Bi-directional Amharic – Afaan Oromo Machine Translation Using Statistical Approach

A Thesis Presented

by

Emebet Girma

to

The Faculty of Informatics

of

St. Mary's University

In Partial Fulfillment of the Requirements
for the Degree of Master of Science

in

Computer Science

February, 2021

ACCEPTANCE

Bi-directional Amharic – Afaan Oromo Machine Translation Using Statistical Approach

By

Emebet Girma

Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

____

Internal Examiner

____

External Examiner

____

Dean, Faculty of Informatics

February 2021

DECLARATION

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been duly acknowledged.

Emebet Girma .
Full Name of Student

Signature

Addis Ababa

Ethiopia

This thesis has been submitted for examination with my approval as advisor.

Michael Melese (PhD) .
Full Name of Advisor

Signature

Addis Ababa

Ethiopia

February 2021

Acknowledgments

First of all, I would like to thank Almightily God from the bottom of my heart that he gave me the strength for this thesis to completion.

I would like to express my deep and sincere gratitude to my research Advisor, Dr. Michael Melese for giving me the opportunity to do this research topic and providing invaluable guidance, support and patience in helping me complete this thesis.

I am extremely thankful to Mr. Fitsum A. who gave his time and provided valuable information and support.

Last and most importantly, I would like to extend my heartfelt gratitude to all my family Members for their love, encouragements and moral support to complete this thesis successfully.

Table of Contents

List of Acronyms

BLEU	Bilingual Evaluation Understudy
EBMT	Example-Based Machine Translation
MT	Machine Translation
NLP	Natural Language Processing
RBMT	Rule-Based Machine Translation
SOV	Subject-Object-Verb
SMT	Statistical Machine Translation
TM	Translation Model
WER	Word Error Rate

List of Figures

List of Tables

Abstract

Machine translation (MT) is an automatic translation from one natural language to another by a computer, without human involvement. The purpose of this study is to develop a bidirectional Amharic- Afaan Oromo machine translation system using statistical machine translation.

In this thesis, to explore the effect of morpheme and word level alignment on bi-Directional Amharic-Afaan Oromo statistical machine translation. In order to conduct the study, the corpus was collected from online source such online documents include Old and new Testament of Holy bible and religious documents for both language and corpus preparation which also involves dividing the corpus for training set, tuning set and test set. A total of 14600 sentences are collected. We use 1460 for testing and 1460 for tuning purpose. For language model we used 11680 parallel sentences sentence for both Amharic and Afaan Oromo language. The experiment was conducted using statistical Machine Translation tool moses, MGIZA++ for word and morpheme alignment toolkit, morfessor were used for morphological segmentation for both Amharic and Afaan Oromo language and IRSTLM language modeling tools. Different experiments were carried out after preparing and designing the corpus and the prototype.

Experiments were conduct based on the morpheme and word level alignment and results were recorded. The experiments were taken separately. The result obtained for the unsupervised morpheme segmentation based level alignment using BLEU score has an average of 19.77 % accuracy for the Amharic to Afaan Oromo and 16.14 % for the Afaan Oromo to Amharic. For word based alignment, the result acquired from the BLEU Score was 13.84 % for Amharic to Afaan Oromo and 9.72`% for Afaan Oromo to Amharic. This result shows that morpheme level alignment translation performs better than word-level alignment translation.

Keywords: *SMT, morpheme level alignment, morfessor, Amharic. Afaan Oromo*

Chapter One

Introduction

1.1 Background

The term "communication" has been derived from the Latin "communis" that means "common" "to communicate" means "to make common" or "to make known", "to share" and includes verbal, non-verbal and electronic means of human interaction [1]. Communication is the act of transferring information from one place, person or group to another. It is a way of interchanging messages or information between two or more people, focusing on the message. Language is a way of communication (verbal or non-verbal codes). Language-communication plays an essential role to exchange or transfer information.

Ethiopia is a multi-lingual country with over 80 distinct languages. Amharic is the official working language of the Federal Democratic Republic of Ethiopia (FDRE) and it is the second most-spoken Semitic language in the world (after Arabic) [2] . Today it is probably the second largest language in Ethiopia after Oromo (Cushitic language) [2] [28]. Afaan Oromo is one of the languages of the Low land East Cushitic within the Cushitic family of the Afro-Asiatic Phylum. It is also one of the major languages spoken in Ethiopia [4].

The syntactic structure is formed by combining different word classes in sequence. The usual word order of Amharic and Afaan Oromo is Subject-Object-Verb (SOV) [24]. Both Amharic and Afaan Oromo have a complex morphology. The word-formation, for instance, involves different formations including prefixation, infixation, suffixation, and reduplication. Morphologically complex languages also tend to display a rich system of agreements between the syntactic part of a sentence like nouns, verbs, person, number, gender, fine and place. This increases the complexity of word generation [3].

Since currently there are a lot of documents available in the Amharic language. Both Amharic and Afaan Oromo language is spoken in Ethiopia, it is obvious that both Amharic and Afaan Oromo speakers need the data or documents written in Amharic or Afaan Oromo and also, they need to communicate with each other.

Thus, there is a need to develop a bi-directional machine translation system from Amharic to Afaan Oromo to translate data from one language to another.

MT has been in existence since the 1940s and has flourished in recent times due to the proliferation of the web. MT was the first computer-based application in natural language processing (NLP), and its history is old. Like computer science, the field of machine translation (MT) is old starting in the days of the Cold War. Because of globalization, tourism, commerce, governance, education, etc., the need for translation has become all-pervading, and the sheer volume of translation has made automation inevitable [4].

Machine Translation means automatic translation, It the field of Artificial Intelligence. Machine translation is a computer program which is design to translate text from one language (source language) to another language (target language) with-out the help of a human. Machine Translation aims to provide a system that translates the text of source language into the target language and translation express the same meaning as it in source language [5]. It helps the people to understand the information of unknown language without the help of a human translator.

The use of machine translation may be broken up broadly into assimilation, dissemination and communication[6]. These category attempt to translate foreign material to understand the content and translating text for publication in different languages. In addition to this, translation of emails, web resource, chat room discussions, and other document translation.

The machine translation approaches may follow a rule-based and corpus-based approach to translate from one source language to another targets language[5] [6]. In rule-based machine translation, human expert define set of rules for the translation process. The human experts specify a set of rules to describe the translation process so that an enormous amount of input from human experts is required. On the other hand, the corpus-based approach, the knowledge is automatically extracted by analyzing translation examples from a parallel corpus built by human experts. Combining the features of the two major classifications of MT systems gave birth to the Hybrid Machine Translation Approach [7]. There are four Machine Translation. In temporal order, they are rule-based machine translation (RBMT), example-based machine translation (EBMT), and statistical machine translation (SMT) and neural machine translation (NMT)[8][16].

Rule-Based Machine Translation (RBMT), also known as Knowledge-Based Machine Translation. First commercial machine translation system and based on the linguistic rule about the source and target languages basically, retrieved from (bilingual) dictionaries and grammars covering the main semantic, morphological, and syntactic regularities of each language respectively [7].

Example-based machine translation (EBMT) is trained from bilingual parallel corpora, which contain sentence pairs, in which translation is the idea of translation by analogy. An EBMT system is given a set of sentences in the source language and corresponding translations of each sentence in the target language with a point to point mapping [7].

Statistical machine translation (SMT) is generated based on statistical models whose parameters are derived from the analysis of bilingual text corpora. The initial model of SMT, based on Bayes Theorem, proposed by Brown et al. takes the view that Any sentence in one language can be translated into any other language, and the most suitable translation is the one with the highest probability assigned by the system [7]. The SMT is a part of corpus-based MT system which requires parallel corpus before undertaking translation.

The main objective of this study is to design and develop Bi-directional Amharic - Oromo machine translation system, is to translate Amharic texts into Afaan Oromo or the vice versa by using a statistical machine approach.

1.2 Statement of the problem

Afaan Oromo and Amharic language are respectively reported to be the 1st and 2nd widely spoken languages in Ethiopia [28]. Amharic is the official and working language of Amhara region of Ethiopia, Afaan Oromo also the official and working language of Oromia region

The following challenges are faced Amharic and Afaan Oromo language speaker.

- Lack of information: Amharic is working language of Federal Government of Ethiopian since currently there are a lot of material or documents are written in

Amharic language so that Afaan Oromo would not be aware the information and also the documents that written in Afaan Oromo needs translation to Amharic language.

- Communication gap: There are massive people at both nations who couldn't speak and understand Amharic/Afaan Oromo language cannot communicate with each other speaker.

Currently, the demand for translation is increasing rapidly but those demands are not fulfil using manually. Therefore, machine translation is rolling a great job by translating to satisfying the demand for translation.

However, only little works have been done on MT in Amharic and Afaan Oromo languages. Some of the studies are carried on Ge'ez -Amharic Machine [3], Afaan Oromo-English [4], English-Amharic[9], English – Afaan Oromo[10] and Geez to Amharic[11].

This also necessitates translation of documents language pair of Afaan Oromo and Amharic languages, However [12], experimented on for seven Ethiopian language pairs by using a statistical MT approach. As noted by the researcher, the negative effect of morphological complexity of the Ethio- Semitic languages on the performance of SMT. the word-level translation process is challenged by many forms of a single word, due to the morphological richness of the languages where a single word in any of the languages composed of many sub-words or morphemes.

Amharic and Afaan Oromo is a morphologically rich language since for morphologically rich languages it is not possible to cover all the words that exist in the language for translation, there is a need to experiment morpheme-based translation. At morpheme level sub-parts of words are specific.

The main aim of this study is to design and develop a bi-directional Amharic-Afaan Oromo machine translation system to solve the above-stated problem. In an attempt to solve the above-stated problem, the following research questions are formulated.

- How to develop a parallel corpus for Amharic Afaan Oromo machine translation system?

- What is the best unit to translate Amharic Afaan Oromo machine translation system in both directions?
- To what extent the Bi-directional Amharic Afaan Oromo machine translation system works?

1.3 Objective of the study

The general and specific objectives of the thesis described below.

1.3.1 General objective

The general objective of this research is to design and develop a bi-directional Amharic - Afaan Oromo machine translation system using a statistical approach.

1.3.2 Specific objective

To achieve the general objective of the study, the following specific objectives are identified:

- To review the related literature on Amharic and Afaan Oromo language and understand the state-of-the-art
- To collect and prepare Amharic- Afaan Oromo parallel corpus for machine translation
- To design an architecture of Amharic – Afaan Oromo machine translation system.
- To develop bi-directional Amharic – Afaan Oromo prototype using statistical machine translation
- To evaluate the performance of Amharic - Afaan Oromo machine translation system.
- To report the finding of the study and recommend for the upcoming research area.

1.4 Methodology of the study

The research methodology is the way which the researcher needs to conduct their researches and systematically solve the research problem [4]. In order to achieve the general and specific objective of this study, different methodologies were employed. The following subsections discuss the methodologies that are applied in this study.

1.4.1 Literature Review

In this study A detailed literature review has been done on machine translation on different language pairs. Published text documents, books, journal articles, and literature review on machine translation also reviewed to explore the principles, methods, techniques and tools employed.

Furthermore, the different algorithms used in implementing them were studied carefully and the syntactic relationship between Amharic and Afaan Oromo languages has been reviewed.

1.4.2 Research design

In order to conduct the study, we follow experimental research design to explore morphemes and words level alignment on Statistical machine translation. Different experiments are conducted for better performance of SMT. Experimental research is a scientific approach to research, where one or more independent variables are manipulated and applied to one or more dependent variables to measure their effect on the latter. It includes a hypothesis, a variable that can be manipulated by the researcher, and variables that can be measured, calculated and compared.

1.4.3 Data Collection

To perform the experiments on corpus-based statistical machine translation, a parallel corpus of Amharic and Oromo is required. Since there are not available parallel corpus and limited resource to prepare parallel corpus for Amharic and Afaan Oromo language.

Therefore we used existing and publicly available documents these documents include the Old, New Testament holy bible and religious document which include Amharic and Afaan

Oromo version they are parallel corpus and prepared parallel corpus which is suitable for SMT easily.

The Size of the corpus for the experiment is 14600, prepared from the above-mentioned source of the corpus. The data set taken from the holy bible. We used 11680 monolingual corpora for language model for Amharic and Afaan Oromo languages which is prepared from the above-mentioned source of the corpus.

1.4.4 Tools and technique

To develop statistical machine translation various tools are available. The basic tool used for accomplishing the machine translation task is Moses. It is a statistical translation system that allows the automatic training of a translation model for any language pair. It is freely available open-source software that is used for statistical machine translation and integrates different toolkits.

To develop the bi-directional Amharic-Afaan Oromo machine translation system we used most popular SMT tools such as MGIZA++ is an implementation of the IBM word-based models MGIZA ++ . MGIZA++ is a multi-threaded tool used for word and morpheme alignment, IRSTLM for building and applying statistical language model [6]. Python programming language is used as a tool for preprocessing in the Ubuntu operating system which is suitable for Moses environment. Also, unsupervised morpheme segmentation tool Morfessor 2.0 is used for morphological segmentation. The BLEU (Bilingual Evaluation Understudy), which is one of the famous evaluation methods is used for evaluation.

1.4.5 Evaluation

The evaluation of machine translation (MT) systems is a vital field of research, both for determining the effectiveness of existing MT systems and for optimizing the performance of MT systems [13].

A fervently debated topic in machine translation is evaluation since there are many valid translations for each input sentence. At some point, we need some quantitative way to assess the quality of machine translation systems or at least a way to be able to tell if one system is better than another or if a change in the system led to an improvement. One way is to ask human judges to assess the adequacy (preservation of meaning) and fluency of machine translation output, or to rank different translations of an individual sentence. Other criteria, such as speed, are also relevant in practical deployments [6].

Machine translation systems are evaluated by using either an automatic evaluation method or a human evaluation method. Human evaluation is time-consuming and expensive. It is also inherently subjective.

Thus, we used a method of automatic machine translation evaluation, which is BLEU score metrics to evaluate the performance of Amharic-Afaan Oromo machine translation system. BLEU was one of the metrics to achieve a high correlation with reference translation and remains one of the most popular automated and inexpensive metrics used in different researches for evaluation purpose. This is the most widely used method of automatic evaluation where we compute n-gram precision concerning reference translation.

1.5 Scope and limitation of the study

There are different types of Machine translation approaches such as example-based approach, rule-based approach, statistical approach and hybrid approach. In this study, we used statistical machine translation approaches.

This thesis focuses on designing Amharic –Afaan Oromo Statistical machine translation approach to a translate sentence only written in Amharic text into Afaan Oromo text and vice versa. The source of the data collected from the Old Testament and New Testament holy bible which include Amharic and Afaan Oromo version. These sources are available, and they are parallel corpus which is suitable for SMT.

Though there are word based, phrased based and tree based SMT approaches, due to time constraint to train, test and analyze the results, but only word a morpheme based SMT is used for this thesis.

Some limitations have been faced during the process of conducting this research. The major limitations are the absence of parallel sentences. Due to the unavailability of the standardized corpus (corpus ready for MT research purpose) thus we prepare a parallel corpus for both target and source language and this takes time by itself. As a result of this, the research is limited to working with very little text.

1.6 Contribution of the study

Machine translation has a great role in exchanging information among different languages around the world, the Machine translation rate is faster than a human translator [4].

The significance of this research can be used to develop machine translation software for Afaan Oromo to Amharic and vice versa and improving efficiency as compared to manual translation and able to translate and understand Afaan Oromo benefited in getting resources that are written in Amharic and vice versa it is possible to address information and solves language barriers between individuals to read and understand different publications.

Also, it contributes for future researcher's and development regarding Amharic – Afaan Oromo language pair used as an additional component in the area of natural language processing specifically in machine translation, Information retrieval, speech processing and text processing

1.7 Thesis Organization

This thesis paper is organized into six chapters. The first chapter discussesed in the above section. The second Chapter presents works of literature that have been conducted on overview of machine translation, approaches of machine translation, different tools used for corpus alignment, automatic evaluation and reviews related works. The third chapter deals with an overview of Amharic and Afaan Oromo language and relationship and difference between Amharic and Afaan Oromo language and discusses Morphology of Amharic and Afaan Oromo language. Chapter four presents a detailed description about the design and development of the prototype statistical machine translation bidirectional Amharic – Afaan Oromo machine translation including, corpus preparation, types of the corpus used for the study, corpus alignment, and briefly discuss the prototype of the system. Chapter five present the experimentation of the study and the results of the experiment with their interpretation of findings. Finally, chapter six deals about conclude the thesis with findings and recommendation for future works.

Chapter Two

2 Literature Review

2.1 Overview

Under this chapter a brief overview of machine translation which includes the approaches of machine translation those are Statistical Machine Translation (SMT), Rule-Based Machine Translation (RBMT), Example-Based Machine Translation (EBMT), Hybrid Machine Translation (HMT), and neural machine translation (NMT), automatic evaluation and it is also review related works that have been done in machine translation for different languages using different approaches and methodologies.

2.2 Machine translation

Machine Translation (MT) is a sub-field of Artificial Intelligence (AI), which is an automated translation system that can translating text or speech from one language (source language) into another language (target language) using a computing system with or without human intervention or assistance [14].

One of the major, oldest, and most active areas in natural language processing is machine translation. machine translation generally starts in the 1950s and 1960s, the impact of the Automatic Language Processing Advisory Committee (ALPAC) report in the mid-1960s that Machine Translation could not produce quality translations as human translators. Research during the 1980s IBM started work in statistical machine translation, new developments in research in the 1990s the parallel text availability had increased the interest in statistical machine translation, and the growing use of systems in the past decade [15]. These resulted in the birth of modern Machine translation.

Machine translation Systems that translate between only two particular languages are called bilingual systems and those that produce translations for any given pair of more than two languages are called multilingual systems. In the case of unidirectional, the translation system from the source language into the target language it is called only in one direction. Bidirectional systems work in both directions in a way that one language can act as a source and the other as a target language and vice versa [7].

In the world, the demands for machine translation become increased due to increase information exchange between different languages. but those demands are not fulfil using manually Because of this manual translation is slower and expensive as compared to machines That are when machine translation (MT) comes in, which can solve this barrier.

2.3 Machine translation approach

Machine translation can be classified according to the methodology into four basic approaches are Rule-Based Machine Translation Approach (RBMT), Corpus-Based Machine Translation Approach (CBMT), Hybrid Methods, and Neural MT (NMT)[16] [17]. The choice of approach for Machine translation depends upon the available resources and the kind of languages involved.

2.3.1 Rule-Based Machine Translation Approach (RBMT)

Rule-based was the earliest approach to machine translation. Rule-Based Machine Translation (RBMT) denotes systems based on linguistic information about the source and target languages and has much to do with the morphological, syntactic, and semantic information about the source and target language. Also, based on millions of bilingual dictionaries for the language pair[18]. In RBMT, the transfer takes place through human-created rules in three different phases through the analysis-transfer-generation (ATG) process [8].

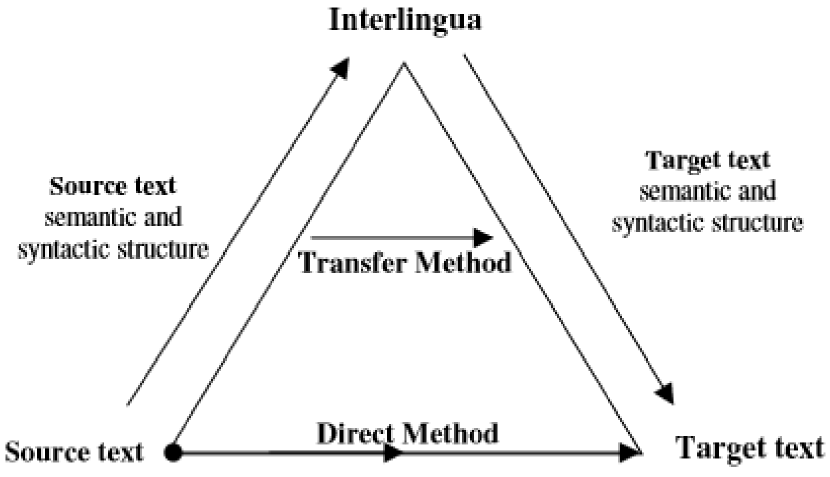


Figure 2-1 Different Methods of Rule-based Machine Translation

The rule-based machine translation approach can be categorized into the direct approach, indirect approach and Interlingua approach.

2.3.1.1 Direct approach

This is the first developed machine translation system. In this approach, the translation is based on large bilingual dictionaries and word-level translation with some simple grammatical adjustments. In this approach, a direct translation system is designed for a specific source and target language pair. These systems depend on well-developed dictionaries, morphological analysis, and text processing software. This approach is suitable for closely related language pairs [16]. Words of Source Language are translated without passing through an additional/intermediary representation.

The process of direct translation includes the following:

- Shallow morphological analysis
- Lexical transfer, based on the bilingual dictionary
- Local reordering
- Morphological generation

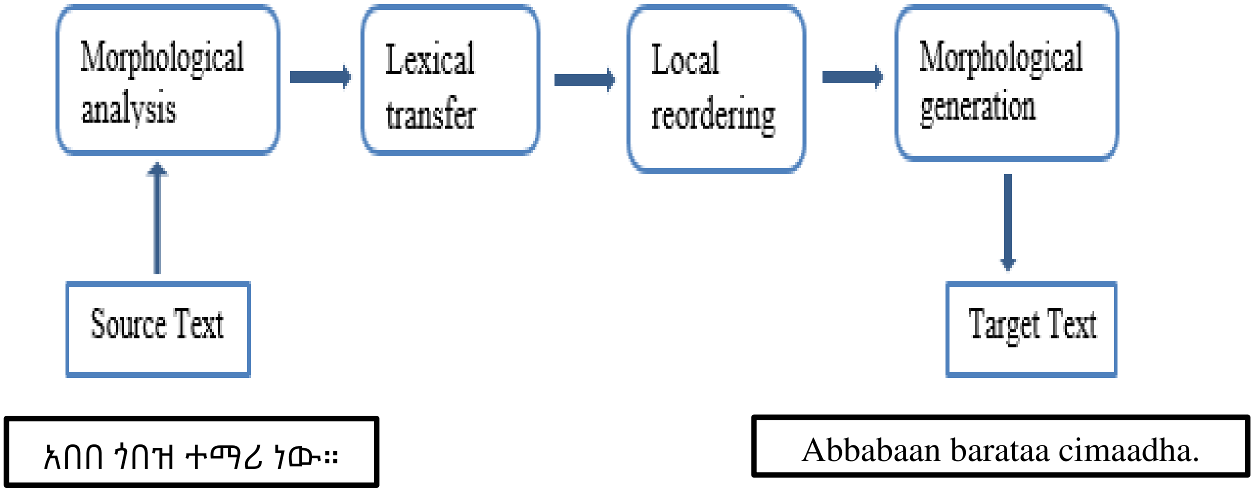


Figure 2-2 Direct machine translation

2.3.1.2 Indirect approach

The indirect approach also called the Transfer Approach or the Linguistic Knowledge (LK) translation[16]. The second generation of machine translation. In this approach, the source language is transformed into less language-specific representation and an equivalent representation is generated for the target language using bilingual dictionaries and grammar rules. The system uses an intermediate representation that captures the structure of the original text to generate the correct translation. Transfer based machine translation process involves three Phases: analysis, transfer and generation [5].

Transfer based approaches need rules for - syntactic transfer, Semantic transfer and lexical transfer

- Syntactic transfer rules – will tell us how to modify the source parse tree to resemble the target parse tree.
- Semantic transfer – using semantic role labeling.
- Lexical transfer rules – based on a bilingual dictionary – The dictionary can be used to deal with lexical ambiguity.

2.3.1.3 Interlingua approach

Interlingua aims to create linguistic homogeneity across the globe. Interlingua is a combination of two Latin words Inter and Lingua which means between/intermediary and language respectively. In Interlingua, the source language is transformed into an auxiliary/intermediary language that is independent of any of the languages involved in the translation. The translated verse for the target language is then derived through this auxiliary representation .interlingua approach required analysis and synthesis modules [16] [18]. The aim of the analysis is the derivation of an Interlingua representation.

This system is that the inter-lingual becomes more valuable as the number of target languages it can be turned into increases. This model translates by performing deep semantic analysis on the input from the source language into the Interlingua representation and the target language generating from the Interlingua. The inter-lingua approach is most attractive for multilingual systems.

2.3.2 Corpus-Based Machine Translation Approach (CBMT)

This is one of the main methods of machine translation is Corpus-Based Machine Translation. The corpus-based approach for machine translation has emerged as one of the widely explored areas in machine translation since 1989 [18]. Corpus-based machine translation overcomes the problem of the knowledge acquisition problem of rule-based machine translation. This approach as its name points uses a large amount of raw data in the form of parallel corpora.

Because of a high level of accuracy achieved during the translation, this method has dominated over other approaches. The Corpus-based approach is further classified into Example-Based Machine Translation and Statistical Machine Translation [18].

2.3.2.1 Example-Based MT (Memory based translation)

Example based MT also known as memory-based MT and this concept of "Translation by Analogy" was first proposed by Makoto Nagao in 1984, but was used by the DLT (Distributed language translation) project in Japan towards the end of the 1980s[16]. It is based on the idea of reusing examples of existing translations as the basis for a new translation. A database of previously analyzed text is stored in the Translation Memory. The basic principle is that, if a previously translated phrase occurs again, the same translation is likely to be corrected again.

An Example-Based Machine Translation (EBMT) system is given a set of the source language sentences and translates each sentence in the target language with a point to point mapping. These examples are used to translate by uses of the bilingual corpus with parallel texts of source-language and the target language [18].

EBMT attempts to choose the best among translation candidates through a syntactic and semantic match. EBMT's method of analyzing the input source sentence is much harder, involving NLP layers of morphology analysis upward, until possibly deep semantic analysis [8]. The main advantage of the EBMT approach is the assurance that the results will be accurate and idiomatic; since the texts have been extracted from databanks of actual translations produced by professional translators. However, the problem arises when one has several different examples each of which matches part of the string, but where the parts they match overlap, and/or do not cover the whole string [10]. May have limited coverage on the size of example database it rise problem.

2.3.2.2 Statistical Machine Translation

The Statistical method of translation was proposed by Warren Weaver in July 1949 but these methods were adopted in the 1950s and 1960s [16]. Statistical machine translation (SMT) deals with automatically translated sentences in one human language (for example, Amharic) into another human language (such as Afaan Oromo). It is a mathematical analysis-based approach.

In this, statistical methods are applied to generate a translated version using bilingual corpora and by using the n-gram approach the translations are generated using a statistical methods system. The Statistical Based MT gives results by picking those word(s) from the given surrounding words which have the highest probability of occupying its current position[16].

Unlike rule-based MT systems, this approach does not require any language-specific linguistic knowledge to perform the translation. The only requirement for the statistical machine translation system is a huge parallel corpus. There is no single system for corporate quality control.

2.3.3 Hybrid Approaches

Hybrid approaches use a linguistic method to parse the source text, and a non-linguistic method, such as statistical-based or example-based, which has proven to have better efficiency in the area of MT systems. It takes the synergy effect of rule-based, SMT and example-based. Hybrid machine translation approach has been developed by combining the positive sides of statistical machine transition and rule-based machine translation methodologies. A hybrid system can be a combination of rule-based and example-based approaches or a combination of RBMT, EBMT and SBMT or a combination of any of these [16].

In HMT there are three components of architecture: identification of source language by observing chunks (words, phrases and equivalents), the transformation of the chunks into the target language, and generation of translated language [19].

2.3.4 Neural Machine Translation Approaches

Neural machine translation is a new breed of corpus-based machine translation it is similar to the statistical machine translation technology that was the state of the art until very recently, but uses a completely different computational approach: neural networks [17]. These systems are also known as sequence-to-sequence models or encoder-decoder networks and were initially fairly simple neural network models made out of two recurrent parts [20]. is an approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, which consists of many small sub-components (words) that are tuned separately, neural machine translation attempts to build and train a single, large neural network that reads a sentence and outputs a correct translation.

Most of the proposed neural machine translation models belong to a family of encoder–decoders [17] [20]. Two recurrent neural networks (RNN) Encoder, is used by the neural network to encode a source sentence into a fixed vector and decoder, used to predict words in the target language.

The main advantage with this encoder-decoder approach is that a neural network needs to be able to compress all the necessary information of a source sentence into a fixed-length vector and There is no separate language model, translation model, and reordering model, but just a single sequence model that predicts one word at a time[20]. The main disadvantages of neural machine translation (NMT) are time-consuming if target vocabulary is large, weak to OOV (out of vocabulary) problem, difficult to debug the errors, and needs high perform computing devices (GPU - graphic process unit [19].

2.4 Statistical Machine Translation

The field of machine translation has recently been energized by the emergence of statistical techniques, which have brought the dream of automatic language translation closer to reality [6]. The general idea in the SMT system is that the translation will be from the most likely translated word.

In statistical machine translation, we use both a translation model and a language model, which ensures fluent output. To finds the most probable target text sentence given a source text sentence based on the noisy channel model. Combining a language model and translation model in this way is called the noisy-channel model. The noisy-channel model developed by Shannon [1948] [6]. In the noisy channel model, we can rewrite this with Bayes' rule: -

$$\hat{T} = \underbrace{argmax}_{\text{Decoding algorithm}}\ \underbrace{P(S/T)}_{\text{translation model}}\ \underbrace{(T)}_{\text{language model}}$$

In SMT approaches there are three components:

- Language model to compute the probability of target language P(T)
- Translation model to compute P(S/T), and
- Decoder attempt to translation process the source text to target-language text.

2.4.1 Language model

Language models measure the fluency of the output and are an essential part of statistical machine translation. They influence word choice, reordering and other decisions. Mathematically, they assign each sentence a probability that indicates how likely that sentence is to occur in a text. For morphologically rich languages like Amharic and Afaan Oromo, better language models may predict individual morphemes [6].

The language model (LM) compute the probability of the target language ‘T’ as probability P (T). The probability is computed using the n-gram model. In the basic sense of the term, is a computable probability distribution over word sequences, typically sentences, which attempts to approximate an underlying stochastic process based on an observed corpus of sequences produced by that process [21]. The statistical language model is predicting the next word is closely related to computing the probability of a sequence of a word.

A sentence is decomposed into the product of conditional probability. By using chain rule an SLM is a probability distribution P(W). A statistical language model is a probability distribution over sequences of words P(w)$(W = W_1, W_2, W_3 ... W_n)$ that model how often each sequences W occurs as a sentence.

A language model gives the probability of a sentence. To calculate sentence probability, it is required to calculate the probability of a word, given the sequence of the word preceding it. The language model can be calculated with an N-gram language model. N-gram language models are based on statistics of how likely words are to follow each other [6].

Based on Markov assumption N-gram model predicts the probability of a word based on few previous words, n indicates the number of the word in a sequence and probability of a word W is calculated on the base of N-1 previous words. The probabilities obtained from the N-gram model could be unigram (size 1), bigram (size 2), trigram (size 3) or higher-order N-grams.

2.4.2 Translation model

To build the translation model we should have a parallel corpus of source and target language. The role of the translation model is to find P(S/T), assigns the probability that a given source text/sentence (S) (Amharic (AM)/Afaan Oromo (OR)) generates target sentence/text (T) (Amharic (AM)/Afaan Oromo (OR)). The training corpus for the translation model is a sentence-aligned parallel corpus of the languages source and target. As mentioned above, for a given source and target sentences AM and OR, it is the way sentences in AM get converted to sentences in OR which is denoted by P (T/S) [3] [4]. It is calculated as:

$$P(T|S) = \frac{\text{Count}(T, S)}{\text{Count}(S)}$$

We cannot compute the above equation from counts of the source and target sentences in the parallel corpus if the sentence is too long, to overcome this problem decomposes the sentence into smaller chunks. To find the sentence translation probability using the translation probabilities of the words in the sentences

$$P\ (\mathrm{S}|\mathrm{T}) = \sum_{x} P\ (\mathrm{X}, \mathrm{S}|\mathrm{T})$$

The variable X represents alignments between the individual chunks in the sentence pair where the chunks in the sentence pair can be morphemes or words or phrases. In morpheme-based translation, the fundamental unit of translation is a morpheme. Phrase-based translations, most commonly used, translates whole sequences of words, where the lengths may differ in which blocks are not linguistic phrases but, phrases found using statistical methods from the corpus.

2.4.3 Decoder

Decoding is a generate-and-score process in which the best translation is searched for in the space of all possible translations [8]. The goal of decoding is to find the translation with the best score. The job of the decoder is to take a source language (either Amharic (am) or Afaan Oromo (or)) from the target language (either Amharic or Afaan Oromo) and produce the best translation using translation model and language model. Finding the sentence that maximizes the translation and the language model probabilities is a search problem. It looks up all translations of every source word or phrase, using a word or phrase translation table and recombines the target language phrases that maximizes the translation model probability multiplied by the language model probability, which is,

$\mathbf{argmax}$ or $(\boldsymbol{p}(\mathrm{or}|\boldsymbol{a}\mathrm{m}) * \boldsymbol{p}(\boldsymbol{a}\mathrm{m}))$. From Afaan Oromo to Amharic translation

$\mathbf{argmax}\ \boldsymbol{a}\mathrm{m}\ (\boldsymbol{p}\ (\boldsymbol{a}\mathrm{m}|\mathrm{or}) * \boldsymbol{p}\ (\mathrm{or}))$ Also for translating Amharic to Afaan Oromo

By the above-mentioned procedures, decoder performs the translations of the input text for both languages.

2.5 Alignment

Alignment is the arrangement of something in an orderly manner with something else[19]. It can be performed at different levels, from paragraphs, sentences, segments, words and characters.

Word alignment: Which determines the translational correspondences at word level given a parallel corpus. It is a mapping between the words in the source sentence and the words in the target sentence. The initial statistical models for machine translation are based on words as atomic units that may be translated, inserted, dropped, and reordered [6].

Word-based statistical machine translation ignores possible morphological relatedness of the words. This is more of a problem for inflectional languages the richer their morphology (Amharic and Afaan Oromo), the larger the training corpus has to be to cover most of the possible word forms [22]. Currently, word alignment models for statistical machine translation do not address morphology beyond merely splitting words.

A morpheme alignment is a function mapping a set of morpheme positions in a source language sentence to a set of morpheme positions in a target language sentence.

Word alignment commonly has done using IBM Models 1-5. IBM Model 1 is weak in terms of conducting reordering or adding and dropping words. The IBM Model 2 has an additional model for an alignment that is not present in Model 1. Some source words may be translated into multiple target words (fertility of the words) this problem is addressed in IBM Model 3. In IBM Model 4, each word is dependent on the previously aligned word and the word classes of the surrounding words. IBM Model 5 reformulates IBM Model 4 by enhancing the alignment model with more training parameters to overcome model deficiency[6].

2.6 Evaluations of machine translation

Machine Translation emerges as an important mode of translation, its quality is becoming more and more important. Judging translation quality is called machine translation evaluation. Machine translation systems are evaluated by using an automatic evaluation method or human/manual evaluation method.

2.6.1 Manual Evaluation method

For evaluating machine translation output is to look at the output and judge by hand whether it is correct or not. Bilingual evaluators who understand both the input and output language are best qualified to make this judgment Manual translation correctness may be too broad a measure. It is, therefore, more common to use the two criteria fluency and adequacy [6].

Adequacy - is the notion of how faithfully the meaning of a sentence in the source language is transferred to its translation in the target language. This factor is also called faithfulness or fidelity in SMT parlance[8].

Fluency: - is native speaker acceptability of the translated sentence. Fluency requires the translation to maintain correct word choice, word order, and register[8].

2.6.2 Automatic Evaluation Method

Automatic machine translation evaluation metrics were developed due to the high costs, lack of repeatability, subjectivity, and slowness of evaluating machine translation output using human judgments, and the desire to enable automatic tuning of system parameters [6]. Human evaluation of translation is subjective and can be too slow for practical purposes. This gave rise to the arrival of the automatic method of MT evaluation. There are different type of automatic evaluation metrics to evaluate machine translation systems, such as Precision and Recall, Word Error Rate (WER) and METEOR [6]. All methods except BLEU requires human translation and time- consuming [3]

- **BLEU (Bilingual Evaluation Understudy): -**

The currently most popular automatic evaluation metric, the BLEU metric, has an elegant solution to the role of word order. The accuracy of the translation results from Amharic to Afaan Oromo MT system has been evaluated using a BLEU (Bilingual Evaluation Understudy) technique. The BLEU score is one of automatic evaluation metric to evaluate the performance of SMT. It works similarly to position-independent word error rate but considers matches of larger n-grams with the reference translation. Given the n-gram matches, we can compute n-gram precision, i.e., the ratio of correct n-grams of a certain order n with the total number of generated n-grams of that order[6].

BLEU is the geometric mean of clipped n-gram precisions for different n-gram lengths (usually from one to four), multiplied by a factor (brevity penalty) that penalizes producing short sentences containing only highly reliable portions of the translation[21]. The BLEU score of system output is calculated by counting the number of n-grams, or word sequences, in the system output that occurs in the set of reference translations.

2.7 Related work

Many types of research have been done in machine translation for different languages using different approaches and methodologies. Some of the studies with a special focus on SMT approach are discussed below.

2.7.1 Morpheme-Based Bi-directional Ge’ez-Amharic Machine Translation

The research was done on morpheme-based bi-directional Ge’ez-Amharic machine translation system using a statistical approach. Conducted by Tadesse KASSA the main objective of the research is to explore the effect of morpheme level translation unit for bi-directional Ge’ez-Amharic machine translation [3].

The size of the parallel corpus that used for the experiment contains a total of 13,833 simple and complex sentences and tools used for this experiment are: - Mosses for the translation process, MGIZA++ for alignment of word and morpheme, IRSTLM for language modelling and for morphological segmentation Morfessor were used.

The research work was implemented using statistical machine translation and compared Experimental performance results between Morpheme-level translation and word-level translation.

The results showed a better performance of 15.14% and 16.15% BLEU scores using morpheme-based from Geez to Amharic and from Amharic to Geez translation, respectively as compared to word-level translation, there is on the average 6.77% and 7.73% improvement from Geez-Amharic and Amharic-Ge’ez respectively [3].

The research shows an additional experiment using unsupervised and rule-based morpheme segmentation approaches. The BLEU score is 0.6% and 1.27% for Ge’ez to Amharic and Amharic to Ge’ez respectively. From the result, rule-based morpheme segmentation approaches are better than the unsupervised approach when Amharic is used as a source language and Ge’ez is used as a target language.

Finally, the researcher recommended, Alignment of Ge’ez-Amharic text is a challenging task because of many-to-many correspondence between words/morphemes of the two languages. Hence, there is a need to identify optimal alignment for Ge’ez-Amharic Machine translation and the researcher used prefix and suffix for rule-based morphological segmentation. However since both languages are morphological rich, there is a need to apply machine learning algorithms for designing an optimal model for segmentation [3].

2.7.2 Optimal Alignment for Bi-directional Afaan Oromo-English Statistical Machine Translation

The research has been conducted by Yitayew Solomon [4]. The main objectives of this study to explore the effect of word level, phrase level and sentence level alignment on bi-Directional Afaan Oromo-English was implemented by using statistical machine translation approach.

The researcher used corpus for the experiment was collected from different sources including criminal code, FDRE constitution, Megleta Oromia and Holly Bible. And prepared 6400 simple and complex sentence and make the corpus suitable for use to training and testing the system and used 9:1 ratio respectively. The system is bidirectional, two language models are developed, for English 1900 monolingual sentence and Afaan Oromo monolingual 12200 sentences used. To develop the system using different tools, such as Mosses for decoding purpose, for alignment MGIZA++, Anymalign and hunalign and IRSTLM for language modelling

The researcher conducted six experiments based on different length of aligned phrases from both directions (from English-Afaan Oromo and Afaan Oromo-English). The first and second experiment is carried out by using 4 Maximum and 1 minimum length of phrases. The result obtained from the experiment has a BLEU score of 21% for English to Afaan Oromo and 42 % for Afaan Oromo to English translation. Third and fourth experiment by using 16 Maximum and 4 minimum length of phrases and the result obtained a BLEU score of 27% for English to Afaan Oromo and 47 % for Afaan Oromo to English translation. Fifth and sixth experiment Sentence level alignment by using 30 Maximum and 20 minimum length of phrases. The BLEU score is 18% and 35% for English to Afaan Oromo and Afaan Oromo to English translation respectively.

From the experimental result, the researcher concludes that the experiment by using 16 Maximum and 4 minimum length of phrases shows better performance the BLEU score of 27% for English to Afaan Oromo and 47 % for Afaan Oromo to English translation and a better translation is acquired when Afaan Oromo is used as a source language and English is used as a target language. The researcher recommended hybrid approach to handle varieties of alignment for developing better SMT result.

2.7.3 Bidirectional English-Amharic Machine Translation: An Experiment using Constrained Corpus

The research was conduct by Eleni Teshome [9] to develop bidirectional English-Amharic translation using the constrained corpus. This research work implemented the statistical machine translation approach. The researcher prepared two different corpora which are simple sentences and complex sentences. For simple sentence prepare (corpus I) 1020 sentence manually and (corpus II) 1951 were collected for a complex sentence from public procurement directive 414 and 1537 from the bible.

The researcher also used two methodology testing. The first one is the BLEU score and the second one preparing a questionnaire manually. BLEU score is an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another. The second testing methodology is manual questioner technic, it has an evaluation method on the scale of 1 to 5 in which if a candidate gives 5, it means that the translation was perfect and if 1 is given, it means it has a very poor translation. Two questionnaires were prepared for the test set i.e. from English to Amharic and from Amharic to English. The questionnaires developed were different because the result obtained from Amharic to English and from English to Amharic were not similar.

From the experimental result for the simple sentence using BLEU score had on the average 82.22% and 90.59% from English to Amharic and Amharic to English respectively and using the manual questionnaire method the accuracy is 91% and 97% form English to Amharic and Amharic to English respectively. For the complex sentences, the result acquired from the BLEU Score was approximately 73.38% for the English to Amharic, 84.12% for the Amharic to English and from the questionnaire method from English to

Amharic was 87% and from Amharic to English was 89%. As we have seen from experimental, manual questionnaire method has better performance for a simple sentence.

Finally, the researcher recommended, further research could be performed machine translation on Amharic to other languages, even using languages in Ethiopia such as Tigrigna, Oromifa while preparing a large corpus.

2.7.4 English – Afaan Oromo Machine Translation: An Experiment Using Statistical Approach.

This study was conducted by Sisay Adugna in 2009, with the objective of to develop a prototype English-Afaan Oromo machine translation system using a statistical approach, i.e., without explicit formulation of linguistic rules [10].

The researcher collected bilingual and monolingual corpus from different sources like a bible, some spiritual manuscript and the United Nation's Declaration of Human Rights. For Afaan Oromo language the monolingual corpus contains 62,300 sentences and bilingual corpus 20,000 sentences. The data is organized into training and testing data in the proportion of 9:1 (90% of the data for training and the remaining 10% for testing). The researcher used different tools for implementation of the system, SRILM toolkit was used for language modelling, for word-alignment, GIZA++ which implements the word alignment methods IBM1 to IBM5, Decoding is done using Moses and the documents were preprocessed using different scripts written for this purpose like the apostrophe. Sentence aligning, tokenization, lowercasing and truncating long sentences that take the alignment to be out of optimality was done by those scripts.

The researcher performs different experiments by a varying number of N-grams, the n-gram score for values of n equals 1, 2, 3, 4, 5, 6, 7, 8 and 9 is observed to be 43.96%, 21.57%, 14.42%, 10.72%, 8.04%, 5.52%, 3.76%, 2.23% and 1.30% respectively.

Finally, the author strongly recommends the addition of more bilingual data for further experimentation and development of spell checker for Afaan Oromo that will help facilitate the document preparation.

2.7.5 Geez to Amharic Automatic Machine Translation: A Statistical Approach

The research has been conducted by Dawit Mulugeta [11] with the general objective to investigate the application of Statistical Machine learning technique to Machine Translation from Geez to Amharic. This researcher uses quantitative experimental methodology.

To perform corpus-based machine translation researcher collected 12860 parallel bilingual corpora from the Holy Bible and some other religious books (Wedase Mariam and Arganon). The collected data were divided into training and testing, more than 90% of the collected data was used as a training set. To develop the system the researcher used difference SMT tools Moses for translation, IRSTLM and SRILM for language modelling, GIZA++ for word alignment toolkit that uses to train IBM Model 1 to Model 5 and the Hidden Markov Model and used the BLEU score for evaluation.

The researcher conducted experiments based on sentence-level alignment the performance of the result and the BLEU score obtained were 8.14%. Further investigation has been done to crosscheck and improve the performance score using 10-fold cross-validation (CV) method. The BLEU score result obtained on the trails are 9.11%, 7.44%, 7.61%, 6.36%, 10.26%, 9.39%, 8.01%, 8.54% and 7.72%. The result verifies that the performance is highly dependent on the training and testing data domain. The researcher also checks the performance of the system after splitting each book of the Bible into the training and testing set. The trials have been done three times to see the result and average accuracy is calculated to compare with the 10-fold CV test result and get better performance than the 10-fold CV result.

Finally, the author recommended extension of this research using the different morphological segmentation and synthesizing mechanisms, using a larger corpus size and various domains of contents other than the religious one and should be undertaken using Example-based Machine Translation approach.

2.7.6 Parallel Corpora for bi-Directional Statistical Machine Translation for Seven Ethiopian Language Pairs

The research https://www.aclweb.org/anthology/W18-3812.pdf describes the development of parallel corpora for Amharic, Tigrigna, Afaan Oromo Wolaytta and Ge’ez language and conducted baseline Bi-directional experiments for seven language pairs by using statistical machine translation approach. They also showed that the morphological complexity of the Ethio-Semitic languages and impact on the performance of the SMT.

The researcher collected data is the only religious domain, it includes Holy Bible and different documents written in spiritual theme for each language pair and each parallel corpus has been organized into training, testing and tuning (80% for the training, 10% for tuning and 10% for test sets). The documents were preprocessed using different automatic methods, during pre-processing: character normalization, sentence tokenization and sentence level alignment have been performed.

The researcher used different tools for implementation of the system, Mosses for the translation process, Giza++ for alignment of words and phrases, SRILM toolkit was used to develop the language models, and Bilingual Evaluation Under Study (BLEU) is used for automatic evaluation.

The researcher conducted 14 experiments based on average sentence length for bi-directions the seven Ethiopian language pairs (Amharic – Tigrigna, Tigrigna – Amharic, Amharic – Afaan Oromo, Afaan Oromo – Amharic, Tigrigna – Afaan Oromo, Afaan Oromo – Tigrigna, Amharic – Wolaytta, Wolaytta – Amharic, Ge’ez – Amharic, Amharic - Ge’ez, Wolaytta – Afaan Oromo, Afaan Oromo- Wolaytta, Tigrigna – Wolaytta and Wolaytta - Tigrigna) the performances of the result the BLEU score obtained were 21.22%, 19.06%, 17.79%, 13.11%,16.82%, 14.61%, 11.23%,

7.17%, 7.31%, 6.29%, 4.73%, 2.73%, 2.2% and 3.8% respectively. The authors conclude that when Ethio-Semitic languages are on the target side the performance of SMT systems decreases.

Finally, the researcher recommended the research could be performed using morphemes instead of words as units for both the translation and the statistical language model and use ANN modelling for better performance.

Summary

The following table presents the summary for the above-mentioned related work. The table shows the title of the study, objective of the study, types of experiment and their BLEU score.

Title	Conducted by	Objective	Types of experiment		BLEU score	
Morpheme-Based Bi-directional Ge'ez -Amharic Machine Translation	Tadesse Kassa	To explore the effect of morpheme level translation unit	Word-based	Geez- AM	8.37%	
				AM- Geez	8.42%	
			Unsupervised morpheme-based	Geez- AM	14.54%	
				AM- Geez	14.88%	
			Rule Morpheme based	Geez- AM	15.14%	
				AM- Geez	16.13%	
Bidirectional English-Amharic Machine Translation: An Experiment using Constrained Corpus	Eleni Teshome	To explore the effects of simple and complex sentence			BLEU	Manually
			Simple sentence	Eng.-AM	82.22%	91%
				AM - Eng.	90.59%	97%
			Complex sentence	Eng.-AM	73.38%	87%
				AM - Eng.	84.12%	89%
Optimal Alignment for Bi-directional Afaan Oromo-English Statistical Machine Translation	Yitayew Solomon	To explore the effect of word, phrase and sentence level alignment	4 max and 1 min phrase Complex sentence	Eng.- Orom	21%	
				Orom- Eng.	42%	
			16 max and 4 min phrase	Eng.- Orom	27%	
				Orom- Eng.	47%	
			30 max and 20 min phrase	Eng.- Orom	18%	
				Orom- Eng.	35%	

English – Afaan Oromo Machine Translation: An Experiment Using Statistical Approach.	Sisay adugna	To develop a prototype English-Amharic SMT without explicit formulation of linguistic rule	experiments by a varying number of N-grams, the n-gram score	Eng.-AM n equals 1, 2, 3, 4, 5, 6, 7, 8 and 9	43.96%, 21.57%, 14.42%, 10.72%, 8.04%, 5.52%, 3.76%, 2.23% and 1.30% respectively
Geez to Amharic Automatic Machine Translation: A Statistical Approach	Dawit Mulugeta	to investigate the application of Statistical Machine learning technique to Machine Translation from Geez to Amharic	Sentence level	Geez –Am	8.14%
			10-fold cross validation (CV)	Geez –Am	9.11%, 7.44%, 7.61%, 6.36%, 10.26%, 9.39%, 8.01%, 8.54% and 7.72%.
Parallel Corpora for bi-Directional Statistical Machine Translation for Seven Ethiopian Language Pairs	Solomon T, Michael M, Martha Y, Million M, Solomon A, Wondwosen M, Yaregal A, Hafte A, Tewodros A, Biniyam E, Wondimagegnhue T, Amanuel L, Tsegaye A and Seifedin S.	Development of parallel corpora for Amharic, Tigrigna, Afaan Oromo Wolaytta and Ge'ez language and conducted baseline Bi-directional experiments	Baseline Bi-directional	Am – Tg, Tg – Am, Am– Orom, Orom– Am, Tg– Orom, Orom– Tg, Am– Wo, Wo– Am, Geez – Am, Am - Geez, Wo –Orom, Orom- Wo, Tg – Wo and Wo – Tg	21.22%, 19.06%, 17.79%, 13.11%, 16.82%, 14.61%, 11.23%, 7.17%, 7.31%, 6.29%, 4.73%, 2.73%, 2.2% and 3.8% respectively

Table 2-1 related works summary

Research gap

As we have seen the above-related works there is only one study conducted Bi-directional experiments for seven language pairs by using statistical machine translation approach [12]. The researchers conducted 7 language pair. Bi-directional Amharic - Afaan Oromo language pair is one of them. The researchers used average sentence for translation and they also showed that due to Ethio-Semitic languages morphological complexity the performance of SMT greatly affected therefore, they recommended further research could be performed using morphemes based for both the translation and the statistical language model. Amharic and Afaan Oromo morphologically richness languages. Therefore, further study to design better translation unit. The aim of this study is to explore the effects of word and morpheme level Amharic- Afaan Oromo Machine translation languages.

Chapter Three

3 Amharic and Afaan Oromo language

3.1 Overview

This chapter briefly discusses an overview of Amharic and Afaan Oromo language. The major Amharic and Afaan Oromo syntactic structure, word classes, which are nouns, adjectives and conjunctions, and linguistic relationship and the difference between Amharic and Afaan Oromo language are described in this chapter.

3.2 Ethiopian Languages

Ethiopian is a multinational and multilingual country where more than 83 languages are spoken by about around 114 million people of different ethnic groups [23]. Ethiopians languages can be classified into four groups. The vast majority of languages belong to the Semitic, Cushitic, or Omotic groups. Most of these languages belong to the Afro Asiatic family (Semitic and Cushitic languages; Omotic languages). Additionally, Nilo-Saharan languages are spoken by what the government calls the "Nilotic" people, though scholars distinguish Nilotic from the Surmic languages, Gumuz languages, and Koman languages and spoken in Ethiopia [2] [23].

The Semitic languages are spoken primarily in the northern, central and eastern parts of the country; they include Ge’ez, Tigrinya, Amharic, Gurage, and Harari. Ge’ez, the ancient language of the Aksumite empire, is used today only for religious writings and worship in the Ethiopian Orthodox Church. Tigrinya is native to the northeastern part of the country. Amharic is one of the country’s principal languages and is native to the central and northwestern areas. Gurage and Harari are spoken by relatively few people in the south and east part of the country [23].

The Cushitic languages are mostly spoken in central, southern and eastern Ethiopia. The Cushitic languages are Oromo, Somali, and Afar. Oromo is native to the western, southwestern, southern, and eastern part of the country. Somali is dominant among peoples of the Ogaden and Hawd, while Afar is most common in the Denakil Plain [2] [23].

The Omotic language is Walayta, this language spoken mostly in the areas of southwest Ethiopia. The Nilotic language group is native to the Western Lowlands, with Kunama speakers being dominant [2].

Under the Ethiopian constitution, all languages enjoy official state recognition. However, Amharic is the "working language" of the federal government. Amharic and Afaan Oromo have most widely spoken languages in the country [2].

3.3 Amharic Language

Amharic is a widely spoken language all over the country and serving as the working language of the government of Ethiopia [2]. It is one of the Ethiopian Semitic languages and a branch of the Afro-Asiatic languages. Some 18.7 million people spoke Amharic in the early 21st century [2].

The Amharic language has its alphabet, ፊደል/ fidäl, inherited from the Geez (Ethiopic) language. Geez is an ancient South Semitic language that now serves only as the liturgical language of the Ethiopian Orthodox Tewahedo Church. The writing system consists of 33 basic characters, each of which has seven orders or shapes depending on which vowel with which a given consonant is combined. Other than those alphabets, there are also around forty labialized characters such as "ቋ ሟ ሎ ቧ ኋ…" [9]. Unlike Arabic, Hebrew or Syrian, Amharic is written from left to right.

Most of the words in Amharic are occurring in its inflected form. For obtaining the root form of the words, the suffixes joined with them are to be removed. Also the morphophonemic change occurring when a root word concatenates with a suffix should be analyzed and generalized. Amharic is morphologically rich language.

The usual word order of Amharic sentence is a subject-object-verb (SOV) format, where the subject comes first, then the object and the verb next to the object. For instance, in Amharic language sentence, "እሱ፡ሸሚዝ፡ገዛ። /ɨsu shɛmizɨ gɛzä/ "እሱ" / ɨsu / is a subject, "ሸሚዝ" / shɛmizɨ / is an object and "ገዛ" / gɛzä / is a verb.

3.4 Afaan Oromo Language

Afaan Oromo is a member of the Cushitic branch of the Afro-Asiatic language family [24]. More than two-thirds of the speakers of the Cushitic languages are Oromo or speak Afaan Oromo, which is also the third-largest Afro-Asiatic language in the world after Hausa and Arabic [25]. The language is widely spoken in Ethiopia and neighboring countries like Kenya and Somalia. Afaan Oromo is an official language of Oromia Regional State and using the Afaan Oromo language in teaching and learning processes for primary and junior secondary schools of the region [4].

Afaan Oromo language writing system based Latin alphabet and It has its own script known as Qubee has been adopted and became an official script of Afaan Oromo since 1991[4]. Like Amharic language Afaan Oromo also written from left to right.

Afaan Oromo syntactic structure is similar to the Amharic language. The sentence structure is subject-object-verb (SOV) format [24], where the subject comes first, then the object and the verb next to the object. For example, if we take Afaan Oromo sentence “inni shamiza bita. “inni” is the subject, " shamiza " is the object and " bita " is the verb of the sentence.

3.5 Nature of the Language pair

Amharic and Afaan Oromo language have differences and also similarities linguistic structure discussed below.

3.5.1 Noun

A noun is a word that functions as the name of some specific thing or set of things. Amharic nouns are marked for any combination of number, definiteness, gender and case. A number have both the singular and the plural (Plural /-ኦች/-occ or /-ዎች/-wocc), Gender (for masculine -ኡ and -ወ for feminine), on the other hand, The Amharic definite article is a suffixed element and has different realizations depending on whether the noun to which it is attached ends in a consonant or a vowel, singular or plural and masculine or feminine.

Nouns are inflected through four cases, equally in the singular and the plural, i.e., the nominative, the genitive, dative and accusative [20]. For example: -

	Singular	Plural
Nom:	ቤት - a house	ቤት-ኦች- houses
Gen:	የ-ቤት- of a house, a house's	የ-ቤት-ኦች- of houses
Dat:	ለ-ቤት- to a house	ለ-ቤት-ኦች to houses
Acc:	ቤት-ን- a house	ቤት-ኦች-ን – houses

Nouns in Afaan Oromo can vary to reflect the number, gender, and case (subjective, objective or possessive). Gender is not marked by any special suffix except -eessa and -eettii which denote mascu1ine and feminine with only some adjectival roots. For Example, dur (rich), for male (dur-eessa) for female (dur-eettii). Gender is also distinguished by different roots for kinship relations, e.g. cibbaa -father, kdadha - mother, ilma - son, inulla – daughter, the gender of animals by using korma for masculine and dhaltuu for feminine, the base and subject form is obtained by suffixing -ni, -n, or -i for masculine, and -n, and -tii for the feminine to noun stems, and -ii for masculine and -n for the feminine to adjectives [26].

3.5.2 Personal Pronouns

Personal pronouns are pronouns that are associated primarily with a particular grammatical person, first-person (as I(እኔ)), second person (as you(አንተ/አንቺ)), third-person (as he, as she(እሱ/እሷ)). The subject form is the nominative case which occurs' as the subject of a simple verb, a verbal phrase, and a nominal sentence [26]. Like Amharic, Afaan Oromo uses different forms of personal pronouns to indicate their role in the sentence.

	Grammatical person	Subject	
Singular	first person	ani, an	እኔ /ɨne/
	second-person	Ati	አንተ /ante/
		Inni	እሱ /ɨsu/
	third person	Isiin	እሷ /ɨswa/
Plural	first person	nuti, nu	እኛ /ɨɲä/
	second-person	Isini	እናንተ,እርስዎ/ɨnänɨtɛ, ɨrɨsɨwo/
	third person	Isaani	እነሱ /ɨnɛsu/

Table 3-1 Personal pronoun in Amharic and Afaan Oromo language

Amharic language are some unique features such as the second person 'you' which may take different agreements when referring 'plural' or 'respected (politeness)'

3.5.3 Articles

The Amharic language has no articles before nouns instead of that suffixes are added to show definiteness instead of using the definite article.

For Example. The girl ➡ልጅቷ,

“Girl” refers to ልጅ and the definite article “the” is replaced by the suffix “ቷ” to show definiteness.

In Afaan Oromo, like the Amharic language, there are no articles that are inserted before nouns. The last vowel of the noun is dropped and suffixes (-icha, -ittii, -attii, -utti) added to show definiteness [10].

For example, karaa ------- 'road', karicha------------ 'the road

Adurree--------‘cat’, adurrattii------------‘the cat’ (feminine)

3.5.4 Conjunction

A conjunction is a word that is used to connect words, phrases, clauses and sentences or coordinate two or more words, phrases, clauses and sentences.

The following are a list of Amharic coordinating conjunction and their Afaan Oromo equivalent.

Amharic coordinating conjunction	Afaan Oromo coordinating conjunction	English coordinating conjunction
እና /ɨnä/	Fi	And
ወይም /wɛyɨmɨ/	Yookin	Or
ግን /gɨnɨ/	haa ta’u malee	But
ለ /lɛ/	Kan	For
ገና /gɛnä/	Siachi	Yet
ስለዚህ /sɨlɛzihɨ/	Kanaafi	so, therefore, As a result, Consequently,

Table 3-2 Amharic and Afaan Oromo coordinating conjunction.

3.5.5 Punctual mark

Amharic language uses its own script, most of the punctuation mark those used in Amharic are different from Afaan Oromo. Afaan Oromo punctuation is the same with English except in the case of Apostrophe mark (‘). Apostrophe mark (‘) in Afaan Oromo is used in writing to represent a glitch sound known as hudhaa appearing between two different consecutive vowels in Afaan Oromo However, in the Amharic language is not used.

The following punctuation marks are used in both languages.

Amharic Symbol	Afaan Oromo symbol	English name	Purpose
: (hulet netib)	White space	Space	To separate individual word
:: (arat netib)	.	Period	To separate sentence
፣ (netela serez)	,	Comma	To separate text in a list
፤ (dirb serez)	;	Semicolon	To indicate a pause to independent clauses
? (tiyake milket)	?	Question mark	To ask a question, placed at the end of the sentence
“ ”	“ ” or ‘ ’	Quotation Mark	Used around direct speeches, quotations or to emphasize a word or phrase
! (kal agano)	!	Exclamation mark	To symbolize the anger, surprise or excitement of that particular sentence.
- (serez)	-	Hyphen	To link the parts of a compound word or phrase
( ) (knef)	( )	Bracket	to enclose an additional inserted word

Table 3-3 some punctuation marks in Amharic and Afaan Oromo language

3.5.6 Adjective

An adjective is a word, come before or after nouns or pronoun in a sentence to modify them and tell about things behaviour or characteristics, like shape, size, colour, type, property. The adjectives position in sentence in Amharic and Afaan Oromo language are not the same. In Amharic adjectives are used before a noun but in Afaan Oromo language adjectives come after the nouns.

For instance: "BLEU pen" ሰማያዊ እስክሪብቶ biirii dooqee

Adjective noun noun Adjective

Amharic Adjectives are generally derived from verbs. The number of simplex adjectives is relatively small. Some simple adjectives ቀይ'red', ደግ'generous'. Adjectives are also derived from nouns or verbal morphemes' 'forceful', from ሀይል'force, energy'. Like nouns, adjectives are inflected for Number, Case, Gender and Definiteness[27]. In Amharic to indicate plural number adjectives repeat themselves,

for example, black (ጥቁር) → ጥቋቁር(blacks).

Like the Amharic language, Oromo adjectives do show gender, number and neutral. Gender morpheme (-aa, - (a) acca, -aawaa and-eessa for masculine suffix and – oo, -tuu, -eettii, -oowtuu and –eettii for feminine suffix) and also which do not show gender differentiation. E.g. adda - different, adii – white, the plurals of adjectives are formed by reduplicating the first syllable of the root [26].

For Example: -

Singular		Plural		
Masculine	Feminine	Masculine	Feminine	
Diim-aa	diim-tuu	diddiimaa (dimdiimaa)	Diddiimtuu	red ones

Table 3-4 Afaan Oromo morphemes of an adjective

3.5.7 Adverb

An adverb is a word that modifies or qualifies a verb. In Amharic and Afaan Oromo there are different types of an adverb. Adverbs have the function to express different adverbial relations such as relations of time, place, degree, manner or measure.

Amharic and Afaan Oromo adverbs precede the verb they modify and More than one adverb can occur in a sequence in both Amharic and Afaan Oromo. Many Afaan Oromo adverbs are derived from other parts of speech in Oromo. However, some words are fit to be categorized as adverbs

3.5.8 Subject-Verb Agreement

Amharic and Afaan Oromo Verbs agree with their subjects. Both Amharic and Afaan Oromo verbs are marked for number, person and gender of subjects of the verb are marked by suffixes or prefixes on the verb.

For example: ልጅቷ መስኮት ሰበረች:: → mucaanyoon foddaa cabsitee

The subject (ልጅ-ቷ), the letter “ቷ” is described feminine and also the verb (ሰበረ-ች), the letter “ች” is marked for the feminine.

In Afaan Oromo the subject mucaanyoon,the suffix –yoon is described female and male and also the verb cabsitee the suffix –tee is marked feminine.

3.6 Morphology

Morphology describes how words are formed in the language and it tries to discover the rules that govern the formation of words in a language [19]. The study of the combining morphemes to form words and deriving the morphemes from words is called morphology. The meaning of a word is refined by placing other words morphology around it. The smallest meaningful unit of a language is a morpheme. A morpheme may either be a word or part of a word. For example, in Amharic we have the word 'ድመት' which is 'adurree' in Afaan Oromo language. 'ድመት’ (adurree) cannot be decomposed into smaller ideas based upon the word and it is a free morpheme that can function independently.

To divide up words in a language is to take the view that have two categories: free morphemes and bound morphemes. content words (free morphemes) that carry meaning by referring to real and abstract objects, actions, and properties (called nouns, verbs, and adjectives/adverbs) that can function independently on the one hand, and function words (bound morphemes) that inform us about the relationships between the content words on the other that cannot function alone [6]. Both types of morpheme occur in Amharic and Afaan Oromo language.

Bound morpheme serves as an affix on a free morpheme. Affix morphemes are divided into two major functional categories, namely derivational morphemes and inflectional morphemes. A derivational morpheme forms new words either by changing the meaning of the base to which it is attached or by changing the word-class that the base belongs to. Inflectional morpheme does not change the referential or the cognitive meaning of the word. Instead, it adds attributes such as a person, gender, number, tense, etc. to the base word such that the word can fit into a particular syntactic slot [22].

According to [6] Morphology poses a special problem to machine translation, especially when dealing with highly inflected languages with large vocabularies of surface forms. If the input language is morphologically rich, it may help to simplify it. If both input and output language models are inflected, we may want to build models that translate lemma and morphemes separately.

The morphological complexity process differs from language to language. Amharic and Afaan Oromo language have a much richer morphology. Rich morphology indicates that for each lemma many word forms exist. Amharic and Afaan Oromo has singular and plural numbers.

Chapter four

4 Development of the MT system

4.1 Overview

As mentioned earlier, the main objectives of the study to develop bidirectional Amharic-Afaan Oromo machine translation using a statistical approach. Therefore, to conduct the experiment corpus preparation, software tools used, the overall architecture of the system has been discussed under this chapter.

4.2 Corpus Collection and Preparation

4.2.1 Corpus collection

Statistical Machine Translation system or corpus-based approaches of machine translation makes use of a parallel corpus of source and target language pairs therefore for this bidirectional Amharic- Afaan Oromo machine translation we collected the corpus from different online sources which contains the parallel text of Amharic and Afaan Oromo language. These sources include the Holly Bible (Old Testament and New Testament) and different documents written in spiritual theme.

We collected 3143 sentence for each Amharic and Afaan Oromo language and we try to get prepared parallel 11,457 sentence for both Amharic and Afaan Oromo language from online source which is github.

4.2.2 Corpus preparation

The major difficulty task was preparing parallel corpus manually for both languages because of the scarcity of parallel corpus. The collected files are in different formats and encoding. Therefore, manipulation of the data to put it into uniform format and encoding was necessary. To prepare parallel corpus we tried to Using manually and automatically. All of the data in the corpus was subsequently converted to plain text, cleaned up from the blank lines and noisy characters, and its encoding was converted to UTF-8 automatically to make it ready for training of the system. There is a number of preprocessing to get a cleaned corpus. These preprocessing includes sentence tokenization, True-case and cleaning. Before undertaking the training of the system, the data must be pre-processed.

Tokenization: is the process of replacing sensitive data with unique identification symbols that retain all the essential information about the data without compromising its security. The tokenized paragraph inserts space between words and punctuation marks. Tokenizing of corpus makes use of a Perl script. True-casing: is the problem in natural language processing (NLP) of determining the proper capitalization of words where such information is unavailable. This apply for Afaan Oromo language. Cleaning: that cleans up a parallel corpus, so it works well with the training script. It performs removes empty lines and removes redundant space characters. The cleaning step to remove longer sentence that has more than 80 words.

Unsupervised morpheme segmentation

The same corpus has been used for morpheme-based translation. But dataset preparation for morpheme-based translation is different from that of word-based, for word segmentation, done using unsupervised segmentation tool called Morfessor.

During word segmentation morfessor follows the following procedure. The first step is to create a model for both corpus using morfessor script and train the morfessor model and then add sentence ending marker at the end (@). The third step Using the created model and morfessor-segment script, segment text corpus as an input for both languages. The forth steps using python script merge segmented words into sentence level using the added sentence marker (@).

For amharic and Afaan Oromo corpus to create the model using training and segment input corpus we use the following syntax.

a. morfessor-train

- Morfessor -t corpus.am -s morphmodelam.bin (for amharic)
- Morfessor -t corpus.or -s morphmodelor.bin (for Afaan Oromo)

b. morfessor-segment

- Morfessor-segment -l morpmodelam.bin corpus.or-am.am > morphed.sent.or-am.am (for amharic)
- Morfessor-segment -l morpmodelor.bin corpus.or-am.or > morphed.sent.or-am.or (for Afaan Oromo)

Amharic	Afaan Oromo
የ ጤና ችግር ን ጨምሮ ከቁጥጥራችን ውጭ የ ሆኑ ሌሎች ተፈታታኝ ሁኔታ ዎች ሲያ ጋጥሙን ይሖዋ እንደሚያስ ብል ን በ መተማመን የሚ ያስጨንቀን ን ነገር ሁሉ በ እሱ ላይ መጣል ይኖርብ ናል።	Ykn rakkin ni humnaa ol ta’e kan biraa n n u mudatu , akka nuuf yaadu amanuu dhaan wanta yaaddoo n utti ta’u Yihowaa irratti gatu u qabna .

Table 4-1 sample morpheme generated for Amharic and Afaan Oromo.

4.3 Architecture of the system

This section is shows about the architecture of the bidirectional Amharic- Afaan Oromo statistical machine translation. The prototype of the system starting from input corpus until the translation. The architecture of the system is shown in Figure 4-1.

The architecture works through the following processes, first input corpus goes tokenization and Then the tokenized dataset divided into monolingual and bilingual dataset. To develop language model for target language we used monolingual corpus and to develop translation model we prepare training corpus for the given language pairs. (Amharic and Afaan Oromo). tuning has been implemented to maximize the translation performance. Test corpus prepared from both languages which was source text and reference text. The Decoder is used to predict words in the target language using the language model, translation model and source text and then produce the target text. Evaluation is conducted by comparing the output of translation system with reference text finally report the performance of the translation system.

We used the architecture used for word level and Morpheme level. For Morpheme, we used the architecture after morphological segmentation (used Morfessor software) for both language (Amharic and Afaan Oromo).

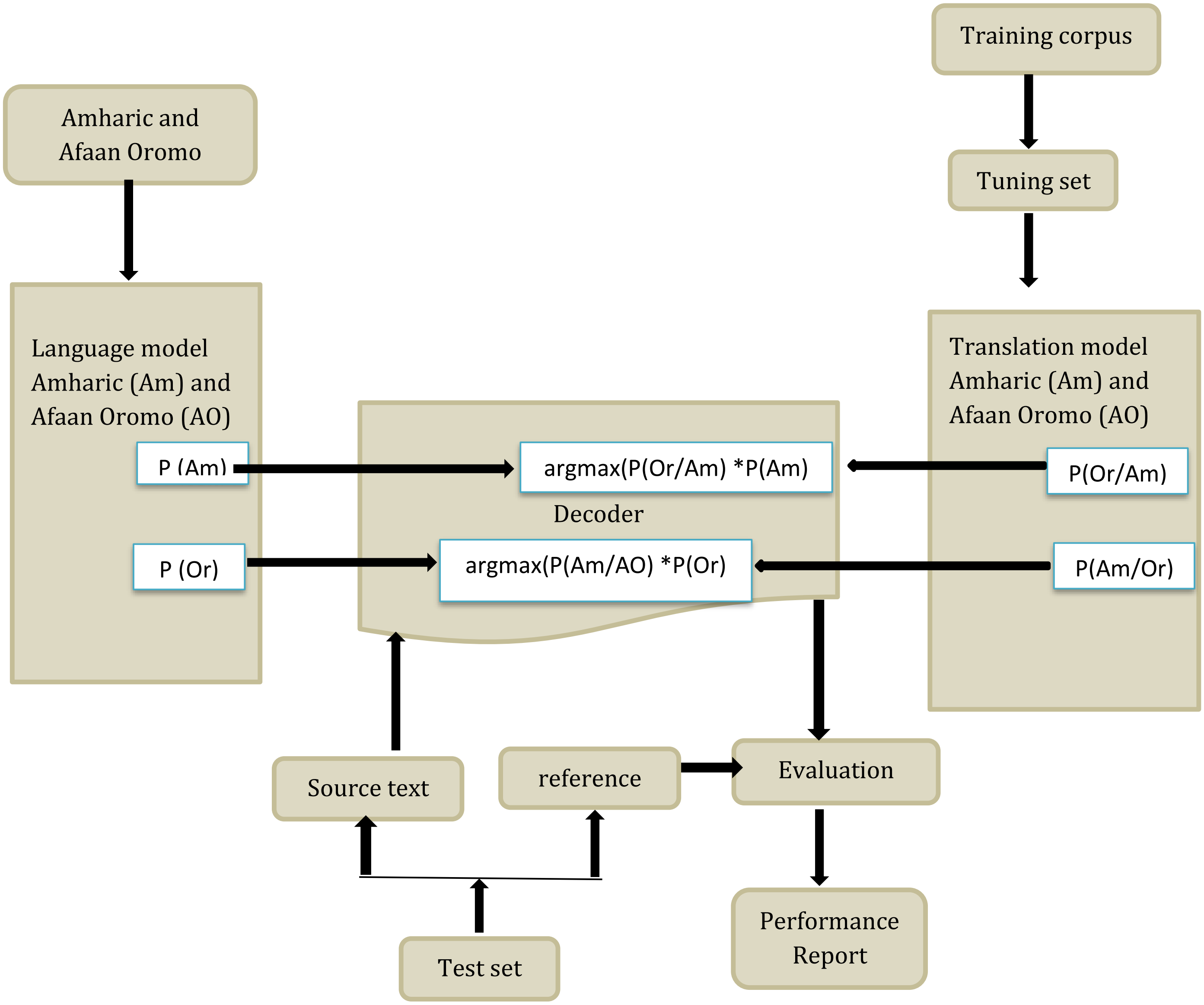


Figure 4-1 Architecture of Statistical Machine Translation system.

4.3.1 Language model

A language modelling is to estimate the probability distribution of various linguistic units, e.g., words, sentences etc. A statistical language model is a probability distribution over sequences of words. There are various software packages available to build a statistical language model. The IRSTLM language modelling toolkit is one of them and used to train the language model for this study. An appropriate 3-gram language model were built.

For the language model, we used monolingual corpora. 11680 sentences are used for both Amharic and Afaan Oromo language. It is the same amount used for both word-based and morpheme-based Machine Translation.

4.3.2 Translation Model

The translation model assigns the probability of a given source language which will generate the target language sentence. For the translation model, we used bilingual corpus.

For the translation model we used the results of MGIZA++ word level and morpheme level aligned corpus.

Alignment tools

MGIZA++ is a multi-threaded word alignment tool. It provides the concept of multithreading and memory optimization. It is used for both word and morpheme level aligned corpus for the translation model by using IBM models (1-5).

4.3.3 Decoder

The decoding is to find the translation with the best score that translates the source sentence to the corresponding target sentence. It starts by searching the phrase table for all possible translations and all possible fragments of the given sources sentences. The decoder uses feature scores and weights to select the most likely translation. It looks up all translations of every source word or phrase translation table and recombines the target language phrases that maximizes the translation model probability multiplied by the language model probability. For this study, the decoder performs the translation process from both directions.

There are many different tools for the decoding stage of the Statistical Machine Translation system. We used the Moses decoder, works with IRSTLM and MGIZA++ toolkit.

4.3.4 Tuning

In order to find the optimal weights from the given possible translation Moses tuning algorithm is used. The optimal weights are those which maximize translation performance on a small set of parallel sentences (the tuning set). We used 1460 sentence for both Amharic and Afaan Oromo sentences. The bilingual corpora used for the tuning are preprocessed with tokenization and cleaning processes.

4.3.5 Evaluation

To evaluate the performance of the prototype, using a reference translation corpus and the translation quality of the system output which was translated can be evaluated by using BLEU score.

Chapter five

5 Experimentation

5.1 Overview

Under this chapter, after design the bidirectional Amharic-Afaan Oromo statistical machine translation and preparing corpus, the experiments are conducted based on aligned word and morpheme from both directions. The experiment of the system discussed in detail as follows.

5.2 Experiment

We conduct four experiment using word and morpheme based translation unit with statistical machine translation for Amhaic –Afaan Oromo language pair. The first two experiment focus morpheme based translation using unsupervised morphological segmentation tool morfessor, and the next two experiment focuses on word based SMT. For each experiment we used 14600 sentences. Out of the total collected sentence, 80 % (11680) for language model, randomly selected sentence pairs have been used for training and from training sentence we used 10%(1460) for testing while the remaining 10% (1460) sentence pairs are used for tuning.

5.2.1 Experiment setup

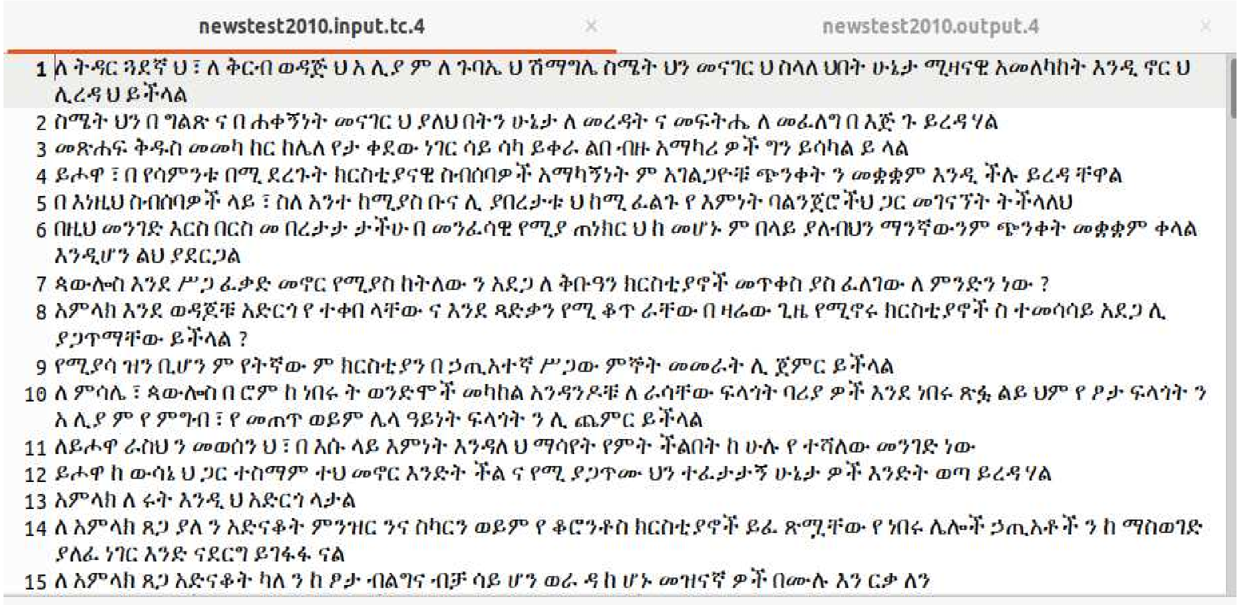
This section describes the toolkit used for conducting the experiments. The experiment of this thesis was conducted on 64-bit Linux machine (Ubuntu 20.04) as an operating system platform Manufacture HP Model, Processor Intel core i3-6006U CPU, Processor speed 2.00 GHZx4 Memory 3.7 G. For this study we used different types of tool (software), such as Moses-Decoder for translation MGIZA++ for word and morpheme alignments IRSTLM to build the language model of words and morpheme, for word segmentation we used Morfessor. The goal is to segment words into morphemes which is the smallest meaningful unit in a language. Morfessor is a family of methods for unsupervised morphological segmentation. To evaluate the output of the system for this study selected the BLEU score metric.

5.2.2 Experiment I: Morpheme-based translation from Amharic to Afaan Oromo

The first experiment is being performed to test the machine translation we used morpheme level aligned corpus. . The source language is Amharic (input text) and the target language is Afaan Oromo (output text).

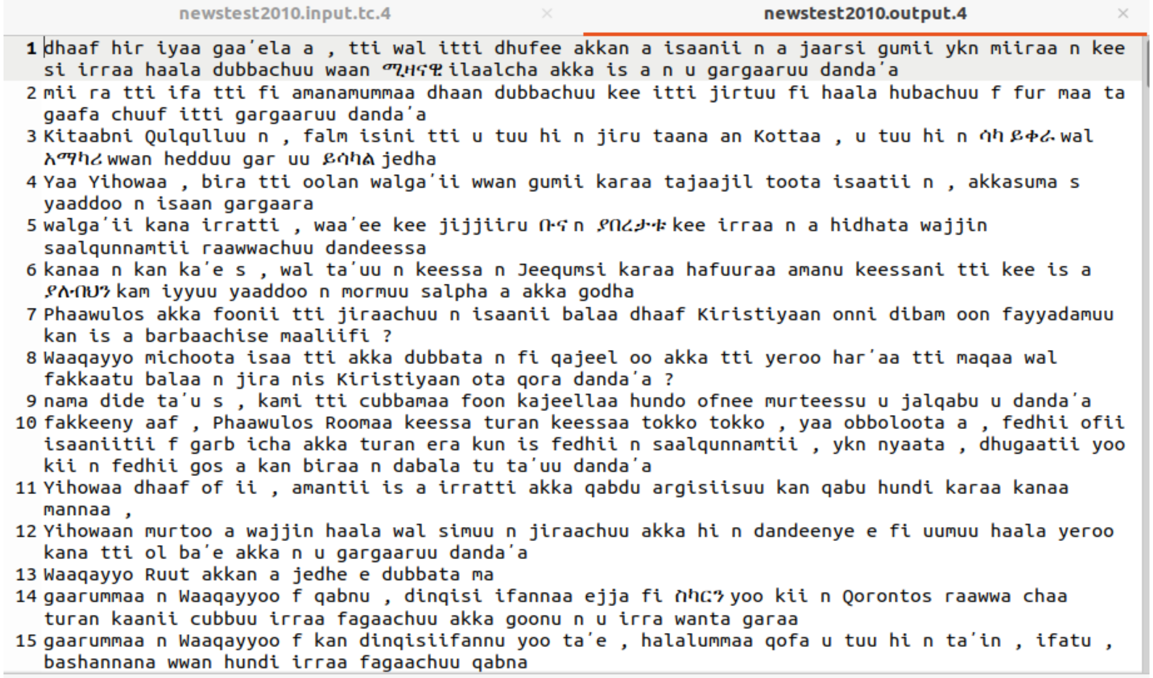
Experimental results show that the machine translation the text into the target language (Afaan Oromo). The performance of the result and BLEU score obtained were 19.77%. Figure 5-1 displays the Amharic language sample translation input text and Afaan Oromo language sample output text.

Input text is Amharic (a)



1 ለ ትዳር ጓደኛ ህ ፤ ለ ቅርብ ወዳጅ ህ አ ሊያ ም ለ ጉባኤ ህ ሽማግሌ ስሜት ህን መናገር ህ ስላለ ህበት ሁኔታ ሚዛናዊ አመለካከት እንዲ ኖር ህ ሊረዳ ህ ይችላል
2 ስሜት ህን በ ግልጽ ና በ ሐቀኝነት መናገር ህ ያለህ በትን ሁኔታ ለ መረዳት ና መፍትሔ ለ መፈለግ በ እጅ ጉ ይረዳ ሃል
3 መጽሐፍ ቅዱስ መመካ ከር ከሌለ የታ ቀደው ነገር ሳይ ሳካ ይቀራ ልበ ብዙ አማካሪ ዎች ግን ይሳካል ይ ላል
4 ይሖዋ ፤ በ የሳምንቱ በሚ ደረጉት ክርስቲያናዊ ስብሰባዎች አማካኝነት ም አገልጋዮቹ ጭንቀት ን መቋቋም እንዲ ችሉ ይረዳ ቸዋል
5 በ እነዚህ ስብሰባዎች ላይ ፤ ስለ እንተ ከሚያስቡ ና ሊ ያበረታቱ ህ ከሚ ፈልጉ የ እምነት ባልንጀሮች ህ ጋር መገናኘት ትችላለህ
6 በዚህ መንገድ እርስ በርስ መ በረታታ ታችሁ በ መንፈሳዊ የሚያ ጠነክር ህ ከ መሆኑ ም በላይ ያለብህን ማንኛውንም ጭንቀት መቋቋም ቀላል እንዲሆን ልህ ያደርጋል
7 ጳውሎስ እንደ ሥጋ ፈቃድ መኖር የሚያስ ከትለው ን አደጋ ለ ቅቡዓን ክርስቲያኖች መጥቀስ ያስ ፈለገው ለ ምንድን ነው ?
8 አምላክ እንደ ወዳጆቹ አድርጎ የ ተቀበ ላቸው ና እንደ ጻድቃን የሚ ቆጥ ራቸው በ ዛሬው ጊዜ የሚኖሩ ክርስቲያኖች ስ ተመሳሳይ አደጋ ሊ ያጋጥማቸው ይችላል ?
9 የሚያሳ ዝን ቢሆን ም የትኛው ም ክርስቲያን በ ኃጢአተኛ ሥጋው ምኞት መመራት ሊ ጀምር ይችላል
10 ለ ምሳሌ ፤ ጳውሎስ በ ሮም ከ ነበሩ ት ወንድሞች መካከል አንዳንዶቹ ለ ራሳቸው ፍላጎት ባሪያ ዎች እንደ ነበሩ ጽፏ ልይ ህም የ ፆታ ፍላጎት ን አ ሊያ ም የ ምግብ ፤ የ መጠጥ ወይም ሌላ ዓይነት ፍላጎት ን ሊ ጨምር ይችላል
11 ለይሖዋ ራስህ ን መወሰን ህ ፤ በ እሱ ላይ እምነት እንዳለ ህ ማሳየት የምት ችልበት ከ ሁሉ የ ተሻለው መንገድ ነው
12 ይሖዋ ከ ውሳኔ ህ ጋር ተስማም ተህ መኖር እንድት ችል ና የሚ ያጋጥሙ ህን ተፈታታኝ ሁኔታ ዎች እንድት ወጣ ይረዳ ሃል
13 አምላክ ለ ሩት እንዲ ህ አድርጎ ላታል
14 ለ አምላክ ጸጋ ያለ ን አድናቆት ምንዝር ንና ስካርን ወይም የ ቆሮንቶስ ክርስቲያኖች ይፈ ጽሟቸው የ ነበሩ ሌሎች ኃጢአቶች ን ከ ማስወገድ ያለፈ ነገር እንድ ናደርግ ይገፋፋ ናል
15 ለ አምላክ ጸጋ አድናቆት ካለ ን ከ ፆታ ብልግና ብቻ ሳይ ሆን ወራ ዳ ከ ሆኑ መዝናኛ ዎች በሙሉ እን ርቃ ለን

output text is Afaan Oromo (b)



```
 1 dhaaf hir iyaa gaa'ela a , tti wal itti dhufee akkan a isaanii n a jaarsi gumii ykn miiraa n kee
   si irraa haala dubbachuu waan ሚዛናዊ ilaalcha akka is a n u gargaaruu danda'a
 2 mii ra tti ifa tti fi amanamummaa dhaan dubbachuu kee itti jirtuu fi haala hubachuu f fur maa ta
   gaafa chuuf itti gargaaruu danda'a
 3 Kitaabni Qulqulluu n , falm isini tti u tuu hi n jiru taana an Kottaa , u tuu hi n ሳካ ይቀራ wal
   አማካሪ wwan hedduu gar uu ይሳካል jedha
 4 Yaa Yihowaa , bira tti oolan walga'ii wwan gumii karaa tajaajil toota isaatii n , akkasuma s
   yaaddoo n isaan gargaara
 5 walga'ii kana irratti , waa'ee kee jijjiiru ቡና n ያበረታቱ kee irraa n a hidhata wajjin
   saalqunnamtii raawwachuu dandeessa
 6 kanaa n kan ka'e s , wal ta'uu n keessa n Jeequmsi karaa hafuuraa amanu keessani tti kee is a
   ያሉበህን kam iyyuu yaaddoo n mormuu salpha a akka godha
 7 Phaawulos akka foonii tti jiraachuu n isaanii balaa dhaaf Kiristiyaan onni dibam oon fayyadamuu
   kan is a barbaachise maaliifi ?
 8 Waaqayyo michoota isaa tti akka dubbata n fi qajeel oo akka tti yeroo har'aa tti maqaa wal
   fakkaatu balaa n jira nis Kiristiyaan ota qora danda'a ?
 9 nama dide ta'u s , kami tti cubbamaa foon kajeellaa hundo ofnee murteessu u jalqabu u danda'a
10 fakkeeny aaf , Phaawulos Roomaa keessa turan keessaa tokko tokko , yaa obboloota a , fedhii ofii
   isaaniitii f garb icha akka turan era kun is fedhii n saalqunnamtii , ykn nyaata , dhugaatii yoo
   kii n fedhii gos a kan biraa n dabala tu ta'uu danda'a
11 Yihowaa dhaaf of ii , amantii is a irratti akka qabdu argisiisuu kan qabu hundi karaa kanaa
   mannaa ,
12 Yihowaan murtoo a wajjin haala wal simuu n jiraachuu akka hi n dandeenye e fi uumuu haala yeroo
   kana tti ol ba'e akka n u gargaaruu danda'a
13 Waaqayyo Ruut akkan a jedhe e dubbata ma
14 gaarummaa n Waaqayyoo f qabnu , dinqisi ifannaa ejja fi ስካርን yoo kii n Qorontos raawwa chaa
   turan kaanii cubbuu irraa fagaachuu akka goonu n u irra wanta garaa
15 gaarummaa n Waaqayyoo f kan dinqisiifannu yoo ta'e , halalummaa qofa u tuu hi n ta'in , ifatu ,
   bashannana wwan hundi irraa fagaachuu qabna
```

Figure 5-1 Sample translation input (a) and output (b) for Amharic to Afaan Oromo translation morpheme level alignment.

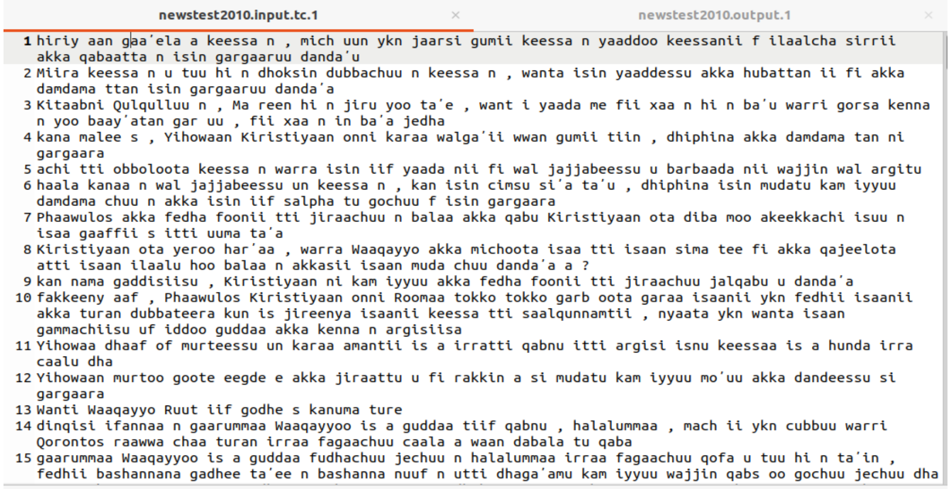
As we have seen from the above figure 5-1(a) there is over and under word segmentation; for example, line 1“ አሊያም” should be segmented “አሊያም ” but you can see it segmented “አ ሊያ ም ” which is over segmented. In the figure 5-1 (a) line 5 “ስብሰባዎች” should be segemented to “ስብሰባ ዎች” but it is segmentation to “ስብሰባዎ ች” which is under segmentation. There are also words that are still unsegmented, for example, in the 5 line“ባልንጀሮችህ”, and also there are words that are segmented which is not need segmentation, for example line 15 “ወሯ ዳ” . Finally there is also perfect segmentation like line 15 “መዝናኛዎች ” to “መዝናኛ ዎች ”

As we have seen the output of the experiment there are sentences or words are untranslated into Afaan Oromo such as “ሚዛናዊ” in the first sentence, “አማካሪ”and,“ይሳካል”in line 3 and “ስካርን” in line 14.These occurred because of alignment problem the source language text and the target language text not align perfectly.

5.2.3 Experiment II: Morpheme-based translation from Afaan Oromo to Amharic

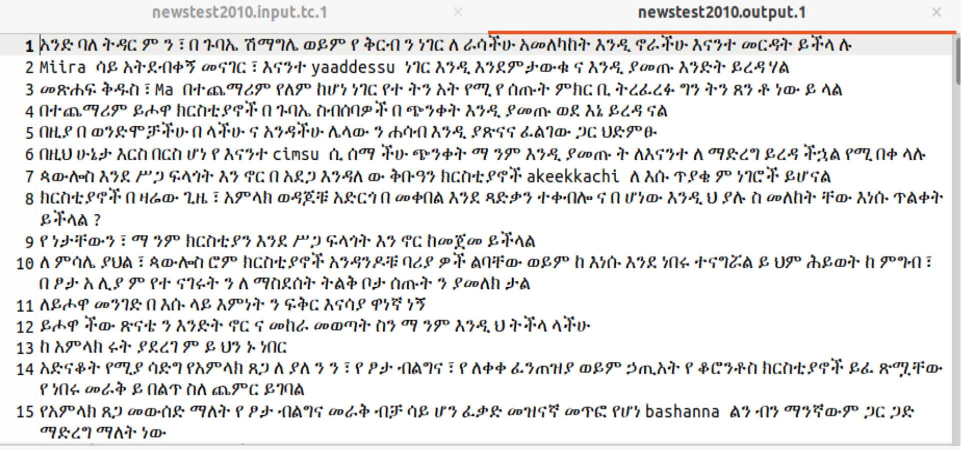
The system works bi-directional so this experiment checks the performance of the system with the same corpus used in the experiment I. This experiment checks the performance of the system with the same corpus used in the experiment I based on morpheme translation. The experiment performed from source language Afaan Oromo to target language Amharic.

The input text is Afaan Oromo (a)



```
 1 hiriy aan gaa'ela a keessa n , mich uun ykn jaarsi gumii keessa n yaaddoo keessanii f ilaalcha sirrii
   akka qabaatta n isin gargaaruu danda'u
 2 Miira keessa n u tuu hi n dhoksin dubbachuu n keessa n , wanta isin yaaddessu akka hubattan ii fi akka
   damdama ttan isin gargaaruu danda'a
 3 Kitaabni Qulqulluu n , Ma reen hi n jiru yoo ta'e , want i yaada me fii xaa n hi n ba'u warri gorsa kenna
   n yoo baay'atan gar uu , fii xaa n in ba'a jedha
 4 kana malee s , Yihowaan Kiristiyaan onni karaa walga'ii wwan gumii tiin , dhiphina akka damdama tan ni
   gargaara
 5 achi tti obboloota keessa n warra isin iif yaada nii fi wal jajjabeessu u barbaada nii wajjin wal argitu
 6 haala kanaa n wal jajjabeessu un keessa n , kan isin cimsu si'a ta'u , dhiphina isin mudatu kam iyyuu
   damdama chuu n akka isin iif salpha tu gochuu f isin gargaara
 7 Phaawulos akka fedha foonii tti jiraachuu n balaa akka qabu Kiristiyaan ota diba moo akeekkachi isuu n
   isaa gaaffii s itti uuma ta'a
 8 Kiristiyaan ota yeroo har'aa , warra Waaqayyo akka michoota isaa tti isaan sima tee fi akka qajeelota
   atti isaan ilaalu hoo balaa n akkasii isaan muda chuu danda'a a ?
 9 kan nama gaddisiisu , Kiristiyaan ni kam iyyuu akka fedha foonii tti jiraachuu jalqabu u danda'a
10 fakkeeny aaf , Phaawulos Kiristiyaan onni Roomaa tokko tokko garb oota garaa isaanii ykn fedhii isaanii
   akka turan dubbateera kun is jireenya isaanii keessa tti saalqunnamtii , nyaata ykn wanta isaan
   gammachiisu uf iddoo guddaa akka kenna n argisiisa
11 Yihowaa dhaaf of murteessu un karaa amantii is a irratti qabnu itti argisi isnu keessaa is a hunda irra
   caalu dha
12 Yihowaan murtoo goote eegde e akka jiraattu u fi rakkin a si mudatu kam iyyuu mo'uu akka dandeessu si
   gargaara
13 Wanti Waaqayyo Ruut iif godhe s kanuma ture
14 dinqisi ifannaa n gaarummaa Waaqayyoo is a guddaa tiif qabnu , halalummaa , mach ii ykn cubbuu warri
   Qorontos raawwa chaa turan irraa fagaachuu caala a waan dabala tu qaba
15 gaarummaa Waaqayyoo is a guddaa fudhachuu jechuu n halalummaa irraa fagaachuu qofa u tuu hi n ta'in ,
   fedhii bashannana gadhee ta'ee n bashanna nuuf n utti dhaga'amu kam iyyuu wajjin qabs oo gochuu jechuu dha
```

The Output text is Amharic (b)



```
 1 አንድ ባለ ትዳር ም ን ፣ በ ጉባኤ ሽማግሌ ወይም የ ቅርብ ን ነገር ለ ራሳችሁ አመለካከት እንዲ ኖራችሁ እናንተ መርዳት ይችላ ሉ
 2 Miira ሳይ አትደብቀኝ መናገር ፣ እናንተ yaaddessu ነገር እንዲ እንደምታውቁ ና እንዲ ያመጡ እንድት ይረዳ ሃል
 3 መጽሐፍ ቅዱስ ፣ Ma በተጨማሪም የለም ከሆነ ነገር የተ ትን አት የሚ የ ሰጡት ምክር ቢ ትረፈረፉ ግን ትን ጸን ቶ ነው ይ ላል
 4 በተጨማሪም ይሖዋ ክርስቲያኖች በ ጉባኤ ስብሰባዎች በ ጭንቀት እንዲ ያመጡ ወደ እኔ ይረዳ ናል
 5 በዚያ በ ወንድሞቻችሁ በ ላችሁ ና እንዳችሁ ሌላው ን ሐሳብ እንዲ ያጽናና ፈልገው ጋር ህድምፁ
 6 በዚህ ሁኔታ እርስ በርስ ሆነ የ እናንተ cimsu ሲ ሰማ ችሁ ጭንቀት ማ ንም እንዲ ያመጡ ት ለእናንተ ለ ማድረግ ይረዳ ችኋል የሚ በቀ ላሉ
 7 ጳውሎስ እንደ ሥጋ ፍላጎት እን ኖር በ አደጋ እንዳለ ው ቅቡዓን ክርስቲያኖች akeekkachi ለ እሱ ጥያቄ ም ነገሮች ይሆናል
 8 ክርስቲያኖች በ ዛሬው ጊዜ ፣ አምላክ ወዳጆቹ አድርጎ በ መቀበል እንደ ጻድቃን ተቀብሎ ና በ ሆነው እንዲ ህ ያሉ ስ መለከት ቸው እነሱ ጥልቀት
   ይችላል ?
 9 የ ነታቸውን ፣ ማ ንም ክርስቲያን እንደ ሥጋ ፍላጎት እን ኖር ከመጀመ ይችላል
10 ለ ምሳሌ ያህል ፣ ጳውሎስ ሮም ክርስቲያኖች አንዳንዶቹ ባሪያ ዎች ልባቸው ወይም ከ እነሱ እንደ ነበሩ ተናግሯል ይ ህም ሕይወት ከ ምግብ ፣
   በ ፆታ አ ሊያ ም የተ ናገሩት ን ለ ማስደሰት ትልቅ ቦታ ሰጡት ን ያመለክ ታል
11 ለይሖዋ መንገድ በ እሱ ላይ እምነት ን ፍቅር እናሳያ ዋነኛ ነኝ
12 ይሖዋ ችው ጽናቲ ን እንድት ኖር ና መከራ መወጣት ስን ማ ንም እንዲ ህ ትችላ ላችሁ
13 ከ አምላክ ሩት ያደረገ ም ይ ህን ኑ ነበር
14 አድናቆት የሚያ ሳድግ የአምላክ ጸጋ ለ ያለ ን ን ፣ የ ፆታ ብልግና ፣ የ ለቀቀ ፈንጠዝያ ወይም ኃጢአት የ ቆሮንቶስ ክርስቲያኖች ይፈ ጽሚቸው
   የ ነበሩ መረቅ ይ በልጥ ስለ ጨምር ይገባል
15 የአምላክ ጸጋ መውሰድ ማለት የ ፆታ ብልግና መረቅ ብቻ ሳይ ሆን ፈቃድ መዝናኛ መጥፎ የሆነ bashanna ልን ብን ማንኛውም ጋር ጋድ
   ማድረግ ማለት ነው
```

Figure 5-2 Sample translation input (a) and output (b) for Afaan Oromo to Amharic translation morpheme level alignment

From the above experiment due to alignments problem of the corpus we used morpheme level aligned some of the word and sub words in the paragraph are not translated. The result recorded from the translation process BLEU Score was 16.14 % for the Afaan Oromo to Amharic translation.

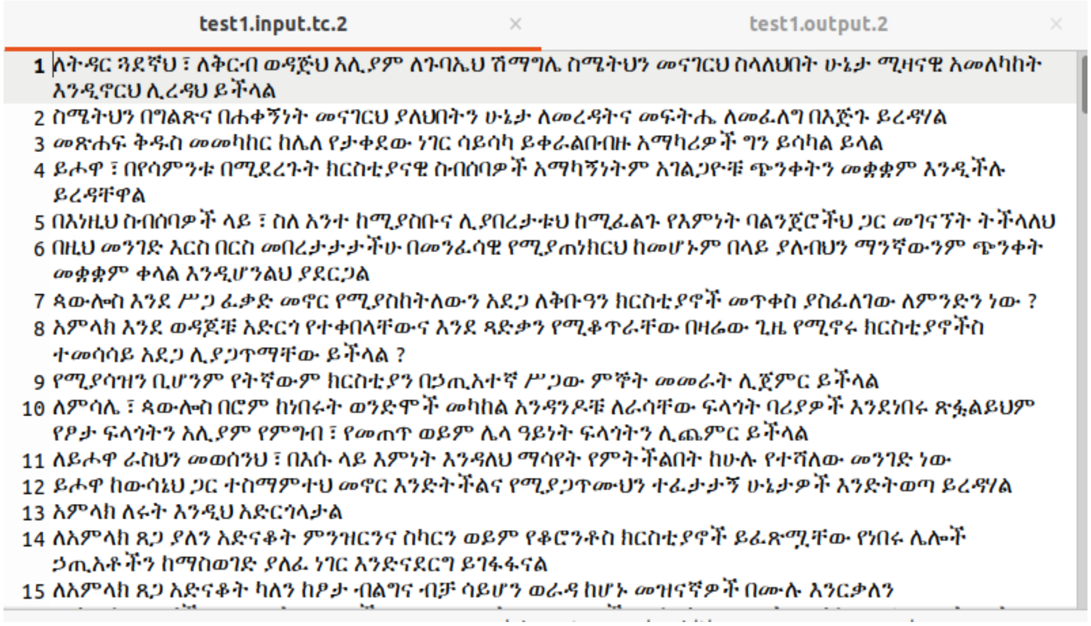
As you can see from the above figure 5-2 there are under and over segmentation problems for poor performance of the system.

From the above two experiments (I and II) we consider that the system achieves better performance when Amharic is source language, because of, alignment probability of the words.

5.2.4 Experiment III: word-based translation from Amharic to Afaan Oromo

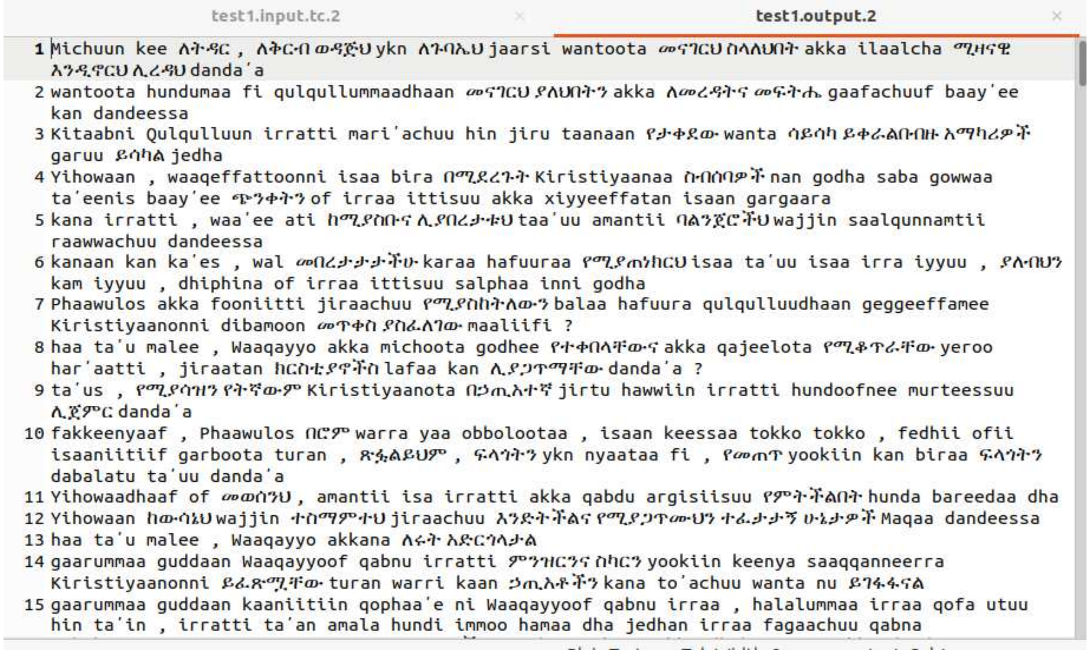
The III experiment shows the word-based aligned translation. The experiment conducted by taking Amharic as source language and Afaan Oromo as target language. We used the same Amharic input text as Experiment I, the result of the experiment shown in the figure 5-3:

The Input text is Amharic (a)



1 ለትዳር ጓደኛህ ፣ ለቅርብ ወዳጅህ አሊያም ለጉባኤህ ሽማግሌ ስሜትህን መናገርህ ስላለህበት ሁኔታ ሚዛናዊ አመለካከት እንዲኖርህ ሊረዳህ ይችላል
2 ስሜትህን በግልጽና በሐቀኝነት መናገርህ ያለህበትን ሁኔታ ለመረዳትና መፍትሔ ለመፈለግ በእጅጉ ይረዳሃል
3 መጽሐፍ ቅዱስ መመካከር ከሌለ የታቀደው ነገር ሳይሳካ ይቀራልበብዙ አማካሪዎች ግን ይሳካል ይላል
4 ይሖዋ ፣ በየሳምንቱ በሚደረጉት ክርስቲያናዊ ስብሰባዎች አማካኝነትም አገልጋዮቹ ጭንቀትን መቋቋም እንዲችሉ ይረዳቸዋል
5 በእነዚህ ስብሰባዎች ላይ ፣ ስለ አንተ ከሚያስቡና ሊያበረታቱህ ከሚፈልጉ የእምነት ባልንጀሮችህ ጋር መገናኘት ትችላለህ
6 በዚህ መንገድ እርስ በርስ መበረታታታችሁ በመንፈሳዊ የሚያጠነክርህ ከመሆኑም በላይ ያለብህን ማንኛውንም ጭንቀት መቋቋም ቀላል እንዲሆንልህ ያደርጋል
7 ጳውሎስ እንደ ሥጋ ፈቃድ መኖር የሚያስከትለውን አደጋ ለቅቡዓን ክርስቲያኖች መጥቀስ ያስፈለገው ለምንድን ነው ?
8 አምላክ እንደ ወዳጆቹ አድርጎ የተቀበላቸውና እንደ ጻድቃን የሚቆጥራቸው በዛሬው ጊዜ የሚኖሩ ክርስቲያኖችስ ተመሳሳይ አደጋ ሊያጋጥማቸው ይችላል ?
9 የሚያሳዝን ቢሆንም የትኛውም ክርስቲያን በኃጢአተኛ ሥጋው ምኞት መመራት ሊጀምር ይችላል
10 ለምሳሌ ፣ ጳውሎስ በሮም ከነበሩት ወንድሞች መካከል አንዳንዶቹ ለራሳቸው ፍላጎት ባሪያዎች እንደነበሩ ጽፏልይህም የፆታ ፍላጎትን አሊያም የምግብ ፣ የመጠጥ ወይም ሌላ ዓይነት ፍላጎትን ሊጨምር ይችላል
11 ለይሖዋ ራስህን መወሰንህ ፣ በእሱ ላይ እምነት እንዳለህ ማሳየት የምትችልበት ከሁሉ የተሻለው መንገድ ነው
12 ይሖዋ ከውሳኔህ ጋር ተስማምተህ መኖር እንድትችልና የሚያጋጥሙህን ተፈታታኝ ሁኔታዎች እንድትወጣ ይረዳሃል
13 አምላክ ለሩት እንዲህ አድርጎላታል
14 ለአምላክ ጸጋ ያለን አድናቆት ምንዝርንና ስካርን ወይም የቆሮንቶስ ክርስቲያኖች ይፈጽሟቸው የነበሩ ሌሎች ኃጢአቶችን ከማስወገድ ያለፈ ነገር እንድናደርግ ይገፋፋናል
15 ለአምላክ ጸጋ አድናቆት ካለን ከፆታ ብልግና ብቻ ሳይሆን ወራዳ ከሆኑ መዝናኛዎች በሙሉ እንርቃለን

The output text is Afaan Oromo (b)



1 Michuun kee ለትዳር , ለቅርብ ወዳጅህ ykn ለጉባኤህ jaarsi wantoota መናገርህ ስላለህበት akka ilaalcha ሚዛናዊ እንዲኖርህ ሊረዳህ danda'a

2 wantoota hundumaa fi qulqullummaadhaan መናገርህ ያለህበትን akka ለመረዳትና መፍትሔ gaafachuuf baay'ee kan dandeessa

3 Kitaabni Qulqulluun irratti mari'achuu hin jiru taanaan የታቀደው wanta ሳይሳካ ይቀራልቡበዙ አማካሪዎች garuu ይሳካል jedha

4 Yihowaan , waaqeffattoonni isaa bira በሚደረጉት Kiristiyaanaa ስብሰባዎች nan godha saba gowwaa ta'eenis baay'ee ጭንቀትን of irraa ittisuu akka xiyyeeffatan isaan gargaara

5 kana irratti , waa'ee ati ከሚያስቡና ሊያበረታቱህ taa'uu amantii ባልንጀሮችህ wajjin saalqunnamtii raawwachuu dandeessa

6 kanaan kan ka'es , wal መበረታታታችሁ karaa hafuuraa የሚያጠነክርህ isaa ta'uu isaa irra iyyuu , ያሉብህን kam iyyuu , dhiphina of irraa ittisuu salphaa inni godha

7 Phaawulos akka fooniitti jiraachuu የሚያስከትለውን balaa hafuura qulqulluudhaan geggeeffamee Kiristiyaanonni dibamoon መጥቀስ ያስፈለገው maaliifi ?

8 haa ta'u malee , Waaqayyo akka michoota godhee የተቀበላቸውና akka qajeelota የሚቆጥራቸው yeroo har'aatti , jiraatan ክርስቲያኖችስ lafaa kan ሊያጋጥማቸው danda'a ?

9 ta'us , የሚያሳዝን የትኛውም Kiristiyaanota በኃጢአተኛ jirtu hawwiin irratti hundoofnee murteessuu ሊጀምር danda'a

10 fakkeenyaaf , Phaawulos በሮም warra yaa obbolootaa , isaan keessaa tokko tokko , fedhii ofii isaaniitiif garboota turan , ጽፏልይህም , ፍላጎትን ykn nyaataa fi , የመጠጥ yookiin kan biraa ፍላጎትን dabalatu ta'uu danda'a

11 Yihowaadhaaf of መወሰንህ , amantii isa irratti akka qabdu argisiisuu የምትችልበት hunda bareedaa dha

12 Yihowaan ከውሳኔህ wajjin ተስማምተህ jiraachuu እንድትችልና የሚያጋጥሙህን ተፈታታኝ ሁኔታዎች Maqaa dandeessa

13 haa ta'u malee , Waaqayyo akkana ለሩት አድርጎላታል

14 gaarummaa guddaan Waaqayyoof qabnu irratti ምንዝርንና ስካርን yookiin keenya saaqqanneerra Kiristiyaanonni ይፈጽሟቸው turan warri kaan ኃጢአቶችን kana to'achuu wanta nu ይገፋፋናል

15 gaarummaa guddaan kaaniitiin qophaa'e ni Waaqayyoof qabnu irraa , halalummaa irraa qofa utuu hin ta'in , irratti ta'an amala hundi immoo hamaa dha jedhan irraa fagaachuu qabna

Figure 5-3 Sample translation input (a) and output (b) for Amharic to Afaan Oromo translation word level alignment.

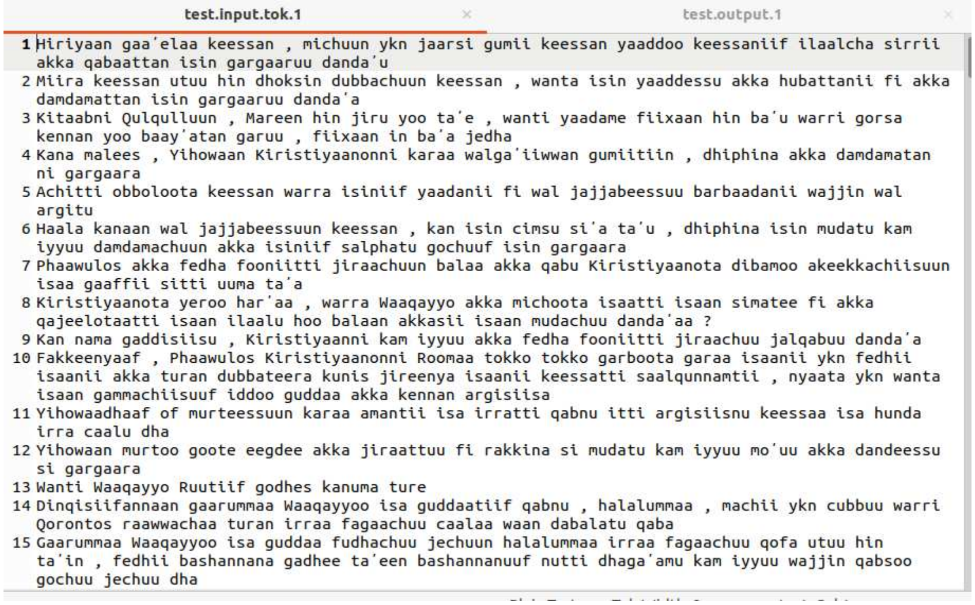
The BLEU score recorded for this experiment is 13.84 %. When we compare this result with experiment I, the performance of experiment I (morphem based Amharic to Afaan Oromo) the translation system is enhanced by 5.93%. Due to morpheme level alignment quality the result is improved during the training of system.

5.2.5 Experiment IV: word -based translation from Afaan Oromo to Amharic

In this experiment is word-based translation is done using Afaan Oromo and Amharic as the source and target languages respectively.

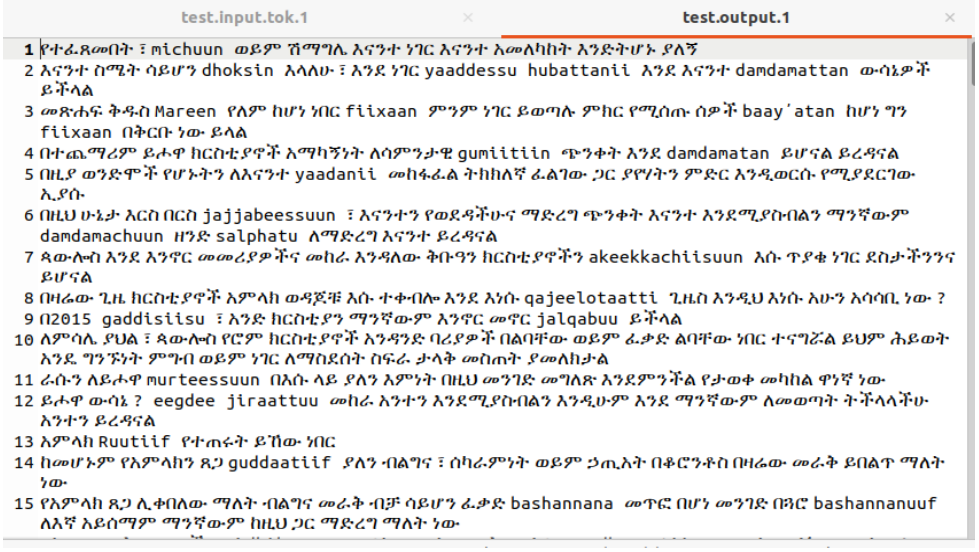
We used the same corpus input text with experiment II, the translation model is trained by using word level aligned corpus like experiment III. The Sample result of the experiment shown in the figure 5-4.

Input text is Afaan Oromo (a)



```
1 Hiriyaan gaa'elaa keessan , michuun ykn jaarsi gumii keessan yaaddoo keessaniif ilaalcha sirrii akka qabaattan isin gargaaruu danda'u
2 Miira keessan utuu hin dhoksin dubbachuun keessan , wanta isin yaaddessu akka hubattanii fi akka damdamattan isin gargaaruu danda'a
3 Kitaabni Qulqulluun , Mareen hin jiru yoo ta'e , wanti yaadame fiixaan hin ba'u warri gorsa kennan yoo baay'atan garuu , fiixaan in ba'a jedha
4 Kana malees , Yihowaan Kiristiyaanonni karaa walga'iiwwan gumiitiin , dhiphina akka damdamatan ni gargaara
5 Achitti obboloota keessan warra isiniif yaadanii fi wal jajjabeessuu barbaadanii wajjin wal argitu
6 Haala kanaan wal jajjabeessuun keessan , kan isin cimsu si'a ta'u , dhiphina isin mudatu kam iyyuu damdamachuun akka isiniif salphatu gochuuf isin gargaara
7 Phaawulos akka fedha fooniitti jiraachuun balaa akka qabu Kiristiyaanota dibamoo akeekkachiisuun isaa gaaffii sitti uuma ta'a
8 Kiristiyaanota yeroo har'aa , warra Waaqayyo akka michoota isaatti isaan simatee fi akka qajeelotaatti isaan ilaalu hoo balaan akkasii isaan mudachuu danda'aa ?
9 Kan nama gaddisiisu , Kiristiyaanni kam iyyuu akka fedha fooniitti jiraachuu jalqabuu danda'a
10 Fakkeenyaaf , Phaawulos Kiristiyaanonni Roomaa tokko tokko garboota garaa isaanii ykn fedhii isaanii akka turan dubbateera kunis jireenya isaanii keessatti saalqunnamtii , nyaata ykn wanta isaan gammachiisuuf iddoo guddaa akka kennan argisiisa
11 Yihowaadhaaf of murteessuun karaa amantii isa irratti qabnu itti argisiisnu keessaa isa hunda irra caalu dha
12 Yihowaan murtoo goote eegdee akka jiraattuu fi rakkina si mudatu kam iyyuu mo'uu akka dandeessu si gargaara
13 Wanti Waaqayyo Ruutiif godhes kanuma ture
14 Dinqisiifannaan gaarummaa Waaqayyoo isa guddaatiif qabnu , halalummaa , machii ykn cubbuu warri Qorontos raawwachaa turan irraa fagaachuu caalaa waan dabalatu qaba
15 Gaarummaa Waaqayyoo isa guddaa fudhachuu jechuun halalummaa irraa fagaachuu qofa utuu hin ta'in , fedhii bashannana gadhee ta'een bashannanuuf nutti dhaga'amu kam iyyuu wajjin qabsoo gochuu jechuu dha
```

Output text is Amharic (b)



```
1 የተፈጸመበት ፣ michuun ወይም ሽማግሌ እናንተ ነገር እናንተ አመለካከት እንድትሆኑ ያለኝ
2 እናንተ ስሜት ሳይሆን dhoksin አላለሁ ፣ እንደ ነገር yaaddessu hubattanii እንደ እናንተ damdamattan ውሳኔዎች ይችላል
3 መጽሐፍ ቅዱስ Mareen የለም ከሆነ ነበር fiixaan ምንም ነገር ይወጣሉ ምክር የሚሰጡ ሰዎች baay'atan ከሆነ ግን fiixaan በቅርቡ ነው ይላል
4 በተጨማሪም ይሖዋ ክርስቲያኖች አማካኝነት ለሳምንታዊ gumiitiin ጭንቀት እንደ damdamatan ይሆናል ይረዳናል
5 በዚያ ወንድሞች የሆኑትን ለእናንተ yaadanii መከፋፈል ትክክለኛ ፈልገው ጋር ያየሃትን ምድር እንዲወርሱ የሚያደርገው ኢያሱ
6 በዚህ ሁኔታ እርስ በርስ jajjabeessuun ፣ እናንተን የወደዳችሁና ማድረግ ጭንቀት እናንተ እንደሚያስብልን ማንኛውም damdamachuun ዘንድ salphatu ለማድረግ እናንተ ይረዳናል
7 ጳውሎስ እንደ እንኖር መመሪያዎችና መከራ እንዳለው ቅቡዓን ክርስቲያኖችን akeekkachiisuun እሱ ጥያቄ ነገር ደስታችንንና ይሆናል
8 በዛሬው ጊዜ ክርስቲያኖች አምላክ ወዳጆቹ እሱ ተቀብሎ እንደ እነሱ qajeelotaatti ጊዜስ እንዲህ እነሱ አሁን አሳሳቢ ነው ?
9 በ2015 gaddisiisu ፣ አንድ ክርስቲያን ማንኛውም እንኖር መኖር jalqabuu ይችላል
10 ለምሳሌ ያህል ፣ ጳውሎስ የሮም ክርስቲያኖች አንዳንድ ባሪያዎች በልባቸው ወይም ፈቃድ ልባቸው ነበር ተናግሯል ይህም ሕይወት እንደ ግንኙነት ምግብ ወይም ነገር ለማስደሰት ስፍራ ታላቅ መስጠት ያመለክታል
11 ራሱን ለይሖዋ murteessuun በእሱ ላይ ያለን እምነት በዚህ መንገድ መግለጽ እንደምንችል የታወቀ መካከል ዋነኛ ነው
12 ይሖዋ ውሳኔ ? eegdee jiraattuu መከራ እንተን እንደሚያስብልን እንዲሁም እንደ ማንኛውም ለመወጣት ትችላላችሁ እንተን ይረዳናል
13 አምላክ Ruutiif የተጠሩት ይኸው ነበር
14 ከመሆኑም የአምላክን ጸጋ guddaatiif ያለን ብልግና ፣ ሰካራምነት ወይም ኃጢአት በቆሮንቶስ በዛሬው መሬት ይበልጥ ማለት ነው
15 የአምላክ ጸጋ ሊቀበለው ማለት ብልግና መሬት ብቻ ሳይሆን ፈቃድ bashannana መጥፎ በሆነ መንገድ በጓሮ bashannanuuf ለእኛ አይሰማም ማንኛውም ከዚህ ጋር ማድረግ ማለት ነው
```

Figure 5-4 Sample translation input (a) and output (b) for Afaan Oromo to Amharic translation word level alignment.

The BLEU score recorded for this experiment is 9.72`%. When we compare this result with experiment II result it is poorer because of morpheme level alignment better than word level alignment,

5.3 Discussion of Result

The main purpose of this study is to conduct experiment on bi-directional Amharic -Afaan Oromo, statistical machine translation to explore the effects of word and morpheme alignment for better performance of statistical machine translation in both directions.

We conduct two group experiment which are morpheme and word-Based alignment. The first and second experiment conduct morpheme based using unsupervised morphological segmentation tools Morfessor (from Amharic-Afaan Oromo and Afaan Oromo-Amharic) and the third and fourth experiment conduct word-based alignment. Summary of the experimental result is presented in table5-1 below.

Types of experiment	Result of experiment in BLEU score	
	Amharic to Afaan Oromo	Afaan Oromo to Amharic
morpheme based alignment	19.77%	16.14 %
word-based alignment	13.84%	9.72%

Table 5-1 summary of the experiment result

As we have seen the above table 5-1 for both language (Amharic and Afaan Oromo) unsupervised morpheme-based segmentation translation performed better than word-based with performance improvement of greater than 5.93% and 6.42% BLEU score for Amharic to Afaan Oromo and Afaan Oromo to Amharic respectively. Due to many words aligned are challenged between two morphological rich languages, however morpheme level alignment is not challenged because morphemes are specific this affect the performance of statistical machine translation therefore, this is the challenge for this study.

Regarding direction of translation as Amharic to Afaan Oromo pair machine translation performs better than Afaan Oromo to Amharic machine translation the result recorded BLEU score shows that the unsupervised morpheme-based approach BLEU score 3.63% and 4.12% BLEU score word-based approach.

When we consider all the experiments better BLEU score is achieved or recorded when Amharic is used as source language and Afaan Oromo as target language. This is because of alignment quality is better when Amharic is used as source language whether we used morpheme level alignment and word level alignment during the training of system. Behind those experimental Due to one word Amharic to many Afaan Oromo words alignment affect the performance, this also another challenge observed for this study.

Chapter Six

6 Conclusion and Recommendation

6.1 Conclusion

The purpose of this study was developed bidirectional of Amharic- Afaan Oromo machine translation using statistical approach. In this study we explored word and sub-words called morpheme alignment by considering Amharic and Afaan Oromo language and vice versa translation. In order to explore the alignment first, we studied the sentence structure of both Amharic and Afaan Oromo language. We conduct the experiment by aligning the corpus at morpheme (sub-word) and word level of alignment.

The development process of Amharic – Afaan Oromo a statistical machine translation involves collecting parallel corpus from freely available online sources such as Old, new Testament holly bible and other spiritual book and then corpus preparation which also involves dividing the corpus for training set, tunning set and test set.

We used Morfessor to unsupervised segment morpheme of Amharic and Afaan Oromo Language. MGIZA++ used for word and morpheme level alignment. Moses for used for translation process which integrate all necessary tools for machine translation such as IRSTLM, MGIZA++ and decode.

After designing different experiments are conducted under taking morpheme and word level of alignment. The unsupervised morpheme segmentation experiment it has BLEU score 19.77 % and 16.14 % for Amharic to Afaan Oromo and Afaan Oromo to Amharic translation respectively while word based experiment BLEU score is 13.84 % for Amharic to Afaan Oromo and 9.72`% for Afaan Oromo to Amharic.

Generally, we show that using unsupervised morphological segmentation (morpheme) translation for the Amharic and Afaan Oromo language can improve output translation scores rather than word based and when the source and the target languages are Amharic and Afaan Oromo. This study achieves a promising result that identifies morpheme is the best unit to translate Amharic-Afaan Oromo machine translation system in both directions and it enhances the performance of bi-directional Amharic-Afaan Oromo machine translation.

Both Amharic and Afaan Oromo are morphologically rich languages However, conducting machine translation between those languages, there are a number of challenges observed. One of the challenges is mis-alignment between words/morphemes. The other reason for this is the size of data used for the training, as the larger the size of the corpus used the better the machine is able to do a high-quality translation.

6.2 Recommendation

Statistical machine translation is one of corpus-based approach for translation. This study explores morpheme and word-based bi-directional machine translation for Amharic-Afaan Oromo languages. Based on the above findings, we would like to recommend the following areas could be explored further as a continuation of this study.

- The corpus taken for this study cannot be enough for future research should be conducted using a large set of corpora for better result.
- All corpus used for this study is collected from holly bible and religious documents Therefore, the researcher recommends should be conducted and prepared from different disciplines or domain.
- In this study we use Morfessor for unsupervised morphological segmentation for both Amharic and Afaan Oromo language. Both languages Amharic and Afaan Oromo are morphological rich, therefore, need to be apply rule based morphological segmentation or machine learning algorithms for designing an optimal model for segmentation.
- Better results can be achieved by using the corpus with proper alignment used for training the system. So, by increasing the size of the training data set that properly aligned at morpheme level one can develop a better bi-directional Amharic-Afaan Oromo machine translation.

7 References

[1] P. J. V. DR. GEORGIA BRONI, “Communication cycle: Definition, process, models and examples,” 2020. [Online]. Available: https://pdfs.semanticscholar.org/da4e/69265653057d6f03fdc4ce3692b4e6923a0f.pdf.

[2] Britannica.com, “Ethnic groups and languages.” https://www.britannica.com/place/Ethiopia/Ethnic-groups-and-languages.

[3] T. Kassa, “Morpheme-Based Bi-Directional Ge’ez -Amharic Machine Translation,” Addis Ababa University, Addis Ababa, Ethiopia, 2018.

[4] Y. Solomon, “Optimal Alignment for Bi-directional Afaan Oromo-English Statistical Machine Translation,” Addis Abeba university, Addis Ababa, Ethiopia, 2017.

[5] M. Irfan, “MachineTranslation,” Department of Computer Science Bahria University Islamabad, Oct. 2017, [Online]. Available: https://www.researchgate.net/publication/320730405.

[6] P. Koehn, Statistical Machine Translation. CAMBRIDGE UNIVERSITY PRESS, 2010.

[7] M. D. Okpor, “Machine Translation Approaches: Issues and Challenges,” Department of Computer Science, Delta State Polytechnic, OzoroDelta State, Nigeria, Sep. 2014, [Online]. Available: https://www.ijcsi.org/papers/IJCSI-11-5-2-159-165.pdf.

[8] P. Bhattacharyya, Machine Translation. Indian Institute of Technology Bombay Mumbai, India., 2015.

[9] E. Teshome, “Bidirectional English-Amharic Machine Translation: An Experiment using Constrained Corpus,” Addis Ababa University, Addis Ababa, Ethiopia, 2013.

[10] S. Adugna, “English – Afaan Oromoo Machine Translation: An Experiment Using Statistical Approach,” A.A university school of graduate study in information science, Addis Ababa, Ethiopia, 2009.

[11] D. MULUGETA, “Geez to Amharic Automatic Machine Translation: A Statistical Approach,” Addis Abeba university, ADDIS ABABA, ETHIOPIA, 2015.

[12] solomon tefera, michael melese, marta yifru, and milion meshesha, “Parallel Corpora for bi-Directional Statistical Machine Translation for Seven Ethiopian

Language Pairs," Aug. 2018, [Online]. Available: https://www.aclweb.org/anthology/W18-3812.pdf.

[13] B. Dorr, M. Snover, and N. Madnani, "MachineTranslationEvaluation , Published 2010" .

[14] S. Harold L. and W. John Hutchins, An Introduction to Machine Translation. LONDON SAN DIEGO NEW YORK BOSTON SYDNEY TOKYO TORONTO: Harcourt Brace Jovanovich, 1992.

[15] K. E.F.K and asher R.E, concise history of the language sciences. 1995.

[16] D. Preeti and Devanand, "Machine Translation System for Hindi-Dogri Language Pair," International Conference on Machine Intelligence Research and Advancement, 2013, [Online]. Available: https://shodhganga.inflibnet.ac.in/bitstream/10603/78191/9/09_chapter%202.pdf.

[17] M. L. Forcada, "Making sense of neural machine translation," Translation Spaces 6:2 (2017) 291–309, DOI 10.1075/ts.6.2.06for, 2017, [Online]. Available: https://www.dlsi.ua.es/~mlf/docum/forcada17j2.pdf.

[18] S. Tripathi and J. Krishna, "Approaches to machine translation," May 2014, [Online]. Available: https://www.researchgate.net/publication/228574546.

[19] M. MARA, "ENGLISH-WOLAYTTA MACHINE TRANSLATION USING STATISTICAL APPROACH," ST. MARY'S UNIVERSITY SCHOOL OF GRADUATE STUDIES, ADDIS ABABA, ETHIOPIA, 2018.

[20] D. Bahdanau, K. Cho, and Y. Bengio, "NEURALMACHINETRANSLATIONBYJOINTLYLEARNING TO ALIGN AND TRANSLATE," May 2016, [Online]. Available: https://arxiv.org/pdf/1409.0473.pdf.

[21] C. Goutte, N. Cancedda, M. Dymetman, and G. Foster, Learning Machine Translation. Cambridge, Massachusetts London, England: 2009 Massachusetts Institute of Technology.

[22] "Introduction Natural Language ProceSsing Morphology Syntax Semantics Overview of Amharic Language Problems in Amharic Morphology." [Online].

Available:
https://shodhganga.inflibnet.ac.in/bitstream/10603/17292/6/06_chapter%201.pdf.

[23] “Ethiopian Treasures.”
http://www.ethiopiantreasures.co.uk/pages/language.htm. 2002.

[24] W. Olani, “Inflectional Morphology in Oromo,” [Online]. Available:
https://www.academia.edu/9644628/Inflectional_Morphology_in_Oromo. March 2017.

[25] M. BULCHA, “Onesimos Nasib’s Pioneering Contributions to Oromo Writing,” University of Uppsala, Sweden, Nordic Journal of African Studies (1): 36-59 1995.

[26] M. ALl and A. ZABORSKI, “HANDBOOK OF THE OROMO LANGUAGE,” Univerzitna kniznica V Bratislave, 1990.

[27] S. Amsalu and D. Gibbon, “Finite State Morphology of Amharic,” Fakult¨at f¨ur Linguistik und LiteraturwissenschaftUniversit¨at BielefeldUniversit¨at strasse 25D-33501, Germany, Feb. 2016, [Online]. Available:
https://www.academia.edu/18003881/Finite_State_Morphology_of_Amharic.

[28] getachew Dires, ''Language policy of Ethiopian'' , · september 2019, [online].
Available:
https://www.researchgate.net/publication/335868337_Language_Policy_of_Ethiopia

8 Appendix

Appendix A: Python Scrip for recombine the segmented words

```
import sys

reload(sys)

sys.setdefaultencoding('utf8')

#fh = open("morphed.sent.or-am.am","r")

fh = open("morphed.sent.or-am.or","r")

text = fh.read()

fh2 = open("morphed.corpus.or-am.or","w")

#fh2 = open("morphed.corpus.or-am.am","w")

for sent in text.split("@"):

    tmp = ""

    for w in sent.split("\n"):

      tmp+=w+" "

    tmp2 = tmp.decode('utf8')[:-2]+"\n"

            fh2.writelines(tmp2.lstrip(" "))
```

Appendix B: Sample of sentence

Amharic	Afaan Oromo
የጤና ችግርን ጨምሮ ከቁጥጥራችን ውጭ የሆኑ ሌሎች ተፈታታኝ ሁኔታዎች ሲያጋጥሙን ይሖዋ እንደሚያስብልን በመተማመን የሚያስጨንቀንን ነገር ሁሉ በእሱ ላይ መጣል ይኖርብናል።በተጨማሪም ከክርስቲያናዊ ስብሰባዎችና ከሌሎች መንፈሳዊ ዝግጅቶች ጥቅም ለማግኘት አቅማችን የፈቀደውን ሁሉ እናድርግ።ከእውነት ቤት የወጡ ልጆች ስላሏቸው ታማኝ ወላጆችስ ምን ማለት ይቻላል?አረጋዊው ሳሙኤል ትላልቅ የሆኑ ልጆቹ እሱ ያስተማራቸውን የጽድቅ መሥፈርቶች እንዲጠብቁ ማስገደድ አይችልም ነበር።ሳሙኤል ጉዳዩን ለይሖዋ ከመተው በቀር ምንም አማራጭ አልነበረውም።ሆኖም ንጹሕ አቋሙን ይዞ በመቀጠል የሰማይ አባቱን ማስደሰት ይችላል።በዛሬው ጊዜም እንዲህ ባለ ሁኔታ ውስጥ የሚገኙ በርካታ ክርስቲያን ወላጆች አሉ።ይሖዋ ንስሐ የሚገቡ ኃጢአተኞች ወደ እሱ የሚመለሱበትን ጊዜ በጉጉት እንደሚጠባበቅ ይተማመናሉ።በሌላ በኩል ደግሞ የእነሱ ታማኝነት ልጃቸው ወደ ይሖዋ እንዲመለስ እንደሚረዳው በመተማመን ምንጊዜም ለይሖዋ ታማኝ ለመሆን ከፍተኛ ጥረት ያደርጋሉ።ለምትቀርበው ሰው ስሜትህን አውጥተህ መናገርህ ጭንቀትህን ለማቅለል ይረዳሃል።	Yommuu dhibeen fayyaa ykn rakkinni humnaa ol ta’e kan biraan nu mudatu, akka nuuf yaadu amanuudhaan wanta yaaddoo nutti ta’u Yihowaa irratti gatuu qabna. Akkasumas, walga’iiwwan gumii fi qophiiwwan hafuuraa kan biroo irraa fayyadamuuf wanta nuuf danda’ame hunda gochuu qabna.Waa’ee warra amanamoo, ijoolleen isaanii Yihowaa tajaajiluu dhiisanii hoo maal jechuutu danda’ama? Saamu’el maanguddichi ilmaan isaa ga’eessota ta’an ulaagaalee Yihowaa warra qajeeloo inni isaan barsiiseef amanamoo ta’anii akka jiraatan isaan dirqisiisuu hin dandeenye. Dhimma kana harka Yihowaatti gatee dhiisuu qaba ture. Ta’us, Saamu’el amanamaa ta’ee jiraachuu fi Abbaa isaa isa samii Yihowaa gammachiisuu danda’eera. Yeroo ammaattis, Kiristiyaanonni ijoollee qaban hedduun haala wal fakkaatu keessatti argamu.Isaan Yihowaan cubbamoota yaada geddaratan simachuuf hawwiidhaan eeggachaa akka jiru amanannaa qabu. Hammasitti garuu, fakkeenyi isaanii ijoolleen isaanii gara karrichaatti akka deebi’an akka isaan gargaaru abdachuudhaan, Yihowaadhaaf amanamoo ta’anii jiraachuuf carraaqqii cimaa gochuu qabu. Wanta isinitti dhaga’amu nama amantanitti himuun keessan dhiphina dandamachuu akka dandeessan isin gargaaruu danda’a.

Appendix C: Sample of morphem based alignment

# Sentence pair (11) source length 17 target length 15 alignment score : 1.50234e-33

ሌሎች ም የ ራሳቸው ን ውሳኔ የ ማድረግ ነፃነት እንዳ ላቸው አምነ ን መቀበል አለብን

NULL ({ 5 }) bilisummaa ({ }) warri ({ }) kaan ({ 1 }) jireenya ({ }) isaanii ({ }) keessa ({ }) tti ({ 2 }) murtoo ({ 6 }) mataa ({ 3 4 7 }) isaanii ({ }) gochuu ({ 8 }) f ({ }) qaba ({ }) n ({ }) kabaju ({ 9 10 11 12 14 }) u ({ 13 }) qabna ({ 15 })

# Sentence pair (12) source length 2 target length 2 alignment score : 0.0626827

ለምን ?

NULL ({ }) Maaliif ({ 1 }) ? ({ 2 })

# Sentence pair (13) source length 22 target length 19 alignment score : 7.00358e-46

እያንዳንዱ ክርስቲያን የ መምረጥ ነፃነት ስለተ ሰጠው ፣ ሁለት ክርስቲያኖች ሁልጊዜ ፍጹም አንድ ዓይነት ውሳኔ ያደርጋሉ ብሎ መጠበቅ አይቻልም

NULL ({ }) hundi ({ 1 }) keenya ({ }) iyyuu ({ }) mirga ({ 3 }) filannoo ({ 4 5 6 7 16 17 18 19 }) waan ({ }) qabnu ({ }) uf ({ }) , ({ }) Kiristiyaan ({ 2 10 }) onni ({ 8 }) lama ({ 9 }) yeroo ({ }) hunda ({ 11 }) murtoo ({ 12 15 }) wal ({ }) fakkaatu ({ 13 14 }) gochuu ({ }) hi ({ }) n ({ }) danda'a ({ }) n ({ })

# Sentence pair (14) source length 14 target length 16 alignment score : 1.3993e-31

ይ ህ ደግሞ በምና ሳየው ምግባር ና በምና ቀርበው አምልኮ ላይ ም እንኳ ሊ ንጸባረቅ ይችላል

NULL ({ }) kun ({ 1 2 }) a ({ }) malaa ({ 3 4 5 6 8 9 15 16 }) fi ({ 7 }) waaqeffannaa ({ 10 13 14 }) keenyaa ({ }) wajjin ({ }) haala ({ }) wal ({ 11 }) qaba ({ }) tee ({ }) nis ({ 12 }) ni ({ }) hojjeta ({ })

# Sentence pair (15) source length 38 target length 44 alignment score : 9.09596e-120

እያንዳንዱ ክርስቲያን የራሱ ን የ ኃላፊነት ሸክም መሸከም እንዳለ በት ከተ ገነዘብ ን ሌሎች ሰዎች ያ ን ያህል ትልቅ ቦታ በማይ ሰጣቸው ጉዳዮች ም ረገድ የ መምረጥ ነፃ ነታቸውን ተጠቅመ ው የ ራሳቸው ን ውሳኔ የ ማድረግ መብት እንዳ ላቸው አምነ ን እንቀበላ ለን

```
NULL ({ 4 13 17 34 42 }) Kiristiyaan ({ 2 }) ni ({ }) hundi ({ 1 }) dhuunfaa ({ 5 }) tti ({
}) ba'aa ({ 6 7 }) ofii ({ 3 33 }) isaa ({ }) baachuu ({ 8 }) akka ({ }) qabu ({ 9 10 31 44
}) yoo ({ }) hubanne ({ 11 12 }) , ({ }) yeroo ({ }) wantoota ({ }) baay'ee ({ })
barbaachisa ({ }) a ({ 32 }) hi ({ }) n ({ }) taane ({ }) irratti ({ }) murtoo ({ 35 })
dhuunfaa ({ 43 }) goonu ({ }) tti ({ }) illee ({ 24 }) mirga ({ 36 38 41 }) warri ({ 15 })
kaan ({ 14 }) filannoo ({ 27 28 }) gochuu ({ c 37 }) f ({ }) qaba ({ 16 39 40 }) n ({ 26 })
ni ({ }) kabajna ({ 18 19 20 21 22 23 25 29 30 })
```

Appendix D: Sample of word based alignment

```
# Sentence pair (1) source length 9 target length 5 alignment score : 8.53012e-07
አንዳንዶቹ የየዋሆችን ልብ እያታለሉ ነበር
NULL ({ }) warri ({ }) tokko ({ }) tokko ({ }) namoota ({ }) hin ({ }) shakkine ({ 1 2 3 4
}) sossobanii ({ }) gowwoomsaa ({ }) turan ({ 5 })
# Sentence pair (2) source length 16 target length 13 alignment score : 5.88776e-24
በቆሮንቶስ ጉባኤ ፣ ለተወሰነ ጊዜ ያህል ከአባቱ ሚስት ጋር የሚኖር ወንድም እንደነበረም
እናስታውስ
NULL ({ 3 }) gumii ({ 2 }) Qorontos ({ 1 }) keessa ({ }) yeroo ({ 5 }) muraasaaf ({ 4 })
obboleessi ({ }) haadha ({ }) manaa ({ }) abbaa ({ }) isaa ({ }) wajjin ({ 9 }) jiraatu ({
}) akka ({ }) tures ({ 6 7 8 10 11 }) haa ({ }) yaadannu ({ 12 13 })
# Sentence pair (3) source length 20 target length 15 alignment score : 1.20368e-24
እንግዲያው አምላክ ፣ ክርስቲያኖች በሥጋዊ ነገሮች ላይ ማተኮር እንደሌለባቸው በጳውሎስ በኩል
ማሳሰቢያ መስጠቱ የሚያስገርም አይደለም
NULL ({ 3 }) Kanaaf ({ 1 }) , ({ }) Waaqayyo ({ 2 }) Phaawulositti ({ }) fayyadamee ({
}) waa'ee ({ }) yaanni ({ }) ofii ({ }) foon ({ }) irratti ({ }) akka ({ }) xiyyeeffatu ({ 5 6 7
8 }) gochuu ({ }) Kiristiyaanota ({ 4 }) akeekkachiisuun ({ 9 10 11 12 13 }) isaa ({ })
kan ({ }) nama ({ }) dinqisiisu ({ 14 15 }) miti ({ })
# Sentence pair (4) source length 7 target length 5 alignment score : 5.8537e-07
ይህ ማስጠንቀቂያ ዛሬም ቢሆን ይሠራል
NULL ({ }) Akeekkachiisni ({ }) sun ({ }) , ({ }) yeroo ({ }) har'aatiifis ({ 1 2 3 4 5 }) ni
({ }) hojjeta ({ })
# Sentence pair (5) source length 30 target length 28 alignment score : 2.6367e-55
የይሖዋ ድርጅት አባል መሆን ምንኛ መታደል ነው ! እርግጥ ነው ፣ የአምላክን መሥፈርቶች
እንዲሁም እሱ የሚጠብቅብንን ነገሮች ማወቃችን ፣ ትክክል የሆነውን ነገር የማድረግና የእሱን
ሉዓላዊነት የመደገፍ ኃላፊነት ያስከትልብናል
NULL ({ 10 11 }) kutaa ({ }) jaarmiyaa ({ 2 3 }) Yihowaa ({ 1 }) ta'uun ({ 4 }) mirga ({
}) guddaa ({ }) dha ({ 7 }) ! ({ 8 }) ulaagaa ({ 9 13 }) fi ({ }) qajeelfama ({ })
Waaqayyoo ({ 12 }) beekuun ({ }) keenya ({ }) , ({ 19 }) wanta ({ 22 }) sirrii ({ 20 21
}) ta'e ({ }) raawwachuu ({ }) fi ({ 14 }) olaantummaa ({ 5 6 16 17 18 23 }) isaa ({ 15
}) deggeruuf ({ 24 25 }) itti ({ }) gaafatamummaan ({ 26 27 28 }) akka ({ }) nu ({ })
irra ({ }) jiraatu ({ }) godha ({ })
```