Addis Ababa University

Addis Ababa Institute Of Technology

School Of Electrical And Computer Engineering

Bi-Directional English-Afan Oromo Machine Translation Using Convolutional Neural Network

By

Arfaso Birhanu

Advisor

Mr. Menore Tekeba

A Thesis submitted to the school of graduate studies of Addis Ababa University in partial fulfillment of the requirement for the Degree of Master of Science in Computer Engineering

October 14, 2019

Addis Ababa, Ethiopia

Addis Ababa University

Addis Ababa institute of Technology

School of Electrical and Computer Engineering

Bi-Directional English-Afan Oromo

Machine Translation

Using Convolutional Neural Network

By

Arfaso Birhanu

Approved by the board of Examiners

Mr. Menore Tekeba ____________________

Advisor Signature

____________________________________ ____________________

Internal Examiner Signature

____________________________________ ____________________

External Examiner Signature

Declaration

I, Arfaso Birhanu, hereby declare that “Bi-Directional English-Afan Oromo Machine Translation Using Convolutional Neural Network” is my original work and has not been presented for a degree in any other University and that all sources of materials used for the work have been only acknowledged.

Declared by

Arfaso Birhanu	________________________	October 14, 2019
Student	Signature	Date

Confirmed by

Mr. Menore Tekeba	________________________	October 14, 2019
Advisor	Signature	Date

Acknowledgement

First of all I want to thank the Almighty God whose protection is always with me and that he gave me the opportunity and strength to achieve whatever I have achieved so far.

Next, I would like to thank my Advisor Mr. Menore Tekaba for his supportive advising from beginning up to the end of this thesis work that helped me to complete my study.

Thirdly, I would like to express my gratitude to Dr. Fitsum Andargie for his kindness to support me by shaping the structure of my document.

Fourthly, I would like to forward my deepest appreciation and gratitude to Dr. Getacho Alemu for taking time to see my document which contributed a lot to this study.

Next, I would like to thank all those helped me through the progressive of this study by showing me a direction in their comments.

Lastly, I would like to thank my friends those supported me during data collection for their contribution is a lot to this study.

Table of Contents

List of Figures and Tables

List of Figures

List of Tables

Acronyms and Abbreviation

BLEU	Bi-Lingual Evaluation Understudy
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DNN	Deep Neural Network
FDRE	Federal Democratic Republic of Ethiopia
GLU	Gated Linear Unit
GRU	Gated Recurrent Neural Network
LSTM	Long Short Term Memory
NMT	Neural Machine Translation
POS	Part of Speech
RBMT	Rule Based Machine Translation
RAM	Random Access Memory
RNN	Recurrent Neural Network
RLU	Rectified Linear Unit
RMSProp	Root Mean Square Prop
STM	Statistical Machine Translation

Abstract

Many languages are spoken across the world which can bring communication gaps where two people that speak different languages cannot communicate. Usually, this communication gap is solved by using a human interpreter. However, the use of human interpreters is expensive and inconvenient. Many researches are being done to resolve this problem using machine translation techniques. Machine translation is an automatic translation of a source language to a target language. This can be speech to speech or text to text translation. In this work, a bi-directional text based machine translation for English and Afan Oromo languages pair using convolutional neural networks is proposed. We started our study with objective of improving the previous work on English to Afan Oromo machine translation by making the translation bi-directional by applying convolutional neural network on translations between these language pair. In order to achieve our objective, we collected parallel corpus data from different sources and divided into training and testing sets. We have used 80% of total dataset for training and 20% of total dataset for testing. Three systems were implemented where the first system uses a word based statistical approach that used as a baseline, while the second system with recurrent neural network approach is used as a competitive model and lastly, the third system with convolutional neural networks for the bi-directional translation between Afan Oromo and English languages.

After training and testing these systems on corresponding training and testing datasets, the convolutional neural network achieved 3.86 BLEU score improvement on translation from English to Afan Oromo and 3.32 BLEU score on translation from Afan Oromo to English translation than baseline system. Also convolutional neural network approach has shown an improvement of 1.58 BLEU score on translation from English to Afan Oromo and 1.51 BLEU score on translation from Afan Oromo to English translation than recurrent neural network approach. The convolutional neural network approach is faster on training than recurrent neural network approach.

Keywords: *Machine translation, English-Afan Oromo machine translation, recurrent neural network, convolutional neural network bi-directional machine translation*

Chapter One

1. Introduction

1.1. Background

Languages are used as a system of communication between one or more persons either in spoken or written form of communications. In fact, many languages are spoken across the world in which even a single country has more number of different language speakers. As a result, when speakers of two different languages meet each other, the communication gap will occur, unless either one of them knows the other's language or an interpreter helped them to understand each other. In the case of using written form of communication, the written materials in original language give less contribution to speakers those do not understand the language of the writers, unless the written material is translated into language of the target speakers[1].

In recent years, natural language processing has become one research focus area in the field of artificial intelligence and machine learning. This focus was enabled by the availability of datasets and computational capabilities required by artificial intelligence and different new machine learning algorithms. The field of natural language processing consists of the sub-fields such as text summarization, speech recognition, speech to text (or text to speech) conversion, machine translation and similar tasks. Among these sub-fields, machine translation can facilitate social interaction and flow of information among different language speakers[1], [2].

In order to enable information exchange between speakers of different languages, the translation of information from source language to targeted language is required[3]. Therefore, the task of translation between two languages can be performed in two ways. These are experienced human based translation and machine translation. Even though human translation is being used for practical translation, it is very slow as human-expert reads the sentence of source language; tries to understand the concept and thinks over to find its equivalent concept in target language, and finally writes the decided equivalent meaning in target language[3].
Although a lot of work has been done on the atomization of translation task, still in order to properly translate information from one language to another language, the only choice is translating by experienced human-based method, because there is no well-developed automated

translation method. This idea of machine translation was started in 1950s by Georges Artsrouni and Petr Troyanskii, but after the system designed by IBM in 1954s had shown interesting translation result US government released funds to facilitate research works that attracted public interest to work on machine translation. However, the successes of these findings had not gone far as expected to replace human based translation[4], [5].

The main reasons for unsuccessfulness of the atomization work are that the work of atomization needs the considerations of linguistic skills of the languages under study, general knowledge about the world to fully understand the intention of the writer which can be derived from proximity of the cultures and similarity of the languages based structural relatedness. Therefore, developing practical automated machine translation needs the feeding of all these requirements to machine with appropriate algorithm that guides machine on translation to act like that of human beings[5].

However, as getting such appropriate algorithm became difficult and machine translation shown no progress; US government Automatic Language Processing Advisory Committee (ALPAC) announced that it is difficult to implement the automatic machine translator to replace human translator and stopped funding machine translation research studies in 1966s. Nevertheless, other countries like Canada, Germany and France continued their work on machine translation and in mid 1970s great change had been seen on the progress of machine translation because of administrative and commercial demands of multilingual translation which changed the progress of machine translation[5], [6].

Moreover, in 1988s the IBM research team lead by Peter F. Brown suggested statistical based machine translation which attracted many authors because it had shown improvements in translation quality which measured by translation quality measuring metrics than previous rule based technique that uses linguistic rules of languages to implement machine translation[7]. In statistical based approach, machine uses the human translated parallel corpus to learn the structures of the languages under translation work by using the statistical probability of that the certain target sentence in a corpus to be the possible translation of a given source sentence[8].

Even though statistical approach had shown some promise, it is less effective for translation between languages of different structures[9]. As a result, the work done by [9] introduced pure neural network based machine translation which solved problems with using statistical approach. The major problems with statistical approach are less predicting rarely occurred words and accuracy measuring metrics level degradation with increase of sentence length. Therefore, using neural network approach is better to keep longer context that found in the sentence of parallel corpus[9], [10].

Therefore, the availability of several algorithms is showing an attraction to work on the study of machine translation task. The progressive of machine translation study result can be judged by measuring the level of quality improved by comparing against the previous works in which the quality is measured as closeness of system translation output compared against human translation that can be evaluated either by human based or by automatic measuring metrics. Therefore, the study on machine translation is still active research area in the field of artificial intelligence[11].

Generally, in order to improve the performance of machine translation in terms of translation accuracy, the selection of better techniques among existing algorithm is important. According to the study by [11], there is no perfect approach that is free from problems. But if one selects the best technique out of existing machine translation algorithms, then the problem of imperfection will be minimized [11]. Consequently in this study, convolutional neural network based bi-directional English-Afan Oromo text-based machine translation is studied. The reason for choosing convolutional neural network is that it has inherited human-like learning behavior and faster at training time.

Therefore, this development of machine translation between English and Afan Oromo is done with the expectation to improve the previous work from English to Afan Oromo machine translation which was studied by [12] using Statistical Machine Translation(SMT) by adding Afan Oromo to English translation. Therefore, in order to implement our bi-directional neural network based machine translation, we have implemented neural network to translate information in source language to target language.

1.2. Statement of the problem

Now a day, people of the world are joining together to discuss different global issues in order to improve their environments. As an example, different countries in European Union participate on European Union meetings. At the time of their meeting, they discuss different issues like climate change adaptation, urban nature-based solutions, social and economic development integrations and the like. After the meeting is completed; each point of discussion will be translated into official language of each country.

Also, as the need of technology is increasing, the number of social media and mass media are also increasing. These media are re-reporting their news in different languages to get more followers from different language speakers and they are also re-reporting the news of another media repeating in their own media channel's language. In such case, translating the news of one channel to another targeted channel's language is necessary. If they rely only on expert based translation, then the slowness and expensiveness of using expert based translation can challenge them. So they need to change their translation mechanism from expert-based to automated machine based translation.

Moreover, different inspirational books and academic materials are being translated from the writer's own language to other target languages. As an example, there are several books being translated from original language like English to different targeted language like Afan Oromo. However, all these listed tasks are difficult for expert based translation as it takes much time and requires many experts. But still there is no well-developed automated machine translation [11]. So, to simplify this translation task and to improve the translation accuracy level of machine translation, still making strong study is necessary.

More specifically, the study of machine translation between English and Afan Oromo was studied by using statistical machine translation approach, but has not been studied by using neural network based approach like that of machine translation between English and other European languages like German, French or Romanian. However, neural network based approach has better learning ability which captures and apply the features of human learning behavior[2], [12].

The study between English and Afan Oromo was started by [12] who has used statistical approach that uses statistical probability to get the possible target sentence given that a set of source sentences by learning from parallel corpus. He started his study of machine translation from English to Afan Oromo with the objective of developing English to Afan Oromo SMT prototype. After implementing and testing the system, the author got 17.74% translation accuracy in Bi-Lingual Evaluation Understudy (BLEU) score, which is a standard automatic machine translation measuring metrics.

However, when we compare the English to Afan Oromo translation, with the result of machine translation between English and other languages like Amharic or Japanese, it is not enough to be satisfied. According to the study by [13] on the neural network based machine translation between English and Japanese, the score is 53% BLEU score. So, it is necessary to go forward with the study of machine translation between English and Afan Oromo using convolutional neural network in order to improve translation quality. As human translators are working in both directions translation, we need to make bi-directional translation between these languages in our study.

In general, the consideration of the points discussed above will raise the following central questions of the study. These are:

- Which translation method is better in performance between the method using CNN and the method using RNN?
- Can we improve machine translation between English and Afan Oromo by using CNN when compared with statistical approaches used so far?

1.3. Objectives

1.3.1. General objective

The general objective of this study is to design and implement a bi-directional machine translation between English and Afan Oromo based on convolutional neural network.

1.3.2. Specific objectives

This study has the following specific objectives:

- To make the one-directional machine translation from English to Afan Oromo [12] bi-directional machine translation.
- To implement the bi-directional machine translation between Afan Oromo and English using:

  - Statistical approach.
  - Recurrent neural network approach.
  - Convolutional neural network approach.
- To compare the performance of these different machine translation methods using BLEU score metrics.

1.4. Research methodology

In order to successfully accomplish this study, we have followed the following steps. As a first step, we collected datasets from different sources like religious texts such as the Holy Bible, data from governmental sectorial sources like Federal Democratic Republic of Ethiopia (FDRE), Oromia regional state constitution, data from Ethiopian Revenue and Customs Authority (ERCA), and medical sources. This collected data includes both online available textual data and textual data from published written materials.

The main criteria considered to include these materials as a source depended on whether the source is available in both languages or not. The other criterion is that these sources have to be from different subjects like religious material, political, medical and so on. The necessity of this criterion is that it enables the inclusiveness of all the possible words of the languages which allows for effective measurement of the translation quality in BLEU score.

After collecting the necessary data for the study, we designed an experimental system as second step of our work. This design phase includes the designing of both baseline system which is in statistical approach, and the proposed system of the thesis which is based on convolutional neural network. We have also designed a third system by using recurrent neural network approach to compare our proposed system with.

Designing these systems require the consideration of the required components for training and testing. This designing part also includes the specification of implementation environmental domain for our systems. At this stage we also have set out the required parameters of the study such as evaluation metrics needed to measure the quality of translation and techniques of evaluation.

After we had completed the designing phase, we continued with the step of system implementations. At implementation step, we have implemented three systems. The first system is the proposed system which is system under study by using convolutional neural network. The second system is the baseline system which is statistical approach and used to compare the performance level of our system with its performance in term of accuracy of translation. The other system is the competitive system which is in Recurrent Neural Network (RNN) and has similar components like our proposed system with different architecture of neural network.

Therefore, at this implementation stage, we implemented all components required at training stage and at testing. Our baseline system is implemented by using statistical approach which has similar structure like that of the system implemented by [12] and its translation result is compared with our neural network based machine translation results. We have designed two different systems by using neural network to apply on our data in order to see the effect of using different types of neural network architecture.

Additionally, as the neural machine translation approach uses encoder-decoder language modeling which is proposed by [9] with recurrent neural network based architecture at beginning and then recently its architecture is changed into convolutional neural network. Therefore, first we implemented encoder-decoder language modeling by recurrent neural network to apply on our data. Next, we implemented using convolutional neural network in order to compare the results of both architectures to see the improvement resulted with changing of architecture. Finally we made comparison between the results of all our systems.

At the final step of our work, we continued with system training and testing. At this phase, we trained each implemented system by using training dataset and tested the systems using testing dataset. In order to train our system, we have used 80 percent of total datasets as training datasets, while the 20 percent is used as testing datasets. Therefore, the sequence steps for our study are described by Figure 1.1 in which each activity is described by box and the repetitions of steps are decided from result of testing.

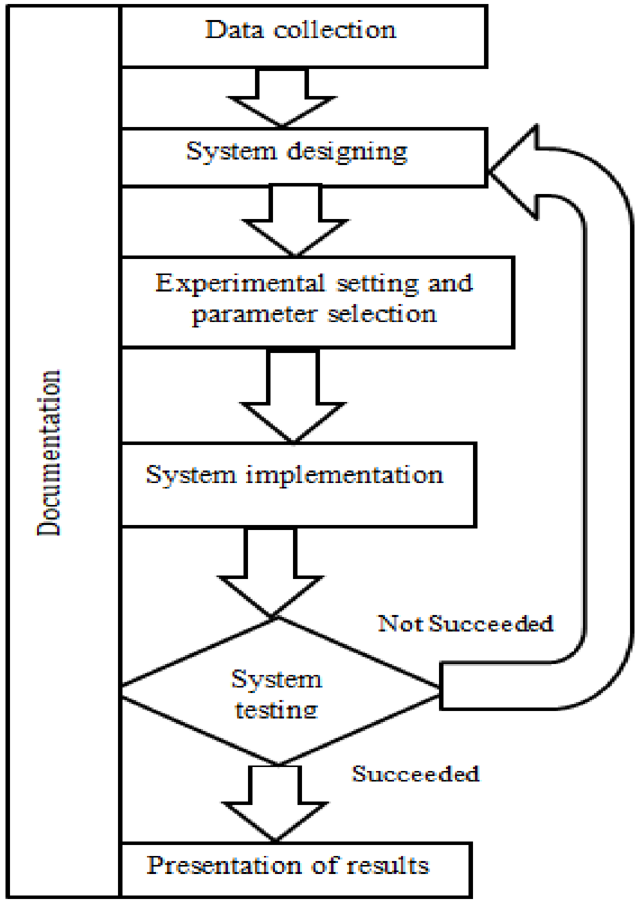


Figure 1.1 diagrammatical representation of research methodology

1.5. Significance of study

The central idea of the translation, either human based translation or machine based translation, is to facilitate the exchange of information between two different language speakers at the level of speech or text. Consequently, the basis of the study on machine translation is to automate this mechanism of language translation. So, the development of machine translation gives great contribution to scarce resource language if the translation is done from resource rich language to scarce resource language. English is resource rich language, while Asian and African languages are scarce resource language [14].

Although Afan Oromo is one of largely spoken East African language, it is in the category of scarce resource language[15]. In order to overcome this scarcity of written resource, one of the possible solutions is the act of facilitating the translation task from resource rich language to this language. Therefore, English is the best choice to translate material from, because it is in the categories of resource rich language. Furthermore for that most of Afan Oromo speakers do not well-understood English language, it is necessary to provide the society with updated information about the world by translating from English sources.

Also as Afan Oromo is the official working language of Oromia regional state, applying machine translation on the translation of different educational books or other materials can contributes to different government institutions like elementary education. Moreover, as a number of Medias are transmitting their programs in both languages, it is necessary to work on machine translation in order to help them on translation data from the source languages to their target language rather than translating their daily news manually from one language to another.

1.6. Scope of the study

This research study focuses on implementation and performance analyzing of bidirectional machine translation between English and Afan Oromo using convolutional neural network based algorithm. The intention beyond the study is to confirm whether convolutional neural network can improve the translation between both languages by comparing against baseline (STM based) system.

Therefore, as convolutional neural network consists of the nature which offers parallel computation, using the concept of parallel computing can improves speed of machine's training to facilitate translation between both languages. If this concept of parallel computing will be added to system in future work, then the amount of training time needed by CNN based approach will be minimized effectively.

1.7. Limitations of the study

This research study is limited to the design and implementation of bi-directional textual translation between Afan Oromo and English language by using convolutional neural network. It does not include speech to speech translation. It is focused on text to text translation from source language to target language. But the translation work may include speech to speech translation for further increasing of application of translation.

1.8. Application of the study

This thesis work may have different applications, but some of them are list below.

- The data set collected for this thesis work can be used for different task like sentence classification, part of speech tagging, spam analyzer or so on.
- This thesis work can be a basis for different studies like speech to text translation, text to speech translation or speech to speech translation for these language pair
- The study accomplished by this work can be used as guiding and teaching material for the topic related to what is covered by the study work.

1.9. Organization of the paper

This research document is organized into the following structures. This Chapter discusses the introduction part of the research in short manner.

The Second Chapter discusses different approaches that used to implement machine translation. It discusses different available algorithm to solve machine translation problems and shows different design options available.

The Third Chapter discusses the literature review of the thesis work. The chapter explains the works that directly or indirectly related to this thesis work.

The Fourth Chapter discusses the system designing part of the study. It describes design of baseline system (STM), other competitive system by using RNN approach and system under study (CNN approach).

The Fifth Chapter discusses the experimental results of the study and gives discussions on the result. The last Chapter discusses conclusion and recommendation of the research work.

Chapter Two

2. Machine Translation Approaches: Background

This chapter discusses the basic ideas of machine translation and the different models used to develop machine translations. It also describes the progress of English-Afan Oromo Machine translation.

2.1. Historical back ground of Afan Oromo and English languages

Afan Oromo, which is Cushitic family of Afro-Asiatic languages, is one of the widely spoken languages in Ethiopia and it is the official working language in Oromia regional state [16]. It is also spoken in some parts of the country and more than half of the population of the country speaks the language[17]. Also Afan Oromo is spoken in other African countries like Somalia, Kenya, Tanzania and Sudan[18]. It is the second widely spoken language in Africa following the Hausca which is one of the languages spoken in Nigeria[15], [17], [19].

Although Afan Oromo has high number of speakers; its usage was limited at the level of oral conversation for longer time in the past. The reason was that writing script had not been derived for the language. Driving the writing script for one language plays a vital role for the development of that language by extending the usage of the language from only oral usage to the full usage level which includes both oral and writing usage levels. In order to derive writing script for Afan Oromo language, different people tried different writing scripts at different times. From such trial, the work of Sheik Bakri Sapalo and Onesmos Nasib(Abbaa Gammachiis or Hiikaa Hawaajii) are non-forgettable[16], [20].

Therefore, Sheik Bakri Sapalo used the Arabean script for Afan Oromo writing and started to write different religious poems and literacy, while Onesmos Nasib used sabean script of Ethiopic script for Afan Oromo writing. By using Sabean script Onesmos Nasib has translated the Holy bible from Hebrew language to Afan Oromo language in the time between 1885 and 1898. Although both Sabean and Arabean scripts were used for practical writing in Afan Oromo language, they did not fit to the structure of the language. The problem with using Arabean script is that there is difference between vowels needed for Afan Oromo writing and that of Arabic language. The sabean script has three deficiencies. The first problem is that it does not differentiate consonant on some words[21].

In Afan Oromo, the tone of sound needed to pronounce some words matters on their meaning, but Sabean script does not handle this problem. For example, '*corraa*' [stream of light came through the hole of something] and '*coraa*' [left over of something] have similar writing in sabean script. The second problem is that it cannot handle glottal stops. In Afan Oromo writing, two consonant will not come next to each other, unless vowel is entered between them. However, there is a case when single i-sound is ignored and two different consonants come next to each other.

For example, in '*harika*' by ignoring 'i' gives '*harka*' [hand]. In other case, two or more than two different vowels, or two or more than similar vowels with different tone never come together, unless joined together by use of apostrophe (') which is used as glottal symbol in Afan Oromo writing. For example, '*mo'aa*' [winner] is formed by two different vowels joined together by glottal symbol and '*ba'aa*' [load] is formed when the same vowel is repeated with different tone.

The last drawback of sabean script is that sabean script has seven vowels while Afan Oromo needs five vowels in two different forms in which totally ten vowels are required for Afan Oromo writing. These two forms of vowels, which longer sound (sagalee dheeraa) and shorter sound(sagalee gabaabaa), are required to differentiate longer sound from shorter sound [19]. Consequently, Qubee script, which is the Latin script with some modification, is selected as official writing symbol of Afan Oromo language in 1991[17].

Therefore, Qubee script consists of twenty four consonants with three additional consonants are included to write the borrowed foreign words and has ten vowels. Therefore table 2.1 shows all symbols of Qubee Afan Oromo script with its nearly similar sound in English language. In the table consonants are given at top, next are double consonants while additional letters to write borrowed words follow and lastly required vowel are given at the bottom.

Table 2.1. Afan Oromo writing script[21]

<table>
<tr><td colspan="11">The basic consonant letters</td></tr>
<tr><td>Capital</td><td colspan="2">B</td><td colspan="2">C</td><td colspan="2">D</td><td colspan="2">F</td><td>G</td><td>H</td></tr>
<tr><td>Small</td><td colspan="2">b</td><td colspan="2">c</td><td colspan="2">d</td><td colspan="2">f</td><td>g</td><td>h</td></tr>
<tr><td>Sound relative to English word</td><td colspan="2">as in bad</td><td colspan="2">tch sound</td><td colspan="2">as in dad</td><td colspan="2">as in far</td><td>as in gap</td><td>as in hall</td></tr>
<tr><td>Capital</td><td colspan="2">J</td><td colspan="2">K</td><td colspan="2">L</td><td colspan="2">M</td><td>N</td><td>Q</td></tr>
<tr><td>Small</td><td colspan="2">j</td><td colspan="2">k</td><td colspan="2">l</td><td colspan="2">m</td><td>n</td><td>q</td></tr>
<tr><td>Sound relative to English word</td><td colspan="2">as in jacket</td><td colspan="2">as in coco</td><td colspan="2">in</td><td colspan="2">as in man</td><td>as in narrow</td><td>Glottal sound k</td></tr>
<tr><td>Capital</td><td colspan="2">R</td><td colspan="2">S</td><td colspan="2">T</td><td colspan="2">W</td><td>X</td><td>Y</td></tr>
<tr><td>Small</td><td colspan="2">r</td><td colspan="2">s</td><td colspan="2">t</td><td colspan="2">w</td><td>x</td><td>y</td></tr>
<tr><td>Sound relative to English word</td><td colspan="2">as in rat</td><td colspan="2">as in sun</td><td colspan="2">as in task</td><td colspan="2">as in what</td><td>Glottal sound t</td><td>As in year</td></tr>
<tr><td colspan="11">The six double consonant letters</td></tr>
<tr><td>Capital</td><td colspan="2">Ch</td><td colspan="2">Dh</td><td colspan="2">Ny</td><td colspan="2">Sh</td><td>Ts</td><td>Ph</td></tr>
<tr><td>Small</td><td colspan="2">ch</td><td colspan="2">dh</td><td colspan="2">ny</td><td colspan="2">sh</td><td>ts</td><td>ph</td></tr>
<tr><td>Sound relative to English word</td><td colspan="2">as in chase</td><td colspan="2">glottal sound d</td><td colspan="2">Like in cognac</td><td colspan="2">As in shall</td><td>Like in</td><td>Pope</td></tr>
<tr><td colspan="11">The three consonant letters to write the borrowed foreign words</td></tr>
<tr><td>Capital</td><td colspan="3">P</td><td colspan="4">V</td><td colspan="3">Z</td></tr>
<tr><td>Small</td><td colspan="3">p</td><td colspan="4">v</td><td colspan="3">z</td></tr>
<tr><td>Sound relative to English word</td><td colspan="3">as in past</td><td colspan="4">as in value</td><td colspan="3">as in zero</td></tr>
<tr><td colspan="11">The vowel letters in longer and shorter sound</td></tr>
<tr><td>Capital</td><td>A</td><td>AA</td><td>E</td><td>EE</td><td>I</td><td>II</td><td>O</td><td>OO</td><td>U</td><td>UU</td></tr>
<tr><td>Small</td><td>a</td><td>aa</td><td>e</td><td>ee</td><td>i</td><td>ii</td><td>o</td><td>oo</td><td>u</td><td>uu</td></tr>
<tr><td>Sound relative to English word</td><td>as in apple</td><td>as in far</td><td>as in egg</td><td>as in eight</td><td>as in hip</td><td>as in teepee</td><td>as in or</td><td>As in hole</td><td>As in bull</td><td>As in book</td></tr>
</table>

Even though writing script problem is solved for Afan Oromo language and it is spoken by more than 50 million people, there is less number of materials written in this language[15]. Despite Afan Oromo is official working language of Oromia regional state, there are not a lot of written materials available in order to support socio-economy and education of the society of the region. Therefore, out of many solutions to improve scarcity of written resources in Afan Oromo language, translating different supportive documents from resource rich languages are expected solutions[19].

Moreover, as need of people for the new technology is currently increased; a lot of media are emerging both with in the country and outside of the country. These media are competing with each other by offering different programs in different languages. So, now a day several media are using Afan Oromo language as the medium of program transmission including British Broadcasting (BBC) and Aljazeera. They need the act of translation from other language to Afan Oromo to further increase the capacity of their information transmission.

In other direction, English is one of the resource rich languages and has a wide variety of written resource materials either as published or available on a websites[6]. Also it is international languages by which the societies of the world are communicating at international context. Moreover, it is the medium of instruction for higher educational institutions with in the country as well as at the international level. Lastly, it is the language that dominates the transmission time of higher media like BBC and Aljazeera. So it is the leading language in the availability of reading resources.

Generally as Afan Oromo is one of scarce resource language and English resource rich language, if resources of English language are translated into Afan Oromo then it can contribute to socio-economic and education of Afan Oromo speakers. Moreover, as many media are using both languages for transmission of their programs, automating the way of available translation task contributes great support for development these media channel.

Lastly, as the societies of the world are becoming closer together on the need of information, a lot of media are delivering similar news like currency exchange rate news, global metrological news and so on. Using only expert based translation is more expensive for such kind daily translation. So, if there is a mechanism by which one translates from one language to other language, it can ease that cost of translation.

2.2. Machine translation

Since the beginning of research study in the field of natural language processing, the work on machine translation has become the main concern to replace the need for expert to translate between different languages. When the expert of languages start the act of translation, the expert should understand the concept in source language and then try to replace that concept in target language. Similarly the machine translation should work to perform the task of changing the concept available in source language, either in textual form or speech, to the concept in another targeted language[5].

In order to accomplish this translation task by machine, several preconditions should be satisfied. The first precondition is that machine must be trained on the dataset that has been previously translated from source language into target language by the person who understands both languages. Therefore, such dataset must be prepared to be fed into the machine after collected from variety of acceptable sources. The reason of collecting data from variety of sources is that it helps to include all possible word of the language pair. After the datasets are collected from such acceptable sources, it must be aligned with its corresponding translation in and must be stored as parallel corpus[22], [23].

After preparing parallel corpus, the next step is the step of pre-processing data which includes the process of identifying the maximum length of the sentence with in the data set, clearing the sentence that is longer than the selected maximum length of sentence to be fed into the machine and clearing unnecessary character out of the data. After pre-processing parallel corpus, the next step is the step of creating a vocabulary from a data set and representing each word of the datasets with a unique numerical form representation. This step of converting datasets into numerical form representation is known as tokenization[23]. Therefore, the pre-processing step includes the act of cleaning data, the technique of preparing vocabulary and the activity of tokenizing a sequence of sentences in a data set[9].

Therefore, after data has been pre-processed, the step of selecting language modeling for translation is followed. At this stage, methodology of translation with proper language modeling technique that clearly describes the structure of both languages must be selected. The act of language modeling can be related to the act of expert to translate language which includes the task of understanding the concept in both languages and trying to convert idea in one to another. So, like that of expert, machine must learn the structures of both languages in order to translate the idea in a given language into other language[23], [24].

Therefore, for machine to perform such activity, the machine must pass through out a set of required processes. These processes include, the process of learning the structure of both source and target languages by taking enough training in order to understand the structures of both languages and the process of translating from one language to another based on experience acquired during training. Although these sequences of activity look easy at theoretical level, it needs a well-developed form of strategic procedures. For this reason different researchers were following different approaches to deal with machine translation. These approaches are categorized into five different categories[11], [25]. These are:

i. Rule based Machine Translation
ii. Example based Machine Translation
iii. Statistical Machine Translation
iv. Hybrid Machine Translation
v. Neural network based machine translation(with different architectures)

Therefore, Figure 2.1 describes the available approaches of machine translation in which source language is given at left and target language is given at right side, while different possible translation passes are shown by arrow and the possible inter-mediatory language form are shown by blocks which are used in the Figure [Amine, 2012].

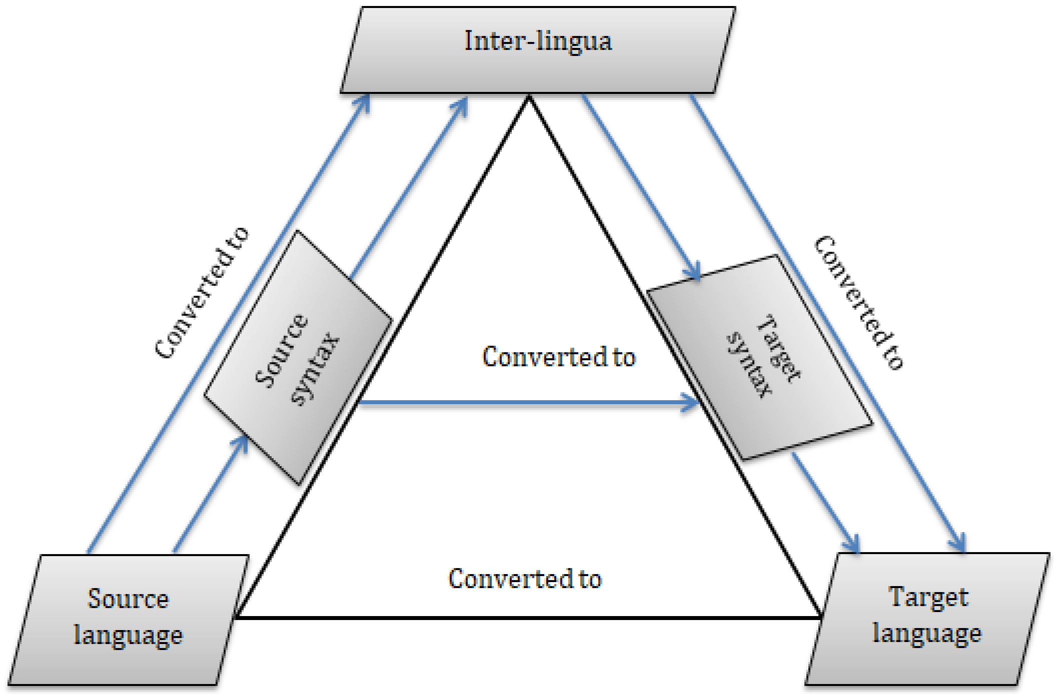


Figure 2.1. Vazquez triangles to describe design options[26]

2.2.1. Rule-Based Machine translation

The rule-based machine translation (RBMT) is the machine translation approach which mainly concerns with the linguistics of both the target and source languages. Using this technique requires a lot of human effort on the preparation of rules and linguistic resources of both languages. The rules required are the syntactic rules of both target and source languages, and the transferring rules that rearrange the meaning of words in source language into the manner of target language. The linguistic resources required for using RBMT include morphological analyzers, part-of-speech (POS) taggers, syntactic parsers and morphological generators Therefore, using RBMT requires the implementation of these separated components and integrating the components by providing all necessary rules of the languages and datasets[25], [27].

Therefore, from these components, the first component is morphological analyzer. The function of this component is to identify each input word used to create the sentence in the given language. The morphological analyzer works on the identification of the structure of sentences used in both source and target languages by breaking down the sentences into smallest units. This process helps to describe the grammatical structures of one language by identifying base words, prefixes and suffixes of the sentences used in the language. This aims on the identifications of basic unit of sentence either at level of word differentiation from prefix and suffix or at the level of phrase by selecting different number of words[28].

The second component is the part of speech (POS) tagger that helps to put description on each word of the sentence. This description tells that a given word has been used as either of subject, verb, object or other. Hence, each word in the sentence of each language has to be described by such POS tagger. The third component is syntactic parser which accepts the tagged sentence to re-arrange the words to enable them to conform into syntactic rule of the language. The words arrangement in source language has to be in the manner of target language structure. After syntactic parser is applied, each word of the sentence will be independently translated by using matching rule from the defined rules of the language pairs[28], [29].

Finally the last component, which morphological generator, generates the morphological arrangement of words for target language. The generated target language sentence highly depends on the selected RBMT language modeling. Indeed before deciding to choose RBMT to implement for machine translation, the decision of whether getting all the necessary rules and required linguistic resources should be decided early at beginning. Also as it needs large amount of dictionaries of both languages, the availability of these dictionaries must be confirmed before start[29].

Therefore all the components required to develop RBMT are described in Figure 2.2 where translation takes place from source sentence to target sentence. Hence, all components are independent and connected in sequential style until finally translated target sentence is generated as described in the figure. This RBMT can be modeled in three different ways. These are:

i. Dictionary-based Machine Translation
ii. Transfer-based Machine Translation
iii. Interlingua based machine translation

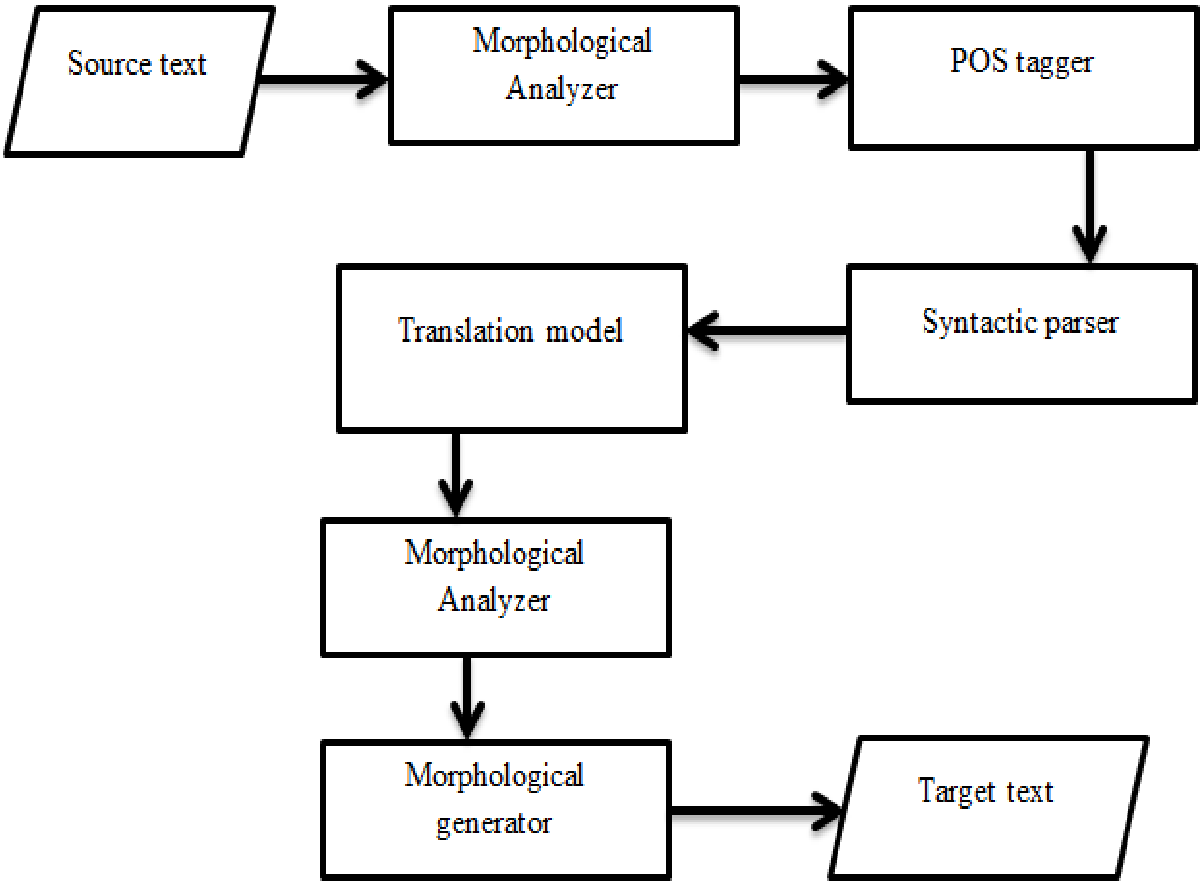


Figure 2.2. Design of Rule Based Machine Translation (RBMT)[28]

2.2.1.1. Dictionary-based machine translation

The dictionary-based machine translation technique is word-by-word translation which uses dictionary for machine translation. Unless applied to the language of the same structure, the result of its translation does not give a sense. Preferably, it is applied to implement dictionary software[29].

2.2.1.2. Interlingua-based machine translation

The machine translation with this language modeling approach works with the idea of changing the concept in the source language into inter-mediatory neutral language, known as Interlingua, in order to later target sentence will be derived by selecting other translation technique like dictionary based modeling. This Interlingua is neutral representation and related to the structure of both languages. This Interlingua based translation needs to drive the inter-mediatory representation between languages of having similar structure and difficult for the languages of having different structures[28], [29].

2.2.1.3. Transfer-based machine translation

Like interlingua-based approach, this transfer-based technique also uses inter-mediatory language for translation. But unlike that of Interlingua-based translation the Interlingua of this technique is partially dependent to language pairs[28].

2.2.2. Example-Based Machine translation

Another interesting machine translation approach is Example-based approach which uses the idea of analogy between prepared bilingual corpuses and sentences to be translated for the purpose of translation. The parallel corpus prepared should keep correspondence between each element of data for both source and target languages. Therefore, sentence can be translated by searching matching translation in previously translated examples. Thus, this machine translation approach was used in the early machine translation by Nagao (1984), Brown et al. (1990) and Sato and Nagao (1990) at starting of parallel corpus-based translation[8].

2.2.3. Statistical Machine translation

In the case of statistical machine translation, translation can be done by doing statistical probability on parallel corpora. According to the idea of this technique, translation is processed by finding the probability of matching of target sentence with in a given source sentence. More formally, given a source language sentence S, statistical machine translation (SMT) searches throughout all the target language sentences T and finds the one with the highest probability as P (T|S)[8].

Since the beginning of statistical based language modeling, a lot of works are done to improve the progressive of machine translation accuracy with use of the application of statistical probability that applied on machine translation by applying in different approaches. Therefore, statistical machine translation can be categorized into three different categories depending on the basic unit of modeling. These are word-based language modeling, phrase based language modeling and syntactical-based language modeling[8], [30].

In the case of word-based SMT modeling, the probabilities of the co-occurrences of words are calculated as language modeling, while the probabilities of how each individual word can be translated from source language to target language are calculated as translation modeling. In order to model STM using syntactical modeling, the syntax of the language is modeled either by word based-modeling or phrase-based modeling. At the time of translation, the highest probability of the matching word is selected from the prepared translation model and language model is used to arrange words into right contextual matching[8], [29].

For phrase based modeling, the phrase or sequence of words will be considered to find the probabilities of co-occurrence of phrases for language modeling and how individual phrases are translated from source to target sentences are used as translation modeling. The modern statistical machine translations are focused on phrase based modeling and phrases are estimated from segmentation of the aligned bilingual corpora by using relative frequencies[8], [25].

Therefore, SMT consists of some basic components which have some similarity in each category of modeling except the change of the basic unit mentioned above. These components are language modeling, translation modeling, tuning (optional) and decoding. The language modeling assigns the probability of co-occurrence to the words in the target sentence using n-gram calculation. For phrase-based translation n-gram calculation depends on more than one word like 2-grams calculation, 3-gram calculation, and so on. After assigning probability to each word of the sentence, the word with its probability is stored as language model[29], [31].

$P(T) = P(w_n|w_1, w_2, w_3 \ldots w_{n-1})$

$P(T) = P(w_1)P(w2|w1)P(w3|w2,w1)P(w4|w3, w2, w1)$ [Equation 2.1]

Therefore, translation modeling assigns probability to each word of the target sentence given all the words of source sentence to describe the translation relationship between words of both languages. After assign the probability, the translation table is prepared for the machine for later use the information for translation of new sentences. Hence this translation modeling probability can be calculated by using equation 2.2.Then, as translation modeling is completed, it must be given as input into decoder part either through tuning or directly connected into decoder[29].

$P(t|s) = P(t1|s1, s2, \ldots)$ [Equation 2.2]

The tuning component is used as parameter optimization mechanism for SMT and can be optional component. But if GIZA++ is used for sentence alignment and the size of training data small, it is better to choose tuning mechanism than non-tuning SMT technique. For large size data, the cost of tuning is expensive as a large parameter need to be optimized. The other component of SMT is decoder which used as the searching mechanism the translation model. This component receives language modeling and translation model to search the matching translation of input sentence[29], [32].

$T_{best} = argmax_e P(s|t)P(e)$ [Equation 2.3]

Therefore, Figure 2.3 discusses the components required to develop STM in which the left side shows target language modeling in which only target sentences of parallel corpus used while right hand side shows the language modeling part which needs both source and target sentence, and finally decoder combines both model and gives translation which can be measured by BLEU score metrics.

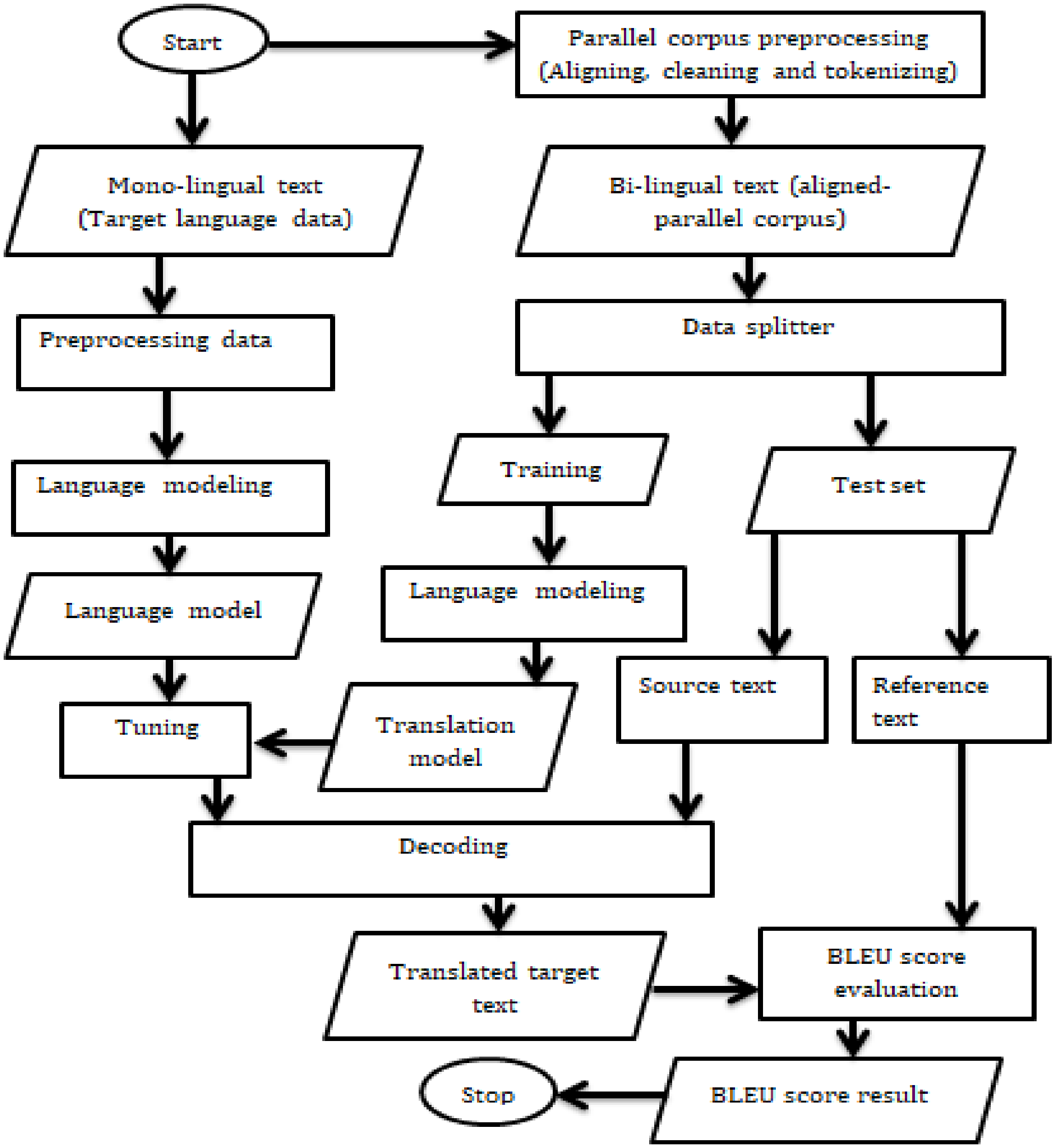


Figure 2.3. Design of Statistical Machine Translation[12], [29]

2.2.4. Hybrid Machine translation

The Hybrid machine translation (HMT) approach combines the strength of both statistical machine translation approach and that of Rule-based approach to get better result of translation accuracy. In order to implement machine translations by using HMT, we have to select from two well-known approaches[32]. The first approach is rule post-processed by statistics in which translations are performed by using rules of both languages like syntax rule of categorizing word into either subject or verb or object, and statistics are then used to adjust the candidate words in correct order of target language sentence[32].

The second HMT approach is statistics guided by rules in which rules are used to preprocess data by grouping words into its class of either subject or verb or object groups, while statistics is used to generate the correct translation of words based on statistical computation of matching word and latter rules are applied on candidate words to drive translated target language sentence. Therefore, the selection between these two approach of HMT translation depend on whether statistics is preferred for word translation or rules are preferred for translation each word of language under study[32].

2.2.5. Neural Machine translation

The last interesting approach of machine translation is neural network approach which is a recently developed machine translation technique by use of inter-connected neural network to perform translation activity. At beginning of neural network based machine translation, neural network was introduced as supportive tool for SMT to facilitate the computation of statistical probability that assigned to each word in a sequence [Schuster et al., 1997][Cho et al., 2014]. After further work of different authors, the pure neural network based machine translation was modeled with the idea of jointly training and translating of data from one language to other language[9], [10].

Although SMT was dominated the technology of machine translation for several decades, it cannot handle longer context with the sentence. This problem make difficult to use SMT for longer sentences which became the reason for emerge of the pure neural network based machine translation[9]. Neural network based language modeling involves the building of an end-to-end neural network based language modeling that neural network is trained to map aligned bilingual texts from source sentence to target sentences without additional external linguistic information[9], [10].

In recent time, neural machine translation becomes the application of deep neural networks (DNNs) to build end-to-end encoder-decoder models which a translation system consists of subcomponents that are separately optimized[33]. Therefore, figure 2.4 shows this encoder-decoder language modeling in which source sentence pre-processed to be fed into encoder, then the encoder generates contextual relationship and feed data enhancement mechanism which separately shows decoder the relevant context for decoder to generate target sentence that match the given context[9].

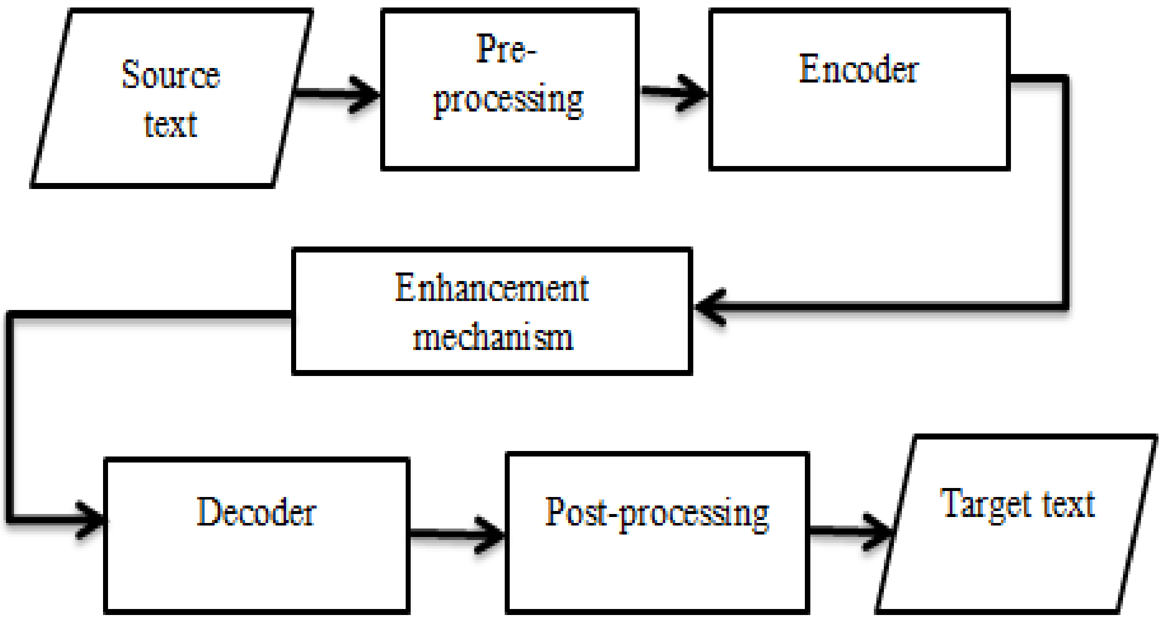


Figure 2.4. Design of Encoder-Decoder language modeling[9]

In fact the neural machine translation varies in terms of their exact architecture of neural network used for machine translation. In order to work on sequential data like text to text translation, the familiar architecture is recurrent neural network that depend on current input text (word) of source language and the previously generated output to decide current output. Therefore, it takes longer time to generate output for longer sentence as it looks for internal states for several times, so that currently convolutional architecture is being in consideration for machine translation [33].

There are different types of recurrent neural network applied for machine translation. The first one is gated recurrent neural network (GRU) that has gated unit to control information flow in the neural network[10]. The other type of recurrent neural network is bidirectional recurrent neural network which the concatenated form of hidden forward state to read data in forward manner hidden backward state to read data reverse manner[2]. Hence, in neural machine translation, neural network is used to implement the encoder and decoder part of encoder-decoder end to end language modeling[9].

Therefore, the original sequential data must be changed into the form suitable to read for neural network. This format is word embedding format which changes sequential data in to vector form representation. The encoder part receives this embedded form of data and generates contextual relation between words. The decoder part receives the context generated by encoder and searches for matching words with in target language. When the sequence includes longer sentence, enhancement mechanism is needed to help storing of longer context with in the sentence[2], [10].

2.2.5.1. Word embedding

The parallel corpus of sequential data cannot be directly fed into the neural network for processing. It needs some form of representation in a suitable format. To change the data into this suitable format, each word of the sentence in the data must be identified and represented by unique id number. After assigning unique number (identity number or id) to each word, each sentence has to be represented by these sequences of words' id numbers. This form of representation is known as one-hot vector representation[34].

From these sequences of one-hot vector representation, word embedding will be formed. The word embedding changes the dimension of one-hot vector into the matching size of designed input layer neural network. In order to represent each vocabularies of parallel corpus data set of machine translation, word embedding represents each word with specified dimension vector in a given vector space. Therefore, this vector representation is read by the first layer of neural network in neural machine translation[34], [35].

2.2.5.2. Encoder

In designing of Encoder – decoder language modeling, Encoder is the connection of neural network which is developed to convert the word vector (sequences of vectors) into sentence matrix. Therefore, each row of sentence matrix encodes the meaning of each word in the context of the sentence. The encoder can use either of deep neural network architecture. Some of these architectures are long-short-term memory (LSTM), gate recurrent unit (GRU), gated linear unit (GLU) or bi-directional recurrent neural network[13].

Therefore, most familiar architectures of encoder part of neural machine translation on text based translation are GRU and LSTM. The difference between these two architectures is that GRU has no output gate, while LSTM has both output gate and backward gate for storing the internal state of the network. Therefore GRU is LSTM that has no output gate, so writes the contents from its memory cell to the larger network at each time step[2], [23].

In order to work on textual data by encoder part of encoder-decoder language modeling, the data must be given in embedded form of sequential data. If $X=\{x_1, x_2, x_3 \ldots x_n\}$ is the input sequence to the encoder with n sequence length, the encoder will produce $Y= \{y_1, y_2, y_3 \ldots y_m\}$ context in the sequence of input data. In order to drive this context, the selection of network architecture is necessary. In this study, we have seen two different architecture of neural network for machine translation between these language pair. Hence, the recurrent neural network uses two parameters to drive the sentence context. These are the collection of currently inputted data and the previous output of the network[13].

Therefore, when using recurrent neural network the previous output is considered as internal state of the neural network and if the current input is the starting word, then the input is marked by the sign of the start of token symbol indicator and given as input. From these two parameters the context c_i derived as $c_i=f(x(t), h(t-1))$ where $x(t)$ represents current input data and $h(t-1)$ represents previous output of the system. This context vector will be fed into decoder part either by using enhancement mechanism which is attention-mechanism or without using attention mechanism. If attention mechanism is left, the last layer of encoder network will be connected to decoder part of the network[13], [23].

2.2.5.3. Decoder

Similar to the encoder part of the NMT, the decoder part is also built from the interconnection of neural network. The decoder part receives context derived from the data either through attention mechanism or directly from the last layer of encoder part of the network. The decoder part searches for the word that matches to the derived context of the sequence of word representation. In order to search the matching word, beam searching mechanism with support softmax function works for translation activity[36], [37].

Therefore, at one time of decoding only one targeted word is searched and the searching activity has to be repeated up to the end of source sentence which has been marked by end of token symbol. For translation from source sentence to target sentence, decoder receive the sequence of context vector as z = $\{z_1, z_2, z_3 \ldots z_n\}$ and computes the probability of matching word as $y(t)=softmax(k(t)/z(t), k(t-1), c(t))$ *where* k(t) is current decoder state, c(t) is current context of sequence of data, z(t) is target sequence of words and k(t-1) is previous internal state of the decoder.

2.2.5.4. Attention

To model neural machine translation by using encoder-decoder language modeling, two critical issues must be considered. The first one is how to recover performance degrade as the number of sequence length increases. The context with in a sentence is derived as the inter-dependency of nearby words in a given sequences words in the sentence. But as the length of the sentence increase, the inter-dependency of words at the beginning of sentence and at end of the sentence is loosely related. This problem also high in the statistical machine translation approach, as statistical machine translation is weak on the translation of longer sentence than the short sentence[2], [38].

The second problem of encoder-decoder model is how to handle larger number of vocabulary size available within the data. As each word in the sentence is visited, it must be assigned new identity number in order to identify word by unique id number at time it encountered in the data. But when the length of dictionary increases, the number used for word representation becomes higher and the dimension of word vector needed becomes higher. So, adding attention mechanism to encoder-decoder language modeling solves this problem by minimizing changing of higher dimension vector into lower dimension vector[10], [39].

Therefore, the attention mechanism minimizes the size of required sentence matrixes paying attention only to part needed to keep during vector dimension reduction. There are different types of attention mechanisms. Some of these are matrix, matrix plus context, matrix plus vector or matrix plus matrix attention. However, if no attention mechanism is used, either sum or average or max-pooling can be used for the network[10].

2.3. Evaluation metrics

In order to evaluate the performance of machine translation in terms of translation accuracy, the machine translated sentence must be evaluated either by human (expert) or some automated measuring metrics. Human based evaluation is more expensive, because in order to evaluate translation accuracy we need to find for the expert of both languages. This expert compares the system output against given reference sentence and describes the result of comparison in percentage computation. Therefore, doing this processes for number of sentences take much time. Therefore, in order to solve this slowness of expert based evaluation, the best choice is to use automatic measuring metrics. One of such automatic metrics is bi-lingual evaluation understudy (BLEU) that is used to measure translation accuracy by comparing the system's translation output against human translated reference sentences[40].

Moreover, in order to generalize the evaluation result of machine translation, the selection of correct testing technique is necessary. Therefore, the total dataset must be split into training set and testing set. This splitting of data into these two sets helps to train system on training dataset and to test the system on testing dataset in order to get convincing report evaluation result of translation accuracy. Therefore, the best technique of testing is Pareto principle (80/20) in which 80 percent of total data is used for training set, while 20 percent of total dataset is left for testing the system[41].

2.3.1. Bi-lingual Evaluation Understudy (BLEU) score

Even though human-based evaluation is good to measure translation quality of machine translation by considering different conditions like grammatical correctness or real intention of speakers, it requires much time as expert of the languages compares each system's output sentence against reference sentence and computes percentage accuracy[40]. The proper way to evaluate the translation accuracy is measuring the BLEU scores of the translation. Therefore, BLEU score is automatic evaluation technique that calculates n-grams precision by working on total words of translated sentence compared against words of reference sentence.

Therefore, BLEU score does not take any consideration for the grammatical correctness of the textual data. Sometimes, the output of system may be the repeated copy of a single word. In such case the unnecessary repeated word may generate incorrect matching probability. In order to minimize such problem, it is necessary to use modified n-gram method which limits the number repeated words if a single word is occurred more than the number of its occurrence in reference sentence[40]. Therefore, standard form of BLEU score metrics is the form of modified n-gram type and can be evaluated as:

$P(i) = \frac{Match(i)}{H(i)}$ where *H (i)* is the number of i-gram tuples, i=1, 2… n and

$Match(i) = \sum_{i=1}^{n} min\left(Ch(t_i)\middle|, max\left(Ch_j(t_i)\right)\right)$ where t_i is number lists, $Ch(t_i)$ is number t_i occurs in hypothesis and $Ch_j(t_i)$ is number of t_i occurs in reference sentence.

Therefore, combining all the values computed above, the BLEU score can be the value of:

BLEU = $\{\{\prod_{i=1}^{n} \boldsymbol{p(i)}\}^{\frac{1}{n}}\}$ [Equation 2.4]

where n represents the maximum n-gram (usually set to 4-gram) [40].

Chapter Three

3. Literature Review

This chapter gives the review of literature on machine translation between English and Afan Oromo. Also, it discusses some literature reviews on machine translation between English and other languages those are used neural network based machine translation.

3.1. Machine translation between English and Afan Oromo

Although the studies on machine translation between English and Afan Oromo have been started, the results of the studies have not gone far like that of machine translations between English and other languages like European languages[2], [12], [42] .

The research conducted by [12] is the starting work on machine translation between English and Afan Oromo language pair. The study was based on statistical machine translation approach[12]. This study work was accomplished based on two main objectives. Therefore, the first objective is to test how far to go with available limited corpus by testing applicability of existing SMT on translation from English to Afan Oromo. The second objective is to analyze the output of the system with the objective of analyzing the challenge to be tackled. The author had collected parallel corpus of 20,000 pairs of bi-lingual data and 62,300 Afan Oromo mono-lingual data, and 10% of bi-lingual data for testing the system. After implementing and testing the system on the collected parallel corpus, the result of evaluating the performance of his system on English to Afan Oromo translation has shown an average BLEU score of 17.74%.

The other study between these language pairs is the one conducted by [43] which deals with Bi-directional English-Afaan Oromoo machine translation using hybrid approach. The author collected 3000 pairs of parallel corpus and used 2,900 for training while he used 100 pairs of sentences for testing purpose. Therefore, after collecting data, the author conducted two experimental studies. First, the author implemented the system based on statistical approach. Hence, by testing this system, the author got 32.9% BLEU score for English to Afan Oromo and 41.5% BLEU score for Afan Oromo to English translation. Second, the author implemented the system based on hybrid approach. After testing this system, the author got 37.41% BLEU score result for English to Afan Oromo translation and 52.02% BLEU score for Afan Oromo to English translation.

The other study is the study by [44] which is also based on statistical approach. This study work was focused on identification of the effectiveness of different corpus alignment techniques like word-level, phrase-level and sentence level corpus alignments on English-Afaan Oromo statistical machine translation. Therefore, by collecting 6,400 pairs of bi-lingual data, 19,300 English mono-lingual data and 12,200 Afan Oromo mono-lingual data; they did three different experiments. These are experiment of word-level alignment, experiment of phrase-level alignment at different length of phrase and experiment of sentence-level alignment. Hence, after doing their experiments, the authors reported that they have got best result of 27% BLEU score on phrase-level alignment with the phrase-length between 5 up to 16 and they got minimum result of 18% BLEU score on sentence-level alignment of parallel corpus.

3.2. Machine translation between English and other language

The machine translations between English and other languages have gone far. In order to look machine translation between English and other languages, we have to classify existing machine translation between the language pairs into two big categories. These categories are machine translation between English and Europeans languages, and machine translations between English and Asians languages.

The automated machine translations between English and European languages have been come nearly closer to practical level[2]. The reason for its development is highly dependent on the excessively availability of the data set resources which is derived from European parliamentary data. This European parliamentary data are the result of discussion of the meeting between the members of European Union which will be translated into the official languages of each member country after each time of the meeting. Everyone who needs this data for study, can access the data through internet from their websites[2], [24], [42].

The second category is machine translations between English and Asian countries. In recent time, interesting progresses have been seen in the areas of machine translation between English and Families of Asian languages. As an example, the machine translation between English and Japanese has been studied by using Example-based machine translation approach, Rule-based machine translation approach, neural network based approach(feed forward architecture) and recurrent neural network based approach[13], [24]. Also, many works are done on machine translation between English and Chinese by using different approaches[45].

Especially, Google has recently bridged old phrase-based statistical machine translation model into neural network based model of machine translation. After changing their system into neural network based model, they have solved the problems with conventional phrase-based statistical based language modeling such as loss of words that occur less frequently and unknown words. In their paper the Google research team reported that they have got competitive result as their earlier state-of-the art system by solving the problems with conventional phrase-based statistical system by 60%[2], [45].

Therefore, the neural network based machine translation has shown an interesting progressive because of that it can be modeled in an end-to-end modeling (synonym of encoder-decoder modeling) which all the components that are needed to implement the system are connected in cascaded form. The work by [9] has brought an interesting sequence to sequence language modeling, which is also known as encoder decoder modeling, by using recurrent neural network architecture[9].

Later on this encoder-decoder language modeling got good attention in neural network based machine translation and the issue becomes how to minimize the long training time of recurrent network. The research conducted by Lamb shows the steps taken to change the architecture of encoder-decoder modeling into the use of convolutional neural network[33]. The work by Lamb uses encoder-decoder language modeling which uses encoder to generate context of source language sequence and uses decoder to generate target language sequence that matches to the context[9], [33].

More recently, the convolutional neural network architecture has been used to build encoder part which uses convolutional neural network to accept input sentence and performs multiple convolution to produce fixed-length encoding of the sentence as output. This output of encoder will be given as input into to a recurrent neural network based decoder layer which also uses either gated recurrent unit or Long Short Term Memory type of RNN. Encoder and decoder are jointly trained to maximize the conditional probability of the target sentence given the source sentence from word embedding form representation of data[33], [39], [46].

The other study is the one which was conducted by [42]. This study work deals with convolutional sequence to sequence learning. This research focused on changing the architecture of encoder-decoder model to entirely convolutional neural network architecture. The authors designed both encoder and decoder on the basis of convolutional neural network based architecture, while intermediate states are computed based on fixed number of input elements. Different blocks of encoder and decoder are used and each block contains a one dimensional convolutional layer followed by a non-linearity layer.

In order to compute the contextual relevance with in the sentences, the separate attention mechanism is used for encoder network. After implementing convolutional neural network based machine translation, its operational performance has been tested against the accuracy of machine translation of recurrent neural network based architecture. The convolutional neural network based architecture has shown better performance for machine translations between English-Romanian by 1.9 BLEU, between English-French by 1.6 BLEU and between English-German by 0.5 BLEU on WMT'14 data than recurrent neural network based architecture[42].

3.3. Summary

In general, the study of machine translation from English and Afan Oromo has been conducted before by using statistical and hybrid approaches. So, the neural network machine translation is new technique on the translation between these two language pair. The studies on other language pairs have used neural machine translation techniques and they have benefited from advantage of using encoder-decoder language modeling to simplify the complicated tasks like separate language modeling, translation modeling, tuning and decoding in STM by making jointly learning and translating based on neural network approach of machine translation. In addition to simplifying the complexity of the previous methods, they have got the benefit of improving translation either in BLEU scores or by recovering of lost words (rarely occurred words) in the case of using neural network based machine translation.

Therefore, several authors have completely outperformed their previous conventional approach which is statistical-based machine translation approach by using RNN based architecture of neural machine translation. After outperforming SMT by RNN architecture, now they are on the process of finding a solution to the inherent problem which came with approach of using RNN based architecture of neural machine translation. This inherent problem is related to the time inefficiency of recurrent neural network at a training time. In order to solve this problem , they are working on way of changing the architecture encoder-decoder language modeling from RNN based architecture to CNN based architecture which requires less training time compared to that RNN based architecture training time.

Therefore, this study is focused on the design and implementation of English-Afan Oromo bi-directional machine translation using CNN based architecture of neural machine translation using statistical machine translation as baseline system. The designing of this system starts from the designing of RNN based architecture of encoder-decoder language modeling to compare whether CNN based architecture shows better BLEU score improvement on these language pair and then to compare the level of training time required by both architectures on the resulted BLEU score.

Chapter Four

4. Designing Deep Learning Based English-Afan Oromo Machine Translation

This chapter discusses the designing and modeling of English-Afan Oromo machine translation based on encoder-decoder language modeling by using convolutional neural network based architecture. Then, it describes the components of each system in details.

4.1. System designing

In recent time, the encoder-decoder end-to-end language modeling is becoming attractive language modeling for machine translation task. It is a language modeling which is based on deep learning algorithm. According to the work of different researchers, in the early emergence of neural machine translation the recurrent neural network was used to generate word co-occurrence probabilities for statistical machine translation. Through time, different researchers continued to work on its gradual development to develop the pure recurrent neural network based machine translation[10].

After a few years, different authors outperformed statistical machine translation by using recurrent neural network. Now a day, rather than using tedious steps like preparation of language modeling, preparation of translation modeling, tuning and decoding steps of statistical machine translation, using encoder-decoder based modeling became best approach. Therefore, encoder-decoder based machine translation became the better choice for the simplicity of the steps of modeling. In encoder-decoder modeling, the steps are interconnected. Therefore, human-intervention at training time is not necessary like that of statistical approach[2], [10].

Hence, the problem with statistical machine translation is not only the complexity of the approach, but also using statistical probability by itself depends on similarity of language structure[11]. However, the structures of these languages are different as Afan Oromo is Subject-Object-Verb (SOV) form, while English is Subject-Verb-Object (SVO) form. Therefore, using recurrent neural network based machine translation can solve these problems, because neural network has ability to learn syntactically related words of sentences[10]. Therefore, using neural network based approach solves the problem of inconvenience of statistical probability for different structure language pair, but other problem was co-emerged with recurrent neural network based machine translation. This problem is the slowness of recurrent neural network on the training which takes longer training time for even small sized data[2].

Therefore, in this research work, we focused on designing bi-directional machine translation using convolutional neural network based machine translation, because of that neural network based has human-like learning and convolutional neural network is faster during training. Therefore, in order to show that convolutional neural network approach is better than other approach, we compare CNN based translation performance with that of RNN based approach. Moreover, we have used STM based system, which is similar with the one applied on the language pair before by [12], as a baseline system for machine translation between English and Afan Oromo language pair. We are also interested to compare the training time of RNN based model with the training time of CNN based model to show that CNN approach gives time-efficiency during training than RNN based approach.

Therefore to accomplish the study objectives, first we designed recurrent neural network based system to apply on the translation between the language pair, because different authors are using encoder-decoder language modeling by the use of RNN architecture. Then, we designed CNN based machine translation for the language pair which has similar structure like that of recurrent neural network based architecture.

Therefore at beginning of designing, RNN based system with encoder-decoder machine translation is designed with all supporting components. After RNN based system design is completed, similar encoder-decoder language model but with CNN architecture is designed. This CNN based architecture is the design of our proposed system which is designed by expecting that it will improves the translation between English and Afan Oromo in both direction of translation. Since we need to compare performance in terms of both translation quality and time efficiency of both architectures on the translation of the language pair, we need to use similar components in designing of both systems except replacement of RNN architecture with CNN architecture.

Lastly, as our collected datasets are in text format, we have designed supporting components that can change data from text format to vector form representation. The purpose of this conversion is that neural networks require vector form representation to work on data. So, the original text format must be converted into a one-hot form representation style as seen in Table 5.1 by use of unique word id and zero-padding. This representation style uses a fixed length of sentence. Therefore, to set the length of sentences to the fixed length, zero is added to short sentences. So, adding this at the end of each short sentence is known as zero-padding. This conversion process has been done in reverse direction after getting output from the system to form the translated output sentence.

Therefore, the designing part considers the proper integration of supporting components with the core machine translation system. Also to conclude that we improved machine translation between the language pairs, the performance comparison against the baseline system in terms of BLEU score metrics must be done by designing the scorer component. Therefore, designing includes the BLEU score measuring component at the end of the designed machine translation system.

From Table 4.1 we can see how to convert original data into one-hot form representation. In Table 4.1a each word of the dataset is visited to give word id for each word of the sentences. Therefore, for this small sample data, we have 22 different words so that our vocabulary size is 22. In Table 4.1b we have marked the start and end of each sentence by start of the sentence (SOS) and end of sentence (EOS) indicators respectively. After marking the sentence, first, we have converted the data into integer form representation which has different lengths and finally, we have converted this integer form representation into a one-hot style by using zero-padding to set the length of sentence to fixed length.

Table 4.1 Process of data preparation to feed into the neural network based systems a) word counting and vocabulary preparation b) Text format to vector form data conversion

a)

Sample data before pre-processing	Word counting and Vocabulary formation									
they knew that they were naked.	Padding	0	that	5	are	10	do	15	is	20
blessed are the meek.	SOS	1	were	6	the	11	you	16	clean	21
for what do you ask?	EOS	2	naked	7	meek	12	ask	17	will	22
the room is clean.	they	3	.	8	for	13	?	18	die	23
they will die in the room.	knew	4	blessed	9	what	14	room	19	in	24

b)

The pre-processed of the sample dataset	Sentence Tokenization	One-hot form using zero-padding
SOS they knew that they were naked . EOS	[3 4 5 3 6 7 8]	[1 3 4 5 3 6 7 8 2]
SOS blessed are the meek . EOS	[9 10 11 12 8]	[1 9 10 11 12 8 2 0 0]
SOS for what do you ask ? EOS	[13 14 15 16 17 18]	[1 13 14 15 16 17 18 2 0]
SOS the room is clean . EOS	[11 19 20 21 8]	[1 11 19 20 21 8 2 0 0]
SOS they will die in the room . EOS	[3 22 23 24 11 19 8]	[1 3 22 23 24 11 19 8 2]

The designing phase of machine translation system takes a consideration of two-step processes. These are training phase and testing phase. In encoder-decoder language modeling the training phase is followed by testing phase. The availability of these two phases is related to that the human translator need to learn the structures of language pairs before starting translate from one language to another. In similar manner, machine translation system must take training on the structure of both source and target languages like that of human translator learns one language through practicing. Therefore, like human translator, machine also learns the structure of language pair by repeatedly looking into the structures of the languages from datasets. So, the design of neural network based systems at training phase is shown in Figure 4.1.

After designing training time components, we continued on the designing of components required at testing time to test the systems on testing datasets after training systems. This testing step is the step in which the system must translate the unseen text of the source language to target language from the experience of the trained model of the system gained at training time. In order to measure the translation quality of the system, the translated output of the system is compared against human translated reference sentence. Therefore, for this testing step we have prepared the design of system components which is referred to as system design during testing phase as shown in Figure 4.2.

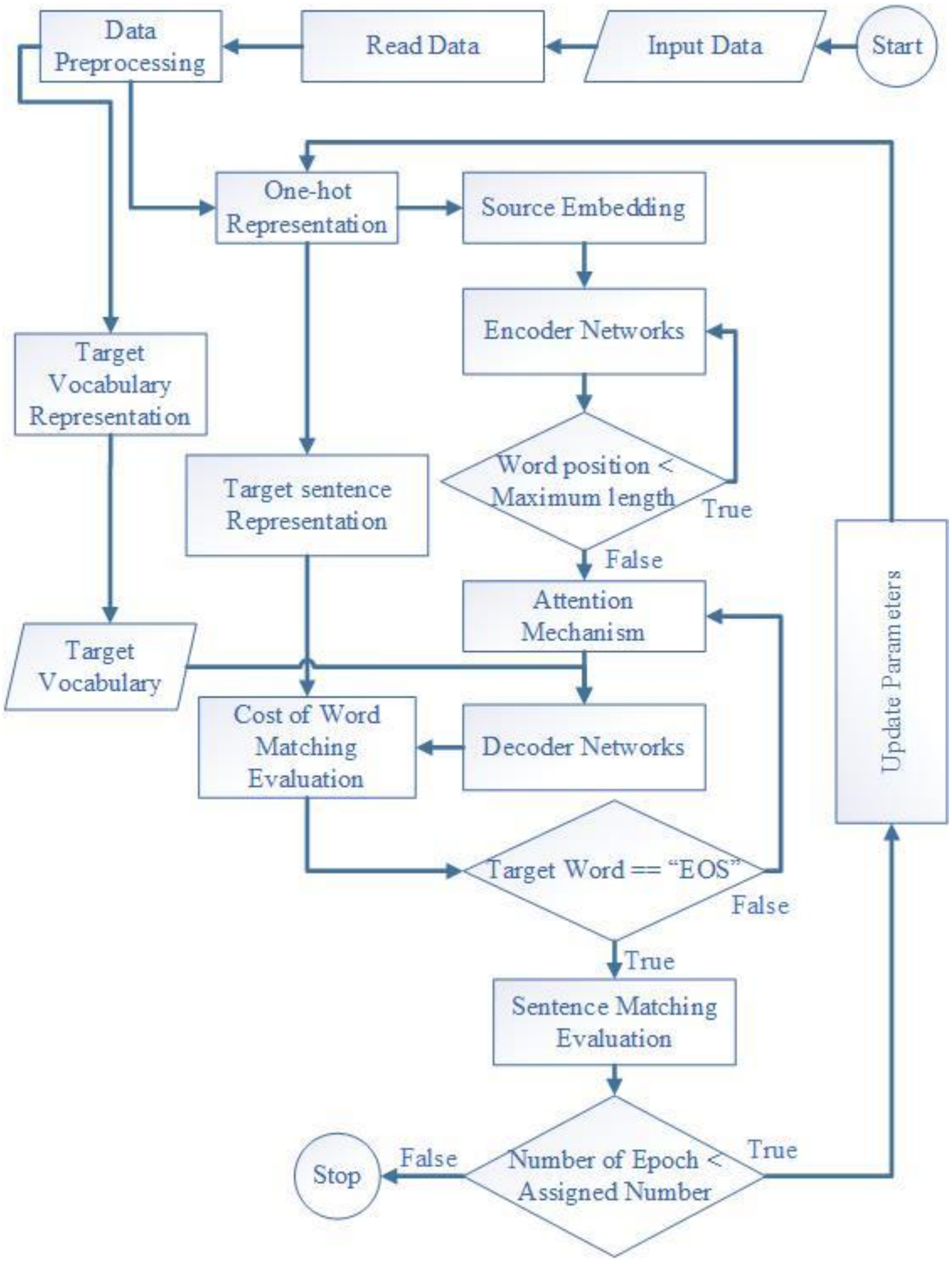


Figure 4.1. System design for training both RNN and CNN based models

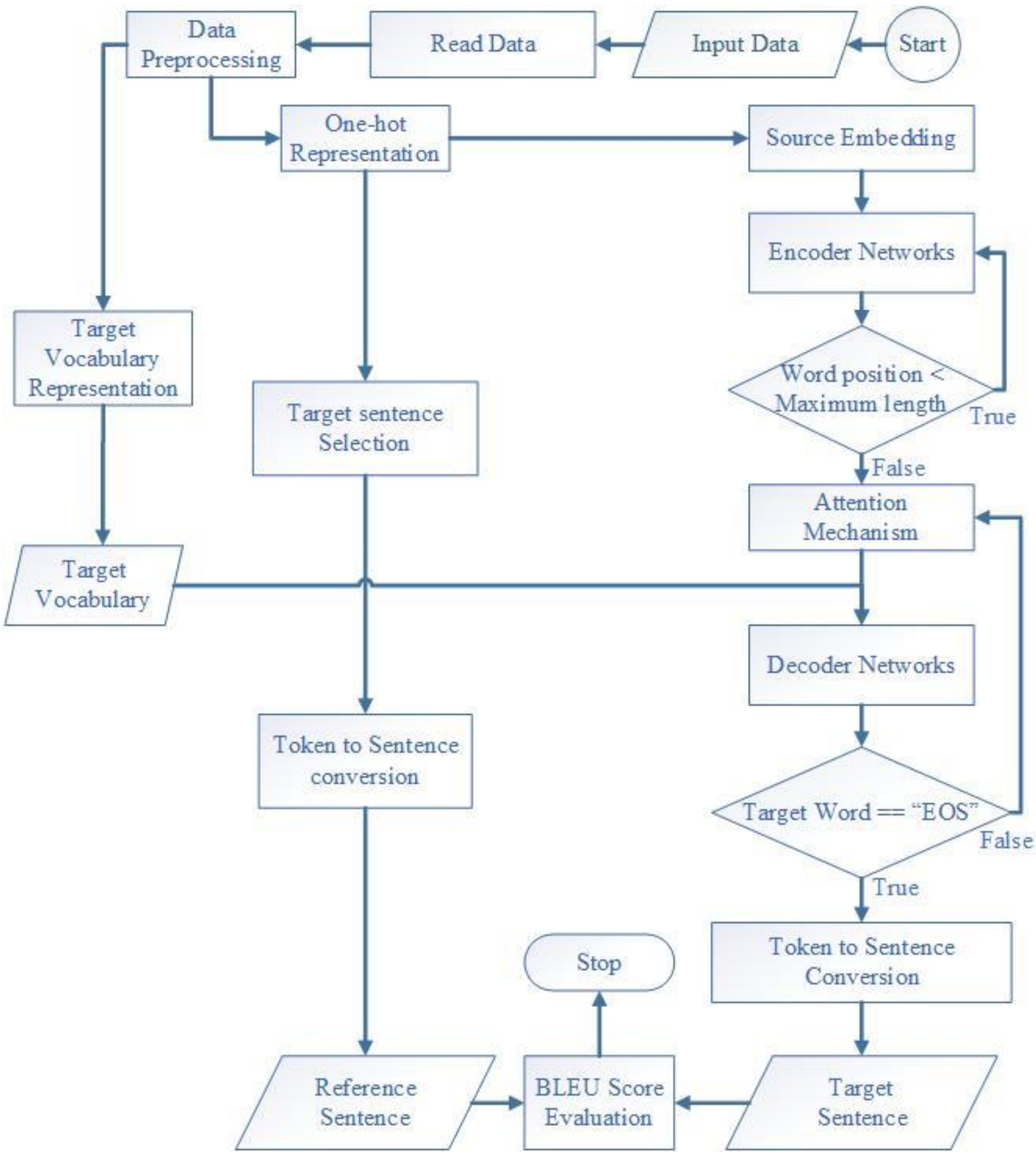


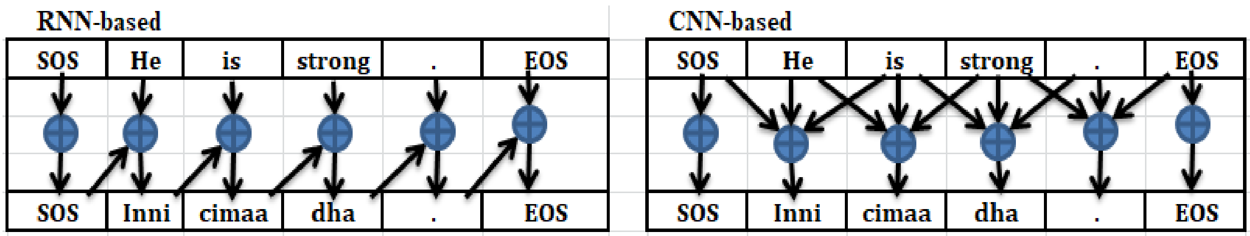
Figure 4.2. System design for testing both RNN and CNN based models

Therefore, figure 4.1 and figure 4.2 describe the designs of our neural network-based systems using encoder-decoder language modeling without specifying the architecture of the neural network. Figure 4.1 shows the training stage in which the system learns the structure of the languages pair and figure 4.2 shows how to test the system to get BLEU-score reading. These components are similar for RNN based approach and CNN based approach except the change of the neural network type and some supporting parameters like parameter optimization functions.

Here, the encoder networks generate contextual relevance between words by breaking a sentence into word(s). The encoder network with RNN-based architecture uses the previously outputted word and the current input word of the source language to give the current output word of the target language. The encoder network of CNN-based architecture uses a fixed number of candidate words (n-grams) to visit to give one output word at a time. Thus, looking at the fixed number of words start from left and go forward up to the end of the sentence is reached. When the end of the sentence has been reached, the encoder ignores current information and will begin the translation of the new sentence (if more sentences are under translation) from the start of the sentence indicator.

From Table 4.2, we how Encoder and decoder of both models operate on input data to give their respective outputs. In the table, we can how RNN based system reads input and can give output at the left-hand side. The encoder of the RNN-based system uses the current input word which gives at top and its previous output to give current output as shown in the table (left-hand side). The decoder of the RNN-based model uses the context-matrix derived by the encoder layer and uses its previous output to give current output like that of the encoder. In the table, we also see how the CNN-based model operates on the data which is described in the table on the right-hand side. The CNN-based model does not use the previous output to give the prediction of the current target. Rather it looks at the relationship between groups of words to give the prediction of target word of the current position.

Table 4.2 Example of context-derivation and word-prediction



The decoder receives contextual matrixes from the encoder network and guided by attention mechanisms to predict the translated target word of the target language. To do this task, the RNN-based decoder uses both currently received input word context and previously predicted word to give the prediction of target word for the current position. Here if the current word-context is at the beginning, the start of sentence indicator will be used as the previous word. But if current word-context is not at the beginning and previous is not relevant to the candidate words, the teaching force technique which forces the neural network to ignore this irrelevant input and considers other alternatives. These predictions of target word will be continued up to end-of-sentence (EOS) will be reached.

The CNN-based model uses only context-matrix by selecting a fixed size of context-matrix a time starting from the start of the sentence and go forward up to the end of the sentence will be reached. Here, using a group of word context-matrix for prediction of current position word in the CNN-based model solves the problem of delaying time for looking backward to check the previous output with that RNN model. Using this group of words also solves the problem of dependency on an only single component of context-matrix which needs teaching force technique when RNN-based model.

4.1.1. Components of designed systems

The designed system composed of several components as described in the figure 4.1 and figure 4.2 for both at training and testing stage respectively. One of our objectives in this study is to improve the translation between English and Afan Oromo language pair in terms of translation quality within effective training time required for training the system. In order to show these improvement RNN based architecture is first applied on the translation between the language pair, and then CNN is applied on the translation between both languages.

Here, the reasons of having RNN model are that we need to show whether RNN architecture of neural network shows translation improvement over baseline system and then we need to compare its performance testing result with the testing result of our CNN based model. The designs of both RNN based model and CNN based model are similar in structure as we have used encoder-decoder language modeling[10]. Therefore, the only difference between both models is that we have changed the architectures of neural networks for both encoder and decoder layers of both models.

Therefore, the next sections explain the components of the systems which are used both at training phase and at testing phase. Except the change of neural network architecture both in encoder and decoder components, we have similar structures both in our RNN based and CNN based models both at training and testing phases. Therefore, we have separately explained only the encoder part and decoder part of both models. Hence, components like preprocessing, word-embedding and attention mechanism are the similar in both RNN and CNN approaches.

4.1.2. Data pre-processing

Before feeding machine with data, we have to preprocess the data to facilitate the training by simplifying the repeatedly reading of data by changing data into appropriate format. The task of data pre-processing includes several activities like vocabulary preparation, selection of the maximum-length of sentences, cleaning the data by removing unnecessary characters and longer sentence, and representing words with unique number. In order to represent each word by unique number, we have to open and visit all the collected parallel corpus data. Therefore, once we have seen a given word, we assign unique number to that word and later the word will be represented by the number. Hence, the last number assigned to the lastly visited word is taken as the length of vocabulary in each language.

Therefore, the availabilities of these activities support the preparation of data for training by changing the original datasets into the representation form which is based integer number representation form. This process of changing original data into this representation form is known as tokenization. The original data set can be in text format, but the tokenized data must be in a format which can be easily readable for machine. Hence, at the step of data pre-processing the raw data will be tokenized and the vocabulary size will be defined.

Therefore, based on the defined vocabulary size and fixed sentence length, each sentence of the data will be represented in numerical representation which is known as one-hot vector representation. In one-hot vector representation each word with in the sentence is represented by integer value as describe in Table 4.1. At the start of tokenization, as soon as a given word is visited, it will be represented by using unique number. Therefore, the machine later uses this unique number at each the time of visiting the word in the parallel corpus dataset.

4.1.3. Embedding

For neural network-based machine translation, the cleaned parallel corpus has to be changed into the format which is suitable to train machine the structures of both languages from the collected parallel corpus. In the process of language modeling; word embedding plays a great role in adjusting the data into a form suitable for the machine to learn from. The machine cannot directly work on a sequence of sentences. Rather, these sentences must be split into a sequence of words, and then these sequences of words must be changed into integer form representation as shown in Table 4.1a. This integer form representation represents each word in the sentence with its unique integer number derived at the time of vocabulary formation. This integer form representation, which described in Table 4.1b, is known as tokenization.

After changing the dataset into integer representation form representation, it has to be changed into some form of numerical representation which is known as one-hot vector representation. This one-hot representation uses a unique vector representation for each word of the sentence of the dataset and uses zero-padding to set lengths of sentences to fixed lengths as described in Table 4.1b(right-hand side). Therefore, by representing our data, now we have a longer vector of a two-dimensional vector with most elements have zero for shorter sentences. If such vectors are directly given to neural networks, the neural network learns nothing rather than wasting training when zero values encountered. To solve this problem, using word embedding form representation is the only choice. Therefore, word embedding changes a higher dimensional one-hot vector into a lower-dimensional vector by keeping related words in a closed area of vector space by using algebraic vector representation. This word-embedding form representation keeps related words in a similar position to simplify representation as described in Figure 4.3. Therefore, neural networks can read the input data through word embedding form representation.

Figure 4.3 shows the word-embedding form representation of small data by using two-dimensional spaces. In the figure, only two dimensions of word embedding are used and words are scattered in space based on the frequency of occurrence and relatedness relative to other words in the datasets.

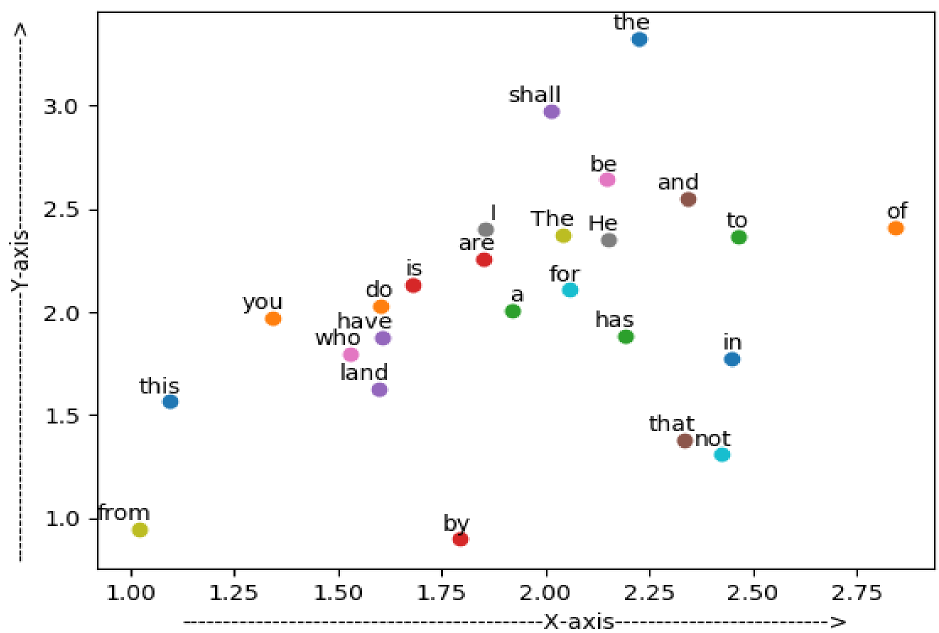


Figure 4.3 Example of word embedding representation form in two-dimensional space

4.1.4. Encoder

In this study the main concern is the architecture of both encoder and decoder components. In our encoder-decoder model, we have used RNN based model to design encoder for our competitive model, while CNN based architecture is used for our proposed system. These designing are needed to compare the performance in terms of translation quality and training time needed for both architectures.

Therefore, designing of encoder part considers the adjustment of number of unit needed to receive and process the data, choices of number layers needed to read the total words of source language sentence, choice of layer types whether recurrent neural network or convolutional neural network is selected, and the choice of required technique for network optimizations like choosing weight initialization or updating mechanisms, selecting appropriate formulation algorithm or choosing proper learning rate[46].

4.1.4.1. Designing Encoder for RNN based system

The encoder of RNN based system is designed based on gated recurrent neural network based architecture which is proposed in[10]. The gated recurrent neural network (GRU) uses gate unit to control the flow of information. It uses the current input and its previous output, which can be considered as the current internal state of the network, to give current output[47]. The output of encoder is computed as context vector that derived from source sequence, x_t , *current internal state,* h_t, *and gives current context* z_t *as:*

$$z_t = f(x_t, h_t) \quad \text{[Equation 4.1]}$$

In order to implement encoder with this architecture, we designed the number of internal units required to read the input sequence from the word embedding layer with equal size of the word embedding. Therefore, each element of vector from word embedding element is read by each unit of the network. The following figure shows the arrangements of gated recurrent neural network to form the encoder part of RNN based system.

In figure 4.3, we can see the structure of encoder network to read data from word embedding layer and use the start position indicator in order to generate context vector for RNN model from sequence of input sentence. The read of sentence starts from start token and continues up to end token. In order to use similar length, zero padding is used after end of sentence indicator. This zero padding fills zero to shorter sentence to equate with longer sentence.

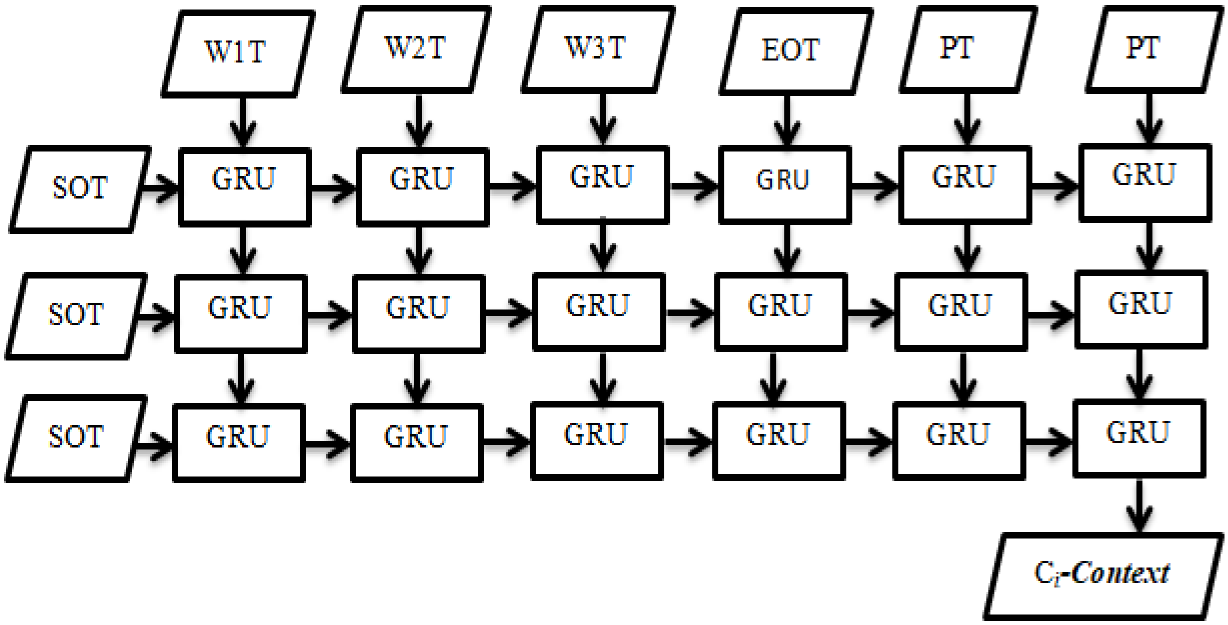


Figure 4.4 Design of encoder network structure for RNN based model

4.1.4.2. Encoder for proposed system

The encoder of our proposed system which is designed from convolutional neural network architecture takes the vector representation of source sentence which may have variable length and converts into fixed size vector representation. In order to make this conversion, the convolution layer works convolution operation on input vector by using the fixed size kernel window which is used as filter which selects only inputs that enters with in the size of kernel window by ignoring the part of input which is found at outside of the window.

Therefore to design the encoder part of our proposed system, we selected the convolutional neural network which is created with the convolution layer followed with rectified linear unit (RLU) non-linearity function with fixed number of layers[48], [49]. The convolutional layer reads sequence of words in sentence with specified length of words, which we have fixed as a length of longer sentence in our data set and performs convolution operation by selecting only a fixed length of sequence of words, which equal to the kernel width, for one time of visiting the data[39].

In order to visit the total sequence of words; the sliding window (fixed size filter) moves all over the sequences of words of source sentences. When words with in filter are visited at particular, the RLU non-linearity function gives the currently generated word context vector at each time of visiting the sequence words in the sentence. Therefore, this activity has to be repeated by different layers for detail construction of context vector

So, in order to derive the context of source language, our encoder layer accepts source data through word embedding layer which is vector representation of source language sentence. After deriving the new vector which represents the contextual relationship between words of the sentences by generating one word at time, it will feed the context vector into attention mechanism layer.

Therefore, these set of components are shown in figure 4.4 in which the word embedding layer is shown at top with its dimension and the width of kernel window, and the structure of encoder is shown at middle, while bottom component shows the derived context vector of the source sentence. The last component represents an attention mechanism layer which inter-connects the encoder network to the decoder layer of our system.

Therefore, in Figure 4.4 we can see that our CNN based model reads word embedding vector from the start to end with the use of fixed kernel window (filter) size. Hence, after reading input, it changes this feature vector from previous vector form to new vector form by passing in convolution layer and generate context vector through RLU non- linearity function. Therefore, this context vector is later fed into decoder through attention mechanism.

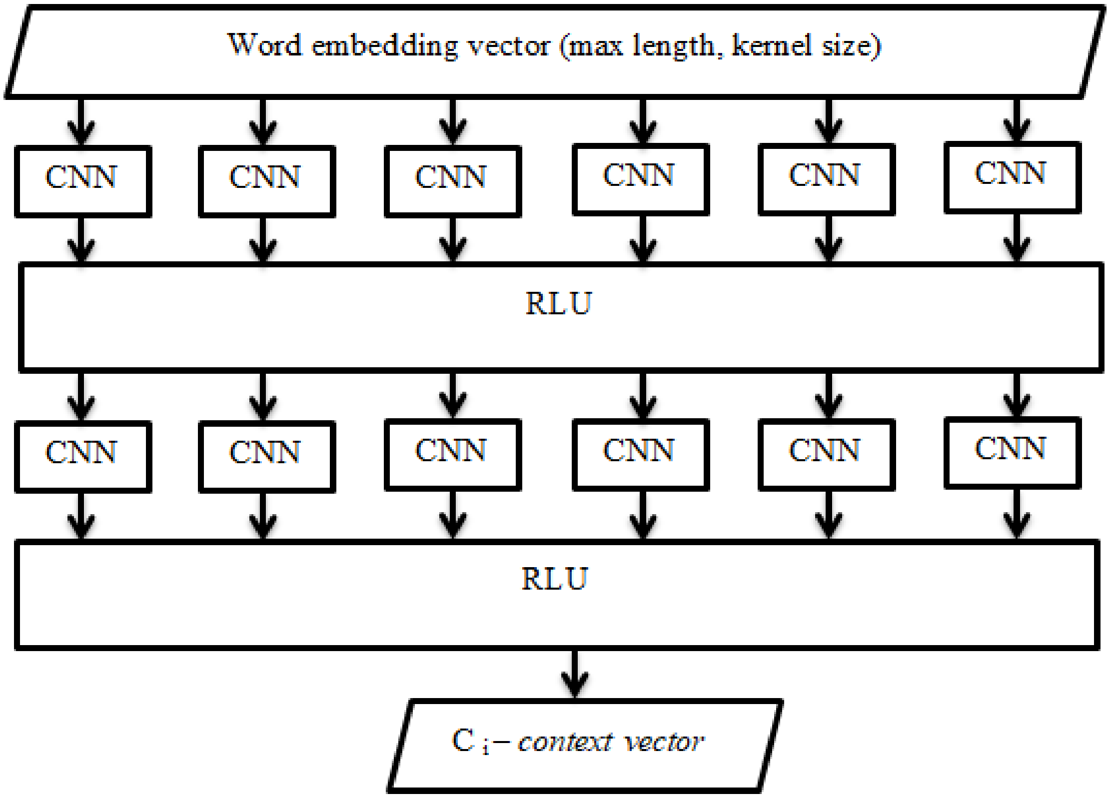


Figure 4.5. Design of encoder part CNN based model

4.1.5. Attention Mechanism

The attention mechanism is important to support the storing of longer context vectors with in the sequence in order to help the network to remember all the visited words of a given sentence[50]. Therefore by using this attention mechanism, the network pays more attention on some part of source sequence which is more relevant to sequence of target sentence by ignoring the other parts of the input data.

In order to evaluate the attention value required to deliver context vector to decoder from encoder part, computing the context vector derived from the last layer of the network with use of attention formula is required. Therefore, the context vector within sequence data can be evaluated as

$$Ct = \frac{\sum_{1}^{n} exp(Hj.Ht)Hj}{\sum_{1}^{n} exp(Hj.Ht)}$$

$Ct = (\sum_{1}^{n} \boldsymbol{a(t,j)Hj})$ [Equation 4.2]

After getting attention information, the new created decoder part combines the context vector with previous internal state to generate new internal state required to search for appropriate target word which fit to derived context.

4.1.6. Decoder

The decoder part in encoder-decoder language modeling is the component which is dedicated to searching of the word of the target language that matches the contextual vector received from encoder layer. In order to perform this task, the decoder must first train on sentences of the target language to get the structural information of the target language. Therefore, we designed the structure of decoder uniquely for both RNN based and CNN based models in similar way we had designed the structures of encoder parts. So, the next sections discuss decoder for both RNN and CNN based models.

4.1.6.1. Decoder for RNN based system

The decoder part of RNN based system is designed by using gated recurrent neural network (GRU) architecture which uses the input of currently generated context vector and previous output of the decoder network architecture[47]. The number of GRU used is set as maximum length of the sequence in the target sentence data and when the shorter sentence entered into the network, the end of sentence must be marked with end of string indicator symbol and the other units are fed with padding which is used to change the variable length vector in to fixed length vector by completing the length of shorter sentence by using zero value.

Therefore the decoder receives each element of context vector through each unit of the network and uses its previous output, which is considered as its internal state, to generate the current output of the system. However, when the word under consideration is the starting word, then that word must be marked with start of string indicator to ignore the internal state consideration as the word must be translated into the beginning word of the target sentence.

In order to find the matching word, decoder uses the searching mechanism of best search type with the use of softmax function which searches for the best matching word with in less expensive cost. Therefore the decoder generates one word at a time and repeats the searching until encounter end of string indicator which shows the last word of the sentence under translation. So figure 4.5 describes our RNN based model that receive context vector from attention mechanism and gives target word output.

Therefore, figure 4.5 shows the decoder part of RNN model which receives context vector with start of token indicator to generate target word through softmax function given at the bottom.

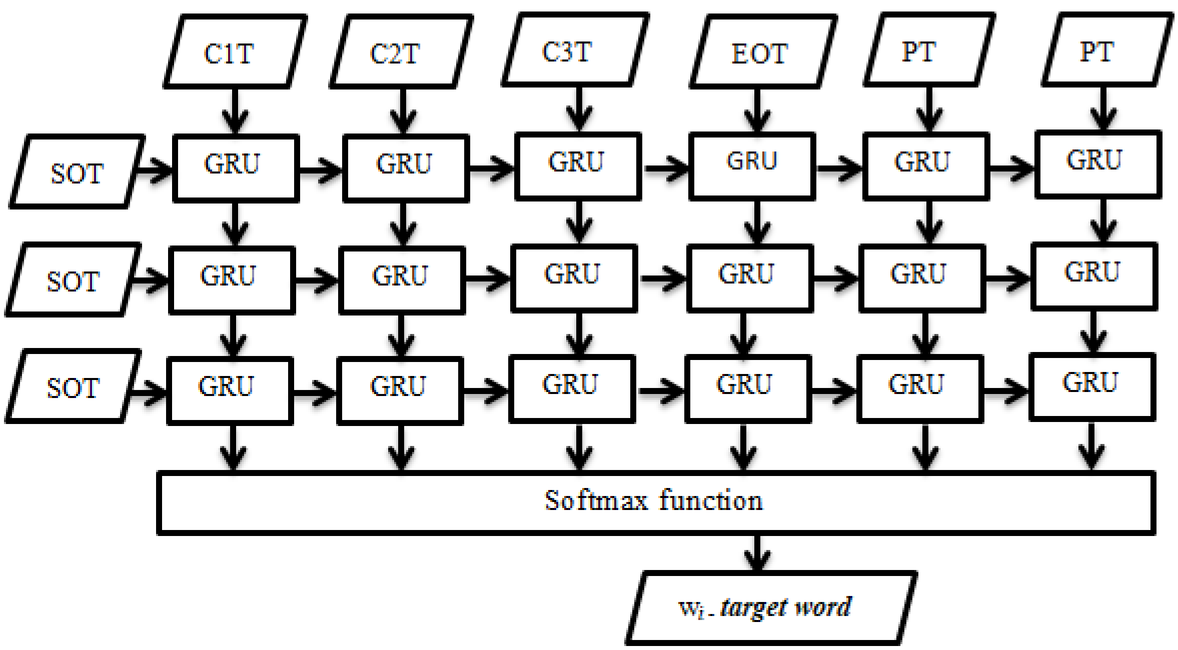


Figure 4.6 Design of decoder part for RNN based model

4.1.6.2. Decoder for proposed system

The decoder part of our proposed system, which is CNN based, is designed by using convolutional neural network based architecture. The decoder receives the context vector derived by encoder part of the system through attention mechanism. After receiving the context vector, the decoder predict the target word that match to context by use of softmax function which is activation function that supports output modeling for deep neural network models[51].

In order to design the internal structure of our CNN based model, we used the convolutional units which are equal with the size of our word embedding and we have tried different dimensions to select from. Frist, we begun by setting the size of both decoder units and embedding size to 128 dimension and we have seen that the loss quickly converges but starts to diverge before arriving at required level. Then we tried 256 dimensions and it works better than the one before. Again we tried 512 dimensions, but this became worse on loss convergence at the time of neural network training.

Therefore we selected both the dimension of word embedding and internal units of the decoder as 256 dimensions. Also like that of our CNN based encoder network, the decoder part also built by convolutional layer followed by RLU non-linearity function to get output from decoder layer. In order to produce the translation, the decoder receives the input context through attention mechanism by reading the context vector with help of kernel window which is used as filter to select the part of input context for neural network.

Therefore, the decoder uses this fixed size kernel window to read a context and uses beam searching mechanism which is searching for best matching word of target language within less cost mechanism. Therefore, softmax function facilitates this searching mechanism by generating one target word at a time and the searching is repeated until the end of string indicator is generated. Thus, the translation of the sentence is the collection translated words starting from start of token indicator goes up to the end of token indicator.

Hence, from figure 4.6 we can see the designed internal structure of decoder for CNN model which receives derived context vector of source sentence to read with kernel window size in order to generate target word through softmax function after doing convolution operation on context vector and pass through RLU layer to feed softmax function.

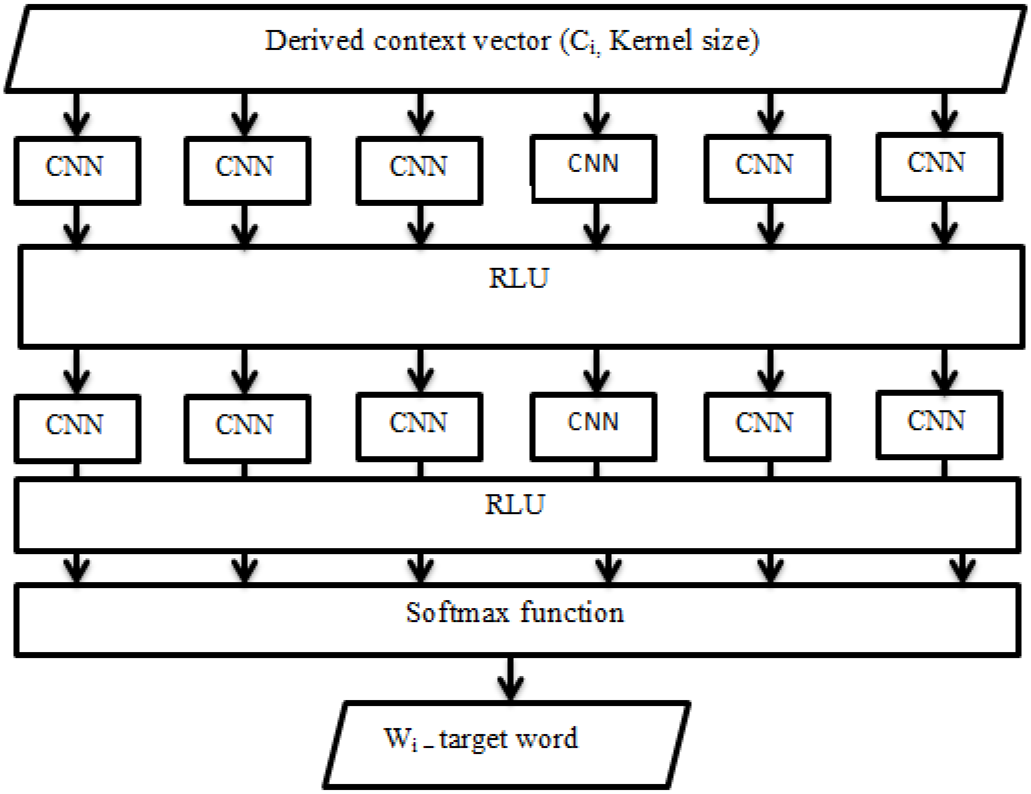


Figure 4.7 Design of decoder part for CNN based model

4.2. Inter-connection of components of the system

Here, the implementation of designed systems needs to properly inter-connecting of the components of the designed systems. This considers the interconnecting word embedding with encoder and providing context vector to decoder either through last layer or through attention mechanism from encoder network. In order to train the implemented systems, the consideration of parameter initialization and optimization is necessary[35].

Therefore, as the weight initialization affects the total operation of the networks and for the network to operate appropriately, the careful weight initialization plays a vital role[42]. Thus, the uniform random distribution (N(0, 0.5)) works better for weight initialization when activation function like tanh and RELU are used in network[35], [48].

Once training has started, the consideration of how to evaluate the quality of prediction of the network is necessary. This can be done by selecting loss evaluation and minimization technique by selecting appropriate algorithm. Hence, we selected RMSProp algorithm which is the type of Adaptive algorithm recommended for that it helps the quick convergence of loss gradient [35].

4.3. Design of the baseline system

As one directional statistical machine translation from English and Afan Oromo has been done by Sisay, here similar design of word-based statistical machine translation is required to be used as baseline of our system in order to compare the performance improvement made by applying convolutional neural network based architecture. However, in our case the designed system must include both direction of translation, because our goal is to design bi-directional machine translation by using of neural network approaches.

The statistical machine translation uses statistical probability of n-gram co-occurrences words within a sentence of a given language to learn the structure of the language. This statistical probability evaluation is done on more than two words like 2-gram co-occurrence of two words, 3-grams co-occurrence of three words and so on[31], [52].The designing of this model needs the consideration of language modeling part, translation modeling part, tuning part and decoder part.

The language modeling part receives the target language data and computes the probability of the co-occurrences of word for each sentence and assigns this probability to each individual words of the target language. The language modeling part is done by using IRSTLM which is free available tool for language modeling. The language modeling part generates the probability of the co-occurrences words with the language sentences as

$$P(wn|w1, w2,\ldots,wn-1) \quad \text{for 1-gram}$$

$$P(wn, wn-1|w1, w2,\ldots,wn-2) \quad \text{for 2-gram}$$

$$P(wn, wn-1, wn-2|w1, w2,\ldots,wn-3) \quad \text{for 3-gram , so on.}$$

The translation modeling part accepts the bilingual parallel corpus and assigns the corresponding translation probability available between sentences of both languages. This probability is computed by equation 4.3 by taking *s* as source language sentence and *t* as target language sentence.

P(s|t) = $\frac{p(s)p(t|s)}{p(t)}$; p(t) = 1

P(s|t) = $(\boldsymbol{p}(\boldsymbol{s})\boldsymbol{p}(\boldsymbol{t}|\boldsymbol{s}))$ [Equation 4.3]

Therefore the derived language model and the generated translation model are combined by decoder component which performs the searching mechanism based on the probability maximization. This probability maximization means that the decoder searches for highest generated probability derived from the combination of language model and translation model components. Therefore the decoder uses argmax function which generates the highest possible probability of matching sentence for translating source sentence to target sentence. This probability can be computed as:

P(t) = ê = $argmaxp(s|t)$

P(t) = $(\boldsymbol{argmaxp}(\boldsymbol{s})\boldsymbol{p}(\boldsymbol{t}|\boldsymbol{s}))$ [Equation 4.4]

In order to implement our baseline system using STM model, we designed word-statistical machine translation with the architecture shown in Figure 4.7 in which language modeling uses only target sentences to make language model (in left side of Figure 4.7) and translation model uses both source and target sentence (right side), and finally decoder combines both models to give target translated sentence which finally measured by BLEU score metrics.

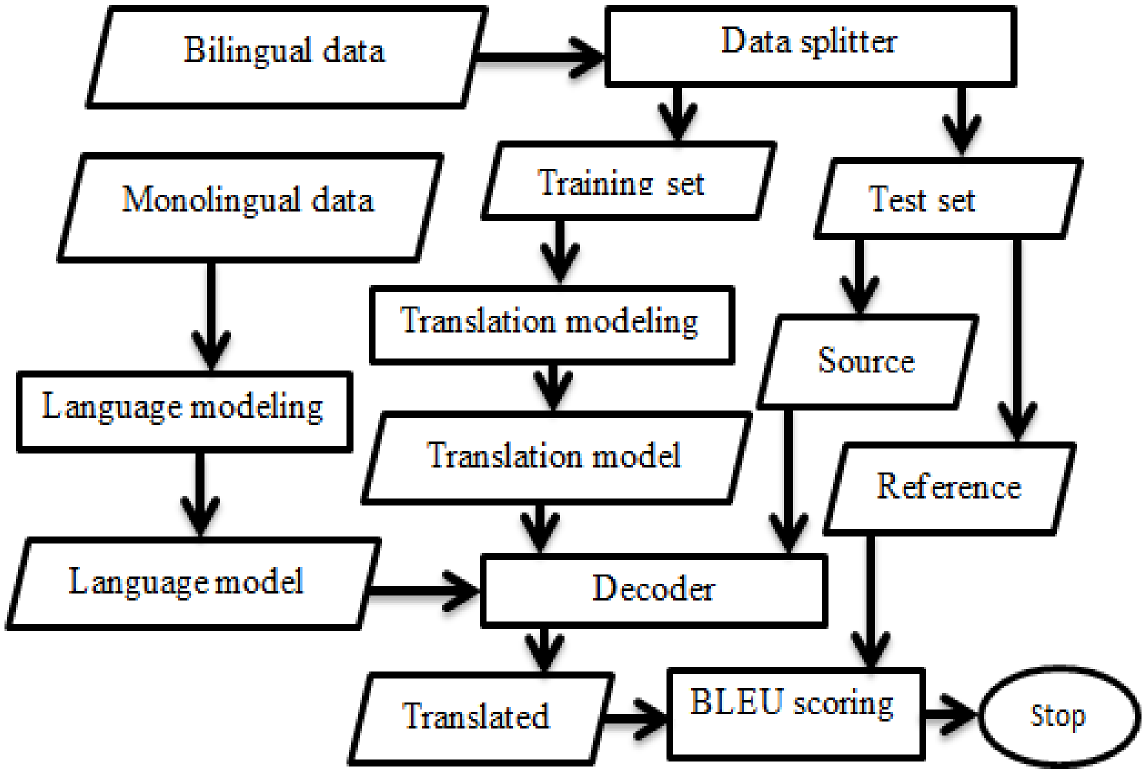


Figure 4.8 Design of the baseline using statistical machine translation

4.4. Data collection for the study

For our study, the parallel corpus is prepared from the data collected from different sources that are available in both languages. Some of these sources are published books while others are online available data. So, the sources used for data collection are:

- Holy Bible
- FDRE Criminal Code, FDRE constitution
- Megeleta Oromia (Constitution of Oromia state)
- Published Conversational books: Geda conversation, Effortless English-Ahmaric-Afan Oromo phrase based translation
- Data of Ethiopian Revenue and Customs Authority
- Online available data: phrase based translated data for Afan Oromo language learners, online guide for job seeker

As there are only few materials available in these language pair, the total number of parallel corpus collected is 5550 sentence pairs. To keep structural similarity between the structures of both languages, the sentences are carefully selected. This is important, because in some part of the materials, the translators had not given attention to keep similarity for the structures of the languages even not context, rather the translator focused on transmission of idea of the story. However, in the case of machine translation machine has to deeply learn the structure of both languages before the act of translation.

4.5. System environment

For this research work, the designed systems have to be implemented by using two different deep learning algorithms. These are RNN based architecture and CNN based architecture. To implement these systems, preparing the required environment is necessary. At the first step, the sequence of text data has to be changed into one-hot vector representation. This activity uses numpy library which is freely available.

After data is changed into one-hot vector, the vector will be fed into the model of the system. The model of the system is built by using torch tensor. Torch is freely available library for deep learning algorithm in python. Tensors are standard form of vector representation to facilitate machine training in deep learning algorithm. These tensors are multidimensional arrays which are an extension of two-dimensional matrices. After setting up the environment, the designed system will be coded. We have chosen python programming language to implement the system. Python programming language supports these set of freely available library.

After implementing system, we have run and tested on core i5 CPU based computer. Finally, to compare the level of improvement of machine translation between both languages in comparison with previously used technique for translation between both languages, we have implemented word-based statistical machine translation. For this SMT technique, we have used freely available libraries like Giza++ for sentence-based aligning of parallel corpus, IRSTLM which estimates probability for language modeling and Moses decoder which searches for matching target sentence for translation.

Chapter Five

5. Experimental Results and Discussion

This chapter discusses the experimental results of this thesis work by showing the experimental setups and performance testing results of the experimental systems using BLEU score metrics and time units. Also, it gives comparison of the results among all systems on both directions of translation between the language pair.

5.1 Dataset preparation for training

In order to perform our experimental study, we have collected datasets of total 5550 sentences of parallel corpus from different sources such as religious books like Holy Bible, data collected from some published conversational books, data of online available conversational sentences and data from governmental sources like constitutions (both regional and federal), data from Oromia regional revenue and data from Oromia health sector. After we had finished collecting the datasets, then we aligned this data of parallel corpus along its translation with the use of tab separation.

Therefore, this aligned data is stored in text format to feed into systems. After collecting datasets, aligning and storing the data, we shuffled the data elements before classifying into training set and testing set. After shuffling the data element, we have classified the data elements into training set and testing. The training set consists of 4440 sentences of parallel corpus which is 80% of total collected data, while the test set consists of 1110 sentences which is 20% of total datasets.

In order to feed the system each set of parallel corpus, the system opens the total textual formatted datasets and splits the data into set of sentences for both training and testing data sets. After splitting all elements of datasets into sequences of sentences for both sets, then again splits sentences into sequences of words. Therefore, these sequences of words in sentences can be fed into the system by use of integer form representation, which is known as tokenization. In order to form integer form representation, each word in the data is assigned with one unique integer number as soon as visited. So, the maximum number that used as word id is also used to define the size of vocabulary of the dataset for one language. Therefore, our dataset consists of 8527 vocabulary size for English language and 4221 for Afan Oromo language.

5.2 Experimental setups for the models of systems

In order to accomplish our experiments, we have implemented our designed systems by specifying required parameters based on different experimental tests. In order to implement our baseline system, we have used Giza++ to pre-process our parallel corpus data by aligning and tokenizing each parallel corpus sentence. As our baseline system is implemented by using word based statistical machine translation, we have implemented the language model and translation model by using IRSTLM language modeling tool which is freely available library for statistical machine translation language modeling task.

After creating both language model and translation model by use of IRSTLM tool, we fed these models into decoder to train the model of the system. The decoder component is implemented by freely available Moses library. This decoder component uses Greedy searching mechanism to search for matching word of target language. Therefore, after training has been completed, we tested the system on testing dataset. Hence, after all the sentences of testing set has been translated, the performance measuring in BLEU score is recorded as our baseline result to compare with the results of other neural network based experiments.

In our second experiment, we implemented our RNN based model with use of GRU type of RNN architecture using encoder-decoder based language modeling designed in Chapter 4. For this system, we adjusted parameters based on the idea of loss reduction which is the process of reducing the difference between exact value and system predicted value to supervise the training of neural network. In order to uniquely represent the words of parallel corpus, we have used word embedding representation with size of 256 as our vocabulary is small by comparing against vocabularies of other European language pairs which have large datasets.

Therefore, in order to read each word representation with individual neuron unit; we have selected 256 numbers of neurons for each hidden layer of neural network in both encoder and decoder networks. In each encoder and decoder networks, we have used 3 layers of hidden layers for neural networks to strengthen the learning of the contextual relevance with in the datasets. In order to support longer contextual dependency between words of the sentences we have used attention mechanism which is dot product type in order to feed the output of encoder network into decoder network. Therefore, the decoder network uses its internal state and attention mechanism output to search target word at a time with a use of beam search algorithm by using logarithmic softmax activation function.

Lastly, we implemented our CNN based model which is our proposed model for the study of machine translation between these language pair. In order to implement this system, we have used encoder- decoder language modeling which is similar to our RNN architecture model in structure but with CNN architecture. In order to adjust the network, we have used different numbers for both embedding size and neural network elements. We have used 128, 256 and 512 sizes at different times of experiments and finally we have chosen the 256 for both embedding size and network units, because the gaps between systems predicted value and exact value have been smoothly converged over a number of epochs as shown in Figure 5.1a.

5.3 Parameter optimization and training the experimental systems

In our experimental study, first we trained and tested our baseline system which is word-based STM model on our training and test datasets respectively. The training part of this model includes sequence of steps like preparation of language model for target language, preparation of translation model which defines how each word will be translated from source language to target and decoding which is the process of searching for target word from available vocabularies of target language to generate translation output. The training of this model with these tasks took an average time of 4 hours and 18 minutes on core i5 CPU with 4GB RAM based computer for English to Afan Oromo Translation, while it took 4 hours and 12 minutes for Afan Oromo to English Translation.

After completing our experiment with the first system, we continued the training and testing of our RNN architecture based system on our parallel corpus. As a training of neural network needs the adjustment of parameters, we have initialized parameter by using uniform distribution randomly between -0.5 and 0.5, and we selected sigmoidal optimizer to optimize parameter. We have used small learning rate of 0.1 to train the neural network. Therefore, training the model on our training dataset for 14 epochs took 7 hours and 07 minutes on translation from English to Afan Oromo translation, while it took 6 hours and 58 minutes for translation from Afan Oromo to English translation.

Finally, we implemented and trained our CNN based model to train on training dataset and test on our testing dataset. Therefore, in order to get competitive result as that of our recurrent neural network and better than that of our baseline models, we have done different experiments on different occasion to adjust the parameters of the model. First, we selected between different number of size which are used for size of neuron units and word embedding size. At this stage we have seen that small size of neural network faster on training depending loss level versus number of epochs(number of looking back into a single dataset element) graph. The bigger size shows a sign of faster convergence at beginning, but starts to diverge very soon. Therefore, we chose the medium size of word embedding and number of neuron as shown by Figure 5.1 a.

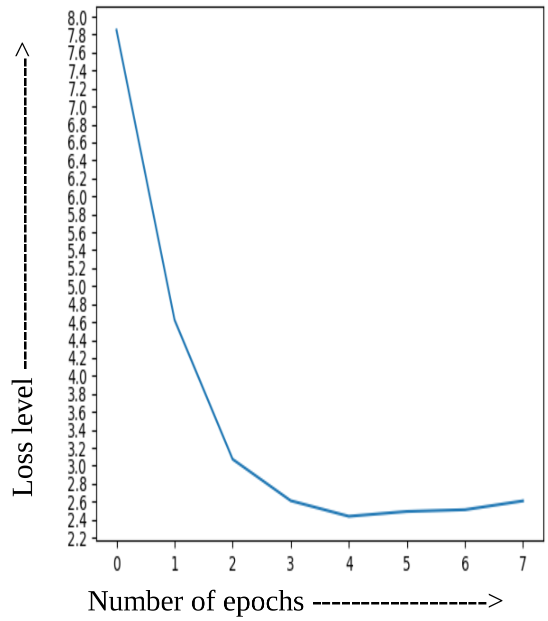
After adjusting network size and work embedding, we chose between different learning rate and batch size of training data. Therefore, we have got smooth convergence with small learning rate of 0.2 as seen on Figure 5.1b. In order to train CNN based model, we have used batch training with batch size of 32 and trained for 14 epochs on CPU based core i5 with 4GB RAM computer. Therefore, the training of our CNN based architecture took 5 hours and 47 minutes for translation from English to Afan Oromo, while it took 5 hours and 29 minutes for translation from Afan Oromo to English translation.

Figure 5.1 describes effects of different parameters on adjusting neural networks for implementation both neural network based models. In Figure 5.1a we are looking for best-fit size word-embedding and number of neurons. This fitness of size can be decoded by convergence of the graph. The result in the graph is found by training the CNN model for 7 epochs which means that the model has looked the training data for seven times. Therefore, from graphs we can see that the 128-size of word-embedding and neuron numbers fit to our training data.

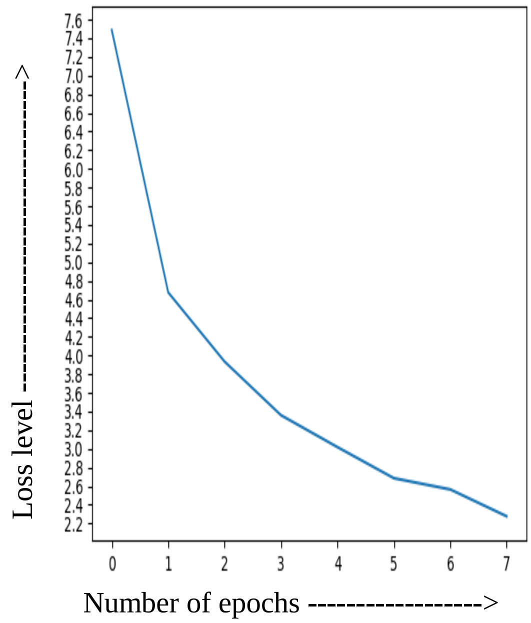
In Figure 5.1b shows the effect of selecting different learning rates to train the neural network models. The necessity of selecting best-fit learning rate is that if appropriate learning rate is not selected, the neural network will face the problem of not fitting. In the Figure, the learning rate with 0.5 was fast on training, but did not appropriately minimize the loss function to predict the sentence translation to relevant. However, the learning rate with 0.2 is best-fitted to our data by reducing the loss function appropriately.

In Figure 5.1c shows the graphical comparison between both architecture of neural network on machine translations between these language pairs. From the graph, we can see that CNN based model is faster on training by using similar parameters. In the graph, we have used 8 epochs (times of looking into total training data) and at the top both graphs RNN and CNN based models have high amount of loss which has been minimized through the increase of epoch number.

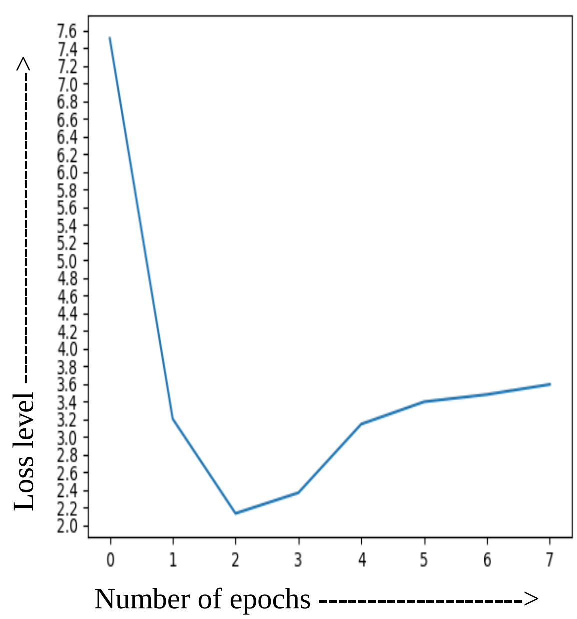
Loss level versus Number of epoch graph when using 128-size for word-embedding and neuron units



Loss level versus Number of epoch graph when using 256-size for word-embedding and neuron units

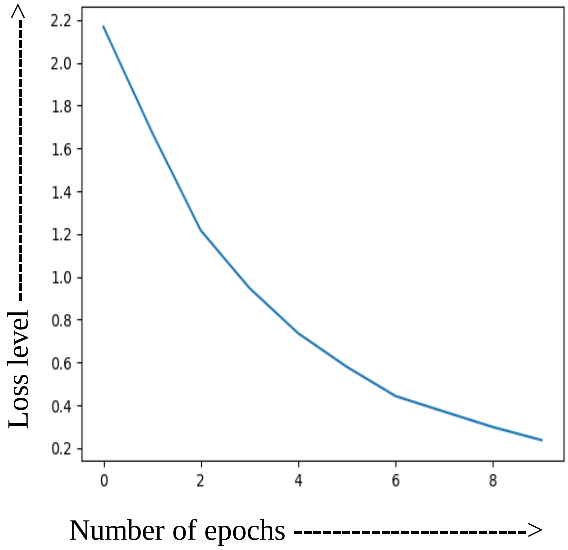


Loss level versus Number epoch for training by 512-size of word-embedding and neuron units

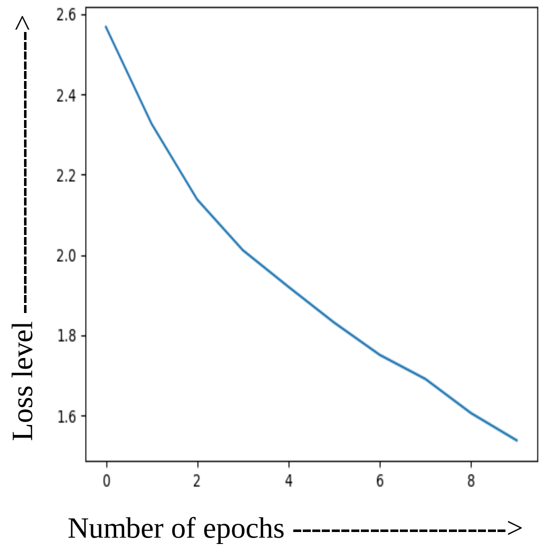


a)

CNN: 256-units and embed, 3-layers
Learning rate = 0.2

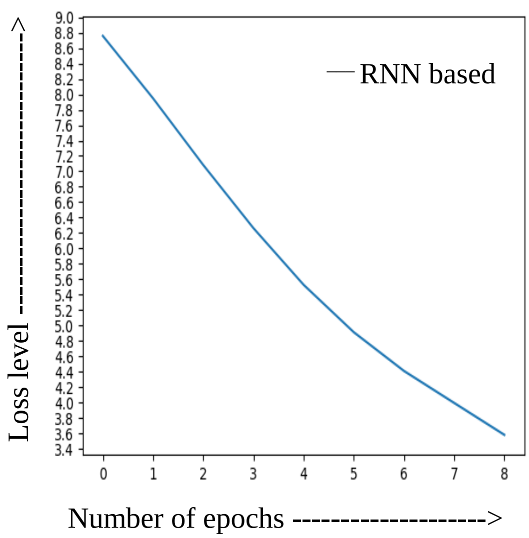


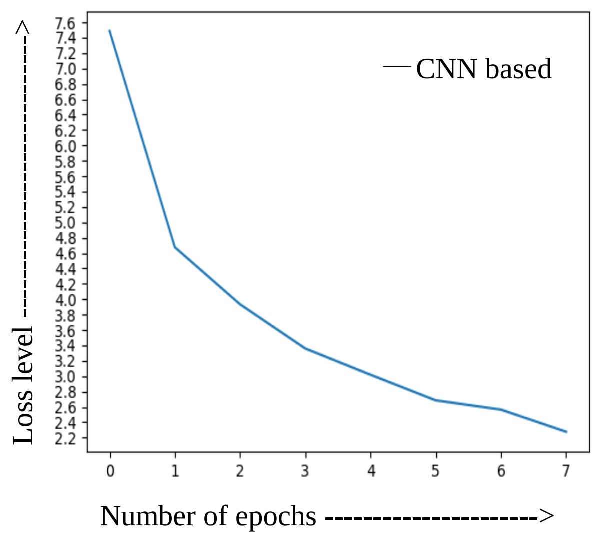
CNN: 256-units and embed, 3-layers
Learning rate = 0.5



b)

Parameters: 256-units and embed size, 3-layers, learning rate = 0.2





c)

Figure 5.1 Graphs to show loss-level with respect to number epochs to adjust parameters

5.3 Experimental results

In our experimental study, we have implemented the three designed systems of our study and trained the systems in similar way to get correct comparison of their performance. After training and testing each system, we have measured the result of testing the system by using BLEU score metrics to see the difference in their scores. In order to report the result of our testing, we have used the average BLEU score result of five times of training and testing the systems. Therefore, this average result of baseline (STM) system, RNN based system and CNN based system is described by table 5.1.

Table 5.1 Experimental result in BLEU score metrics

System model type	English to Afan Oromo in BLEU score	Afan Oromo to English in BLEU score
Baseline (STM) model	20.51	19.86
RNN based model	22.79	21.67
CNN based model	24.37	23.18

Therefore, from table 5.1 we can see that our RNN based model shows better translation accuracy in BLEU score than that of our baseline system in both direction of translation. Furthermore when we compare the result of CNN based model with that of both our RNN based model and STM model, our CNN based model shows better BLEU score results in both direction of translation. Lastly, when we compare both direction of translation, English to Afan Oromo translation shows better BLUE score result than Afan Oromo to English translation.

Table 5.2 Sample outputs of our systems a) translation from Afan Oromo to English b) translation from English to Afan Oromo

a)

Translating the same sentence	Input	Kunoo ilmi namaa harka cubbamootaatti dabarfamee kennamuufi .
	Reference	behold the son of man is betrayed into the hands of sinners .
STM Model	Translation	behold man is betrayed during came next to him .
	BLEU score	32.37
RNN Model	Translation	the son of man came about great man whom you in .<EOS>
	BLEU score	37.015
CNN Model	Translation	Behold the son of man is betrayed into hands . <eos>
	BLEU score	46.274
Translating the same sentence	Input	yisaaq yaaqoobin dhalfate .
	Reference	isaac begat jacob .
STM Model	Translation	he begat child Jacob.
	BLEU score	61.27
RNN Model	Translation	matthan begat jacob . <EOS>
	BLEU score	50.53
CNN Model	Translation	Isaac begot Jacob. <eos>
	BLEU score	74.40
Translating the same sentence	Input	Torbee dhufu boqonnaa qabdaa?
	Reference	Are you free next week?
STM Model	Translation	You are free tomorrow?
	BLEU score	47.32
RNN Model	Translation	Can you have anything next week? <EOS>
	BLEU score	32.43
CNN Model	Translation	Are you free?? ? ? <eos>
	BLEU score	50.25
Translating the same sentence	Input	Dhiifama yaada kee tasuma hin hubanne .
	Reference	Sorry i didn't quite catch what you said.
STM Model	Translation	I didn’t quite sorry you are there the.
	BLEU score	62.34
RNN Model	Translation	Sorry i didn't catch that your eyes. <EOS>
	BLEU score	61.90
CNN Model	Translation	Sorry i didn't catch what you said . you . <eos>
	BLEU score	73.47

b)

Translating the same sentence	Input	the details shall be determined by law .
	Reference	tarreeffamni isaa seeraan murtaa a .
STM Model	Translation	tarreeffamni dubbichaa seeraan ta a .
	BLEU score	36.27
RNN Model	Translation	tartiibni isaas akkaataa seeraan murtaa a . <EOS>
	BLEU score	54.4
CNN Model	Translation	tarreeffamni isaa seeraan murtaa a . <eos>
	BLEU score	85.53

Translating the same sentence	Input	arise and take up your stretcher and walk.
	Reference	ka'iitii siree kee baadhuu deemi.
STM Model	Translation	uffata kee baadhuu kottu ka'i .
	BLEU score	32.37
RNN Model	Translation	afata kee baadhuu deemi kee .. <EOS>
	BLEU score	57.32
CNN Model	Translation	ka'iitii siree kee baadhuu deemi . <eos>
	BLEU score	83.98

Translating the same sentence	Input	If I bear witness of myself my testimony is not true.
	Reference	Ani waa'ee ofii koo yoon dhugaa ba'e dhugaan koo dhugaa hin ta'u.
STM Model	Translation	Waa'ee ofii yoon gale ba'e koo dhugaa hin bareedu qabu.
	BLEU score	41.27
RNN Model	Translation	ani waa'ee ofii koo yoon koo isa hin . <EOS>
	BLEU score	39.20
CNN Model	Translation	Ani waa'ee ofii dhugaa yoon ba'e dhuga hin beeku.<eos>
	BLEU score	47.94

Translating the same sentence	Input	Which is born of the flesh is flesh.
	Reference	Wanti foon irraa dhalate fooni.
STM Model	Translation	Namni foon irraa dhalate foon nyaate.
	BLEU score	62.32
RNN Model	Translation	wanti foon irraa dhalate irraa . <EOS>
	BLEU score	66.87
CNN Model	Translation	Wanti foon dhalate irraa .<eos>
	BLEU score	73.74

Table 5.2 shows some sample outputs of these three systems by selecting some interesting results of translation. In Table 5.2a the translation is done from Afan Oromo to English, while Table 5.2b shows the output of translation from English to Afan Oromo. In both case when translating input sentence by system to system output sentence, we use human translated reference sentence to get BLEU score of the translation.

5.4 Discussion on the result of the study

In order to accomplish the objective of this thesis work, we mainly focused on design and implementation of machine translation between English and Afan Oromo by using CNN approach in which the act of translation is from either of one to other language. The reason of its implementation is to improve translation accuracy, which can be measured in BLEU score metrics, over previous approach of machine translation between English and Afan Oromo machine translation.

In order to confirm that using CNN approach worked better on translation between the language pair, we have implemented STM based baseline system to compare our CNN approach with this baseline. Also, in order to confirm that our CNN based model even better in currently available approaches, we have implemented RNN based model to compare with our CNN based model with expectation to get better result in both BLEU score and time efficiency.

According to the result of our experiments, our baseline system which is based on STM model has shown interesting translation accuracy of 20.51 BLEU score for English to Afan Oromo translation and 19.86 BLEU score for Afan Oromo to English translation which much interesting with respect to our small datasets. However, we have even got great improvement by using our RNN based system which has shown 1.81 BLEU score improvement on translation from Afan Oromo to English translation and 2.28 BLEU score for translation from English to Afan Oromo compared with our baseline system.

Although our RNN based model shows interesting BLEU score improvement, our target was to apply convolutional neural network on machine translation between the language pair. Therefore, the results of our experiments show that our CNN based system has shown better

BLEU score than both of our baseline and RNN based systems. Our CNN based system outperformed the baseline system (STM model) by 3.32 BLEU score on translation from Afan Oromo to English and by 3.86 BLEU score on translation from English to Afan Oromo. Also when we compared our RNN based model with our CNN based model, our CNN based model shows better BLEU score and time efficiency (faster) than our RNN based model on translation between these language pair in both direction of translation.

According to the result of our experiment, one of interesting features of neural network based systems is that the neural network based systems can stores longer contextual relevancies of words found in longer sentence of our datasets. Therefore, the comparison of BLEU score metrics for some range of word length shows that all the systems have shown good translation result for shorter sentences than translating longer sentences. This feature has been described by Figure 5.2 which compares all three models on translation of sentences of different lengths.

Therefore, as we can see from Figure 5.2, when the length of word in source sentence is increased, their BLEU score level is decreased with increase of sentence length for all models. However, the results of the experiment confirmed that RNN based model works better than baseline system with increase of sentence length, while CNN based model able to store longer context in sentences better than both RNN based model and baseline system. Therefore, when longer sentence is the candidate, CNN based model is the better choice than both systems.

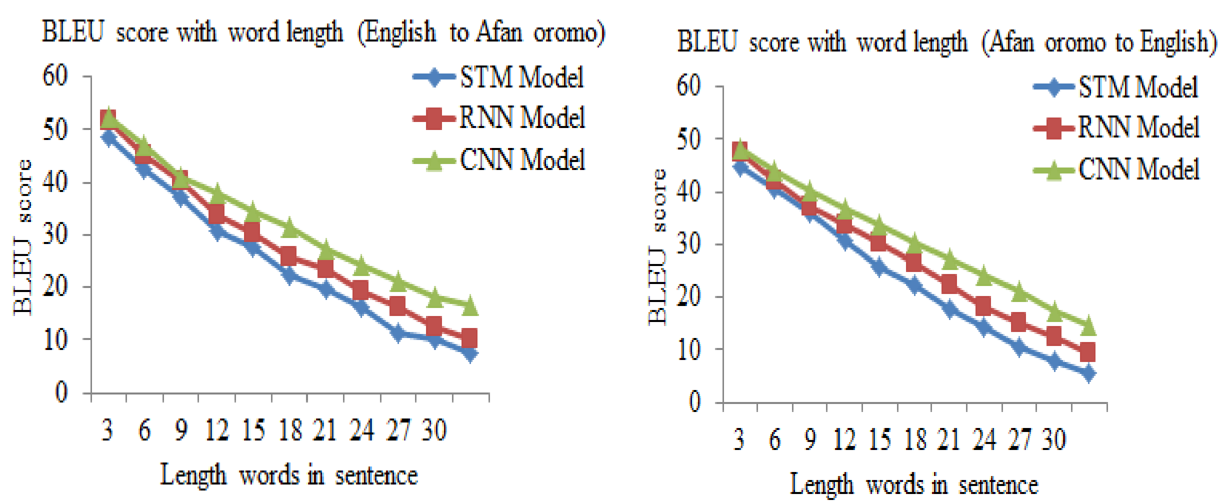


Figure 5.2 Effect of increasing length of sentence on BLEU score for all systems

Moreover, the other interesting feature seen on using CNN based model is that the CNN based model keeps contextual relatedness in which related words are used for translation rather than only using exact target word for translation. In that case BLEU score can be affected as it is dependent on exact matching word of reference sentence rather than using related dictionary words. Therefore, the result of our CNN based model has been affected by this property of BLEU score metric and would have better than what is reported here. This feature of CNN based model is described by sample output given in Table 5.3(top).

Table 5.3 Sample output to show drawback of BLEU score

Cases in which BLEU under scores	When related words are used(dictionaries or contracted phrases)	Input-1	Dhiifama maal jette?
		Reference	**Excuse me** what did you say?
		Translation	**I'm sorry** what did you say? <eos>
		BLEU score	55.78
		Input-2	Boodan deebisee bilbila.
		Reference	**I'll** call back later.
		Translation	**I will** call back later. <eos>
		BLEU score	65.64
	Evaluation of shorter sentence (input-1) compared with longer sentence (input-2) both are missing only one word	Input-1	Maal naan jetta?
		Reference	What do you **mean**?
		Translation	What do you **say**? <eos>
		BLEU score	52.54
		Input-2	Yoom akka deebi'an dhageesseettaa?
		Reference	**Have** you heard when they will return?
		Translation	**It** you heard when they will return? <eos>
		BLEU score	80.81

Lastly, although BLEU score metrics is the best technique to measure accuracy of machine translation, it shows some discrimination on correctly evaluating shorter sentences compared with longer sentences of having similar number of missing words in which BLEU score of shorter sentence is reduced. However, the method of our evaluation in which we have used 20% of total dataset can minimize the effect of this problem. This property of BLEU score metrics is shown in Table 5.3(bottom) in which translation of different sentences with different length with similar number missing words are compared against each other.

Therefore, from Table 5.3(top) we can see that using of dictionaries has affected the BLEU score by comparing system's translation against reference sentence in which both sentences have similar meaning in our daily conversation. In the bottom of the table we can see that BLEU score evaluation on shorter sentence is affected by comparing against evaluation of longer sentence that has equal number of missing words like that of shorter sentence, but having different word length in which input1 versus input2 and input3 versus input4 input sentences are compared against each other.

Finally, before summarizing the discussions of the results of our study, we need to answer the initial questions of the study. In our study process, we were working to answer two main research questions. The first question is to identify which method can give better performance between the method using RNN and CNN approaches. According to the results of our experiments, we have seen that the CNN approach gave better performance in the BLEU score. From Table 5.1 we can see that the CNN approach has shown better BLEU score than RNN approach on both directions of translation.

Here, the CNN approach has shown 1.60 BLEU score improvement than the RNN approach on translation from English to Afan Oromo, while the CNN approach has shown 1.50 BLEU scores than the RNN approach. Also, the CNN approach is faster on training by comparing against training time of RNN based approach. So, the CNN approach is faster by 1hour and 20 minutes than the RNN approach on our training datasets at 14 epochs on translation from English to Afan Oromo, while the CNN approach is faster by 1 hour and 30 minutes on Afan Oromo to English translation. Therefore, these results show that the CNN approach is better than the RNN approach on these language pairs based on the size of our datasets.

The second research question was to confirm whether we can improve machine translation between these language pairs by using the CNN approach when compared against SMT used so far. The previous studies between these language pairs did not use a neural network approach. The study by [12] used the SMT approach and worked on the varying lengths of sentences. The author used 90% datasets for training and 10% for the testing set for evaluation. The study by [43] used the Hybrid method for machine translation between these language pairs. In his study, the author used only 100 sentences for testing. Finally, the other study on machine translation between these languages is the study by [44] which is focused on identification effective alignment method for SMT between these languages pairs. In their study, they did different experiments based on different alignment methods and finally, they have reported that they got the best result on phrase-based dataset alignment with the using phrase length between 5 up to 16 sizes. Therefore, all these works have used methods that use statistics and the study by [12] testing has tried to generalize the performance testing based on technic of evaluation and by defining the different lengths of the sentence on evaluation.

Therefore, here in our study, we have used the SMT approach as a baseline system by testing on our datasets. To test our system, we have the currently active testing method which is 80% of the dataset for training and 20% for the testing based type of data classification method. Also, to generalize our performance evaluations, we have included all the sentences of our dataset which includes a sentence of having at a maximum of 55 words. Hence, the report of our work has used generalized evaluation methods. Therefore, by using this type of standardized evaluation technique, we have tested our systems in environments. Therefore, the results of our experiment show that the CNN approach has shown 3.9 BLEU score improvements on translation from English to Afan Oromo, and it has shown 3.3 BLEU score improvements on translation from Afan Oromo to English by comparing against SMT approach.

5.6 Summary

According to the result of the experiments, the CNN based model shows interesting improvement of translation accuracy in BLEU score over baseline and other competitive current active approach. More importantly the convolutional based architecture shows great result in case of translating longer sentence than both RNN and STM based models. Also, the CNN approach is faster by comparing against RNN approach in both direction of translation.

Generally, the result of our experiments have shown that our CNN based model has given great translation accuracy improvements on both direction of translation between the language pair, while these improvements are seen in terms BLEU score when comparing against baseline and RNN approach, time efficiency by comparing against RNN approach and driving better contextual relevance in sentences of the languages by comparing against both baseline and RNN approach.

Chapter Six

6. Conclusion and Recommendation

This chapter discusses the conclusion driven from this research work and the recommendation for any person or organization that interested to work on the machine translation between English and Afan Oromo language pair (or any other language pair) or related task.

6.1 Conclusion

The main goal of our study is to design and implement automated bidirectional machine translation between English and Afan Oromo language pair by using convolutional neural network algorithm in order to solve the problem of depending only on human translation. Therefore this thesis work is done on scarce resource language pair with a data of consisting 5,550 parallel corpuses collected from diversified data resources such as religion issues, publicly available governmental data including Ethiopian and Oromia regional constitutions, different conversational books, medical issues, Ethiopian revenue and customs authority data and online available conversational phrases.

As the objective of the work is to improve the machine translation between the language pair by implementing bi-directional convolutional neural network based architecture, the system based on statistical model is implemented as the baseline of our CNN based system to compare the BLEU score of our CNN based model with BLEU score of STM based model. We have also implemented RNN based model as we need to apply on our dataset to compare its BLEU score with the BLEU score of our CNN based model to use our RNN model as competitive model of CNN based model, because of that RNN based is also interesting approach for machine translation.

Therefore, results of our experiments have shown that even our RNN based model has shown better BLEU score in both directions of translations, which is 2.28 improvements for translation from English to Afan Oromo and 1.81 improvements for Afan Oromo to English translations, than baseline system. Also, when we compare our CNN based system with RNN based system;

the CNN based system shows 1.58 BLEU score improvements on translation from English to Afan Oromo, while it has shown 1.51 BLEU score improvements on the translation from Afan Oromo to English. Therefore, the convolutional neural network based architecture has shown interesting improvement both in translation quality which has been measured by BLEU score metrics, and uses less training time in comparison with our RNN based model.

Generally, when we compare both our RNN and CNN based models, they have shown better BLEU score than baseline (STM) system. Other than BLEU score improvement, both RNN based model and CNN based model are better on generating contextual relevance by using different word dictionaries for translation and by working better than baseline system on translation of longer sentences. Hence, CNN based model worked better than other models on machine translation task between the language pair by evaluating with BLEU score. Also CNN based model is faster in training on our datasets than our competitive (RNN based) system RNN based model.

Finally, as other question of our experiment is to see which direction of translation shows better BLEU score improvement, we have compared both direction of translation by reversing the direction of translation for datasets. Therefore, the results of the experiment shows that the translation from English to Afan Oromo by our CNN model has shown better BLEU score which is 3.86 than Afan Oromo to English language translation which is 3.32 BLEU score over the result of our baseline system.

Lastly, when we compare time efficiency of both direction of translation for CNN based model, it uses higher time efficiency (shorter time) on translation from Afan Oromo to English than Afan Oromo to English translation. The reason for taking longer time for translation from English to Afan Oromo is that English has used higher number of vocabulary which must pass through encoder during training time of English to Afan Oromo translation.

6.2 Recommendation

As the objective of our study is to implement machine translation only at level of text to text translation, we recommend that speech to speech machine translation between these language pair will be studied in future work.

Also we recommend that the future work will include a higher number of dataset to further increase the translation quality and also test the systems on GPU based computer to further minimize the required training time of the systems.

Additionally, the prepared data set can be used for other tasks like sentence classification, part of language tagging or similar activities.

Lastly, we recommend that the implemented system can be used for the study of machine translation between other language pairs.

References

[1] Y. Bar-hillel, “The Present Status of Automatic Translation of Languages •,” vol. 1, 1960.

[2] Y. Wu *et al.*, “Google ’ s Neural Machine Translation System : Bridging the Gap between Human and Machine Translation,” pp. 1–23.

[3] “MACHINE TRANSLATION IN EUROPE,” no. 5, pp. 28–33, 1964.

[4] P. F. Brown *et al.*, “Statistical approach to machine translation,” vol. 16, no. 2, pp. 79–85, 1990.

[5] L. Ahrenberg, “Comparing Machine Translation and Human Translation : A Case Study,” 2014.

[6] K. Puchała-ladzińska, “MACHINE TRANSLATION : A THREAT OR AN OPPORTUNITY FOR HUMAN TRANSLATORS ?,” no. January 2015, pp. 89–98, 2016.

[7] H. Li, A. C. Graesser, and Z. Cai, “Comparison of Google Translation with Human Translation,” pp. 190–195, 2014.

[8] P. E. Brown, S. A. Della Pietra, V. J. Della Pietra, and R. L. Mercer, “The Mathematics of Statistical Machine Translation : Parameter Estimation,” vol. 10598, 1993.

[9] I. Sutskever, “Sequence to Sequence Learning with Neural Networks,” pp. 1–9.

[10] D. Bahdanau, K. Cho, and Y. Bengio, “N EURAL M ACHINE T RANSLATION,” pp. 1–15, 2015.

[11] M. D. Okpor, “Machine Translation Approaches : Issues and Challenges,” vol. 11, no. 5, pp. 159–165, 2014.

[12] A. Eisele, “English – Afaan Oromoo Machine Translation: An Experiment Using a Statistical Approach.”

[13] Z. Zhu, “Evaluating Neural Machine Translation in English-Japanese Task,” no. October, pp. 61–68, 2015.

[14] Y. Ren, N. Kaji, N. Yoshinaga, M. Toyoda, and M. Kitsuregawa, “Sentiment Classification in Resource-Scarce Languages by using Label Propagation,” pp. 420–429, 2011.

[15] I. Bedane, “The Origin of Afaan Oromo : Mother Language,” vol. 15, no. 12, 2015.

[16] “THE JOURNAL OF OROMO STUDIES.”

[17] M. I. Aguilar *et al.*, “Oromo Studies,” *J. Oromo Stud. III Number 1/2 Winter/Summer 1996 ISBN 1070-2202*, vol. III, pp. 1881–1900, 1996.

[18] D. Tesfaye, “Designing a Rule Based Stemmer for Afaan Oromo Text,” no. 1, pp. 1–11.

[19] W. Tegegne, “The Development of Written Afan Oromo and the Appropriateness of Qubee , Latin Script , for Afan Oromo Writing,” *Hist. Res. Lett.*, vol. 28, no. 1976, pp. 8–14, 2016.

[20] O. Writing, “Onesimos Nasib ’ s Pioneering Contributions to,” vol. 1989, no. 1, pp. 36–59, 1995.

[21] E. Note, M. Bulcha, B. Deressa, T. Gamta, and B. Reviews, “Oromo Studies,” vol. I, no. 1, 1993.

[22] C. Espa, “Statistical Machine Translation A practical tutorial,” 2010.

[23] R. Sennrich, “Today ’ s Lecture so far we discussed RNNs as encoder and decoder we discussed some architecture variants : Machine Translation 10 : Advanced Neural Machine Translation Architectures Rico Sennrich today some important components of neural MT architectures,” 2018.

[24] E. Greenstein and D. Penner, “Japanese-to-English Machine Translation Using Recurrent Neural Networks,” pp. 1–7.

[25] M. R. Costa-juss and A. R. Fonollosa, “STUDY AND COMPARISON OF RULE-BASED AND STATISTICAL CATALAN-SPANISH MACHINE TRANSLATION SYSTEMS ´ s Mireia Farr u ˜ o , Jos ´ Jos ´ e B . Mari n,” vol. 31, pp. 245–270, 2012.

[26] M. A. Ch, “Theoretical Overview of Machine translation 2 History of Machine Translation,” pp. 160–169.

[27] D. Stein and D. Stein, “Machine translation : Past , present and future,” pp. 5–17.

[28] A. Hurskainen, “Rule-based Machine Translation from English to Finnish,” vol. 2, pp. 323–329, 2017.

[29] S. Sreelekha, “Statistical Vs Rule Based Machine Translation ; A Case Study on Indian Language Perspective.”

[30] G. Hanneman, V. Ambati, J. H. Clark, A. Parlikar, and A. Lavie, “An Improved Statistical Transfer System for French – English Machine Translation,” no. March, pp. 140–144, 2009.

[31] H. Yulian, “Statistical machine translation based on translation rules,” vol. 6, no. 7, pp. 1628–1635, 2014.

[32] V. Machines and H. Karlbom, “Hybrid Machine Translation Choosing the best translation with Support,” no. September, 2016.

[33] A. Lamb and M. Xie, “Convolutional Encoders for Neural Machine Translation,” pp. 1–8.

[34] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations ofwords and phrases and their compositionality,” *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2013.

[35] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” 2015.

[36] R. Shu and H. Nakayama, “Improving Beam Search by Removing Monotonic Constraint for Neural Machine Translation,” vol. 1, no. 6, p. 6.

[37] M. Post and D. Vilar, “Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation,” no. Table 1, pp. 1314–1324, 2018.

[38] G. Neubig, “Neural Machine Translation and Sequence-to-sequence Models : A Tutorial,” pp. 1–65.

[39] Y. Liu, L. Ji, R. Huang, T. Ming, and C. Gao, “An attention-gated convolutional neural network for sentence classification,” pp. 1–19.

[40] K. Papineni, S. Roukos, T. Ward, and W. Zhu, “B LEU : a Method for Automatic Evaluation of Machine Translation,” no. July, pp. 311–318, 2002.

[41] L. Kaplow, “Issn 1045-6333,” 2005.

[42] J. Gehring and Y. N. Dauphin, “Convolutional Sequence to Sequence Learning,” 2016.

[43] H. Approach and J. Daba, “Addis Ababa University School of Graduate Studies College

of Natural Sciences Department of Computer Science,” no. November, 2013.

[44] M. Meshesha and Y. Solomon, “English-Afaan Oromo Statistical Machine Translation,” no. 9, pp. 26–31, 2018.

[45] Y. Zhao, J. Zhang, Z. He, C. Zong, and H. Wu, “Addressing Troublesome Words in Neural Machine Translation,” pp. 391–400, 2018.

[46] Y. Kim, “Convolutional Neural Networks for Sentence Classification,” pp. 1746–1751, 2014.

[47] J. Chung, “Gated Recurrent Neural Networks on Sequence Modeling arXiv : 1412 . 3555v1 [cs . NE] 11 Dec 2014,” pp. 1–9.

[48] C. E. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation Functions : Comparison of Trends in Practice and Research for Deep Learning,” pp. 1–20.

[49] A. F. M. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” no. 1, pp. 2–8.

[50] M. Luong and C. D. Manning, “Effective Approaches to Attention-based Neural Machine Translation,” 2015.

[51] S. Kanai, “Sigsoftmax : Reanalysis of the Softmax Bottleneck,” no. NeurIPS, 2018.

[52] E. Simon and R. E. N. Au, “Calculation of phrase probabilities for Statistical Machine Translation by using belief functions,” no. December, pp. 1101–1110, 2012.

Appendices

Appendix I: Sample of parallel corpus from collected dataset

Oh that we had died in the land of Egypt!	Nuyis biyya Gibxii keessatti utuu dhumneerra ta'ee!
He sees the light of the world.	Ifa addunyaa kanaa ni arga.
Abba, Father, all things are possible with You.	Aabbaa,wanti hundi siif ni danda'ama.
When did your brother come?	Obboleessi kee yoom dhufa?
I became hungry.	Nan beela'e.
You could do it with your eyes closed.	Ija dunuunfattee hojjetta.
Identify yourself immediately.	Dafii of beeksisi.
Who is your country is prime minister?	Mummichi ministeera biyya keesani eenyuu?
The members of the Committee shall be elected and shall serve in their personal capacity.	Miseensonni Koree kanaa kan filatamanii fi mataa isaaniitiin kan tajaajilan ta'u.
Good morning teacher?	Attami bultan barsiisa?
I feel down and my leg is dislocated	Kufee miilli na meelate.
Actually, it isn't that long. It's almost a week.	Baayyee narra hin turre, Yoo baayyate torbee ta'a
The Caffee shall adopt rules and procedures regarding its working procedures and legislative process.	Caffeen haala hojiisaafi seerri itti tumamu ilaalchisee labsii baasuu ni danda'a.
I don't want it extracted.	Akka buqa'u hin barbaadu.
We were to learn English for a year.	Afaan Inglizii waggaa tokkoof barachuuf turre.
In the mount of the Lord it will be provided.	Inni tulluu isa Waaqayyo irratti waa argamsiisudha.
We leave now.	Nu amma deemna.
Study problems encountered regarding implementation of rural land administration.	Raawwii bulchiinsaa itti fayyadama lafa baadiyaa irratti rakkoowwan qunnaman qorata.

The girl showed her mother her new dress.	Intalattiin wandaboo ishee isa haaraa haadha isheetti agarsiiste.
She competes with me.	Isheen naan morkatti.
Are there not twelve hours in the day?	Guyyaan tokko sa aatii kudha lamaaf ifa qaba mitii?
And the entire congregation lifted up their voice and cried.	Halkan sana guutummaan waldaa sagalee isaa ol fudhatee iyyee in boo'e.
After the receiver picks up the phone	Bilbilamaafiin erga bilbila kaassee booda
He was about to go.	Deemuuf jedha.
Submits lands used for different services to administrative councils that found at different levels.	Lafa tajaajila adda addaatiif oolu mana maree bulchiinsa sadarkaa addaa addaan jiraniif ni dhiyeessa.
Would you mind repeating what you said?	Waan amma dubbatte utuu irra naf deebite.
Why are they doing that which is not lawful on the Sabbaths?	Isaan wanta guyyaa Sanbataatti seeraan dhorkame maaliif hojjetu?
Organize Police and Security forces for maintenance of peace and security of the Region.	Nagaafi tasgabbii Naannichaa eeguuf humna poolisiifi nageenyaa ni gurmeessa.
He did not believe them.	Waan isaan dubbatan hin amanne.
The sheep did not hear them.	Hoolonni garuu isaan hin dhageenye.
Do you mind leaving me alone?	Kophaa ko taa'un barbaada utuu na bira deemte.
Show me by waving your hand.	Harka kee raasuun na argisiisi.
All things are possible to the one who believes.	Nama amantii qabuuf wanti hundi ni danda'ama.
Cause the book account of the organization audited.	herrega dhaabbatichaa odiitii ni taasisa.
May I have your name?	Maqaa kee natti himtaa?
Then Daniel said to the king, O king, live forever.	Daani'el immoo deebisee mootichaan, Yaa mootii, bara baraan jiraadhu jedhe..

Do you have vacancies?	Bakka duwwaa qabdaa?
Is there a doctor here?	Doktorri naannoo kana jiraa?
They breathe air.	Isaan hafuura fudhatu.
Would you call back then after 3 hrs?	Sa'a sadii booda bilbiluu dandeessuu?
And Moses cried to the Lord because of the frogs which He had brought against Pharaoh.	Museen waa'ee fattee isa inni mooticha irratti fide sanaaf gara Waaqayyootti in iyye.
But when you hear of wars and rumors of wars, do not be troubled.	Kana malees, waraanaa fi oduu waraanaa yommuu dhageessan, akka hin sodaanne of eeggadhaa.
The committee shall consist of eighteen members.	Koreen kun miseensota 18 kan qabaatu ta'a.
He gave them bread to eat that came down from heaven.	Akka isaan nyaataniif samii irraa nyaata isaaniif kenne.
Next dial the number.	Itti aansuun lakkoobsa feene bilbilla.
Are you menstruating?	Lagu qabdaa?
Deliberate upon and approve plans and programs with regard to economic development, social services and public administration of the District.	Karooraawwaniifi sagantaalee misooma dinagdee, tajaajiloota hawaasummaafi hojii bulchiinsaa xiinxalee ni raggaasisa.
I will leave tomorrow morning.	Bor ganaman dema
Yes, it is my sister's Wedding.	Eeyyee, cidha obboleettii kootiiti.
And Noah went out.	Nohis gad ba'e.
My favorite singer is Ali Birra.	Weellisaan ani jalladhu Ali birra dha .
Any land holder has the right to get land holding certificate on his holding.	Abbaan qabiyyee kamiyyuu lafa seeraan kennameef irratti waraqaa ragaa argachuuf mirga qaba.
Giving birth to a baby.	Daa'iima yoo argatu.
And no one puts new wine into old wineskins.	Kana malees, eenyu illee daadhii wayinii haaraa qalqalloo moofaatti hin naqu.
Let's go hit some bars to celebrate	Mana bashannanaa haa deemnu malee

And when he had gone a little farther thence, he saw James the son of Zebedee, and John his brother, who also were in the ship mending their nets.	Erga xinnoo adeemee booda, Yaaqoob ilma Zabdewosii fi obboleessi isaa Yohaannis bidiruu isaanii keessatti utuu kiyyoowwan isaanii suphanii isaan arge.
And the seven good ears are seven years.	mataan midhaanii tolan torbanis waggootuma torban sana.
Our school has taken only brilliant student in this year.	manni barumsaa keenya waggaa kana keessa barattoota cimoo qofa fudhate.
And when Jesus arrived, He saw a great multitude and He was moved with compassion toward them.	Yeroo bidiruu irraa bu'us, namoota hedduu argee baay'ee isaaniif gadde.
Why does this generation seek a sign?	Dhaloonni kun maaliif mallattoo barbaada?
Why are you doing this?	Maaliif kana gootu?
What did Moses command you?	Museen maal isin ajaje?
Gives necessary supports regarding preparation of land use plan.	Karoora itti fayyadama lafaa qopheessuu ilaalchisee deeggarsa barbaachisaa ta’e ni kenna.
Fill four water jars with water and pour on the burnt sacrifice and on the wood.	Bishaan okkotee afur guutaatii qalmicha gubamuu fi qoraanicha irratti dhangalaasaa!
Everyone has the right to seek and to enjoy in other countries asylum from persecution.	Namni kamiyyuu miidhaa duraa baqachuu fi biyyoota biroo keessatti irkataa tahee jiraachuuf gaafachuuf mirga ni qaba.
I work for Ethiopia television as a reporter.	Ani gabaasaa oduu Televiishinii Itiyoophiyaati.
He advised me repeatedly for my wrong deeds.	balleessaa koof irra deddeebi'ee na gorsee ture.
Now Nahaman is commander of the army of the king of Syria.	Baras Na'amaan kan jedhamu tokko abbaa duulaa mootii Sooriyaa ture.
They shall condemn Him to death, and shall deliver Him up to the Gentiles.	Isaanis du'a isatti murteessu, namoota Yihudoota hin taanettis dabarsanii isa kennu.
I am God, the God of your fathers.	Ani Waaqayyo, Waaqayyo isa kan abbaa keetii ti.

What about sometime next week?	Torbee dhufu guyyaa ta'e?
All that the Lord has spoken we will do.	Wanta Waaqayyo jedhe hundumaa in goona.
The Constitutional Inquiry Council shall have powers to investigate constitutional disputes.	Gumiin Calaltuu Dhimmoota Heera Naannichaa, aangoo dhimmoota Heerichaa falmisiisaa ta’an calaluu ni qabaata.
And Isaac prayed to the Lord for his wife because she was barren.	Yisihaaq waa’ee haadha manaa isaa ishee dhabduu turteef Waaqayyoon kadhate.
And Jonah prayed to the Lord his God out of the fish’s belly.	Yoonaas immoo garaa qurxummichaa keessaa Waaqayyo isaa goofticha in kadhate.
I fear the Lord, the God of heaven, Who has made the sea and the dry land.	Waaqayyoon isa gooftaa ta'ee waaqa irra jiraatu, isa galaanaa fi lafa uume nan waaqeffadha.
And the seed should sprout and grow.	sanyichis biqilee guddata.
Meeting of the District Council shall be public.	Walga’iin Mana Maree Aanaa ifaan ni adeemsifama.
Cry aloud with a great voice, for he is a god.	sagalee guddisaatii guddisaatii iyyaa, inni waaqayyolii keessaa tokko waan ta'eef!
The three colors shall be set horizontally in equal dimension.	Bifti sadanuu walqixa ta’anii dalgaan taa’u.
What will you have?	Maal siif ha dhufu?
Because all the shops are closed today.	Sababni isaa suuqiin hundi cufaadha.
Wait a moment.	Yero gabaabduuf na eegi.
The new council begins its session within fifteen days after the expiry of the tenure of the previous Council.	Barri hojii Mana Maree duraanii akka xumurametti, Mana Maree haarawaan guyyaa kudha shan keessatti hojiisaa ni jalqaba.
Rejoice and be filled with joy, for great is your reward in heaven.	Badhaasni isin samii irratti argattan guddaa waan ta'eef, gammadaa, ililchaas.
Nullifying the authority of the Word of God by your tradition which you have passed down.	Kanaaf, aadaa keessan isa dhalootatti dabarsitaniin dubbii Waaqayyoo gatii dhabsiistu.

Your hands are very cold.	Harkikee baay'ee diilallaa'adha.
Discharge such other functions assigned to him by the President.	Hojiiwwan biraa Pireezidaatichaan kennamaniif ni raawwata.
We have one at 500 dollars a day on the 5th floor.	Darbii shanaaffaa irraa doolaara 500 tokko qofatu jira.
Are calling for Eleni?	Eleniif bilbilaa jirtaa?
For now I will stretch out My hand that I may strike you and your people with plagues, and you shall be cut off from the earth.	Egaa akkan harka koo ol fudhadheen si, saba kees utuun dha'eera ta'ee silaa yowwana ati lafa irraa baddeetta.
I don't agree.	Ani tole hin jedhuu.
They feared the people, because everyone held that John was indeed a prophet.	Namoonni hundi Yohaannisiin akka raajiitti waan ilaalaniif namoota sana sodaatan.
I am sitting on a chair.	Barcuma irran taa'utti jira.
My fever is gone, but I still have a cough.	Dhaqna gubaan na dhiiseera, garuu qufaatu narratu hin citne.
The tops of the mountains were seen.	Roggeen tullootaas ni mul'atan.
What is your business?	Hojiin kee maali?
And everyone who lives and believes in me shall not die forever.	Lubbuudhaan kan jiruu fi anatti kan amanu hundis matumaa hin du'u.
Keep your heart with all diligence.	yaadakee ba'eessa godhii eeggedhu.
Is not the whole land before you?	Kunoo, lafichi guutummaan isaa si dura jira mitii ree?

Appendix II: STM based model: Language modeling and Translation modeling

The statistical machine translation depends on two main components. The first is the language model which describes co-occurrence relationship between words the target sentence. Therefore, this information is required to arrange the words of target sentence to form correct sentence from predicted words of statistical machine translation. Therefore, this can have the following form in the case of English is target sentence.

```
! ||| the kingdom of god ! ||| 1 0.482353 0.00740741 1.2674e-09 2.718 ||| 0-4 ||| 2 270 2
! ||| the lord ! ||| 1 0.482353 0.0037037 0.00012272 2.718 ||| 0-2 ||| 1 270 1
! ||| the lord . ||| 0.0769231 0.0327264 0.0037037 0.000398093 2.718 ||| 0-2 ||| 13 270 1
! ||| the midst of the sea . ||| 1 0.0327264 0.0037037 2.92393e-10 2.718 ||| 0-5 ||| 1 270 1
! ||| the mountain . ||| 0.166667 0.0327264 0.0037037 2.59649e-05 2.718 ||| 0-2 ||| 6 270 1
! ||| the people round about . ||| 1 0.0327264 0.0037037 2.4025e-11 2.718 ||| 0-4 ||| 1 270 1
! ||| the sea . ||| 0.125 0.0327264 0.0037037 0.000129816 2.718 ||| 0-2 ||| 8 270 1
! ||| the wood . ||| 0.333333 0.0327264 0.0037037 4.32749e-05 2.718 ||| 0-2 ||| 3 270 1
! ||| them . ||| 0.0434783 0.0327264 0.0037037 0.00269695 2.718 ||| 0-1 ||| 23 270 1
! ||| themselves . ||| 1 0.0327264 0.0037037 0.00027664 2.718 ||| 0-1 ||| 1 270 1
! ||| this ? ||| 0.333333 0.0024361 0.0037037 3.74367e-05 2.718 ||| 0-1 ||| 3 270 1
! ||| to me . ||| 0.5 0.0327264 0.0037037 0.000127004 2.718 ||| 0-2 ||| 2 270 1
! ||| to the mountain . ||| 0.166667 0.0327264 0.0037037 7.81745e-07 2.718 ||| 0-3 ||| 6 270 1
! ||| today . ||| 1 0.0327264 0.00740741 0.0003458 2.718 ||| 0-1 ||| 2 270 2
```

Another main component of STM is Translation model that is used to know how individual word is translated from source sentence into target sentence depending on probability of matching evaluation. Therefore, it is core of statistical evaluation as a highest probability is selected to choose the word of translation. Thus, from the following figures we can see the probability of translation from source to target sentences in which left side is from English to Afan Oromo translation model and the right hand side is from Afan Oromo English translation model.

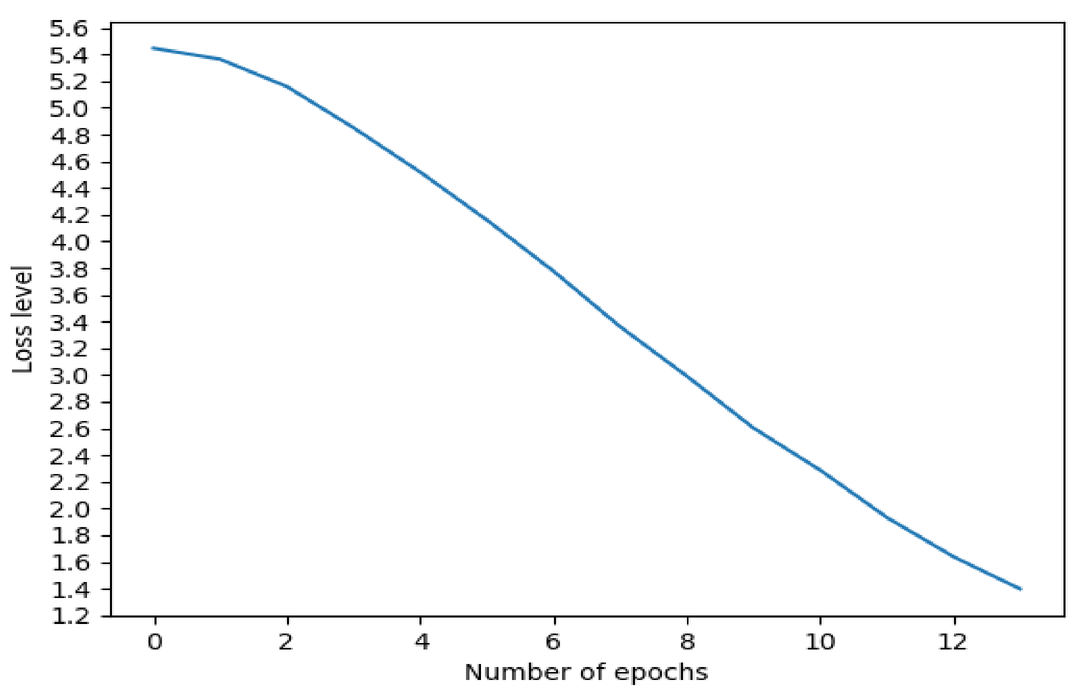
```
number lakkoobsa 0.5833333
dialed lakkoobsa 0.0833333
telephone lakkoobsa 0.1666667
awfully  lakkoobsa 0.0833333
dial  lakkoobsa 0.0833333
protection  eegumsi 0.3333333
and  eegumsi 0.3333333
NULL  eegumsi 0.3333333
remained  markabicha 0.2000000
go  markabicha 0.2000000
alive  markabicha 0.2000000
ark  markabicha 0.4000000
```

```
ibricha  am 0.0058824
ibricha  a 0.0017065
ibricha  hebrew 0.2000000
galee  house 0.0243902
barataan  his 0.0034483
barataan  you 0.0008889
barataan  disciple 0.1666667
barataan  are 0.0033784
daraartuu  is 0.0014749
daraartuu  derartu 0.2500000
qulqulleessi  board 0.0454545
qulqulleessi  the 0.0003683
qulqulleessi  erase 0.5000000
```

Appendix III: Neural Network Training based on loss minimization

```
Reading data...
________________________________________________________________
=======================================================
Read 5550 sentence pairs.
=======================================================
Creating language dictionaries.....
________________________________________________________
Vocabulary in Afan Oromoo:  4221
Vocabulary in English:  8527
=======================================================
['you are a youth .', 'ati ijoollee dha .']
Training time: 0:22:9 ==>Epoch: 1    ==>Loss: 5.45
Training time: 0:45:46 ==>Epoch: 2    ==>Loss: 5.36
Training time: 1:10:0 ==>Epoch: 3    ==>Loss: 5.16
Training time: 1:34:19 ==>Epoch: 4    ==>Loss: 4.85
Training time: 1:58:45 ==>Epoch: 5    ==>Loss: 4.52
Training time: 2:23:32 ==>Epoch: 6    ==>Loss: 4.16
Training time: 2:48:21 ==>Epoch: 7    ==>Loss: 3.78
Training time: 3:13:14 ==>Epoch: 8    ==>Loss: 3.36
Training time: 3:38:20 ==>Epoch: 9    ==>Loss: 2.99
Training time: 4:3:23 ==>Epoch: 10    ==>Loss: 2.60
Training time: 4:28:31 ==>Epoch: 11    ==>Loss: 2.29
Training time: 4:53:47 ==>Epoch: 12    ==>Loss: 1.93
Training time: 5:19:14 ==>Epoch: 13    ==>Loss: 1.64
Training time: 5:44:42 ==>Epoch: 14    ==>Loss: 1.40

=======================training is completed==========================
Training has taken 5 hrs, 44 minutes and 42 seconds.
======================================================================
```



Appendix IV: drawback of BLEU score

From the following output, we can see that neural network can transfer the context in the sentence by using available dictionaries or related words like using 'i'm sorry' in place of using 'excuse me' which is understandable for human, but the available automatic evaluator like BLEU score metrics cannot consider this concept. So, BLEU score of such sentence can be affected.

```
=====================================================================
Source language input sentence::>>    dhiifama maal jette ?
------------------------------------------------------------------
Human translated reference sentence::>>    excuse me what did you say ?
------------------------------------------------------------------
System translated output Sentence::>>   i'm sorry what did you say ? <EOS>
------------------------------------------------------------------
Bleu Score Value: 55.77503997480662
=====================================================================
```

Also, from the following two sample output we can understand that the first sentence is shorter and the second sentence is longer. Although the all required words are correctly translated, except not cutting end of string indicator which is similar for both outs, the BLEU score of the shorter sentence not equal to the longer sentence so that the BLEU score of the shorter sentence is affected by BLEU scoring metrics.

```
=====================================================================
Source language input sentence::>>    nu kana ni fudhanna
------------------------------------------------------------------
Human translated reference sentence::>>    we admit it .
------------------------------------------------------------------
System translated output Sentence::>>   we admit it . <EOS>
------------------------------------------------------------------
Bleu Score Value: 65.53609623522635
=====================================================================
```

```
=====================================================================
Source language input sentence::>>    ani amma murteessu nan danda'a .
------------------------------------------------------------------
Human translated reference sentence::>>    i can decide now .
------------------------------------------------------------------
System translated output Sentence::>>   i can decide now . <EOS>
------------------------------------------------------------------
Bleu Score Value: 73.25510644096215
=====================================================================
```

Therefore, the above figures show us that the evaluation of the result of translation is even under scored by BLEU score metrics on some translation outputs, but since average result is used the total is not affected.