast Shan States. The Parauk language is spoken in the southwest of China.

The Mon-Khmer language family includes the Wa language, which presents machine translation with a number of linguistic nuances, opportunities, and challenges. A detailed analysis of these distinguishing characteristics is essential to a sophisticated comprehension of Wa-Myanmar translation.

One key trait is word order flexibility: Wa exhibits a significant amount of word order diversity by displaying both Verb-Subject-Object (VSO) and Subject-Verb-Object (SVO) structures. A thorough comprehension of context is essential for successful translation [38]. Another important aspect is Wa's tonal complexity; as a tonal language, meaning is conveyed through intonation and pitch changes. Accurately interpreting these tonal subtleties is crucial to maintaining intended meaning in translation [26].

Wa's agglutinative morphology also adds complexity, as it uses affixes to modify root words to convey various grammatical functions. Knowledge of these affixes is necessary for accurate translation [59]. Additionally, Wa verbs and adjectives frequently undergo nominalization, which is vital for precise translation.

Lastly, the cultural nuances embedded in the Wa language are integral to Wa culture and identity. Terms and phrases with cultural significance require extra care in translations to preserve cultural integrity. Table 2.5 shows the consonants of the Wa language, whereas Table 2.6 shows its vowels.

Table 2.5 Consonants of Wa-Language.

Consonants	Pronunciation
K	ကာ့
KH	ခါ့
NG	ငါ့
S	စာ့
SH	ရှ
NY	ည
T	တာ့
TH	ထာ့
N	နာ့
P	ပါ့
PH	ဖာ့
M	မာ့
Y	ယာ့
R	ရာ့
L	လာ့
V	ဗာ့
W	ဝါ့
H	ဟာ့
X	အာ့
C	ကျ
CH	ချ
D	ဒါ့
G	ဂ
Q	ဂါ့
B	ဘ
F	ဖာ့
J	ဂျ
Z	ဇ

Table 2.6 Vowels of Wa-Language

Vowels	Pronunciation
A	အာ
E	အေ
IE	အဲ
AW	အော်
OI	အွဲ
AO	အောင်း
AU	အော့ပ်
AI	အိုင်

I	အီ
O	အို
U	အူ
EE	အေီး
EU	အေား

2.2 Segmentation on Nature of Language

The Myanmar language, also known as Burmese, which is the official Myanmar language (formerly known as Burma) and is spoken by a significant portion of the population. Unlike languages with spaces between words, Myanmar script lacks explicit word boundaries, making tokenization a non-trivial task. Standard tokenization strategies can misinterpret these script intricacies, leading to improper segmentation and adversely affecting translation quality. Myanmar lacks extensive linguistic resources and large-scale parallel corpora for training NMT models. This scarcity makes it crucial to optimize the utilization of available data by incorporating language-specific insights, such as proper segmentation.

The syllable structure of the Myanmar language is an essential aspect of its linguistic nature. The language's syllabic structure is characterized by the arrangement of consonants, vowels, and tone markers within a syllable. A typical Myanmar syllable consists of one or more consonants, followed by a vowel or a combination of vowels, and potentially a tone marker. Syllable segmentation takes into account the context of the text, including the arrangement of characters within words and sentences. Understanding this context is important for correct segmentation, especially in cases where characters may have different forms depending on their position within a word.

Phase structure segmentation, also known as phrase structure parsing, involves dividing a sentence into its constituent phrases or grammatical units, which include phrases like verb phrases, noun phrases, prepositional phrases and more. This type of segmentation provides insights into the syntactic structure of the sentence and is a crucial step in understanding the grammatical relationships between words.

Word-based segmentation of the Myanmar language script involves dividing text into individual words or lexemes. Unlike languages that use spaces to delineate words, Myanmar script requires specific methods to identify word boundaries due to the absence of spaces between words. Proper word segmentation is essential for the various NLP tasks, including Machine Translation, text analysis, and information retrieval. Developing an effective word-based segmentation approach for Myanmar script requires a combination of linguistic insights and computational techniques. It is crucial to consider the unique characteristics of Myanmar script, such as agglutinative morphology and tone marks, when designing segmentation algorithms. Proper word segmentation is a foundational step toward enabling accurate language processing and understanding in the context of the Myanmar language.

2.2.1 Word Level Style Myanmar Language Segmentation

A fundamental task and a significant issue in natural language processing is word segmentation. For certain languages lacking word separators, the boundaries of words must be deduced from the fundamental character sequence rather than being indicated by white space. The majority of research on this topic has been concentrated on segmenting Asian languages, for which the traditional, cutting-edge method employing conditional random fields has produced adequate results.

Similar to other Asian languages, word segmentation is not an easy operation for Myanmar texts since, unlike English, there is no white space to indicate word boundaries. Since it is the first stage in the processing of language, it is also vital to all languages. Allowing several accurate segmentations of the same text might also be required, contingent on the demands of additional Natural Language Processing stages, as Myanmar to other language Machine Translation. High-level language analysis, such as syntactic parsing and name entity recognition, which are employed in numerous NLP applications like Machine Translation, require it.

Formally, the practice of introducing gaps into textual material without changing or rewriting it is known as Burmese word segmentation. Because syllables in Burmese are unbreakable writing units, all algorithms investigated in this note first apply a syllable segmentation process to input, which involves inserting spaces into syllable borders, before deciding how the syllables form words, which involves deleting gaps between syllables. The job becomes a binary classification issue for syllables depending on whether the next space is to be eliminated using a typical machine learning technique.

2.2.2 Syllable Level Style Myanmar Language Segmentation

White space may occasionally be placed between phrases in Myanmar scripts, but regular white space is not used between words or between syllables. Myanmar scripts are written in sequence from left to right. In the Myanmar language, words can have one or more syllables. A Myanmar syllable can also be thought of as being made up of many characters. The 75 characters that make up the typical Myanmar scripts may be divided into 12 divisions. In the Myanmar language, a syllable might consist of one or more characters in addition to one or more syllables. The creation of a Myanmar syllable is extremely obvious and easy. A Myanmar syllable typically comprises of one beginning consonant followed by zero or more medials, zero or more vowels, and optional dependent varied indicators, however the elements can exist in different sequences.

Syllables are the smallest linguistic units of the Myanmar language, and a word can have one or more of them. In general, words used in Myanmar may be divided into two categories: (1) standard words (i.e., words having a conventional syllable structure) and (2) irregular words (i.e., words with condensed characters or words written in unique traditional writing styles). A crucial stage in tasks requiring natural language processing for the Myanmar language is syllable-level segmentation of the language. A syllable can comprise multiple consonants, multiple medials, multiple vowels and various signs. Words in Myanmar (Burmese) are often made up of one or more syllables because it is a syllable-based language.

In Myanmar writing, each syllable corresponds to a single vowel sound, a consonant, a consonant cluster, or a combination of both. The syllable boundary for each series of characters must first be divided. The appropriate syllabic order might then be combined to establish the proper word boundary. As a result, word segmentation is significantly influenced by the goal of syllable segmentation. The primary emphasis of our proposed study is the segmentation of formal and informal text at the syllable level. A fundamental unit of sound or a sound is a syllable. There may be one or more syllables in a word. A syllable is generated in Myanmar according to criteria that are quite clear cut and unambiguous. Multiple medials, multiple consonants, and multiple vowels can all be found in a same syllable. These components can show up in a variety of combinations, such as consonants, medials, vowels, additional consonants and vowels, or consonants, medials, vowels, and more consonants.

Words in Myanmar (Burmese) are often made up of one or more syllables because it is a syllable-based language. In Myanmar writing, each syllable corresponds to a single vowel sound, a consonant, a consonant cluster, or a combination of both. Syllable segmentation is crucial for a number of NLP tasks, such as tokenization and language comprehension in text-to-speech synthesis, voice recognition, and Machine Translation.

2.2.3 Wa Language Segmentation

In the Wa language, segmentation refers to the process of dividing a piece of text into individual words. This segmentation is typically done at the word level, meaning that the text is separated into discrete units representing individual words.

The segmentation process in the Wa language involves identifying boundaries between words based on whitespace. Whitespace refers to any spaces, tabs, or line breaks that separate words within the text. When writing or typing in the Wa language, authors and speakers insert spaces between words to indicate where one word ends and the next one begins.

For example, consider the sentence: "nawh tien". In this sentence, "nawh" means "သူ", "tien" means "အား တယ်". Each of these words is separated by whitespace, making it clear where one word ends and the next one begins.

Segmentation at the word level is essential for readability and comprehension in written and spoken communication. It allows readers and listeners to distinguish between individual words, understand their meanings, and interpret the overall message of the text or speech.

Overall, in the Wa language, segmentation at the word level involves dividing text into separate words based on whitespace, enabling effective communication and understanding among speakers and writers.

2.3 Reviews on Machine Translation

Machine Translation (Machine Translation) is one of the trickiest and most difficult tasks in the literature on natural language processing. The fundamental process of Machine Translation is natural language processing. Its development process has historically been nearly identical to that of Machine Translation, and the two are complementary. When statistical Machine Translation is applied, the problem is solved by separately calculating the probability of the target language model and the translation model for the language pair. The translation model must also handle issues with alignment and other things. However, using a broad, end-to-end model to do this task is now achievable thanks to neural Machine Translation (NMT). The review will now focus on certain attention-based NMT research methodologies. The relevant earlier efforts on this subject are thoroughly explained in this section.

In [30], Wang Ling et al. proposed a character-based NMT system that uses an attention model on both the source and destination sides. The system eliminates the difficulties of tokenization in the preprocessing stage and is capable of comprehending and producing unseen word forms. Two datasets are implemented by the system. There are 600K sentence pairs from Europarl for training in the English-Portuguese language pair, 500 sentence pairs for development, and 500 sentence pairs for testing. The system demonstrated that its techniques could outperform comparable word-based NMT models.

In [45], To increase the performance of NMT systems, Vaswani and et al. created the Transformer design. A cascade approach can be utilized to provide translation across remote and non-English language pairings for which a big parallel corpus is not available. Two models are trained in the cascade method: a source language to English model and an English to target language model. The source sentence is then sent through the two models to convert it to the target sentence.

In [41], weaver first proposed the idea of computer-assisted language translation in a ground-breaking 1949 memo that the author reported, establishing Machine Translation as a separate field of study and providing the framework for further advancements. The foundation of Machine Translation technology was laid by early Machine Translation initiatives, most notably the Georgetown-IBM experiment in the 1950s. Even though these early algorithms were still in their infancy, they faced significant difficulties and frequently produced inaccurate and strange translations. A major obstacle was the complexity of human language, which includes quirks, context-dependency, and cultural subtleties. Machine Translation systems started to show promise as computational power and linguistic theory developed.

In [49], Jing Wu et al. used a NMT correction model together with three sub-word training techniques and monolingual data. According to attention-based NMT models, the system intends to improve low-resource Mongolian-Chinese language pairings. A phrase-based statistical Machine Translation model was also developed by the system utilizing Moses toolkits. The system trained the attention-based NMT with 1024 dimensions for word embedding, 1024 hidden units per layer for both the encoder and decoder, 50K for the source vocabulary, 10K for the target vocabulary, and 10 for beam search on the GPU of the NVIDIA Tesla K80. The Mongolian-Chinese NMT model might be improved using the suggested ways.

In [25], the author presented "Cross-lingual Language Model Pretraining" (Advances in Neural Information Processing Systems, 2019) as a noteworthy contribution to the subject. This idea focuses on cross-lingual pretraining of language models to improve the transferability of knowledge between languages. Researchers are investigating creative techniques to improve inclusivity and address the challenges faced by low-resource languages.

The historical evolution of Machine Translation systems is traced in this study, with a focus on how these systems were tailored to the Myanmar-Wa corpus. The study clarifies tactics and inventions meant to accomplish effective Wa language translation through an analysis of technology developments. Most importantly, this research contributes to more general fields like linguistics and language preservation, going beyond the field of Machine Translation. Examining the difficulties and solutions in Machine Translation adaptation for the Myanmar-Wa corpus, the work adds important knowledge to the expanding corpus on underrepresented languages. In an increasingly globalized world, this endeavor is essential to the survival and rebirth of languages.

2.4 Reviews on Transfer Learning

There is no significant agreement in the scientific literature on Machine Translation as to what size corpus is considered low-resource. However, in general, we may state that a low-resource condition occurs when the size of the parallel training corpus is insufficient to get a satisfactory outcome using the conventional Machine Translation methods. This is typically evaluated using the BLEU standard automatic evaluation metric, which has a strong correlation with evaluations from human translators.

Conversely, low-resource Machine Translation handles corpora containing a few thousand sentences or less. Despite the initial impression given by this figure that nothing valuable can be obtained for low resource languages, even little data sets can be leveraged. Among these is a deep learning method known as transfer learning, which applies the knowledge discovered while addressing one problem to another that is related but distinct.

Comprehending the reasons behind the success of transfer learning can enhance optimal approaches and facilitate the exploration of methods to get analogous advantages without necessitating parental role models. To help you understand what information is being transferred, this section includes a number of ablation research on transfer learning.

Zoph et al. were the first to incorporate transfer learning to NMT, where only the source language is exchanged before/after the transfer. To work with multiple languages and assist target language shifts, Nguyen and Chiang, Kocmi and Bojar employ common sub word vocabularies [59] [24].

In [14], Neubig and Hu provided a thorough study on the application of multilingual models. They used Machine Translation for four low-resource languages (Azeri, Belarusian, Galician, and Slovakian) using a "massively multilingual" corpus of 58 languages. They were able to acquire a BLEU score of up to 29.1% with a parallel corpus size of just 4500 phrases for Galician, as opposed to 22.3% and 16.2% with a traditional single-language training with Statistical Machine Translation (SMT) and NMT, respectively. In situations where there are no training data available for the target language, transfer learning also permits what is known as a "zero-shot" translation. The authors report a BLEU score of 15.5% for Galician on their test set, without the model having previously seen any Galician sentences.

Kim et al. proposed further strategies in 2019 to enable NMT transfer even in the absence of common vocabulary. To the best of knowledge from the previous research, the first to offer transfer learning techniques are applied in their system that specialize in transferring a source encoder and a target decoder at the same time using a pivot language. Furthermore, they demonstrated effective zero-shot translation results for the first time using solely pivot-based NMT pre-training [20].

Gheini and May developed a general vocabulary for transfer learning to address this problem. By simultaneously training the sub-word tokens across several languages and applying Romanization to languages written in scripts other than Latin, they were able to acquire a global vocabulary. However, this global vocabulary could only be able to represent unknown languages with an extremely aggressive and perhaps ineffective sub-word segmentation [20].

Lin and colleagues (2019) conducted a grid-search including multiple parent languages to ascertain the most suitable selection criteria for the ideal parent in the context of transfer learning. Their research revealed that, in addition to language relatedness, other characteristics like corpus size may also play a role in determining which language parents are the best. The BLEU score indicates that there is typically little difference between different parents [53].

According to the earlier research, approaches like linguistically informed data mixing and transfer learning can facilitate language communication for all participants. So, transfer learning approach is applied in this research work.

2.5 Reviews on Pivot-Based Machine Translation

When there is little or no direct translation data available, there are two approaches to address the difficulty of translating between two languages: pivot-based Machine Translation and triangulation-based Machine Translation. The technique we employ to construct A-to-C and/or C-to-A Machine Translation systems without (or with minimal) parallel data of the A-C language pair is called pivot Machine Translation. If there exist sizable A-B and B-C parallel corpora that can be utilized, a "pivot" language B could be used to assist in the development of A-C Machine Translation systems [42]. Translation is completed in two stages when using pivot-based Machine Translation: from source language (SL) to pivot language (PL) and from pivot language (PL) to target language (TL). Translations of the source text are first made into a pivot language, which is subsequently translated into the target language. Two language pairs—SL to PL and PL to TL—need simultaneous data. The quality of the pivot language translation has a significant impact on the pivot-based translation's quality. The quality of the translation can be lowered by this method's sensitivity to faults introduced in both translation steps. The earlier studies on pivot-based Machine Translation are provided in this section. A pivot-based Machine Translation system forms the foundation of this system.

The importance of pivot language selection for SMT was underlined by the authors of [35]. This project aimed to provide data to facilitate future studies on Machine Translation between language pairings with limited resources. As a result, the writers looked at the possibility of employing pivot languages other than English. The results of an experiment using SMT techniques to translate 12 languages demonstrate that the translation quality of 61 out of 110 language pairings improved when a non-English pivot language was utilized. The results demonstrate that the pivot source language and pivot target language translation performance have a significant influence on the optimal pivot language, especially for small training corpora. Additionally, the results demonstrate that pivot languages with a strong connection to the source language perform better overall in pivot translation than pivot languages with a weak connection to the destination language.

The authors of [18] provided a novel approach to enhancing pivot-based SMT through the use of machine learning. For language couples with little bilingual data, pivot-based SMT can construct source-target translations by using a different language as a "bridge." However, if the corresponding source and target phrases are linked to different pivot phrases, then certain suitable source-target translations cannot be produced. To get over this problem, they used Markov random walks to link likely translated terms in both the source and destination languages. The method performs significantly better than the baseline system across all tasks, according to the experimental results utilizing spoken language, web data, and European Parliament data.

In order to enable the translation of language pairings with no resources, including pivot-based and multilingual translations, a great deal of research has been done on the NMT [14]. Multilingual models generally generalize better since they incorporate different languages [45]. Nevertheless, this is sometimes not a feasible form or phologically rich language due to variances in morphological complexity. Pivot-based Machine Translation is another standard technique for translating language pairings without any resources. However, when the model is trained using a pivot-based method, fluency issues occur. To address the aforementioned problems, this system proposes a Transfer Learning-based Semi-Supervised Pseudo Corpus Generation (TLSPG) approach for translating zero resource languages. TLSPG makes use of semi-supervised learning to capitalize on similarities between low and zero resource language pairings.

A pivot language, which is typically a rich-resource language, is chosen as a bridge in pivot-based techniques. The source-target translation can then be constructed by utilizing the source-pivot and pivot-target corpora and model. One method is to create a source-pivot-target model by directly combining the pivot-target and source-pivot models once they have been trained [52]. Using pivot language to create pseudo-parallel data and train the source-target model is another popular technique. While Chen et al. [54] construct pseudo-parallel corpora using the source-pivot corpus and pivot-target model, Zheng et al. [56] use a pivot-source NMT model to translate the pivot language in a pivot-target parallel corpus to source language. Pseudo-parallel corpora can be created using source and destination language monolingual data in addition to parallel corpora [57]. Furthermore, another technique to make use of the pivot language is to leverage the parameters of the source-pivot and pivot target models. Kim and associates [2019b] transfer to the source target model the pivot-target model's decoder and the source-pivot model's encoder [21]. Based on cross-lingual pre-training [30], Ji et al. [35] pre-train a universal encoder for source and pivot languages, then subsequently train on pivot-target parallel data with a portion of the encoder frozen. The choice of pivot languages is crucial for pivot translation since it has a big impact on the translation's quality. Based on past information, one pivot language is typically chosen. Additionally, a learning to route (LTR) technique is available that automatically chooses one or more pivot languages for multi-hop translation [27].

To train source-to-pivot and pivot-to-target translation models, respectively, pivot-based techniques presuppose the existence of source-pivot and pivot-target parallel corpora. Creating a source-to-target phrase table by fusing source-to-pivot and pivot-to-target phrase tables is one of the most representative methods, known as the triangulation approach. A pivot-based translation mechanism is used in another exemplary approach [53]. Thus, source-to-target translation can be split into two stages: the pivot-to-target model is used to translate the source sentence to a target sentence, and the source-to-pivot model is used to translate the source sentence to a pivot sentence.

Because pivot-based techniques are easy to apply, efficient, and require little in the way of multilingual data, they are frequently employed in SMT. Johnson et al. have recently adapted pivot-based techniques to NMT and demonstrated that pivot-based NMT provides significantly better translation performance than their universal model without incremental training. On the other hand, pivot-based methods frequently experience the issue of error propagation, whereby mistakes in the translation from source to pivot will transfer to the pivot to target. The reason for this mismatch between source-pivot and pivot-target parallel corpora, which are typically unrelated or only weakly related, can be partially explained [18].

Character-based pivot translation for under-resourced languages and domains is the title of the research [41], which looked into the employment of character-level translation models using closely related pivot languages to facilitate translation from and to textual domains and languages with limited resources. These low-level models perform well even with limited training data, according to experiments. In a domain adaptation job, the method is based on movie subtitles for three language pairings and legal texts for a fourth language pair. The pivot translations considerably outperformed the baselines.

To enhance pivot-based SMT, the authors of [58] put forth a unique strategy that makes use of a machine learning technique. Using a different language as a "bridge" to provide source-target translation is one method for pivot-based SMT that may be used for language pairs with little bilingual data. However, if the relevant source phrase and target phrase relate to distinct pivot phrases, then certain beneficial source-target translations cannot be constructed. They employed Markov random walks to link potential translated phrases across the source and target languages in order to tackle this challenge. Based on spoken language, web data, and the European Parliament, the experimental results show that the method greatly outperforms the baseline system on all tasks.

The scenario is made worse by the independent training of the pivot-to-target and source-to-pivot translation models, which widens the linguistic divide between the source and target languages.

2.6 Reviews on Myanmar Language Translation

This section examines earlier studies on statistical Machine Translation from English to Myanmar. There has been some research on the SMT of the Myanmar language up to this point.

The first Hierarchical Phrase-Based Statistical Machine Translation, Phrase-Based Statistical Machine Translation, Open-Source Machine Translation evaluations between Myanmar and Rakhine were made possible by Thazin Myint Oo et al. (2018)

in relation to dialects of the Myanmar language. In order to examine the behavior of a dialectal Myanmar-Rakhine Machine Translation, the experiment made use of an 18K parallel corpus. The findings demonstrated that even with the restricted data, higher BLEU (57.88 for Myanmar-Rakhine and 60.86 for Rakhine Myanmar) and RIBES (0.9085 for Myanmar-Rakhine and 0.9239 for Rakhine-Myanmar) scores could be obtained for the Rakhine-Myanmar language pair [39].

The initial SMT assessments between the language pairs of Dawei (Tavoyan) and Myanmar were also contributed by Thazin Myint Oo et al. (2019). Higher BLEU (21.70 for Myanmar-Dawei and 29.56 for Dawei Myanmar) and RIBES (0.78 for Myanmar-Dawei and 0.82 for Dawei-Myanmar) scores were attained using the OSM technique, according to the SMT results using the created 9K Myanmar-Dawei parallel corpus [38]. PBSMT, HPBSMT, and OSM were used in this paper's Machine Translation tests between the language pairs of Kayah and Myanmar, which were based on the experimental findings of earlier publications.

In [35], the statistical Myanmar phrase translation system with morphological analysis was described by Thet Thet Zin et al. (2011). The 13,042 total amount of data used for these trials consisted of 12,827 parallel sentences for training and 215 additional sentences for testing. Additionally, the translation likelihood for converting Myanmar sentences into English phrases was reformulated using Bayes' rule. F-measure, precision, and recall were the Machine Translation evaluation criteria. In the original baseline system, there were issues with a large number of out-of-vocabulary (OOV) words, including proper nouns, nouns, and verb phrases. In order to solve the aforementioned OOV issue, morphological analysis is used in the second step of the translation process on the pre-processing phrase. The morphological analysis method achieved a good comparison with the baseline, based on the results. Nonetheless, the majority of post-positional marker errors continued to have unclear meaning. Consequently, the part-speech (POS) tagging technique was used as one solution to that issue. For the greatest outcomes and lowest OOV rates, the baseline system was enhanced with the morphology and POS of the Myanmar language. However, there were still 95 mistakes in 215 examined sentences, including segmentation errors, untranslatable phrases, unknown foreign terms, translation failures, and missing English particles.

Ye Kyaw Thu and colleagues (2016) provided the first comprehensive analysis of the translation of the Myanmar language in [54]. The study included 40 language pairs that included languages that were fundamentally different from Myanmar as well as those that were related to it. 457,249 sentences were utilized in this experiment for training, 5,000 for development, and 3,000 for evaluation. Based on the BLEU [23] and RIBES scores [42], the results indicated that the hierarchical phrase-based SMT (HPBSMT) [41] technique produced the best translation quality.

The first comparison analysis of five popular Machine Translation techniques used with low-resource languages was done by Win Pa Pa et al. (2016) in [51]. Limited amounts of travel domain data were translated in both directions between English and {Thai, Laos, Myanmar} using PBSMT, HPBSMT, tree-to-string (T2S), string-to-tree (S2T), and Operation Sequence Model (OSM) translation techniques. In this instance, 20,000 sentences were used for training, 500 for development, and 300 for assessment. The trial findings showed that the PBSMT approach produced the greatest quality translations in terms of appropriateness (as determined by BLEU score). Here, the S2T and T2S experiments exclusively employ the annotated tree for the English language. This is due to the lack of a tree parser for Thai, Lao, and Myanmar languages that is available to the general public. We found that the OSM technique produced the best Machine Translation results for translating from Myanmar to English based on their RIBES scores.

In [24], Sari Dewi Budiwati et al. examined different setups for the multiple pivots of four phrase tables on Japanese-Indonesian, one of the SMT's resource-poor language pairs. In a single pivot translation, the pivot languages were English, Myanmar, Malay, and Filipino. There were four pivot techniques used: fill-up interpolation, cascade, linear interpolation, and triangulation. When using multiple pivot techniques, the term tables of the best pivot approaches were employed in the subsequent phase, which involved combining multiple pivots, after the resulting BLEU of each approach was first taken into account in a single pivot. Since there isn't yet a corpus available in the Burmese dialect, it is challenging to take into account several pivot translation techniques. Furthermore, compared to the single pivot approach, the multiple pivot translation method incurs higher computing costs.

Pivot-based Machine Translation systems have produced better translation results in previous Machine Translation system experiments. For the Myanmar-Wa

Machine Translation system to achieve high BLEU scores, the transfer learning strategy serves as the pivotal transfer learning mechanism.

2.7 Chapter Summary

In this chapter, detailed description of the Myanmar Language and Wa Language are presented. The basic consonants and vowels of Myanmar Language and Wa Language are shown in detail using tables. Word Level Style Myanmar Language Segmentation is described in this chapter to support the Myanmar Language Translation. And then, literature reviews on Machine Translation, Transfer Learning, Pivot-Based Machine Translation are discussed. Moreover, the related works on Myanmar Language translation are described in this chapter.

CHAPTER 3

BACKGROUND THEORY

Technology advances to new heights when it moves from the conception of concepts to their widespread application in real-world situations. One such path aims to remove language barriers in order to create social communication that is seamless across all domains. In this sense, the development of pertinent domains including machine learning, artificial intelligence, and natural language processing (NLP), along with AI-based language modeling (LM), is crucial to the creation of a faultless autonomous Machine Translation system [58]. It remains difficult to achieve the necessary fluency, sufficiency, accent, and overall accuracy even with a variety of heuristic approaches to preserve both lexical and contextual interpretation of the source language(s) onto the translated target language(s). With the development of contemporary NLP (artificial intelligence) techniques, it is now possible to train an effective translation system using a parallel corpus of translation pairs in both the source and target languages that is both high-quality and well resourced (i.e., a big number of corpora available). For high resource languages with a large global digital footprint, Machine Translation systems can be highly effective when properly trained. Conversely, low-resource languages with no digital presence and widespread awareness have significant challenges. When there are low-resource languages present in the destination or source language, this imbalance frequently results in poor-quality translation. Consequently, Machine Translation systems must comprehend the morphology (rules to cover morphemes, the smallest meaningful components, into words), semantics (meaning of words and combinations), and syntax (rules to combine words) of such low-resource languages.

Rule-based Machine Translation (RBMT), example-based Machine Translation (EBMachine Translation), statistical Machine Translation (SMT), and neural Machine Translation (NMT) systems are the four categories into which Machine Translation models are divided depending on heuristic paradigms. Each has benefits and drawbacks of its own. According to a set of guidelines, RBMT models define a language and how its many linguistic devices—words, phrases, and sentences—interact with one another. The three machines are hard-coded with these sets of rules and processes that are defined for a translation in a pair of languages. The target and source languages gathered from unilingual (one language), bilingual (two languages), or multilingual (more than two languages) dictionaries constitute the majority of the linguistic data utilized in an RBMT model. The model also makes use of grammar, which covers each language's morphological, semantic, and syntactic regularities.

However, because of its complexity, a well-built RBMT model is difficult to construct and requires highly skilled and professional human labor. Furthermore, languages tend to require more time and effort to solve due to their ambiguous nature, particularly in large and complicated models. It takes a lot of work to make RBMT models practical in daily life. Therefore, there is still a need for translation systems that are more effective than RBMT. Many translation examples are used in EBMachine Translation approaches. EBMachine Translation models are notable for their utilization of bilingual corpora manipulation, which involves segmenting a bilingual corpus into smaller sections, translating those segments into the target language, and then reassembling the parts to create entire translated phrases. Unlike RBMT models, they do not take into account the syntax, semantics, or morphological analysis of the source and destination languages. However, since SMT doesn't require human involvement, it performs better than RBMT and EBMachine Translation models. This method of translation involves applying a statistical learning algorithm to a sizable bilingual corpus in order to assist the computer in learning the translation. By using this technique, the machine can also translate sentences that it did not come across during testing and training. Converting an input word sequence from the source language to the target language is the aim of SMT. In fewer than 20 years, it has taken the lead in both commercial Machine Translation and university Machine Translation research.

However, a neural network (NN) is used to carry out NMT. NMT lacks a unique language model, translation model, and reordering model in contrast to SMT. Rather, it employs a single sequence model that makes decisions about individual words. The prediction is predicated on the source sentence's previously produced target language sequence. NMT is a machine learning technique based on deep learning that makes use of a large NN and word vector representations. Even though the NMT has achieved remarkable results in a few translation experiments using high-resource language, researchers are unsure if the NMT could actually replace SMT and if its success would extend to other tasks. Ultimately, the reality is brought to light by the experiment conducted by Michał (2016) on the United Nations corpus, which consists of 15 low-resource languages. It is clear from the outcome of his experiment that, in most circumstances, 4 SMT performs better than NMT, as indicated by the BLEU score. Using low resource language, numerous studies (Lohar et al., 2019; Zhou et al., 2017; Wang et al., 2017; Castilho et al., 2017) have highlighted the shortcomings of NMT over SMT, including the fact that NMT needs a larger corpus and more resources than SMT.

In this chapter, the background theory of this exploration is described as the sub-titles. About the Machine Translation in NLP and types of Machine Translation are discussed in this section. And then Transfer Learning approaches for the NLP and Transfer Learning for Transformers are presented in this chapter.

3.1 Machine Translation in Natural Language Processing

The goal of the computational linguistics subfield of Machine Translation is to create automated systems that can translate speech or text between languages. The aim of Machine Translation in Natural Language Processing (NLP) is to generate translations that accurately capture the original content's meaning while adhering to grammatical rules. In linguistics, Machine Translation is the automated use of computational models and algorithms to translate text or speech between two languages. To enable translation without the need for human participation, this technology makes use of sophisticated artificial intelligence techniques and vast databases of multilingual text. Rule-based Machine Translation and statistical Machine Translation are the two basic categories of Machine Translation techniques. While statistical Machine Translation (SMT) generates translations using statistical models, rule-based Machine Translation (RBMT) generates translations based on linguistic rules.

Every method, including SMT and RBMT, has benefits and drawbacks. When rules are carefully crafted, RBMT can yield more accurate translations; nonetheless, it is frequently challenging to create rules that address every scenario. Although SMT is less accurate than RBMT, it is significantly simpler to create and frequently yields superior results for practical uses.

The most common method for Machine Translation is based on statistical models, which are trained on large parallel corpora (collections of texts in different languages). This approach has become very popular in recent years due to the increase in available computing power and the availability of large parallel corpora. Like any technology, Machine Translation comes with its own set of challenges. One challenge is that, because Machine Translation relies on algorithms, it can be difficult to create translations that are completely accurate. Additionally, Machine Translation can be expensive and time-consuming to set up and maintain.

Finally, there can be a loss of meaning or context when translating using a machine, which can lead to mistranslations. Machine Translation is a powerful tool in natural language processing that can help bridge language barriers to create more efficient communication. It is important to understand the various methods used for Machine Translation and evaluate which technique might be best suited for your particular needs. With the right implementation, you can leverage Machine Translation technology to reduce manual labour and cost associated with traditional human-based translations as well as enable quicker access of information for global audiences. Bernard Vauquois' pyramid is depicted graphically in Figure 3.1, which displays the relative depths of intermediary representation, interlingua Machine Translation at the top, transfer-based translation in the middle, and direct translation at the bottom.

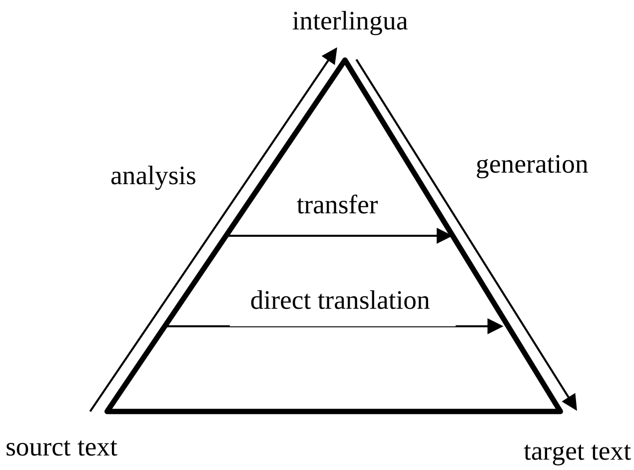


Figure 3.1 Bernard Vauquois' pyramid

These are various types of Machine Translation **in NLP:**

- Statistical Machine Translation (SMT)
- Rule-based Machine Translation (RBMT)

- Hybrid Machine Translation (HMT)
- Neural Machine Translation (NMT)

The detailed relationship between various Machine Translation techniques can be seen below:

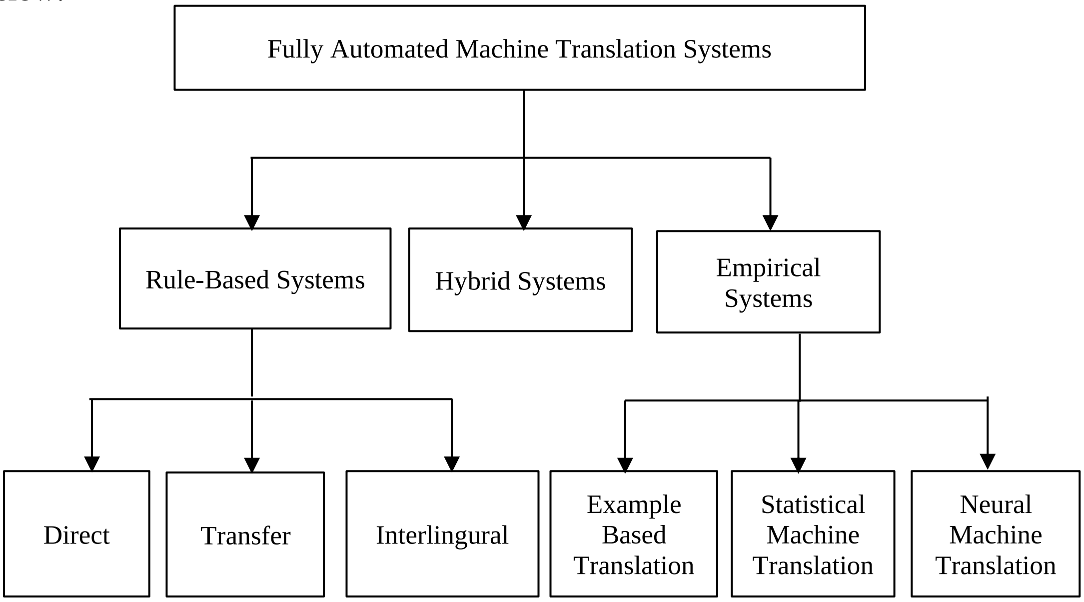


Figure 3.2 relationship between various Machine Translation techniques

Language localization for software and websites, worldwide corporate interactions, and government and diplomatic communications are just a few of the many uses for Machine Translation. Even though Machine Translation has advanced significantly in recent years, more study and research is still required to properly handle idioms and slang, effectively capture context and nuance in language, and work with low-resource languages that have little training data.

3.1.1 Statistical Machine Translation or SMT

In the 1990s, statistical Machine Translation (SMT) became the predominant Machine Translation technique. SMT systems learn and forecast the most likely translation of a given input by utilizing vast databases of parallel texts from source and target languages. Because of this, SMT systems are able to translate text more accurately than rule-based systems, despite the fact that they are still hampered by problems with idiomatic expressions and context recognition. The task of automatically translating sentences from one human language—such as French—into another—such as English—is known as statistical Machine Translation, or SMT. The terms "source" and "target" refer to the first and second languages, respectively. You could think of this process as stochastic. Several SMT variations exist, based on the translation model used. Certain methods include mapping strings to strings, while others employ trees to strings and tree-to-tree models. The fundamental concept across all of them is automatic translation, using models that are estimated using both monolingual corpora (examples of target phrases) and parallel corpora (source-target pairs).

In the 1990s, Statistical Machine Translation (SMT) became the predominant Machine Translation technique. SMT systems learn and forecast the most likely translation of a given input by utilizing vast databases of parallel texts from source and target languages. Because of this, SMT systems are able to translate text more accurately than rule-based systems, although they are still hampered by problems with idiomatic expressions and context recognition.

A Machine Translation paradigm known as "Statistical Machine Translation " (SMT) bases translations on statistical models, the parameters of which are determined by analyzing vast amounts of multilingual text data. A collection of many organized writings written in two different languages is referred to as a bilingual text orpus. Information theory, which examines the quantification, storing, and transmission of information, forms the foundation of SMT. Statistical models are built using supervised and unsupervised machine learning methods. The statistical models include well-formed sentences and statistical data on the correlation between the SL and TL. Statistical methods were used in the translation process to determine the most accurate translation of the original text.

Statistical Machine Translation uses machine learning to translate text instead of relying on linguistic rules. Large-scale human translations are analyzed by machine learning algorithms, which then find statistical trends. The software makes educated estimates when translating a new source text by taking into account the statistical probability that a given word or phrase will be connected to another in the destination language.

SMT is a Machine Translation method that determines the most likely translation for a given input by utilizing vast amounts of multilingual data. By examining the statistical connections between source texts and their previously published human translations, statistical Machine Translation systems acquire the ability to translate. The language model and the translation model are the most crucial elements in statistical Machine Translation. The output language monolingual data is used to build the language model. Based on the translation language, the language model selects the best option among the possible translations. Because the language model provides the translated text with its natural language flow, fluency in translation can be linked to it. In parallel data, the statistical Machine Translation model is trained. A table with aligned phrases and their translations is called a translation model. We refer to these expressions as n-grams. The translation model's goal is to forecast potential translations for particular input texts. Because the translation model maintains the source's meaning, it can be linked to adequacy.

To train its models, SMT uses massive amounts of multilingual text data, or "parallel corpora." Pairs of sentences or documents in the source language and their matching translations in the target language make up these corpora. SMT algorithms discover statistical correlations and patterns between words, phrases, and sentence structures in several languages by examining these multilingual texts. The probability principle forms the foundation of SMT. SMT algorithms determine which translation has the highest probability given a given source sentence by calculating the likelihood of each possible translation. With this method, SMT systems may produce translations that accurately reflect the context and idiom of the original language while also capturing its complexities.

One of the key advantages of SMT is its ability to adapt and improve over time. As more bilingual data becomes available, SMT models can be retrained to incorporate this new information, resulting in enhanced translation quality. This adaptability makes SMT a valuable tool for industries such as e-commerce, travel, and global communication, where accurate and efficient translation is essential. However, it is important to note that SMT is not without its limitations. SMT models heavily rely on the quality and quantity of the training data. If the parallel corpora used for training are limited or of poor quality, the translation output may suffer from inaccuracies and inconsistencies.

Additionally, SMT may struggle with translating rare or domain-specific terminology, as it relies on statistical patterns that may not be well-represented in the training data. To overcome these limitations, researchers and developers have been exploring various approaches to improve SMT. This includes incorporating linguistic knowledge and rule-based systems into the statistical models, as well as leveraging neural networks and deep learning techniques to enhance the translation quality.

In conclusion, SMT is a powerful technology that revolutionizes the way we communicate across different languages. By harnessing the power of statistics and probability, SMT systems can generate translations that are both accurate and fluent. While there are challenges to overcome, ongoing advancements in this field continue to push the boundaries of Machine Translation, making it an indispensable tool in our increasingly globalized world.

3.1.2 Rule-based Machine Translation or RBMT

The initial methods of Machine Translation translated text using dictionaries, rule-based systems, and linguistic principles. These systems, however, were constrained by the intricacy of language and the challenge of identifying each rule and exception in a given language. These resources are necessary for RBMT T to guarantee accurate translation of particular content. The software parses the incoming text, creates a transitional representation, and then uses dictionaries and grammar rules to translate it into the destination language.

The first commercial Machine Translation systems were called Rules Based Machine Translation (RBMT) systems. These systems are built on linguistic rules that permit words to have diverse meanings and be placed in different contexts. Large sets of linguistic rules can be applied to using RBMT technology in three stages: generation, transfer, and analysis. Human language specialists and programmers have worked hard to comprehend and map the rules between two languages in order to construct the rules. RBMT uses manually constructed translation lexicons, some of which users can modify and improve for better translation quality.

With the development of the vocabulary and user dictionaries, RBMT yields output that is reasonably predictable and gives some control. But because of word ambiguity, these improvements might occasionally result in translations that are of worse quality. They can also be exceedingly time-consuming to implement and maintain. Because translations are generated based on rules, the resultant writing style is frequently more "machine-like," and while translations can be understandable, they are frequently not fluid. This is commonly referred to as "gist" quality, where the translation's content is understandable but requires a significant amount of post-editing to fit the target audience and writing style.

Recently, RBMT developers have begun selling their products as Hybrid Machine Translation models in an effort to solve some of the shortcomings of RBMT by adding additional SMT approaches to their core RBMT technology. There are several "Hybrid Machine Translation" versions, and it is important to comprehend the advantages and disadvantages of each.

3.1.3 Hybrid Machine Translation or HMT

Hybrid Machine Translation technologies combine different Machine Translation models into one software program, using a mix of techniques to improve a single translation model's overall efficacy. Typically, rule-based and statistical Machine Translation subsystems are integrated into this process, and the final translation output is a combination of the outputs produced by each subsystem.

Certain cases lend themselves more to the use of certain Machine Translation methods. To find the best translation quality, Language Studio combines several translation technologies at once and calculates a translation quality confidence score. Compared to older Statistical Machine Translation (SMT) technology, Neural Machine Translation (NMT) often produces translations that are more fluid and natural-sounding. When the text being translated differs greatly from the content the Machine Translation engine was trained on, NMT is known to perform erratically and yield inaccurate results. It has also been demonstrated that SMT performs better on extremely brief phrases that are only one or two words long.

Every translation in Language Studio is given a confidence score, and when an NMT translation's score falls below a user-configurable threshold, it can effortlessly transition to SMT. The ultimate output is then chosen based on which translation of each technology has the highest score. This straightforward yet efficient method produces translations that are far better than those produced by employing either technology separately. Consequently, this resolves one of the primary issues with NMT that typically irritate editors and linguists who are responsible for assessing and improving the Machine Translation output before publication.

3.1.4 Neural Machine Translation or NMT

A more recent method is called neural Machine Translation, which trains artificial neural networks to translate text between languages. Compared to earlier methods, NMT systems can translate more fluently and accurately and comprehend more intricate linguistic patterns. NMT is now the most popular method for Machine Translation. A neural network is a network of connected nodes that operates as an information system; it is modeled after the structure of the human brain. These nodes process input data and output the result. Large-scale datasets are processed by neural network-based Machine Translation software. Each node in the system adds a distinct change from the source text to the target text until the desired outcome is reached at the output node.

Neural Machine Translation (NMT) is a novel method to language translation and localization problems. It trains neural models using deep neural networks and artificial intelligence. It has only taken three years for NMT to overtake SMT as the most used Machine Translation technique. Statistical Machine Translation techniques often yield translations that are less adequate and more fluent than those produced by NMT.

The amount of memory used by neural Machine Translation is significantly less than that of standard SMT models. Since every component of the neural translation model is trained collaboratively (end-to-end) to optimize translation performance, this NMT approach varies from traditional translation SMT systems. Neural Machine Translation aims to construct and train a single, massive neural network that can read a sentence and produce an accurate translation, in contrast to the conventional phrase-based translation system, which is made up of numerous little sub-components that are modified individually. SMT methods should not be entirely disregarded, nevertheless, as there are numerous instances in which they will yield translation results that are of higher performance than those of NMT. Because of this, Omniscien has adopted the Hybrid Machine Translation strategy, which seamlessly combines the advantages of both technologies to produce translations of a better performance.

3.1.5 Deep NMT

A contemporary technology built on Artificial Intelligence (AI) and machine learning is called deep neural Machine Translation. A branch of machine learning called "deep learning" draws inspiration from the composition and operations of the human brain. An expansion of Neural Machine Translation (NMT) is called Deep Neural Machine Translation (Deep NMT).

The Deep NMT processes more than one neural network layer, as opposed to just one, like the Shallow NMT does. We thus witnessed the highest level of Machine Translation quality ever achieved. A sizable neural network is used by both Shallow and Deep NMT; however, Deep NMT analyses several neural network layers as opposed to only one. It has been demonstrated that deep encoders can effectively enhance NMT systems.

Deep NMT typically involves a few hidden layers, inputs, and outputs that must be perceived, processed, and delivered appropriately. Put otherwise, deep learning is nothing more than a collection of neural network algorithms that mimic certain aspects of the human learning process, such as pattern recognition and object, person, and object recognition. It uses numerical data transformed from real-world photos, movies, texts, etc., in place of brain electric signals.

With fewer layers, Shallow NMT served as the foundation for early NMT systems. With the development of technology, it became possible to process data using more layers, increasing translation quality and accuracy.

3.2 Transfer Learning in NLP

An overview of transfer learning in natural language processing (NLP) and its many benefits—think of it as the best thing since sliced bread—are provided in this article. You can do some really advanced NLP with practically no programming expertise if you click the link to the code at the end of the article and explore what it is like to play around with these fantastic pre-trained models.To give you a quick overview of this topic, natural language processing, or NLP, is the practice of applying machine learning to analyze text that is considered "natural," which in this case refers to material that is found in books and newspapers rather than, say, computer programming code (okay, some of the models are learning to code, but we’ll stick to talking more generally about ‘natural’ language here). This technology is driving amazing things from automatic article summarization, to responsive chatbots and even creative writing generation. In NLP, pre-trained language models aid us in doing this, and in the realm of deep learning, this concept is referred to as transfer learning. These models give data scientists a base model to expand on in order to complete a specific

NLP task, allowing them to work on new problems. The effectiveness of pre-trained models has already been established in the field of computer vision.

Traditional machine learning models, like Naive Bayes, logistic regression, and support vector machines, were widely used in the early days of natural language processing (NLP) to solve text-related tasks. To get strong performance, these models usually needed a lot of labeled data and carefully designed features. Because deep learning models can automatically extract features from unprocessed text, models like CNNs and LSTMs have gained popularity for natural language processing (NLP) tasks. These models suffered from the problem of vanishing gradients when trained on lengthy sequences, and they still needed a significant amount of labeled data to function well. Transformers established a revolutionary self-attention mechanism that made it possible for models to learn word contextual representations and digest lengthy sequences more quickly. As a result, pre-trained language models were created. These models were trained utilizing unsupervised learning objectives such as next-sentence prediction and masked language modeling on large text corpora.

Utilizing these pre-trained models to handle particular NLP tasks with comparatively minimal extra training has become possible with the rise in popularity of transfer learning. Through fine-tuning the pre-trained models on smaller datasets tailored to a particular task, practitioners could attain cutting-edge outcomes with minimal training time and computer resources. Transfer learning increases a new model's performance by expediting the training process overall. It is mostly employed when training a model calls for a significant investment of time and money. For these reasons, transfer learning is used in many deep learning projects, like sentiment analysis or neural networks that do NLP or CV tasks.

It is important to note that machine learning is not mentioned specifically in transfer learning. It resembles active learning in that it leans more toward a design process. It refers to a method that addresses issues with idea drift or multi-task learning rather than a specific study. Concept drift in machine learning describes the alterations that a task's statistical features go through over time and that the model attempts to forecast. Consequently, the model's forecast accuracy suffers. Since transfer learning depends on a plethora of data and knowledge to provide accurate predictions, it can be useful at this point. Furthermore, transfer learning is essential in situations where there is not enough training data available since it initializes the second model's weights using the weights that were recorded from the previous model. In order to transfer features from one task to another, transfer learning depends on feature generalization. It follows that in this case, datasets are essential. It has been noted that when the datasets used in the first and second trainings are comparable, transfer learning can produce optimal outcomes.

The capacity of transfer learning to use information from one task or dataset to enhance performance on a related job has drawn a lot of interest in natural language processing (NLP) in recent years [58]. By allowing models to learn general language representations that capture syntactic, semantic, and contextual information, this method has completely changed natural language processing (NLP). Humans use natural language, whether spoken or written, to organize their thoughts, interact with one another, and explain their surroundings. It is the clearest and most accurate depiction of how we interact with the outside environment [47]. In its essence, it is human data. Although machine learning algorithms that learn from data can already be built, natural language poses unique difficulties. The majority of data is numerical, such as GPS signals, temperature readings, or picture data. But unlike humans, machines cannot comprehend language. Characters, words, and punctuation make up the orderly yet peculiar collection of symbols that is language. It is disorganized and hardly has any significance on its own. A lot of the time, meanings are implied or only make sense in broader contexts. Therefore, developing machine learning algorithms that comprehend English is difficult.

The sequence models may be created that process language as it naturally occurs—as a sequence of symbols, maintaining order and structure, and taking context into account—by using deep learning to natural language processing (NLP). Sequence models are capable of producing language that sounds human, translating between languages, detecting emotions, and answering questions automatically. They are cumbersome and intricate, though. Adding them to real-world systems and products would be prohibitively expensive. Large amounts of data, highly qualified human experts, and costly infrastructure are required to develop these models. Although sequence models and transfer learning are not novel methods or instruments, recent advances combine them. Their combination proves to be particularly potent. [44].

More significantly, transfer learning lowers the expense of utilizing sophisticated methods while simultaneously increasing the sequence models' robustness and accuracy. Using these strategies is now feasible thanks to transfer of learning. Globally, transfer learning has drawn a lot of interest and is being investigated extensively in a number of fields, including as natural language processing and computer vision. By utilizing past knowledge from data with diverse distributions, researchers and practitioners worldwide are improving performance through the use of transfer learning approaches. The potential of transfer learning to improve machine learning and deep learning applications is becoming increasingly recognized, as evidenced by this global trend. At the regional level, some nations and regions are actively funding transfer learning-related research and development. The understanding and use of transfer learning techniques are being advanced through collaboration between academic institutions, research organizations, and industry participants. To promote innovation in transfer learning and to exchange information, regional conferences, workshops, and seminars are being planned. Transfer learning is becoming popular locally and is being used in particular fields and applications.

In this exploratioon, transfer learning-based Myanmar-Wa language Machine Translation refers to the process of creating pre-trained models and knowledge from another related language pairs (such as Myanmar to English) to get the better performance of Machine Translation particularly for the Myanmar - Wa language. This approach is especially useful when there is limited parallel data available for direct training translation model of Myanmar-Wa language. The Myanmar language, also known as Burmese, is the conclusive language of Myanmar and is spoken by the generali-ty of the population in Myanmar. Wa is an Austroasiatic language spoken by the Wa people, an ethnic group living in Myanmar and China. Since Wa is a less-resourced and lesser-studied language compared to Myanmar. So, the developing a productive Machine Translation directly translation from Myanmar language to Wa language may be complicated due to the deficiency of parallel data. Transfer learning comes into play in such scenarios. Instead of building a Myanmar-Wa translation model from scratch, transfer learning allows to benefit from the representation and knowledge learned using a pre-trained model on a related language pair (such as Translation Myanmar to English).

3.3 Transfer Learning with Transformers

The technique of using previously trained transformer-based models, such as BERT, GPT, or RoBERTa, to enhance natural language processing (NLP) task performance on novel datasets or tasks is known as transfer learning using transformers. The efficacy of pre-trained transformer models in collecting rich linguistic patterns and semantic information has led to the standardization of this method in NLP.

With very few training data examples, transformer-based transfer learning models can achieve excellent prediction accuracies on text-based supervised learning tasks. Thus, social scientists who aim for the most accurate text-based metrics feasible but have limited resources for training data annotation are likely to profit from these models. Large transformer-based models are pre-trained on vast amounts of text data using unsupervised learning objectives, such as masked language modelling (e.g., BERT) or autoregressive language modelling (e.g., GPT). These pre-trained models learn general-purpose representations of language and capture rich contextual information. To adapt these pre-trained models to specific tasks or datasets, transfer learning involves fine-tuning the parameters of the pre-trained models on labelled task-specific data.

During fine-tuning, the task-specific heads are added or modified to suit the target task. For example, for a sentiment analysis task, a classification head might be added on top of the pre-trained transformer model. The fine-tuning procedure involves feeding task-specific labeled data to the pre-trained model, computing the loss between the model predictions and the ground truth labels, and then updating the model parameters using backpropagation and gradient descent. The process is typically repeated for multiple epochs until convergence.

Transfer learning with transformers offers several advantages: such as efficient use of resources, improved performance and reduced annotation effort. By leveraging pre-trained models, transfer learning allows practitioners to benefit from the large amounts of compute and data used during pre-training, making it more efficient to train models on smaller datasets or with limited computational resources. Transfer learning with transformers often leads to improved performance on downstream tasks compared to training models from scratch. The pre-trained models capture rich linguistic patterns and semantic information, which can generalize well to a wide range of tasks. Since transfer learning requires labeled data only for fine-tuning, it can significantly reduce the amount of annotation effort required for training models on specific tasks, especially in cases where labeled data is scarce or expensive to obtain.

Overall, transfer learning with transformers has become a standard approach in NLP, enabling practitioners to build high-performance models for a variety of tasks with minimal effort.

3.4 Pre-trained Model Hugging Face

Hugging Face is a popular platform and library for Natural Language Processing (NLP), known for its pre-trained models and libraries like Transformers. Pre-trained models offered by Hugging Face are trained on large datasets and fine-tuned for specific NLP tasks, making them powerful tools for various applications. Hugging Face's pre-trained models often achieve state-of-the-art performance on various NLP benchmarks and tasks, such as text classification, named entity recognition, and language generation. The Hugging Face library provides a user-friendly interface for loading, using, and fine-tuning pre-trained models, making it accessible to both beginners and experienced practitioners. Hugging Face offers a wide variety of pre-trained models, including those based on architectures like BERT, GPT, RoBERTa, and many more, each suitable for different tasks and scenarios. The Hugging Face community actively contributes to the ecosystem by developing new models, sharing fine-tuning scripts, and providing support, resulting in a rich and vibrant ecosystem.

However, there are also some limitations to consider: Fine-tuning or using large pre-trained models may require significant computational resources, including GPU or TPU accelerators, making them less accessible to users without access to such resources. While pre-trained models excel in many NLP tasks, they may not always generalize well to domain-specific or niche tasks without additional fine-tuning on task-specific data. Some pre-trained models can be quite large in size, which may present challenges in terms of memory and storage requirements, especially for deployment in resource-constrained environments. Pre-trained models trained on large corpora may inadvertently capture biases or sensitive information present in the training data, raising concerns about privacy and fairness in certain applications. Overall, Hugging Face's pre-trained models are powerful tools for NLP tasks, offering state-of-the-art performance and ease of use, but it is essential to consider the trade-offs and limitations associated with their use in specific contexts. The applied pre-trained models in this exploration are described in this section.

The goal of Hugging Face's Natural Language Processing (NLP) challenges is not merely to recognize words but also to comprehend their meanings and contexts.

Computer systems require a pipeline, or a succession of stages, to process texts since they do not process information in the same manner as people do. In 2017, Hugging Face initially introduced its conversation platform. They developed an NLP library that offers numerous resources, such as datasets, transformers, and tokenizers, etc., to standardize NLP and make models available to everyone. Hugging Face's Transformers NLP libraries and other tools were released, and they quickly gained a lot of traction with major IT businesses.

For improved contact experiences, many businesses are already integrating NLP technology into their systems. It is now more crucial than ever to keep communication as near to the human experience as possible. Hugging Face enters the scene in this situation. The Hugging Face NLP library offers two methods to get started: either utilizing pipeline or reusing any pre-trained model that is readily available to work on your ideas. These models take up a lot of storage space, and the model will be downloaded when the aforementioned code is performed for the first time.

Hugging Face transformer library is available through high-level APIs and includes many models for various activities. Transformer models are difficult to construct since they need to have tens of billions of parameters fine-tuned and extensive training. The purpose of the hugging Face transformer library was to make it simple, flexible, and easy to use these intricate models by utilizing a single API. The models may easily be loaded, trained, and stored. All pre- and post-processing operations are carried out by the Hugging Face Transformer pipe-line on the input text data. These pipelines, which are the most fundamental object in the Transformer library, contain the whole process of any NLP solution. This makes it easier to integrate a model with the necessary pre- and post-processing procedures, as the suggested system simply needs input texts.

3.4.1 Helsinki Model

"Helsinki-NLP/opus-Machine Translation-en-ROMANCE" is one of the pre-trained models offered by the Helsinki NLP (University of Helsinki) team. This model is particularly designed for Machine Translation tasks from English to Romance languages (e.g., French, Spanish, Italian, Portuguese, etc.). Here are some key points about the Helsinki pre-trained models. The Helsinki pre-trained models focus on multi-lingual translation tasks, enabling translation between various language pairs. The Helsinki models are built upon Transformer architectures, which have shown state-of-the-art performance in natural language processing tasks, including Machine Translation. While the pre-trained models are already trained on large datasets, they can be further fine-tuned on specific datasets or tasks to improve their performance for particular use cases. These models are publicly available through the Hugging Face model hub, making them easily accessible to developers and researchers. The Helsinki NLP team actively engages with the community, providing support, updates, and contributions to the open-source NLP ecosystem. These pre-trained models have been used in various applications such as text translation, cross-lingual understanding, and more.

Text translations from one language to another are done using the Helsinki-NLP models. Thus, a block of text is entered into the model, and the model outputs the translated block of text. This specific model generates English text from German text as its output. The Helsinki-NLP group created the single- or bilingual-direction Machine Translation models known as OPUS-Machine Translation. There are now more than 1000 OPUS-Machine Translation variants available. Each type has six layers in the encoder and decoder and is based on a transformer. With the use of OPUS parallel data, each model is trained from scratch. As an illustration, the model Helsinki-NLP/opus-Machine Translation-en-de is trained using a dataset of parallel phrases in both German and English.

This model can translate an English-language text sequence into German. The initial OPUS-Machine Translation models were created using the C++-based Marin-Machine Translation framework, and they were then translated to PyTorch so that they could be used with the transformers library.

3.4.2 mT5 Model

mT5, short for "Multilingual Translation with T5," is a pre-trained model developed by Google Research. It is an extension of the T5 (Text-To-Text Transfer Transformer) architecture, specifically tailored for multilingual translation tasks. mT5 is designed to translate between multiple languages in both directions (e.g., English to French and French to English). Here are some key points about the mT5 pre-trained model: mT5 is trained to perform translation tasks across multiple languages. It supports translation between various language pairs, making it versatile for multilingual applications. mT5 is built upon the T5 architecture, which treats all NLP tasks as text-to-text problems. This unified approach allows for seamless adaptation to various tasks, including translation. Like T5, mT5can be fine-tuned on specific downstream tasks or datasets to further improve its performance for particular use cases. mT5is trained on a large corpus of multilingual text data, enabling it to learn robust representations for translation across different languages. mT5has demonstrated competitive performance in Machine Translation benchmarks, achieving high accuracy and fluency in translating text between multiple languages. mT5is open-source, and the pre-trained models are publicly available, allowing researchers and developers to use and build upon them for various applications. mT5has been used in a wide range of multilingual applications, including cross-lingual information retrieval, multilingual conversational agents, and more.

For the latest updates, information, and access to mT5models, you can refer to Google Research publications, GitHub repositories, or platforms like the Hugging Face model hub, where pre-trained mT5 models may be available for use. The fundamental component of the Original Transformer concept is closely reflected in the encoder-decoder Transformer design T5 [45]. The earlier unifying frameworks for downstream NLP activities are inherited by and transformed into a text-to-text format by the T5 network architecture [37]. In other words, the T5 design enables the use of the encoder-decoder method to combine all NLP tasks into a single network. As a result, each job uses the same loss function and hyper-parameters.

All of the T5 model's capabilities are inherited by mT5. A modified version of the C4 dataset with more than 10,000 web page contents in 101 languages (including Persian) was used to train mT5 over the course of 71 monthly scrapes. mT5 achieves state-of-the-art on all the tasks [47]– [53], particularly on the summarizing job, when compared to other multilingual models such multilingual BERT [25], XLM-R, and multilingual BERT (no support for Persian) [28].

3.5 Pivot-Based Transfer Learning

In the context of Natural Language Processing (NLP), pivot-based transfer learning involves transferring knowledge from a source language to a target language through an intermediary, or pivot, language. This approach is particularly useful when direct transfer between the source and target languages is not feasible due to lack of parallel data or linguistic differences. Pivot-based Machine Translation and triangulation-based Machine Translation are two ways for dealing with the challenge of translating between two languages when direct translation data is limited or absent.

The technique we employ to construct A-to-C and/or C-to-A Machine Translation systems without (or with minimal) parallel data of the A-C language pair is called pivot Machine Translation. If there exist sizable A-B and B-C parallel corpora that can be utilized, a "pivot" language B could be employed to assist in the development of A-C Machine Translation systems [51].

Translation is completed in two stages when using pivot-based Machine Translation: from source language (SL) to pivot language (PL) and from pivot language (PL) to target language (TL). Translations of the source text are first made into a pivot language, which is subsequently translated into the target language. Two language pairs—SL to PL and PL to TL—need simultaneous data. The quality of the pivot language translation has a significant impact on the pivot-based translation's quality. This approach is sensitive to errors introduced in both translation steps, which can accumulate and degrade translation quality.

In machine learning, transfer learning refers to utilizing the knowledge gained for performing one task for some other task. This is done by using the machine learning model trained to perform one task as initialization for performing some other task. In neural Machine Translation, transfer learning can be performed by using the source to pivot and pivot to target models. The parameters of these models can be used to initialize the source to target models in various ways. Also, the process in which these models are trained, and the parameters are initialized can also be performed in various ways. Initializing the encoder of the source to target model with the encoder of the pivot to target model and the decoder of the source to target model with the decoder of the pivot to target model are two methods for setting the parameters of the source to target model [56]. This kind of initialization is carried out because the source to target model's encoder can be initialized because it has acquired representations or knowledge for the source language from the source to pivot model. Similar to this, the decoder of the pivot to target model can be used to initialize the decoder of the source to target model since it has acquired knowledge of the target model's representations. The model is trained on source-target parallel data after the encoder and decoder of the source to target model are initialized. A problem with the first approach is that the encoder in the source to pivot model is trained to produce outputs for the pivot decoder and not the target decoder. And the decoder of the pivot to target model is trained on the outputs of the pivot encoder and not the source encoder. In order to overcome this drawback, a step wise pre-training strategy is followed to train the models. In the first step, a source to pivot model is trained on source pivot parallel data. In the next step, the encoder of the source to pivot model is used to initialize the encoder of the pivot to target model.

Now the pivot to target model is trained on the pivot to target data, but the encoder is frozen. This means that the parameters of the encoder are not updated. This retains the source language representations in the encoder learned in the first step. This also prevents the encoder from adapting to the pivot language. Now the encoder is producing representations from the source encoder which is used by the target decoder. In this way, the drawback of first transfer learning approach is mitigated. In the next step, the encoder and decoder of the model from the second step is used to initialize the encoder and decoder of source to target model. The source to target model is then trained on source to target parallel data. Figure 3.3 shows the processes of plain transfer learning.

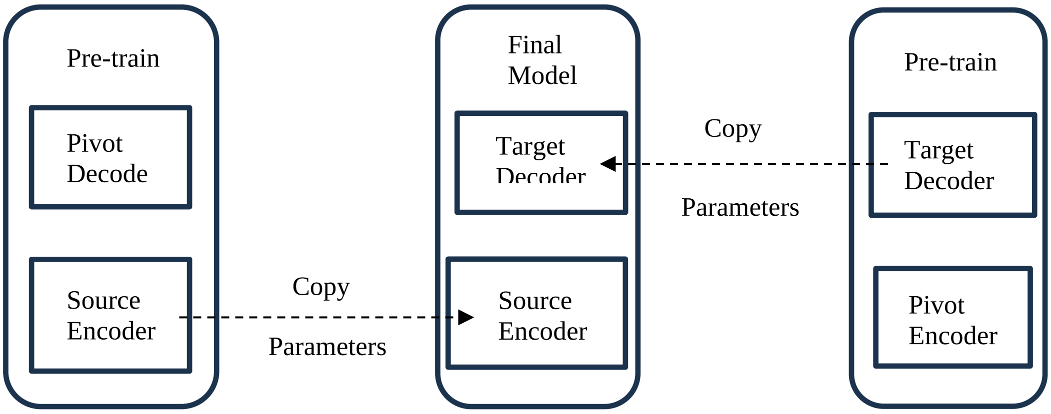


Figure 3.3 Plain Transfer Learning [56]

The process of pivot-based transfer learning in NLP involves the following steps:

Train a language model or pre-train a neural network on the data from the source language. This step allows the model to learn general linguistic patterns and features that are useful across languages.

Fine-tune the pre-trained model on data from the pivot language. This adaptation step helps the model adjust its representations to better align with the linguistic characteristics of the pivot language, which may differ from both the source and target languages.

Finally, transfer the knowledge learned from the pre-trained model, now adapted to the pivot language, to the target language. This can be achieved by fine-tuning the model on the limited labeled data available in the target language or by directly using the adapted model for inference in the target language.

By leveraging a pivot language, pivot-based transfer learning allows models to effectively transfer knowledge and adapt to the target language, even in cases where direct transfer is challenging due to linguistic differences or lack of data. This approach has been successfully applied in tasks such as Machine Translation, cross-lingual document classification, and sentiment analysis across languages.

3.6 Transformer Architecture and Cascade Models

The Transformer architecture and Cascade models are two distinct approaches in machine learning, often used for different purposes. Transformers are based on the self-attention mechanism, which allows the model to weigh the importance of different input tokens when making predictions. This mechanism enables capturing long-range dependencies in sequences efficiently. The architecture consists of encoder and decoder layers, with multi-head self-attention and feedforward neural networks. Transformers have been applied to various NLP tasks, including Machine Translation, text generation, sentiment analysis, and more. Pre-trained transformer models like BERT, GPT, and RoBERTa have achieved state-of-the-art performance on numerous NLP benchmarks and have become the foundation for transfer learning in NLP tasks. Cascade models, also known as cascaded classifiers or cascade classifiers, are a type of ensemble learning technique where multiple classifiers are used sequentially, with each classifier refining the predictions of its predecessors.

Cascade models are commonly used in computer vision tasks, particularly object detection, where they help improve both speed and accuracy. In object detection, cascade models typically consist of a series of classifiers, each trained to perform a specific task, such as detecting regions of interest (e.g., regions containing objects) at different levels of confidence. The output of each classifier is used to filter out negative samples or refine the predictions, reducing the number of false positives and improving overall detection performance. Cascade models have been widely used in popular object detection frameworks like Viola-Jones, and variants have been developed for more modern deep learning-based detectors like Faster R-CNN and Cascade R-CNN. Transformers are primarily used for sequence modelling tasks in NLP, while Cascade models are used for object detection and related computer vision tasks. However, it is worth noting that with the increasing versatility of deep learning models, there may be opportunities to combine or adapt these techniques in novel ways to address new challenges.

Cascade models, on the other hand, are a class of models that consist of multiple stages or layers where the output of one stage serves as the input to the next stage. In the context of NLP, cascade models can refer to a series of models applied sequentially to solve a specific task or to a hierarchical model architecture where information flows through multiple levels of abstraction. Cascade models in NLP often consist of multiple processing stages, such as feature extraction, representation learning, and classification/regression. Each stage can have its own set of parameters and may be trained independently or jointly with other stages. In some cases, cascade models can be combined with Transformer architectures. For example, a cascade model might use a Transformer-based language model for feature extraction and representation learning, followed by additional layers for classification or regression tasks. Overall, while the Transformer architecture is a specific model architecture used primarily for sequence processing tasks like NLP, cascade models represent a broader class of models that involve multiple stages or layers of processing, which can include Transformer components.

3.7 Tokenization and Encoding

In the realm of translation, the significance of tokenization and encoding cannot be overstated. While they might appear as technical minutiae, they form the foundational pillars upon which the entire translation process stands.

Tokenization, the process of breaking down text into smaller units such as words or subwords, is fundamental for accurate translation. By dissecting the text into manageable chunks, tokenization enables the translation system to comprehend the structure and semantics of the input language, facilitating more precise rendering into the target language.

Encoding, on the other hand, involves representing these tokens in a numerical format suitable for processing by machine learning algorithms. Through encoding, each token is transformed into a numerical vector, allowing the translation model to analyze and manipulate the text mathematically. This numerical representation is crucial for the model to learn patterns, relationships, and nuances within the language data, thereby enhancing the quality and fluency of translations.

Together, tokenization and encoding lay the groundwork for the intricate dance of translation, empowering machines to navigate the complexities of language and bridge the gap between diverse linguistic landscapes. While seemingly mundane, their role in the translation process is indispensable, underscoring the critical importance of these technical mechanisms in achieving accurate and effective translations.

3.8 Chapter Summary

In this concluding segment of Chapter 3, we encapsulate the intricate groundwork laid out in the preceding sections. We delve into the depths of background theories, meticulously examining their nuances and implications within the realm of Machine Translation (Machine Translation). Not only do we explore the multifaceted landscape of Machine Translation methods, dissecting their varied approaches and methodologies, but we also provide a comprehensive exploration of their nuances. Additionally, we shine a spotlight on the burgeoning domain of Transfer Learning methods in Machine Translation, illuminating the innovative strategies and techniques employed to bridge linguistic gaps and enhance translation quality. Through meticulous analysis and insightful discussion, this chapter cements a foundational understanding of the diverse methodologies and theoretical underpinnings that propel the field of Machine Translation forward.

CHAPTER 4

METHODOLOGY OF THE PROPOSED MODEL

In this chapter, step by step of approaches of the proposed model are described in detail. The applied methods and models are presented and discussed in this section. Moreover, about Transformer NMT, Transfer Learning with Transformer NMT and Pivot-Based Transfer Learning are discussed in detail in this section.

4.1 Machine Translation Models

Text translation from one language to another is the responsibility of Machine Translation models. From English to Spanish, for instance. Transformer sequence-to-sequence architecture serves as the foundation for the models. Text can be automatically translated between languages using Machine Translation models. Over time, these models have undergone substantial changes, with improvements in translation quality being largely attributed to developments in deep learning and neural network architectures. Linguistic patterns and translation rules are specifically established by human specialists in rule-based Machine Translation. These guidelines control how text is translated across languages. To carry out translation, RBMT systems often use morphological analysers, grammar rules, and dictionaries. While RBMT systems can be precise in certain scenarios, they often struggle with handling the complexity and nuances of natural language.

Statistical Machine Translation (SMT): Statistical Machine Translation models learn translation patterns and relationships from large amounts of bilingual text corpora. They use statistical methods to estimate the probability of generating target language translations given source language input. SMT models often rely on techniques such as phrase-based translation and language modelling. While SMT systems have shown significant improvements over RBMT, they still face challenges in capturing long-range dependencies and handling rare or unseen translation patterns.

The development of neural network designs and deep learning has led to the emergence of a new paradigm known as neural Machine Translation (NMT). The capacity of NMT models to learn intricate translation patterns straight from data, without depending on manually created linguistic characteristics or alignment models, has made them the state-of-the-art in Machine Translation. Recurrent neural networks

(RNNs), convolutional neural networks (CNNs), and transformer architectures—arguably the most important—are essential parts of NMT models.

Vaswani et al.'s section "Attention is All You Need" established the transformer architecture, which is now the foundation of many contemporary NMT systems. In order to provide more precise and contextually relevant translations, transformers use self-attention processes to identify long-range dependencies and links between words in the input text. Overall, Machine Translation models aim to bridge language barriers and facilitate communication across different linguistic communities by automatically translating text or speech between languages. They play a crucial role in various applications, including cross-border communication, international business, localization of software and content, and improving accessibility for multilingual users.

4.1.1 Statistical Machine Translation Models

Statistical Machine Translation (SMT) models were investigated in the early stages, but the absence of parallel corpora hindered development. The constraints imposed by the Myanmar-Wa corpus's lack of aligned sentences led to the investigation of substitute techniques. Linguistic patterns and translation rules are specifically established by human specialists in rule-based Machine Translation. These guidelines control how text is translated across languages. RBMT systems typically involve components such as dictionaries, grammar rules, and morphological analyzers to perform translation. While RBMT systems can be precise in certain scenarios, they often struggle with handling the complexity and nuances of natural language. SMT models learn translation patterns and relationships from large amounts of bilingual text corpora. They use statistical methods to estimate the probability of generating target language translations given source language input. SMT models often rely on techniques such as phrase-based translation and language modeling. While SMT systems have shown significant improvements over RBMT, they still face challenges in capturing long-range dependencies and handling rare or unseen translation patterns. Predictable methods are used in SMT to train machines to translate using parallel bilingual text corpora. The machine predicts the translation of the foreign languages by using the translated text that it has been taught. Because it is data-driven, all that is required is the corpus of the target and source languages. Nevertheless, during translation, the word or phrase alignment divides the sentences into separate words or phrases. The word cannot be taken into consideration or translated before the completion of the preceding one. Additionally, the time and effort required for corpus gathering are high. Since the statistical approach mostly involves the human creation of huge multilingual dictionaries, it cannot be the dominating method. Figure 4.1 shows the pipeline of SMT models.

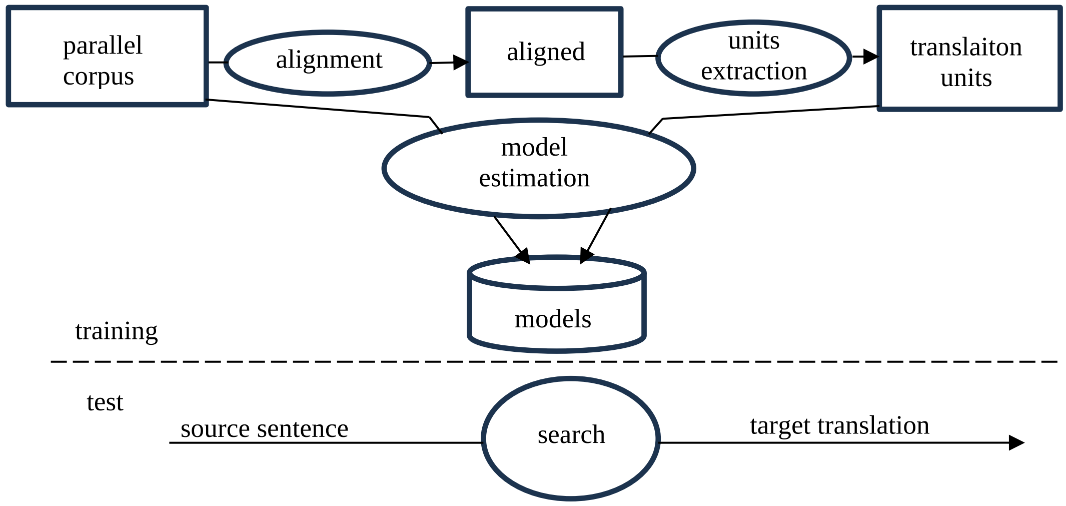


Figure 4.1 Statistical Machine Translation Model

With the advancement of deep learning and neural network architectures, a new paradigm called Neural Machine Translation (NMT) has emerged. NMT models have become the state-of-the-art in Machine Translation due to their ability to learn complex translation patterns directly from data without relying on handcrafted linguistic features or alignment models. Key components of NMT models include recurrent neural networks (RNNs), convolutional neural networks (CNNs), and most notably, transformer architectures.

The architecture of transformer, introduced in their work "Attention is All You Need" by Vaswani et al., has become the backbone of many modern NMT systems. Transformers leverage self-attention mechanisms to capture long-range dependencies and relationships between words in the input text, allowing for more accurate and contextually relevant translations.

Overall, Machine Translation models aim to bridge language barriers and facilitate communication across different linguistic communities by automatically translating text or speech between languages. They play a crucial role in various applications, including cross-border communication, international business, localization of software and content, and improving accessibility for multilingual users.

4.1.2 LSTM-based Machine Translation Models

Neural Machine Translation (NMT) models that use Long Short-Term Memory (LSTM) units, a kind of recurrent neural network (RNN) architecture, are called LSTM-based models. The vanishing gradient issue that traditional RNNs frequently face is addressed by LSTM units, which makes it possible for them to more successfully capture long-range dependencies in sequential data. An LSTM encoder is used to convert the input sequence in the source language into a fixed-length vector representation. Usually, each word in the input sequence is represented by a word embedding, which is fed one token at a time into the LSTM encoder.

The LSTM encoder modifies its internal hidden state as it processes each token in the input sequence. The context of the complete input sequence is captured by the encoder LSTM's final hidden state, which is then fed into the decoder. Based on the encoded input and the previously created tokens, the decoder—which is likewise constructed using LSTM units—generates the target sequence one token at a time. To forecast the next token in the target sequence, the decoder LSTM uses the previously created token and the current hidden state (initialized with the final encoder hidden state) as input at each time step. In many LSTM-based NMT models, an attention mechanism is employed to allow the decoder to focus on different parts of the input sequence while generating the output sequence. This mechanism helps improve the model's ability to align source and target language words and capture relevant information from the input sequence. LSTM-based Machine Translation models are trained using parallel corpora consisting of source language sentences paired with their corresponding target language translations. The model parameters, including the weights of the LSTM units and any additional components such as attention mechanisms, are optimized to minimize a loss function that measures the discrepancy between the model's predictions and the ground truth translations. LSTM-based Machine Translation models have been widely used in the past due to their effectiveness in capturing sequential dependencies and generating fluent translations. However, more recently, transformer-based architectures, such as those based on the attention mechanism and self-attention mechanisms, have largely surpassed LSTM-based models in terms of translation quality and efficiency, leading to their widespread adoption in modern Machine Translation systems. Figure 4.2 shows the basic architecture of LSTM-based Machine Translation Model.

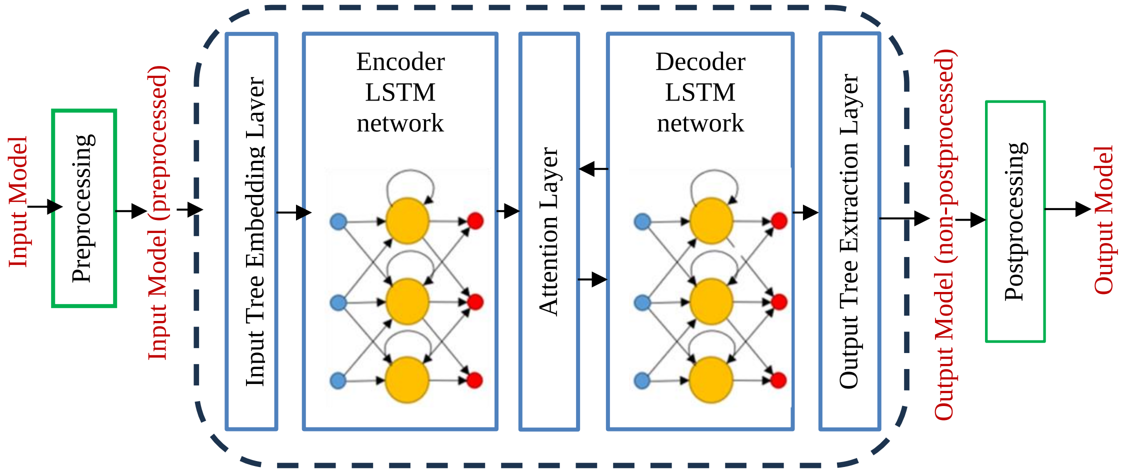


Figure 4.2 Basic Architecture of LSTM-based Machine Translation Model

The introduction of Long Short-Term Memory (LSTM) models was a response to the difficulties encountered by SMT. With their capacity to record sequential dependencies, LSTMs demonstrated potential in managing the intricacies of the Wa language. Even with its improvements, the LSTM architecture was unable to handle Wa's particular language features, therefore more work was needed. The construction of the LSTM encoder-decoder is shown in Figure 4.3.

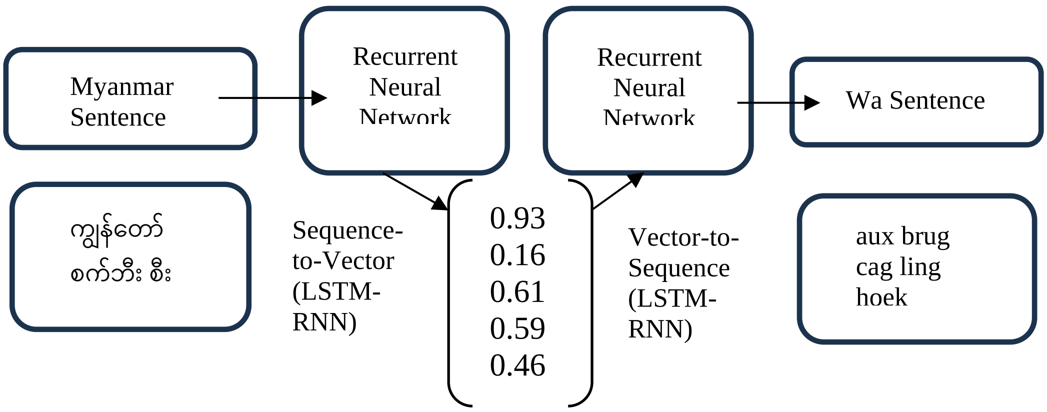


Figure 4.3 Architecture of LSTM based Encoder-Decoder for Myanmar-Wa Corpus

4.1.3 Transformer-based Machine Translation Models

A subset of neural Machine Translation (NMT) models known as transformer-based models make use of the transformer architecture, which was first presented in the 2017 paper "Attention is All You Need" by Vaswani et al. Because transformer-based models are better at capturing word associations and long-range dependencies than classic recurrent neural network (RNN)-based models like LSTM, they have emerged as the state-of-the-art in Machine Translation.

In recent years, the discipline of natural language processing (NLP) has seen a revolution thanks to a particular kind of deep learning architecture called transformers. They are frequently employed for jobs like sentiment analysis, text categorization, and language translation, among others. The history of transformer architecture, its essential parts, and some of the most widely used transformer models in use today will all be covered in this blog post. Transformer-based versions have an encoder and a decoder, just like other NMT variants. The sequence input in the source language is processed by the encoder, while the output sequence in the target language is produced by the decoder. The self-attention mechanism is the central part of the transformer architecture. By weighing the significance of each word in the input sequence as it is processed, the model with self-attention is able to accurately capture long-range dependencies. The self-attention mechanism in transformers is usually built with many attention heads to capture various components of context. The input sequence is learned by each head with a different attention distribution, and the output is concatenated and linearly transformed. Since transformers do not inherently capture the order of words in a sequence, positional encoding is used to provide the model with information about the positions of words. Positional encoding vectors are added to the input embedding to convey positional information to the model.

After processing the input sequence in the source language, the transformer encoder generates a series of contextualized representations for every word. Position-wise feedforward neural networks come after a multi-head self-attention mechanism in each encoder layer of the transformer. The encoder's contextualized representations are used by the transformer decoder to construct the output sequence in the target language. Every decoder layer has feedforward and multi-head self-attention neural networks, same like the encoder, with an extra cross-attention mechanism to handle the encoder's output. Transformer-based Machine Translation models are trained using parallel corpora, which are collections of texts in the source language and their translations into the target language. The model parameters are optimized to minimize a loss function that measures the discrepancy between the model's predictions and the ground truth translations. The transformer-based model's overall architecture is depicted in Figure 4.4.

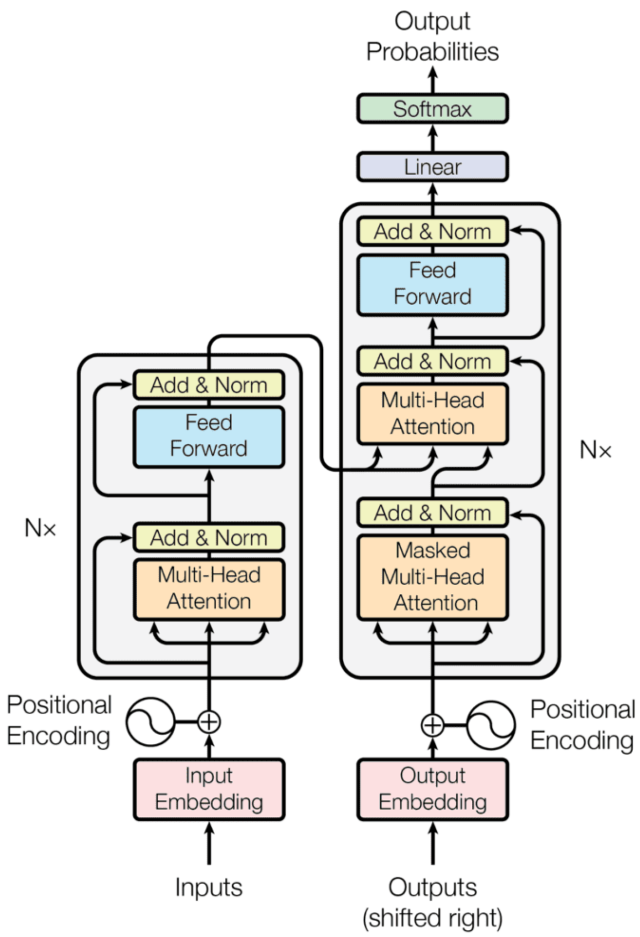


Figure 4.4 Overview Architecture of Transformer Based Model

Transformer-based Machine Translation models have been widely used in modern Machine Translation systems due to their remarkable success in improving translation quality and efficiency. Examples of these models include Google's Transformer, BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer), and T5 (Text-to-Text Transfer Transformer).

A significant turning point in the development of Machine Translation for Myanmar-Wa was the introduction of transformer-based models. With its positional embedding and attention methods, the transformer design brought a fresh method of managing sequential dependencies. The dynamic representation of the transformer's output, which was contingent on the length of the input sequence, resolved problems with different word ordering and subtleties in context, which enhanced the quality of the translation.

The architecture of the Transformer Encoder-Decoder is shown in Figure 4.5, with particular attention paid to the interconnected layers and information flow between the encoder and decoder. With its novel attention methods and positional embedding, the transformer represents a major breakthrough in Machine Translation architectures.

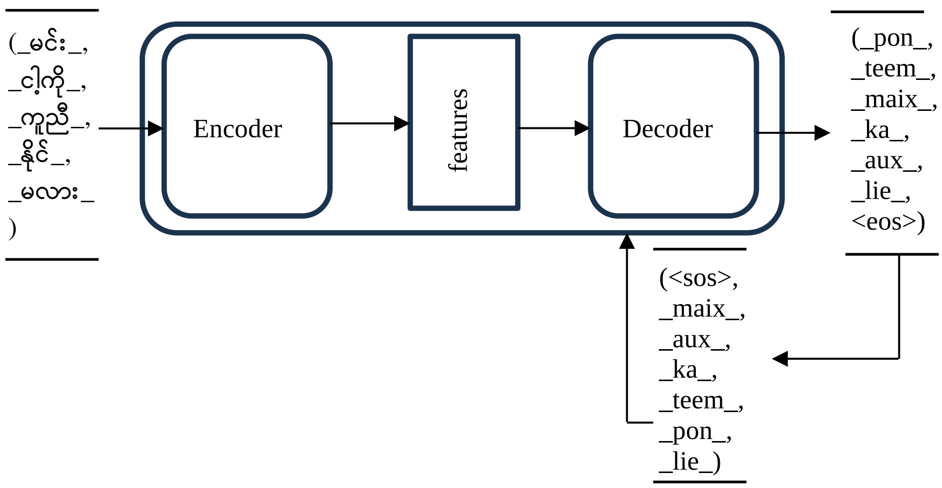


Figure: 4.5 Architecture of Transformer Encoder-Decoder for Myanmar-Wa corpus

4.2 Transfer Learning on mT5 Pre-Trained Model

Traditional transfer learning involves taking a pre-trained model and fine-tuning it on a new task or domain. In the context of Machine Translation using the mT5(Multilingual Translation with T5) pre-trained model, traditional transfer learning can be applied by fine-tuning the mT5 model on a specific translation task or domain. The mT5 model is a variant of the T5 (Text-to-Text Transfer Transformer) model pre-trained on a large corpus of multilingual text data. It is capable of performing various text-to-text tasks, including Machine Translation, text summarization, question answering, and more. The mT5 model is a variant of the T5 (Text-to-Text Transfer Transformer) model pre-trained on a large corpus of multilingual text data. It is capable of performing various text-to-text tasks, including Machine Translation, text summarization, question answering, and more.

Traditional transfer learning is based on the idea of using information from an existing model to improve a machine's performance in a related activity. This is done in Neural Machine Translation (NMT) by training a sparse parallel corpus and a previously trained model (parent) to provide the initial parameters for a new model (child). This system’s ground breaking Machine Translation project, the Myanmar-Wa corpus, demonstrates the application of transfer learning. Low-resource languages like Wa were difficult to align, but creative solutions like crowdsourcing and data augmentation were used to increase the corpus's size and diversity.

The translation quality between Myanmar and Wa is greatly improved by using the mT5 model. Language dynamics are captured by LSTM-based models, and transfer learning combined with SMT overcomes data scarcity. The Traditional Transfer Learning using mT5 Pretrained Model on Myanmar-Wa corpus is shown in Figure 4.6.

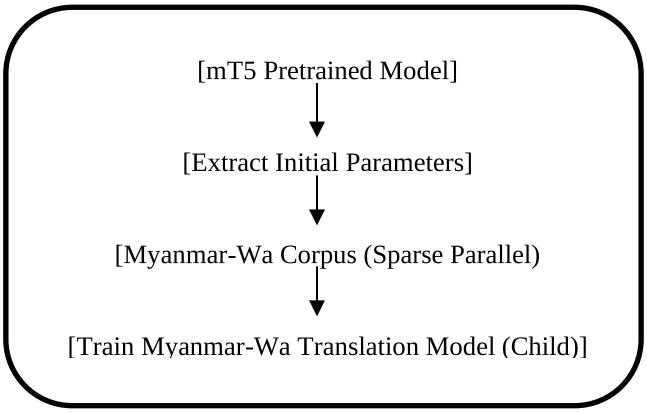


Figure 4.6 Traditional Transfer Learning on mT5 Pre-Trained Model

4.3 Pivot-Based Transfer Learning NMT Models with mT5

Pivot-Based Transfer Learning NMT Model is a model that is translation from the source language (SL) to the pivot language (PL) and then from the pivot language (PL) to the target language (TL) constitute the two stages of pivot-based Machine Translation [33]. Since parallel data for Myanmar-Wa is scarce, pivot-based NMT using mT5 models has been thoroughly investigated. The architecture is based on transfer learning and is specially designed for language pairs other than English, especially Myanmar-Wa.There are three key phases in the operating framework.

Pretraining is the initial stage, which makes use of parallel source-pivot and pivot-target corpora. In the context of the pivot language, this first phase enables the model to understand the subtleties of the source and target languages. The next step is to incorporate a source-target parallel corpus for iterative model optimization, whereby the model is adjusted to improve its translation performance for the particular task.

Using the pivot language to pretrain source encoders and target decoders is the last step toward achieving a comprehensive comprehension of language dynamics.The mT5 model is a strong substitute for the conventional mT5 model in this situation. The approach leverages the common linguistic features of Wa and Filipino by deliberately choosing pivot language pairs, English-Fil (Philippines) and Myanmar-English, based on their similar sentence patterns [47]. This tactical decision captures the synergy between these languages and greatly improves the quality of translation. Figure 4.7. mT5-Based Pivot-Based Transfer Learning NMT Models.

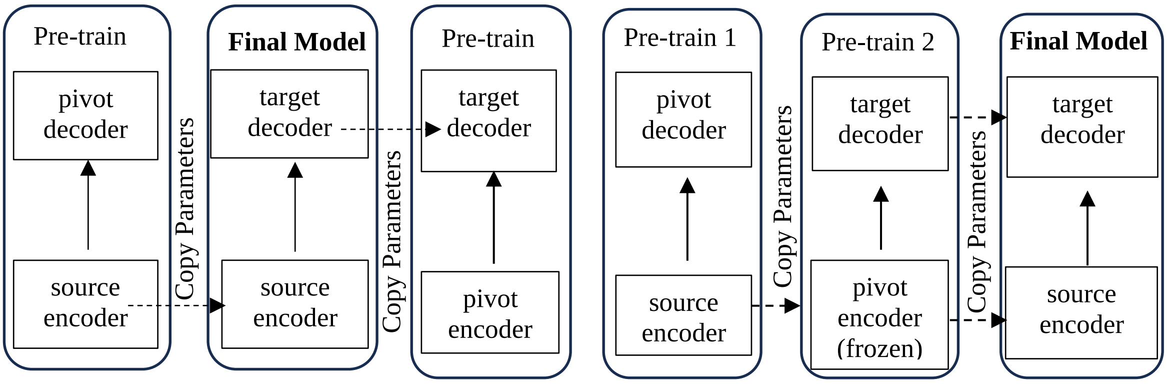


Figure 4.7 Pivot-Based Transfer Learning Model

The technique of transfer learning is similar to the Python code that is given in that it involves merging the parameters from two different models ({model1} and `model2}) to create a new model ({new_model}). Both the decoder parameters from {model2} and the encoder parameters from `model1} are smoothly transferred into the new model. This combination shows how the language expertise from the two models has been combined into a single, cohesive architecture.

The transfer learning paradigm's use of pre-trained models, particularly mT5, highlights how versatile and adaptable this method is across a wide range of language pairs. This novel approach, similar to the model parameter merging in Python code, has great potential to improve Machine Translation performance, especially in language contexts with limited resources.

The mT5model is a strong substitute for the mT5model in this situation. The method takes advantage of the similarities between Wa and Filipino by choosing the pivot language pairings, English-Fil (Philippines) and Myanmar-English, which have comparable sentence structures. This enhances the quality of the translation.

The use of pre-trained models (mT5, in particular) in transfer learning highlights how flexible this method is for a variety of language combinations. This novel approach has the potential to improve Machine Translation performance, especially in low-resource language contexts. The Pivot Based Transfer Learning Model using mT5 and english as a pivot language on Myanmar-Wa corpus is shown in Figure 4.8.

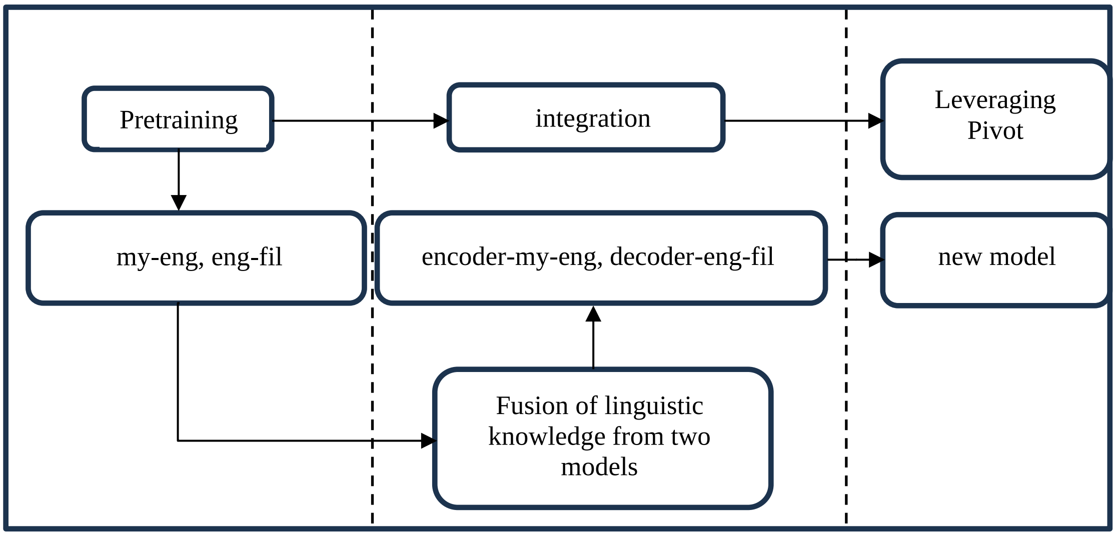


Figure 4.8 Pivot Based Transfer Learning Model with mT5 on Myanmar-Wa Corpus

4.4 Pivotal Role of Language-specific Tokenizer in Transfer Learning

Machine Translation has played a key role in bridging language gaps, and its development has been continuous, with notable technological advances. In order to improve translation skills, this section describes how a Machine Translation model is built and trained using a pivot transfer learning strategy. The study focuses on the Myanmar (Burmese)-Wa corpus, which is a subset of the language family known as Austro-Asiatic.

Tokenization and encoding of the source and target texts using cutting-edge models are the first phases. Tokenization is preceded by segmentation for the Myanmar language. A Myanmar word segmentation tool [49] is used to segment, and the Myanmar BERT tokenizer (UCSYNLP/MyanBERTa) is then used to encode the sentences. The target sentences are encoded using the facebook/bart-base tokenizer.

Training an encoder-decoder architecture for Machine Translation is the goal of transfer learning. The encoder tokenizer is the Myanmar BERT tokenizer. As the encoder tokenizer, the Myanmar BERT tokenizer and also Myanmar BERT model—which was first created for natural language comprehension—extracts context from the divided source phrases. Built specifically for the Myanmar language, the MyanBERTa language model is based on the BERT architecture. A 528,000-step pre-training phase was conducted for MyanBERTa using a word-segmented dataset that was unique to Myanmar. This large corpus had 136 million words, or 5,992,299 sentences. Following the implementation of word segmentation, a byte-level Byte Pair Encoding tokenizer with 30,522 subword units is utilized.

The decoder tokenizer is the pre-trained facebook/bart-base tokenizer, which uses the encoded data to provide target translations. For sequence-to-sequence (seq2seq) tasks, the Bidirectional AutoRegressive Transformer (BART) represents a transformer design made up of an encoder-decoder structure. When optimized for text generating tasks like translation and summarization, the BART model shows remarkable effectiveness [56].

This two-step procedure, which involves knowledge transfer from the facebook/bart-base encoder-decoder model and a tokenizer fusion strategy (combining MyanBERTa tokenizer and facebook/bart-base tokenizer), significantly enhances translation performance. By facilitating efficient knowledge transfer, this approach achieves more accurate and nuanced translations, effectively managing the diverse linguistic characteristics of different languages.

The training procedure is divided into several epochs, where the training dataset is iterated through throughout each epoch. The AdamW optimizer is used to improve the model, and a step-wise learning rate scheduler makes sure that optimal convergence is achieved. Gradient accumulation steps are used to alleviate GPU memory constraints, enabling the model to accumulate gradients over multiple batches prior to parameter updates.

To evaluate the model's generalization on untested data, validation is an essential step. The validation loss and perplexity are calculated using the validation dataset, which consists of the Myanmar-Wa corpus. The exponential of the average validation loss is used to compute ambiguity, a metric used to assess translation quality.

This measure shows how effectively the model represents the intricacy of the context and language.

At the conclusion of each epoch, snapshots are saved in order to record the model's evolution. Model reproducibility is facilitated and checkpoints are provided by the saved models. Understanding the model's learning trajectory and pinpointing possible areas for improvement requires an understanding of the model's evolution.

Using two specialized tokenizers on one encoder-decoder model to get the best translation performance, this model-building procedure captures the essence of tokenizer fusion and transfer learning. The study using the Myanmar-Wa corpus demonstrates how flexible and successful the suggested strategy is. The field of linguistics and language preservation benefit greatly from the ongoing development of Machine Translation models, which are propelled by creative approaches and flexibility in the face of various linguistic difficulties.

4.5 Proposed Model Training Process

The Fusion Tokenizer (FusionTOK) Model represents our novel approach to Myanmar to Wa Machine Translation, which combines the strengths of tokenizer fusion and transfer learning techniques.

The proposed model, FusionTOK, represents a significant advancement in the field of machine translation. By leveraging transfer learning with a pre-trained translation model and incorporating language-specific tokenizers, FusionTOK enhances the quality of translations. Its architecture is meticulously designed to effectively capture cross-linguistic dependencies, thereby facilitating seamless communication between the encoder and decoder components. The experimental setup and training process for the FusionTOK model are detailed in the following sections.

4.5.1 Data Collection and Myanmar-Wa Corpus Preparation

In this pursuit of enhancing the Myanmar-Wa Neural Machine Translation (NMT) through the introduction of a novel Transformer architecture combined with advanced segmentation techniques tailored for non-space languages, it is essential to meticulously outline the experimental framework that underpins this research. The design and configuration of experiments are vital to the integrity and validity of the study's findings.

An important first step toward creating a corpus specifically designed for Machine Translation was the creation of the Myanmar-Wa corpus. Strict attention to linguistic details was paid throughout construction, and the Myanmar-Wa-Chinese dictionary proved to be a useful resource. The lack of parallel data for the Myanmar-Wa language pair presented a significant barrier during the data collection phase, even with our dedication to accuracy.

Low-resource languages, like Wa's, are distinct from more widely used language pairings like English-Spanish or English-French, which have access to larger bilingual corpora. As such, there are less opportunities for text alignment in these languages. To overcome this obstacle, we used cutting-edge techniques like data augmentation, crowdsourcing, and cross-lingual transfer learning to increase the corpus's size and variety.

It is important to remember that the Myanmar-Wa corpus is a valuable resource for future research projects in addition to being a groundbreaking effort in our Machine Translation activities. Through tackling the unique data limitations associated with low-resource languages, these endeavors sought to establish a foundation for the creation of Machine Translation models that exhibit enhanced accuracy and contextual suitability.

Myanmar-Wa corpus, representing the pioneering effort in Machine Translation for this language pair, is strategically divided for training, validation, and testing. The training set comprises 26970 instances, the validation set 4030, and the test set 4000, totalling 35,000 instances. This meticulous segmentation ensures a robust and comprehensive foundation for developing precise and contextually appropriate Machine Translation models for Myanmar-Wa. Additionally, the corpus size table in table 4.1 gives an idea of the amount of data used in the training, validation, and testing stages for the Myanmar-Wa language pair.

Table 4.1 Corpus Size for Myanmar-Wa

Dataset	**Number of Sentences**
Training	26970
Validation	4030
Testing	4000
Total	35000

4.5.2 Segmentation

In the intricate landscape of Machine Translation tailored for the Myanmar-Wa corpus, navigating segmentation challenges emerges as a critical endeavor preceding tokenization. This is particularly significant for the Myanmar language, where the absence of white space necessitates careful consideration to determine the most suitable segmentation style—whether word, syllable, etc.—for the target language [40]. Effective segmentation, whether at the word or syllable level, is essential to encapsulate the essence of both languages, laying the groundwork for accurate translation outcomes. We prepared the Myanmar corpus using two distinct segmentation approaches to find the most effective method for translating to Wa:

1. Syllable-Style Segmentation (Myanmar Language to Word-Wa Language): Syllables, representing smaller linguistic units, are used to segment the text, accommodating languages with intricate phonetic features. This approach is particularly useful for capturing the nuances of languages where syllable boundaries are clearer than word boundaries.

2. Word-Style Segmentation (Myanmar Language to Word-Wa Language): This approach focuses on segmenting the text into individual words, making it relevant for languages with more apparent word boundaries. This method leverages the direct word correspondences identified during word alignment, ensuring that the segmentation aligns with the natural linguistic structures of both languages. Sample syllable segmentation and word segmentation on myanmar language are shown in Table 4.2.

Table 4.2 Syllable Segmentation and Word Segmentation on Myanmar Language

Word: ခင်ဗျား_ ဘယ်လောက်_ စောစော_ လို_ ချင်_ လဲ_ ။ Syllable :'ခင် ဗျား ဘယ် လောက် စော စော လို ချင် လဲ'
Word: ကျွန်တော်တို့_ မီးပူတိုက်_ ထား_ တဲ့_ အဝတ်အစား_ တွေ_ ဝတ်_ သ_ င_ ့်_ တယ်_ ။ Syllable: ကျွန် တော် တို့ မီး ပူ တိုက် ထား တဲ့ အ ဝတ် အ စား တွေ ဝတ် သ င့် တယ်
Word: ဝမ်းနည်း_ ပါတယ်_ ကျွန်တော်တို့_ အဲ_ လို_ မ_ လုပ်_ ပေး_ နိုင်_ ပါ_ ဘူး_ ။ Syllable: ဝမ်း နည်း ပါ တယ် ကျွန် တော် တို့ အဲ လို မ လုပ် ပေး နိုင် ပါ ဘူး

Word: ကောင်လေး_ဘာ_လုပ်_နေလဲ_။ Syllable: ကောင် လေး ဘာ လုပ် နေ လဲ
Word: ဘယ်တော့_လောက်_လဲ_။ Syllable: ဘယ် တော့ လောက် လဲ

After testing traditional transfer learning with both word and syllable style segmentation on myanmar corpus, we decided to use word segmentation on later experiments based on the results. Word segmentation often yielded higher METEOR scores, indicating better BLEU according to the following results.

Sample Sentence Scores of Train with Syllable Segmentation on Myanmar Language:

Sample 1
Translated Sentence: tah maox rhiem ka aux sang jah yam mawx kawx
BLEU Score: 0.5387551338654779
METEOR Score: 0.7585268884703913
Sample 2
Translated Sentence: nawh tien tix
BLEU Score: 1.1334226886629 42e-154
METEOR Score: 0.625
Sample 3
Translated Sentence: jao pa tix maix ah yuh nin
BLEU Score: 0.6147881529512643
METEOR Score: 0.7114285714285714

Sample Sentence Scores of Train with Word Segmentation on Myanmar Language:

Sample 1
Translated Sentence: maix tah maox rhiem ka aux sang jah yam mawx
BLEU Score: 0.6432188699036832
METEOR Score: 0.8440677966101694
Sample 2
Translated Sentence: nawh tien
BLEU Score: 9.047424648113057e-155
METEOR Score: 0.6465517241379309
Sample 3
Translated Sentence: jao pa tix maix ah nin
BLEU Score: 0.6431870218238024
METEOR Score: 0.7217391304347825

4.5.3 Tokenization and Encoding on Myanmar-Wa

Tokenization and encoding play a crucial role in the effectiveness of the FusionTOK. By leveraging MyanBERTa for tokenization, we effectively represent input sentences in the Myanmar language as sequences of tokens, preserving their semantic and syntactic information. The encoded representations are then fed into the decoder, facebook/bart-base tokenizer, which decodes the tokens and generates target language sentences. This seamless integration of tokenization and encoding, facilitated by FusionTOK, ensures smooth communication between the encoder and decoder, ultimately leading to better translation results. Tokenization examples provided:

Before segmentaion on myanmar text: ဒီမြို့ကတကယ့်ကိုစိတ်ဝင်စားဖို့ကောင်းတယ်

Tokenized with MyanBERTa: [14375, 71309, 1910, 203421, 170987, 114273, 59392, 24561, 45086, 69301, 78615]

After segmentaion on myanmar text: ဒီ မြို့ က တကယ့် ကို စိတ် ဝင် စား ဖို့ ကောင်း တယ်

Tokenized with MyanBERTa: [14375, 259, 71309, 259, 1910, 259, 203421, 6697, 259, 1975, 121633, 259, 59392, 145636, 259, 45086, 259, 69301, 259, 78615]

Wa text: veng in mawh pa mhawm laigrhawm tete telai

Tokenized with facebook/bart-base: [14375, 259, 71309, 259, 1910, 259, 203421, 6697, 259, 1975, 121633, 259, 59392, 145636, 259, 45086, 259, 69301, 259, 78615]

4.6 Process flow of Proposed Model

In this section, we provide a process flow of the proposed model. The process flow of the model building used in the Myanmar (Burmese)-Wa corpus for Machine Translation is shown in Figure 4.9. By employing Tokenizer Fusion on MyanBERTa + facebook/bart-base, with Facebook BART as the decoder and Myanmar BERT as the encoder. This fusion strategy involves the following steps.

4.6.1 Corpus Preparation

A Myanmar-Wa parallel corpus is meticulously prepared, ensuring that each sentence in Myanmar has a corresponding and accurately aligned translation in Wa. This parallel corpus serves as the foundational dataset for training the translation model.

4.6.2 Segmentation

The Myanmar text undergoes a segmentation process. Myanmar Word Segmentation Version 1.0 is used for word segmentation for myanmar text [55]. This step involves breaking the continuous flow of characters into meaningful units, typically words or phrases, which is crucial for subsequent tokenization.

4.6.3 Tokenizing

The input text in the source language (Myanmar) is tokenized using the MyanBERTa. MyanBERTa is adept at breaking down sentences into tokens that capture the linguistic nuances specific to the Myanmar language. Concurrently, the target language text in Wa is tokenized using the facebook/bart-base. This ensures compatibility with the model's decoder and prepares the Wa text for effective translation.

4.6.4 Model Training

The model training process involves utilizing the pre-trained facebook/bart-base model as a foundational framework. FusionTOK integrates the token sequences from both the MyanBERTa and facebook/bart-base tokenizers, aligning them to form a unified input-output token sequence pair. This fusion allows the model to simultaneously capture the linguistic characteristics of both the source (Myanmar) and target (Wa) languages, thereby enabling more accurate and contextually relevant translations.

4.6.5 Model Saving and Translation

Upon completion of the training process, the FusionTOK model is preserved and readied for operational deployment. This involves saving the trained model to a persistent storage medium, allowing for its subsequent use in translation tasks. When translating Myanmar text into Wa text, the input Myanmar text must first undergo segmentation. This segmentation step is critical to ensure that the text aligns with the tokenization process employed during model training. Properly segmenting the Myanmar text prepares it for accurate processing by the model, enabling effective and contextually appropriate translations into Wa.

Summary, the training process for the FusionTOK involves fine-tuning the pre-trained facebook/bart-base model on our My-Wa corpus, a parallel corpus of Myanmar and Wa languages, using transfer learning techniques. We fine-tune the model parameters to adapt them to the characteristics of the target language pair. During training, we optimize a combined loss function that incorporates both reconstruction loss and translation loss to ensure the generation of accurate and contextually relevant translations.

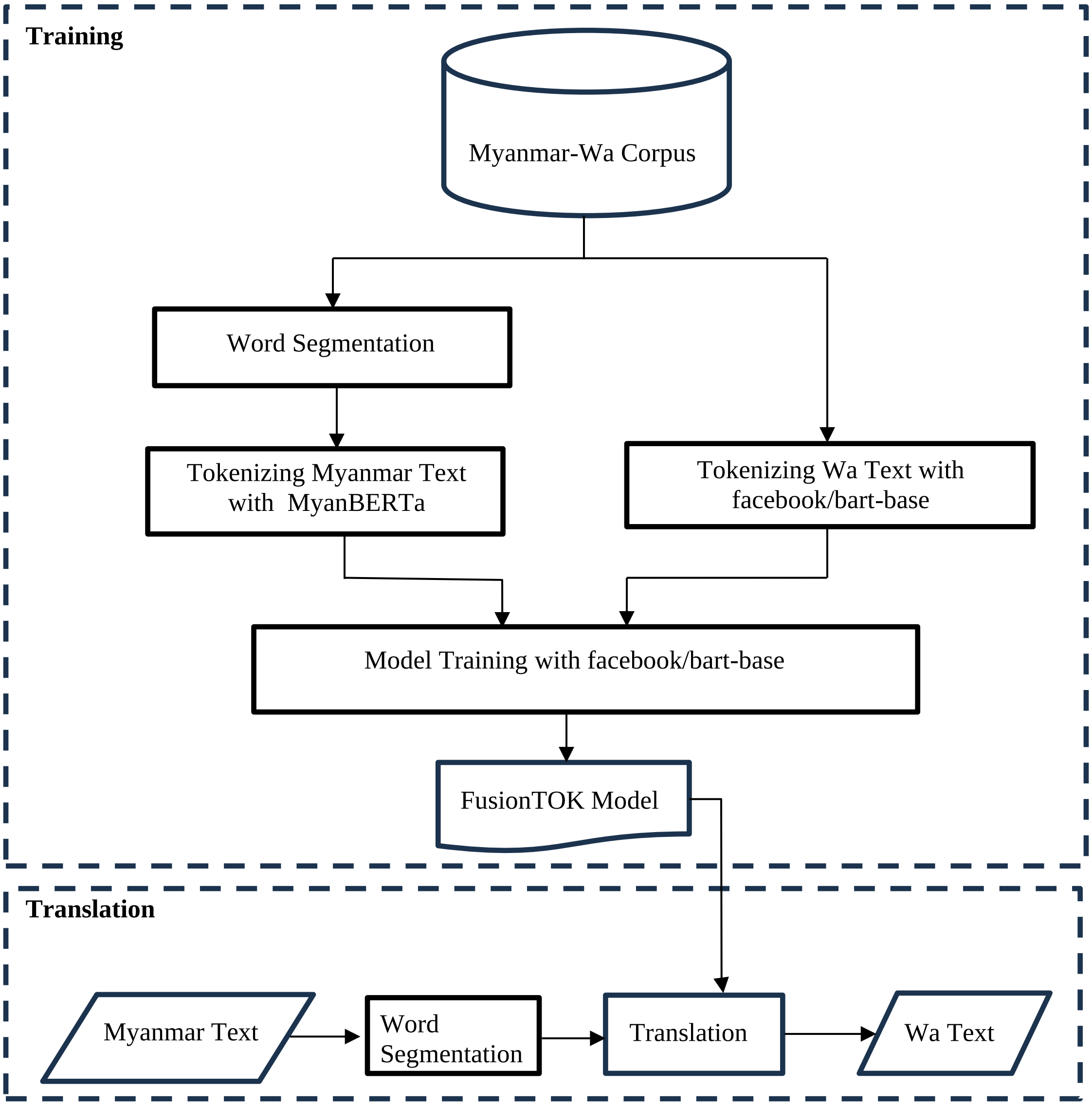


Figure 4.9 Process Flow Diagram of the Proposed Model

The effectiveness of this tokenizer fusion is crucial, as selecting the correct tokenizer is vital for accurately capturing the nuances and structures of both source and target languages. Highlighted are the validation procedures, training process, and model evolution, which demonstrate how flexible and successful the suggested strategy is at bridging language obstacles.

Advances in Natural Language Processing (NLP) and the availability of large parallel datasets have propelled the extraordinary success of Machine Translation (Machine Translation). Notwithstanding these advancements, there have been significant obstacles in Machine Translation's adaptation to the particular linguistic demands of the Myanmar-Wa language pair. The insufficient number of aligned sentences in the Myanmar-Wa corpus hampered the creation of trustworthy translation models in the early attempts, which were limited and flawed.

The study of next portion delves into linguistic analysis, breaking down the distinctive qualities of the Wa language and looking at the metrics and processes used to assess the effectiveness of the Machine Translation system. This analysis focuses on the practical applications and repercussions of enhanced Machine Translation for the Myanmar-Wa language pair, highlighting the critical role that improved Machine Translation plays in cultural preservation and the larger goal of language revitalization. Ultimately, a review of the key findings, a discussion of their ramifications, and closing thoughts on the dynamic field of Machine Translation and its crucial role in bridging linguistic gaps round out the model’s examination.

4.7 Chapter Summary

In this chapter, we provide a comprehensive overview of the methodologies underlying the proposed FusionTOK model. We begin with a detailed design and description of the model, meticulously outlining its architecture, core components, and the rationale behind its development. This includes an exploration of the innovative aspects of FusionTOK, such as its integration of transfer learning and language-specific tokenizers, which enhance its ability to capture cross-linguistic dependencies. We further delve into the experimental setup, describing the process of corpus preparation, segmentation, tokenization, model training, and the fusion of token sequences from both source and target languages. The chapter concludes with an explanation of the model saving and translation process, ensuring that the model is optimally prepared for deployment and practical use in translating Myanmar text into Wa text. Through this detailed exposition, we aim to elucidate the technical intricacies and innovative strategies that underpin the FusionTOK model, demonstrating its potential to significantly advance the field of machine translation.

CHAPTER 5

EXPERIMENTAL RESULTS

This chapter provides an in-depth exploration of the intricate process of constructing Machine Translation models leveraging transfer learning techniques. It carefully outlines each stage of development, from data preprocessing to model architecture selection, placing a particular emphasis on the significant role that Machine Translation plays in bridging linguistic divides. Furthermore, it presents comprehensive evaluation results, including metrics such as accuracy, fluency, and computational efficiency, shedding light on the effectiveness of the developed models. A thorough performance analysis is conducted, examining the strengths and limitations of the models in different contexts and scenarios.

5.1 Sentence Length Analysis

Sentence length has a significant impact on the efficiency of Machine Translation models, especially for languages with different syntactic patterns like Wa. This subsection examines the effect of sentence length on Myanmar-Wa Machine Translation. The examination entails a thorough analysis of the distribution of sentence lengths in the source and destination languages, which may indicate associations with translation quality.

Maintaining context and meaning in the target language can be difficult when dealing with longer sentences in the source language. Comprehending the distribution of phrase lengths facilitates the efficient adaptation of Machine Translation models. This visual aid functions as a first investigation, opening the door to additional quantitative examination and improvements in models specifically designed for the Myanmar-Wa language combination.

Longer phrases in the original language might be difficult to translate, especially when it comes to keeping the target language's meaning and context. Knowing the distribution of sentence lengths makes it easier to modify Machine Translation models so they can better handle the linguistic differences found in the Myanmar-Wa corpus. By providing a basic investigation of sentence length features, this visualization paves the way for additional research and advancements in Machine Translation models for the Myanmar-Wa language pair.

5.2 Baseline Models

We establish a foundation for comparison by implementing three fundamental models: a Transformer model, a plain text transfer learning model and a Transfer Learning model using bridge language. The Transformer model serves as a widely recognized benchmark for sequence-to-sequence tasks, offering a standard reference point for evaluating the performance of our proposed Fusion Token Model (FusionTOK). On the other hand, the plain text transfer learning model represents a simpler approach, relying solely on pre-trained embeddings without specialized tokenization techniques. In addition to the Transformer model and the plain text transfer learning model, we introduce Bridge Language Transfer Learning (BiT5 model) as another baseline model. This approach leverages a bridge language to facilitate translation between source and target languages. By first translating the source language into the bridge language and then translating it into the target language, this model aims to bridge the linguistic gap between language pairs with limited parallel data. We implement and evaluate this approach to assess its effectiveness in handling translation tasks, particularly in scenarios where direct translation between the source and target languages is challenging due to data scarcity or linguistic differences. These baseline models enable us to assess the efficacy and superiority of our FusionTOK architecture that is Tokenizer Fusion strategy against established methods, providing valuable insights into the advancements achieved in Machine Translation technology.

5.3 Evaluation Metrics

Evaluation of Machine Translation for Myanmar-Wa faces particular difficulties because of linguistic differences and small parallel corpora. Performance evaluation makes use of a range of measures and methods.

In assessing the efficacy of our translation models, we employ the BLEU Score as a fundamental metric, providing a quantitative measure of the similarity between the machine-generated translations and human reference translations. While BLEU is widely used and provides valuable insights into the overall performance of Machine Translation systems, its effectiveness may be limited in language pairs with sparse resources, such as the Myanmar-Wa language pair. The nuances and complexities inherent in these languages may not be fully captured by BLEU, leading to potential discrepancies between the scores and the actual translation quality. Despite its limitations, BLEU remains a valuable tool for gauging the general adequacy of translations and serves as a benchmark for comparison.

In addition to the BLEU Score, we leverage the METEOR Score to obtain a more nuanced evaluation of translation quality. METEOR takes into account a broader range of linguistic features, including synonyms, word order, and stemming, offering a more comprehensive assessment compared to BLEU. This metric is particularly beneficial for capturing the intricacies of translation in language pairs with limited resources, where subtle variations in language usage can significantly impact translation accuracy. By incorporating METEOR alongside BLEU, we aim to provide a more holistic perspective on the performance of our translation models, ensuring that they not only produce fluent translations but also maintain fidelity to the original meaning and context.

By combining these quantitative and qualitative assessments, Machine Translation for Myanmar-Wa may be thoroughly understood in terms of their advantages and disadvantages, which will direct future developments in the sector.

5.3.1 Evaluation on Model Performance

This aspect of the evaluation strategy involves segmenting sentences into different length categories: short (1-5 words), middle (1-10 words), and long (1-15 words). By doing so, the evaluation aims to understand how the model performs across varying sentence lengths, which is crucial for assessing its robustness and effectiveness in real-world translation scenarios.

The performance range refers to the specific segment of sentence lengths (1-10 words) where the model is analyzed in detail. This range is identified as the model's optimal performance zone, indicating where it demonstrates superior translation accuracy and fluency. Understanding this optimal range helps to highlight the strengths and weaknesses of the model more effectively.

5.3.2 Model Performance Comparison on BLEU Scores

FusionTOK Model: Achieves the highest BLEU score among all models evaluated, indicating a significant n-gram overlap and commendable alignment with translations from human references. This suggests that the FusionTOK Model performs exceptionally well in capturing linguistic nuances and achieving translation accuracy.

BiT5 (Bridge Language Transfer Learning) and Traditional Transfer Learning Model: These models show balanced performance across parameters but have slightly lower BLEU scores compared to the FusionTOK Model. However, they still demonstrate proficiency in translation, making them robust options for various translation tasks.

TransformerNMT: This model performs well in terms of BLEU scores but lags behind in METEOR scores. While it indicates good performance in terms of n-gram overlap, it may face challenges in accurately capturing semantic accuracy and linguistic nuances. Figure 5.1 shows the BLEU score analysis on four Myanmar-Wa Machine Translation Models.

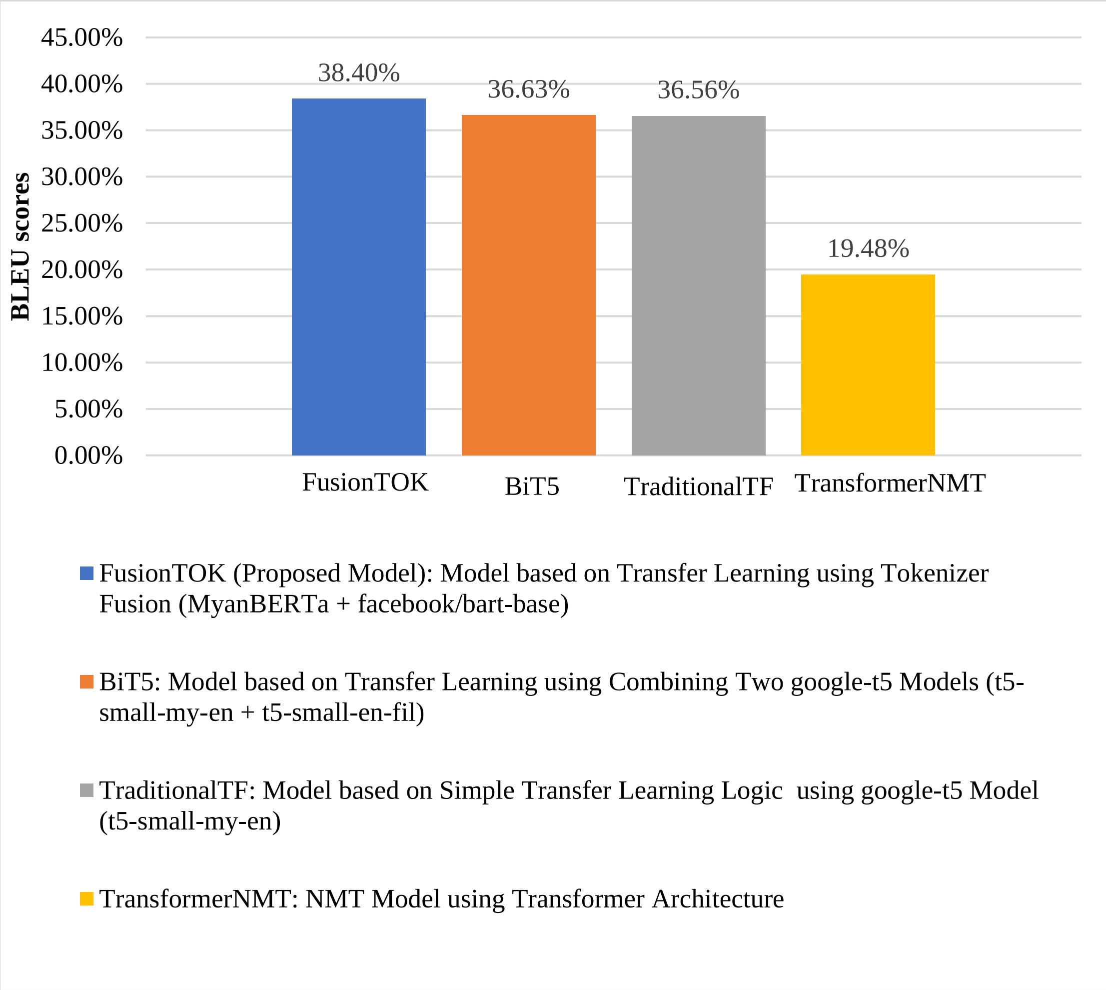


Figure 5.1 BLEU Score Analysis on Myanmar-Wa Machine Translation Models

5.3.3 Model Performance Comparison on METEOR Scores

FusionTOK Model: Similarly, the FusionTOK Model achieves the highest METEOR score, indicating its superior performance in terms of semantic accuracy and fluency. This suggests that the model not only captures n-gram overlap effectively but also excels in preserving the meaning and coherence of the translated text.

BiT5 (Bridge Language Transfer Learning) and Traditional Transfer Learning Model: These models show comparable METEOR scores, suggesting that they perform well in terms of semantic accuracy and fluency, although slightly lower than the FusionTOK Model. They still demonstrate proficiency in capturing the meaning of the translated text.

TransformerNMT: This model lags behind in terms of METEOR scores compared to the other models. While it may perform adequately in terms of n-gram overlap, it may struggle with preserving semantic accuracy and fluency in the translated text. Figure 5.2 shows the METEOR score analysis on four Myanmar-Wa Machine Translation Models.

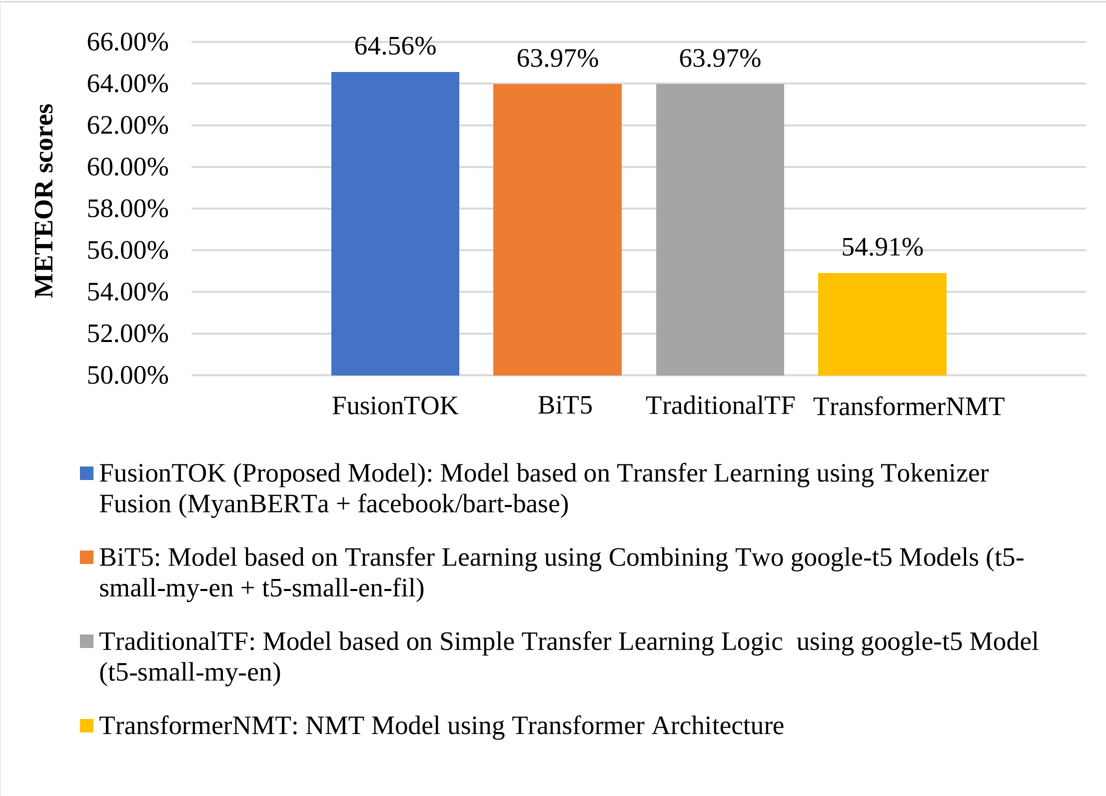


Figure 5.2 METEOR Score Analysis on Myanmar-Wa Machine Translation Models

5.4 Discussion

The FusionTOK Model, which combines MyanBERTa and facebook/bart-base tokenizers, outperforms the other models for several reasons:

1. Effective Tokenizer Fusion: By leveraging two distinct tokenizers, the FusionTOK Model can capture a broader range of linguistic nuances and syntactic structures present in the Myanmar-Wa Corpus. This allows it to achieve higher alignment with translations from human references, resulting in superior translation accuracy.

2. Comprehensive Representation: MyanBERTa and facebook/bart-base tokenizers provide comprehensive representations of the source and target languages, respectively. The FusionTOK Model's ability to fuse these representations enables it to capture both the intricacies of the source language and the target language, resulting in more accurate translations.

3. Semantic Accuracy: The FusionTOK Model excels in preserving semantic accuracy, ensuring that the translated text maintains the intended meaning and coherence of the original sentences. This is crucial for delivering high-quality translations that effectively convey the intended message to the target audience.

4. Fluency: In addition to semantic accuracy, the FusionTOK Model demonstrates fluency in the translated text, making it more natural and readable. This fluency enhances the overall user experience and ensures that the translated content is accessible and engaging to the target audience.

5. Optimal Performance Zone: The FusionTOK Model's performance is particularly noteworthy for short to medium-length sentences (1-10 words), which constitute the optimal performance zone. Within this range, the model demonstrates superior translation accuracy and fluency, highlighting its effectiveness in capturing linguistic nuances and semantic accuracy.

Overall, the FusionTOK Model's success can be attributed to its innovative tokenizer fusion architecture, which effectively combines the strengths of MyanBERTa and facebook/bart-base tokenizers to deliver high-quality translations with superior linguistic accuracy, semantic precision, and fluency.

More training data may improve the performance of models. These nuanced evaluations could help researchers choose models for Machine Translation jobs by acknowledging that different metrics highlight different elements of performance. Further research could look into ways to improve semantic preservation and use more data to improve each model's ability to refine itself. The ongoing development of Machine Translation models facilitates communication and language understanding.

It is imperative to note that the Myanmar-Wa corpus used in these experiments is the very first corpus created by the research team. The corpus size was expanded to enhance the Machine Translation performance and evaluation.

5.5 Chapter Summary

This chapter serves as a comprehensive exploration into the proposed model's performance, elucidating the meticulous steps taken in data collection and the subsequent implementation of these models. The chapter meticulously dissects the Machine Translation models, delving into their efficacy through a nuanced analysis. With the aid of detailed charts and analytical findings, it not only highlights the models' translation accuracy and fluency but also offers a deeper understanding of their strengths and limitations. By presenting a thorough evaluation of the models' performance, this chapter significantly advances the field of Machine Translation research, providing researchers and practitioners with valuable insights into the intricacies of translation technology.

CHAPTER 6

CONCLUSION

Our journey through the landscape of Machine Translation (MT) techniques within the intricate domain of Myanmar-Wa languages has been an intellectual voyage, marked by the pursuit of understanding, the discovery of new frontiers, and the illumination of pathways towards linguistic inclusivity. As we embark upon the final leg of our journey, it is imperative to delve deeply into the multifaceted terrain we have traversed, unraveling the complexities, nuances, and implications of our findings.

6.1 Myanmar-Wa Machine Translation based on Transfer Learning

At the heart of our exploration lies the paradigm of transfer learning, a concept that has revolutionized the field of MT and redefined our understanding of language processing. Transfer learning, in essence, involves leveraging knowledge acquired from one domain or task and applying it to another, thereby accelerating the learning process and improving performance. In the context of MT, transfer learning has emerged as a transformative approach, offering a pathway to enhanced translation quality, adaptability across diverse language pairs, and improved efficiency.

Our examination of various MT models, including Transformer, Plain text (traditional) Transfer Learning, BiT5, and FusionTOK, has underscored the pivotal role of transfer learning in shaping the landscape of MT technology. Models like BiT5, which combine multiple pre-trained models and fine-tune them on domain-specific data, exemplify the power of transfer learning in achieving superior translation performance. Similarly, FusionTOK, through its integration of Tokenizer Fusion techniques, demonstrates the potential of transfer learning in enhancing linguistic precision and contextual fidelity.

The implications of transfer learning extend far beyond the realm of MT, permeating various domains of artificial intelligence and machine learning. By capitalizing on the wealth of knowledge encoded in pre-trained models and refining it through domain-specific fine-tuning, transfer learning has the potential to unlock new frontiers of innovation and discovery in language processing and beyond.

6.2 Pros and Cons of Myanmar-Wa Machine Translation Models

Our comparative analysis of MT models has provided valuable insights into the nuances and trade-offs inherent in translation quality metrics such as METEOR and BLEU scores. While each model exhibits strengths in specific areas, it is evident that there is no one-size-fits-all solution in the realm of MT. Instead, the selection of an appropriate MT model depends on various factors, including the characteristics of the target language pair, the availability of training data, and the specific requirements of the task at hand.

Models like BiT5 have demonstrated exceptional proficiency in integrating diverse pre-trained models, leveraging their collective strengths to achieve superior translation performance. Through the fusion of various techniques such as transfer learning and fine-tuning, BiT5 has showcased remarkable versatility and adaptability across different language pairs.

However, it is FusionTOK that stands out as a beacon of innovation in the realm of MT. This model showcases the power of Tokenizer Fusion techniques in enhancing linguistic precision and contextual fidelity, thereby elevating translation quality to new heights. By seamlessly blending MyanBERT Tokenizer with facebook/bart-base Tokenizer through FusionTOK, we have witnessed a quantum leap in the accuracy and coherence of translations.

In our comparative analysis, FusionTOK has emerged as a trailblazer, pushing the boundaries of MT technology and setting a new standard for linguistic excellence. Its ability to capture subtle nuances and preserve contextual richness underscores its pivotal role in bridging linguistic divides and fostering cross-cultural communication.

However, our exploration has also revealed significant challenges that must be addressed to fully harness the potential of MT technology. Chief among these challenges is the issue of data scarcity, particularly for less-resourced languages like Wa. The limited availability of training data poses a significant obstacle to achieving optimal translation quality, highlighting the need for innovative strategies for data collection and augmentation.

Furthermore, the complexity of MT models presents a barrier to accessibility and usability, requiring specialized technical expertise for effective implementation and customization.

6.3 Future Research

Looking to the future, there are several promising avenues for further research and development in the field of MT within the Myanmar-Wa context. Fine-tuning the intricacies of MT models represents a critical area of focus, as we seek to enhance translation quality and bridge the gap between languages. This entails delving deeper into the architecture of these models and exploring novel techniques for capturing semantic richness and contextual fidelity.

Moreover, the integration of additional data sources holds immense potential for augmenting the learning process and improving the robustness of MT systems. By incorporating diverse datasets that capture the nuances of language usage across different contexts and domains, we can enhance the adaptability of MT models and ensure more accurate and contextually relevant translations.

Furthermore, the expansion of the Myanmar-Wa corpus emerges as a strategic imperative for advancing MT research in this domain. By curating a comprehensive dataset that encompasses a wide range of linguistic expressions and cultural nuances, we can provide MT models with the necessary training data to better understand and accurately translate the nuances of the Myanmar-Wa language pair.

In conclusion, our expedition into the realm of MT techniques within the Myanmar-Wa context has provided valuable insights and opened new horizons for research and application. By addressing the challenges and seizing the opportunities that lie ahead, we can continue to push the boundaries of MT technology and foster greater communication and understanding across linguistic divides.

AUTHOR'S PUBLICATIONS

1. F. Yune and K. M. Soe, "Myanmar-Wa Machine Translation using LSTM-based Encoder-Decoder Model," presented at the 2023 IEEE Conference on Computer Applications **(ICCA)**, Yangon, Myanmar, 2023, pp. 1-5. doi: 10.1109/ICCA51723.2023.10181692.

2. F. Yune and K. M. Soe, "TransLingua: A Transfer Learning Approach to Enhancing Myanmar-Wa Neural Machine Translation," presented at the **iccr2023** conference. Available online: [TransLingua Paper] (https://public.thinkonweb.com/sites/iccr2023/media?key=site/iccr2023/abs/B-6-1.pdf).

3. F. Yune and K. M. Soe, "Advancing Transfer Learning. Paradigms for Myanmar (Burmese) to Wa. (Austroasiatic Language Family) Language Translation," presented at the 26th INTERNATIONAL CONFERENCE ON Oriental-COCOSDA **(OCOCOSDA 2023)**, December 4-6, 2023. Conference website: [OCOCOSDA 2023](https://www.igdtuw.ac.in/uploads/OCOCOSDA%202023%20Brochure%20May%2023.pdf).

4. F. Yune and K. M. Soe, "Tracing the Evolution of Machine Translation: A Journey through the Myanmar (Burmese)-Wa (sub-group of the Austro-Asiatic language) Corpus," Advances in Science, Technology and Engineering Systems Journal, vol. 9, pp. 79–90, 2024. ISSN: 24156698. doi: 10.25046/aj090108. (**Scimago index –Q3**).

REFERENCES

[1] Aye Mya Hlaing, Win Pa Pa, "MyanBERTa: A Pre-trained Language Model For Myanmar", In Proceedings of 2022 International Conference on Communication and Computer Research (ICCR2022), November 2022, Seoul, Republic of Korea.

[2] Callison-Burch, C., & et al. (2008). "The META-SHARE Metadata Schema for the Description of Language Resources". In LREC.

[3] Chen, Y., et al., "A teacher-student framework for zero-resource neural Machine Translation," in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2017.

[4] Cheng, Y., et al., "Joint training for pivot-based neural Machine Translation," in Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2017.

[5] Conneau, A., and Lample, G., "Crosslingual language model pretraining," arXiv:1901.07291, 2019.

[6] Dabre, R., et al., "Exploiting multilingualism through multistage fine-tuning for low-resource neural Machine Translation," in Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.

[7] David, C., "Hierarchical Phrase-Based Translation," Computational Linguistics, vol. 33, no. 2, pp. 201-228, 2007.

[8] Deng, J., et al., "ImageNet: A comprehensive image database for visual object recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[9] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding". arXiv preprint arXiv:1810.04805.

[10] Doddington, G. (2002). "Automatic Evaluation of Machine Translation Quality Using N-gram Co-occurrence Statistics". Association for Computational Linguistics.

[11] Dwei, B. S., and Masayoshi, A., "Multiple pivots in statistical Machine Translation for low resource languages," in Proc. PACLIC, Hakodate, Japan, 2019, pp. 345–355.

[12] Firat, O., et al., "Zero-resource translation with multi-lingual neural Machine Translation," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016.

[13] Gheini, M., and May, J., "A universal parent model for low-resource neural Machine Translation transfer," arXiv:1909.06516, 2019.

[14] Graham Neubig, Junjie Hu, "Rapid Adaptation of Neural Machine Translation to New Languages", EMNLP 2018, 13 Aug 2018.

[15] Hua Wu and Haifeng Wang. Pivot language approach for phrase-based statistical machine translation. Machine Translation, 2007.

[16] Isozaki, H., et al., "Automatic evaluation of translation quality for distant language pairs," In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, MIT, Massachusetts, USA, October, 2010, pp. 944-952.

[17] Ji, B., et al., "Cross-lingual pre-training based transfer for zero-shot neural Machine Translation," in Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 2020.

[18] Jhnson, M., et al., "Google's multilingual neural Machine Translation system: Enabling zero-shot translation," arXiv:1611.04558, 2016.

[19] Karakanta, A., et al., "Neural Machine Translation for low-resource languages without parallel corpora," Machine Translation, pp. 167-189, 2018.

[20] Kim, Y., Gao, Y., and Ney., H., 2019. "Effective cross-lingual transfer of neural machine translation models without shared vocabularies". In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1246–1257, Florence, Italy.

[21] Kim, Y., et al., "Pivot-based transfer learning for neural Machine Translation between non-English languages," in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019.

[22] Kim, Y., Petrov, P., Petrushkov, P., Khadivi, S., & Ney, H. (2019). "Pivot-based transfer learning for neural Machine Translation between non-English languages". In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 866–876). Association for Computational Linguistics.

[23] Kishore, P., et al., “BLEU: a Method for automatic evaluation of Machine Translation,” IBM Research Report rc22176 (w0109022), Thomas J. Watson Research Center, 2001.

[24] Kocmi, T., and Bojar, O., "Trivial transfer learning for low-resource neural Machine Translation," in Proceedings of the Third Conference on Machine Translation: Research Papers, 2018.

[25] Lampe, G., et al., "Cross-lingual Language Model Pretraining," in Advances in Neural Information Processing Systems, 2019.

[26] Lavie, A., & Agarwal, A. (2007). "An Automatic Metric for Machine Translation Evaluation with Improved Correlation with Human Judgments". Association for Computational Linguistics.

[27] Leng, Y., et al., "Unsupervised pivot translation for distant languages," in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019.

[28] Lewis, A., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension". arXiv preprint arXiv:1910.13461.

[29] Lin, Y., et al., "Choosing transfer languages for cross-lingual learning," in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019. “Choosing transfer languages for cross-lingual learning”. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 3125–3135, Florence, Italy. Association for Computational Linguistics.

[30] Ling, W., et al., "Character-based Neural Machine Translation," in Proceedings of the International Conference on Learning Representations (ICLR), 2016.

[31] Locke, W. N., and Booth, A. D., "Machine Translation of languages," in Fourteen essays, MIT Press, 1955.

[32] Martin, J. H., and Jurafsky, D., “Speech and language processing,” International Edition, 2000.

[33] Neubig, G. and Hu, J., "Rapid Adaptation of Neural Machine Translation to New Languages," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018.

[34] Papineni, K., & et al. (2002). "A Method for Automatic Evaluation of Machine Translation". Association for Computational Linguistics.

[35] Paul, M., et al., "On the importance of pivot language selection for statistical Machine Translation," in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2009.

[36] Raffel, C., et al., "Exploring transfer learning limits with a unified text-to-text transformer," Journal of Machine Learning Research, vol. 21, pp. 140:1-140:67, 2020.

[37] Raffel, C., et al., “Exploring the limits of transfer learning with a unified text-to-text transformer,” J. Mach. Learn. Res., 21:140:1–140:67, 2020.

[38] Thazin Myint OO, Ye Kyaw Thu, Khin Mar Soe and Thepchai Supnithi, “Statistical Machine Translation between Myanmar (Burmese) and Dawei (Tavoyan)”, The First Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019), 11-13 September 2019, Trento, Italy.

[39] Thazin Myint Oo, Ye Kyaw Thu, Khin Mar Soe, “Statistical Machine Translation between Myanmar (Burmese) and Rakhine (Arakanese)”, In Proceedings of ICCA2018, February 22-23, 2018, Yangon, Myanmar, pp. 304-31.

[40] Thet Thet Zin, Khin Mar Soe, and Ni Lar Thein, “Myanmar phrases translation model with morphological analysis for statistical Myanmar to English translation system”, In Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation, Singapore, December, 2011, pp. 130-139.

[41] Tiedemann, J., "Character-based pivot translation for under-resourced languages and domains," in Proceedings of the European Chapter of the Association for Computational Linguistics (EACL), 2012.

[42] Utiyama, M., and Isahara, H., "A comparison of pivot methods for phrase-based statistical Machine Translation," in Proceedings of the Human Language Technologies, 2007.

[43] Utiyama, M., & Isahara, H. (2007). "A comparison of pivot methods for phrase-based statistical Machine Translation". In Proceedings of Human Language Technologies, The Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2007) (pp. 484–491). Rochester, USA.

[44] Vaswani, A., et al., "Attention is All You Need," arXiv:1706.03762, 2017.

[45] Vaswani, A., et al., "Attention is All You Need," CoRR, abs/1706.03762, 2017.

[46] Vaswani, A., et al., “Attention is all you need”, 2017.

[47] Vaswani, A., et al., “Attention Is All You Need,” In Advances in Neural Information Processing Systems, 5998–6008, 2017.

[48] Wu, H., and Wang, H., "Pivot language approach for phrase-based statistical Machine Translation," Machine Translation, 2007.

[49] Wu, J., et al., "Adapting Attention-based Neural Network to Low-resource Mongolian-Chinese Machine Translation," Springer Verlag Berlin Heidelberg, 2011.

[50] Win Pa Pa, Ni Lar Thein, "Myanmar Word Segmentation using Hybrid Approach", In Proceedings of 6th International Conference on Computer Applications, Yangon, Myanmar, pp. 166-170, 2008.

[51] Win Pa Pa., et al., "A study of statistical Machine Translation methods for under-resourced languages," 5th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU Workshop), Procedia Computer Science, Yogyakarta, Indonesia, Volume 81, May 09-12, 2016, pp. 250–257.

[52] Xie, T., "Burmese (Myanmar): An Introduction to the Script," Southeast Asian Language Resource Center, 2008.

[53] Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2020). "mT5: A massively multilingual pre-trained text-to-text transformer". arXiv preprint arXiv:2010.11934.

[54] Ye Kyaw Thu, Andrew Finch, Win Pa Pa, and Eiichiro Sumita, “A large-scale study of statistical machine translation methods for Myanmar language”, In Proceedings of Symposium on Natural Language Processing 2016, February 10-12, 2016.

[55] Zettlemoyer, L. (2020). "Multilingual denoising pre-training for neural Machine Translation".

[56] Zheng, H., et al., "Maximum expected likelihood estimation for zero-resource neural Machine Translation," in Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2017.

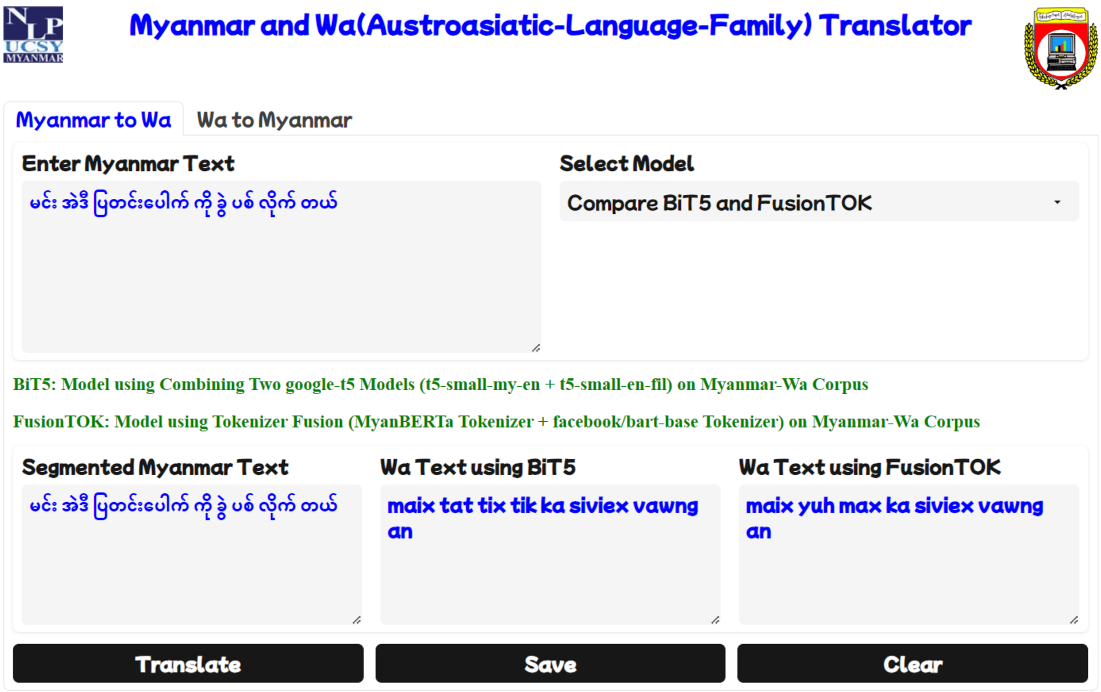
[57] Zhu, X., et al., "Improving pivot-based statistical Machine Translation using random walk," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2013.

[58] Zhu, X., et al., "Improving pivot-based statistical Machine Translation using random walk," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2013.

[59] Zoph, B., et al., "Transfer learning for low-resource neural Machine Translation," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016.

Appendix I: Interface Designs of Myanmar-Wa Translator

The following figures show the interface designs to translate the Myanmar Language to Wa Language. This step accepts the Myanmar Language to translate and then generate the Wa Language text. Wa to Myanmar interface is supposed to check the translation. By following this structured approach and incorporating back translation as a quality assurance measure, the translation model can ensure accurate and reliable translations, thereby enhancing its utility and effectiveness in practical applications.



Back Translation for Quality Checking

To ensure translation accuracy, the process begins by inputting the Wa text. Following this, users click on the "Translate" button, initiating the translation from Wa to Myanmar. Subsequently, the translated Myanmar text is copied and pasted into the assigned text box for translation back to Wa. Users then have the option to select between the BiT5 and FusionTOK Model or Compare_BiT5_ and_FusionTOK. Finally, clicking the "Translate" button executes the translation of the Myanmar text back to Wa, completing the process.

Our interface offers FusionTOK and BiT5 models, both based on transfer learning principles and trained on the Myanmar-Wa Corpus. FusionTOK utilizes tokenizer fusion techniques, combining the MyanBERTa Tokenizer with the facebook/bart-base Tokenizer, while BiT5 combines two google-t5 models, specifically t5-small-my-en and t5-small-en-fil. These models provide users with options for accurate and efficient language processing tasks.

Additional Information: The back translation strategy is employed as a quality assurance measure to validate the accuracy of the translation. By translating the output back to the original language, discrepancies or errors in the translation can be identified. This iterative process helps refine the translation model and improve its overall performance.

Manual Input and Translation

In the manual input and translation process, users follow a structured sequence to initiate and execute translations. Firstly, in Step 1, users have the option to manually input Myanmar text or select from a predefined set of available options. Then, in Step 2, users choose their desired translation model, selecting from options such as BiT5 (my-en+en-fil), FusionTOK (tokenizer fusion on MyanBERTa Tokenizer + facebook/bart-baseTokenizer), or Compare_BiT5_ and_FusionTOK. Finally, in Step 3, users trigger the translation process by clicking the "Translate" button, prompting the system to generate the corresponding Wa translation.

Output Saving

Regarding output saving, the system provides functionality for preserving translated outputs for future reference or analysis. Users can achieve this by clicking the "Save" button, which stores the translated output in a retrievable format. These saved outputs serve multifaceted purposes, including acting as valuable records for tracking translation history, evaluating model performance over time, and facilitating ongoing research and analysis endeavors.