Addis Ababa University

College of Natural Science

Amharic-to-Tigrigna Machine Translation Using Hybrid Approach

Akubazgi Gebremariam

A Thesis Submitted to the Department of Computer Science in Partial Fulfillment for the Degree of Master of Science in Computer Science

Addis Ababa, Ethiopia

October 7, 2017

Addis Ababa University

College of Natural Sciences

Akubazgi Gebremariam

Advisor: Yaregal Assabie (PhD)

This is to certify that the thesis prepared by Akubazgi Gebremariam, titled: *Amharic-to-Tigrigna Machine Translation Using Hybrid Approach* and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name	Signature	Date

Advisor: Yaregal Assabie (PhD)_____

Examiner: (PhD) _____

Examiner: (PhD) _____

Abstract

Machine Translation is one of the applications of Natural Language Processing that studies the use of computer software to translate a natural language into another language in the form of text or speech. People use human translation and they tend to be slower as compared to machines. Sometimes it can be hard to get a precise translation that reveals what the text is about without everything being translated word-by-word. In addition, it can be more important to get the result without delay which is hard to accomplish with a human translator. It also leads to unwanted expenses like, time and cost. Thus, this research works on Amharic-to-Tigrigna machine translation system using a hybrid approach i.e. the combination of rule based and statistical approaches to solve the problems. Though, Amharic and Tigrigna are from the same family of language and uses similar sentence structure, they have also difference in constructing various types of phrases. Therefore, the study proposes syntactic reordering approach which aligns the structural arrangement order of words in the source sentence to be more similar to the target sentences. So, reordering rules are developed that fulfils for both simple and complex Amharic sentences that have difference in the structural arrangement order of words. As the researcher knowledge is concerned, there is no prior work conducted on machine translation between Amharic and Tigrigna which is in need to solve this currently. In order to achieve the objective of the study, a corpus is collected from different domain and prepared in a format suitable in the development process and classified as training set and test set. Reordering rules are applied on both the training and testing set in a pre-processing step. One language model is developed, since the system is unidirectional i.e. Amharic-to-Tigrigna. Translation model which assign a probability that a given source language sentence generates target language sentence are built and decoder which searches for the best sequence of translation probability is used. Two major experiments are conducted using two different approaches and their results are recorded. The first experiment is carried out using a statistical approach and the result obtained from the experiment has a BLEU score of 7.02%. The second experiment is carried out using hybrid approach and the result obtained has a BLEU score of 17.47% s. From the result, it can be concluded that the hybrid approach is better than the statistical approach for Amharic-to-Tigrigna machine translation system.

Keywords: Machine Translation, Statistical Machine Translation, Hybrid Machine Translation, Reordering Rule.

Acknowledgments

Above all I would like to thank the almighty God, who gave me the opportunity and strength to achieve whatever I have achieved so far. I would like to express my gratitude to all the people who supported and accompanied me during the progress of this thesis.

Next, I would like to express my deep-felt gratitude to my advisor, Dr. Yaregal Assabie, for his excellent and enduring support shaped this work considerably and made the process of doing this thesis to learn invaluable experience.

I would also like to thank the experts of the languages who were involved in checking the correctness of the parallel corpus and POS tagging work for this thesis: Mr. Semere Tesfagabr, Teklay Guesh, Berhe Gebremichael and Dr. Getahun Amare.

I would like also to thank Mr. Yrgalem G, Kidane W/mariam and Mr. Mekonnen Ybrah as the second reader of this thesis and I am gratefully anticipated to them for their very valuable comments on this thesis.

Finally, all my parents deserves special thanks in every moment of my life.

Table of Contents

List of Tables

List of Figures

List of Algorithms

Acronyms and Abbreviations

A	Amharic
BLEU	Bilingual Evaluation Understudy
DOP	Data-oriented parsing-based
EBMT	Example Based Machine Translation
FDRE	Federal Democratic Republic of Ethiopia
HMT	Hybrid Machine Translation
IBM	International Business Machines
LM	Language Model
MT	Machine Translation
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
POS	Part of Speech
rA	Locally reordered Amharic
RBMT	Rule Based Machine Translation
SMT	Statistical Machine Translation
SOV	Subject Object Verb
T	Tigrigna

Chapter 1: Introduction

1.1 Background

Natural languages are languages spoken by people [1]. Natural Language Processing (NLP) girdles everything a computer needs to understand natural language and also generates natural language [2]. Natural Language Processing is a subfield of Artificial Intelligence and linguistic, devoted to make computers understand human language [2]. A Natural language also known as ordinary language that is spoken or written by people (human) for general purpose communication. Natural language came into existence because when user wishes to communicate with the computer you cannot force users to learn machine specific language so this basically provides to managers or children's who do not have enough time to learn new specific languages or get skilled in them [1], like Amharic, Tigrigna, English, and Chinese etc. A language is a system, a set of rules or set of symbols [1].

In the past decade, successful natural language processing applications have become part of our everyday experience, spelling and grammar correction in word processors, machine translation, information retrieval, question answering, information retrieval, text segmentation, and dialogue system [3]. At present, computers do not understand ordinary language or converse naturally with human about arbitrary topics and the goal of research in NLP and in natural language understanding, and in dialogue system in particular is to break down the barrier [4]. Many researches been conducted on topics related to the application of natural language processing such as, Machine Translation (MT), text summarization, knowledge extraction, tutoring system, question–answering, catalogue shopping, reservation; help desk and dialogue with animated characters in computer games [3]. Machine translation is the process by which computer software is used to translate a text from one natural language (such as English) to another (such as Spanish) [5, 6]. There are three major machine translation approaches; Rule Based Machine Translation (RBMT), Statistical Machine Translation (SMT) and Hybrid Machine Translation (HMT).

The Ethiopic writing system is used to represent different Semitic languages including Amharic and Tigrigna [7]. These languages are mainly spoken in Ethiopia and Eritrea. Amharic and Tigrigna are closely related languages to each other.

The writing system of both languages are from left to right and have similar alphabets i.e. Fidel. Amharic and Tigrigna languages have 33 and 35 base symbols (consonants) with seven orders

which represent seven vowels for each base symbol respectively [7, 8]. The vowels of the alphabet are not encoded explicitly but appear as modifiers of the base characters. It does not make any distinction between upper and lower case letters. There are no systematic variations in the form of the symbol according to its position in the word. These two languages also have similar numbering system and punctuation marks [7].

1.2 Motivation

There are over 80 languages spoken in Ethiopia which has a population of over 100 million at present [9, 10]. Among these languages, Amharic is the working language of the federal government and official language of four regions: Amhara, Benishangul-Gumuz, Gambella, Southern Nations, Nationalities, and People's and two chartered city administration, Addis Ababa and Dire Dawa [11, 12]. Amharic is spoken as a mother tongue by a large segment of the population and it is the most commonly learned second language throughout the country and as a result, Amharic is the lingua franca of the country in the modern era [13]. Tigrigna is the official and working language of Tigray region [11, 12]. It is also one of the languages in Eritrea [4]. The language is spoken approximately by 6 million people in the northern part of Ethiopia (Tigray) and central part of Eritrea [14]. According to [15], the number of Amharic first mother tongue speakers are 29.3%, Tigrigna first mother tongue are 5.9% and the federal government permanent employees of both ethnic groups are 50.29% and 7.79 % Amhara and Tigray respectively. In general, the fact that motivated us to work on Amharic to Tigrigna machine translation is stated as follows.

Lack of information: As Amharic is the federal language of Ethiopia [11], official documents, news and vacancies are written in Amharic language. The constitution [11] recommends it is better if every regional official documents are translated and documented in Amharic language parallel with the local language. However, there is no any description that talk about the reverse way i.e. from Amharic to other local languages, instead every official federal documents such as proclamation and constitution are prepared in Amharic and English. Therefore, it is good if there is a mechanism to change the federal written documents and literal news into local languages which can prevent peoples from lack information and face to unwanted expenses, such as time and cost.

Communication gap: Speakers of Tigrigna who are unable to speak and understand Amharic cannot communicate and interact with Amharic speakers in easy way without finding human

translators which could be hard sometimes since Amharic has large segment of speakers [11, 15] comparing to Tigrigna speakers across the country.

Thus, it is good contribution if there is a way in which Amharic texts are translated into Tigrigna texts.

1.3 Statement of the Problem

Several studies and applications have been done for foreign languages using different methodologies and approaches. Most of the works have been done on language pair of English and other languages, such as Arabic [16] and Japanese [17], French [18], Spanish [19], etc. This is because English languages is the dominant languages spoken across the world next to Mandarin and it is the language of international communication [20]. However, only little work has been done on machine translation system among English and Ethiopian languages. Some of the studies are carried out on English-Amharic language pair [21, 6, 22] and English-Affan Oromo language pair [23, 24].

Literatures tells us Amharic and Tigrigna are closely related languages [7, 8]. The reason is, because both language have similar sentence structure (Subject-Object-Verb) and writing system, have similar phrasal categories, use similar punctuation and numbering system. However, there are several differences between the two languages. For instance, the way in constructing phrases. There are several Amharic phrases that differ the order of words in a given phrase from Tigrigna phrases. So to solve this problem, syntactical reordering rules are proposed to change the order of words in a given Amharic phrase in a sentence to have more similar structural order of words as the target language which can be cosidered as a pre-processing step to statistical approach.

People use human translation and they tend to be slower as compared to machines. Sometimes it can be hard to get a precise translation that reveals what the text is about without everything being translated word-to-word. In addition, it can be more important to get the result without delay which is hard to accomplish with a human translator. That is when machine translation comes in, that solves most of the problems caused by a human translator.

As the researcher knowledge is concerned, there is no prior study conducted on the development of Amharic-to-Tigrigna machine translation system.

1.4 Objective

General objective

The general objective of this study is to design and develop Amharic-to-Tigrigna machine translation system using hybrid approach.

Specific objective

The specific objectives are:

- Review related systems and literature
- Identify the linguistic behaviors of Amharic and Tigrigna languages
- Develop parallel bilingual corpus for Amharic and Tigrigna languages
- Identify syntactic relationships between Tigrigna and Amharic
- Design a general architecture for Amharic-to-Tigrigna machine translation
- Develop prototype
- Test and evaluate the performance of the system.

1.5 Methods

In order to achieve the objectives of this study, the following methods will be employed.

Literature review

For this study, secondary data sources, like books, articles, publications and other resources related to the topic will be reviewed. This helps to have a better understanding about the subject of the study. Studies related to this study will be compiled so as to know the pros and cons of various machine translation techniques. Machine translation system on different languages will studied with respect to the closeness and difference among the languages. The details of the approaches and algorithms followed to build the translation system will be reviewed. The linguistic behavior of Amharic and Tigrigna languages will also investigated and identified.

Data Collection

Amharic-Tigrigna corpora will be collected from relevant data sources with parallel text; namely from the Holy Bible, the Constitution of Federal Democratic Republic of Ethiopia (FDRE), and relevant websites as well as from previously used texts translated by language experts. The study uses two sampling technique, Random Sampling and Convenience Sampling**. Convenience sampling** is used in investigative research where the researcher is interested in getting an

inexpensive approximation of the truth. The samples are selected because they are convenient. This technique is used for Holy Bible sentences because all the writing system and words in the sentences are not similar with the other literal documents and news which is difficult to model rules. **Random sampling** is the purest form of probability sampling. Each member of the sentence has an equal and known chance of being selected. Since there is scarcity of well-organized parallel document, data will be collected randomly from previously revisit data and other sources manually with help of language translators.

Software Tools

Since the objective of this research is to develop Amharic-to-Tigrigna machine translation system using hybrid approach, which is a combination of rule based and statistical approach, the freely available software's such as IRSTLM [25] toolkit for the purpose of language modeling, GIZA++ [25] will be used for translation model, and Moses [25] will be used for the translation process taking language model and translation model into account for the statistical part will be used. The Python [26] programming language will be used for the rule part; moreover, Ubuntu 14 operating system suitable for Moses environment will be used.

Experiment and Evaluation

Machine translation system can be evaluated either using by human (manual) or automatic evaluation methods. Manual evaluation is time consuming and expensive to perform, so comprehensive comparisons of multiple systems are rare. BLEU score will be used to evaluate the performance of the system, which is an automatic evaluation technique.

1.6 Scope and Limitation of the Study

Scope of the study

Amharic-to-Tigrigna machine translation system is designed which enables to translate Amharic texts into Tigrigna text either at word or phrase or sentence level based on the input provided. Speech translation is not included in this study.

Limitation of the study

Absence of publically available automatic Amharic part-of-speech tagger is the main limitation of this study. Therefore, Amharic sentences are tagged manually which took much time and effort. Scarcity of bilingual parallel corpus, checking grammatical correctness of input text, unable to translate text figures and PDF file documents are other limitation in this study.

1.7 Application of Result

The results of this research work have many applications. The system can be used for translation from Amharic-to-Tigrigna texts, and the translation system can be used as a tool for the teaching learning process of the languages. This study can be used to simplify the barrier of language difficulty among the language users. It enables to access information and interaction easily and fills communication gap between peoples using the two languages; moreover, the study can be used as a component for other NLP applications such as speech translation.

1.8 Thesis Organization

This thesis is organized into six chapters. Chapter Two presents literature review which includes an overview of the languages and machine translation. Chapter Three presents different related works on machine translation. Chapter Four presents design of Amharic-to-Tigrigna machine translation using hybrid approach. The experiments and results are discussed in Chapter Five and Chapter Six presents conclusions and future works.

Chapter 2: Literature Review

2.1 Introduction

In this chapter, a brief overview of Tigrigna and Amharic, and machine translation (MT) are discussed. Additionally, the state of the art of machine translation which includes statistical machine translation (SMT), rule based machine translation (RBMT), example based machine translation (EBMT) and hybrid machine translation (HMT) are described.

2.2 Amharic Morphology

The roots of verbs and most nouns in the Semitic languages are characterized as a sequence of consonants or "radicals" (hence also the term consonantal root) [6, 8, 28] Such abstract consonantal roots are used in the formation of actual words by adding the vowels and non-root consonants (or "transfixes") which go with a particular morphological category around the root consonants, in an appropriate way, generally following specific patterns. It is a peculiarity of Semitic linguistics that a large majority of these consonantal roots are trilateral (although there are a number of quadrilaterals and in some languages, also bilateral). A trilateral root is a root containing a sequence of three consonants. As with many Semitic languages, Amharic uses triconsonantal roots in its verb morphology. The result of this is that a fluent speaker of Amharic can often decipher written text by observing the consonants, with the vowel variants being supplemental [6, 27, 28].

2.2.1 Inflection

Inflection is a morphological variation that does not change the part of speech category and general meaning, but the grammatical function [6, 27, 28]. Since Amharic language is highly inflectional, a given root of a language word can be found in different forms. The highly inflected parts of speech are discussed as follows.

Nouns: Nouns of the language are marked for gender, person and number, and results an infected word with affixes to the noun [27, 28]. Grammatically, the language specify two types of genders: Feminine and masculine. Therefore, the inflected nouns help to express possession, pluralism, tribe and gender. Table 2.1 shows how these affixes are used to inflect Amharic nouns.

Table 2.1: Amharic inflection of nouns

No.	Amharic inflected nouns			Remarks	
1.	Noun stem	Affixes	Inflected Noun	Number	Gender
2.	ላም/lam	-ኦች/^oc	ላሞች/ lamoc	Plural	Feminine
3.	ፊደል/fidel	-ኣት/^et	ፊደላት/fidelat	Plural	Masculine
4.	ቅጠል/qTel	-ኣ/ ^e	ቅጠላቅጠል/qTelaqTel	Plural	Common
5.	ቤት/bEt -	-ኦች/^oc	ቤቶች/ bEtoc	Plural	Masculine

Verb: For the purpose of investigating the inflection of the language verbs, Amharic grammar Book by Baye Ymam [27, 28] is used. Amharic verbs are found in different forms like, perfective, imperfective, gerundive, jussive and imperative by employing affixes. For a better understanding of the forms of verbs, the researchers have stated each of them one by one below:

The morphological inflection of a perfective verb is formed by adding suffixes like ኩ/ku, ክ/k, ሽ/, ኧ/, ኦች/, ኣቹ, ኡ/ that indicates person, and gender and number to the perfect verb stem (ሄድ). E.g., ሄድኩ, ሄድክ, ሄድሽ, ሄደ, ሄደች, ሄዳቹ, ሄዱ, ሄዱ (hEdku, hEdeh, hEdx, hEde, hEdec, hEdacu, hEdu, hEdu).

These are perfective verbs its stem ሄድ/hEd in Amharic which means 'to go'. In perfective verb example the verb ሄዱ/hEdu is used for both plural male and female.

The other verb type is imperfective verb and is formed by affixing morphemes like, ል-ኢ. ት-, ት-ኢ, ይ- ኢ, ን - ኢ, ት- ኡ and ይ- ኡ that indicates gender, person and number markers to the imperfective verb stem ሄዱ. E.g, the following are examples of imperfective verbs.

ልሂድ, ትሄድ, ትሄጂ, ይሂድ, ንሂድ, ትሄዱ and ይሄዱ (lhid, thid, thEji, yhid, nhid, thEdu, and yhEd)

These are imperfective verbs that illustrate how the affixes are used to inflect the imperfective verb stem. In these examples, the second person plural form of imperfective verbs are similarly represented as ትሄዱ for both masculine and feminine. Similar things are also observed for third person plural form representation of the language.

The gerundive form is inflected by adding suffixes at the end of the gerundive verb to indicate person, gender and number. The morphemes, አሁ, አህ, አሽ, አ, አች, አን, አችሁ and እተዋል can be taken as an example for the stem ሰር.

ሰራሁ, ሰራህ, ሰራሽ, ሰራ, ሰራች, ሰራን, ሰራችሁ, ሰራችሁ and ሰርተዋል (serahu, serah, serax, sera, serac, seran, serachu, serachu, and sertewal)

These shows how the gerundive Amharic stem verbs morphologically varies. The gerundive Amharic verb has similar behavior as perfect and imperfect representing for 2nd and 3rd person feminine and masculine in the plural form of the given verb.

Jussive and imperative verbs are sometimes called mood and jussive verbs are used to express a command for first and third persons whereas imperative verb is used to express second person in the singular and plural form [27, 28].

Adjectives: Adjectives of Amharic language are marked for gender and number, and results an inflected word with affixes [27, 28]. In Amharic, some of the morphemes that are used to inflect given adjectives are, -ኦ/^o,-ኢት/^it and ኦች/^oc. Table 2.2 shows examples of inflected Amharic adjectives.

Table 2.2: Amharic inflection of adjectives

No.	Adjectives	Masculine	feminine	Pluralisation
1.	ሞኝ	ሞኞ	ሞኚት	ሞኞች
2.	ብልጥ	ብልጦ	ብልጢት	ብልጦች
3.	ቂል	ቂሎ	ቅሊት	ቅሎች

2.2.2 Derivation

Derivational morphology deals with adding affixes to words to bring a change in meaning and category of the original word. Although derivational morphology is not the core issue of the research work, the highly derivational parts of speech are discussed as follows.

Nouns: unlike inflection, nouns can be derived and have new structure and meaning as well as word classification [27, 28]. The following are examples of derivational Amharic nouns:

ቤተኛ, ሰውነት, ልጅነት, ክብረት and etc are dreivated from the original nouns, ቤት, ሰው, ልጅ and ክብር by attaching the morphemes like, ኛ, ነት, and ኛት respectively. Nouns can be derived from the root and stem forms of verbs by infixing vowels between consonant and affixing morphemes respectively. It can be also derived from adjectives.

Verbs: verbal stems [28] can be derived from verbal roots by affixing the vowel -ኧ- between consonant as shown below.

E.g. ስ-ብ-ር, ስኧብኧር which gives ሰበር and similarly, it can be constructed from verbal stems by affixing the morphemes like, ተ, አስ, and አ with the stems ሰበር, ለመን and ወረድ which gives ተሰበር, አስለመን and አወረድ.

Adjectives: Amharic adjective can be derived from verbal roots by infixing vowels between consonants [27]. For example, ድ-ር-ቅ by applying the vowel, ድኧርኧቅ it gives the adjective ደረቅ. It can be also derived from nouns, stems by suffixing bound morphemes, and compound words.

2.2.3 Affixation

Affixes in Amharic language are morphemes which can be classified as prefix, suffix and infix [6, 28]. These are used during inflection and derivation of nouns, verbs and adjectives from different aspects. However, there are two types of affixes, namely the preposition and morphemes that are attached to the words. Amharic prepositions like የ, ከ, ስለ, and ለ are not morphemes but they cannot stand separately like morphemes. Whereas morphemes are minimal units of morphology that can bring meaningful change over the words attached to them. Affixation also includes about definiteness issue which is part of a sentence that plays great role to construct meaningful information in any language. However, definiteness can appear at different places and in different form for individual languages. In this research, in Amharic, most of the definiteness do not independently appear instead they are appended to the noun, verb or the adjective. Table 2.3 shows model for Amharic noun, verbs and adjectives with definiteness and prepositions.

Table 2.3: Model for Amharic nouns, verbs and adjectives with definiteness and prepositions

No.	Amharic noun and adjective with definiteness	General form : Amharic noun stem +morphemes	Stem +morphems
1.	ተማሪዎቹ/temariwocu	Noun +ኣቹ | ዉ | ዋ | ኡ | ታ | ኣቹን | ውን | ዋን | ኡን | ታን/Noun +^acu | wu | wa | ^u | tWa | ^acun | wn | wan | ^un |tWan	ተማሪው_ ኣቹ
2.	ተማሪው/temariw		ተማሪ_ ው
3.	ተማሪዋtemariwa		ተማሪ_ ዋ
4.	ልጁ/lju		ልጅ_ ኡ/ lj_ ^u
5.	ልጅቷ/ljtWa		ልጂ_ታ/lje_tWa
6.	ተማሪዎቹን/temariwocun		ተማሪች_ ኣቹን/ temaric_ ^ocun
7.	ተማሪውን/ temariwn		ተማሪ_ውን/ temari_ wn
8.	ተማሪዋን		ተማሪ_ዋን/ temari_wn
9.	ቤቱ/ bEtu		ቤት_ኡ/ bEt_^u
10.	All adjective +morphemes	Adjuctive +ኣቹ | ዉ | ዋ | ኡ | ታ | ኣቹን | ውን | ዋን | ኡን | ታን/ Adjuctive +^acu | wu | wa | ^u | tWa | ^acun | wn | wan | ^un | tWan	Adjective suffixed with morphemes
11.	Noun, verb and adjective with preposition	General form	Preposition +noun or verb or adjective
12.	የተማ, ከበላ, በቀዩ	Preposition + verb or adjective or verb	የ_ተማረ, ከ_በላ, በ_ቀዩ

2.3 Amharic Phrases

A phrase is a structure in a language that is constructed from one or more words in the language. In Amharic, phrases are categorized into five categories, namely noun phrase, verb phrase, adjectival phrase, adverbial phrase and prepositional phrase [27, 28, 29]. Each phrase type can be categorized into "simple" (where only one word class is represented) and "complex" (where more than one word classes are represented).

2.3.1 Noun Phrases

A noun phrase is a syntactic unit in which the head (H) is a noun or a pronoun. It can be simple or complex. The simplest noun phrase consists of a single noun or pronoun such as እሱ (he), እሷ (she), እነሱ (they), etc. A complex noun phrase can consists of a noun (called head) and other constituents (like complements, specifires, adverbial and adjectival modifiers) that modify the head from different aspects [8, 27, 29]. Table 2.4 shows examples of noun phrases of Amharic language.

Table 2.4: Amharic noun phrases

No.	Amharic sentence
1.	ለምለም ወደ አክሱም ሄደች/lemlem wede ^eksum hEdec
2.	ያቺ ረጅሟ ለምለም ወደ አክሱም ሄደች/yaci rejmWa lemlem wede ^eksum hEdec
3.	ነጭ እርግብ /neC ^rgb
4.	የዘንድሮ ተማሪዎች/yezendro temariwoc
5.	ህዝብ የሚታከምበት ሆስፒታል/hzb yemitakembet hospital
6.	ካሳ የሳር ቤት ነበረው/ kasa yesar bEt neberew
7.	ዳግማዊ ሚኒሊክ ትምህርት ቤት/dagmawi minilik tmhrt bEt

Generally, the structural rule for noun phrase of both languages can be written as:

NP => Spec AdvP Adj NP |NP|Adj NP|NP N

Note that, noun phrase can be also constructed from another different possible combination.

2.3.2 Verb Phrases

A verb phrase is composed of a verb as a head and other constituents such as complements, modifiers and specifiers [27, 28, 29]. For example in the verb phrase, ወደ አክሱም ሄደዋል/ wede ^eksum hEdewal 'he went to Axum', ወደ አክሱም 'to Axum' is prepositional phrase modifying the verb ሄደዋል 'went' from place point of view. In general the structural rule of verb phrases can be formulated as: VP => PP V|V|AdjP V|NP VP|NP PPVP|AdvP PP VP

Similarly as noun phrase, verb phrases can also be constructed from another forms of combination.

2.3.3 Adjectival Phrases

The construction of Amharic adjectival phrase is similar to that of a noun phrase and verb phrase. It can be composed of an adjective (head), and other constituents such as complements, modifiers and specifires [27, 28]. For example, ያ በጣም ጎበዝ/ya beTam gobez,'That very clever', ያ 'that' is a specifires, በጣም 'very' is a modifier modifying the head of the adjective, ጎበዝ 'clever'. Generally, the structural rule for Amharic languages can be formulated as follows:

AdjP=>Adj|Spec Adv Adj|PP Adj|NP Adj.

Adjectival phrase can be derived from nouns and verbs [27, 28]. Adjectival nouns phrase are phrases that are constructed from nouns and called as adjectival phrase. The underlined adjectival phrases given below are examples of Amharic language:

አስቴር እንደ አባትዋ እልኸኛ ነች/^estEr ^nde ^ebatwa ^lKeNa nec

አስቴር እንደ አባትዋ አመለኛ ነች/^estEr ^nde ^ebatwa ^emeleNa nec. In this examples, the head words are እልኸኛ/^lKeNa and አመለኛ/^emeleNa derived from the nouns እልህ/^lh and አመል/^emel. Adjectival phrase can also be derived from verbs called as adjectival verb phrase. አስቴር እንደ ካሳ ሸማግሌ አታላይ ነች/^estEr ^nde kasa xmaglE ^etalay nec. In this examples, እንደ ካሳ ሸማግሌ አታላይ/^nde kasa xmaglE ^etalay is an adjectival phrase, its head word is the adjective አታላይ/^etalay derived from the verb አታለለ/^etalele.

2.3.4 Prepositional Phrase

Prepositional phrase is constructed from a preposition (P) head and other constituents such as nouns, noun phrases, verbs, verb phrases, etc [6, 27, 28]. In the prepositional phrase, እንደ እንሰሳ በዱር "like an animal on the forest", for instance, እንደ "like" and በ "on" are prepositions which are combined with the nouns እንሰሳ, "an animal" and ዱር "the forest", respectively to form their prepositional phrase. The two prepositional phrase, in turn, combine to result in the bigger preposition that is provided in the example. In general, the structural rule for Amharic can be written as: PP => PP PP |PP NP|PP NN |PPV|N PP

2.3.5 Adverbial Phrases

An Adverbial phrase is a phrase in which its head word is adverb [6, 28]. It can be constructed from one or more adverbs in the languages. In the example, ክፉኛ ታማለች/kfnya tamalec, 'she is severely ill', ክፉኛ 'severely' is the only adverb which describes the degree of the verb. The general structural rule for adverbial phrase is:

AdvP => Adv|Adv Adv

2.4 Amharic Sentences

The sentence structure for Amharic language is a Subject-object-verb (SOV) structure unlike English with a subject-verb-object combination [6, 27, 28]. We can take the following as an Example, አበበ እንጀራ በላ/Abebe ate enjera, in this example the sentences is composed of አበበ: subject, እንጀራ: Object and በላ: Verb which is different from English.

Like English, Amharic nouns are words used to name or identify any class of things, people, places or ideas or a particular one of these. An important property of the Amharic is that any word that comes at the end of a complete grammatical Amharic sentence is a verb. As an outcome of this property, a word at the end of such a sentence is expected to be tagged as a verb by an Amharic tagger. Amharic verbs are also known for taking such subject markers as "ሁ, ህ, ክ, ሽ…" and so on. Adjectives in Amharic usually precede the nouns that they modify or describe [27]. The following example shows a sentences with different part word classes and their role.

እሱ በጣም ጎበዝ ተማሪ ነበር/ he was very clever student

As it can be understand from the above Amharic and its translation in English, the adjective ጎበዝ "clever" precedes the noun ተማሪ "student" in which it modifies the student characteristics. But it does not mean that a word is an adjective just because it precedes a noun. For instance, in ያ ልጅ "that boy", the word ያ "That" precedes the noun ልጅ "boy". Based on the number of verbs or tasks involved in the sentence, a given sentence can be classified as simple and complex.

2.4.1 Simple Sentences

A simple Amharic sentence consists of a noun phrase, which is the subject, followed by a verb phrase that comprises the predicate [6, 27, 28].

በሩ ክፍት ነው /the door is open

The morphemes like /-ኡ/ that is attached to a verb (e.g. በር-ኡ) refers to definiteness and number of the subject, and objects of the sentences. The morpheme that indicates the subject always precedes the morpheme that indicates the object. A given simple sentence may describe the state of being of the subject or an action that takes place in the sentence.

Example: አበበ ዶክተር ነው/ Abebe is a doctor; this sentence describes the present state of being of Abebe.

Researchers classify simple sentences into four [6, 27] namely: declarative sentences, interrogative sentences, negative sentences and imperative sentences. Declarative sentences are used to convey ideas and feelings that the speaker has about things, happenings, feelings, etc., that could be physical, mental, real or imaginary.

Example: ራሄል ዶክተር ሆነች/Rahel became a doctor

A sentence that questions about the subject, the complement, or the action the verb specifies, is called an interrogative sentence.

Example: ራሄል መቼ ዶክተር ሆነች? /When did Rahel become a doctor?

In order to construct interrogative sentences, Amharic sentences usually involve such interrogative pronouns as ማን "who", ምን "what", የት "where", ስንት "how many", and መቼ "when". These interrogatives can then be combined with prepositions to produce some more interrogative prepositional phrases like ከማን "from whom", ለምን "why", etc. [6]

Negative sentences simply negate a declarative statement made about something.

Example: ራሄል ዶክተር አይደለችም/Rahel is not a doctor

Simple imperative sentences convey instructions and mostly their subject is a second person pronoun that is usually but implied by the suffix on the verb.

Example: ዝም በል/Shut up!

For the purpose of this study, both types of sentences are used.

2.4.2 Complex Sentences

Complex sentences in Amharic are those sentences that are composed of complex phrases such as noun phrase (NP), verb phrase (VP), or adjectival phrase [6, 27, 28]. The pattern of combination could take the form of a complex NP and a simple VP, a simple NP and a complex VP or both complex NP and VP. A complex NP is one that contains an embedded sentence with in it. The phrase, for instance, ራሔል የገዛችው የተበላሸ እንቁላል "The rotten egg that Rahel bought" is a complex

NP whose head is እንቁላል “egg”. This head has been combined with the complement የተበላሸ “rotten” in order to produce the simple Np የተበላሸ እንቁላል “a rotten egg”. This simple Np, in turn, was combined with the dependent clause ራሔል የገዛችው “that Rahel bought” to produce the above complex NP. The presence of the የ “that” in it indicates that the clause is a subordinate clause and it cannot stand alone. Verb phrase is complex if it contains more than one verb or a sentence within it like complex noun phrase [6, 27 and 28]. That is, like a complex NP, a complex VP also contains an embedded sentence that plays the role of a compliment or a modifier. Example: ናሆም ወደ ኮሌጅ ስለ ሄደ ራሔል ተደሰተች /Rahel was happy because Nahom went to college. The dependent clause here is ናሆም ወደ ኮሌጅ ስለሄደ “because Nahom went to college” and ስለ is the word that made the clause dependent. As this clause indicates the reason for “Rahel’s happiness”, it is used as an adverbial clause of reason. Simple sentences are composed of simple noun phrase and simple verb phrases while complex sentences can consist of a complex NP and a simple VP, a simple NP and a complex VP, or a complex NP and a complex VP.

2.5 Tigrigna Morphology

The roots of verbs and most nouns in Tigrigna are characterized as a sequence of consonants or "radicals" (hence also the term consonantal root) like Amharic language. Such abstract consonantal roots are used in the formation of actual words by adding the vowels and non-root consonants which go with a particular morphological category around the root consonants, in an appropriate way [30, 31, 32].

It is a particularity of Semitic linguistics that a large majority of these consonantal roots are trilateral (although there are a number of quadrilaterals and in some languages, also bilateral). A trilateral or triconsonantal root is a root containing a sequence of three consonants [6, 31 and 32. As with many Semitic languages, it uses triconsonantal roots in its verb morphology. The result of this is that a fluent speaker of the language can often decipher written text by observing the consonants, with the vowel variants being supplemental detail

2.5.1 Inflection

Inflection is a morphological variation that does not change the part of speech category and general meaning, but the grammatical function [31, 32]. Like Amharic, Tigrigna language is highly

inflectional, a given root of a language word can be found in different forms. The highly inflected parts of speech are discussed as follows.

Nouns: similarly like Amharic, Nouns of Tigrigna language are marked for gender, person and number and results an inflected word with affixes to the noun [31, 32]. The language specify two types of genders: Feminine and masculine. Therefore, the inflected nouns help to express possession, pluralism, tribe and gender. Table 2.5 shows how these morphemes are used to inflect Tigrigna nouns.

Table 2.5: Tigrigna nouns of inflection

No.	Tigrigna inflected nouns			Remarks	
1.	Noun stem	Affixes	Inflected Noun	Number	Gender
2.	ላሕሚ/laHmi	-ኣ/^a	ኣላሕም/^alaHm	Plural	Feminine
3.	ፊደል/fidel	-ኣት /^at	ፊደላት/fidelat	Plural	Masculine
4.	ቐፅሊ/Qe^Sli	-ታት/tat	ቐፅሊታት /Qe^Slitat	Plural	Masculine
5.	ገዛ/geza	-ውቲ/ wti	ገዛውቲ/ gezawti	Plural	Masculine
6.	ኢትዮጵያ/ ^ityoPya	-ዊ/wi	ኢትዮጵያዊ/^ityoPyawi	Single	Masculine

Verbs: For the purpose of investigating the inflection of the language verbs, different books like Tigrigna grammar books by Daniel Teklu [32] and Kasa Gebrehiwot. [31], are used. Like Amharic, Tigrigna verbs are found in different forms like, perfective, imperfective, gerundive, jussive and imperative by employing morphemes. For a better understanding of the forms of verbs in both languages, the researcher's has stated each of them one by one below:

The morphological variations of a perfective verb is formed by adding suffixes that indicate person, gender and number to the perfect verb stem. For example,

ከድኩ, ከድካ, ከድኪ, ከደ, ከደት, ከድክን, ከዱ, ከዳ/kedku, kedka, kedki, kede, kedet, kedkn, kedu, keda

These are perfective verbs formed from the stem ከድ/ked which means 'to go'. As it can be observed from the above table, the suffixes attached are ኡ/u/, ን/en/, ኩ /ku/, ኪ/ki/, ካ/ka/, ና/na/, ኡ/ae/, ክን/kn/, ኣ/a/ and indicate person, gender and number of the subject and the pronoun that indicates the person.

The other verb type is imperfective verb and is formed by affixing gender, person and number markers to the imperfective verb stem. For example,

እኽድ, ትኽድ, ትኽዲ, ይኽድ,ንኽድ, ትኽዱ, ትኽዳ, ይኽዱ and ይኽዳ /^Ked, tKed, tKedi, yKed,nKed, tKedu, tKeda, yKedu and yKeda

These are imperfective verbs that illustrate how the affixes are used to inflect the imperfective verb stem. As it can be understand from the above example, the morphemes attached ት(t), ይ(y), እ(i), and ን(n) are prefixes and ኢ(e), ኡ(u), and ኣ(a) are suffixes. It is also possible to find the negative form of the above mentioned root by using ኣይ/ay/, ኣይት/ayt/, ኣይን/ayn, as prefixes and ን/n/, ኣን/an/, ኡን/un/ as suffixes.

The gerundive form is inflected by adding suffixes at the end of the gerundive verb to indicate person, gender and number. E.g.,

ሰሪሐ,, ሰሪሕኻ, ሰሪሕኺ, ሰሪሑ, ሰሪሓ, ሰሪሕና, ሰሪሕኹም, ሰሪሕኽን, ሰሪሖም , ሰሪሐን/ seriHe, seriHKa, seriHKi, seriHu, seriHa, seriHna, seriHKum, seriHKn, seriHom , seriHen

These shows how the gerundive Tigrigna verbs ሰሪሕ/seriH morphologically varies. As it can be understand from the above example, the affixes like ኣ/a, ኻ/ Ka, ኪ/ki, ኡ/u, ና/na, ኩም/kum, ኩም/ken…, etc are attached to inflect the stem.

Jussive and imperative verbs are sometimes called mood and jussive verbs are used to express a command for first and third persons whereas imperative verb is used to express second person in the singular and plural form [30, 31, 32].

Adjectives: Adjectives of Tigrigna language are marked for gender, person and number, and results an inflected word with affixes to the adjective. It can be inflected by adding a ት/t/, ቲ/ti/, ኣዊት /^awit /, ኣዊ/^awi/, ኣት/^at/, ኣዊያን/^awiyan/, etc morphemes [31, 32].

Table 2.6 shows an example of inflected form of Tigrigna adjectives.

Table 2.6: Tigrigna inflection of adjective

No.	Masculine	Feminine	Plural form
1.	ቀይሕ	ቀያሕ	ቀያሕቲ
2.	ቀታሊ	ቀታሊት	ቀተልቲ
3.	ኢትዮጵያዊ	ኢትዮጵያዊት	ኢትዮጵያውያን

As it is described the inflectional property of the language categories in the above, it can be understood that the three part of speech; noun, verb and adjective are highly inflectional to person, gender and number. This tells us Tigrigna language is morphological reach and complex like Amharic language.

2.5.2 Derivation

Derivational morphology deals with adding affixes to words to bring a change in meaning and category of the original word [31, 32]. Tigrigna derivational morphology have similar property as Amharic language and the highly derivational parts of speech are discussed as follows.

Nouns: in Tigrigna, the derivational property is similar with Amharic. It can be constructed by compounding two nouns [31, 32]. For example, combining ቤት and ብልዒ gives new name ቤት-ብልዒ which means restaurant. Additionally it can be constructed new nouns from nouns and adjectives by affixing morphemes like -ነት and -ኛ.

E.g. ሰብ +-ነት, this gives ሰብነት and ከንቱ +-ነት gives ከንቱነት, these are examples of noun and adjective combined with morphemes which gives different nouns respectively. Nouns can be also derivated from root and stem verbs by infixing vowels and adding morphemes respectively. For example, ቅ-ር-ፅ (root) by adding vowels ኣ, ቅአርአፅ it gives ቀረፅ.

Verbs: unlike nouns and adjectives, new verbs are constructed from verbal roots and stems [32]. E.g. ቅ-ብ-ር, ቅኽብኻር which gives ቀበር and similarly, it can be constructed from verbal stems by affixing the morphemes like, ተ and ኣ with the stems ሰረቅ and በደል which gives ተሰረቅ and አበደል.

Adjectives: adjectives are similar with nouns, it can be created from nouns, verbs and adjectives itself [31, 32]. Nouns combined with morphemes can give derivational adjectives. By adding the morphemes like ኣዊ, ዊ, ኣም and ታይ, to the nouns ፍትሒ, ኣፍሪካ, ጠንቂ, ማእኸል, and ሽረ. We can have the following adjective, ፍትሓዊ, ኣፍሪካዊ, ጠንቃም, ማእኸላይ and ሽረታይ. Similarly it can be constructed from the root verb. For example, for the root verb ሕስር, by infixing the vowel -ኡ- we can have the adjuctive called ሕሱር.

2.5.3 Affixation

Like Amharic, Affixes in Tigrigna language are morphemes which can be classified as prefix, suffix and infix [32]. These are used during inflection and derivation of nouns, verbs and adjectives from different aspects. However, there are two types of affixes, namely the preposition and morphemes that are attached to the words. Tigrigna prepositions like ን, ስለ and ብ are not

morphemes but they cannot stand separately like morphemes. Whereas morphemes are minimal units of morphology that can bring meaningful change over the words attached to them. Definiteness are also included in affixation that plays great role to make meaningful information in any language. However, definiteness can appear at different places and in different form for individual languages. In Tigrigna, definiteness are appeared independent of the noun, pronoun or the adjective. Table 2.7 shows model for Tigrigna noun, verbs and adjectives with definiteness and prepositions.

Table 2.7: Model for Tigrigna nouns, verbs and adjectives with definiteness and prepositions

No.	Definiteness +noun	General form :morphemes+ Tigrigna	Tigrigna noun and adjective with definiteness
1.	እቶም_ተማሃሮ	እቶም | እቲ | እታ | ነቶም | ነቲ | ነታ + “ “ + Noun	እቶም ተማሃሮ
2.	እቲ_ተማሃራይ		እቲ ተማሃራይ
			እታ ተማሃሪት
3.	እታ_ተማሃሪት		
4.	እቲ _ቆልዓ/^ti qol`a		እቲ ቆልዓ/^ti qol`a
5.	እታ_ቆልዓ/^ta qol`a		እታ ቆልዓ/^ta qol`a
6.	ነቶም_ተማሃሮ/netom temaharo		ነቶም ተማሃሮ/netom temaharo
7.	ነቲ _ ተማሃራይ/ temaharay neti		ነቲ ተማሃራይ/ temaharay neti
8.	ነታ_ተማሃሪት/ neta temaharit		ነታ ተማሃሪት/ neta temaharit
			እቲ ገዛ/ ^ti geza
9.	እቲ_ገዛ/ ^ti geza		
10.	Definiteness + Adjective		Definiteness + Adjective
11.	Noun, verb and adjective with preposition	General form	Preposition +noun or verb or adjective
12.	ንኣበበ, ብመኪና, ስለዝኾነ	Preposition+ Tigrigna words	General form: preposition_ Tigrigna words

2.6 Tigrigna Phrases

A phrase is a structure in a language that is constructed from one or more words. Like Amharic, Tigrigna phrase are constructed in similar way and are categorized into five, namely noun phrase, verb phrase, adjectival phrase, adverbial phrase and prepositional phrase [30, 31, 32]. Each phrase type can be categorized into simple and complex.

2.6.1 Noun Phrase

A noun phrase is a syntactic unit in which the head (H) word is a noun or a pronoun [31, 32]. It can be simple or complex. Noun phrase can be constructed from noun (called head) and other constituents like complements, specifires, adverbial and adjectival modifiers that modify the head from different aspects. Table 2.8, shows examples of noun phrases of Tigrigna language.

Table 2.8: Tigrigna noun phrase

No.	Tigrigna sentence
1.	ለምለም ናብ ኣክሱም ከይዳ/lemlem nab ^aksum keyda
2.	እታ ነዋሕ ለምለም ናብ ኣክሱም ከይዳ/^ta newaH lemlem nab ^akesum keyda
3.	ፃዕዳ ርግቢት/^Sa`da rgbit
4.	ናይ ለብዘበን ተማሃሮ/nay lebzeben temaharo
5.	ህዝቢ ዝሕከመሉ ሆስፒታል/ hzbi zHkemelu hospital
6.	ካሳ ናይ ሳዕሪ ገዛ ነይርዎ/kasa nay sa`ri geza neyrwo
7.	ቤት ትምህርቲ ዳግማዊ ሚሊክ/bEt tmhrti dagmawi milik

Generally, the structural rule for noun phrase of Tigrigna language can be written as:

NP => Spec AdvP Adj NP |NP|Adj NP|NP N

Note that, noun phrase can be also constructed from another different possible combination.

2.6.2 Verb Phrases

A verb phrase is composed of a verb as a head and other constituents such as complements, modifiers and specifires [31, 32]. For example in the verb phrase, ናብ ኣክሱም ከይዱ/ nab ^aksum keydu ‘he went to Axum’, ናብ ኣክሱም ‘to Axum’ is prepositional phrase modifying the verb ከይዱ‘went’ from place point of view. In general the structural rule of verb phrases can be formulated as follows.

VP => PP V|V|AdjP V|NP VP|NP PPVP|AdvP PP VP

Similarly as noun phrase, verb phrases can also be constructed from another forms of combination.

2.6.3 Adjectival Phrase

The adjectival phrase can be composed of an adjective (head), and other constituents such as complements, modifiers and specifies [31, 32], like noun and verb phrases. Its main objective is to describe nouns but also it can describe verbs. For example, እቲ ብጣዕሚ ንፉዕ/eti beTami nfuE, 'That very clever', እቲ 'that' is a specifier, ብጣዕሚ 'very' is a modifier modifying the head of the adjective, ንፉዕ 'clever'. Generally, the structural rule for Tigrigna language can be formulated as follows:

AdjP=>Adj|Spec Adv Adj|PP Adj|NP Adj.

Similarly like Amharic, Tigrigna adjectival phrases can be derived from noun and verb [32]. Accordingly, adjectives derived from noun are called adjectival noun phrase and adjectives derived from verbs are called adjectival verb phrase. The following underlined phrases are examples of adjectival noun phrase.

አስቴር ከም ኣቡኣ ህልኽኛ እያ (^esetEr kem ^abu^a hlKeNa ^ya)

አስቴር ከም ኣቡኣ ኣመለኛ እያ (^esetEr kem ^abu^a ^ameleNa ^ya)

As it is shown in the above, the head words are ህልኽኛ/ hlKeNa ^ya and ኣመለኛ/^ameleNa derived from ህልኽ/hlK and ኣመል/ ^amel Tigrigna nouns. Adjectival phrase can also be derived from verbs called as adjectival verb phrase. The following underlined phrase are examples adjectival phrase.

አስቴር ከም ካሳ ሸማግለ መታለሊት እያ (^astEr kem kasa xmagle metaleli ^ya)

In this example, the head word is መታለሊት derived from ኣታሊላ.

2.6.4 Prepositional Phrases

Prepositional phrase is constructed from a preposition as a head and other constituents such as nouns, noun phrases, verbs, verb phrases, adjectives, etc. [31, 32]. Unlike English, the head word of the phrase cannot stand alone, at least there must be complements with it. Prepositions in Tigrigna language appear all time in front of the nouns and other words. For instance the following underlined phrases are prepositional phrase. እሱ ብመኪና ካብ ኣዲስ ኣበባ which means "he came from Addis Ababa on bus", in this example there are two preposition, ብ/on and ካብ/from which are combined with መኪና/bus and ኣዲስ ኣበባ/Addis Ababa respectively to construct their prepositional phrase. The two prepositional phrase, in turn, combine to result in the bigger preposition that is provided in the example. In general, the structural rule can be written as follows:

PP => PP PP |PP NP|PP NN |PPV,

The structural rule is similar with Amharic except ‘N PP’ combination.

2.6.5 Adverbial Phrases

An Adverbial phrase, is a phrase in which its head word is adverb [31, 32]. It can be constructed from one or more adverbs in the languages. Some examples are like, ብጣዕሚ ሓሚማ ኣላ/she severely ill, the word ብጣዕሚ/severely, is an adverb which describes the verb. Similarly like Amharic, the number of adverbs are few and their task is done by other complements and constitutes. The general structural rule for adverbial phrase is:

AdvP => Adv|Adv Adv

2.7 Tigrigna Sentences

So far, the types of phrases and morphological variation of the language have been discussed. In this section the construction of Tigrigna sentence and types of sentences are described. Like Amharic, the sentence structure for Tigrigna language is a Subject-object-verb (SOV) [31, 32] unlike English that is subject-verb-object combination. We can take the following as an Example, አበበ እንጀራ በሊዑ/Abebe ate enjera, in this example the sentences is composed of አበበ: subject, እንጀራ: Object and በሊዑ: Verb which is similar with Amharic. So mostly, in the sentences structure the words come at the end of the sentence is verb, at the begging is noun and in the middle is object may be classified as noun or other. Adjectives in Tigrigna usually precede the nouns that they modify or describe. The following example shows a sentences with different part word classes and their role.

ንሱ ብጣዕሚ ንፉዕ ተምሃራይ ነይሩ/ he was very clever student

As it can be understand from the above example, the adjective ንፉዕ “clever” precedes the noun ተምሃራይ “student” in which it modifies the student characteristics. But it does not mean that a word is an adjective just because it precedes a noun. For instance, እዚ ቆልዓ “this boy”, the word እዚ “This” precedes the noun ልጅ “ቆልዓ”. Based on the number of verbs or tasks involved in the sentence, a given sentence can be classified as simple and complex.

2.7.1 Simple Sentences

A sentences composed of simple (single) noun phrase and verb phase is simple sentences [31, 32]. However, complements and modifiers can be included in the simple phrases. እቲ ተምሃራይ ናብ ቤት

ትምህርቲ ኬይዱ/the student went to school, this is simple sentences because it composed of single subject that is ተምሃራይ and ኬይዱ that is a verb which conveys single information. Whereas, ናብ ቤት ትምህርቲ is prepositional phrase which modifies the verb from past action.

Like Amharic, Tigrigna simple sentences are classified into four [31, 32], namely: declarative sentences, interrogative sentences, negative sentences and imperative sentences. Declarative sentences are used to convey ideas and feelings that the speaker has about things, happenings, feelings, etc, that could be physical, mental, real or imaginary.

Example: ራሄል ዶክተር ኮይና/Rahel became a doctor

A sentence that questions about the subject, the complement, or the action the verb specifies, is said to be an interrogative sentence.

Example: ራሄል መዓዝ ዶክተር ኮይና?/When did Rahel become a doctor?

In order to construct interrogative sentences, Tigrigna sentences usually involve such interrogative pronouns as መን "who", እንታይ "what", አበይ "where", ክንደይ "how many", and መዓዝ "when". These interrogatives can then be combined with prepositions to produce some more interrogative prepositional phrases like ካብ መን "from whom", ስለምንታይ "why", etc.

Negative sentences simply negate a declarative statement made about something.

Example: ራሄል ዶክተር ኣይኮነትን/Rahel is not a doctor

Simple imperative sentences convey instructions and mostly their subject is a second person pronoun that is usually but implied by the suffix on the verb.

Example: ቀስ በል/slow down!

2.7.2 Complex Sentences

Basically complex sentences in Tigrigna are those sentences that are composed of complex phrases such as noun phrase (NP), verb phrase (VP), or adjectival phrase that contains two subjects and verbs [31, 32]. The pattern of combination could take the form of a complex NP and a simple VP, a simple Np and a complex VP or both complex NP and VP. For example the following sentences are complex sentences. ኣነ ናብ ቤት ትምህርቲ እናኸድኩ ሃርያ ማይ ትቐድሕ ነይራ/ when I going to school harya was fetching water. From this example we can have two subject's ኣነ "I" and ሃርያ "Harya", and two verbs እናኸድኩ "going" and ነይራ "was" as well as two different information. These are the movement to school and fetching water. Generally, anybody who have knowledge of Amharic language can understand the behavior of Tigrigna sentences. Tigrigna sentences either simple or

complex are constructed like Amharic sentences which can have either similar or different structural arrangement of order of words in the given phrase.

2.8 Comparison of Amharic and Tigrigna Languages

The study tried to describe the nature and properties of Tigrigna and Amharic languages with regard to their writing system, sentence structure, punctuation and numbering system and morphology in general. As it is discussed before, the two languages has similar sentence structure (SOV) and morphological behavior. The nouns, verbs and adjectives of both languages are highly inflectional and derivational due to they are marked to person, gender and number. This makes the languages morphological complex and rich. The phrase categories and formation of the phrase in the languages are also similar. However, the local order of words in a given phrases in both languages have differences. The following are some examples that shows differences in local word order in a given phrase or sentences between the languages.

A: ትምህርት ቤቱ/CN አልታደሰም/V
T: እቲ ቤት ትምህርቲ/CN ኣይተሓደሰን/

A: የሰሜን ጎንደር/N2 ዞን/N ዋና/ADJ አስተዳደር/N እንደ/PRP ገለፁት/V
T: ዋና/ADJ ኣማሓዳሪ/N ዞባ/N ሰሜን ጎንደር/N2 ከም/PRP ዝገለጽዎ/V

A: እሱ/PRON ቤት/N ውስጥ/IN ነበር/V
T: ንሱ/PRON ኣብ ውሽጢ/IN ገዛ/N ነይሩ/V

Therefore, from the above examples it can be understood that, the two languages have difference in local order of words in a given phrase and sentence. Generally, we can learn knowing the general literature of the language and especially the structural arrangement order of words in different phrases helps the study to identify and develop proper syntactic local reordering rules for Amharic texts that converts the structure of a given Amharic text in to the form of the target language text structure.

2.9 Machine Translation

Machine translation (MT) is the process by which computer software is used to translate a text from one natural language (such as English) to another (such as Spanish) [33]. To process any translation, human or automated, the meaning of a text in the original (source) language must be fully restored in the target language [6, 24]. While on the surface this seems straightforward, it is far more complex [33]. Translation is not a mere word-to-word substitution. A translator must

interpret and analyze all elements of a text and know how each word may influence another and this requires extensive expertise in grammar, syntax (sentence structure/word order), semantics, etc., in the source and target languages, as well as familiarity with each local region in which syntax and semantic means of sentence structure and meanings respectively [6].

Machine translation systems can be bilingual systems or multilingual systems depending on the number of languages involved in the translation process [34, 35]. Bilingual systems are designed specifically for two languages (single pair of languages) and multilingual systems are designed for more than two languages [24]. The translation can be unidirectional or bidirectional [34]. In case of unidirectional, the system translates from the source language into the target language only in one direction [34]. Bidirectional systems work in both directions in a way that one language can act as source language or a target language [24]. Bilingual systems can be unidirectional or they can be bidirectional, but multilingual systems are usually designed to be bidirectional. Machine translation has its own advantage in allowing communication between users who speak different languages which advances globalization of the information highway [36]. Since most translation systems work in an online environment, a fast and speedy communication can be made between people who speak different languages, leave in different environments and locations which will contribute to the growth of information technology [24]. A number of methods in machine translation are being used by different researchers, and the basic approaches and methodologies according to [5] are: rule based machine translation (RBMT), statistical machine translation (SMT), example based machine translation (EBMT) and hybrid machine translation (HMT). These approaches and methodologies are going to be describing in the section bellow.

2.9.1 Rule-Based Machine Translation Approach

Rule-Based Machine Translation (RBMT) [5], also known as Knowledge-Based machine Translation and Classical Approach of MT. It is a general term that denotes machine translation systems based on linguistic information about source and target languages, basically retrieved from (bilingual) dictionaries and grammars covering the main semantic, morphological, and syntactic regularities of each language respectively [34].

Having input sentences (in some source language), RBMT system generates them to output sentences (in some target language) on the basis of morphological, syntactic, and semantic analysis of both the source and the target languages involved in a concrete translation task [5].

RBMT contains several approaches within it; these are direct approach, transfer based approach and Interlingua approach. Though they are all belongs to RBMT, they differ in the depth of analysis of the source language and the extent to which they attempt to reach a language-independent representation of meaning or intent between the source and target languages. Figure 2.1 shows Vauqouis triangle of RBMT approach.

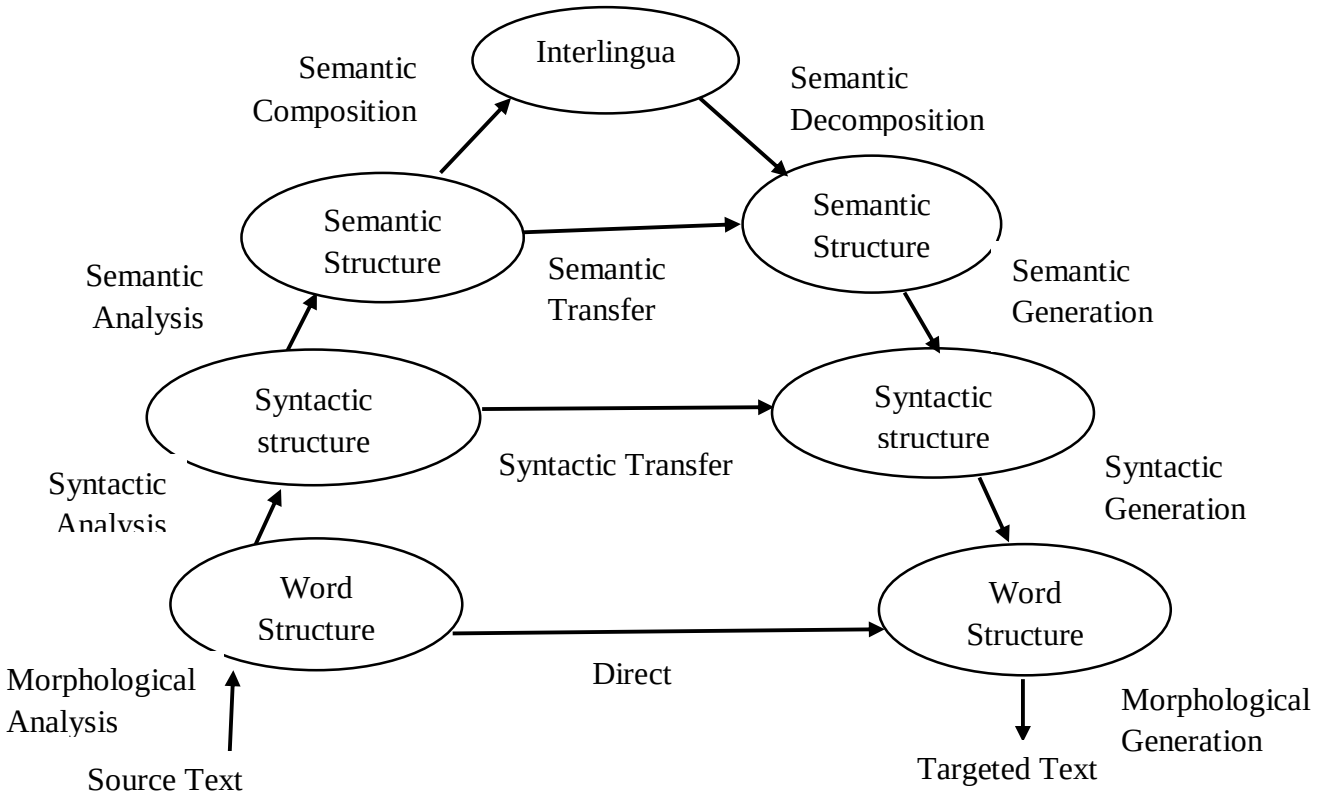

Figure 2.1: The Vauqouis triangle for Machine translation

Figure 2.1, the Vauqouis triangle [37] illustrates levels of analysis starting with the shallowest level at the bottom, direct transfer is made at the word level. Moving upward through syntactic and semantic transfer approaches, the translation occurs on representations of the source sentence structure and meaning respectively. Finally, at the Interlingua level, the notion of transfer is replaced with a single underlying representation the Interlingua that represents both the source and target texts simultaneously. Moving up the triangle reduces the amount of work required to traverse the gap between languages, at the cost of increasing the required amount of analysis (to convert the source input into a suitable pre-transfer representation) and synthesis (to convert the post-transfer representation into the final target surface form). For example, at the base of the triangle, languages can differ significantly in word order, requiring many permutations to achieve a good

translation. However, a syntactic dependency structure expressing the source text may be converted more easily into a dependency structure for the target equivalent because the grammatical relations (subject, object, and modifier) may be shared despite word order differences. Going further, a semantic representation (Interlingua) for the source language may totally abstract away from the syntax of the language, so that it can be used as the basis for the target language sentence without change.

Direct machine translation approach (DMT): Starting with the shallowest level at the bottom of the pyramid is the Direct Machine Translation Approach. According to the [5, 24] DMT approach is the oldest and less popular approach. Direct translation is made at the word level and MT systems that use this approach are capable of translating a language, called source language (SL) directly to another language, called target language (TL). Words of the SL are translated without passing through an additional/intermediary representation. The analysis of SL texts is oriented to only one TL. Direct translation systems are basically bilingual and unidirectional. Direct translation approach needs only a little syntactic and semantic analysis. SL analysis is oriented specifically to the production of representations appropriate for one particular TL. DMT is a word-by-word translation approach with some simple grammatical adjustments. Figure 2.2 shows the architecture of direct approach.

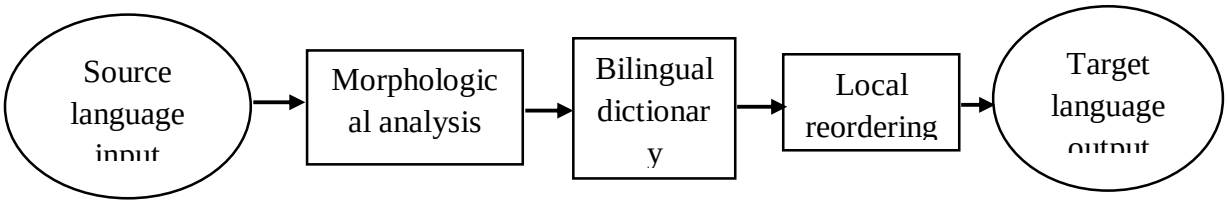

Figure 2.2: Direct machine translation

Words of source language are translated without passing through intermediately information.

Transfer-based Machine Translation Approach: it belongs to the second generation of machine translation [5]. The source language is transformed into an abstract, less language representation and an equivalent representation is then generated for the target language using bilingual dictionaries and grammar rules. It has three major components, these are:

- **Analysis:** analysis of source text is done based on linguistic information such as morphology, part-of-speech, syntax, semantics, etc.

- **Transfer**: the syntactic/semantic structure of the source language is transferred into the syntactic/semantic structure of the target language.
- **Synthesis:** its task is to replace the constituents in the source language to the target language equivalents.

Interlingua based machine translation approach: it aims at creating linguistic homogeneity across multiple languages which can be used as an intermediary between the languages. Interlingua machine translation is one instance of rule-based machine-translation approaches. In particular [5, 24], there was increasing support for the analysis of source language texts into some kind of intermediate representation, a representation of its meaning in some respect which could form the basis of generation of the target text. In this approach, the source language, i.e. the text to be translated, is transformed into an interlinguas language, i.e. a language neutral representation that is independent of any language. Finally, the target language is then generated out of the Interlingua.

In general Rule-based machine translation have its own strength and weakness. Some of its strengths are; it bases on linguistic theory, good for languages with limited resource and requires less computational resource. The need for large bilingual dictionary, need for rules, difficulty to extend as well as maintain are drawbacks of the approach.

2.9.2 Statistical Machine Translation

Statistical machine translation (SMT) is generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora [41]. The initial model of SMT based on Bayes Theorem, proposed by Brown *et al.* [38] takes the view that every sentence in one language is a possible translation of sentence in the other and the most appropriate is the translation that is assigned the highest probability by the system. SMT is a machine translation approach that uses human produced translations known as parallel corpus [24]. According to [39], the translation process by using SMT is considered as a machine learning problem. After examining the parallel corpus, SMT algorithms automatically learn how to translate new sentences. The algorithms are machine learning algorithms which learn how to translate new sentences from the parallel corpus which is a collection of previously translated texts. The translation accuracy of these systems mainly depends on the parallel corpus regarding its domain, quantity and quality. So, in order to have a good translation quality, the data must be preprocessed consistently.

SMT is a MT approach that builds probabilistic models of faithfulness and fluency, so that the most probable translation can be selected by combining models [24, 34, 41]. SMT focuses on the result of translation rather than the process. So, true translation which is both, faithful to the channel equation shows that two components are needed. These are a translation model *P(F|E)*, and a language model *P(E)*. SMT works based on the Bayesian model which translates foreign language F to English (E) or source language and the best translation is selected depending on the highest value of the translation model (*P (E|F)*) [34, 38]. Therefore the noisy channel via Bayesian rule is given as shown below.

$$\begin{aligned} E &= argmax_E p(E|F) \\ &= argmax_E \frac{p(F|E)p(E)}{p(F)} \\ E &= argmax_E\, p(F|E)p(E) \end{aligned} \tag{1}$$

Applying the noisy channel model to machine translation requires thinking of things backwards [24, 34]. It needs to pretend that the foreign (source language) input F must be translated in a corrupted version of some target (e.g. English) sentence *E*, and that the task is to discover the hidden (target language) sentence *E* that generates the observation sentence *F*. The noisy channel model of statistical MT thus requires three components to translate from a foreign sentence *F* to an English sentence *E* [5, 24, 41]. These are, language model to compute *P (E)*, translation model to compute *P (F |E)* and decoder, which is given *F* and produces the most probable *E*.

Language Model

The language model (LM) used for the International Business Machines (IBM) models is an n-gram model, which has proven to be a very strong source of information across natural language processing [34, 41]. It builds on the Markov assumption that the probability of a given word appearing in a sentence only depends on the previous *n-1* words. Even though these decisions are very local, the model provides a good evaluation of the entire sentence, since the overlapping n-grams make up a chain of events. If *n* is set to three, the model is called a trigram model. Evaluating the probability of given sentence is shown below:

$$P\,(W_n / W_1, \ldots, W_{n-1}) \tag{2}$$

The model is learned by examining a large amount of monolingual text. This is done fairly simple by maximum likelihood estimation [36]. As an example for the sentence "Mary loves John", the probability of *P(Mary| loves John)* can be estimated by observing how often 'Mary' is followed

by 'loves John'. A better result is achieved by assuming that the training data perhaps does not reflect the language perfectly.

Translation Model

Statistical methods are applied to generate translated version using bilingual corpora. This methodology uses different kinds of translation models [38, 41], namely statistical word-based translation model, statistical phrase-based model and statistical syntax-based model.

Statistical Word-Based Translation Model: Statistical machine translation is based on the idea that every target sentence is a possible translation of every source sentence [24]. A reasonable assumption is that the probability of a given target sentence being a good choice for translation relies heavily on which source sentence is under consideration for translation. It is, therefore, possible to condition the probability on the source sentence, yielding the posterior probability of the target sentence given the source sentence *P (t|s)* [40]. As it is described by [40], this model of translation correlates with an intuitive view of translation; given a source sentence that we want to translate, which is the best target sentence to choose as translation, which is target sentence that leads to the highest value for *P(t|s)*.

***Statistical Phrase-Based Translation Model*:** The job of the translation model, given source sentence *E* and a foreign sentence *F*, is to assign a probability that *E* generates *F* [23]. Modern statistical MT is based on the intuition that a better way to compute these probabilities is by considering the behavior of phrases [24, 41]. The intuition of phrase-based statistical MT is to use phrases (sequences of words) as well as single words as fundamental units of translation. The generative story of phrase-based translation has three steps [24, 41]. First grouping the source language words into phrases *e, e ...e*. Next step is to translate each source phrase $\boldsymbol{e}_i$ into a foreign phrase $\boldsymbol{f}_j$. Finally, each of the target phrases (optional) is reordered. The probability model for phrase-based translation relies on a translation probability and distortion probability. The factor $\varphi(f_j|e_i)$ is the translation probability of generating foreign phrase *fj* from source phrase e_i [34]. The reordering of the foreign phrase is done by the distortion probability of d. Distortion in statistical machine transition refers to a word having a different ('distorted') position in the foreign sentence than it had in the source sentence; it is thus a measure of the distance between the positions of a phrase in the two languages. The distortion probability in phrase-based MT means the probability of two consecutive source language phrases being separated in foreign words by a span of

particular length. Formally, the distortion is parameterized by a $d(a_i\text{-}b_{i-1})$ where a_i is the start position of the source phrase generated by the i^{th} source phrase e_i and b_{i-1} is end position of foreign phrase generated by $i\text{-}1^{th}$ source phrase of e_{i-1}. It is possible to use very simple distortion probabilities, in which it simply raises some small constants α to the distortion $D(ai\text{-}bi\text{-}1)=\alpha^{|a_i\text{-}b_{i-1}\text{-}1|}$. This distortion model penalizes large distortions by giving lower and lower probability to larger the distortion.

Statistical Syntax-Based Model: One drawback of phrase-based models is that the phrases are simply sequences of words rather than being syntactically motivated [41]. Changes in word order usually occur for syntactic reasons, and the lack of syntactically motivated translation units makes it difficult to model changes in word order effectively. If there is a need to use information about the syntax and structure of the sentence, particularly when translating between languages with different sentence structures and word order. For example, re-ordering is needed when translating between languages with subject-verb-object sentence structure such as English and languages with subject-object-verb sentence structure such as Amharic and Tigrigna. It is possible to include and use the syntax structure information of the input language, output language or in both languages by constructing trees to model the sentence structure. It can be viewed these as tree-to-string, string-to-tree and tree-to-tree models respectively. A string-to-tree syntax-based translation model that views the input language string as being generated by an output language parse tree is introduced and passed through a noisy channel; the translation system can assign P (FE) for an English parse tree E and input sentence F [42]. An English string is assigned a parse tree by a syntactic parser, then the process by which translation occurs is modelled as a reordering of the parse tree, followed by insertion of words at each node and translation of leaf words. Gains over IBM Model 5 for alignment of parallel text sentences are reported. A decoder for this syntactic model, with the addition of phrasal translation, is presented [42].

A tree transducers is developed as a generalization of the finite state transducer able to compute transformations of trees [43]. Other work builds on this using increasingly complex rules extracted from parallel text to build models of string-to-tree alignment [44, 45]. Joshua [46] is an open source toolkit for parsing-based machine translation. Other syntax-based translation systems use tree-to-tree translation to make use of syntax in both languages [47, 48]. In general the idea of Syntax-based translation is based on the idea of translating syntactic units, rather than single words or strings of words (as in phrase-based MT), i.e. (partial) parse trees of sentences/utterances.

Examples of this approach include Data-oriented parsing-based (DOP) MT and, more recently, synchronous context-free grammars.

The Decoder

Recall that, the job of decoder is to take a foreign source sentence *F* and produce the best (English) translation *E* according to the product of the translation and language model [24]. Decoders in MT are based on best-first search, a kind of heuristic or informed search; these are search algorithms that are informed by knowledge from the problem domain. Best-first search algorithms select a node *n* in the search space to explore based on an evaluation function *f (n)*. MT decoders are variants of a specific kind of best-first search called A∗ search.

A∗ search was first implemented for machine translation by IBM [24], in which its basic intuition is to maintain a priority queue with the entire partial translation hypothesis, together with their scores. In order to limit the search space in decoding, it does not want to search through the space of all source sentences; it only want to consider the ones that are possible translations for foreign sentence *F*. To help reduce the search space, only it want to consider sentences that include words or phrases which are possible translations of words or phrases in the foreign sentence *F*. This works by searching the phrase translation table, for all possible source translations of all possible phrases in *F*. the decoders can also produce ranked list of translation candidate that indicates the possibility of foreign sentences.

Alignment

All statistical translation models are based on the idea of a word alignment. A word alignment is a mapping between the source words and the target words in a set of parallel sentences [34]. Phrase alignment in MT can be also achieved by phrase translation probabilities ϕ (f_i, e_i) which needs to be trained and distortion constant α could be set if there is large bilingual training set, the foreign sentence was paired with the source sentence and if it is further known that which phrase in the foreign sentence is translated by the phrase in the source sentence [41]. The different alignment models which can help for machine translation are the IBM models.

In general, the advantage of SMT are; easy to build, no linguistic knowledge is required, reduces human cost, easy to maintain and it is translated with human translation. However, it requires large bilingual parallel corpora's, have no linguistic background, difficult to error analysis, unable to

deal with order of words and requiring large computational resource are drawbacks of the approach.

2.9.3 Example-Based Machine Translation Approach

Example-Based Machine Translation (EBMT) is characterized by its use of bilingual corpus with parallel texts as its main knowledge, in which translation by analogy is the main idea [36]. An EBMT system is given a set of sentences in the source language (from which one is translating) and corresponding translations of each sentence in the target language with point to point mapping. These examples are used to translate similar types of sentences of source language to the target language. There are four tasks in EBMT [31]: example acquisition, example base and management, example application and synthesis. The foundation of example-based machine translation is the idea of translation by analogy. The principle of translation by analogy is encoded to example-based machine translation through the example translations that are used to train such a system. Its advantages is; it uses fragments of human translation which can result in higher quality. However, it may have limited coverage depending on the size of the example database and flexibility of matching heuristics.

2.9.4 Hybrid Machine Translation Approach

Hybrid machine translation approach is developed taking advantage of both statistical and rule based translation methodologies, a new approach was developed, called hybrid based approach, which has proven to have better efficiency in the area of MT systems [5, 24, 36]. At present, several governmental and private based MT sectors [49] use this hybrid based approach to develop translation from source to target language, which is based on both rules and statistics. The hybrid approach can be used in a number of ways [5]. In some cases, translations are performed in the first stage using a rule based approach followed by adjusting or correcting the output using statistical information. In other words, rules are used to pre-process the input data as well as post-process the statistical output of a statistical-based translation system. This technique is better than the previous and has more power, flexibility, and control in translation.

In this research work, hybrid approach is chosen based on linguistic of the languages, the data availability, difficulty, error analysis and maintenance, extendibility and cost. It is explained that, hybrid approach can be used in number of ways. In this research work, the rule part is used to pre-

process the input data (Amharic corpus) of statistical based machine translation system. The reason is, the two languages are morphologically reach and less resourced; so it is costly and inappropriate to choose rule based only because it requires millions of bilingual dictionary, needs linguistic rules, difficult to maintain and extend; you need to start from the scratch all times, and it is unreliable since human languages are not consistent. Similarly, choosing statistical based approach for the study requires large bilingual parallel corpus (but not yet organized), it has no information background, and it does not deal with grammatical order and requires high computational resource. Therefore, taking the rule for pre-processing the input data of statistical machine translation can improve the statistical approach and minimizes the limitations instead of taking the individual approaches separately. As it has been described, the main objective of the rules are for pre-processing Amharic corpus not for translation purpose, i.e. to transfer the structural arrangement order of words in source sentence to the structural arrangement order of words in target sentence. This makes the sentences more similar in their structural word order, which is good for statistical approach since its probability distribution are derived based on parallel bilingual corpora.

2.9.5 Evaluation of Machine Translation

Evaluating the quality of a translation is essential, and research on evaluation methodology has played an important role from the earliest days of MT to the present. The criteria commonly used for evaluation of machine translation by human evaluators are fluency and adequacy [41, 50]. Fluency is a measure of the quality of language of the hypothesis translation ranging from a grammatically flawless sentence into incomprehensible/inconvisable [41]. Whereas adequacy measures how much of the meaning of the reference translation is expressed in the hypothesis translation [41]. The evaluators are provided with different sources of information in different areas of the screen: the hypothesis to be evaluated, the source sentence, the context of the source sentence (previous and next sentences in the same source document), the reference translation for the source sentence and the context of the reference translation (previous and next sentences in the same reference document).

Human evaluation of machine translation output is expensive in terms of both time taken and financial expense of paying bilingual and monolingual evaluators [41]. There are also issues with inter-evaluator agreement (different evaluators may give different results) and intra-evaluator consistency (the same evaluator may produce different results at different times) [41]. Inter-

evaluator is group of evaluators that evaluates individually based on the screening methods which can be difficult to make analysis. Whereas in case of intra-evaluators, once evaluated result is recorded, next time with the same evaluator and criteria the recorded result may be differ. Therefore it may not be possible to measure small changes in the quality of the output. When developing machine translation systems, it require feedback as rapidly as possible after a change has been made to the system, in order to determine quickly the success of the approaches applied. Therefore, it requires an automated method of evaluating the output of a system. For a corpus C of translation hypotheses, it uses a corpus R of human-generated reference translations. Each sentence $e \in C$ may have one or more reference translations; write R(e) for the reference sentences corresponding to e [41].

The Bilingual Evaluation Understudy (BLEU) score [51] is a metric widely used for automatic evaluation of machine translation output. The basic premise is that a translation of a piece of text is better if it is close to a high-quality translation produced by a professional translator. The translation hypothesis is compared to the reference translation, or multiple reference translations, by counting how many of the n-grams in the hypothesis appear in the reference sentence(s); better translations will have a larger number of matches. The ranking of sentences using BLEU score has been found to closely approximate human judgment in assessing translation quality [41].

Chapter 3: Related Work

3.1 Introduction

In this chapter, it is reviewed previous works done on statistical machine translation, rule based machine translation and hybrid machine translation approach. Studies that involves English with Ethiopian languages and some studies on foreign language combination are all reviewed.

3.2 Machine Translation Systems for European Languages

Greg Hanneman *et al* [52] conducted research work on translating French and German into English based on statistical transfer approach. Stat-XFER framework is a general search based and syntax driven framework for developing MT system under a variety of data conditions [53]. Statistical methods within their framework that allow for the principle extraction of syntax-based transfer rules from parallel corpora given word alignments and constituency parses is introduced. It involves constructing a bilingual translation lexicon and a transfer grammar. Methods are developed for extracting syntax-based translation lexicon and transfer rules fully automatically from parsed and word-alignment parallel corpora. Transfer grammars for statistical transfer is developed manually by the experts of the language, however it was very small. Parallel sentences are used from the Europarl training data parsed the English side with Stanford parser [54], on the French side with the Xerox XIP parser [55] and for German side pairs parsed with English and German versions of Stanford parser. Performance is evaluated on test sets from the 2007 WMT shared task. However, the Stat-XFER systems' performance lags behind the state-of-the-art scores on the 2007 test data. Low volume of training data is used for rule learning. It also unable to differentiate low-frequency correct and useful transfer rules from noisy rules that are due to parser errors and incorrect word alignments.

Preslav Nakov [19] worked on English-Spanish machine translation: basically experiments in domain adaptation, sentence paraphrasing, tokenization, and recasting based on statistical approach. Experiment with domain adaptation, combining a small in-domain news bi-text and a large out-of-domain one from the Europarl corpus, building two separate phrase translation models and two separate language models are included in their work. They have made experiment with different tokenization and recasting rules, achieving 35.09% Bleu score on the WMT'07 news test data when translating from English to Spanish, which is a sizable improvement over the highest

Bleu score achieved on that dataset at WMT'07: that is 33.10%. On the WMT'08 English to Spanish news translation, they achieve 21.92%, which makes best on Bleu score. From this work, it can be understood that building translation model and language model for each language which is different from the traditional statistical approach can produce better result in the translation system. However, it does not clearly describe how the interaction and combination between the two separate phrase translation model and language model for each language pairs.

3.3 Machine Translation Systems for Asian Languages

Mossa Ghurab *et al* [56] conducted research on Bidirectional Arabic-Chinese Machine Translations systems using phrase-based statistical approach. The core of the system implements standard phrase-based statistical machine translation architecture, where corpus data used for the systems was collected from the United-Nations website and various news engine websites. Two different metrics, BLUE and NIST are used to measure the accuracy of the translation systems. Accordingly, BLUE and NIST (National Institute of Standards and Technology) result of Arabic to Chinese translation system was obtained 0.4916 and 7.9905 similarly, Chinese to Arabic was obtained 0.4678 and 7.0643 respectively. The strength of the work is, models are then softly integrated into statistical machine translation architecture so they can interact with other models without modifying the basic architecture and as a result, phrase translation probabilities can learn directly rather than deriving them heuristically. However, there is no description clearly stating about the results obtained which are very small in case of BLUE comparing to NIST.

3.4 Machine Translation Systems Involving Ethiopian Languages

Rule based system for English-Amharic translation done by Michael Gasser [22], describes an implementation of a rule based bidirectional Amharic-to-English machine translation system within L3 framework based on Extensible Dependency on Grammar that relies on constraint satisfaction for parsing and generation. Additionally, it focuses on the features as well as the advantages that L3 offers for handling structural divergences between English and Amharic and its capacity to accommodate shallow and deep translation within a single system. It tries to show the translation in both direction between English and Amharic given simple sentence but it does not explain for complex Amharic sentence, inflections and derivations in detail. Reliability of the work is also another drawback of the research work.

Mulu Gebreegziabher and Laurent Besacier [21] conducted Preliminary Experiments on English-Amharic machine translation based on statistical approach. The experiment involves a total of 632 parliamentary corpora of which 115 is used in the experiment. A pre-processing task is performed on the parallel documents in order to retain and convert the full content into a valid format suitable for the system. Some of these pre-processes include text conversion, trimming, sentence splitting, sentence aligning and tokenization. The process of trimming was performed before and after aligning at document level. Sentence splitting was done before starting aligning at sentence level while tokenization is performed after aligning at the sentence level. The alignment at the sentence level has been done using a sentence aligner called Hunalign. The system result shows the ability to translate the basic meaning of the English sentence into Amharic sentence. Accordingly, the baseline phrase-based BLEU score result is 35.32% and a 0.32% increase in BLEU is achieved by applying morpheme segmentation to the tokens of Amharic output result and the reference of the baseline system. The work shows adding some other components like morphological analyzer and generator for Amharic language can increase the BLUE score result. This is because, applying morpheme segmentation shows increment over the prior result. However, the result with segmentation process is not much more than the baseline result and it is not clearly described in the research work.

Jabesa Daba and Yaregal Assabie [24] conducted Bidirectional English-Oromiffa machine translation system based on hybrid of rule based and statistical approaches. Since the two language has different sentence structure, reordering rules are developed for simple, interrogative and complex English and Oromiffa sentences that changes the order of words to their corresponding targeted language. The researcher conducted two groups of experiments using purely statistical approach and hybrid approach. The Oromiffa-English SMT yields a BLEU score of 41.50% whereas English-Oromiffa SMT have a BLEU score of 32.39%. The second experiment was carried out using Hybrid approach and the results are; a BLEU score of 37.41% for English to Afaan Oromo translation and 52.02% for Afaan Oromo to English translation. From the result, we can understand hybrid approach is better than the statistical approach for the language pairs. Disambiguation of words in the translated results are observed.

Eleni Teshome [6] conducted bidirectional English-Amharic machine translation system using statistical approach. In the study, two different English-Amharic parallel corpuses are collected

from different sources; the first corpus consists of 1,020 simple sentences and the other one consists of 1951 complex sentences. Two different experiments are conducted and the evaluation was performed using two different methodologies. The result obtained for simple sentences using BLEU score has an average score of 82.22% accuracy for English to Amharic translation and 90.59%, for the Amharic to English translation. For complex sentences, the result obtained from the BLEU score is 73.38% for English to Amharic translation and 84.12% for Amharic to English translation. The studies shows Amharic to English translation have a better accuracy than English to Amharic translation. The drawback is, testing data's for measuring performance of the translation are included at the training data and it makes the system performance biased.

3.5 Summary

The studies have been done using statistical machine translation approach, rule-based machine translation and hybrid machine translation approach. In most of the studies, statistical approach is used as a base machine translation system. However, for example, in bidirectional English-Oromiffa machine translation system using hybrid approach, syntactic reordering approach in which rules are written in order to make the structure of the source sentence to be similar to the target language is used. Then, translation model and language modeling are done using statistical approach. From the study, it can be understood that, adding syntactic reordering approach as a pre-processing step to input data of statistical approach can advance the performance of the translation system.

Apart from the above related work, no prior work on machine translation system involving Tigrigna language is found to knowledge of the researcher. Moreover, statistical methods are proven to be better than the other approaches. Local reordering benefits especially for similar language and we have seen that Amharic and Tigrigna are syntactically similar. Thus, we expect that local reordering improves the performance of Amharic-to-Tigrigna statistical machine system. Accordingly, we hypothesize a hybrid approach involving rule based and statistical based machine translation can improve the translation system.

Chapter 4: Design of Amharic-to-Tigrigna Machine Translation

4.1 Introduction

The objective of the study is to develop Amharic-to-Tigrigna machine translation system. In order to be accomplished the study, some measures are taken. Appropriate approach is selected, parallel corpus are collected from various domains, appropriate syntactical local reordering rules for Amharic texts are identified and different tools are used to develop the translation system. In general, this chapter briefly describes how the system work and the overall system architecture including its components.

4.2 Architecture of the System

The Amharic-to-Tigrigna Translation System is hybrid system where a given Amharic text is locally reordered before applying statistical methods for translation. We use Amharic POS tagger to identify word classes which help to apply local reordering rules. After the Amharic text is locally reordered takes the form of target language text, we apply the statistical methods for translation. The translation model is built using the parallel corpus of Tigrigna text and locally reordered Amharic (rA) text. The language model is built using corpus of Tigrigna text only since the translation is one way direction which is from Amharic to Tigrigna. The architecture of the system is shown in Fgure 4.3.

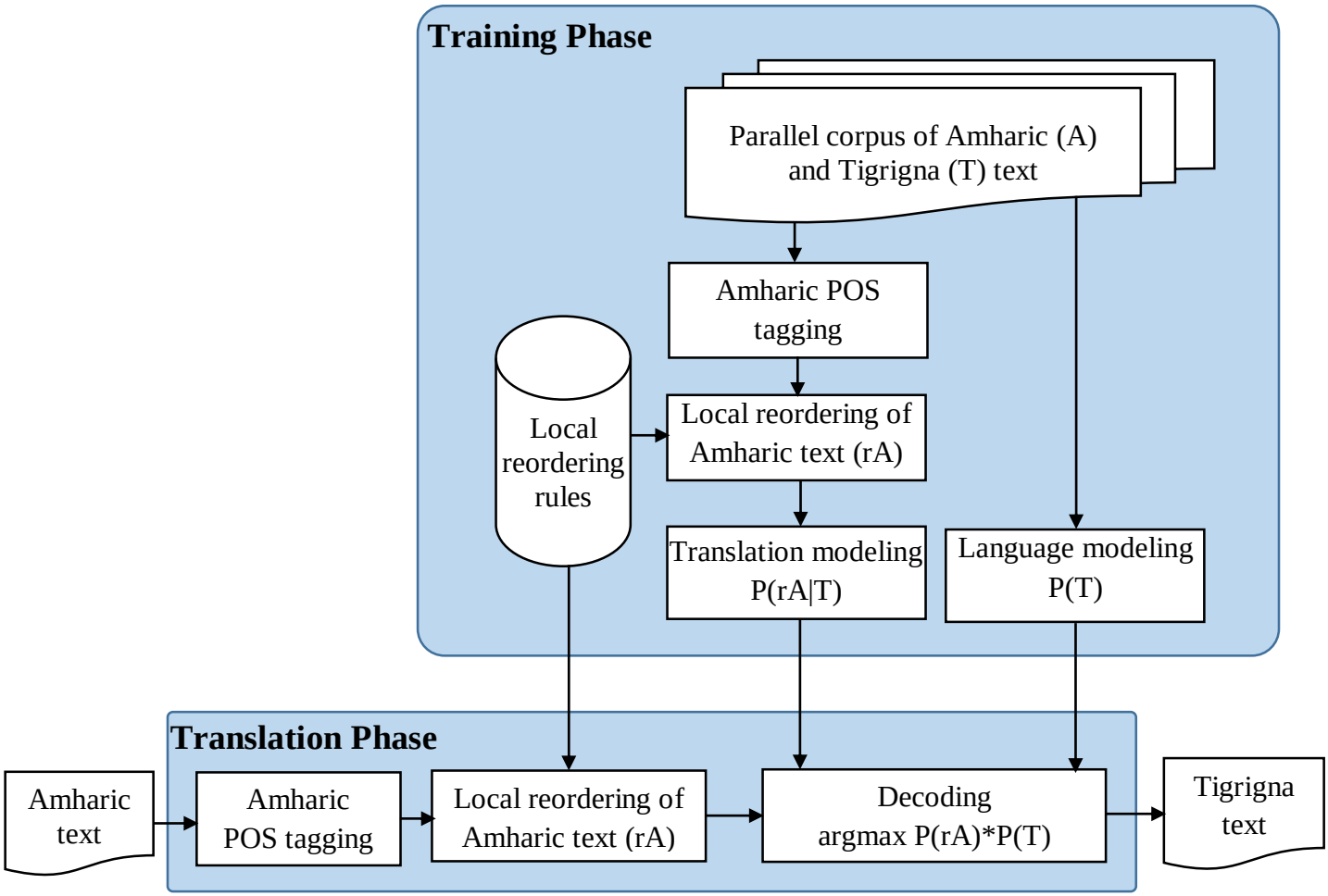

Figure 4.3: Architecture of the System

4.3 Training Phase

4.3.1 Parallel Corpus

Corpora are developed for the purpose of this study since most of the tasks are done in statistical machine translation approach. The corpus are composed of simple sentences and complex sentence collected manually from different sources such as literal documents, social Medias, and websites and revisit data's. The sentences are translated to each other with help of language expert and prepared in parallel. Other sentence are taken from Bible, federal and regional constitution which contains Amharic and Tigrigna sentences in parallel.

4.3.2 Language Modeling

The translation system is designed to work from Amharic-to-Tigrigna using hybrid of statistical and rule based approach. Thus, the corpus that contains Tigrigna texts is an input to the language model i.e *P (T)*. The role of the language model is to give an estimate of how probable a sequence of words is to appear in the target language. The language model helps the translation system with selecting words or phrases appropriate for the local context and with combining them in a sequence with better word order. It builds on the Markov assumption that the probability of a given word appearing in a sentence depends on the previous *n-1* words. Particularly in this study the language model uses N-gram model which takes monolingual corpus containing texts of the target language and compute the probability of each word, which can be phrase or sentence. The probability obtained from the N-gram model could be unigram, bigram trigram or other. If *n* is set to three, the model is called a trigram model. Generally evaluating the probability of given text is calculated as follows:

$$P\ (W_n/W_{1\ldots}W_{n-1}) \quad (3)$$

The following examples are simple Tigrigna sentences:

ሃይላይ እንጅራ በሊዑ /haylay ^njera beli`u/Hailay ate injera

ሃይላይ እንጅራ ኣይበለዐን/haylay ^njera ^aybele`en/Hailay did not ate injera

ሃይላይ ናብ ኣክሱም ከይዱ /haylay nab ^aksum keydu/Hailay went to Aksum

ኣበባ ንፍዕቲ እያ/^abeba nf`ti ^ya/Abeba is clever

The unigram probability can be computed as follows:

Considering the Tigrigna words represented as ***t*** and Amharic words as ***a***, it calculates in the following forms.

$$p(t_1) = \frac{count\ (t_1)}{total\ words} => \ p(\text{ሃይላይ}) = \frac{count\ (\text{ሃይላይ})}{count(total\ word)} = \frac{3}{13} = 0.23$$

Where, 3 indicate the number of the word "ሃይላይ" appeared in the given sentence and 13 represent the total words.

The bigram probability can be also computed as shown below:

$$p(t_2/t_1) = \frac{count\ (t_1t_2)}{count(\ t_1)} => \ p(\text{እንጀራ}/\text{ሃይላይ}\ ) = \frac{count\ (\text{ሃይላይ እንጀራ})}{count(\ \text{ሃይላይ})}\ \frac{2}{3} = 0.67$$

In this case, 2 indicate the number of "*ሃይላይ*" *and* "*እንጀራ*" words used together and 3 indicates the number of times word "*ሃይላይ*" used.

The last but not the least, trigram is computed by:

$$p(t_3/t_1t_2) = \frac{count\ (t_1t_2\ t_3)}{count(\ t_1\ t_2)} => \ p(\text{ለሊው}/\text{ሃይላይ እንጀራ}\ ) = \frac{count\ (\text{ሃይላይ እንጀራ ለሊው})}{count(\ \text{ሃይላይ እንጀራ})} = \frac{1}{2}$$
$$= 0.5$$

Where, "*1*" indicates the number of time's "*ሃይላይ*", *እንጀራ*" *and* "*ለሊው*" words used together and 2 represents the number of times words "*ሃይላይ*" *and* "*እንጀራ*" used together.

4.3.3 Amharic POS Tagging

This component is the first step in the rule part and its purpose is to assign part of speech to each word/phrase for Amharic sentence since the translation is from Amharic to Tigrigna. The tagged sentence are used as an input in Amharic reordering rules. Though different researches have done on Amharic language, there is no publically available POS tagger tool for Amharic at present. The following are the POS tag sets used in Amharic language. These are: PN(personal noun), CN(compound word that changes order of words in the target language) , N2(compound words that never change order of words in the target language), N(noun), VN(verbal noun), PRON(pronoun), V(verb), AXU(auxiliary verb), VREL(relative verbs), ADJ(adjective), NUM(number), NUMCR(cardinal number), NUMOR(ordinal number), PRP(prepositions that have similar positional order like the target language), IN(prepositions that have different positional order from the target language) , ADV(adverb), PUN(punctuation), CC(Conjunctions and subordinate conjunction) and UNC (unclear). All except VN, AXU, VREL, NUMOR, NUMCR, PRP, IN, ADV, CC and UNC have further classification by attaching either prepositions or conjunction or both at a time. For example, noun and verb attached to prepositions are tagged as NP and VP respectively. Similarly, the later tags attached with preposition and conjunction are included in the former classification. For example, NUMOR attached with preposition is represented as NUMP. In general, different POS tag sets are prepared and the Amharic corpus are tagged manually using the tag sets shown in Annex I.

The most relevant tags in this research work are; PN (personal noun, it can be attached with preposition and conjunction), CN, CNP, CNC, CNPC (compound word (CN), CN attached with preposition, CN attached with conjunction, CN attached with both at a time respectively), N2, N2P, N2C, N2PC(compound words (N2), N2 attached with preposition, N2 attached with conjunction, N2 attached with both at a time respectively), N, NP, NC, NPC (noun (N), N attached with preposition, N attached with conjunction, N attached with both at a time respectively), ADJ (adjective), NUMOR (ordinal number), NUMCR (cardinal number), NUMP/number attached with preposition and NUMPC (number with preposition and conjunction). Most of the POS tag sets are taken from Tagging and Verifying Amharic News Corpus [57] that is tagged by Ethiopian language research center of Addis Ababa and some of them are added by the researcher's with help of language experts.

4.3.4 Amharic Local Reordering Rules

Amharic and Tigrigna belongs to same family of languages. These two languages have much commonality. The sentence structure of both languages is Subject-Object-Verb (SOV) that could be suitable for statistical machine translation since it is based alignment of words or phrases. As described in Chapter Two, the noun, adjectives and verbs are highly marked for number, person, gender and tense in the same way in both languages. The derivational and inflectional behavior of words in Amharic and Tigrigna languages is also similar with different form of affixes as described in the literature part. In the construction of Amharic and Tigrigna noun phrases and prepositional phrase, there are differences in the structural arrangement order of words in various types of sentences. However, there are also Amharic noun phrase and prepositional phrase that have similar structural arrangement order of words as Tigrigna noun phrase and prepositional phrase.

Therefore, rules are developed to perform local reordering on Amahric texts i.e. (rA). the main purpose of the rules are identifying Amharic sentences in the corpus that have difference in their structural arrangement order of words and change the order of words in it to have a more similar sentence structural order of words with Tigrigna sentence structural order of words. Local reordering rules are not applied for Amharic sentence in the corpus that have similar structural arrangement order of words as the target language in the corpus. The following sentences are examples of Amharic language that do not require local reordering rule:

A: ብዙ/ADJ ህፃናት/N ያለረዳት/NP ያድጋሉ/V። (bzu h^Sanat yaleredat yadgalu/ Many children growing without help).

T: ብዙሓት/ADJ ህፃናት/N ብዘይሓጋዚ/NP ይዓብዩ እዮም/V።/PUN (bzuHat h^Sanat bzeyHagazi y`abyu ^yom).

As it can be clearly seen from these examples, Amharic sentence and their translation in Tigrigna, the structural order of words in both languages are similar with slight differences on the number of words to represent one idea in both languages (e.g. an Amharic verb, ያድጋሉ/V is represented with two words in Tigrigna ይዓብዩ እዮም/V. Additionally, in Some Amharic imperfective verbs and a word either noun or adjective with definiteness, sometimes it is represented in two words in case of Tigrigna. Thus, for these kinds of sentences, local reordering rule are not applied.

Generally, the study discovers thirteen local reordering rule for Amahric texts that needs to be reordered. The purpose for the rules are to perform local reordering of Amharic texts into the form of the target language. Thus, local reordering rules are shown as follows.

- Rule 1: Local Reordering Rule for Amharic Prepositional Phrases
- Rule 2: Local Reordering Rule for Compound Word (N2), Adjective, Number and Noun Combination of Amharic Noun Phrase
- Rule 3: Local Reordering Rule for Compound Word (N2), Compound Word (CN), and noun (N) Combination of Amharic Noun Phrase
- Rule 4: Local Reordering Rule for Noun (N), Compound Word (CN) and Compound Word (CN) Word Combination of Amharic Noun Phrase
- Rule 5: Local Reordering Rule for Compound Word (CN), Adjective and Noun Combination of Amharic Noun Phrase
- Rule 6: Local Reordering Rule for Noun and Adjective Word Combination of Amharic Noun Phrase
- Rule 7: Local Reordering Rule for Compound Word (CN) and Compound Word (CN) Word Combination of Amharic Noun Phrase
- Rule 8: Local Reordering Rule for Compound Word (N2) and Compound Word (N2) Combination of Amharic Noun Phrase
- Rule 9: Local Reordering Rule for Compound Word (N2) and Compound Word (CN) Combination of Amharic Noun Phrase

- Rule 10: Local Reordering Rule for Number and Noun Combination of Amharic Noun Phrase
- Rule 11: Local Reordering Rule for Compond Word (N2) and Noun Combination of Amharic Noun Phrase
- Rule 12: Local Reordering Rule for Noun and Compound Word (CN) Combination of Amharic Noun Phrase
- Rule 13: Local Reordering Rule for Compound Word (CN) Amharic Noun Phrase

The above local reordering rules discovered in this research needs a mechanism how they are applied to the corpus. This is to give a highlight of the local reordering rules how they are applied. So, it presents about the priority given to local reordering rules. Accordingly, the priority of Amharic local reordering rules are given below;

High level procedure for local reordering rules is done using algorithm 4.1 given below.

```
Load Amharic sentences S from the tagged corpus
  For each sentence s_i in S, i=1, 2, 3….n, where n is the number of sentences in S
       Extract patterns Pt in s_i // Pt is the set of consecutive tag set
       If Pt== (‘PN’ or ‘N’ or ‘CN’ or ‘N2’, ‘IN’) or (‘NUMP’, ‘N’, ‘IN’)
              Apply RULE 1
        End if
        If Pt== (‘NP’, ‘ADJ’, ‘ADJ’, ‘N’, ‘N2’) or (‘NP’, ‘NUMOR’, ‘ADJ’, ‘N’)
              Apply RULE 2
        End if
        If   Pt == (‘N2’, ‘CN’, ‘N’) or (‘N2’, ‘CN’, ‘CN’)
              Apply RULE 3
        End if
       If Pt == (‘NP, ‘CN’, ‘CN’) or (‘CN’, ‘N’, ‘CN’) or (‘CN’, ‘CN, ‘N’ or ‘NC’)
           Apply RULE 4
        End if
       If Pt== (‘CN’, ‘ADJ’, ‘N’) or (‘NP’, ‘ADJ’, ‘CN’)
             Apply rule 5
         End if
     If Pt== (‘NP’, ‘ADJ’, ‘N’ or ‘NC’)
           Apply RULE 6
       End if
       If Pt== (‘CN’, ‘CN’)
                Apply RULE 7
         End if
      If Pt== (‘N2’, ‘N2’)
            Apply RULE 8
        End if
      If Pt== (‘N2’, ‘CNP’)
           Apply RULE 9
          End if
       If Pt== (‘NUMOR’ or ‘NUMPC’, ‘N’, ‘N’)
           Apply RULE 10
        End if
      If Pt== (‘N2, ‘N’) or (‘NP’, ‘N2’) or (‘NP’, ‘N2’, ‘N’ or ‘NC’)
            Apply RULE 11
        End if
     If Pt== (‘NP’, ‘CN’, ‘N’ or ‘NC’) or (‘CN, ‘N’ or ‘NC’) or (‘NP’, ‘CN’)
            Apply RULE 12
      End if
     If    Pt== (‘CN’ or ‘CNP’ or ‘CNC’ or ‘CNPC’)
           Apply RULE 13
      End if
End for
```

Algorithm 4.1: Algorithm for high level procedure local reordering

The overall local reordering rules applied in this study are stated as follows.

Rule 1: Local Reordering Rule for Amharic Prepositional Phrases

A prepositional phrase is a phrase that begins with a preposition and ends with a noun or pronoun of the preposition. The object of the preposition will often have one or more modifiers to describe it. According to literatures, Amharic Prepositions can appear before or after noun and adjectives. The difference happen at prepositions appear after noun and adjective. Literatures of Tigrigna language describes that every preposition are appear before noun/pronoun or an adjective. In this section Amharic preposition that have different order from the target language are identified and reordering rule is applied. For example, in the following phrases, the underlined words are Amharic prepositions that appear after noun:

A: ቤት/N ውስጥ/IN (bEt wsT/In the house)

T: ኣብ ውሽጢ/IN ገዛ/N (^ab wxTi geza)

A: በአንድ/NUMP ወር/N ውስጥ/IN (be^and wer wsT/within one month)

T: ኣብ ውሽጢ/IN ሓደ/NUM ወርሒ/N (^ab wxTi Hade werHi)

As it is shown above Amharic phrases and their translation in Tigrigna, the prepositions appear after nouns in Amharic whereas in Tigrigna the prepositions appear before the noun. Thus, to avoid this kind of problems, words in Amharic phrases are reordered in similar structure to Tigrigna. After applying the reordering rule to the above examples, the following reordered Amharic phrase are obtained:

rA: ውስጥ ቤት (wsT bEt)

T: ኣብ ውሽጢ ገዛ (^ab wxTi geza)

rA: ውስጥ በአንድ ወር (wsT be^and wer)

T: ኣብ ውሽጢ ሓደ ወርሒ (^ab wxTi Hade werHi)

Reordering the above prepositional phrase is done using algorithm 4.2

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1, 2, 3….n, where n is the number of sentences in S
  Extract word W in si //segmentation process based on tag set
  For each word Wj ε si, j=1, 2, 3….k, where k is the number of words in si
     If POS (Wj) is 'PN' or 'N' or 'CN' or 'N2'and POS (Wj+1) is 'IN'
           Swap (Wj, Wj+1)
      End if
      Else if POS (Wj) is 'NUM' and POS (Wj+1) is 'N' and POS (Wj+2) is 'IN'
           temp =Wj+2
           temp1=Wj+1
           Wj+2=temp1
           Wj+1=Wj
           Wj= temp
       End else if
      End for
    End for
```

Algorithm 4.2: Reordering Algorithm for Amharic prepositional phrase

The algorithm also works for the given tag sets attached with preposition and conjunction. The algorithm has three major tasks, first load POS tagged sentence, extract the words in each sentence, next the part of speech of each word is checked. Finally local reordering is applied.

Local Reordering Rule for Amharic Noun Phrase

As it has been discussed in the literature part, Section 2.3.1, a noun phrase is a syntactic unit in which the head is a noun or a pronoun and it can be simple or complex. In fact, there might be noun phrases in Amharic and Tigrigna that have similarity in structural order, but when Amharic noun phrases are constructed especially from compound words, more complements and modifiers, there are several conditions in which structural order of words in the phrase can be changed in Tigrigna language. Therefore the main objective of syntactic reordering is to model this difference. Compound words of both langauges can be constructed from the combination of noun, adjuctive and verbs. In the modern Amharic and Tigrigna linguistic books [27, 32], the syntactical structure, morphology, category of phrase and part-of-speech are all described. A noun can be constructed from more than two words classified as noun in the two languages and named as simply noun and similarly a single word can be also named as noun. In similar manner, a phrase that contains an

adjective followed by noun can be called as noun phrase and a phrase contains an adjective followed by compound noun can be also called as noun phrase. In doing so, the challenge encountered is lack of *a* consistent name for noun phrase constructed from nouns and other constitutes like complements and modifiers. The complements and modifiers can be compound word, adjectives, number and prepositions. The reason behind is several Amharic noun phrases have different word order in Tigrigna even though they are classified as noun phrase. So, in order to avoid this problem it is proposed to name the several types of Amharic noun phrases with their combination, just like if the noun phrase is constructed from compound noun (CN), it is called as reordering rule for compound word (CN) Amharic noun phrase or if it is constructed from compound word and noun, it is called as reordering rule for compound word and noun combination Amharic noun phrase and so on. Based on the nature of the two languages and the idea stated before, reordering rules have been provided for different Amharic noun phrase to solve the structural divergence of order of words in a given Amharic sentences in the following section.

Rule 2: Local Reordering Rule for Compound Word (N2), Adjective, Number and Noun Combination of Amharic Noun Phrase

Amharic sentence can contain noun phrase constructed from compound word (N2) that never change its word order in the target language, adjuctive (ADJ), number (NUM) and noun (N) word classes at the same time which is structurally different from the target language. Particularly, a sentence with a sequence of NP, N2, ADJ, ADJ, N and NP, NUMOR, ADJ, N are to be considered in this reordering rule. The other combination are either included in the other reordering rules or they do not need reordering technique. The following underlined phrases are examples of the two languages constructed from these words:

A: <u>የኢትዮጵያ/NP ፌዴራላዊ/ADJ ዲሞክራሲያዊ/ADJ ሪፐብሊክ/N ሕገ መንግስት/N2</u> (ye^ityoPya fEdEralawi dimokrasiyawi ripeblik Hge mengst) (NP, ADJ, ADJ, N, N2) ከኅሳሴ 15 1987 ዓ/ም ጀምሮ ሙሉ በሙሉ በስራ ላይ ወሏል።/ keneHasE 15 1987 `am jemro mulu bemulu besra lay welWal/ The Constitution of the Federal Democratic Republic of Ethiopia fully at work starting from August 21 1995.)

T: <u>ሕገ መንግስቲ/N2 ፌዴራላዊ/ADJ ዲሞክራሲያዊ/ADJ ሪፐብሊክ/N ኢትዮጵያ/N</u> (Hge mengsti fEdEralawi dimokrasiyawi ripeblik ^ityoPya) (N2, ADJ ADJ, N, NP) ካብ ነሓሰ 15 1987 ዓ/ም ኣትሒዙ ብዕሊ ኣብ ስራሕ ውዒሉ ኣሎ። (kab neHase 15 1987 `am ^atHizu b`li ^ab sraH w`ilu ^alo)

A: የኢህአዲግ/NP 4ኛ/NUMCR ድርጅታዊ/ADJ ጉባኤ/N ማምሻውን ተከፈተ (NP, NUMOR, ADJ, N) (ye^ihadig 4Na drjtawi guba^E mamxawn tekefete/The EPRDF council has started its meeting at the evening)

T: 4ይ/NUMCR ውድባዊ/ADJ ጉባኤ/N ኢህወዴግ/N ኣማስዩ ተኸፊቱ (NUMOR, ADJ, N, NP) (4y wdbawi guba^E ^ihwedEg ^amasyi teKefitu)

One can clearly understand from these Amharic sentences and their translation in Tigrigna, there are two possible combinations of Amharic noun phrases that has different structural order in Tigrigna. The combinations are (NP, ADJ, ADJ, N, and N2) and (N2P, CN and CN) and (NP, NUMOR, ADJ, N). In Tigrigna, for the first combination words found at the begging and last of Amharic phrases will swap their order and in the second combination, the first word of the Amharic phrase will move to the right end of the phrase. Thus, to solve such kind of problems we have provided reordering rule so that it will have similar structure as Tigrigna language. After applying the reordering rule we can get the following Amharic phrases:

rA: ሕገ መንግስት ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ የኢትዮጵያ (Hge mengst fEdEralawi dimokrasiyawi ripeblik ye^ityoPya)

T: ሕገ መንግስቲ ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ ኢትዮጵያ (Hge mengsti fEdEralawi dimokrasiyawi ripeblik ^ityoPya)

rA: 4ኛ ድርጅታዊ ጉባኤ የኢሃዲግ (4Na drjtawi guba^E ye^ihadig)

T: 4ይ ውድባዊ ጉባኤ ኢህወዴግ (4y wdbawi guba^E ^ihwedEg)

Reordering the above task is done using Algorithm 4.3 given below:

Load Amharic sentences S from the POS Tagged corpus
For each sentence s_i ε S, i=1, 2, 3….n, where n is the number of sentences in S
  Extract words W in s_i //segmentation process based on tag set
  For each word W_j ε s_i, j=1, 2, 3….k, where k is the number of words in s_i
   If POS (W_j) is 'NP' and POS (W_{j+1}) is 'ADJ' and POS (W_{j+2}) is 'ADJ' and POS (W_{j+3}) is 'N' and POS (W_{j+4}) is 'N2'
        Swap (W_j, W_{j+4})
   End if
   Else if POS (W_j) is 'NP' and POS (W_{j+1}) is 'NUMOR' and POS (W_{j+2}) is 'ADJ' and POS (W_{j+3}) is 'N'
        temp =W_{j+3}
        temp1=W_{j+2}
        temp2=W_{j+1}
        W_{j+3}=W_j
        W_j=temp2
        W_{j+1} =temp1
        W_{j+2}=temp
    End else if
   End for
End for

Algorithm 4.3: Reordering Algorithm for Amharic noun phrase N, NUM, N2 and ADJ word combination

Rule 3: Local Reordering Rule for Compound Word (N2), Compound Word (CN), and noun (N) Combination of Amharic Noun Phrase

Amharic noun phrase can be constructed from compound word (N2); words that never change its word order in the target language, compound word (CN); words that changes its word order in the target language and noun (N) combination. In this section it is studied the structural word order difference between the languages and provided a rule to localize the order of Amharic words in the sentences to Tigrigna sentences. The rule also works for the tag set attached with their corresponding preposition and conjunction. The following underlined phrases of both languages are combination of N2, CN and N:

A: <u>የአዲስ አበባ/N2P ውሀና ፈሳሽ/CN ባለስልጣን/N</u> (ye^edis ^ebeba whena fesax baleslTan/Addis Ababa water and sewerage authority)

T: <u>በዓልስልጣን/N ፈሳስን ማይን/CN ኣዲስ ኣበባ/N2</u> (be`alwana fesasn mayn ^adis ^Abeba)

A: <u>የአዲስ አበባ/N2P ውሀና ፈሳሽ/CN ባለስልጣን ቢሮ/CN</u> (ye^edis ^ebeba whena fesax baleslTan biro/Office of Addis Ababa water and sewerage authority)

T: <u>ቢሮ በዓልዋና/CN ፈሳስን ማይን/CN ኣዲስ ኣበባ/N2</u> (biro be`alwana fesasn mayn ^adis ^Abeba)

As it is shown in these Amharic phrases and their translation in Tigrigna, there are two possible combinations of Amharic noun phrases that have different structural order in Tigrigna. The combinations are (N2P, CN and N) and (N2P, CN and CN). In Tigrigna, these Amharic phrases are shown similarly in reverse order. Thus, to solve such kind of problems we have provided reordering rule so that it will have similar structure as Tigrigna language. After applying the reordering rule we can get the following Amharic phrases:

rA: ባለስልጣን ፈሳሽ ውሀና የአዲስ አበባ (baleslTan fesax whena ye^edis ^ebeba)

T: በዓልስልጣን ፈሳስን ማይን ኣዲስ ኣበባ be`alslTan fesasn mayn ^adis ^abeba)

rA: ቢሮ ባለስልጣን ፈሳሽ ውሀና የአዲስ አበባ (biro baleslTan fesax whena ye^edis ^ebeba)

T: ቢሮ በዓልስልጣን ፈሳስን ማይን ኣዲስ ኣበባ (biro be`alwana fesasn mayn ^adis ^Abeba)

Reordering the above task is done using Algorithm 4.4:

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1,2,3,….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3,….k, where k is the number of words in si
      If POS (Wj) is 'N2' and POS (Wj+1) is 'CN' and POS (Wj+2) is 'N'
            temp =Wj+2
            str =Apply reordering algorithm 4.10
            W(j+1,j+2)=str
            Wj=temp
         End if
         Else if POS (Wj) is 'N2' and POS('Wj+1') is 'CN' and POS('Wj+2') is 'CN'
            temp=Pj+2
            str=Apply reordering algorithm 4.10
            P(j+1,j+2)=str
            Pj=temp
         End else if
      End for
   End for
```

Algorithm 4.4: Reordering Algorithm for Amharic noun phrase N2, CN and N word combination

Rule 4: Local Reordering Rule for Noun (N), Compound Word (CN) and Compound Word (CN) Word Combination of Amharic Noun Phrase

Amharic noun phrase can be constructed from noun with more than one compound words, particularly two in this section. There are Amharic noun phrases combination of the above words which have different structural order in Tigrigna.

For example, the following underlined phrase of both languages are combination of CN and N:

A: የተማሪዎች/NP ኣገልግሎት መስጫ/CN ጽህፈት ቤት/CN (yetemariwoc ^agelglot mesCa Shfet bEt/ Office of the student service)

T: ቤት ፅሕፈት/CN መውህቢ ግልጋሎት/CN ተምሃሮ/N (bEt SHfet mewhebi glgalot temaharo)

A: የማከላዊ ዞን/CNP ፖሊስ/N ጽህፈት ቤት/CN (yema^kelay zon polis Shfet bEt/ Central zone police office)

T: ቤት ፅሕፈት/CN ፖሊስ/N ዞባ ማእኸል/CN (bEt SHfet polis zoba ma^Kel)

A: የጤና ጥበቃና/CNPC ትምህርት ሚኒስቴር/CN ሰራተኞች/N (yeTEna Tbeqana tmhrt ministEr serateNoc/ Workers of the ministry of health and ministry of education)

T: ሰራሕተኛታት/N ሚኒስቲሪ ትምህርትን/CNC ሓለዋ ጥዕናን/CNC (seraHteNatat ministiri tmhrtn Halewa T`nan)

As shown in these examples, Amharic phrases and their translation in Tigrigna, there are three possible combinations of Amharic noun phrases that have different structural order in Tigrigna. The combinations are (NP, CN and CN), (CN, N and CN) and (CN, CN and N). In Tigrigna, the three Amharic phrases are shown similarly in reverse order. Thus, to solve such kind of problems we have provided reordering rule so that it will have similar structure as Tigrigna language. After applying the reordering rule we can get the following Amharic phrases:

rA: ቤት ጽህፈት መስጫ ኣገልግሎት የተማሪዎች (bEt Shfet mesCa agelglot yetemariwoc ^)

T: ቤት ፅሕፈት መውህቢ ግልጋሎት ተምሃሮ (bEt SHfet mewhebi glgalot temaharo)

rA: ቤት ጽህፈት ፖሊስ ዞን የማእከላይ (bEt Shfet polis zon yema^kelay)

T: ቤት ፅሕፈት ፖሊስ ዞባ ማእኸል (bEt SHfet polis zoba ma^Kel)

rA: ሰራተኞች ሚኒስቴር ትምህርት ጥበቃና የጤና (serateNoc ministEr tmhrt Tbeqana yeTEna)

T: ሰራሕተኛታት ሚኒስቲሪ ትምህርትን ሓለዋ ጥዕናን (seraHteNatat ministiri tmhrtn Halewa T`nan)

Reordering the above task is done using Algorithm 4.5.

Load Amharic sentences S from the POS Tagged corpus
For each sentence s_i ε S, i=1, 2, 3….n, where n is the number of sentences in S
Extract word W in s_i //segmentation process based on tag set
For each word W_j ε s_i, j=1, 2, 3….k, where k is the number of words in s_i
If POS (W_j) is 'NP' and POS ('W_{j+1}') is 'CN' or 'CNP' or 'CNC' or 'CNPC' and POS (W_{j+2}) is 'CN' or 'CNP' or 'CNC' or 'CNPC'
Temp =W_j
str =Apply reordering algorithm 4.8
W ($_{j, j+1}$) =str
W_{j+2}=temp
End if
Else if POS (W_j) is 'CN' or 'CNP' or 'CNC' or 'CNPC' and POS (W_{j+1}) is 'N' and POS (W_{j+2}) is 'CN' or 'CNP' or 'CNC' or 'CNPC'
temp=W_{j+2}
str=Apply reordering algorithm 4.13
W ($_{j+1, j+2}$) =str
W_j =temp
End else if
Else if POS (W_j) is 'CN' or' CNP' or 'CNC' or 'CNPC' in and POS (W_{j+1}) is 'CN' or 'CNP' or 'CNC' or 'CNPC' and POS (W_{j+2}) is 'N' or 'NC'
temp=W_{j+2}
str=Apply reordering algorithm 4.8
W_j =temp
W ($_{j+1, j+2}$) =str
End else if
End for
End for

Algorithm 4.5: Reordering Algorithm for Amharic noun phrase CN, N and CN word combination

Rule 5: Local Reordering Rule for Compound Word (CN), Adjective and Noun Combination of Amharic Noun Phrase

Some Amharic noun phrase combination of compound words (CNs), adjectives and nouns has different structural word order in Tigrigna noun phrase. For example an Amharic phrase combination of CN, ADJ and N with their corresponding preposition and conjunction for valid sentences and phrases, the order of the compound word will move to the right end of the noun

phrase but the other move back with similar order in the case of Tigrigna language. The following underlined example of both languages are composed of CN, ADJ and N:

A: የአዳማ ከተማ/CNP ዋና/ADJ አስተዳደር/N (ye^edama ketema wana ^estedader/Adama city general administration)

T: ዋና/ADJ ኣማሓዳሪ/N ከተማ ኣዳማ/CN (wana ^amaHadari ketema ^adama)

A: የወረዳው/NP ዋና/ADJ ስራ አስኪያጅ/CN (yeweredaw wana sra ^eskeyaj/The chief executive of the woreda)

T: ዋና ስራሕ መካየዲ እቲ ወረዳ (wana sraH mekayedi ^ti wereda)

These examples clearly indicated that Amharic phrases and their translation in Tigrigna, there are two types of combinations which makes difference between the languages. The first combination is CN, ADJ, and N and the second is NP, ADJ and CN combination. Normally in Tigrigna, for such kind of combination the words before the adjective will move to the right end of the phrase. Thus, to solve such kind of problems we have provided reordering rule so that it will have similar structure as Tigrigna language. After applying the reordering rule we can get the following Amharic phrases:

rA: ዋና አስተዳደር ከተማ የአዳማ (wana ^estedader ketema ye^edama)

T: ዋና ኣማሓዳሪ ከተማ ኣዳማ (wana ^amaHadari ketema ^adama)

rA: ዋና ስራ አስኪያጅ የወረዳው (wana sra ^eskeyaj yeweredaw)

T: ዋና ስራሕ መካየዲ እቲ ወረዳ (wana sraH mekayedi ^ti wereda)

The above reordering task is done by using Algorithm 4.6:

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1,2,3,….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3….k, where k is the number of words in si
       If POS (Wj) is 'CN' and POS (Wj+1) is 'ADJ' and POS ('Wj+2') is 'N'
            temp =Wj+2
            temp1=Wj+1
            Wj+2 =Wj
            Wj+1 =temp
            Wj =temp1
        End if
        If POS (Wj) is 'NP' and POS (Wj+1) is 'ADJ' and POS ('Wj+2') is 'CN'
            temp =Wj+2
            temp1=Wj+1
            Wj+2 =Wj
            Wj+1 =temp
            Wj =temp1
         End else if
      End for
   End for
```

Algorithm 4.6: Reordering Algorithm for Amharic noun phrase combination of CN, ADJ and N.

Note: if the Amharic noun phrase constructed from CN or NP or N2 or nothing followed by ADJ and N or CN then, phrase left to adjective will be shifted to the right position where as the adjective and its head word will not change their order. The following example describes the above descriptions:

rA: አለምአቀፍ/ADJ የሴቶች ቀን/CN (^elem^eqef yesEtoc qen/International women's day): this Amharic phrase is composed of ADJ and CN so only the compound word will change its order i.e. አለምአቀፍ ቀን የሴቶች similar structure with Tigrigna noun phrase i.e. T: ዓለምለኻ/ADJ መዓልቲ ደቂ-ኣንስትዮ/CN /me`alti deqi^anestyo.

Rule 6: Local Reordering Rule for Noun and Adjective Word Combination of Amharic Noun Phrase.

In this section, reordering rule for the valid possible combinations of nouns and adjectives are discussed. There are some combination of adjectives and nouns in Amharic noun phrase that have different word order from Tigrigna. The following underlined Amharic and Tigrigna noun phrase are combination of noun and adjectives:

A: የኢትዮጵያ/NP ዋና/ADJ ከተማ/N (ye^ityoPya wana ketema/Capital city of Ethiopia)

T: ዋና/ADJ ከተማ/N ኢትዮጵያ/N (wana ketema ^ityoPya)

A: የጉምሩክ/NP ዋና/ADJ ዳይሬክቶር/N (yegumruk wana dayrEktor/General Director of custom authority)

T: ዋና/ADJ ዳይሬክቶር/N ጉምሩክ/N (wana dayrEktor gumruk)

As it is shown above Amharic phrases and their translation in Tigrigna, Reordering rule for the above examples combination of NP, ADJ, and N is just move the NP (የኢትዮጵያ/ye^ityoPya) to the right of ADJ, N combination which will looks like as ዋና ከተማ የኢትዮጵያ/wana ketema ye^ityoPya similar structure with Tigrigna language. After applying the reordering rule we can get the following reordered Amharic phrase:

rA: ዋና ከተማ የኢትዮጵያ (wana ketema ye^ityoPya)

T: ዋና ከተማ ኢትዮጵያ (wana ketema ^ityoPya)

rA: ዋና ዳይሬክቶር የጉምሩክ (wana dayrEktor yegumruk)

T: ዋና ዳይሬክቶር ጉምሩክ (wana dayrEktor gumruk)

Reorderig for the above is done by using Algorithm 4.7

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1,2,3,….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3,….k, where k is the number of words in si
      If POS (Wj) is 'NP' and POS (Wj+1) is 'ADJ' and POS ('Wj+2') is 'N' or 'NC'
            temp =Wj+2
            temp1=Wj+1
            Wj+2 =Wj
            Wj+1 =temp
            Wj=temp1
         End if
       End for
    End for
```

Algorithm 4.7: Reordering Algorithm for Amharic noun phrase combination of N and ADJ

Rule 7: Local Reordering Rule for Compound Word (CN) and Compound Word (CN) Word Combination of Amharic Noun Phrase

Amharic noun phrases composed of more than two compound words labeled as CN particularly two compound words in this section has different order in Tigrigna. The compound words can be attached with preposition and conjunction as well. In this case, when such kind of sequence appear in Amharic noun phrase, in Tigrigna the order of the compound words is appeared in reverse order. For example, the following underlined phrases of both languages are compound words:

A: <u>የጤና ጥበቃ ጽሕፈት ቤት</u> (yeTEna Tbeqa SHfet bEt/the office ministry of health)

T: <u>ቤት ፅሕፈት ሓለዋ ጥዕና</u>/bEt ^SHfet Halewa T`na)

As it is shown in the above Amharic phrases and their translation in Tigrigna, Amharic phrase composed of two compound words has different structural order in Tigrigna. Unlike Amharic, in Tigrigna the compound words will swap their order. Thus, to solve such kind of problems we have provided reordering rule so that it will have similar structure as Tigrigna language. After applying the reordering rule we can get the following Amharic phrases:

rA: <u>ቤት ጽሕፈት/CN ጥበቃ የጤና/CN</u> (bEt SHfet Tbeqa yeTEna)

T: <u>ቤት ፅሕፈት/CN ሓለዋ ጥዕና/CN</u> (bEt ^SHfet Halewa T`na)

Reordering task for the above is done using Algorithm 4.8

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1, 2, 3….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3….k, where k is the number of words in si
      If POS (Wj) is 'CN' and POS (Wj+1) is 'CN'
          Swap (Wj, Wj+1)
       End if
    End for
End for
```

Algorithm 4.8: Reordering Algorithm for Amharic CN and CN word combination.

Rule 8: Local Reordering Rule for Compound Word (N2) and Compound Word (N2) Combination of Amharic Noun Phrase

Amharic noun phrases composed of more than two compound words labeled as N2 particularly two compound words in this section has different order in Tigrigna. The compound words can be attached with preposition and conjunction as well. In this case, when such kind of sequence appear in Amharic noun phrase, in Tigrigna the order of the compound words is swapped each other. For example, the following underlined phrases of both languages are compound words:

A: <u>አፄ ዮሀንስ/N2</u> <u>ቤተ መንግስት/N2 i.e. (N2, N2)</u> ለንብኚዎች ክፍት ሆነ ። (^e^SE yohens bEte mengst i.e. (N2, N2) legobNiwoc kft hone/the palace of king Yohanes is opened for visitors.)

T: <u>ቤተ መንግስቲ/N2</u> <u>ሃፀይ ዮሃንስ/N2</u> ንንብነይቲ ክፍቲ ኮይኑ ። (bEte mengsti ha^Sey yohans ngobneyti kfti koynu.)

As it can be clearly seen from these examples, an Amharic phrase combination of (N2, N2) has different word order comparing to Tigrigna phrase. So, to avoid such kind of problems it is provided reordering rule that swaps their order. After applying the rule we get the following reordered Amharic phrase along with Tigrigna:

rA: <u>ቤተ መንግስት</u> <u>አፄ ዮሀንስ</u>

T: <u>ቤተ መንግስቲ</u> <u>ሃፀይ ዮሃንስ</u>

Reordering this task is done using algorithm 4.9;

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1,2,3,….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3,….k, where k is the number of words in si
      If POS (‘Wj’) is ‘N2’ and POS (‘Wj+1’) is ‘N2’
               Swap (Wj, Wj+1)
       End if
    End for
End for
```

Algorithm 4.9: Reordering Algorithm for Amharic noun phrase combination of N2 and N2

Rule 9: Local Reordering Rule for Compound Word (N2) and Compound Word (CN) Combination of Amharic Noun Phrase

N2 is different from CN and it is described earlier. In this reordering rule, the possible combination of compound words labeled as N2 and CN as well as their order is discussed. There are Amharic noun phrase combination of N2 and CN that have different structure in Tigrigna noun phrase. The rule also works for N2 and CN attached with preposition and conjunction. The following underlined phrases are examples of Amharic and Tigrigna languages:

A: <u>የአዲስ አበባ/N2P ምክር ቤት/CN</u> (ye^edis ^ebeba mkr bEt/ Council of Addis Ababa city)

T: <u>ቤት ምኽሪ/CN ኣዲስ ኣበባ/N2</u> (bEt mKri ^adis ^Abeba)

As it can be clearly seen from these examples, an Amharic phrase combination of (N2, CN) has different word order comparing to Tigrigna phrase. So, to avoid such kind of problems it is provided reordering rule that swap the order of words into the form of CN, N2. After applying the rule we can get the following reordered Amharic phrase along with Tigrigna:

rA: <u>ቤት ምክር የአዲስ አበባ</u> (bEt mkr ye^edis ^ebeba (A)

T: <u>ቤት ምኽሪ ኣዲስ ኣበባ</u> (bEt mKri ^adis ^Abeba)

Reordering the above task is done by using Algorithm 4.10 given below.

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1, 2, 3….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3….k, where k is the number of words in si
        If POS (Pj) is 'N2' or 'N2P' and POS (Pj+1) is 'CNP' or 'CN' or 'CNC' or 'CNPC'
                Swap (Pj, Pj+1)
         End if
     End for
  End for
```

Algorithm 4.10: Reordering Algorithm for Amharic noun phrase combination of N2 and CN

Rule 10: Local Reordering Rule for Number and Noun Combination of Amharic Noun Phrase

Noun phrase in both language can be constructed from collection of words belonging to noun and numeric classes. In this reordering, the combination of number labeled as NUM in general and noun N are identified and reordering rule to map the Amharic noun phrase to Tigrigna noun phrase is provided. The rule also works for valid Amharic phrases and sentences taged as NUM and N combination attached with preposition and conjunction.

The following underlined examples of both languages are combination of NUOR and N:

A: 3ኛ/NUMCR ዓመት/N ተማሪዎች/N (3Na `amet temariwoc/Third year students)

T: ተማሃሮ/N 3ይ/NUMCR ዓመት/N (temaharo 3y `amet)

As it is shown above Amharic phrases and their translation in Tigrigna, the single underlined word (N) in Amharic phrase is shown at the end of the phrase whereas in Tigrigna it is shown at the beginning of the phrase. Therefore, to solve such kind of problems we have provided reordering rule so that it will be similar structure with Tigrigna language. After applying the reordering rule we can get the following reordered Amharic phrase:

rA: ተማሪዎች 3ኛ ዓመት (temariwoc 3Na `amet)

T: ተማሃሮ 3ይ ዓመት (temaharo 3y `amet)

The above reordering task is done using Algorithm 4.11:

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1, 2, 3,….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3….k, where k is the number of words in si
      If POS ('Wj') is 'NUMOR' or 'NUMPC or NUM' and POS (Wj+1) is 'N 'and POS
(Wj+2) is 'N'
                  temp =Wj+2
                  temp1= Wj+1
                  Wj+2 =temp1
                  Wj+1 =Wj
                  Wj =temp
         End if
    End for
End for
```

Algorithm 4.11: Reordering Algorithm for Amharic noun phrase combination of NUM and N

Rule 11: Local Reordering Rule for Compond Word (N2) and Noun Combination of Amharic Noun Phrase

It is been described compound word labeled as N2 are combination of words that never change the order of words in both language pairs. Such kind of compound words are most of the time used as name of places and organization in both languages. Based on the description, in this reordering rule, it is provided a rule for Amharic noun phrase with the combination of N2 (it can be attached to preposition and conjunction) and noun that requires reordering to have similar structure as Tigrigna. The following underlined examples are Amharic and Tigrigna phrases with combination of N and N2:

A: ምድር ባቡር/N2 ድርጅት/N (mdr babur drjt/Railway Organization)

T: ትካል/N ምድሪ ባቡር/N2 (tkal mdri babur)

A: የኢትዮጵያ/NP ሕገ መንግስት/N2 (ye^ityoPya Hge mengst/Ethiopian Constitution)

T: ሕገ መንግስቲ/N2 ኢትዮጵያ/N (Hge mengsti ^ityoPya)

A: የኢትዮጵያ/NP ምድር ባቡር/N2 ድርጅት/N (ye^ityoPya mdr babur drjt/Ethiopian railway organization)

T: ትካል/N ምድሪ ባቡር/N2 ኢትዮጵያ/N (tkal mdri babur ^ityoPya)

As it is shown in the above Amharic phrases and their translation in Tigrigna, the combination of compound words (N2) and nouns in Amharic and Tigrigna has different order. In the above example, there are three valid possible combination. These are, the Amharic phrase contain N2 and N and its translation of N and N2, the Amharic phrase contain NP and N2 and its transition of N2 and NP and the Amharic phrase contain NP, N2 and N, and its translation of N, N2 and NP. Thus, to avoid this kind of problems we have reordered the words in Amharic phrases so that they can have similar structure with Tigrigna. After applying the reordering rule to the above examples we can get the following reordered Amharic phrases:

rA: ድርጅት ምድር ባቡር (drjt mdr babur)

T: ትካል ምድሪ ባቡር (tkal mdri babur)

rA: ሕገ መንግስት የኢትዮጵያ (Hge mengst ye^ityoPya)

T: ሕገ መንግስቲ ኢትዮጵያ (Hge mengsti ^ityoPya)

rA: ድርጅት ምድር ባቡር የኢትዮጵያ (drjt mdr babur ye^ityoPya)

T: ትካል ምድሪ ባቡር ኢትዮጵያ (tkal mdri babur ^ityoPya)

Reordering for the above task is done using by Algorithm 4.12 below.

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1,2,3,….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3,….k, where k is the number of words in si
      If POS (Wj) is 'N2' or 'N2P' or 'N2PC'and POS (Wj+1) is 'N'
            Swap (Wj, Wj+1)
        End if
        Else if POS (Wj) is 'NP' and POS (Wj+1) is 'N2' or 'N2P' or 'N2C' or 'N2PC'
            Swap (Wj, Wj+1)
        End else if
        Else if POS(Wj) is 'NP' and POS('Wj+1') is 'N2' or 'N2P' or 'N2C' or 'N2PC' and
POS('Wj+2') is 'N' or 'NC'
            Swap (Wj, Wj+2)
         End else if
      End for
  End for
```

Algorithm 4.12: Reordering Algorithm for Amharic noun phrase combination of N2 and N.

Rule 12: Local Reordering Rule for Noun and Compound Word (CN) Combination of Amharic Noun Phrase

In this section, reordering rule for Amharic noun phrase composed of noun N and compound words CN (it can be attached with preposition CNP or conjunction CNC or both CNPC) are described. It means that, reordering rule is developed only for the possible valid combination that have different word order comparing to the target (Tigrigna) noun phrase. For example, in the following phrases, the underlined words are compound words and nouns:

A: ጤና ጥበቃ/CN ሚኒስቴር/N (TEna Tbeqa minister/ Ministry of health)

T: ሚኒስትሪ/N ሓለዋ ጥዕና/CN (ministri Halewa T`na)

A: የክልል/NP ምክር ቤት/CN (yekll mkr bEt /Regional Council)

T: ቤት ምኽሪ/CN ክልል/N (bEt mKri kll)

A: የሚኒስትሮች/NP ምክር ቤት/CN ሰብሳቢ/N (yeministroc mkr bEt sebsabi/Chairman of the Council of Ministers)

T: ኣካቢ/N ቤት ምኽሪ/CN ሚኒስትራት/N (^akabi bEt mKri ministrant)

As it is shown above Amharic phrases and their translation in Tigrigna, the combination of compound words and nouns in Amharic and Tigrigna has different order. In the above example, there are three valid possible combination. These are, the Amharic phrase contain CN, N and its translation of N, CN, the Amharic phrase contain NP, CN and its transition of CN, NP and the Amharic phrase contain NP, CN, N, and its translation of N, CN, NP. Thus, to avoid this kind of problems we have reordered the words in Amharic phrases so that they can have similar structure with Tigrigna. After applying the reordering rule to the above examples we can get the following reordered Amharic phrases.

rA: ሚኒስቴር ጥበቃ ጤና (ministEr Tbeqa TEna)

T: ሚኒስትሪ ሓለዋ ጥዕና (ministri Halewa T`na)

rA: ቤት ምክር የክልል (bEt mkr yekll)

T: ቤት ምኽሪ ክልል (bEt mKri kll)

rA: ሰብሳቢ ቤት ምክር የሚኒስትሮች (sebsabi bEt mkr yeministroc)

T: ኣካቢ ቤት ምኽሪ ሚኒስትራት (^akabi bEt mKri ministrant)

The above reordering task is done using Algorithm 4.13below.

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1, 2, 3,….n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3,…k, where k is the number of words in si
        If POS (Wj) is 'NP' and POS (Wj+1) is 'CN' and POS (Wj+2) is 'N' or 'NC'
              Swap (Wj, Wj+2)
         End if
        Else if POS (Wj) is 'CN' and POS (Wj+1) is 'N' or 'NC'
              Swap (Wj, Wj+1)
        End else if
        Else if POS (Wj) is 'NP' and POS (Wj+1) is 'CN'
              Swap (Wj, Wj+1)
         End else if
      End for
 End for
```

Algorithm 4.13: Reordering Algorithm for Amharic noun phrase combination of N and CN

The above Algorithm also works for compound word (CN) attached with preposition, conjunction and both at the same time.

Rule 13: Local Reordering Rule for Compound Word (CN) Amharic Noun Phrase

Compound word (CN) is a combination of more than two words that can be treated as single word in a sentence. In this research combination of words labeled as compound words are only words in Amharic language that have different syntactical word order comparing to Tigrigna language. Therefore, the reordering algorithm works on identifyand reordering every compund noun phrase specified as CN. It also works for compound word attached with preposition (CNP) or conjunction (CNC) or both at a time (CNPC). The following phrases, the underlined words are compound words:

A: <u>ትምህርት ቤት/CN</u> (tmhrt bEt/ School)

T: <u>ቤት ትምህርቲ/CN</u> (bEt tmhrti)

As it is shown in the example, Amharic phrases and their translation in Tigrigna, the compound words in Amharic and Tigrigna has different word order. Thus, to avoid such kind of problems the words in Amharic phrases are reordered so that they can have similar structure with Tigrigna. After

applying the reordering rule to the above examples we can get the following reordered Amharic compound word phrase:

rA: ቤት ትምህርት (bEt tmhrt)

T: ቤት ትምህርቲ (bEt tmhrti)

Reordering for the above Amharic compound word is done by Algorithm 4.14

```
Load Amharic sentences S from the POS Tagged corpus
For each sentence si ε S, i=1, 2, 3,…n, where n is the number of sentences in S
   Extract word W in si //segmentation process based on tag set
   For each word Wj ε si, j=1, 2, 3,…k, where k is the number of words in si
      If POS (Wj) is 'CN' or 'CNP' or 'CNC' or 'CNPC'//then apply local reordering
            Swap (Wj, Wj-1)//Wj is the word with the tag set CN
       End if
    End for
End for
```

Algorithm 4.14: Reordering Algorithm for Amharic compound word noun phrase

Note: Combination of reordering rules for different Amharic noun phrase can be made. For example, the following underlined phrase are combinations of different word classes reordered by the combination of different rules:

- A: የኢትዮጵያ ቴሌቪዥን/CNP ድርጅት/N ዋና/ADJ ዳይሬክተር/N (ye^ityoPya tElEvijn drjt wana dayrEkter/General director of Ethiopian Television) is Amharic noun phrase and its order is CN, N, ADJ, N respectively. For such kind of noun phrase we can combine two reordering rule. For የኢትዮጵያ ቴሌቪጅን ድርጅት/Ethiopian Television Organization, noun phrase we have provided reordering algorithm at algorithm 4.13 (CN, N) and assigned to 'str'. So we will have 'str', A, N and apply algorithm 4.6 to have similar structure as Tigrigna noun phrase; ዋና ዳይሬክተር ድርጅት የኢትዮጵያ ቴሌቪጅን and finally apply algorithm 4.14 for words in compound words (CN). After reordering, we can get the following reordered Amharic noun phase:
  - rA: ዋና ዳይሬክተር ድርጅት ቴሌቪጅን የኢትዮጵያ (wana dayrEkter drjt tElEvijn ye^ityoPya)
  - T: ዋና ዳይሬክተር ትካል ቴሌቪጅን ኢትዮጵያ (wana dayrEkter tkal tElEvijn ^ityoPya)
- A: የጤና ጥበቃና/CNPC ሀገር መከላከያ/CN ፅሕፈት ቤት/CN (yeTEna Tbeqana heger mekelakeya ^SHfet bEt/ Office of health and national defense) this Amharic noun phrase is constructed from three compound word i.e. የጤና ጥበቃና (CNPC) ሀገር መከላከያ (CN) and ፅሕፈት ቤት (CN)

which is CNPC, CN, CN order. First, apply algorithm 4.8 (CNPC, CN) and assigned to variable 'str'; then, swap the order of 'str' and CN and call algorithm 4.14. Finally, the Amharic phrase will be shown as ቤት ፅሕፈት መከላከያ ሀገር ጥበቃና የጤና (rA) (bEt ^SHfet mekelakeya heger Tbeqana yeTEna) similar to the structure of Tigrigna noun phrase: ቤት ፅሕፈት ምክልኻል ሃገርን ሓለዋ ጥዕናን (bEt ^SHfet mklKal hagern Halewa T`nan)

- A: <u>የአዲስ አበባ /N2P ከተማ አስተዳደር/CN ጽሕፈት ቤት/CN ባወጣው መግለጫ፡በዚህ ዓመተ ለ10ሺ ወጣቶች ስራ ዕድል እፈጥራለሁ</u> ብለዋል። (ye^edis ^ebeba ketema ^estedader SHfet bEt baweTaw megleCa bezih `amete lexi weTatoc sra `dl ^feTralehu blewal/Addis Ababa city administration office has released a statement of job opportunity for 10,000 young people's.), for this reordering, first apply algorithm 4.10 (N2, CN) and assign to 'str' next swap order of 'str' and CN finally, its structure will be shown as ጽሕፈት ቤት(CN) ከተማ አስተዳደር(CN) የአዲስ አበባ(N2P) and call algorithm 4.14 to have similar structure with the target language:
  - rA: ቤት ጽሕፈት አስተዳደር ከተማ የአዲስ አበባ (bEt SHfet ^estedader ketema ye^edis ^ebeba),
  - T: ቤት ፅሕፈት ኣማሓዳሪ ከተማ ኣዲስ ኣበባ (bEt ^SHfet ^amaHadari ketema ^adis ^abeba)
- A: <u>የሰሜን ጎንደር/N2P ዞን/N ዋና/ADJ አስተዳደር/N</u> እንደገለፀት፤በዞኑ <u>የሆስፒታል እጥረት</u> እንዳለና, ህዝብ ለወጪ እየተዳረገ እንደሆነ አብራርተዋል (yesemEn gonder zon wana ^estedader ^ndegele^Sut; bezonu yehospital ^Tret ^ndalena, hzb leweCi ^yetedarege ^ndehone ^ebrartewal/The north Gondar administration has announced the shortage of the hospital that leads the people into extra expense). The underlined phrases makes difference between the two languages; for such kind of phrase the sequence is N2P (N2 attached to preposition), N, A, N. we have provided rule for N2, N combination at algorithm 4.12 and assigned the result to 'str'; Then apply algorithm 4.6 for 'str', A, N combination and will be shown structurally similar with Tigrigna:
  - rA: <u>ዋና አስተዳደር ዞን የሰሜን ጎንደር</u> (wana ^estedader zon yesemEn gonder)
  - T: <u>ዋና ኣማሓዳሪ ዞባ ሰሜን ጎንደር</u> (wana ^amaHadari zoba semEn gonder)
- A: <u>የከተማዋ ኮሙዩኒኬሽን/CNP ጉዳዮች/N ጽህፈት ቤት/CN ኃላፊ.N</u> አቶ አብይ ኃይሉ እንደገለጹት፥ <u>ከንቲባዎቹ የተሾሙት ያላቸውን የትምህርት ዝግጅት, የሥራ ልምድና</u> ሕዝባዊ ወገንተኝነታቸውን መሰረት በማድረግ ነው። (yeketemawa komuyunikExn gudayoc Shfet bEt lafi ^eto ^eby ylu ^ndegeleSut: kentibawocu yetexomut yalacewn yetmhrt zgjt, ye^sra lmdna Hzbawi wegenteNnetacewn meseret bemadreg new.) The phrases denoted with underline Amharic noun phrase of words makes difference. The first phrase which consists six words i.e. <u>የከተማዋ ኮሙዩኒኬሽን ጉዳዮች ጽህፈት ቤት ኃላፊ,</u> can be break down into two small phrase; <u>የከተማዋ ኮሙዩኒኬሽን ጉዳዮች (CNP, N)</u> apply

algorithm 4.13 and assigned to ‘str1’ and similarly, the same algorithm is applied for the second phrase: ጽህፈት ቤት ኃላፊ, (CN, N), and assign to ‘str2’. Finally swap order of ‘str1, str2’ and apply algorithm 4.14 for CN word order which will be shown similar with the structure of Tigrigna language. After reordering the phrases, we can get the following result:

- rA: ኃላፊ ቤት ጽህፈት ጉዳዮች ኮሙዩኒኬሽን የከተማዋ (halafi_bEt Shfet gudayoc komuyunikExn yeketemawa/The head of communication office, Mr. Abiy Hailu has said, mayors has appointed as per their qualification, experience, and their community services).
- T: ሓላፊ ቤት ፅሕፈት ጉዳያት ኮሙዩኒኬሽን እታ ከተማ (Halafi bEt ^SHfet gudayat komuyunikExn ^ta ketema)

➢ A: ጠቅላይ ሚኒስትሩ የሀገሪቱ/NP ርዕሰ መስተዳድር/N2, የሚኒስትሮች/NP ምክር ቤት/CN ሰብሳቢና/NC የጦር ኃይሎች/CNP ጠቅላይ/ADJ አዛዥ/N ነው/AXU።(Teqlay ministru yehegeritu r`se mestedadr, yeministroc mkr bEt sebsabina yeTor yloc Teqlay ^ezaZ new/Prime minister is the governor of the country, chairman of the cabinet and commander-in-chief of the armed forces.) The deference happen at the underlined phrases. In this Amharic sentence there are three phrases which have to be reordered; first, for የሀገሪቱ ርዕሰ መስተዳድር (NP, N2) we can apply algorithm 4.12, second for የሚኒስትሮች ምክር ቤት ሰብሳቢና (NP, CN, NC) apply algorithm 4.13 and the last is የጦር ኃይሎች ጠቅላይ አዛዥ (CNP, A, N) we can apply algorithm 4.6 and finally algorithm 4.14 is applied for word order in the compound word (CN) to have similar syntactical structure as the target language. After applying the rule we can get the following reordered phrases:

- rA: ርዕሰ መስተዳድር የሀገሪቱ , ሰብሳቢና ቤት ምክር የሚኒስትሮች ጠቅላይ አዛዥ ኃይሎች የጦር (r`se mestedadr yehegeritu, sebsabina bEt mkr yeministroc Teqlay ^ezaZ Hayloc yeTor)
- T: ርእሰ ምምሕዳር እዛ ሃገር , ኣካቢ ቤት ምኽሪ ሚኒስትራትን ዋና ኣዛዚ ሓይሊታት ኩናት (r^se mmHdar ^za hager , ^akabi bEt mKri ministratn wana ^azazi Haylitat kWinat)

➢ A: ጠቅላይ ሚኒስትሩና የሚኒስትሮች/NP ምክር ቤት/CN ለሕዝብ ተወካዮች/CNP ምክር ቤት/CN ተጠሪዎች/N ናቸው። (Teqlay ministruna yeministroc mkr bEt leHzb tewekayoc mkr bEt teTeriwoc nacew/Prime minister and council of ministers are accountable to the House of peoples' representatives). The difference is raised at the underlined phrases. For የሚኒስትሮች ምክር ቤት (NP, CN) we can apply algorithm 4.13, for ለሕዝብ ተወካዮች ምክር ቤት ተጠሪዎች (CNP, CN, N) apply algorithm 4.8 and finally algorithm 4.14 is applied for CN word order to be shown in similar structure as Tigrigna language. After reordering, we can get the following reordered Amharic phrase:

- rA: <u>ቤት ምክር የሚኒስትሮች ተጠሪዎች ቤት ምክር ተወካዮች ለሕዝብ</u> (bEt mkr yeministroc teTeriwoc bEt mkr tewekayoc leHzb)
- T: <u>ቤት ምኽሪ ሚኒስተራትን ተፀዋዕነቶም ንቤት ምኽሪ ተወከልቲ ህዝቢ</u> (bEt mKri ministeratn te^Sewa`netom nbEt mKri tewekelti hzbi)

Reordering Simple and Complex Amharic Sentences

By definition, simple sentences are sentences composed of simple noun and verb phrases with single subject and verbs whereas complex sentences are sentences build from complex noun and verb phrases with more than one subjects and verbs in general. Simple noun phrase might be constructed from noun, modifiers and complements which describes single noun and similarly verb phrase are phrase which can be constructed from complements and modifiers describes a single action. However, in complex noun and verb phrase it can contains small phrases and large phrase which can describes more than one subjects and verbs. In this section, no special reordering rule is prepared for simple and complex sentences, because all the above reordering rules can work for simple and complex sentences. Table 4.9 shows examples of Amharic and its translation in Tigrigna which can be classified as simple and complex:

Table 4.9: Amharic and Tigrigna simple and complex sentence

Amharic sentence	Tigrigna sentence	Remarks	
		Type of sentence	Reordering rule
ካሳ ምሳ እየበላ አስቴር መጣች(kasa msa ^yebela ^estEr meTac/ Aster came when Kasa eating his lunch)	ካሰ ምሳሕ እንዳበለዐ አስቴር መፂአ/kasa msaH ^ndabele`e ^estEr me^Si^a	Complex sentences	No
ያቺ ረጅምዋ ቀይ ሴትዮ መጥታ ነበር(yaci rejmwa qey sEtyo meTta neber/ The long woman was here	እታ ነዋሕ ቀያሕ ሰበይቲ መፂአ ነይራ/^ta newaH qeyaH sebeyti me^Si^a neyra	Simple sentences	No
እሱ ወደ ትምህርት ቤት ሄደዋል(^su wede tmhrt bEt hEdewal/ He went to school)	ንሱ ናብ ቤት ትምህርቲ ከይዱ/ nsu nab bEt tmhrti keydu	Simple sentence	Yes(algorithm 4.3)
የጤና ጥበቃና ትምህርት ሚኒስቴር ሰራተኞች የባንዴራ ቀን በጋራ አከበሩ (TEna Tbeqana yetmhrt ministEr serateNoc yebandEra qen begara ^ekeberu; bemeCerexam sle hegeracew guday wyyt ^ekaydewal.	ሰራሕተኛታት ምኒስቲሪ ትምህርትን ሓለዋ ጥዕናን መዓልቲ ባንዴራ በሓባር ኣኽቢሮም seraHteNatat mnistiri tmhrtn Halewa T`nan me`alti bandEra beHabar ^aKbirom	Complex sentence	Yes (uses algorithm 5 and 4.14)

As it can be shown in Table 4.9, Amharic and its translation in Tigrigna, the last column describes the types of sentences. It also shows whether the sentences require reordering rule or not. So a sentence either complex or simple in Amharic language might or might not require reordering to have similar structure with Tigrigna language. On the other hand, the sentences that in need of reordering rules are not out of the rules that we have used earlier.

4.3.5 Translation Model

In this study, the input of the translation model is a parallel corpus containing Tigrigna and locally reordered Amharic (rA) texts. In general, translations require many to many alignments between words. This means that a group of words in the source language should be translated by a group of words in the target language, and there might not be a word level correspondence between these groups. In phrase based models, the smallest unit used for translation is called a phrase. By translating phrases instead of words, the problem of ambiguity is partially solved, but this cannot be totally avoided. As mentioned above, for a given locally reordered source and target sentences ***rA*** and ***T***, it is the way sentences in r***A*** get converted to sentences in ***T*** which is denoted by *P (rA|T)*.

4.4 Translation Phase

4.4.1 Amharic Input Text

The user input texts in this translation system is Amharic text since the intended translation system is from Amharic-to-Tigrigna. Once the user enters an Amharic text to the system, the texts are tagged by Amharic part-of-speech tagger and local reordering is applied using the identified rules to have similar text structure as the trained model. So the Amharic input texts are changed to the form of locally reordered texts i.e. ***rA***. Therefore, the translation system/decoder accept locally reordered Amharic input text, performs translation process and finally generate better result to end user.

4.4.2 The Decoder

The job of the decoder (mosses) is to find the highest scoring sentence in the target language (according to the translation model) corresponding to a given source sentence. Decoder searches for the best sequence of transformations that translates source sentence to responding target sentence. It is also possible for the decoder to output a ranked list of the translation candidates, and

also to supply various types of information about how it came to its decision (for instance the phrase-phrase correspondences that it used). It looks up all translation of every source word, phrase using word or phrase table and recombines the target language phrases that maximize the translation model probability by the language model. In this research work, hybrid of rule based and statistically based approach is used, so that the model is trained with locally reordered Amharic text and Tigrigna texts. Therefore, the decoder takes locally reordered Amharic texts as an input and displaying Tigrigna as an output.

$\overline{T} = \underset{t}{\operatorname{argmax}} P(rA|T) * P(T)$

$P(rA|T)$ Amharic-to-Tigrigna translation model

$P(T)$ Tigrigna language model

4.4.3 Tigrigna Output Text

In this translation system the expected output is Tigrigna text. The system accepts Amharic text as an input and generate Tigrigna text as a final output.

Chapter 5: Experiment

5.1 Introduction

Based on the design, Amharic-to-Tigrigna machine translation is experimented using a combination of linguistic rule for the purpose of Amharic corpus preprocessing and statistical approaches in a way that the addition of linguistic knowledge into statistical approaches improves the translation. Under this Chapter, the conducted experiments and discussion is presented in a way that states the evaluation and performance of two experiments. The first experiment focuses on statistical approach while the second experiment uses the hybrid (rule followed by statistical) approach.

5.2 Corpus Collection

Hybrid approach which is the combination of statistical based approach and rule based approach requires the availability of bilingual parallel corpus. In this research work, the bilingual parallel corpus of both Amharic and Tigrigna languages that are publically available and previously revisit sentence have used. These parallel corpus includes some chapters of the Holy Bible, the Constitution of FDRE, and other simple sentences. Most of the collected data's are in a txt file format and the others are collected from webs.

5.2.1 Corpus Preparation

After the corpus have collected and pre-processed with rules, it was prepared in a format that is suitable for the translation purpose. Thus, the following two procedures have been applied on the collected preprocessed corpus by linguistic rules to make it ready for training and testing of the translation system at the statistical machine translation part with the help of commands once the necessary machine translation tools and software's are correctly installed and integrated. Therefore, no need of writing scripts are required.

Tokenization: This is a procedure that inserts a space between words and punctuation.

Cleaning; Cleaning is the process of removing the long sentences, empty sentences and misaligned sentences. Long sentences, empty sentences and misaligned sentences can cause problems with the training pipeline. Cleaning process perform only to cut down the length of the sentence and remove the unusual space between words and between sentence also.

In this experimentation, the study used a total of 1582 Amharic-Tigrigna parallel sentences for training and testing the proposed solution and from the total 1582 parallel sentences, 1482 parallel sentences are used for training whereas the rest are used for testing of the system. The classification is based on literatures that, ten percent of the total training data are used for testing purpose. Since the system is unidirectional, the training process is performed from Amharic to Tigrigna. As it has been mentioned under Chapter One, showing the advantage of using hybrid approach rather than using statistical approach in machine translation for the language pair is one of the specific objectives of the research work. Therefore, two major experiments are conducted by using similar parallel sentences using two different approaches. The first experiment is conducted by using a statistical approach and the second experiment is conducted by using a hybrid approach. Similar tools are used for training and testing in both of the experiments. Both of the experiments was discussed briefly in the following sections as Experiment I and Experiment II.

5.3 Experiment I: Statistical Approach

This is the first experiment conducted on Amharic-to-Tigrigna language translation by using a statistical approach. As it was mentioned in Chapter Two, statistical machine translation is an approach which tries to generate translations using statistical methods based on bilingual text corpora.

5.3.1 Training the System

We have used the data set described in Section 5.2 to perform the training and testing procedures. From a total of 1582 Amharic-Tigrigna parallel sentences 1482 sentences were used to train the system. Moses which is freely available software is used to train the system from Amharic-to-Tigrigna machine translation. The training process includes the following procedures.

Language Model Training

The language model is used to ensure fluency of the output, so it is built with the target language, that is, for Tigrigna since the translation is unidirectional from Amharic to Tigrigna. As it has been stated, IRSTLM toolkit is used to perform language modeling task. An appropriate tri-gram language model is built.

Training the Translation System

At this step, word-alignment, phrase extraction and scoring were used and lexicalized reordering tables and Moses configuration file is created. Mainly this step creates a moses.ini file, which is used for decoding. The phrase table was also created which basically contains the probabilities of a word following words to the given word. GIZA++ toolkit is used to train the translation system.

Tuning

As it is mentioned in the previous section, while training the translation system, a 'moses.ini' file was produced which is used for decoding. The querying process could be started right away but the weights used by Moses to weight the different models against each other are not optimized. Therefore, to find better weights, the translation system needs to be tuned. This process produces another ".ini" file that is used for decoding. After training the system, testing is conducted to test the trained translator for the sake of measuring the performance of the trained system.

5.3.2 Result of Test Set on Experiment I

We have used 100 Amharic and Tigrigna parallel sentences in order to test the performance of the system in terms of translation accuracy to translate a single Amharic sentence to Tigrigna sentence. To do so, BLEU score methodology which was discussed in Chapter Two is used in order to see the result of the translation process. The result recorded from the BLEU score methodology shows 7.02% of the translation process are correctly translated an Amharic texts to Tigrigna texts. The result is as shown in the following figure.

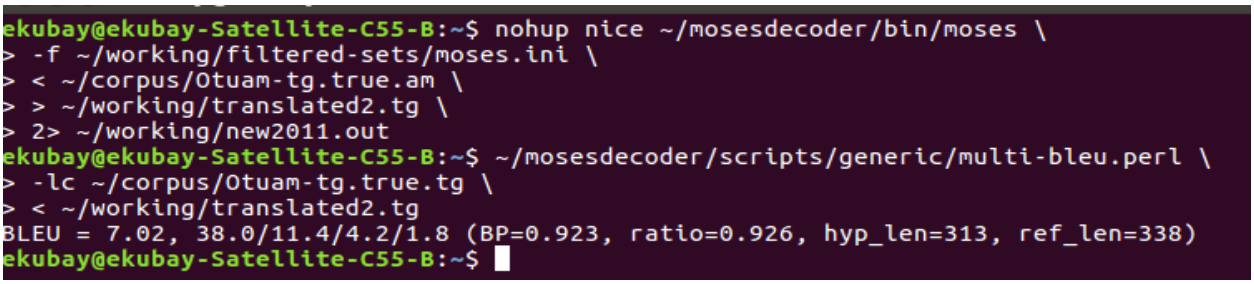

Figure 5.4: Experimental result for statistical approach

5.4 Experiment II : Hybrid Approach

The second experiment that conducted on Amharic-to-Tigrigna language pair is done by using a hybrid approach which is the main objective of the proposed research work. As it is discussed in Chapter One the main objective of this research work is to develop Amharic-to-Tigrigna machine

translation using a hybrid approach. First the rules mentioned in Chapter Four are applied on the training and test dataset; then both data sets are prepared for training and testing of the proposed solution. Since the rules are applied before the training and testing step, the procedure of training and testing of the hybrid approach is similar with that of statistical approach stated in the above section. In the translation, first the Amharic reordering rules which are discussed in Section 4.2.2, are applied on Amharic tagged sentences so that the sentences will have similar syntactic structure with that of Tigrigna sentences pair. After the rules are applied on Amharic tagged sentence and become reordered, the POS tagging labels are removed using python script so that it would have similar structure and free from additional texts like its corresponding language sentences. Finally, the statistical approach is applied on the prepared and reordered corpus.

5.4.1 Training the System

We have used the data set described in Section 5.2 to perform the training and testing procedures. From a total of 1582 Amharic-Tigrigna parallel sentences 1482 sentences are used to train the system. Moses which is freely available software is used to train the system from Amharic-to-Tigrigna machine translation. The training process includes the following components.

Language Model Training: The language model is used to ensure fluent output, so it is built with the target language, that is, for Tigrigna since it is only from Amharic-to-Tigrigna translation. As it has been stated, IRSTLM toolkit is used to perform language modeling task. An appropriate tri-gram language model is built.

Training the Translation System: At this step, word-alignment, phrase extraction and scoring are used and lexicalized Reordering tables and Moses configuration file are created. Mainly this step creates a moses.ini file, which is used for decoding and the phrase table is also created which basically contains the probabilities of a word following another. Training the translation system is done by using GIZA++ toolkit that is freely available on the web.

Tuning: As it is mentioned in the above, while training the translation system, a 'moses.ini' file is produced which is used for decoding. The querying process could be started right away but the weights used by Moses to weight the different models against each other are not optimized. Therefore, to find better weights, the translation system needs to be tuned. This process produces another ".ini" file that is used for decoding. After training the system, the next steps are followed by testing process.

5.4.2 Result of Test Set on Experiment II

We have used 100 Amharic and Tigrigna parallel sentences in order to test the performance of the system in terms of translation accuracy to translate a single Amharic sentence to Tigrigna sentence. To do so, BLEU score methodology which is discussed in Chapter Two is used in order to see the result of the translation process. The result recorded from the BLEU score methodology shows 17.47% of the translation process are correctly translated, that was an Amharic texts to Tigrigna texts. The result is shown graphically below:

```
ekubay@ekubay-Satellite-C55-B:~$ nohup nice ~/mosesdecoder/bin/moses \
> -f ~/working/filtered-newtesset/moses.ini \
> < ~/corpus/Rtuam-tg.true.am \
> > ~/working/Ftranslated.tg \
> 2> ~/working/newtest.out
ekubay@ekubay-Satellite-C55-B:~$ ~/mosesdecoder/scripts/generic/multi-bleu.perl \
> -lc ~/corpus/Rtuam-tg.true.tg \
> < ~/working/Ftranslated.tg
BLEU = 17.47, 49.7/26.6/14.0/5.3 (BP=0.985, ratio=0.985, hyp_len=336, ref_len=341)
```

Figure 5.5: Experimental result for hybrid approach

5.5 Discussion

Two experiments are conducted using two different approaches i.e. statistical and hybrid approach. As it is shown in the above Section 5.2.2 and Section 5.3.2, one can observed that, the result recorded from BLEU score shows the hybrid approach is better than the statistical approach for Amharic-to-Tigrigna machine translation. However, the results recorded from both experiments are less, one major reason is size of the corpus. Both experiments are guided by statistical approach and since statistical approach is based on bilingual corpus, as the size of corpus become increase the accuracy also increase and similarly the BLEU score result can also increase. Similarly, result recorded from experiment II is better than experiment I. These shows us reordering structural syntax of Amharic sentence to a form of the targeted sentences have great impact in the translation process. Ambiguity of word translation are also observed from both experiments, for example an Amharic phrase, “የአዲስ አበባ ምክር ቤት”/ye^edis ^ebeba mkr bEt, is translated as “ገዛ ምኽሪ አዲስ ኣበባ”/“geza mkri ^edis ^abeba” in the translation process. However, the word “ቤት”/bEt in Amharic is translated wrongly as “ገዛ”/”geza” in Tigrigna to mean house. As it can be understood the absence of word sense disambiguation have its own negative impact on the translation system. Therefore, word sense disambiguation is recommended as a future work of this research work.

Chapter 6: Conclusion and Recommendation

6.1 Conclusion

The purpose of this study is to develop Amharic-to-Tigrigna machine translation system based on hybrid approach in which rules are used to reorder the structural difference between the languages and finally the translation is made by using a statistical approach. The study began with brief introduction and behavior of the two languages, Amharic and Tigrigna. Phrasal categories as well as the sentence structure of Amharic and Tigrigna are described and the effect of structural differences on the translation process are also studied. Generally, It have also described the general similarity and differences between the languages with examples and an overview of the morphological information, the articles, numbering, punctuation marks and conjunctions that are used in both languages is also identified.

Amharic-to-Tigrigna machine translation design involves collection Amharic and Tigrigna parallel corpus, corpus preparation which also involves dividing the corpus as a training set and test set, POS tagging, implementing the reordering rules for Amharic sentences by using Python programming language, language modeling by using IRSTLM tool, translation modeling by using GIZA++ and training the system by using Moses. Language modeling, translation modeling and decoding are all components of the statistical approach which are freely available on the web and incorporated in the translation system.

The aim of preparing reordering rules are to make the structure of the source language to be more similar to the structure of the target language using their part of speech. We have prepared manually tagged corpus for Amharic language since there are no publically available Amharic POS tagger tool. The linguistic background and nature of the two languages have been also studied in order to design the reordering rules for different types of sentences.

Finally two experiments were conducted by using the collected data set to check the accuracy of the system using two different approaches. The first experiment is conducted by using a statistical approach and it has a BLEU score of 7.02%. The second experiment is carried out by using a hybrid approach and it has a BLEU score of 17.47%. From the test results of the conducted experiments in this research and other related works as well as literatures, it can be concluded that, the hybrid approach is better than the statistical approach.

6.2 Recommendation

As it can be understood the research work, machine translation is an important research area of NLP application and hybrid approach is better than statistical and rule based approach according to literatures and studies made among different approaches of machine translation system. Since hybrid approach takes the better property of the approaches, it can be useful for related languages especially for Ethiopian language pairs.

The following areas could be explored further as a future work of the study:

- Better results can be accomplished by increasing the size of the data set used for training of the system. Increasing the size of the training data set could increase the performance of Amharic-to-Tigrigna machine translation.
- In this research work, only reordering rules to model the structural differences from the targeted language are developed for Amharic language. Therefore, additional results can be determined by further exploring the rules especially by developing morphological rules.
- The system is developed in order to translate Amharic sentences into Tigrigna sentences. Therefore, this could help to develop speech to text and text to speech translations between the language pairs.
- The hybrid approach that we have used in this research work can be applied on other Ethiopian language pairs, most probability for the languages that uses Geez alphabets i.e. Fidel.
- Furthermore, there are no deep studies done on the relationships of Ethiopian languages in general and Amharic-Tigrigna in particular. Therefore, national wise investigation and studies are recommended to develop best translation system among the language pairs.
- Incorporating components like, automatic POS tagger, morphological analyzer and generation and word sense disambiguation can increase the capacity of the translation system.

References

[1] Chopra Abhimanyu, Abhinav Prashar and Chandresh Sain. "Natural language processing." *Int. J. Technol. Enhanc. Emerg. Eng. Res* 1.4 (2013): 131-134.

[2] Natural language processing, www.ncbi.nlm.nih.gov v.18(5); Sep-Oct 2011.

[3] Ann Copestake, "Natural Language Processing", www.cl.cam.ac.uk, 2004,

[4] University of Rochester, https://www.cs.rochester.edu/~schubert/247-447/lecture01-notes.pdf, last visited March 5, 2017.

[5] Okpor MD. "Machine translation approaches: issues and challenges." *International Journal of Computer Science Issues (IJCSI)* 11, no. 5 (2014): 159.

[6] Eleni Teshome "Bidirectional English – Amharic Machine Translation: An Experiment using constrained corpus", MSc thesis, Addis Ababa University, Ethiopia, 2013.

[7] Gebrehiwot Assefa. "*A two-step approach for Tigrigna text categorization"*. Thesis. Addis Abeba University, 2011.

[8] Abeba Ibrahim and Yaregal Assabie. "Amharic Sentence Parsing Using Base Phrase Chunking", In *Proceedings of the 15th International Conference on Intelligent Text Processing and Computational Linguistics*, Springer Lecture Notes in Computer Science Vol. 8403, pp. 297-306, Kathmandu, Nepal.

[9] Ethnologue lanagauge of the world, www.ethnologue.com/country/ET, last visited June 24, 2017

[10] "Ethiopian population, Worldpopulationreview.com, las visited June 24, 2017.

[11] "Constitution of the Federal Democratic Republic of Ethiopia", www.wipo.int/edocs/laws/en/et/et007en.pdf, December 8, 1994

[12] Fisha Haftetsion, "Choosing a Working Language in Multiethnic Nations", www.aigaforum.com/articles/paper-on-Langauge-Policy.pdf, last visited June 24, 2017.

[13] Lewis M. Paul, Gary F. Simons and Charles D. Fennig. *Ethnologue: Languages of the world*. Vol. 16. Dallas, TX: SIL international, 2009.

[14] Russell Bear, "M.Ed. In TESEL Program, language specific informational reports", Rhode Island College.

[15] Milkessa Midega, "Official Language Choice in Ethiopia: Means Of Inclusion Or Exclusion"?, *open access liberary journal*, Dire Dawa University, Ethiopia october 31, 2014

[16] Badr Ibrahim, Rabih Zbib and James Glass. "Segmentation for English-to-Arabic statistical machine translation".*Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*. Association for Computational Linguistics, 2008.

[17] Takahashi S, Wada H, Tadenuma R, and Watanabe S. English-Japanese machine translation. Readings in Machine Translation. 2003:193.

[18] Holger Schwenk, Jean-Babtiste Fouet and Jean Senellart, “First Steps towards a general purpose French/English Statistical Machine Translation System ”, *In: proceedings of the third workshop on statistical machine translation*, pages 119-122, Ohio, 2008.

[19] Preslav Nakov. “Improving English-Spanish Statistical Machine Translation: Experiments Domain Adaptation, Sentence Paraphrasing, Tokenization, and Recasing”, *in: proceedings of the third workshop on statistical machine translation*, pages 147-150, Ohio, 2008.

[20] “Top ten most spoken language in the world”, listverse.com, last visited jun 24,2017

[21] Mulu Gebreegziabher Teshome and Laurent Besacier. "Preliminary experiments on English-Amharic statistical machine translation." In *SLTU*, pp. 36-41. 2012.

[22] Michael Gasser, “Toward a Rule-Based System for English-Amharic Translation”, *In: proceedings of the 8th International Workshop of the ISCA Special Interest Group on Speech and Language Technology for Minority Languages and the 4th workshop on African Language Technology*, 2012

[23] Sisay Adugna, “English-Oromo Machine Translation: An Experiment Using a Statistical Approach”, MSc thesis, Addis Ababa University, Ethiopia, 2009.

[24] Jabesa Daba and Yaregal Assabie. “A Hybrid Approach to the Development of Bidirectional English-Oromiffa Machine Translation”, *In: Proceedings of the 9th International Conference on Natural Language Processing* (PolTAL2014), Springer Lecture Notes in Artificial Intelligence (LNAI), Vol. 8686, pp. 228-235, Warsaw, Poland, 2014.

[25] Philipp Koehn, “MOSES Statistical Machine Translation System User Manual and Code Guide”, University of Edinburgh, October 13, 2016.

[26] Python software foundation, https://www.python.org/, last visited June 25, 2017.

[27] Baye Ymam , “ የአማርኛ ሰዋሰው ”: ሁለተኛ ዕትም Mega printing enterprise, 2002 ዓ/ም

[28] Amare Getahun.: ዘመናዊ የአማርኛ ሰዋስው በቀላል አቀራረብ (Modern Amharic Grammar in a Simple Approach), Addis Ababa, Ethiopia (2010)

[29] Gochel, Daniel. “An integrated approach to automatic complex sentence parsing for Amharic text”. Diss. Addis Ababa University, 2003.

[30] Hailay Beyen. “Design and Development of Tigrigna Search Engine”, MSc thesis, Addis Ababa University, 2013.

[31] Kasa Gebrehiwot. “ሰዋሰው ትግርኛ, ኣዲስ ኣበባ”: Mega printing enterprise, 2004.

[32] Daniel Teklu, “ዘበናዊ ሰዋሰው ቋንቋ ትግርኛ, አዲስ አበባ ”: Mega printing enterprise, 2002 ዓ/ም

[33] SYSTRAN, http://www.systransoft.com/systran/corporate-profile/translation technology/ what-is-machine-translation/, last visited March 7, 2017

[34] Daniel Jurafsky. and James H. Martin, Speech and Language Processing: an introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, PrenticeHall, 2006

[35] Neeha Ashraf and Manzoor Ahmad. “Machine Translation Optimization using Hybrid Architectures”. *International Journal of Applied Information Systems (IJAIS) – ISSN* :

Foundation of Computer Science FCS, New York, USA,Volume 10 – No.9, 2249-0868 May 2016.

[36] Jacob Elming, “Syntactic Reordering In Statistical Machine Translation”, Copenhagen Business School, a PhD thesis, June 2008

[37] Bonnie J. Dorr, Hovy H. Eduard and Levin S. Lori. "Natural Language Processing and Machine Translation Encyclopedia of Language and Linguistics, (ELL2). Machine Translation: Interlingual Methods."

[38] Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer and Paul S. Roossin. "A statistical approach to machine translation." *Computational linguistics* 16, no. 2 (1990): pp, 79-85.

[39] Lopez Adam. "Statistical machine translation" *ACM Computing Surveys (CSUR)* 40.3 (2008): 8.

[40] Collins Michael. "Statistical machine translation: IBM models 1 and 2." *Columbia Univ* (2011).

[41] James Brunning. “Alignment models and algorithms for statistical machine translation”. Diss. University of Cambridge, 2010.

[42] Yamada, Kenji, and Kevin Knight. "A syntax-based statistical translation model." *In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pp. 523-530. Association for Computational Linguistics, 2001.

[43] Kevin Knight and Graehl Jonathan. "An overview of probabilistic tree transducers for natural language processing." In *International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 1-24. Springer Berlin Heidelberg, 2005.

[44] Chiang David, Kevin Knight, and Wei Wang. "11,001 new features for statistical machine translation." *In Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics*, pp. 218-226. Association for Computational Linguistics, 2009.

[45] Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. (2004). what’s in a translation rule? In Susan Dumais, D. M. and Roukos, S., editors, HLT-NAACL 2004: Main Proceedings, pages 273–280, Boston, Massachusetts, USA. Association for Computational Linguistics.

[46] Li, Zhifei, Chris Callison-Burch, Chris Dyer, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren NG Thornton, Jonathan Weese, and Omar F. Zaidan. "Joshua: An open source toolkit for parsing-based machine translation." *In Proceedings of the Fourth Workshop on Statistical Machine Translation*, pp. 135-139. Association for Computational Linguistics, 2009

[47] Alshawi, Hiyan, Srinivas Bangalore, and Shona Douglas. "Learning dependency translation models as collections of finite-state head transducers." Computational Linguistics 26, no. 1 (2000): 45-60.

[48] Melamed I.Dan. Dan. "Statistical machine translation by parsing." *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics.* Association for Computational Linguistics, 2004.

[49] http://www.systransoft.com/systran/news-and-events/press-release/hybrid-machine-translation-solution-for-enterprises/ last visited June 27, 2017

[50] John S. White, Theresa A. O'Connell, and Lynn M. Carlson. "Evaluation of machine translation." In *Proceedings of the workshop on Human Language Technology*, pp. 206-210. Association for Computational Linguistics, 1993.

[51] Papineni Kishore, Salim Roukos, Todd Ward and Wei-Jing Zhu. "BLEU: a method for automatic evaluation of machine translation." *In Proceedings of the 40th annual meeting on association for computational linguistics,* pp. 311-318. Association for Computational Linguistics, 2002.

[52] Hanneman Greg, Edmund Huber, Abhaya Agarwal, Vamshi Ambati, Alok Parlikar, Erik Peterson and Alon Lavie. "Statistical transfer systems for French--English and German--English machine translation." In *Proceedings of the Third Workshop on Statistical Machine Translation*, pp. 163-166. Association for Computational Linguistics, 2008.

[53] Alon Lavie. Stat-XFER: "A general search-based syntax-driven framework for machine translation". In *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer Berlin Heidelberg. 2008 Feb 17 (pp. 362-375).

[54] Dan Klein and Christopher D. Manning. "Fast exact inference with a factored model for natural language parsing. *"Advances in neural information processing systems,*2003: 3-10.

[55] Salah Ait-Mokhtar, Jean-Pierre Chanod and Claude Roux."A multi-input dependency parser". In *Proceedings of the Seventh International Workshop on Parsing Technologies*, Beijing, China, October, 2001.

[56] Mossa Ghurab, Yueting Zhuang, Jiangqin Wu and Maan Younis Abdullah. "Arabic-Chinese and Chinese-Arabic Phrasebased statistical machine translation systems*". Inform. Technol*. J. 2010;9: 666-72.

[57] Björn Gambäck. "Tagging and Verifying an Amharic News Corpus." Language Technology for Normalisation of Less-Resourced Languages (2012): 79.

Annex I: Amharic POS Tag Set

No.	Tags	Description	Examples
	Noun		**Any entity belonging to noun**
1.	PN	Personal noun(can be single, plural or mass)	አበበ፣ ሀይለ/^ebebe, heyle
	PNP	Personal noun attached to preposition	የአበበ፣ ስለአስቴር፣ ለሀይላይ/ye^ebebe, sle^estEr, leheylay
	PNC	Personal noun attached to conjunction	አበበና ከበደ/ ^ebebena kebede
	PNPC	Personal noun attached to preposition and conjunction.	የአበበና ከበደ/ye^ebebena kebede
2.	CN	Any compound word (can be single, plural and mass) which changes its order of words in the target language	ትምህርት ቤት፣ ምክር ቤት፣ ጽሕፈት ቤት/tmhrt bEt, mkr bEt, SHfet bEt
	CNP	Compound word attached to preposition.	የትምህርት ቤት፣ ለምክር ቤት፣ ከጽሕፈት ቤት/yetmhrt bEt, lemkr bEt, keSHfet bEt
	CNC	Compound word attached to conjunction.	ምክር ቤትና፣ ትምህርት ቤቱና/mkr bEtna, tmhrt bEtuna
	CNPC	Compound word attached to preposition and conjunction.	ለምክር ቤትና፣ ከጽሕፈት ቤቱና/lemkr bEtna, keSHfet bEtuna
3.	N2	Compound words never change their order in both leagues (can be single, plural or mass).	ራስ ዳሸን፣ አዲስ አበባ፣ ሰሜን ጎንደር፣ ምዕራብ ትግራይ፣ ሰሜን ሸዋ/ ras daxn, ^edis ^ebeba, semEn gonder, m`rab tgray, semEn xewa
	N2P	N2 attached to preposition	የራስ ዳሸን፣ ከአዲስ አበባ/yeras daxn, ke^edis ^ebeba
	N2C	N2 attached to conjunction	አዲስ አበባና፣ ሰሜን ሸዋና/^edis ^ebebana, semEn xewana
	N2PC	N2 attached to preposition and conjunction.	የአዲስ አበባና፣ የሰሜን ሸዋና/ye^edis ^ebebana, yesemEn xewana
4.	N	Noun different from the rest (can be single, plural or mass)	ኮኮቦች፣ ድርጅት፣ ትርፍ፣ ሚኒስቴር, መኪና/kokoboc, drjt, trf, ministEr mekina
	NP	Any noun including verbal noun attached with preposition	በመኪና፣ የድርጅቱ፣ ከጎንደር/bemekina, yedrjtu, kegonder
	NC	Any noun including verbal noun attached with conjunction	እጥረትና፣ ፖሊሲና፣ ባህርያትና/^Tretna, polisina, bahryatna

	NPC	Any noun including verbal noun attached with preposition and conjunction	በመሆኑም፣ ከመሆኑም፣ የአባላትና፣ የፖሊሲና፣ በአዋጅና/bemehonum, kemehonum, ye^ebalatna, yepolisina, be^ewajna
5.	VN	Verbal/infinitival noun, formed from any verb form such as active, passive and repetitive.	መከፈላቸውን፣ ማስተዋወቅ፣ ማንቀሳቀስ፣ መጫወት፣ መምጣቴ፣ መውረድ፣ መታየቱን/mekefelacewn, mastewaweq, manqesaqes, meCawet, memTatE, mewred,metayetun
	Pronoun.		
6.	PRON	Personal pronoun	እሱ፣ እሷ፣ እኔ፣ አንተ፣ እነሱ፣ ማን፣ ይኸው፣ ይህ/^su, ^sWa, ^nE, ^ente, ^nesu, man, yKew, yh
	PRONP	Pronoun attached with prepositions	Prepositions like ከ፣ ለ and የ attached to any pronoun. ከእሱ፣ የኔ፣ የንተ፣ በዚህ/ke, le ye, ke^su, yenE, yente, bezihe
	PRONC	Pronoun attached with conjunction	Conjunctions attached to any pronoun እኔና፣ ይህም፣ አምላክም/^nEna, yhm, ^emlakm
	PRONPC	pronoun attached with preposition and conjunction	Prepositions and Conjunctions attached to any pronoun. የእሱና፣ የኔና/ye^suna, yenEna
	Verb		
7.	V	Any verbs out of the rest verbs	በላ፣ ጠጣሁ፣ ሰበረ፣ መጣሁ፣ አስታወቀ/bela, TeTahu, sebere, meTahu, ^estaweqe
	VP	Verbs including relative and auxiliaries attached with prepositions	ባወጣው, እንዳስታወቀው, የተፈራረሙት, እንደተጠቀሰው, እንደሚካሄድ/baweTaw ^ndastaweqew, yeteferaremut, ^ndeteTeqesew, ^ndemikahEd
	VC	Verbs including relative and auxiliaries attached with conjunctions	Verbs with conjunction like ና, ሰራና፣ መጣና/na, serana, meTana
	VPC	Verbs including relative and auxiliaries attached with preposition and conjunction	ባወጣውና, እንዳስታወቀውና, የተፈራረሙትና, እንደተጠቀሰውና, እንደሚካሄድና/baweTawna, ^ndastaweqewna, yeteferaremutna, ^ndeteTeqesewna, ^ndemikahEdna
	AXU	Auxiliary verb	ነች፣ ናት፣ ነኝ፣ ናቸው፣ ነው/nec, nat, neN, nacew, new
	VREL	Relative verbs	የሚያሳየው፣ የወሰዱ፣ የሚያስችል፣ ያለው/yemiyasayew, yewesedu, yemiyascl, yalew
8.	**Adjective**		
9.	ADJ	The adjective different from the rest	All types of adjective; ቀይ፣ ከፍተኛ፣ ትልቅ፣ ጎበዝ፣ ጎባጣ፣ መሪ/qey, kefteNa, tlq, gobez, gobaTa, meri
	ADJP	Adjective attached with preposition	በመሪ፣ ለቀዩ፣ ከትልቁ/bemeri, leqeyu, ketlqu
	ADJC	Adjective attached with conjunction.	ቀይና፣ ቆራጥና፣ ጎበዝና ፣ አለማቀፋዊና/qeyna, qoraTna, gobezna , ^elemaqefawina
	ADJPC	Adjective attached with preposition and conjunction	በጥቃቅንና፣ የአነስተኛና፣ ፣ በመካከለኛና/beTqaqnna, ye^enesteNana, , bemekakeleNana
	Numerals		

10.	NUMCR	Cardinal number	አንድ፣ 2፣ 11ኪ፣ 45፣ 145ቱ/^end, 2, 11xi, 45, 145tu
	NUMOR	Ordinal number	መቶኛ, 2ኛ፣ 3ኛ አስረኛ/metoNa, Na, Na ^esreNa
	NUMP	Numeral attached with preposition	በመቶው, የ1993, የ8ኛ/bemetow, ye1993, ye8Na
	NUMC	Numeral attached with conjunction	ሁለተኛና፣ ሶስተኛና/huleteNana, sosteNana
	NUMPC	Numeral attached with preposition and conjunction	በሁለቱም፣ የሁለቱም፣ ከ12ቱም/behuletum, yehuletum, ketum
11.		**Prepositions**	
12.	PRP	Preposition or subordinating conjunction	ለ፣ ከ፣ ስለ፣ ስለ፣ የ፣ እንደ/le, ke, sle, sle, ye, ^nde
13.	IN	Prepositions appear after nouns and compound nouns	ጋራ፣ ላይ፣ መካከል/gara, lay, mekakel
14.		**Adverbs**	
15.	ADV	Adverb	ትናንት፣ያለፈው ሳምንት፣ በቀጣይ፣ በተለይ፣ ባለፉት፣ ብቻ፣ ዛሬ/tnant, yalefew samnt, beqeTay, beteley, balefut, bca, zarE,
		Punctuation	
16.	PUN	Punctuation at the corpus	፣። ? ፤ ; ፥ /, . ? ; :
17.		**Coordinate conjunction**	
18.	CC	Coordinating Conjunction	እና፣ ወይም፣ ግን፣ እስከ፣ ደግሞ፣ እንጂ፣ እንኳን፣ ስለዚህ፣ ነገር ግን፣ ቢሆንም/^na, weym, gn, ^ske, degmo, ^nji, ^nkWan, slezih, neger gn, bihonm
19.	Interjection	INT	
20.	UNC	Unclear	

Annex II: Sample Parallel Corpus for Training

Amharic sentences	Tigrigna sentences
የኢትዮጵያ ፌዴራል መንግሥት ከፍተኛ የአስፈጻሚነት ሥልጣን የተሰጠው ለጠቅላይ ሚኒስትሩና ለሚኒስትሮች ምክር ቤት ነው ።	ዝለዓለ ስልጣን ላፈፃምነት መንግስቲ ፌዴራል ኢትዮጵያ ዝተውሃበ ንቐዳማይ ሚኒስትርን ንቤት ምኽሪ ሚኒስተራትን እዩ።
የኢትዮጵያ ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ ሕገ መንግስት ከነሓሴ 15 1987 ዓ.ም ጀምሮ ሙሉ በሙሉ ላይ በስራ ወሏል።	ሕገ መንግስቲ ፌዴራላዊ ዲሞክራሲያዊ ሪፐብሊክ ኢትዮጵያ ካብ ነሓሰ 15 1987 ዓ/ም ኣትሒዙ ብዕሊ ኣብ ስራሕ ውዒሉ ኣሎ።
የኢትዮጵያ ቴሌቪጅን ድርጅት ዋና ዳይሬክተር	ዋና ዳይሬክተር ትካል ቴሌቪጅን ኢትዮጵያ
የከተማዋ ኮሙዩኒኬሽን ጉዳዮች ጽህፈት ቤት ኃላፊ አቶ አብይ ኃይሉ እንደገለጹት ከንቲባዎቹ የተሾሙት ያላቸውን የትምህርት ዝግጅት ፣ ሕዝባዊ ወገንተኝነታቸውን የሥራ ልምድና መሰረት በማድረግ ነው።	ሓላፊ ቤት ፅሕፈት ጉዳያት ኮሙኒኬሽን እታ ከተማ ኣይተ ኣብዩ ሃይሉ ከምዝገለፅዎ እቶም ከንቲባታት ዝተሰየሙ ብዘለዎም ምድላው ትምህርቲ ፤ ህዝባዊ ወገናውነቶምን ልምዲ ስራሕን መሰረት ብምግባር እዩ።
ወንበዴዎች በጐ ፈቃደኞች የነቡትን ተቋም ከጥቅም ውጭ አደረጉ	እቶም ዓዋሉ ገበርቲ ሰናይ ዝሃነፅዎ ትካል ካብ ጥቕሚ ወፃኢ ገይሮም
ዛሬ ወደ ስራ አልሄድም	ሎሚ ናብ ስራሕ ኣይከይድን
ንፋሱ ዘላኖች አሸዋ ላይ የሰሩትን ቤት ጨርሶ አፈረሰ	እቲ ንፋስ ሰበኽ-ሳግም ኣብ ሑፃ ንዝሰርሕዎ ገዛ መሊኡ ኣፅንዩዎ
አስቴር እህቷ በጣም ስለወፈረች ከመጠን በላይ ተናደደች	ኣስቴር ሓፍታ ብጣዕሚ ስለዝረጎደት ካብ መጠን ንላዕሊ ተናዲዳ
ከበደ ስለታመመ ትምህርት ቤት አልመጣም	ከበደ ሰለዝሓመመ ቤት ትምህርቲ ኣይመፀን
ዘበኛው አሰሪው ስለፈቀደለት የማታ ትምህርት ተመዘገበ	እቲ ዋርድያ ሓላፊኡ ስለዝፈቐደሉ ምሽታዊ ትምህርቲ ተመዝጊቡ
ልጅቱ ቁርጥ እናቷን ትመስላለች	እታ ጀልዓ ቅምጥ-ኢላ ነዲኣ ትመስል እያ
ወደ አዲስ አበባ መቼ መጣህ	ናብ ኣዲስ ኣበባ መዓዝ መፂእኻ
የትምህርት ሚኒስቴር ለኤችአይቪ ኤድስ መከላከያ 10.8 ሚሊየን ብር መደበ	ሚኒስትሪ ትምህርቲ ኤች ኣይ ቪ ኤድስ ንምክልኻል 10.8 ሚሊየን ቕርሺ መዲቡ
እመጣለሁ ብለህ ለምን ቀረህ ?	ክመፅእ እየ ኢልካ ንምንታይ ተሪፍካ ?
ያዘዝኩህን ብቻ ሰርተህ ጠብቀኝ ?	ዝኣዘዝኩኻ ጥራሕ ሰሪሕኻ ፅንሐኒ ?
ህጻኑ ስለታመመ ምግብ አልተመገበም	እቲ ህፃን ስለዝሓመመ ምግቢ ኣይበልዐን
አበበና ካሳ ለምን ተጣሉ	ኣበበን ካሳን ንምንታይ ተፃሊኦም
የቤት ስራዉን ለነገ ሰርታችሁ እንድትመጡ	እቲ ዕዮ ገዛ ንፅባሕ ሰሪሕኹም ከምትመፁ
ፈተናዉን በትእዛዙ መሰረት ብቻ ስሩ	ነቲ ፈተና ብመሰረት እቲ ትእዛዝ ጥራሕ ስርሑዎ
የካሳ ጓደኛ እንደ ካሳ ጎበዝ ተማሪ ለመሆን ሞከረ	ዓርኪ ካሳ ከም ካሳ ንፉዕ ተምሃራይ ንምኳን ፈቲኑ
ወፎቹ በቀለ ያፈሰሰውን ስንዴ በችኮላ ለቀሙ	እተን ኣዕዋፍ በቀለ ዘፍሰሶ ስርናይ ብቕልጡፍ ኣርየንኡ
አሮጊቱ ሌቦች አምና የገደሉትን ብቸኛ ልጃቸውን አስታወሱ	እተን ሽማግለ ለያቡ ዓሚ ዝቐተልዎ ብሕታዊ ወደን ዘኪረን
ተማሪው ጓደኛው ስለዘገየ ወደ ክፍል ብቻውን ገባ	እቲ ተምሃራይ ዓርኩ ስለዝደንጎየ ናብ ክፍሊ ንበይኑ ኣትዩ
እኛ አስተማሪው የሰጠንን የቤት ስራ ክፍል ውስጥ ሰራን	ንሕና እቲ መምህር ዝሃበና ዕዮ ገዛ ኣብ ውሽጢ ክፍሊ ሰሪሕና
ኤርትራ ከስንት ሀገሮች ትዋሰናለች	ኤርትራ ምስ ክንደይ ሃገራት ትዋሰን
ኢትዮጵያ ከስንት ሀገሮች ትዋሰናለች	ኢትዮጵያ ምስ ክንደይ ሃገራት ትዋሰን
ነጋዴው ገበሬው የሸጠለትን እህል ወደ ከተማ አመጣ	እቲ ነጋዳይ እቲ ገባራይ ዝሸጠሉ እኽሊ ናብ ከተማ ኣምፂኡ
ሚኒስተሩ ወታደሮቹ ሀገራቸውን ከወራሪ ስለታደጉ በጣም አመሰገኑ	እቶም ሚኒስተር እቶም ወተሃደራት ሃገሮም ካብ ወራሪ ስለዘድሐኑ ብጣዕሚ .ኣመስጊኖም
አበራ ወንድሙ የጋበዘውን ጎበዝ የኮሌጅ ተማሪ ተዋወቀ	ኣበራ ሓዉ ዝዓደሞ ንፉዕ ተምሃራይ ኮሌጅ ተፋሊጡ
አባቴ ከስራ ሲመጣ ወንድሜ ማንበቢያ ክፍል ገባ	ኣቦይ ካብ ስራሕ ምስመፀ ሓወይ ናብ መንበቢ ክፍሊ ኣትዩ

የኢትዮጵያ መንግስት በአሜሪካ የደረሰውን አደጋ አወገዘ	መንግስቲ ኢትዮጵያ ኣብ ኣሜሪካ ንዝበፅሐ ሓደጋ ኣውጊዙ
አባላቱ ያካሄዱት ውይይት ዴሞክራሲያዊ ነበር /V	እቶም ኣባላት ዘካየድዎ ዘተ ዲሞክራሲያዊ ነይሩ
በአዳማ በ6.8 ሚሊየን ብር ፕሮጀክቶች ተሰሩ	ኣብ ኣዳማ ብ6.8 ሚሊየን ቕርሺ ፕሮጀክትታት ተሰሪሑም
በትግራይ ክልል 2047 ተማሪዎች የ12ኛ ክፍል መልቀቂያ ፈተና ወሰዱ	ኣብ ክልል ትግራይ 2047 ተምሃሮ ናይ 12 ክፍሊ መልቀቒ ፈተና ወሲዶም
የኢትዮጵያ የትምህርት ሽፋን በ6.4 በመቶ ማደጉ ተገለፀ	ሽፋን ትምህርቲ ኢትዮጵያ ብ6.4 ሚኢታዊ ምዕባዩ ተሓቢሩ
የትምህርት ግባትና ቁሳቁስ ተሟላ	ቁሳቒስን እታወታትን ትምህርቲ ተማሊኡ
መከላከያ ምኒስትሩ እንደገለጹት 11 ሚሊየን ብር በላይ ለተቀናሽ ሰራዊቶች እንደተከፋፈለ ተናግረዋል	እቶም ሚኒስቲር ምክልኻል ከምዝገለፅዎ ካብ 11 ሚሊየን ቕርሺ ንላዕሊ ንተቐነስቲ ሰራዊት ከምዝተኸፋፈለ ተዛሪቦም
ሀኪሙ ልጇ አስቸጋሪ የሆነውን ሴትዮ በቅድሚያ አስተናገደ	እቲ ሓኪም ረባሺ ቆልዓ ንዘለዋ ሰበይቲ ቀዲሙ ኣስተኣናጊዱ
እኔ ካሳ የገዛውን ነጭ በግ ትላንት አየሁ	ኣነ ካሳ ዝገዝኦ ፃዕዳ በጊዕ ትማሊ ሪእ
ቦርሳው ብዙ እቃ በውስጡ ስለያዘ በጣም ከበደኝ	እቲ ቦርሳ ብዙሕ ኣቕሓ ኣብ ውሽጡ ስለዝሓዘ ብጣዕሚ ከቢድኒ
በውጪ የሚኖሩ ኢትዮጵያዊያን 260ሺ ዶላር ለገሱ	ኣብ ወፃኢ ዝነብሩ ኢትዮጵያዊያን 260ሽሕ ዶላር ኣወፍዮም
ሻዕቢያ ተጨማሪ ከፍተኛ ባለስልጣናትን አሰረ	ሻዕቢያ ተወሰኽቲ ላዕለዎት ሰበስልጣናት ኣሲሩ
ምክር ቤቱ ከጣሊያን ተቋም ጋር በመተባበር የአቅም ግንባታ ስልጠና አዘጋጀ	እቲ ቤት ምኽሪ ምስ ጣልያናዊ ትካል ብምትሕብባር ናይ ዓቕሚ ህንፀት ስልጠና ኣዳልዩ
እንግሊዝ በኤርትራ የሚፈፀመውን እስራት ተቃወመች	እንግሊዝ ኣብ ኤርትራ ንዝፍፀም ማእሰርቲ ተቓዊማ
ለፖርክ የሚሆኑ ቦታዎች በቦረና ዞን ተከለሉ	ንፓርክ ዝኾኑ ቦታታት ኣብ ዞባ ቦረና ተኸሊሎም
የእርዳታ እህል ተከፋፈለ	ናይ ረዲኤት እኽሊ ተኸፋፈሉ
ፖስታ ቤቱ ድንገተኛ ጭማሪ አደረገ	እቲ ቤት ፖስታ ሃንደበታዊ ወሰኽ ገይሩ
የድርጅቱ አባላት በጉባኤው ተደሰቱ	ኣባላት ናይቲ ውድብ በቲ ጉባኤ ተሓጎሶም
የኢፌዴሪ የካቢኔ አባላት ሹመት ፀደቀ	ሽመት ኣባላት ካቢኔ ፌዴሪኢ ፀዲቑ
የኢትዮጵያ አየር መንገድ ወደ አሜሪካ ለመብረር ተፈቀደለት	መንገዲ ኣየር ኢትዮጵያ ናብ ኣሜሪካ ንምብራር ተፈቒድሉ
ምክር ቤቱ ሁለት አዋጆችን መረመረ	እቲ ቤት ምኽሪ ኽልተ ኣዋጃት መርሚሩ
አልማ የህክምና መሳሪያዎችን አሰራጨ	አልማ መሳርሕታት ሕክምና ኣከፋፊሉ
አለምአቀፍ የምግብ ቀን ነገ ይከበራል	ዓለምለኻ መዓልቲ ምግቢ ፅባሕ ክኽበር እዩ
ጤና ጥበቃ ምኒስተር የወባ በሽታን ለመከላከል እንቅስቃሴ ጀመረ	ምኒስቲሪ ሓለዋ ጥዕና ሕማም ዓሶ ንምክልኻል ምንቅስቓስ ጀሚሩ
እነሱ መልስ ፈልገዋል ።	ንሳቶም መልሲ ደልዮም።
እሷ መልስ ሰጠችው ።	ንሳ መልሲ ሂባቶ።
ሳራ እያወራች ነው ።	ሳራ ትዛረብ ኣላ።
እሷ የሳራ እናት ነች ።	ንሳ ኣዶ ሳራ እያ።
እሱ እናቱን ሊያይ ሄደ ።	ንሱ እዲኡ ክርኢ ከይዱ።
እሱ እርዳታ ጠየቀ ።	ንሱ ሓገዝ ሓቲቱ።
እሷ ጥፋተኛ አይደለችም ።	ንሳ ጥፍኣተኛ ኣይኮነትን።
የእሱ ጥፋት ቡዙ ነው ።	ናቱ ጥፍኣት ቡዙሕ እዩ።
እሱ የእሷ ነፃነት አረጋገጠ ።	ንሱ ናታ ነፃነት ኣረጋጊፁ።
እሱ እሷን እየረዳት ነው ።	ንሱ ንዓኣ ይሕግዛ ኣሎ።
እሷ ወንድ ልጅ አላት ።	ንሳ ወዲ ቆልዓ ኣለዋ።
እሷ ወንድ ልጇን ጥላው ሄደች ።	ንሳ ንወዳ ጉዲፋቶ ከይዳ።
እሱ ለእሷ ሚስጥር ነገራት ።	ንሱ ንዓኣ ሚስጥር ነጊሩዋ።

እሱ እሷን ማወቅ ይፈልጋል ።	ንሱ ንዓኣ ክፈልጣ ይደሊ እዩ።
እነሱ እየተጣሉ ነው ።	ንሳቶም ይበኣሉ ኣለው።
እሱ ድብድቡን አሸንፏል ።	ንሱ ነቲ ባእሲ ኣሸኒፉ።
እነሱ ሰውየውን ገደሉት ።	ንሳቶም ነቲ ሰብኣይ ቀቲሎምዎ።
እሷ እሱ የት እንዳለ አታውቅም ።	ንሳ ንሱ ኣበይ ከምዘሎ ኣይትፈልጥን።
እሱ ዘጠኝ አመቱ ነው ።	ንሱ ትሽዓተ ዓመቱ እዩ።
እሱ አስር አመቱ ነው ።	ንሱ ዓስርተ ዓመቱ እዩ።
እሷ እሱን በዕድሜ ትበልጠዋለች ።	ንሳ ንዕኡ ብዕድመ ትበልፆ እያ።
እሷ ከእሱ ታጥራለች ።	ንሳ ካብኡ ትሓፅር እያ።
እሱ ከእሷ ይረዝማል ።	ንሱ ካብኣ ይነውሕ እዩ።
እኛ የምንሄደው ወዴት ነው ።	ንሕና እንኸዶ ናበይ እዩ።
ምድርም ባዶ ነበረች አንዳችም አልነበረባትም ፥ ጨለማም በጥልቁ ላይ ነበረ። የእግዚአብሔርም መንፈስ በውኃ ላይ ሰፍፎ ነበር።	ምድሪ ድማ በረኻን ጥራያን ነበረት፡ ፀልማት ከኣ ኣብ ልዕሊ መዓሙቕ ነበረ። መንፈስ ኣምላኽ ድማ ኣብ ልዕሊ ማያት ይዝምቢ ነበረ።
እግዚአብሔርም ፦ ብርሃን ይሁን አለ። ብርሃንም ሆነ።	ኣምላኽ ከኣ፥ ብርሃን ይኹን በለ። ብርሃን ድማ ኾነ።
እግዚአብሔርም ብርሃኑ መልካም እንደ ሆነ አየ። እግዚአብሔርም ብርሃንንና ጨለማን ለየ።	ኣምላኽ ድማ እቲ ብርሃን ፅቡቕ ከም ዝኾነ ረኣየ። ኣምላኽ ከኣ ነቲ ብርሃን ካብ ፀልማት ፈለዮ።
እግዚአብሔርም ፦ በውኖች መካከል ጠፈር ይሁን ፥ በውኃና በውኃ መካከልም ይክፈል አለ።	ኣምላኽ ድማ፥ ንማያት ካብ ማያት ዚፈሊ ጠፈር ኣብ መንነ ማያት ይኹን፡ በለ።
እግዚአብሔርም ጠፈርን አደረገ ፥ ከጠፈር በታችና ከጠፈር በላይ ያሉትንም ውኖች ለየ ፡ እንዲሁም ሆነ።	ኣምላኽ ነቲ ጠፈር ገበሮ። ነቲ ኣብ ትሕቲ ጠፈርን ኣብ ልዕሊ ጠፈር ዘሎ ማያት ፈለዮ ፡ ከምኡውን ኾነ።
እግዚአብሔር ጠፈርን ሰማይ ብሎ ጠራው። ማታም ሆነ ጥዋትም ሆነ ፥ ሁለተኛ ቀን።	ኣምላኽ ከኣ ነቲ ጠፈር ሰማይ ኣውፅኣሉ። ምሸት ኩነ ብጊሓትውን ኩነ ፡ ካልኣይቲ መዓልቲ።
እግዚአብሔርም ፦ ከሰማይ በታች ያለው ውኃ በአንድ ስፍራ ይሰብሰብ አለ ፥ የብሱም ይገለጥ አለ እንዲሁም ሆነ።	ኣምላኽ ድማ ፥ እቲ ኣብ ትሕቲ ሰማይ ዘሎ ማያት ናብ ሓንቲ ቦታ ይተኣኻብ በለ። እቲ ንቐፅ ምእንቲ ኺርኤስ ከምኡ ድማ ኾነ።
እግዚአብሔርም የብሱን ምድር ብሎ ጠራው ፡ የውኃ መከማቻውንም ባሕር አለው። እግዚአብሔርም ያ መልካም እንደ ሆነ አየ	ኣምላኽ ከኣ ነቲ ንቐፅ ምድሪ ኣውፅኣሉ፡ ነቲ እኩብ ማያት ድማ ባሕሪ ኣውፅኣሉ። ኣምላኽ ከኣ ፅቡቕ ከም ዝኾነ ረኣየ።
ምድርም ዘርን የሚሰጥ ሣርንና ቡቃያን እንደ ወገኑ ዘሩም ያለበትን ፍሬን የሚያፈራ ዛፍን እንደ ወገኑ አበቀለች። እግዚአብሔርም ያ መልካም እንደ ሆነ አየ።	እታ ምድሪ ሳዕርን ከከም ዓይነቱ ዘርኢ ዚህብ ብቑልን ዘርኡ ኣብ ርእሱ ዘለዎ ፍረ ዚፈሪ ኣእዋም ከኣ ኣውፅኤት። ኣምላኽ ድማ ፅቡቕ ከም ዝኾነ ረኣየ።
ማታም ሆነ ጥዋትም ሆነ ፥ ሦስተኛ ቀን	ምሸት ኩነ ብጊሓትውን ኩነ ፡ ሳልሳይቲ መዓልቲ
እግዚአብሔርም አለ ፦ ቀንና ሌሊትን ይለዩ ዘንድ ብርሃናት በሰማይ ጠፈር ይሁኑ ፡ ለዘመኖች ለዕለታት ለዓመታትም ምልክቶችም ይሁኑ።	ኣምላኽ ድማ በለ ፤ ንመዓልቲ ኻብ ለይቲ ዚፈልዩ ብርሃናት ኣብ ጠፈር ሰማይ ይኹኑ ፡ ንዘበናትን ንመዓልትታትን ዓመታትን ከኣ ንመፈለጥታ ይኹኑ።
በምድር ላይ ያበሩ ዘንድ በሰማይ ጠፈር ብርሃናት ይሁኑ ፡ አለ። እንዲሁም ሆነ	ኣብ ምድሪ ምእንቲ ከብርሁ ብርሃናት ኣብ ጠፈር ሰማይ ይኹን፡ በለ። ከምኡ ድማ ኾነ።
እግዚአብሔርም ሁለት ታላላቆች ብርሃናትን አደረገ ፥ ትልቁ ብርሃን በቀን እንዲሠለጥን ፥ ትንሹም ብርሃን በሌሊት እንዲሰለጥን ፡ ከዋክብትንም ደግሞ አደረገ	ኣምላኽ ከኣ ኽልተ ዓበይቲ ብርሃናት ገበረ ፥ እቲ ዓብዪ ብርሃን ብመዓልቲ ኺስልጥን፡ እቲ ንእሽቶ ብርሃን ድማ ብለይቲ ኺስልጥን ፡ ከዋኽብቲ እውን ገበረ
እግዚአብሔርም በምድር ላይ ያበሩ ዘንድ በሰማይ ጠፈር አኖራቸው። በቀንም በሌሊትም እንዲሠለጥኑ ፥ ብርሃንንና ጨለማንም እንዲለዩ ፡ እግዚአብሔርም ያ መልካም እንደ ሆነ አየ	ኣምላኽ ድማ ኣብ ልዕሊ ምድሪ ምእንቲ ኼብርሁ ኣብ ጠፈር ሰማይ ገበሮም። ኣብ መዓልትን ኣብ ለይትን ድማ ኪስልጥኑ ፡ ንብርሃንውን ካብ ፀልማት ኪፈልዩ ፡ ኣምላኽ ድማ ፅቡቕ ከም ዝኾነ ረኣየ
ማታም ሆነ ጥዋትም ሆነ ፥ አራተኛ ቀን	ምሸት ኩነ ብጊሓትውን ኩነ ፡ ራብዐይቲ መዓልቲ
እግዚአብሔርም አለ ፦ ውኃ ሕያው ነፍስ ያላቸውን ተንቀሳቃሾች ታስገኝ ፥ ወፎችም ከምድር በላይ ከሰማይ ጠፈር በታች ይብረሩ	ኣምላኽ ድማ፥ ማያት ህያው ነፍሲ ዘለዎ ለመምታ የውፅእ ፡ ኣዕዋፍ ከኣ ኣብ ልዕሊ ምድሪ ኣብ ትሕቲ ጠፈር ሰማይ ይንፈራ ፡ በለ።

እግዚአብሔርም ታላላቆች አንበሪዎችን ፥ ውኃይቱ እንደ ወገኑ ያስገኝቻቸውንም ተንቀሳቃሾቹን ሕያዋን ፍጥረታት ሁሉ ፥ እንደ ወገኑ የሚበሩትንም ወፎች ሁሉ ፈጠረ። እግዚአብሔርም ያ መልካም እንደ ሆነ አየ	ኣምላኽ ድማ ነቶም ዓበይቲ እንስሳ ባሕርን ፥ ነቲ ማያት ዘውፅኦ በብዓይነቱ ህያው ነፍሲ ዘለዎ ውንጅርጅር ዚብል ኵሉን ፡ ክንፊ ዘለወን በብዓይነተን ኵለን ኣዕዋፍን ፈጠረ። ኣምላኽ ከኣ ፅቡቕ ከም ዝኾነ ረኣየ።
ማታም ሆነ ጥዋትም ሆነ ፥ አምስተኛ ቀን	ምሸት ኰነ ብጊሓትውን ኰነ፡ ሓምሲይቲ መዓልቲ።
እግዚአብሔርም አለ ፦ ምድር ሕያዋን ፍጥረታትን እንደ ወገኑ እንስሳትንና ተንቀሳቃሾችን የምድር አራዊትንም እንደ ወገኑ ታውጣ ፡ እንዲሁም ሆነ	ኣምላኽ ድማ በለ ፥ ምድሪ ህያው ነፍሲ ዘለዎ በብዓይነቱ እንስሳን ለመምታን ኣራዊት ምድርን በብዓይነቱ ተውፅእ ፡ ከምኡውን ኰነ።
እግዚአብሔርም ባረካቸው ፥ እንዲህም አላቸው ፦ ብዙ ተባዙ ፥ ምድርንም ሙሉአት ፥ የባሕርን ዓሦችና የሰማይን ወፎች ፣ በምድር ላይ የሚንቀሳቀሱትንም ሁሉ ግዙአቸው።	ኣምላኽ ከኣ ባረኾም። ኣምላኽ ድማ ፡ ፍረዩን ተባዝሑን ፥ንምድሪ ኽኣ ምልእዋን ምለኽዋን ፡ ነዕዋፍ ሰማይን ንዓሳ ባሕርን ፣ ኣብ ምድሪ ለመም ንዝብል ኩሉ እንስሳን ከኣ ግዝኡ
ለምድርም አራዊት ሁሉ ፥ ለሰማይም ወፎች ሁሉ ፥ ሕያው ነፍስ ላላቸው ለምድር ተንቀሳቃሾችም ሁሉ የሚበቅለው ሐመልማል መብል ይሁንላቸው ፡ አለ። እንዲሁም ሆነ።	ንኹሉ ኣራዊት ምድርን ንኹለን ኣዕዋፍ ሰማይን ህያው ነፍሲ ንዘለዎ ኣብ ምድሪ ለመም ንዚብል ኵሉ ኽኣ ኵሉ ለምለም ሳዕሪ ንምግቦም ሂበዮም አሎኹ ፡ በለ። ከምኡ ድማ ኾነ።
ሰማይና ምድር ሠራዊታቸውም ሁሉ ተፈጸሙ።	ሰማይን ምድርን ሰራዊቶምን ኵሉ ተፈፀሙ።
እግዚአብሔርም የሠራውን ሥራ በሰባተኛው ቀን ፈጸመ ፡ በሰባተኛውም ቀን ከሠራው ሥራ ሁሉ ዐረፈ።	ኣምላኽ ከኣ ነቲ ዝገበሮ ግብሩ በታ ሳብዐይቲ መዓልቲ ፈፀሞ ፡ ብሳብዓይቲ መዓልቲ ድማ ኻብቲ ዝገበሮ ግብሩ ዅሉ ዐረፈ።
እግዚአብሔርም ሰባተኛውን ቀን ባረከው ቀደሰውም እግዚአብሔር ሊያደርገው ከፈጠረው ሥራ ሁሉ በእርሱ ዐርፎአልና።	ኣምላኽ ነታ ሳብዐይቲ መዓልቲ ባረኻን ቀደሳን ከኣ ኣምላኽ ኻብቲ ዝፈጠሮን ዝገበሮን ኵሉ ግብሩ ብእኣ ስለ ዝዐረፈ፡
እግዚአብሔር አምላክ ሰማይንና ምድርን ባደረገ አምላክ ሰማይንና ምድርን ቀን፥ በተፈጠሩ ጊዜ የሰማይና የምድር ልደት ይህ ነው።	በታ እግዚኣብሄር ኣምላኽ ምድርን ሰማይን ዝፈጠረላ መዓልቲ ምስ ተፈጥሩ፡ ወለዶ ሰማይን ምድርን እዚ እዩ።
ስራ አስኪያጁ አያይዘውም በዞኖቹ የሚገኙ 43 የቡና አቅራቢ ማህበራት ከ53 ሚሊየን ብር በላይ እዳ በደርቅ ቼክ መጭበርበር ምክንያት መከፈል ባለመቻላቸው ባንኩ ተጨማሪ የእዳ ማራዘሚያ ጊዜ መስጠቱንም ገልጸዋል	እቲ መካየዲ ስራሕ ኣተሓሒዞምውን ኣብተን ዞባታት ዝርከባ 43 ማሕበራት ኣቅረብቲ ቡና ካብ 53 ሚሊዮን ቅርሺ ንላዕሊ ዕዳ ብደረቅ ቼክ ምጭብርባር ምኽንያት ክኽፍሉ ብዘይምኽኣሎም እቲ ባንኪ ተወሳኺ መናውሒ ዕዳ ጊዜ ከምዝሃበ ገሊፁ።
የብድር አቅማቸውን አስቀድመው እንዲገነዘቡና ህጋዊ አሰራርን አጠናክረው እንዲቀጥሉ ጉልህ አስተዋእኦ ያበረክታል።	ናይ ልቓሕ ዓቕሞም ኣቕዲሞም ንኽግንዘቡን ሕጋዊ ኣሰራርሓ ኣጠናኺሮም ንክቕፅሉን ጉሉህ ሚና የበርክት እዩ።
ስራ አስከያጁ እንደገለጹት አሁን ግን ደንበኞች የብድር ትያቄያቸውን በቀጥታ ለቅርንጫፎች የሚያቀርቡበትና ምላሽ የሚያገኙበት አሰራር መመቻቸቱን አመልክቷል	እቲ መካየዲ ስራሕ ከምዝገለፆ ሕጂ ግን ደንበኛታት ሕቶታት ልቓሕ ብቐጥታ ንጨንፈራት ዘቕርቡሉን መልሲ ዝረኽብሉን ኣሰራርሓ ከምዝተመቻቸወ ኣመልኪቶም
የኢትዮጵያ ጠቅላይ ሚኒስቴር ወደ ህንድ የነበራቸውን በረራ ሰረዙ	ቀዳማይ ሚኒስተር ኢትዮጵያ ናብ ህንዲ ዝነበሮም በረራ ሰሪዞም
አለምአቀፍ የሴቶች ቀን ዛሬ ተከብሮ ይውላል	ዓለምለኻ መዓልቲ ደቂኣነስትዮ ለሚዓንቲ ተኽቢሩ ከውዕል እዩ
በቅርንጫፍ ጽህፈት ቤቱ የማህበራዊ መሰረተ ልማት ፕሮጀክት ኦፊሰር አቶ ተሾመ ዘገዬ ትናንት እንደገለጹት ትምህርት ቤቶቹ አገልግሎት መስጠት የጀመሩት ቀደም ሲል የትምህርት አገልግሎት ባልተደረሰባቸው የደቡብ ወሎ ፣ ኦሮሚያ ፣ ወግ በሚገኙ የገጠር ቀበሌዎች ውስጥ ነው።	ኣብቲ ቤት ፅሕፈት ቅርንጭፍ ማሕበራዊ መሰረት ልምዓት ፕሮጀክት ኦፊሰር ኣይተ ተሾመ ዘገየ ትማሊ ከምዝገለፅዎ እተን ቤት ትምህርቲ ግልጋሎት ምሃብ ዝጀመራ ቅድም ክብል ግልጋሎት ትምህርቲ ዘይበፀሐን ደቡብ ወሎ፣ ኦሮምያ፣ ወግ ኣብ ዝርከባ ቀበሌታት ገጠር ውሽጢ እዩ።
ባህላዊና ዘመናዊ የሙዚቃ ዝግጅት ፣ የታሪካዊ ቦታዎች ጉብኝት ፕሮግራም እና ሌሎችን ማካተቱን ከአስተባባሪዎቹ ገለጻ ለመረዳት ተችሏል።	ባህላውን ዘመናውን ምድላው ሙዚቃ ፣ ታሪኻዊ ቦታታት ምርኣይ ፕሮግራምን ካልኦትን ከምዘማለአ ካብኦም መታሓባበርቲ ገለፃ ንምርዳእ ተኻኢሉ።
በጥቃቅንና ኣነስተኛ የተደራጁ ወጣቶች ተጨማሪ የመንግስት ብድር ጠየቁ	ብደቀቕትን ነኣሽተይን ዝተጎጀሉ መናእሰያት ተሳወኺ ልቓሕ መንግስቲ ጠይቕም

Annex III: Sample Parallel Corpus for Testing

Amharicsentences	Tigrigna sentences
እሷ አስተማሪ ነች	ንሳ መምህር እያ
ስብሰባው አዳራሽ ላንድ ነው	እቲ ስብሰባ ኣዳራሽ ሓደ እዩ
እሱ ከሀና ፍቅር ይዞታል	ንሱ ምስ ሃና ፈቅሪ ሒዝዎ
እነሆ ዛሬ ከምድር ፊት አሳደድኸኝ ከፊትህም እሰወራለሁ በምድርም ላይ ኰብላይና ተቅበዝባዥ እሆናለሁ የሚያገኘኝም ሁሉ ይገድለኛል።	እንሆ ሎሚ ኻብ ገጽ ምድሪ አሳጉኺኒ ኻብ ገጽካ ኽሕባእ እየ ኣብ ምድሪ ኰብላልን ቀባሕባሒይን ክኸውን እየ ዚረኽበኒ ዘበለ ቕሉ ኪቕትለኒ እዩ።
በንግስተ ሳባ ትምህርት ቤት ዋና ዳይሬክቶር የሆኑት አቶ ዳኒኤል ተማሪዎቻቸው ለግብረ ገብነት ስርዓት ተገዢና ጎበዞች እንደሆኑ ነግሮውናል	ዋና ዳይሬክቶር ኣብ ቤት ትምህርቲ ንግስተ ሳባ ዝቑኑ ኣይተ ዳኒኤል ተማሃሮኦም ንስነ ምግባራዊ ስርዓት ተገዛእትን ንፉዓትን ከምዝቑኑ ነጊሮምና
የትግራይ ክልል ምክር ቤት ጉባኤ 10 ነባርና 12 አዲስ የያዘ ካብኔ ሾመ	ጉባኤ ቤት ምኽሪ ክልል ትግራይ 10 ነባራትን 12 ሓደሽትን ዝሓዘ ካብኔ ሰይሙ
የአዲስ አበባ ከተማ አስተዳደር ጽሕፈት ቤት ባወጣው መግለጫ ፡ በዚህ ዓመት ለ10ሺ ወጣቶች ስራ ዕድል እፈጥራለሁ ብለዋል።	ቤት ጽሕፈት ኣመሓዳሪ ከተማ ኣዲስ ኣበባ ኣብ ዘውጾ መግለጺ ፡ ኣብዚ ዓመት ን10 ሽሕ መናእሰያት ዕድል ስራሕ ክፈጥርየ ኢሉ።
እሷ የእሱ ሚስት ነች	ንሳ ሰበይቱ እያ
እኛ እየተዝናናን ነበር	ንሕና ንህደስ ነይርና
ብርጭቆው ሙሉ ነው	እቲ ብርጭቆ ሙሉእ እዩ
እሷ ፊልም እያየች ነው	ንሳ ፊልሚ ትርኢ ኣላ
ቤት ምክር ተወካዮች የህዝብ ኣስቸኳይ ኣዋጁን በሙሉ ድምፅ ኣፅድቆታል።	ቤት ምኽሪ ተወከልቲ ህዝቢ ነቲ ህፁፅ ኣዋጅ ብሙልእ ድምፂ ኣፅዲቕሞ።
ዋና ከተማ የኬንያ ከኢትዮጲያ ስንት ኪሎ ሜትር ነች ?	ዋና ከተማ ከንያ ካብ ኢትዮጵያ ክንደይ ኪሎ ሜትር እያ ?
መዝሙር ህዝብ የኢትዮጵያ እየተዘመረ ነው።	መዝሙር ህዝቢ ኢትዮጵያ ይዝመር ኣሎ።
እሱ ተጫዋች ኳስ ቅርጫት ነበር።	ንሱ ተፃወታይ ኩዕሶ ስኪዒት ነይሩ።
እግዚአብሔርም ያደረገውን ሁሉ አየ ፥ እነሆም እጅግ መልካም ነበረ። ማታም ሆነ ጥዋትም ሆነ ፥ ስድስተኛ ቀን።	ኣምላኽ ከኣ ዝገበሮ ዘበለ ቕሉ ረኣየ ፡ እንሆ ብዙሕ ጽቡቕ ኰነ። ምሸት ኰነ ብጊሓትውን ኰነ ፡ ሳድሰይቲ መዓልቲ።
የሁለተኛውም ወንዝ ስም ግዮን ነው። እርሱም ለሁሉም የኢትዮጵያን ምድር ይከበባል።	ስም እቲ ኻልኣይ ርባ ድማ ጊሆን እዩ። ንሱ ንኹላ ምድሪ ኩሽ ይዞራ።
ሴቲቱም ለእባቡ አለችው። በገነት ካለው ከዛፍ ፍሬ እንበላለን።	እታ ሰበይቲ ነቲ ተመን በለቶ፥ ኣብ ገነት ካብ ዘሎ ፍረ ኦም ንበልዕ ኢና።
እግዚአብሔርም ቃየንን አለው። ለምን ተናደድህ ? ለምንስ ፊትህ ጠቆረ ?	እግዚኣብሄር ከኣ ንቃየል ፥ ስለምንታይ ኰፈኻ ፤ ስለምንታይከ ገፅካ ፀለመ ፤
እሷ እረፍት መውሰድ ፈልጋ ነበር።	ንሳ ዕረፍቲ ክትወስድ ደልያ ነይራ።
እሱ ስብሰባ ላይ ነው።	ንሱ ኣብ ኣኼባ እዩ።
የሕዝብ ተወካዮች ምክር ቤትና ጠቅላይ ሚኒስትሩ የሚሰጡትን ሌሎች ተግባሮች ያከናውናል።	ቤት መኽሪ ተወከልቲ ህዝብን ቀዳማይ ሚኒስትርን ዝህብዎ ካልኦት ተግባራት የሳልጥ።
መንግሥት በሀይማኖት ጉዳይ ጣልቃ አይገባም።	መንግስቲ ኢዱ ኣብ ጉዳይ ሃይማኖት ኣየእትውን።
ሕዝብ በመረጠው ተወካይ ላይ እምነት ባጣ ጊዜ ከቦታው ለማንሳት ይችላል። ዝርዝሩ በሕግ ይወሰናል።	ህዝቢ ኣብቲ ዝመረፆ ተወካሊ እምነት ኣብ ዝሰኣነሉ ግዜ ካብ ቦትኡ ከልዕሎ የኽእል። ዝርዝር ብሕጊ ይውሰን።
እግዚአብሔርም ብርሃኑን ቀን ብሎ ጠራው ፥ ጨለማውንም ሌሊት አለው። ማታም ሆነ ጥዋትም ሆነ ፥ አንድ ቀን።	ኣምላኽ ነቲ ብርሃን መዓልቲ ኣውጽኣሉ። ነቲ ጸልማት ከኣ ለይቲ ኣውጽኣሉ። ምሸት ኰነ ብጊሓትውን ኰነ ፡ ሓንቲ መዓልቲ።

Annex IV: Sample Language Model for Tigrigna Language

\data

ngram 1=3174

ngram 2=6762

ngram 3=7951

\1-grams:

-3.883075 <unk> 0

0 <s> -0.58753216

-1.3083168 </s> 0

-3.2891848 ቀዳማይ -0.10480435

-3.2891848 ሚኒስትር-0.04465039

-3.6113107 ርእሰ -0.04465039

-3.6113107 ምምሕዳር -0.04465039

-3.4365904 እዛ -0.04465039

-3.0188699 ሃገር -0.07882098

-2.459479 ፣ -0.05747675

-3.7693648 ኣካቢ -0.04465039

-2.5188622 ቤት -0.40169382

-3.2891848 ምኽሪ -0.116574585

-3.7693648 ሚኒስትራትን -0.04465039

-3.1793323 ዋና -0.07882098

-3.7693648 ኣዛዚ -0.04465039

-3.7693648 ሓይሊታት -0.04465039

-3.6113107 ኩናት -0.04465039

-1.6061265 እዩ -1.0378

-1.2941806 ። -1.1075573

-3.6113107 ዝለዓለ -0.04465039

-3.1793323	ስልጣን	-0.04465039
-3.7693648	ኣፈፃምነት	-0.04465039
-2.7340527	መንግስቲ	-0.08117873
-3.6113107	ፌዴራል	-0.04465039
-2.4981525	ኢትዮጵያ	-0.07318602
-3.7693648	ዝተውሃበ	-0.04465039
-3.6113107	ንቐዳማይ	-0.04465039
-3.6113107	ሚኒስትርን	-0.04465039
-3.091732	ንቤት	-0.47568068
-3.7693648	ሚኒስተራትን	-0.04465039
-3.2891848	ሕገ	-0.23523489
-3.7693648	ፌዴራላዊ	-0.04465039
-3.6113107	ዲሞክራሲያዊ	-0.04465039
-3.7693648	ሪፕብሊክ	-0.04465039
-2.110127	ካብ	-0.07727568
-3.7693648	ነሓስ	-0.04465039
-3.7693648	15	-0.04465039
-3.7693648	1987	-0.04465039
-3.6113107	ዓ	-0.25108692
-3.7693648	/	-0.04465039
-3.7693648	ም	-0.04465039
-3.7693648	ኣትሒዙ	-0.04465039

-3.7693648 ብዕሊ -0.04465039

-1.9353467 ኣብ -0.14744355

-2.700563 ስራሕ -0.044650365

Annex V: Sample Parallel Sentences that Do Not Require Reordering Rules

Amharic sentence	Tigrigna sentence
ብዙ ህፃናት ያለረዳት ያድጋሉ።	ብዙሓት ህፃናት ብዘይሓጋዚ ይዓብዩ እዮም ።
ብዙ ሰዎች ያለትዳር ይኖራሉ።	ብዙሓት ሰባት ብዘይሓዳር ይነብሩ እዮም።
እሷ ለእሱ ስሟን ነገረችው።	ንሳ ንዕኡ ሽማ ነጊራቶ።
እሱ ለእሷ ስጦታ ሰጣት።	ንሱ ንዓኣ ህያብ ሂብዋ።
እሷ ስጦታውን ወዳዋለች።	ንሳ ነቲ ህያብ ፈትያቶ ኣላ።
እሷ እሱን እየፈለገችው ነው።	ንሳ ንዕኡ ትደልዮ ኣላ።
እሱ እሷን እየፈለጋት ነው።	ንሱ ንዓኣ ይደልያ ኣሎ።
እሱ እሷን ማመን አልነበረበትም።	ንሱ ንዓኣ ክኣምን ኣይነበሮን።
እሷ እሱን አጃጃለችው።	ንሳ ንዕኡ ኣዕሽያቶ
እሷ ስህተት ነች።	ንሳ ተጋግያ እያ
እሱ ስለእሷ ተሳስቶ ነበር።	ንሱ ብዛዕባኣ ተጋግዩ ነይሩ።
ሰዎች ያወራሉ።	ሰባት ይዛረቡ እዮም።
ሰዎች በጣም ያወራሉ።	ሰባት ብጣዕሚ ይዛረቡ እዮም።
እነሱ ማውራት አቆሙ።	ንሳቶም ምዝራብ ጠጠው ኣቢሎም።
እሷ ወደ ክፍሉ ገባች።	ንሳ ናብቲ ክፍሊ ኣትያ።
እነሱ ከቢሮ ወጡ።	ንሳቶም ካብ ቢሮ ወፂአም።
እነሱ ስራ እየፈለጉ ነው።	ንሳቶም ስራሕ ይደልዩ ኣለው።
እሱ ምሳውን እየበላ አስቴር ከስራ መጣች	ንሱ ምስሉ እንዳበለዐ ኣስቴር ካብ ስራሕ መፂኣ

Annex VI: Transliteration from Fidel to Latine Characters

First order		Second order		Third order		Fourth order		Fifth order		Sixth order		Seventh order	
ሀ	he	ሁ	Hu	ሂ	hi	ሃ	ha	ሄ	hE	ህ	h	ሆ	ho
ለ	le	ሉ	Lu	ሊ	li	ላ	La	ሌ	lE	ል	l	ሎ	lo
ሐ	He	ሑ	Hu	ሒ	Hi	ሓ	Ha	ሔ	HE	ሕ	H	ሖ	Ho
መ	me	ሙ	Mu	ሚ	mi	ማ	ma	ሜ	mE	ም	m	ሞ	mo
ሠ	^se	ሡ	^su	ሢ	^si	ሣ	^sa	ሤ	^sE	ሥ	^s	ሦ	^so
ረ	re	ሩ	Ru	ሪ	ri	ራ	Ra	ሬ	rE	ር	r	ሮ	ro
ሰ	se	ሱ	Su	ሲ	si	ሳ	Sa	ሴ	sE	ስ	s	ሶ	so
ሸ	xe	ሹ	Xu	ሺ	xi	ሻ	Xa	ሼ	xE	ሽ	x	ሾ	xo
ቀ	qe	ቁ	Qu	ቂ	qi	ቃ	Qa	ቄ	qE	ቅ	q	ቆ	qo
ቐ	Qe	ቑ	Qu	ቒ	Qi	ቓ	Qa	ቔ	QE	ቕ	Q	ቖ	Qo
በ	be	ቡ	Bu	ቢ	bi	ባ	Ba	ቤ	bE	ብ	b	ቦ	bo
ቨ	ve	ቩ	Vu	ቪ	vi	ቫ	Va	ቬ	vE	ቭ	v	ቮ	vo
ተ	te	ቱ	Tu	ቲ	ti	ታ	Ta	ቴ	tE	ት	t	ቶ	to
ቸ	ce	ቹ	Cu	ቺ	ci	ቻ	Ca	ቼ	cE	ች	c	ቾ	co
ኀ	Ke	ኁ	Ku	ኂ	Ki	ኋ	Ka	ኌ	KE	ኍ	K	ኆ	Ko
ነ	^he	ኑ	^hu	ኒ	^hi	ኗ	^ha	ኔ	^hE	ን	^h	ኖ	^ho
ኘ	ne	ኙ	Nu	ኚ	ni	ኛ	Na	ኜ	nE	ኝ	n	ኞ	no
አ	^e	ኡ	^u	ኢ	^i	ኣ	^a	ኤ	^E	እ	^	ኦ	^o
ከ	Ke	ኩ	Ku	ኪ	ki	ካ	Ka	ኬ	kE	ክ	k	ኮ	ko
ወ	We	ዉ	Wu	ዊ	wi	ዋ	Wa	ዌ	wE	ው	w	ዎ	wo
ዐ	'e	ዑ	'u	ዒ	'i	ዓ	'a	ዔ	'E	ዕ	'	ዖ	'o
ዘ	Ze	ዙ	Zu	ዚ	zi	ዛ	Za	ዜ	zE	ዝ	z	ዞ	zo
ዠ	Ze	ዡ	Zu	ዢ	Zi	ዣ	Za	ዤ	ZE	ዥ	Z	ዦ	Zo
የ	Ye	ዩ	Yu	ዪ	yi	ያ	Ya	ዬ	yE	ይ	y	ዮ	yo
ደ	De	ዱ	Du	ዲ	di	ዳ	Da	ዴ	dE	ድ	d	ዶ	do
ጀ	Je	ጁ	Ju	ጂ	ji	ጃ	Ja	ጄ	jE	ጅ	j	ጆ	jo
ገ	Ge	ጉ	Gu	ጊ	gi	ጋ	Ga	ጌ	gE	ግ	g	ጎ	go
ጠ	Te	ጡ	Tu	ጢ	Ti	ጣ	Ta	ጤ	TE	ጥ	T	ጦ	To
ጨ	Ce	ጩ	Cu	ጪ	Ci	ጫ	Ca	ጬ	CE	ጭ	C	ጮ	Co
ጰ	Pe	ጱ	Pu	ጲ	Pi	ጳ	Pa	ጴ	PE	ጵ	P	ጶ	Po
ፀ	^Se	ፁ	^Su	ፂ	^Si	ፃ	^Sa	ፄ	^SE	ፅ	^S	ፆ	^So
ጸ	Se	ጹ	Su	ጺ	Si	ጻ	Sa	ጼ	SE	ጽ	S	ጾ	So
ፈ	Fe	ፉ	Fu	ፊ	fi	ፋ	Fa	ፌ	fE	ፍ	f	ፎ	fo
ፐ	Pe	ፑ	Pu	ፒ	pi	ፓ	Pa	ፔ	pE	ፕ	p	ፖ	po
ሏ	lWa	ሟ	mWa	ሧ	^sWa	ሷ	sWa	ቧ	xWa	ቈ	qWe	ቍ	qWi
ቋ	qWa	ቌ	qWE	ቍ	qW	ቖ	Qwa	቗	QWE	ቘ	QW	ኳ	vWa
ቿ	cWa	ኈ	^hWi	ኋ	^hWa	ኌ	^hWE	ኍ	^hW	ጧ	TWa	ፏ	pWa
ኩ	kWi	ኳ	kWa	ኴ	kWE	ኵ	kW	ኹ	Kwi	ኻ	KWa	ኼ	KWE
ዀ	KW	ዟ	zWa	ዷ	dWa	ጇ	jWa	ጐ	gWe	ጒ	gWi	ጓ	gWa
ጔ	gWE	ጕ	gW	ኍ	HWa	ሯ	rWa	ቈ	Qwe	ቧ	bWa	ቷ	tWa
ጐ	^hWe	ኗ	nWa	ኟ	NWa	ኰ	kWe	ኽ	Kwe	ኧ	ZWa	ጯ	CWa
ጀ	PWa	ጿ	Swa	ፏ	fWa	።	.	፣	,	፤	;	፥	:

Declaration

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

Declared by:

Name: Akubazgi Gebremariam K/Mariam _____

Signature: _____

Date: October 7, 2017 _____

Confirmed by advisor:

Name: Yaregal Assabie (PhD) _____

Signature: _____

Date: October 7, 2017 _____