



**ADDIS ABABA UNIVERSITY**  
**COLLEGE OF NATURAL SCIENCES**

**Trilingual Sentiment Analysis on Social Media**

**Mebrahtu Tadesse G/Medhin**

**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**

**March, 2018**

Addis Ababa University  
College of Natural Sciences

Mebrahtu Tadesse G/Medhin

Advisor: Yaregal Assabie (PhD)

This is to certify that the thesis prepared by *Mebrahtu Tadesse*, titled: *Trilingual Sentiment Analysis on Social Media* 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:

| <u>Name</u> | <u>Signature</u> | <u>Date</u> |
|-------------|------------------|-------------|
|-------------|------------------|-------------|

Advisor: Yaregal Assabie (PhD)

Examiner: \_\_\_\_\_

Examiner: \_\_\_\_\_

# Abstract

The connectivity and availability of social media platforms in the world allows people to express sentiments and share experiences easily by providing sentiments. Analyzing sentiment data from social media helps to understand customers feeling, make decision and design strategies. The social media sentiment data can contain multiple languages and mixed words of languages. However, it is difficult to analyze the sentiments of the data that contain multiple languages and mixed languages of words.

This research work is, therefore, aimed to design sentiment analysis on social media data to analyze trilingual (i.e., English, Amharic and Tigrigna) sentiment sentences based on lexicon approach. The trilingual sentiment analysis system consists of seven main components: preprocessor, language identifier, morphological analyzer, sentence constructor using root words, sentiment word detector, sentiment word polarity weight determiner and sentiment classifier.

The preprocessor cleans the data, tokenizes sentiment sentence into lists of words and performs normalization. The language identifier identifies the language of the texts. The morphological analyzer analyzes morphology of sentiment text words in order to handle the morphological complexity of the languages and to detect sentiment words in the sentence. Sentiment word weight determiner determines the polarity weight of the sentiment words based on the sentiment lexicons. The sentiment classifier classifies sentiment sentences into positive, negative and neutral polarity classes using the polarity weight values of the sentiment words in the sentiment sentences.

The prototype of the system was developed to test and evaluate the functionality of the system. To test and evaluate the system, 564 sentiment sentences were collected from Facebook and YouTube. To measure the accuracy of the system in sentiment classification precision, recall and F-measure evaluation metrics were employed, and an average precision 87.49%, recall 84.78% and F-measure 85.99% were obtained.

**Key words:** Social Media, Sentiment Analysis, Trilingual Sentiment Analysis, Sentiment Words, Multilingual, Sentiment Polarity.

## Dedication

I would like to dedicate this paper to my family.

## Acknowledgments

First and for most, I would like to express my heartfelt thanks to the Almighty God, for giving me the strength, determination, endurance and wisdom to bring this thesis to completion. And also, I would like to express my heartfelt and deepest sense of gratitude to my advisor Dr. Yaregal Assabie for his constructive comments, suggestions, guidance, inspiring and enlightening ideas. Starting from shaping and reshaping the title of this thesis, his support and encouragements were high and without all these efforts, this work wouldn't have been a success at all. Next, I would like to thank Mr. Yirgalem G., Mr. Alemthehay T. and Mr. Redae A. for their help on linguistic parts and valuable comments and suggestions on this study.

I am also grateful to my family, especially my mother Letu Berhe and my father Tadesse G/Medhin for their love and endless support and encouragement throughout my academic life. Using this opportunity, I would like to express my deepest gratitude to my wife Aynadis Dimtsu. Thank you for your love, care and patience and moral encouragement she has given me throughout my study.

Last, but not least, I would like to thank my friends for sharing me their knowledge and experience, and provided me with constructive comments and suggestions in bringing this thesis to an end.

# Table of Contents

|  |     |
|--|-----|
| List of Figures.....                                       | iv  |
| List of Tables .....                                       | v   |
| List of Algorithms.....                                    | vi  |
| Abbreviations and Acronyms .....                           | vii |
| Chapter One: Introduction .....                            | 1   |
| 1.1 Background .....                                       | 1   |
| 1.2 Motivation.....  | 2   |
| 1.3 Statement of the Problem.....                          | 3   |
| 1.4 Objectives.....  | 4   |
| 1.5 Methods.....   | 4   |
| 1.6 Scope and Limitations.....                             | 5   |
| 1.7 Application of the Results.....                        | 6   |
| 1.8 Organization of the Thesis .....                       | 6   |
| Chapter Two: Literature Review .....                       | 7   |
| 2.1 Introduction.....                                      | 7   |
| 2.2 Sentiment Analysis .....                               | 7   |
| 2.3 Levels of Sentiment Analysis .....                     | 9   |
| 2.4 Sentiment Analysis in Multiple Languages .....         | 10  |
| 2.5 Sentiment Analysis and Classification Techniques ..... | 11  |
| 2.5.1 Lexicon Based Approach.....                          | 11  |
| 2.5.2 Machine Learning Approach .....                      | 12  |
| 2.5.3 Hybrid Approach .....                                | 13  |
| 2.6 Sentiment Lexicon Generation.....                      | 13  |
| 2.7 Linguistic Issues in Sentiment Analysis .....          | 15  |
| 2.7.1 Amharic Language.....                                | 16  |
| 2.7.2 Tigrigna Language.....                               | 18  |

|                |   |    |
|----------------|---|----|
| 2.7.3          | English Language .....                                  | 21 |
| Chapter Three: | Related Work .....                                      | 23 |
| 3.1            | Introduction .....                                      | 23 |
| 3.2            | Sentiment Analysis Using Lexicon Based Approach .....   | 23 |
| 3.2.1          | Sentiment Analysis for Amharic Language.....            | 23 |
| 3.2.2          | Sentiment Analysis for English Language .....           | 24 |
| 3.2.3          | Sentiment Analysis for Chinese Language.....            | 25 |
| 3.2.4          | Sentiment Analysis on Multi-Linguals.....               | 25 |
| 3.3            | Sentiment Analysis Using Machine Learning Approach..... | 26 |
| 3.3.1          | Sentiment Analysis for Amharic Language.....            | 26 |
| 3.3.2          | Bilingual Sentiment Analysis .....                      | 27 |
| 3.3.3          | Multilingual Sentiment Analysis .....                   | 28 |
| 3.4            | Summary .....   | 28 |
| Chapter Four:  | Design of Trilingual Sentiment Analysis System .....    | 30 |
| 4.1            | Introduction .....                                      | 30 |
| 4.2            | Architecture of the System.....                         | 30 |
| 4.2.1          | Preprocessing Component .....                           | 32 |
| 4.2.2          | Word-Level Language Identification .....                | 33 |
| 4.2.3          | Morphological Analysis.....                             | 35 |
| 4.2.4          | Sentence Construction Using Root Words .....            | 37 |
| 4.2.5          | Trilingual Sentiment Lexicon .....                      | 38 |
| 4.2.6          | Sentiment Word Detection .....                          | 38 |
| 4.2.7          | Sentiment Word Polarity Weight Determination .....      | 42 |
| 4.2.8          | Sentiment Classification .....                          | 49 |
| Chapter Five:  | Experiment.....   | 51 |
| 5.1            | Introduction .....                                      | 51 |
| 5.2            | The Development Environment and Tools .....             | 51 |

|  |                                 |    |
|--|---------------------------------|----|
| 5.3  | Prototype of the System .....   | 52 |
| 5.4  | Data Collection .....           | 54 |
| 5.5  | Results and Discussion.....     | 55 |
| 5.5.1  | Evaluation Procedures .....     | 55 |
| 5.5.2  | Evaluation Results .....        | 56 |
| 5.5.3  | Test Results.....               | 57 |
| 5.5.4  | Discussion.....                 | 57 |
| Chapter Six: Conclusion and Future Work.....                     |                                 | 59 |
| 6.1  | Conclusion .....                | 59 |
| 6.2  | Contribution of the Thesis..... | 60 |
| 6.3  | Future Work .....               | 61 |
| References.....  |                                 | 62 |
| Appendix-1: Approval Letter for Linguistic Expertise.....        |                                 | 69 |
| Appendix-2: Sample English Positive Sentiment Words .....        |                                 | 70 |
| Appendix-3: Sample English Negative Sentiment Words .....        |                                 | 72 |
| Appendix-4: Sample Amharic Positive Sentiment Words.....         |                                 | 74 |
| Appendix-5: Sample Amharic Negative Sentiment Words .....        |                                 | 76 |
| Appendix-6: Sample Tigrigna Positive Sentiment Words .....       |                                 | 78 |
| Appendix-7: Sample Tigrigna Negative Sentiment Words .....       |                                 | 80 |
| Appendix-8: Inflection of Tigrigna Sentiment Words .....         |                                 | 82 |
| Appendix-9: Inflection of Amharic Sentiment Words.....           |                                 | 83 |
| Appendix-10: Sample Data of Trilingual Sentiment Sentences ..... |                                 | 84 |
| Appendix-11: Ethiopic Scripts/Fidel Vs Latin Scripts .....       |                                 | 85 |

## List of Figures

|   |    |
|---|----|
| Figure 4-1: General Architecture of Trilingual Sentiment Analysis System .....          | 31 |
| Figure 5-1: Trilingual Sentiment Analysis System UI .....                               | 52 |
| Figure 5-2: Analyzed Sentiment Information Result.....                                  | 53 |
| Figure 5-3: Result of Input Sentiment Text with its Sentiment and Polarity Values ..... | 53 |

## List of Tables

|   |    |
|---|----|
| Table 2-1: Sentiment Wordlists .....                              | 14 |
| Table 2-2: Amharic Inflection of Adjectives .....                 | 17 |
| Table 2-3: Tigrigna Inflection of Adjectives .....                | 19 |
| Table 2-4: English Inflection of Adjectives .....                 | 22 |
| Table 4-1: Negation Words in Sentiment Sentences .....            | 40 |
| Table 4-2: Intensifier Terms in Sentiment Sentences .....         | 41 |
| Table 4-3: Polarity Categories of Sentiment Sentences .....       | 50 |
| Table 5-1: Sentiment Words in Trilingual Sentiment Lexicons ..... | 54 |
| Table 5-2: Evaluation Result of the Experiment .....              | 57 |

## List of Algorithms

|  |    |
|--|----|
| Algorithm 4.1: Algorithm for Trilingual Word-Level Language Identification .....                 | 34 |
| Algorithm 4.2: Algorithm for Morphological Analysis of Amharic Words .....                       | 36 |
| Algorithm 4.3: Algorithm for Morphological Analysis of Tigrigna Words.....                       | 37 |
| Algorithm 4.4: Algorithm for Sentiment Word Detection and Polarity Weight<br>Determination ..... | 46 |
| Algorithm 4.5: Algorithm to Determine the Polarity Weight of Sentiment Sentence.....             | 49 |

## Abbreviations and Acronyms

|      |  |
|------|--|
| LIGA | Language Identification Graph Algorithm  |
| MCC  | Matthew Correlation Coefficient          |
| MPQA | Multi-Perspective Question and Answering |
| NLTK | Natural Language Toolkit                 |
| PMI  | Pointwise Mutual Information             |
| RBEM | Rule Based Emission Model                |
| SVM  | Support Vector Machine                   |
| UTF  | Unicode Transformation Format            |

# Chapter One: Introduction

## 1.1 Background

Social media is a group of Internet based applications that builds ideological and technological foundation of web 2.0 and allows for creation and exchange of user generated contents [1]. In recent years, social media shows tremendous growth all over the world [2]. Various social media like Facebook, LinkedIn, Twitter and YouTube have millions of users which connect these users in a single village of the world. Through social media, users can easily share their experience and influence each other by providing sentiments [2].

People's idea and thought towards some specific objects or a view, judgment and appraisal formed in a mind for a particular thing in the real world is an opinion which is the key influence of human behavior in daily activities [3]. When we need to make a decision, we want to know others opinions. In real world governments, political parties, manufactories and business sectors often want to find public and customers opinions on the governance system, policy structure/strategy, products and services [5, 28]. To determine the public opinions, in these days, due to the excessive availability of social media services, governments, political parties and individuals can easily use and access the public opinions/contents in the social media in addition to the survey and opinion polls for the purpose of decision making [3, 6]. Different peoples can easily participate and express their opinion and feedback on topics posted on social media. However, it is difficult to find structured and summarized information from the huge amount of opinion data in social media [15]. To solve this problem, automated sentiment analysis system is needed.

Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to systematically identify and extract subjective information from source data [4]. Sentiment analysis is widely applied for analyzing reviews like movie reviews, consumer product reviews and other applications on social media for the purpose of determining the attitude and sentiments of the writer/opinion holder with respect to some topics or the overall contextual polarity of the document. It determines the subjectivity whether the expression is subjective or objective and uses automated tools to detect the subjective expressions of opinion holders. In natural language, subjective expressions are the aspects of the language that are used to express opinions,

feelings, evaluations and sentiments. Subjective expression adds an additional level of granularity by further classifying the subjective texts into positive, negative and neutral polarities. In general, sentiment analysis is widely used in various fields to analyze people's sentiments on various application domains can be movies, products, politics, hotels and others specific topics with their attributes. According to [3,5] there are three levels of sentiment analysis.

- i. Document level sentiment analysis
- ii. Sentence level sentiment analysis
- iii. Aspect level sentiment analysis

In over the world including Ethiopia, different social media are available which include: Facebook, YouTube and Twitter. Ethiopian social media users can use different languages like Amharic, Tigrigna, Afaan Oromoo and English to express their sentiments on topics posted on social media. Hence, the sentiments can contain multiple languages because of the availability of various languages. Therefore, it is difficult to automatically analyze the sentiments containing multiple languages [15]. Regarding this, this study deals with trilingual sentiment analysis to simplify the difficulties of sentiments written in English, Amharic, and Tigrigna languages.

## 1.2 Motivation

Sentiment analysis involves text analysis, process natural languages to determine and evaluate the subjective information from source data [4]. In real world business companies, governments, political campaigns, politicians and other organizations post their product, service, agenda, strategy and other issues on social media platforms in order to get opinions and feedbacks from consumers, public and citizens. Social media users can also express their feeling, attitude on the posted topics through various natural languages. But, there are challenges to analyze the sentiments which are written in multiple languages on the posted topics. To the best of our knowledge, there is no research done in trilingual sentiment analysis on social media for English, Amharic and Tigrigna languages. Due to the absence of trilingual sentiment analysis on social media thousands of comments, feedbacks and opinions written in English, Amharic and Tigrigna texts cannot be analyzed. In addition to, business companies, governments and other organizations cannot know consumers, public and citizen's opinions and feedbacks

on their product, service, agenda and strategies when the sentiment texts are written in English, Amharic and Tigrigna languages. This motivated us to develop trilingual sentiment analysis system that helps to analyze sentiment texts written in English, Amharic and Tigrigna languages on social media platforms.

### 1.3 Statement of the Problem

The growth of Internet technologies and the availability of smartphones provide the benefit for peoples to actively express their sentiments on governmental policy, governmental service and political campaigns issue through social media platforms. Similarly, they can also provide sentiments, feedbacks and feelings on business companies, product companies and public services through social media platforms. Social media users can write and express their sentiments, feedbacks and feelings by using multiple natural languages like Amharic, Tigrigna, Afaan Oromoo and English languages on topics posted on social media. It is difficult to understand, analyze and make decisions from sentiments, feedbacks and feelings written or expressed in multiple languages in specific topics of domains. Therefore, to solve the problem researchers studied sentiment analysis in different domains using various languages. For instance, various researchers studied monolingual sentiment analysis using English [6,7], Amharic [8,9,10] and Arabic[11,12,13] languages, and also multilingual sentiment analysis [14,15]. Although monolingual sentiment analysis system cannot analyze the sentiments written in different languages, they can miss the sentiment information written in other languages and unable to extend to other languages because they are developed only for single languages.

Sentiment analysis developed in multilingual also cannot analyze sentiment texts written in English, Amharic and Tigrigna languages since the languages have their own linguistic behaviors, and there is a need to study and identify the sentiment patterns of these languages. In addition to this, the trilingual sentiment analysis needs to have its own special system that handles and integrates these three languages in a single system.

## 1.4 Objectives

### General Objective

The general objective of this research is to design a trilingual sentiment analysis system on social media using English, Amharic and Tigrigna languages.

### Specific Objectives

This study is specifically aimed to:

- Review literatures on sentiment analysis, techniques of sentiment analysis and the linguistic behaviors of English, Amharic and Tigrigna languages
- Identify the language of input texts
- Analyze the morphology of English, Amharic and Tigrigna sentiment words
- Design the general architecture of trilingual sentiment analysis system
- Build trilingual sentiment lexicons for English, Amharic and Tigrigna sentiment words
- Develop the prototype for trilingual sentiment analysis system
- Collect trilingual sentiment sentences for performance evaluation
- Test and evaluate the system

## 1.5 Methods

In order to accomplish the objective of the research, the following methods and procedures will be used.

### *i. Literature Review*

The literature review deals with reviewing the related literature with the current research work, sentiment analysis. Sentiment analysis studies in social media using different languages and levels of sentiment analysis will be reviewed in details. In order to get better knowledge on sentiment analysis and polarity classifications various techniques will be reviewed in details. Finally, the linguistic behaviors of English, Amharic and Tigrigna languages will be reviewed; sentiment pattern of English, Amharic and Tigrigna languages will be collected and checked by linguistic expertises.

## ***ii. Data Collection***

Sentiment sentences will be collected from Facebook and YouTube on domain of movie, music, politics and hotels for this research work. The sentiment sentences will be collected manually from the topics posted on Facebook and YouTube. In addition, preprocessing activities will be done on the collected sentiment sentences to enable language identification, morphological analysis and sentiment term detections.

## ***iii. Analyze and Design***

Language of input texts will be identified. To identify the language of input texts, English, Amharic and Tigrigna wordlists will be built in the trilingual lexicons. English, Amharic and Tigrigna sentiment words will be analyzed their morphologies and stored the analyzed sentiment words in the trilingual sentiment lexicons. At last, the general architecture of the trilingual sentiment analysis system will be designed and modeled each components of the system.

## ***iv. Prototype Development***

Different open source tools and programming languages will be selected to implement the system prototype. Python programming language will be used primarily to implement the system because of the following reasons:

- It is suitable for natural language processing
- Simple and powerful programming language with excellent features and extensive libraries
- It contains a lot of packages that help us to do code reusability

## ***v. Evaluation***

Experiment will be conducted to test the functionality of trilingual sentiment analysis system. The performance of the system will be evaluated in terms of precision, recall and F-measure.

## **1.6 Scope and Limitations**

This work deals with designing a general architecture for trilingual sentiment analysis system and developing a working system based on the design. This research work considers only the inputted sentiment texts grammatically correct, analyzes and classifies

sentiment texts written in English, Amharic and Tigrigna languages and focus only sentiment analysis at sentence level. Hence, grammatically incorrect texts, slang of words, sentiments expressed through an image/picture, an audio, video and other emotional symbols are not the concerns of this study. In addition to, words or texts that have indirect or hidden meanings such as an idiom/ቅኔዖዊ አነጋገር/ብቅኔ ዝዘረብ and ambiguous words are not incorporated in this research work.

## 1.7 Application of the Results

Considering the growth of the Internet and the availability of social media services, the application of the results of this research work can have its own benefits as follows.

- Sentiment analysis helps for decision making on activities like to design new policy and strategies, public service distributions, new product launches
- Sentiment analysis helps for decision making in various domains such as politics, movies, music and hotels
- It also provides structured and summarized sentiment information about public opinion that use mixed languages of English, Amharic and Tigrigna

## 1.8 Organization of the Thesis

The remaining part of this thesis is organized as follows. Chapter Two presents the review of related literature which includes: sentiment analysis, approaches of sentiment analysis and the linguistic behaviors of English, Amharic and Tigrigna languages. Chapter Three presents related works done on monolingual and multilingual sentiment analysis system and related areas. The Fourth Chapter deals with the design of a trilingual sentiment analysis system on social media. The Fifth Chapter presents the experimental results of trilingual sentiment analysis system. Finally, conclusion, recommendation and future works are presented in Chapter Six.

# Chapter Two: Literature Review

## 2.1 Introduction

This Chapter explores techniques for discovering the sentiment patterns, the approaches used for analyzing and classifying sentiment texts written in particular topics either in monolingual or multi-languages on social media. The Second Section, 2.2, explains natural language processing and its applications in the field of computer science and the concepts of sentiment analysis. Section 2.3 describes the levels of sentiment analysis. The general concept of multilingual sentiment analysis is presented in Section 2.4. The sentiment analysis and classification approaches are discussed in Section 2.5. In Section 2.6, the methods helpful for sentiment lexicon generation are discussed. Finally, Section 2.7 describes the linguistic issues in sentiment analysis and linguistic behaviors of English, Amharic and Tigrigna languages.

## 2.2 Sentiment Analysis

NLP is the field of computer science, artificial intelligence and computational linguistics concerned with the interaction between computer and human languages, specifically concerned with programming computers to fruitfully process huge natural language corpora [16,17]. It is a way for computers to analyze, understand and derive meaning from human languages in a smart and useful way. Using NLP, developers can organize and structure knowledge to perform tasks such as sentiment analysis, automatic text summarization, machine translation, named entity recognition, speech recognition and automated question answering.

Social media is one of the most significant information exchange technologies of the 21<sup>st</sup> century [18]. Social media users use social media platforms to post messages, photos, videos and others about their daily activities. Various social media channels such as Facebook, Twitter, YouTube and LinkedIn provide convenient and efficient ways of communication and sharing of information publically. Due to this reason, the users of Internet increase from time to time and the high availability of user texts on the Internet, extraction of useful information automatically from abundant documents and unstructured texts obtains an interest to many researchers in many fields, in particular in NLP community [19].

Sentiment analysis also known as opinion mining [3,5] was firstly proposed in the early 2000 and has become an active research area gradually in natural language processing. Moreover, various practical applications of sentiment analysis such as market prediction [20,21], election forecasting [22,23], national relationship analysis [24] and risk detection in banking system [25] draw extensive attentions from industrial communities. On the other hand, the growth of social media, electronic commerce and online review sites such as Twitter, Amazon and Yelp provide a large amount of corpora which are crucial resources for academic and industrial researches. An interest from both academic and industrial research promotes the development of sentiment analysis.

The term sentiment analysis first appeared in [26] to extract sentiments associated with polarities of positive or negative for specific subject from a document, instead of classifying the whole document into positive or negative. Sentiment analysis is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes. There are also many other names for sentiment analysis though these do have slightly different tasks. For example, sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review analysis, review mining, etc. However, all these names are under the umbrella of sentiment analysis. Sentiment analysis mainly focuses on opinions which express or imply positive or negative sentiments.

The sentiment consists of three basic components [3,5,26]: opinion holder, object and opinion.

- **Opinion holder** is a person or an organization that expresses the opinion on a particular object, event, topic or services. It is also called an opinion source [27].
- **Object** is an entity that can be a product, a person, an event, an organization or a topic which an opinion has been expressed on that object by opinion holders.
- **An opinion** is the view, attitude, opinion, sentiment or appraisals on a particular object from an opinion holder. The opinion expressed in the object can be positive, negative or neutral sentiments or can be a numeric rating scores that expressing the strength or intensity of the sentiments. During sentiment analysis the positive, negative and neutral are called the semantic orientations (polarities).

An opinion can be expressed on an object in two ways [28], either in direct opinion or comparative opinion. Direct opinion: opinions are directly expressed on the object or features of the object. For example, the bed is good in Axum hotel. Whereas, comparative opinion expresses a relation of similarities or differences between two or more objects and/or object preferences of the opinion holder on some of the shared features of the objects. The comparative opinion is usually expressed using the comparative or superlative form of an adjective or adverbs, although not always [26,29,30].

## 2.3 Levels of Sentiment Analysis

In general, sentiment analysis has been investigated mainly at three levels [5,19,28]. In this Section, these three levels of sentiment analysis are discussed in details.

- i. **Document level:** the task at this level is to classify whether the whole opinion document expresses a positive or negative sentiments [31]. For example, in a given movie review or product review, the system determines whether the review expresses an overall positive or negative about the movie or the product. This task is commonly known as document level sentiment classification, because it considers only the whole document as the basic information unit. This level of analysis assumes that each document expresses an opinion on a single entity (e.g., a single movie or single product). Hence, it is not applicable to documents which evaluate or compare multiple entities.
- ii. **Sentences level:** the task at this level is to determine whether each sentence expresses positive, negative or neutral sentiments. In this level, besides sentiment classification, subjectivity classification is another problem [19]. Since, subjectivity classification aims to determine whether the sentence is subjective or objective. A subjectivity sentence expresses a subjective view and determines either the sentence is positive or negative sentiment, whereas, an objective sentence expresses a factual information from a sentence.
- iii. **Aspect level:** both document level and sentence level sentiment analysis provide useful information in many applications. However, they don't provide the necessarily detail needed information for some other advanced applications. Aspect level sentiment analysis is proposed to discover detail opinionated texts in the specific objects. It performs a finer-grained sentiment analysis. Aspect level has several

naming variations including entity (feature, attribute or fine-grained) [6,32] level sentiment analysis. Aspect level directly looks at the opinion itself instead of looking on the language constructs (documents, paragraph, sentences, clause or phrases). It is based on the concept that an opinion consists of a sentiment either positive or negative and a target of opinion. During aspect level sentiment analysis, identifying an opinion target helps to understand the sentiment analysis problem better. For example, “although the service is not great, I still love this restaurant”, this sentence contains positive polarity, but we cannot say that this sentence is entirely positive. In fact, the sentence is positive about the restaurant, but negative to the service. In different applications, opinion targets are represented by entities and/or their different aspects because the aim of this level of sentiment analysis is to discover the sentiments on the specific object entities and/or their aspects. Finally, it provides the structured summary of sentiments about entities and their aspects from unstructured texts.

## 2.4 Sentiment Analysis in Multiple Languages

Social media users can express their comments or opinions on the same topic or event via social media platforms; as a result comments or opinions can contain multiple languages on those specific topics/events. Previously, many researchers conducted their studies on the sentiment analysis of texts written in single languages such as, in English [6], Arabic [33] and Spanish [34], however, with the grow of the Internet and high availability of social media platforms around the world, social media users’ comments contain different languages. So, to analyze data in different languages, multilingual sentiment analysis techniques have been developed [14], to analyze sentiments, frameworks and tools for different languages are being built. Sentiment analysis on multiple languages can be performed using various techniques, such as: by transferring knowledge from resource rich language to resource poor languages [35], or by using machine translation system between resource rich and resource poor languages [36], or directly analyze multiple languages without translation or using another resources [37]. In multi-lingual sentiment analysis system, a language that contains extensive resources and tools available can be used to build sentiment analysis for other languages which have few resources or tools. This can be done in one of the following ways:

- By translating test sentences in resource poor language into resource rich language and analyze the sentiment sentences using the resource rich language.
- By translating a resource rich language train corpus into the resource poor language and build a corpus-based sentiment analysis in the resource poor language.
- By translating a sentiment or lexicon terms in the resource rich language to the resource poor language and build lexicon-based sentiment analysis in the resource poor language.

## 2.5 Sentiment Analysis and Classification Techniques

Classification is used to assign a sentiment sentence into a particular category (also referred as a class) or categories based on their sentiment polarities. Sentiment analysis system can classify sentiments in the sentence using different approaches. This Section discusses the sentiment analysis and classification techniques.

### 2.5.1 Lexicon Based Approach

Lexicon based approach involves calculating the sentiment polarity of a sentence from the semantic orientation of words or phrases in the sentence, and uses sentiment of lexicons, collection of known and precompiled sentiment terms with their polarity values [38,39]. The system classifies the sentiments into specific categories of the sentiment classes which are positive, negative or neutral based on the positive or negative sentiment terms occurred in the sentence. This can be done using rule based classifier methods, and if there are positive words only or more number of positive words than negative words in the sentence, the semantic orientation of the sentence is classified into positive. If there are negative words only or more negative words than positive words in the sentence, the semantic orientation is classified into negative. If there are equal numbers of positive and negative sentiment terms or there are no sentiments terms in the sentence, the semantic orientation of the sentence will be neutral and categorized in neutral class. In this approach, overstatement and understatement words are considered (i.e., overstatement words increase the semantic orientations and understatement decrease the semantic orientations of the sentiment words in the sentence).

In this approach, sentiment words or phrases can be easily added to a lexicon and tag as positive or negative terms with their semantic orientation values; for example, in [6], small list of seed adjectives tagged with positive and negative labels manually and grows these lists using WordNet. It is easily understood and modified by humans, viable and practical approach to sentiment analysis without needs of training. The performance of this technique depends on the effectiveness of the lexicon sentiment terms. This method requires less resource relative to other techniques (e.g., machine learning) and suitable for languages that have lack of resources and no any annotated corpora. It is possible to add extra sentiment terms for scaling up from different sources in order to strengthen the efficiency of the lexicon. In this research work, lexicon based approach is chosen based on the linguistics behavior, data availability, difficulty, effectiveness and extendibility.

### 2.5.2 Machine Learning Approach

Machine learning is the sub-field of computer science that gives a computer to learn without being explicitly programmed [40]. Machine learning approach performs sentiment classifications based on training algorithms, the classification is on a set of selected features for specific mission and test on another sets whether it is able to detect the right feature and give the right classifications. In sentiment classification, machine learning algorithm uses the supervised and unsupervised sentiment classification models. The supervised sentiment classification model uses large number of labeled training datasets, while unsupervised sentiment classification model uses when it is difficult to find the labeled training datasets. Supervised sentiment classification model can classify sentiment sentence based on the algorithms like Naïve Bayes, Maximum Entropy and SVM classifiers [31].

*Naïve Bayes classifier* is a simple probabilistic model based on Bayes theorem [41] along with a strong independence assumption. It computes the posterior probability of class based on the distribution of words in the document. It uses Bayes' theorem to predict the probability that a given feature set belongs to a particular label.

$$p(\text{label}|\text{features}) = \frac{p(\text{label}) * p(\text{features}|\text{label})}{p(\text{features})} \text{-----} (1)$$

Where,  $p(\text{label})$  is the prior probability of a label or the likelihood that a random feature set the label,  $p(\text{features}|\text{label})$  is the prior probability that a given feature set is being

classified as a label,  $p(\text{features})$  is the prior probability that a given feature set is occurred.

**Maximum Entropy classifier** is a probabilistic classifier which belongs to the class of exponential models. The models that fit the training examples are computed, where each feature corresponds to a constraint on the model. The model with the maximum entropy over all models that satisfies these constraints is selected for classification. Maximizing the entropy ensures that no biases in the classification system [42].

**SVM classifier** performs by constructing a hyper plane with maximal Euclidean distance to the closet training examples. This can be seen as the distance between the separating hyper plane and two parallel hyper planes at each side, representing the boundary of the examples of one class in the feature space. It is assumed that the best generation of classifier is obtained when the distance is maximal. If the data is linearly not separable, a hyper plane will be chosen that splits the data with the least error possible. SVM yields best accuracy in sentiment classification [31].

### 2.5.3 Hybrid Approach

Both lexicon based and machine learning based approaches are combined to benefit from their synergy effect that rises to hybrid approach. Researchers have proved that this combination gives improved performance in sentiment classification [43]. In hybrid approach, sentiment lexicons are used as seed resources and to detect sentiment polarities and the results from the lexicon based method are used to train machine learning algorithms. Then, sentiment sentences are analyzed using machine learning classifiers based on the knowledge acquired from the training and the lexicon resources.

## 2.6 Sentiment Lexicon Generation

Words or phrases that convey positive or negative sentiments are vital for sentiment analysis and classification system. Positive sentiment words are used to express some desired states or qualities while negative words are used to express some undesired states or qualities. Table 2-1 shows an example of sentiment wordlists.

Table 2-1: *Sentiment Wordlists*

| Language | Positive sentiment words            | Negative sentiment words         |
|----------|-------------------------------------|----------------------------------|
| English  | beautiful, good, wonderful, amazing | bad, poor, awful , terrible      |
| Amharic  | ቆንጅ/qonjo, ብልህ/blh, ጥሩ/Tru          | መጥፎ/metfo, ኃላቀር/halaqer, ሞኝ/ moN |
| Tigrigna | ጸቡቕ/Sbuq, ለባም/lebam, ወሐላ/weHale     | ሕማቕ/HmaQ, ሃሳስ/hasas, በሰሮ/bsero   |

Overall, positive and negative sentiment words are called sentiment lexicons. The sentiment lexicons can be generated using three different approaches [5]: manual approach, dictionary based approach and corpus based approach.

- i. **Manual approach** involves generating of sentiment bearing word lists manually from various resources and label their sentiment polarities (i.e., positive or negative) and strength of polarity values. This approach is an effective for resource poor languages and provides best performance for sentiment analysis system under specific domains.
- ii. **Dictionary based approach** involves using a dictionary which contains synonyms and antonyms of words. In this method, it uses a few seed of sentiment words to bootstrap based on the synonym and antonym structure of the dictionary. Specifically, this method works as first a small set of sentiment words (seeds) with known positive and negative semantic orientation is collected manually, then the algorithm grows these set of words by searching in WordNet [44], thesaurus [45,46] and other online available dictionaries for their synonyms and antonyms. The seed lists will be added with the newly found words. The process iteratively keeps on adding the words until no more new words are found. At last manual inspection is used to remove errors in the lists.
- iii. **Corpus based approach** helps to solve the problem of finding sentiment words with context specific orientations. This method depends on syntactic patterns or patterns that occur together along with a seed list of sentiment words to find other sentiment words in a large corpus. The corpus based approach can be performed using statistical method and sematic method [47].

**Statistical method:** this method helps to find the co-occurrence pattern or seed sentiment words using statistical techniques. The polarity of the word is determined based on the occurrence frequency of the word in the large annotated corpus of texts. If the word occurs more frequently among positive texts, then its polarity is positive. If the word occurs frequently among negative texts, then its polarity is negative. But, if it has equal occurrence frequency, then it is a neutral word. If two words appear together frequently within the same context, then there is high probability that they have same polarity. In this case, the polarity of unknown word can be determined by calculating the relative frequencies of co-occurrence with other words. This can be done using PMI [7], the semantic orientation of a given phrase is calculated by comparing its similarity to a positive word ('excellent') and its similarity negative word ('poor').

**Semantic method:** this technique gives sentiment values directly and uses different principles to compute the similarities between words. In this case, the principle gives similar sentiment values to semantically close words. These semantically close words can be obtained by getting the list of sentiment words and iteratively expanding the initial set with synonyms and antonyms and then determine the sentiment polarity for unknown word by the relative count of positive and negative synonym of the word [48].

## 2.7 Linguistic Issues in Sentiment Analysis

Language is the ability to acquire and use complex system of communication [49]. It is the human system of communication that uses arbitrary signals, such as voice, sound, gesture and/or written symbols that enable humans to express their feeling, sentiment, thought, idea and experience. Natural languages like English, Amharic and Tigrigna are used for communication purposes, whereas these languages have their own linguistic behaviors. The linguistic behaviors include: phonology, morphology, syntax, semantics and pragmatics.

In linguistics, morphology involved in word formation, deals with words, their internal structure and how they are formed with the help of morphemes. One fundamental computational task for a language analysis of its morphology is to derive the root and grammatical properties of the word based on its internal structures. The analysis of word

structure performs using the process of inflection and derivation. In general, morphological analysis plays the vital role in sentiment analysis applications.

**Inflection** is a morphological variation that never changes part of speech or the meaning of the stem word upon which the morphemes are attached, whereas, it changes the grammatical functions of the word [50,51,52]. It also indicates the syntactic or semantic relations between different words in a sentence.

**Derivation** deals with adding of morphemes in the stem word to generate new words, that changes the meaning, part speech and lexical function of the given stem words [50,51,52].

In this Subsection, the linguistic behaviors of Amharic, Tigrigna and English languages are discussed particularly on their morphological characteristics.

### 2.7.1 Amharic Language

Amharic is a Semitic family language which is mainly spoken in Ethiopia and it is the second most spoken Semitic families next to Arabic in the world [53]. Amharic is an official working language of the federal government in Ethiopia and several regional states. It has its own writing script known as Fidel “ፊደል“ having 34 consonants (base characters) and other six characters represented for each base characters derived from the combinations of vowel and consonant of base characters. Amharic alphabets are conveniently written in tabular format of seven columns from left to right side where the first column represents the base character and the others represent the derived characters that are derived from their vocal sounds. The vowels of the alphabet in Amharic language are not encoded explicitly, but appear as modifiers in the base character. For example, the following characters show consonants and vowels modify the base character ሙ(mä) as “ሙ(mu), ሚ(mi), ማ(ma), ሜ(mE), ሞ(m), ም(mo)”. In addition to this, there are about two scores of labialized characters like” ሚ (m<sup>w</sup>a), ላ (l<sup>w</sup>a), ፋ (f<sup>w</sup>a), ፕ (p<sup>w</sup>a), ቁ (k<sup>w</sup>e)” in Amharic language.

Like other Semitic languages, Amharic is one of the most morphologically rich languages and it exhibits a root-pattern morphological phenomenon [54]. Root is the set of consonant (also called radical) which has the basic lexical meaning [54]. The pattern consists of set of vowels which are inserted among the consonant of the root to form a

stem. Amharic verb stem consists of a “root + vowels + template” merger. For example, the root verb sbr + ee + CVCVC leads to a stem seber (‘breaking’). For non-concatenate morphological features, Amharic uses various affixes to create inflectional and derivational morpheme. Affixation can be prefix, infix and suffix.

Amharic language is highly inflectional in various features and the given Amharic word can occur in different forms. The highly inflected Amharic word classes are nouns, adjectives and verbs.

**Nouns:** Amharic nouns are inflected for number, gender, person and cases [50] and the result of the inflected word is with affixes. The inflected noun helps to express pluralism, gender and possession. Morphemes like -ኦች, -ኦት, -ኦት, -ኦ are used to inflect Amharic nouns by attaching to the stem words. For Example, adding the suffixes -ኦች, -ኦት and -ኦት to the words ቤት, መምህር and ህጻን generates ቤቶች, መምህራን and ህጻናት respectively.

**Adjective:** Amharic adjectives similar affixation process like Amharic nouns can be marked for number, gender, definiteness and cases [50]; and results of the inflected word with affixes. In Amharic, morphemes that used to inflect Amharic adjective words are -ኦ/ኦ, -ኦት/ኦት and -ኦች/ኦች. Table 2-2 shows an example of inflected Amharic adjective words.

Table 2-2: Amharic Inflection of Adjectives

| No. | Adjectives | Male | Female | Pluralism |
|-----|------------|------|--------|-----------|
| 1   | ሞኝ         | ሞኞ   | ሞኚት    | ሞኞች       |
| 2   | ብልጥ        | ብልጦ  | ብልጢት   | ብልጦች      |
| 3   | ቂል         | ቂሎ   | ቂሊት    | ቂሎች       |

**Verbs:** Amharic verbs are inflected for any combination of person, gender, number, case, tense, aspect and mood [50,55]. As a result, from a single Amharic verb thousands of Amharic verbs (in surface forms) can be generated. In Amharic, the verbs are found in different forms such as perfective, imperfective, gerundive, jussive and imperative by using affixes [55]. The morphological inflection of the perfective verbs have suffixes such as -ኦ/a, -ኦ/k, -ኦ/ku, -ኦ/h, -ኦ/u, -ኦቹ/uc, -ኦ/x, -ኦ/n and serves for the expression of

person, gender and number to the perfect verb stem. For example, the perfect verb stem ሰበር/seber (breaking) generates the inflected words such as ሰበር, ሰበርከ, ሰበርኩ, ሰበርህ, ሰበራችሁ, ሰበርሽ, ሰበርኝ (seber, seberk, seberku, seberh, seberachu, seberx and sebern). We can also add prefix like የ/ye, በ/be, ከ/ke, እንደ/ind/ in the perfective verbs regardless of the gender, number and person.

Amharic word classes can be derived from other Amharic word classes [50,55]. For example, nouns can be derived from adjectives, verb roots and stems (by inserting vowels between consonants) and nouns themselves. For instance, ሰውነት, ልጅነት, ቤትነት (sewnet, ljet, bEtnet) are derived from the nouns ሰው, ልጅ and ቤት (Sew, lj, bEt) respectively by adding the suffix -ነት/ net. In a similar way, verbs are also derived from different verbs by adding affixes in many ways. Amharic adjectives are also derived from nouns, verbal roots and compound words [50,55] by adding affixes. For instance, the adjectives ቁመኛ, ሀይለኛ and ጉልበተኛ (qimeNa, heyleNa and gulbeteNa) are derived from the nouns ቁም, ሀይል and ጉልበት (qim, heyl and gulbet) respectively by adding the suffix – አኛ/ oNa.

## 2.7.2 Tigrigna Language

Tigrigna is Semitic families of language like Amharic and Arabic languages, which is mainly spoken in Tigray region in Ethiopia and central Eritrea. It is an official language and medium of instruction for primary schools in Tigray region in Ethiopia. This language written with the version of Ge'ez (Ethiopic) scripts, originally developed from Ge'ez language which is appeared in writing during the 13<sup>th</sup> century [56]. Tigrigna has 34 base characters and other six characters represents for each base character derived from the combination of consonants and vowels. Like Amharic alphabets, Tigrigna alphabets are conveniently written in tabular format of seven columns from left to right side where the first column represents the base character and others represent their derived vocal sounds.

Tigrigna is morphologically rich language like Amharic, since both are Semitic families and originated from Ge'ez [57]. The root of verbs and most nouns in Tigrigna are characterized as a sequence of consonants which is called root/radicals like Amharic language. Roots are used for formation of other actual words by concatenating vowels and non-root consonants which goes with particular morphological category around the

root consonants. Hence, for non-concatenating morphological features, Tigrigna uses different affixes to create inflectional and derivational morphemes.

Like Amharic, Tigrigna language is highly inflectional language and the given root of the Tigrigna word can be found in different forms. The highly inflected Tigrigna word classes are nouns, adjectives and verbs and these word classes are discussed as follows.

**Nouns:** Tigrigna nouns are inflected for number, gender, person, definiteness and cases and the result is the inflected word with affixes [51]. The inflected words are used to express possession, plural and gender.

**Adjectives:** Tigrigna adjectives are marked for number, gender, person and degree and the results is the inflected word with affixes to the given adjectives [51]. Adjectives can be inflected by adding affixes like -ኣ/, -ት/t, -ቲ/ti, -ኣቶም/'atom, -ኣዊ/'awi, -ኣት/'at, -ኣዊት/'awit, ከ/k-, ዝ/z- etc. Table 2-3 shows the inflected Tigrigna adjective words.

Table 2-3: *Tigrigna Inflection of Adjectives*

| No. | Male | Female | Plural form | Comparative degree | Superlative degree |
|-----|------|--------|-------------|--------------------|--------------------|
| 1   | ዕጉስ  | ዕግስቲ   | ዕጉሳት        | ይዕገስ               | ዝተዓገሰ              |
| 2   | ቀይሕ  | ቀዖሕ    | ቀየሕቲ        | ይቀይሕ               | ዝቀይሖ               |
| 3   | ቀጢን  | ቀጣን    | ቀጠንቲ        | ይቀጥን               | ዝቀጠነ               |

**Verbs:** like Amharic, Tigrigna verbs also found in different forms, such as perfective, imperfective, gerundive, jussive and imperative by employing affixes. Tigrigna verbs inflected for number, gender, person, tense, mood and aspects, and the result is an inflected word with affixes to the verb stem. The morphological variation of the perfective verbs are generated by adding suffixes like ኣ/'a, ኣት/'at, ኩ/ku, ና/na, ኣቶም/'atom, ኣትን/'aten and ኣ/'a that indicates for person, gender and numbers to the perfect verb stem. For instance, ሰበረ, ሰበረት, ሰበርኩ, ሰበርና, ሰበራቶም, ሰበራትን and ሰበራ (sebere, seberet, seberku, sebirna, sebiratom, sebiraten and sebera). This perfect verb is formed from the stem verb ሰበር/seber which means ‘to break’. Imperfective verbs also formed by adding affixes on the verb stem and markers for gender, person and number. For example,

ትሰብር, ከሰብር, ከሰብራ, ከሰብሩ, ዝሰበረ, ዝሰበረት, ዝሰበረና, ይሰብሩ, ይሰብራ (tsebr, ksebr, ksebra, ksebru, zsebere, zseberet, zseberena, ysebru, ysebara), in this example, the stem verb ሰብር/sebr which means ‘let’s break’ uses both prefix and suffixes. The morphemes that attach in the stem verb such as ት/t, ከ/k, ዝ/z, and ይ/y are prefixes and ኣ/a, ኣ /e, ኣት/et, ኣና/ena and ኣ/u are suffixes.

The gerundive verb from inflected by adding suffixes at the end of the gerundive verb to indicate person, gender and number. For example, ሰሪሐ, ሰሪሐኻ, ሰሪሐኺ, ሰሪሐሉ, ሰሪሐና, ሰሪሐኹም, ሰሪሐኸን, ሰሪሐም and ሰሪሐን (seriHa, seriHXa, seriHXi, seriHu, seriHna, seriHXum, seriHXn, seriHom, seriHan), from this example, the stem verb ሰሪሐ/seriH varies its morphology and generate various verbs that markers of the person, gender and number. The suffixes ኣ/a, ኻ/Xa, ኺ/Xi, ኣ/u, ና/na, ኹም/Xum, ኸን/Xn, ም/m and ን/n attached with the inflected stem verb.

Jussive verbs sometimes called mood verbs are used to express a command for first and third persons, whereas imperative verbs used to express second person in singular and plural forms.

Tigrigna words are highly derivational from other Tigrigna word classes. Some of the highly derivational Tigrigna words classes are nouns, adjectives and verbs.

**Nouns:** Tigrigna nouns derived from other word classes by adding affixes and using compound words [51]. In the case of compound words, the new noun is constructed from two separate words. Example ቤት and ትምህርቲ, ቤት and ብልዒ provided ቤት ትምህርቲ which means school and ቤት ብልዒ which means restaurant or nouns constructed by adding affixes like -ነት and -ኛ. Example, ሰብ + -ነት gives ሰብነት, ዓፋር + -ኛ generates ዓፋርኛ. In addition, nouns can be derived from verbs and adjectives.

**Adjectives:** similar to nouns, Tigrigna adjectives can be derived from nouns, verbs and adjectives itself [51]. Adding morphemes like -ኣዊ/awi, -ዊ/wi, -ኣም/am and -ታይ/tay to nouns such as: ኢትዮጵያ/ityoPya, ሃገር/hager, ነገር/neger and መቐለ/meQele that generates the following new adjectives: ኢትዮጵያዊ/ityoPyawi, ሃገራዊ/hagerawi, ነገራም/negram and መቐለታይ/meQeletay respectively. Similarly, it can be constructed from the root verbs. For instance, the root verb ሕስር by infixing the vowel -ኡ-, it generates the adjective ሕሰር.

**Verbs:** unlike nouns and adjectives, Tigrigna verbs can be derived only from verbal roots and stems [50,51].

Example, ቅ-ብ-ር, ቅኣብኣር which provides ቀብር, and it can also be constructed from verbal stems by adding affixes like ተ and ኣ to the stem verb.

### 2.7.3 English Language

English language is a West Germanic language that was originated in England and now widely spoken in many countries around the world [58]. It is the third most common native language in the world, after Mandarin and Spanish [58]. In most countries in the world including Ethiopia, English is used as a foreign language. The English language contains twenty six letters in upper and lower cases, out of the twenty six letters there are five vowels and twenty one consonants.

Relative to Amharic and Tigrigna, English is not morphologically complex language. Hence, it uses different affixes to construct inflectional and derivational morphemes.

Unlike Amharic and Tigrigna, English is not highly inflected and has small number of inflectional affixes. Most inflected word classes in English are nouns, adjectives and verbs. These word classes are discussed as follows.

**Nouns:** English nouns are inflected for number, but not for case and gender [52,59], and the result is the inflected word with affixes to the noun. English nouns can have regular and irregular forms when they are changed into plural forms. The regular English inflections add –s morphemes to stem nouns, whereas the irregular inflections add like –en, -ices morphemes to stem nouns and the inflected word expresses pluralities.

**Adjectives:** English adjectives are inflected for degree, but not gender or case [52,59], and the result is an inflected adjective word with morphemes to the adjective. Most monosyllabic and disyllabic adjectives have synthetic comparatives in –er and superlatives in –est, whereas adjectives with two syllables and nearly all with three or more form analytic comparatives with more and superlatives with most. Table 2.4 shows inflected adjective words.

Table 2-4: *English Inflection of Adjectives*

| No. | Adjective | Comparative    | Superlative    |
|-----|-----------|----------------|----------------|
| 1   | Happy     | Happier        | Happiest       |
| 2   | Sad       | Sadder         | Saddest        |
| 3   | Beautiful | More beautiful | Most beautiful |

**Verbs:** English verbs inflected for tense and number [52,59], and the result is an inflected verb with morphemes to the verb. English verb stem is the base for other inflected words and may stand alone as an infinite, imperative or general present forms. Like adjectives, English verbs have regular or irregular forms. The regular verb contains three forms based on the stem: a form in –ing that functions a present participle and gerund, a form in –ed that functions as a past tense and past participle and present tense which uses –s for third person singular forms. Irregular verbs also have the stem based form in –ing and –s, but exhibit distinctive pattern in past tense. For example, the verb stem prove can inflected into proves, proved, proving and proven by adding the morphemes –es, -ed, -ing and –en in the stem verb prov.

English words are derivational from other English word classes [52], and some of the derivational English word classes are nouns, verbs and adjectives.

**Nouns:** English nouns can be derived from nouns, verbs and adjectives by adding morphemes in the root of the given words [52]. For example, the root word kill is a verb and when we added –er, it generates a noun word class killer. Various morphemes like -dom, -er, -let, -cy and -hood helps to generate new words from other word categories.

**Adjective:** English adjectives derived from nouns, verbs and adjective itself by adding morphemes in the stem/root of the nouns, verbs and adjectives itself [52]. For example, in the noun root system, when we add the suffix –atic, it generates systematic or adding the prefix un- and suffix –atic, it generates unsystematic. In addition, in the verb root count, when we add the morpheme –less, it generates countless. In general, morphemes like in-, un- and re- are prefixes and morphemes like –able, -ly, -en, -ful and –less are suffixes that helps to generate new adjective words. In the same way, English verbs also derived from nouns and adjectives by adding morphemes in the stem/root of the given words.

## Chapter Three: Related Work

### 3.1 Introduction

Sentiment analysis is one of the popular methods which have been used to analyze user generated texts on social media. Multi-language comments on same topic have more common on social media platforms; however, it is challenging to analyze the comments which have multiple languages in the particular topics. To solve this challenge, a number of researches have been studied so far. In this Chapter, various studies related to the current research work are reviewed and classified based on the approaches used for sentiment analysis in monolingual and multi-languages, and finally show the knowledge gaps in general and propose a solution that this research work clearly addressed.

### 3.2 Sentiment Analysis Using Lexicon Based Approach

#### 3.2.1 Sentiment Analysis for Amharic Language

Selama G/Meskel [8] has developed sentiment mining model for opinionated Amharic texts. In this research, to determine the sentiment of reviews, the author used term counting techniques. Based on this technique, it counts the number of sentiment words in the sentence by checking each word in the prepared lexicon, if the word appears in the lexicon, it considered as the word is sentiment word and assigns polarity weight values automatically, whereas it considered as non-sentiment word. The author developed general and specific Amharic sentiment lexicons which include sentiment terms and contextual valence shifter terms. Sentiment terms were tagged as '+' and '-' in the lexicon, which is '+' for positive and default value assigned +2 and '-' for negative and default value assigned -2. But, the final polarity weight value was calculated by considering the effects of context valence shifters like diminisher, intensifier and negation terms around the sentiment terms in the sentence and finally classified the sentiment sentence into positive, negative and neutral classes based on polarity weight values. For the purpose of experiment, the researcher collected 303 review sentences from two different domains and 955 opinion terms: 411 positive and 544 negative terms. However, the prepared opinion terms are not sufficient and inflected Amharic sentiment words are not handled in this research work.

Tulu Tilahun [9] developed feature level opinion mining model for Amharic texts. The objective of the study was to determine an opinion on features of the domains. In this study, the author first extracted features of the domain and then determined the opinions in the extracted features by employing some rules. An opinion word in the sentence was detected from Amharic general purpose opinion word lexicons that contain a total of 1001 sentiment words which are 578 negative and 423 positive words. The author collected 484 Amharic reviews manually from Hotel, University and Hospital for experimental activities. The effectiveness of the system was evaluated using precision, recall and F-measure metrics through two different experiments. From experiment one, the author got the result of an average precision of 95.2%, an average recall of 26.1% for feature extraction and an average precision of 78.1%, average recall of 66.8% for opinion word determination. From experiment two, an average precision of 79.8%, an average recall of 34% for feature extraction and average precision of 80%, an average recall of 92.7% for opinion word determination. The strength of this study was to determine an opinion on features of the domain in the review sentence. However, the author used only adjectives as sentiment words to determine the opinions of the review sentences, but sentiment words are not only adjectives, but also include adverbs, verbs and nouns. In addition to this, the sentiment words are not sufficient.

### 3.2.2 Sentiment Analysis for English Language

Kundi *et al.* [60] developed a framework for sentiment classification using lexicon based approach in social web sentiment analysis. In this study, the authors integrated various lexicons like general purpose opinion lexicon that contains 1967 positive and 4783 negative sentiment words, dadvar opinion lexicon which contains 136 positive and 109 negative sentiment words and dictionary resources to extract sentiment words from target texts automatically. Semantic orientation was calculated by summing up the semantic orientation of the words and phrases in the document or target texts. The authors have used 308,316 tweets of product dataset for experimental purposes and the experiment was done in two type sentiment classifications using confusion matrix method classifiers. The performance of the classifier was evaluated through precision, recall, F-score and Matthew Correlation Coefficient (MCC). At last, the authors achieved an accuracy of 92% in binary classification and 87% in multi-class classification although the framework can't analyze non-English tweet datasets.

### 3.2.3 Sentiment Analysis for Chinese Language

Sentiment analysis of Chinese reviews using bilingual knowledge and ensemble techniques were studied in [61]. The main aim of the study was to improve sentiment analysis of the target language (i.e., Chinese) by using English rich knowledge. In this study, Chinese reviews were translated into English reviews using machine translation services and then the Chinese reviews and English reviews were analyzed based on Chinese resources and English resources respectively. Then, the individual results were ensembled to obtain better results. To compute the semantic orientation value of the review sentences, the authors have used sentiment lexicons by collecting positive, negative, negation and intensifier sentiment terms for both Chinese and English languages. The experiment was conducted on 836 Chinese product reviews and evaluated the effectiveness of the approach using precision, recall and F-measure to measure the performance of the sentiment classification in positive and negative polarity classes and used an accuracy to evaluate the overall performance of the system. However, the polarity identification of the review sentences in this study was based on unsupervised ways and the classifier can predicts incorrect sentiment polarities, since there was lack of Chinese resources, the performance of the system was degraded because it depends on the machine translation performance between the target language and knowledge rich language and Chinese sentiment words can be missed during translation the Chinese reviews into English reviews.

### 3.2.4 Sentiment Analysis on Multi-Linguals

Volkova *et al.* [62] conducted a study to explore the sentiments in social media and create Twitter specific lexicons via bootstrapping sentiment bearing terms from multilingual Twitter streams. They used bootstrapping approach for subjectivity clue from Twitter data and evaluate their approach on English, Spanish and Russian Twitter streams. The authors used the MPQA lexicon to bootstrapping sentiment lexicons from large pool of unlabeled data using small amount of labeled data to guide the process. Terms that are strongly subjective in translation were used as a seed terms in the new language with term polarity projected from English lexicons. The approach is easily scalable, better able to handle informality and nature of social media, although it depends on the availability of bilingual dictionary for translating an English subjectivity lexicon and crowdsourcing for selecting seed words.

Denecke [35] investigated the method to automatically determine polarity of a sentence in multilingual frameworks. In this work, the method depends on English resources to determine sentiment of texts for other languages. For classifying sentence of the document regarding their sentiments, it comprises three processing pipeline steps: first determine the language of the document, translate non-English document into English and classifying sentiment of the document into sentiment class (i.e., positive or negative). During sentiment analysis of the document, the polarity of the sentences was determined based on SentiWordNet using machine learning algorithm, classification rule and lingPipe classifier algorithm. The performance of the method was evaluated in English and non-English documents using precision and recall evaluation metrics. The author has done the experiment using different corpora for training and testing the classifier algorithms. For training purpose, the author used movie reviews that collected from IMDB's archive `rec.arts.movies.reviews` which contains 1000 positive and 1000 negative reviews in English language, and for testing purpose the author collected 100 positive and 100 negative movie reviews in German language from `Amazon.de` webpages. After conducting the experiment, the author achieved an accuracy of 66% in polarity classification. Whereas, in this work negation terms in SentiWordNet classifier structure were not considered, translation of documents degrades the performance of the system and during translation key sentiment words of the target text can be missed.

### 3.3 Sentiment Analysis Using Machine Learning Approach

#### 3.3.1 Sentiment Analysis for Amharic Language

Multi-scale sentiment analysis model for Amharic online posts written in Ethiopic scripts were studied by Wondwossen Mulugeta [10] using supervised machine learning approach. The objective of the study was to determine multi-scale sentiment sentence based on the polarity weight value of sentiment words. To achieve the objective, the author has prepared sample corpus which contains 608 posts. The corpus was collected from social media sources such as Facebook, Twitter, DireTube and Ethiopian reporter websites. The author employed preprocessing activities in the corpus dataset before the actual sentiment classification. After the preprocessing activities were done, the corpus was manually annotated by giving polarity values and sentiment intensity scale values. The author adopted two scale schemes which are scale positive sentiment further as +1

and +2 for less and more positive respectively and negative sentiment polarities as -1 and -2 for less and more negative respectively, hence neutral sentences were annotated as 0. To distinct the polarity strength of Amharic sentiment words, the author limited into five polarity rank scales. Naïve Bayes algorithm was used to classify sentiment texts based on the trained on the entire corpus. The algorithm used n-gram features to acquire the knowledge from the training corpus, and then the algorithm classifies the sentiments of the test posts based on the acquired knowledge. Among the sample corpus, 486 posts were selected for training dataset and 122 posts were for testing dataset. The author achieved an accuracy of 43.6%, 44.3% and 39.3% for unigram, bigram and hybrid language models respectively. However, the accuracy of the system was very low since the training dataset was very insufficient and the tool used for morphological analysis was not effective.

### 3.3.2 Bilingual Sentiment Analysis

Yan *et al.* [37] developed a bilingual approach to conduct sentiment analysis on English and Chinese social media. In this approach, English and Chinese review comments were processed simultaneously as a stream of texts rather than separately to mine and extract social opinions from social media. To process the social media data, the authors performed: first collected social media data and saved in text document as samples, applied word segmentation and trim down words to extract features, and then select features that represent the data. After selected features, SVM and N-Gram classification models were used to analyze and classify the bilingual sentiments. For the purpose of experiment, the authors collected 2000 English comments (1000 positive and 1000 negative comments) and 2000 Chinese comments (1000 positive and 1000 negative comments) manually from social media sites. Out of the total datasets, 80% of the comments were used for training and 20% of the comments were used for testing their classification models. They used precision, recall and F-score performance measures for evaluating the effectiveness of their approach. Finally, the authors achieved an accuracy of 98.90% and 82.42% sentiment classification using SVM and N-Gram models respectively. However, in this study, the approach was developed for English and Chinese only, so it cannot extend to other languages and can't analyze mixed sentiment texts of the languages.

### 3.3.3 Multilingual Sentiment Analysis

Tromp [2] investigated multilingual sentiment analysis on social media. The objective of the study was to solve the problems of sentiment analysis on multilingual short texts typically present in various social media like Twitter, Facebook and Hyves. The author presented four-step approach to automatically analyze sentiments of the multilingual texts. The approach comprises four basic components: language identification, part of speech tagging, subjectivity detection and polarity detection. For language identification and polarity detection, the author presented new algorithms called LIGA and RBEM respectively. In this work, an extensive experimental study was presented to support and illustrate the benefit of the steps in the four-step approach and it allowed for quantifying the importance of having the output of the corresponding techniques at each steps. In this study, the author conducted the experiments in two main categories: comparative experiment and traditional survey experiment and evaluates the performance of the sentiment classification using accuracy, precision and recalls. However, this study couldn't analyze mixed language of texts; an automated sentiment analysis developed for multi-linguals cannot appropriately analyze the sentiments of trilingual texts written in English, Amharic and Tigrigna languages since these languages have their own linguistic characteristics and needs to study their sentiment patterns. This is the knowledge gap or problem that the current study addressed.

## 3.4 Summary

In this chapter, some of the previous studies related to the current work were discussed in details. These works are reviewed based on the techniques used for sentiment analysis and classification of the sentiment sentences. The techniques presented and discussed here are lexicon based and machine learning based approaches successfully used for sentiment analysis on monolingual and multi-linguals. However, monolingual sentiment analysis used either lexicon based or machine learning approaches cannot be extended to trilingual sentiment analysis. Similarly, the multilingual sentiment analysis that used either lexicon or machine learning approaches relies on machine translation system or bilingual dictionaries for the purpose of using resource rich language to other languages in order to eliminate the gaps on resources used for sentiment analysis. However, using machine translation system has an impact on sentiment analysis system performance and

sentiment terms of the target languages can't be appropriately translated. Since, using machine translation system or bilingual dictionary depends on the quality of the translation machine system or the dictionaries.

In general, machine learning approach depends on labeled corpus to train the model, but this is very costly whereas, lexicon based approaches determine sentiments based on sentiment lexicons which contain positive and negative sentiment terms. This is an effective approach and doesn't require training corpus and capable of detecting and classifying sentiments of the given sentiment sentences. So far, although there are studies conducted on the Amharic sentiment analysis, little attention is given on Tigrigna and Trilingual (i.e., using Amharic, Tigrigna and English) sentiment analysis.

# Chapter Four: Design of Trilingual Sentiment Analysis System

## 4.1 Introduction

The core objective of trilingual sentiment analysis system is to provide a tool that helps to analyze trilingual sentiment texts written in English, Amharic and Tigrigna languages. To achieve this objective, the architectural design of trilingual sentiment analysis system, the components of trilingual sentiment analysis system and the algorithms used for sentiment term detection and polarity weight determination in trilingual sentiment sentences are discussed in this Chapter.

## 4.2 Architecture of the System

The proposed trilingual sentiment analysis system has seven main components. These are preprocessor, language identifier, morphological analyzer, sentence constructor using root words, sentiment term detector, sentiment term polarity weight determiner and sentiment classifier. The preprocessor component contains two subcomponents: tokenization and normalization of texts. Trilingual sentiment lexicon is the backend component of trilingual sentiment analysis system that stores English, Amharic and Tigrigna sentiment words. Each components of the system with their function are discussed in detail in this Section. The general architecture of trilingual sentiment analysis system is shown in the Figure 4-1.

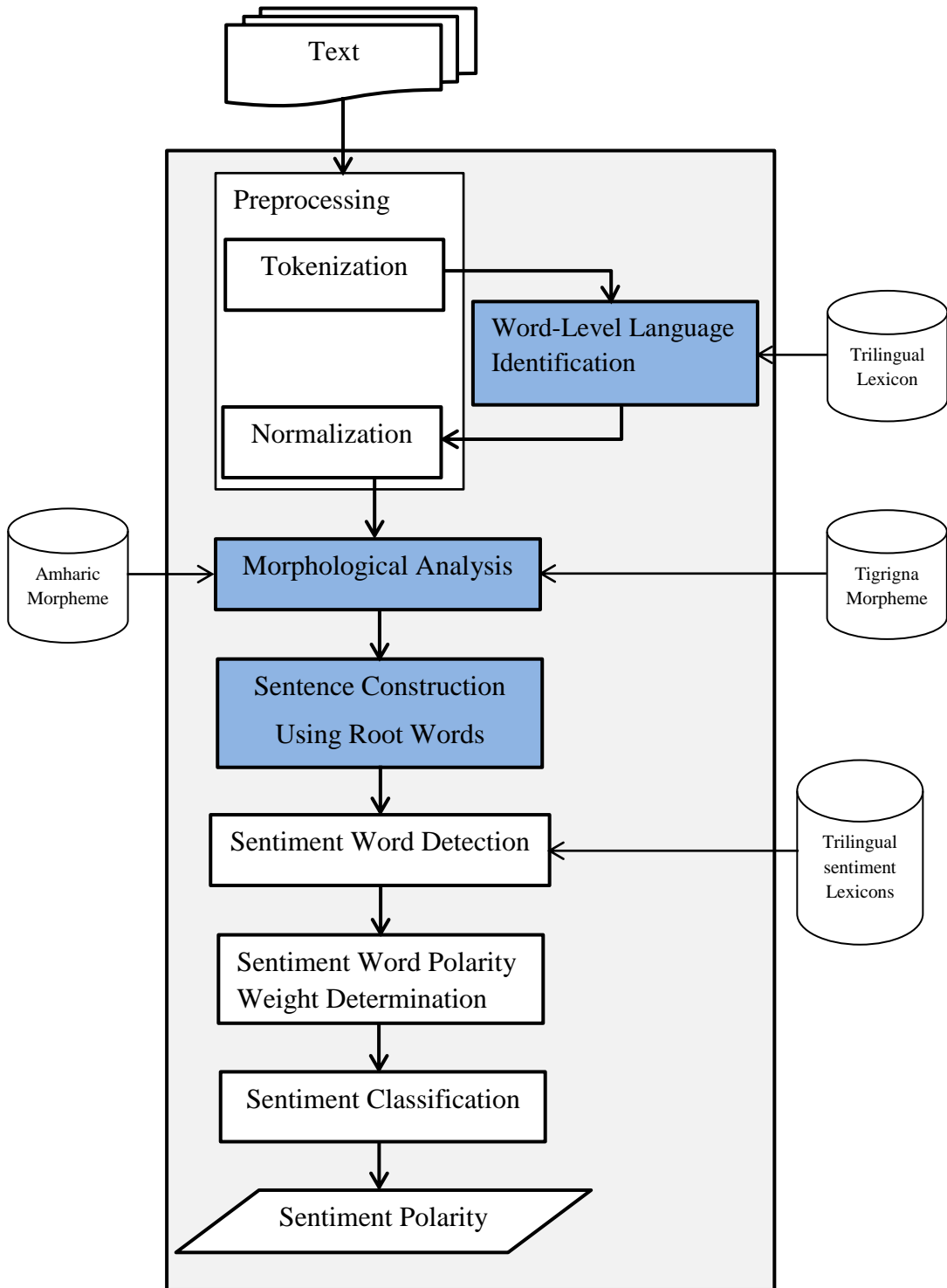


Figure 4-1: General Architecture of Trilingual Sentiment Analysis System

## 4.2.1 Preprocessing Component

The preprocessing component is used for the process of cleaning the input texts to further analysis. Social media texts usually contain a lot of unnecessary texts such as hashtags, HTML scripts, special characters, URL and others. The noisy and unnecessary texts are removed in tokenization component and prepare the texts into suitable formats. This component contains two subcomponents: tokenization and normalization.

### 4.2.1.1 *Tokenization*

Tokenization is the process of chopping the input texts into pieces of tokens. In tokenization process, input texts are tokenized into a sequence of tokens by detecting word boundaries from the written texts. In most languages, white space and punctuation marks are used as boundary markers. For example, English language has its own punctuation marks to demarcate words, sentences etc. into a stream of characters that include punctuation marks like period (.), comma (,) and semicolon (;). Amharic and Tigrigna languages also have their own punctuation marks to demarcate words, sentences etc. into a stream of characters that include punctuation marks like ‘ārati net’ibi’(::), ‘net’ela šerezi’(፣), ‘diribi šerezi’(፣) and question mark(?).

In our case, tokenization is the first step in the preprocessing component. In this step, input texts are tokenized into stream of characters using white spaces and punctuation which helps to convert into list of words. For example, when the sentence “the music is very nice” tokenized, the result will be ['the', 'music', 'is', 'very', 'nice']. In this research work, NLTK is used for tokenizing input texts into lists of words.

### 4.2.1.2 *Normalization*

Normalization component is used in trilingual sentiment analysis system for two purposes. First, it uses to convert Amharic and Tigrigna letters/’fidelat’ that have the same use and pronunciation, but different writing representation symbols to a common word into common representation. For instance, the word ‘clean’ in English can be represented in Amharic in many ways, such as ‘ገጹሀ’, ‘ገፁሀ’, ‘ገጹኅ’, ‘ገጹሐ’, ‘ገጹኸ’. These words differ only by the second and third character ‘ጹ’, ‘ፁ’, ‘ኅ’, ‘ሀ’, ‘ሐ’ and ‘ኸ’, but all have the same meaning and pronunciation. In Amharic language characters like ሀ, ሐ, ኅ and ኸ, ከ and ፀ, ሰ and ሠ, ጸ and ፀ represent the same consonant, have the same pronunciation and can be used interchangeably without any meaning differences. Similar

to Amharic, in Tigrigna also characters like ሀ and ኅ, ሰ and ሠ, ፀ and ጸ have the same pronunciation and meaning which can be used interchangeably without any meaning alteration. For example, the word ‘good’ in English can be written in Tigrigna in two different ways either ‘ሰቡቕ’ or ‘ጸቡቕ’. In these two words, the difference is in the first character only, but both have the same meaning and pronunciation. Therefore, in this research work, the system is designed to understand those words that have the same meaning, but different writing representation symbols by converting into common representation. In Amharic, the letters ሀ, ሐ, ኅ and ኘ converted into ሀ, ኦ and ፀ converted into ኦ, ሰ and ሠ converted into ሰ and ጸ and ፀ converted into ፀ; similarly in Tigrigna also the letters ሀ and ኅ converted into ሀ, ሰ and ሠ converted into ሰ, ፀ and ጸ converted into ፀ. Therefore in this research work, the letters ሀ, ኦ, ሰ and ፀ are taken as common representations for the letters that have the same meaning and pronunciation in Amharic and Tigrigna languages. The second task of normalization is translating Ethiopic script representation texts into Latin script representations in order to maintain data uniformity and consistency. Latin script representation is easy to represent Amharic and Tigrigna vowels and consonant characters and there are no words that have the same meaning and reading pronunciations with different writing spellings in Latin scripts.

#### 4.2.2 Word-Level Language Identification

Word-level language identification component is used to detect and identify the language of the texts. In the world, there are a number of multilingual speakers. Those multilingual speakers can use multiple languages in their daily activities and communications through social media platforms. However, the data from multilingual is difficult to identify and determine the content of its message. Language identification solves the problem in multilingual. The languages of the texts are detect at word-level because the texts can contain mixed language of words or social media users can write their sentiment using mixed language of words in their expressions. For example, “በዚህ ዘፈን ስኬት ከዚህ ስኬት ከዚህ ስኬት” this sentence contains mixed language of words containing English and Amharic words. In this case, the system needs to understand the language of each word for further processing of the sentence.

In this study, lexicon based approach is chosen for language identification due to its high accuracy in identification, its scalability, its detection of mixed language of words in a text and effectiveness for under explored languages (i.e., Amharic and Tigrigna). In this

study, trilingual lexicon is built for language identification that contains English, Amharic and Tigrigna lexicons independently, and stored at backend component of the system<sup>1</sup>. After the tokenization process is completed every  $n^2$  meaningful word of the texts are looked up in the prepared trilingual lexicon to detect their languages. Algorithm 4.1 shows language identification of input texts.

Input: sentiment sentence S

For every sentiment sentence  $s_i \in S$ ,  $i=1, 2, \dots, n$  where n is number of sentiment sentence in S

For each  $s_i$  tokenized into m valid words,  $\forall w_j \in s_i$ , where  $j=1, 2, 3, \dots, m$  where m is number of valid words in  $s_i$

For each  $w_j$  checked in trilingual lexicon

If  $w_j$  in English lexicon

Append  $w_j$  in temp1

Else if  $w_j$  in Amharic lexicon

Append  $w_j$  in temp2

Else if  $w_j$  in Tigrigna lexicon

Append  $w_j$  in temp3

Else

Jump  $w_j$  and consider  $w_j$  is unknown language

End if

End For

End For

Output: English words, Amharic words, Tigrigna words

**Algorithm 4.1: Algorithm for Trilingual Word-Level Language Identification**

---

<sup>1</sup> Trilingual lexicon is not the same as Trilingual sentiment lexicons

<sup>2</sup> n- is the number of meaningful words after the sentence is tokenized into  $n_i$  words, i-index term of the  $n^{\text{th}}$  word in the tokenized sentence

### 4.2.3 Morphological Analysis

The morphological analyzer component is used to analyze English, Amharic and Tigrigna words into their constituent morphemes. Morphological analysis is important especially for morphologically rich languages like Amharic and Tigrigna because it is practically difficult to store all possible words in a lexicon. In this study, the morphology of English, Amharic and Tigrigna words are analyzed independently.

#### 4.2.3.1 *English Morphological Analysis*

English is not morphologically complex language as compared to Amharic and Tigrigna languages. In this research work, to analyze English words into their constituent morphemes, the Porter stemmer algorithm [63] was adopted. This algorithm helps to remove English inflectional morphemes by stripping the suffixes in the stem word. For example, the word ‘love’ can be inflected into words like ‘love’, ‘loves’, ‘loved’ and ‘loving’. In this case, all the words are replaced into a single stem word ‘lov’. Porter stemmer algorithm is used due to its simplicity, good in speed and easily integrated with NLTK packages. The stem words of English sentiment words are stored in trilingual sentiment lexicon for English sentiment term detection during sentiment texts analysis.

#### 4.2.3.2 *Amharic Morphological Analysis*

Amharic is a morphologically rich language and highly inflected in number, gender, tense, person etc. Therefore, to handle those inflected words, a morphological analyzer tool is needed. So far, various researches have been carried out on Amharic morphological analysis although there is no any effective Amharic morphological analyzer tool available. Even if there is HornMorpho morphological analyzer tool, it is not inclusive to all Amharic word classes since it is developed only for Amharic verbs and nouns. However, most of Amharic sentiment words are adjectives and adverbs. Therefore, to analyze all Amharic sentiment words, it is important to develop Amharic morphological analyzer. In this study, Amharic morphological analyzer is developed manually by using Amharic sentiment words. The sentiment words are collected from [8,9,64] and online Amharic dictionary [65] for morphological analysis purpose. The analyzed Amharic sentiment words are represented by their root words. It is used to represent the polarity values of the sentiment words. The root word is generated after segmenting every morpheme of the words such as prefix, suffix and stems. For instance,

the word ‘ሰኬት’ can be inflected into “ሰኬተኛ፣ ሰኬታሞች፣ በሰኬት፣ የሰኬት፣ የተሳካ፣ ሰኬታሞቹ” etc., in this example, all those words have the same root that is ‘ሰኬ’. The root word is used to represent the sentiment polarity values of every Amharic sentiment words and stored at the backend component of trilingual sentiment analysis system. Algorithm 4.2 shows morphological analyzer for Amharic words.

Input: Amharic words

For each Amharic word translate in Latin

Check the word morphology exists in Amharic morphology database

    If the word is analyzed by Amharic morphological analyzer

        Return root of the word

    Else

        Not analyzed the word

    End if

End for

Output: Amharic root words

Algorithm 4.2: *Algorithm for Morphological Analysis of Amharic Words*

### **4.2.3.3 Tigrigna Morphological Analysis**

Like Amharic, Tigrigna is a morphologically rich language and inflected in various features with respect to number, gender, grammatical aspect, person and tense. So, to handle those inflected words, a Tigrigna morphological analyzer tool is needed. In the previous, different researches have been carried out to develop Tigrigna morphological analyzers, but there is no any effective Tigrigna morphological analyzer tool available. So, to solve this gap or problem, this study developed Tigrigna morphological analyzer manually. In this research work, first, [66] Tigrigna sentiment words were collected and

[67] online English-Tigrigna dictionary was used, and then all the sentiment words were analyzed. The morphologically analyzed sentiment words were assigned the sentiment polarity values using the root words. For example, the word ‘ጸቡቕ’ can be inflected into “ጸቡቕቲ፣ ጸቡቃት፣ ትጸቡቕ፣ ብጸቡቕ፣ ምጸባቃቶም፣ ከትጸቡቕ” etc., whereas all these words have the same root that is ‘ጸ-ቡ-ቕ’. The root word represents the sentiment polarity values of the sentiment words and stored in trilingual sentiment lexicons at backend component of the system. Algorithm 4.3 shows the algorithm for morphological analysis of Tigrigna words.

Input: Tigrigna words

For each Tigrigna word translate into Latin scripts

Check the word morphology exists in Tigrigna morphology database

If the word analyzed by Tigrigna morphological analyzer

Return root of the word

Else

Not analyzed the word

End if

End for

Output: Tigrigna root words

*Algorithm 4.3: Algorithm for Morphological Analysis of Tigrigna Words*

#### 4.2.4 Sentence Construction Using Root Words

Sentence construction using root words component is used to aggregate the English, Amharic and Tigrigna root words after the morphological process is completed. Languages of input texts are identified for the purpose of the texts to further analyze and in morphological analysis phase English, Amharic and Tigrigna words are analyzed

independently. However, to detect and determine the semantic orientation of the texts, it needs to reconstruct the independent root words. For instance, “በቀልድ የተዋዘ፣ ትልቅ ቁም ነገር አዘል ፊልም- የወንዶች ጉዳይ 2 is excellent!” This example contains mixed source of text in Amharic and English words. In this situation, after detecting the language of the words in the text, morphological analysis process is employed on each word independently. Behind the morphological analysis, the root words are reconstructed into single sentence in order to detect the sentiment words in the sentence and to determine its sentiment classes.

#### 4.2.5 Trilingual Sentiment Lexicon

The trilingual sentiment lexicon is an external component that stores English, Amharic and Tigrigna sentiment words, contextual valance shifters terms, negation terms and conjunction terms of these languages. In the lexicon, lists of words are grouped into positive words, negative words, overstatement words, understatement words, negation words and conjunction words with their corresponding polarity values.

In trilingual sentiment lexicon, the sentiment words are modeled for the purpose of sentiment analysis to any domains since the sentiment words are not restricted for specific domains. The sentiment words in the lexicon are tagged as ‘+’ and ‘-’ for positive and negative sentiment words respectively and contextual valance shifter terms are tagged as: ‘Inc’ for overstatement, ‘Dec’ for understatement, ‘Neg’ for negation and ‘Cont’ for contrast conjunction words respectively.

#### 4.2.6 Sentiment Word Detection

Sentiment word detection component is used to detect sentiment words and contextual valance shifter words in the sentence. In this research work, sentiment words and contextual valance shifter words are detected using dictionary of sentiment words with their polarity values annotated in the lexicon. The words in the sentence are checked in the trilingual sentiment lexicons to detect their sentiments. If the word is found in the lexicon and its corresponding value is ‘+’, then it is a positive sentiment word. Similarly, if sentiment word is found in the lexicon and its corresponding value ‘-’, then it is a negative sentiment word.

Example:

- ጥሩ ዘፈን ነው/Tru zefen new (Amharic)
- ሰሎሞን ሓውና ፅቡቅ ባህላዊ ደርሬ ቀፅለሉ/Selomon Hawna Sbuq bahlawi derfi qeSlelu (Tigrigna)
- I like the music (English)
- ወዲ ባይረ nice song አጃካ በርትዕ/wedi bayre nice song 'ajoka bert` (Mixed)

In this example, ጥሩ/Tru, ፅቡቅ/Sbuq, like and nice are positive sentiment words, because of the positive sentiment words all the sentences have positive sentiments.

### ***Context Valance Shifter Words***

Contextual valance shifter words change the semantic orientations of their neighboring sentiment words in the sentence. For example, they reverse the positive sentiment words into negative sentiments or change the intensity of the polarity values. Contextual valance shifter words are negations, intensifiers, diminishers and conjunction words.

***Negations*** are used to reverse the semantic orientation of the sentiment words in the sentence. Words like no, not, never (in English), አይደለም/'aydelem, አይደለሁም/'aydelehem (in Amharic) and አይኮነን/'aykonen, አይኮነን/'aykonun (in Tigrigna) are examples of negation words which reverse the polarity of the sentiment words in the sentence.

If a sentiment sentence contains a negation word and negative sentiment word, then the polarity of the sentence is changed into positive sentiment. But, if the sentence contains negation word and positive sentiment word, then the polarity of the sentence is changed into negative sentiment. And also, if the sentiment sentence contains negation word and neutral word, then the polarity of the sentence is negative. Consider this sentence, “the food is amazing” versus “the food is not amazing”. In the first one, amazing is a positive word and the sentence has positive sentiment, but when not is added to the clause, amazing is used as a negation context and the sentence has negative sentiment. Table 4-1 shows the effect of negation words in sentiment sentences.

Table 4-1: *Negation Words in Sentiment Sentences*

| Amharic   | Tigrigna                                   | English                           | Mixed                         |
|---|--|-----------------------------------|-------------------------------|
| ቆንጆ ሆኖል ነው/qonjo hotEl<br>new (+ve)                     | ጥዑም ምግብ እዩ/T`um<br>mgbi 'yu (+ve)          | Amazing food<br>(+ve)             | Delicious የሆነ ምግብ<br>ነው (+ve) |
| ቆንጆ ሆኖል አይደለም/qonjo<br>hotEl 'aydelem (-ve)             | ጥዑም ምግብ አይኮነን/ T`um<br>mgbi 'aykonen (-ve) | It is not amazing<br>food(-ve)    | ምግቡ delicious<br>አይደለም (-ve)  |
| የሚያስጠላ ፍልም ነው<br>/yemiyasTela film new (-ve)            | ረሳሕ ገዛ እዩ/resaH geza 'yu<br>(-ve)          | The food is<br>terrible (-ve)     | አስጠሊ movie ነው<br>(-ve)        |
| የሚያስጠላ ፍልም አይደለም<br>/yemiyasTela film 'aydelem<br>(+ve) | ረሳሕ ገዛ አይኮነን/resaH geza<br>'aykonen (+ve)  | The food is not<br>terrible (+ve) | አስጠሊ movie<br>አይደለም (+ve)     |

**Intensifiers and diminishers** are words that change the degree of the positive or negative sentiment words in the sentence. For example, ‘the movie is very good’, the words very good are more positive together than just good alone. Another example of an intensifier word in a sentence is ‘the music is barely interesting’. The word barely makes the sentiment less positive. Therefore, very and barely are examples of overstatement and understatement words respectively. An overstatement word increases the intensity of the positive or negative sentiment words, while understatement word decreases the intensity of the positive or negative sentiment words by modifying the polarity strength of the sentiment words in the sentence. In general, words like very, extremely (in English), በጣም/beTam, እጅግ/Ejg (in Amharic), ብጣዕሚ/bTa`mi, ማራ/mara (in Tigrigna) and a little bit, less (in English), ትንሽ/tnx, ጥቂት/Tqit (in Amharic), ቁሩብ/qurub, ንእሽተይ/n'xtey, ውሑድ/wHud (in Tigrigna) are overstatement and understatement words respectively.

The overstatement and understatement words are amplifying and attenuating their neighboring sentiment words in the sentence. This is performed based on the polarity values of the sentiment word modeled in the sentiment lexicon. All positive and negative sentiment words in sentiment lexicon have the default value of +2 and -2 respectively. If the positive sentiment word is preceded by an overstatement word in the same clause, it gives a value of +3 or if the positive sentiment word is preceded by understatement word in the same clause then it gives a value of +1. Similarly, for negative sentiment words, if

the negative sentiment word is preceded by overstatement in the same clause, then it gives a value of -3 or if it is preceded by understatement word in the same clause, then it gives the value of -1. In this case, the sentiment sentence is classified into strong positive, positive, strong negative and negative polarity classes. Table 4-2 shows the effects of an overstatement and understatement words in sentiment sentences.

Table 4-2: *Intensifier Terms in Sentiment Sentences*

| Amharic  | Tigrigna  | English                                   |
|--|---|---|
| ጥሩ ፊልም ነው/Tru film new<br>(+ve)                                  | ጽቡቅ ፊልሚ እዩ/Sbuq filmi 'yu<br>(+ve)                    | The movie is nice (+ve)                   |
| በጣም ጥሩ ፊልም ነው/beTam Tru<br>film new (strong +ve)                 | ብጣዕሚ ጽቡቅ ፊልሚ እዩ/bTa`mi<br>Sbuq filmi 'yu (strong +ve) | The movie is very nice.<br>(strong +ve)   |
| የሚያስጠላ ፊልም ነው<br>/yemiyasTela film new (-ve)                     | ሕማቕ ፊልሚ እዩ/HmaQ filmi 'yu<br>(-ve)                    | The movie is boring<br>(-ve)              |
| በጣም የሚያስጠላ ፊልም ነው<br>/beTam yemiyasTela film<br>new (strong -ve) | ምሒር ሕማቕ ፊልሚ እዩ/mHir<br>HmaQ filmi 'yu (strong -ve)    | The movie is very<br>boring. (strong -ve) |

**Conjunction words:** conjunctions are highly functional and important for constructing sentences. Basically, conjunctions link words, phrases and clauses together in a sentence and it may be used to indicate the relationship between the ideas expressed in a clause and the idea expressed in the rest of the sentence. Conjunction words play the vital role in determining the overall sentiment polarity of the sentence since conjunctions often change the sentiment into opposite orientations or add in the strength of the sentiments. The coordinating and subordinating conjunction words are modeled in the lexicon with their corresponding polarity values. Conjunction words like: and (in English), እና/Ena (in Amharic) and ከምእውን/kem'wn (in Tigrigna) usually link sentiment words or phrases which have the same polarities, whereas conjunction words such as: but, however, although (in English), ግን/gn, እንጂ/Enji, ቢሆንም/bihonm (in Amharic) and ሽሕኳ/xHkWa, እንትኾነ/ntXone, ገና/gEna, ይኹንእምበር/yXun'mber (in Tigrigna) usually link sentiment words, phrases or clauses that have opposite polarities.

#### 4.2.7 Sentiment Word Polarity Weight Determination

Sentiment word polarity weight determination component is used to determine the polarity weight of the sentiment sentences. The polarity weight of the sentiment sentence is determined based on the polarity weight of the sentiment words in the sentence. The sentiment words can be either positive or negative. Sentiment sentences can also contain contextual valance shifters and negation words. The overall polarity weight of the sentiment sentence is determined based on the positive and negative sentiment terms with the effects of contextual valance shifters and negation words in the sentence. The effect of valance shifters and negation words for positive and negative sentiment terms are decided based on the window size of the sentiment words in the sentence. In general, polarity weight of sentiment sentence is determined according to the following rules.

##### ***Rule One***

If a sentence contains only positive sentiment words and the positive sentiment words are not linked with any contextual valance shifter words, then the positive sentiment words have maintained their default polarity weight values.

Example:

- This is the nice music (English)
- ቆንጆ ዘፈን ነው/qonjo zefen new (Amharic)
- ሓሪፍ ሙዚቃ እዩ/Harif muziqa 'yu (Tigrigna)
- ቆንጆ ዘፈን ነው I like it/qonjo zefen new I like it (Mixed)

In this example, nice, ቆንጆ/qonjo, ሓሪፍ/Harif and like are positive sentiment words which have positive sentiments in the sentences. In this case, each sentence has positive sentiments with +2 polarity weight values.

##### ***Rule Two***

If a sentence contains only negative sentiment words without any modifier words, then the negative sentiment word is assigned its default polarity weight values.

Example:

- The hotel is boring (English)
- አስጠሊ ሆቴል ነው/ osTeli hotEl new (Amharic)
- ሕማቅ ቤት ብልጫ እዩ/HmaQ bEt bl`i `yu (Tigrigna)
- እቶም ሰባት awful ዝኾነ food እዮም ኡብሊያምና/`tom sebat awful zXone food `yom `abli`omna (mixed)

In this example, all the sentences have negative sentiments with -2 polarity weight values since the words boring, አስጠሊ/ osTeli, ሕማቅ/HmaQ and awful are negative sentiment words and make the sentence to have negative sentiments.

### ***Rule Three***

If a positive sentiment word in a sentence is preceded by overstatement word, then the default polarity weight value of the positive sentiment word is increased by one (i.e., shifts from +2 to +3).

Example:

- Very nice music (English)
- በጣም ቆንጆ የሆነ ስራ ነው/beTam qonjo yehone sra new (Amharic)
- ብጣዕሚ ተባዕ ስራ-ሕተኛ እዩ/bTa`mi teba` seraHteNa `yu (Tigrigna)
- የእቤ እና ከቤ Comedy በጣም fun አድርጎኛል!!!/ye`obE na kebE comedy beTam fun `odrgoNal!!! (mixed)

The words very, በጣም/beTam and ብጣዕሚ/bTa`mi modify the polarity weight values of the sentiment words in each sentiment sentences. This example shows an overstatement words and amplify the polarity strength of the sentiment sentences.

### ***Rule Four***

If a negative sentiment word in a sentence is preceded by an overstatement word, then the default polarity weight value of the negative sentiment word shifts from -2 to -1 and the sentiment expression will be strong negative.

Example:

- The food is extremely nasty (English)
- የሰው እና የመንግስት ንብረት መዘረፍ እጅግ መጥፎ ተግባር ነው/yesew `Ena yemengst nbret mezref `Ejg meTfo tegbar new (Amharic)
- ሽጋራ ምትካኽ ምሥር ሐማቕ ተግባር ተባሂሉ ኣብ ክልልና ብዕሊ ተኸልኺሉ/xgara mtkaX mHir HmaQ tegbar tebahilu 'ab kllna b`li teXeIXilu (Tigrigna)
- In our region ገለ ገለ ሓለፍቲ very corrupted እዮም!!!/In our region gele gele Halefti very corrupted 'yom!!! (mixed)

In this example, nasty, መጥፎ/meTfo, ሐማቕ/HmaQ and corrupted are negative sentiment words. While extremely, እጅግ/Ejg, and ምሥር/mHir are amplifier sentiment words which amplify the semantic orientation of the negative sentiment words in the sentence.

### **Rule Five**

If a positive sentiment word in a sentence is preceded by understatement word, then the polarity weight value/strength of the positive sentiment word is decreased by one. The understatement word makes to attenuate the polarity strength of the sentiment words in the sentence. In this case, the semantic orientation of the sentiment will be weak positive.

Example:

- The music is a little bit interesting (English)
- The movie I have seen is less astonishing (English)
- ትንሽ ደስ የሚል ዘፈን ነው/tnx des yemil zefen new (Amharic)
- ንእሽተይ ጽርይ ዝበለ ዝላ እዩ/n'xtey Sry zbele geza 'yu (Tigrigna)

In this example, a little bit, less, ትንሽ/tnx and ንእሽተይ/n'xtey are an attenuator words that modify the original polarity weight vale of the positive sentiment words in the sentence. The strength of sentiment become weakens and the polarity weight values shift from +2 to +1.

### **Rule Six**

If the negative sentiment word in a sentence is preceded by understatement word, then the polarity weight value of the negative sentiment word is decreased by one. In this case, the sentiment expression of the sentence becomes weak negative and polarity weight value shifts from -2 to -3.

Example:

- The music is a little bit boring (English)
- ትንሽ የሚያስጠላ ዘፈን ነው/tnx yemiyasTela zefen new (Amharic)
- ቁሩብ ዘጽልእ ደርፊ እዩ/qurub zeSI' derfi 'yu (Tigrigna)

The terms boring, የሚያስጠላ/yemiyasTela and ዘጽልእ/zeSI' are negative sentiment terms and when a little bit, ትንሽ/tnx and ቁሩብ/qurub are added in the sentence, then the polarity weight values are altered and the sentiment expressions become weak negative.

### ***Rule Seven***

If a positive sentiment word is preceded or followed by negation terms in the sentence, then the sentiment of the positive word reverses into negative sentiments.

Example:

- The Hotel is not good (English)
- ጥሩ ሆቴል አይደለም/Tru hotEl 'aydelem (Amharic)
- ፅብቅ ምግብ አይኮነን/Sbuq mgbi 'aykonen (Tigrigna)
- በዚህ food satisfied አይደለም/bezih food satisfied 'aydelehum (mixed)

In this example, the words not, አይደለም/'aydelem, አይኮነን/'aykonen and አይደለም/'aydelehum are served as negations and changes the positive sentiment expression into negative sentiments and the polarity weight value of these sentences will be -2.

### ***Rule Eight***

If a negative sentiment word is preceded or followed by negation words in the sentence, then the polarity of the sentiment sentence is shifted into positive sentiments.

Example:

- The music is not boring (English)
- መጥፎ ምግብ አይደለም/meTfo mgbi 'aydelem (Amharic)
- ሕማቕ ምግብ አይኮነን/HmaQ mgbi 'aykonen (Tigrigna)

In this example, the sentences have positive sentiments since the negation words change the semantic orientation of the negative sentiment words in the sentence into positive sentiments.

Algorithm 4.4 shows the algorithm of sentiment word detection and determination of polarity weight values of the sentiment words in the sentence.

Input: English, Amharic, Tigrigna root words

For each root word checked in trilingual sentiment lexicon

    If the root word exists in sentiment lexicon

        Assign the default PWV of the root word //PWV-Polarity Weight Value

        If the root word preceded by overstatement word

            PWV+1 //PWV of the root word increases by one

        Else if the root word preceded by understatement word

            PWV-1 //PWV of the root word decreases by one

        Else if the root word preceded/followed by negation

            PWV\*-1 //PWV of the root word reversed

        End if

    Else

        Assign zero PWV to the root word

    End if

End for

Output: total PWV of sentiment words

*Algorithm 4.4: Algorithm for Sentiment Word Detection and Polarity Weight Determination*

### ***Rule Nine***

If a positive sentiment word preceded or followed by negation and intensifier words, then the semantic orientation (i.e., polarity and strength) value of the sentiment sentence is unchanged.

Example:

- The music is not very good (English)
- እጅግ ጥሩ ዘፈን አይደለም/ Ejj Tru zefen 'aydelem (Amharic)
- ብጣዕሚ ፅቡቅ ሙዚቃ አይኮነን/bTa`mi Sbuq muziqa 'aykonen (Tigrigna)

In this example, the sentences have positive sentiments with polarity weight value of +2.

### ***Rule Ten***

If a negative sentiment word in a sentence preceded or followed by negation and intensifier words, then the sentiment of the sentence has still negative sentiments and its polarity weight value is unchanged.

Example:

- The movie is not very boring. (English)
- እጅግ የሚያስጠላ ፊልም አይደለም/ Ejj yemiyasTela film 'aydelem (Amharic)
- አዝዩ ዘሕርቕ ፊልሚ አይኮነን/'azyu zeHrQ filmi 'aykonen (Tigrigna)

### ***Rule Eleven***

If sentiment words, phrases or clauses of a sentence connected by conjunctions like but, yet in English, ግን/gn, እንጂ/Enji, ቢሆንም/bihonm in Amharic and ሽሕኳ/xHkWa, ይኸንእምበር/yXun'mber in Tigrigna, then the sentiment polarity of the sentence changes into the polarity of the phrase or clause expressed after the conjunction. In sentiment sentence, these conjunctions always link words or phrases that have opposite polarity values and the sentiment polarity of the sentence is determine based on the polarity value of the clause or the phrase expressed after the conjunction.

Example:

- The bed is smart, but the mattress is dirty (English)
- ሰፊ ቤት ነው ግን ቆሻሻ ነው/sefi bEt new gn qoxaxa new (Amharic)
- ሰነፍ ተምሃራይ እዩ ይኸንእምበር ንስድራኡ ኡብ ምሕጋዝ ጎበዝ እዩ/senef temharay 'yu yXun'mber nsdra'u 'ab mHgaz gobez 'yu (Tigrigna)

In this example, the sentiments of the clauses before and after the conjunction have opposite polarity values.

### ***Rule Twelve***

If the sentiment words, phrases or clauses in a sentence connected by connectors like and (in English), እና/Ena (in Amharic) and ከምእውን/kem'wn (in Tigrigna), then the sentiment polarity of the sentence is not changed since these conjunctions always connect words, phrases or clauses that have the same polarities and they don't affect the overall sentiment polarity of the sentence.

Example:

- The song is nice and interesting (English)
- ምቹ እና ንጹህ አልጋ ነው/mcu `Ena nSuh `olga new (Amharic)
- ናይ ወዲ ባየረ ሙዚቃ ፅቡቕ ከምእውን ጥዑም ባህላዊ ደርፊ እዩ/nay wedi bayere muziqa Sbuq kem'wn T`um bahlawi derfi 'yu (Tigrigna)

In this example, all the sentiment terms in the sentences have similar semantic orientations.

In sentiment sentence words, phrases and clauses are connected with conjunctions and the overall semantic orientation of the sentence and its polarity weight values is determined using algorithm 4.5.

Input: Amharic, Tigrigna, English root words

For each root word check in trilingual sentiment lexicon

```
    If the root word exist in the sentiment lexicon
        Assign default PWV of the root word
        If root word preceded/followed with negation
        and intensifier words
            Return default polarity weight value of
            the root word
        Else if root word in a sentence connected with
        contrast conjunction
            Assign the PWV of the root word after the
            conjunction
        Else if the root word in a sentence connects
        with and, እና/`Ena, ከምእውን/kem'wn
            PWV of the root word maintain its default
            values
        End if
    End if
```

End for

Output: Polarity weight value of the sentiment sentence

*Algorithm 4.5: Algorithm to Determine the Polarity Weight of Sentiment Sentence*

#### 4.2.8 Sentiment Classification

Sentiment classification component is used to classify the sentiment sentence into positive, negative and neutral sentiment categories based on the average polarity weight value of the sentiment terms in the sentence. In this research work, sentiment classification performs based on lexicon-based techniques which takes the clue from sentiment lexicons. It extracts sentiment terms from the sentence using the dictionary of sentiment terms and its polarity weight values in the sentiment lexicon. The polarity value of the sentence is calculating from the polarity value of the sentiment words in the

sentence. This is performed by adding individual polarity weight values of the sentiment words in the sentence using Equation 1 and the final result used as the polarity weight values of the sentiment sentence.

$$PWV = \sum_{i=0}^m Wi \text{ ----- (2)}$$

Where, PWV is the total polarity weight value of the sentiment sentence S, m is the total number of positive or negative words in the sentence S,  $W_i$  is the individual polarity value of the sentiment words (i.e., positive, negative) in S.

*Sentiment Polarity* is the polarity of the sentiment sentence. It is categorized based on the average polarity weight values in the sentences which are positive, strong positive, weak positive, negative, strong negative, weak negative and neutral polarity classes. The neutral polarity class has zero polarity weight values. The system classifies the sentence into neutral class is either there are equal number of positive and negative sentiment terms in the sentence, for instance “enjoy the crazy movie” or the sentence is non-sentiment sentence since, if there are equal numbers of positive and negative sentiment terms in sentiment sentence, then the average polarity weight value will be zero. In this case, the system is considered as the sentence as neutral and assigns zero polarity weight values. Table 4-3 shows the sentiment classes with their corresponding polarity weight values in the sentiment sentences.

Table 4-3: *Polarity Categories of Sentiment Sentences*

| # | polarity weight value | Sentiment polarity | Example                    |
|---|-----------------------|--------------------|----------------------------|
| 1 | If $0 < pwv \leq 1$   | Weak positive      | The movie is less nice     |
| 2 | If $1 < pwv \leq 2$   | Positive           | The movie is nice          |
| 3 | If $pwv > 2$          | Strong positive    | The movie is very nice     |
| 4 | If $pwv < -2$         | Weak negative      | The movie is less boring   |
| 5 | If $-2 \leq pwv < -1$ | Negative           | The movie is boring        |
| 6 | If $-1 \leq pwv < 0$  | Strong negative    | The movie is very boring   |
| 7 | If $pwv = 0$          | Neutral            | This is an Ethiopian movie |

# Chapter Five: Experiment

## 5.1 Introduction

Section 5.2, the development environment and tools used to develop the prototype of trilingual sentiment analysis system are discussed. Section 5.3 shows the prototype of the system. Section 5.4 deals the data collections used for language identification, to construct trilingual sentiment lexicon and to test the performance of the system. Section 5.5 presents the results and discussions.

## 5.2 The Development Environment and Tools

The trilingual sentiment analysis system is developed and tested on Toshiba laptop computer:

- Laptop computer with windows 7 ultimate operating system, Intel core i5 with 2.6 GHz processor speed, 6.0GB RAM and 465 GB hard disk capacity
- Other software components used to develop and test our system is Python, NLTK, Notepad++ and PyMySQL.

Python<sup>3</sup> is a simple yet powerful programming language with excellent functionality for processing linguistic data [68]. It is highly readable, allows data and methods to encapsulate and contains extensive library including components for graphical programming, numerical programming and web connectivity. For this work, the version 3.5 of python 32-bits is used. It is used to develop the prototype of trilingual sentiment analysis system.

NLTK<sup>4</sup> is natural language processing toolkit that can be used to build NLP programs in python. It is an open source toolkit that contains open source python module, linguistic data and documentation for research and development in natural language processing field [68]. It provides basic class for representing data relevant to NLP, standard interface for performing tasks such as text classification, part-of-speech tagging and syntactic parsing. In this work, NLTK is used for sentence and word tokenization in order to split the input sentiment sentences into lists of words.

---

<sup>3</sup> <http://www.python.org/>

<sup>4</sup> <http://www.nltk.org/>

Notepad++<sup>5</sup> is a free source code editor and the replacement of Notepad that supports several languages. In this work, we used Notepad++ version of 7.5.1 with 32-bits which is used to build and edit the lexicons.

PyMySQL<sup>6</sup> is the package that contains pure-Python MySQL client library and used to connect and access MySQL databases.

### 5.3 Prototype of the System

The system prototype is developed using python programming language. The developed system provides the functionalities for analyzing user generated sentiment texts. Users can provide the sentiment texts into the system in two ways. One by preparing the sentiment texts in UTF-8 file formats in specific location and imported into the system or by providing sentiment sentences one by one on the provided GUI. Then, the system generates the sentiment polarity with corresponding polarity weight values. Figure 5.1 shows the UI of trilingual system

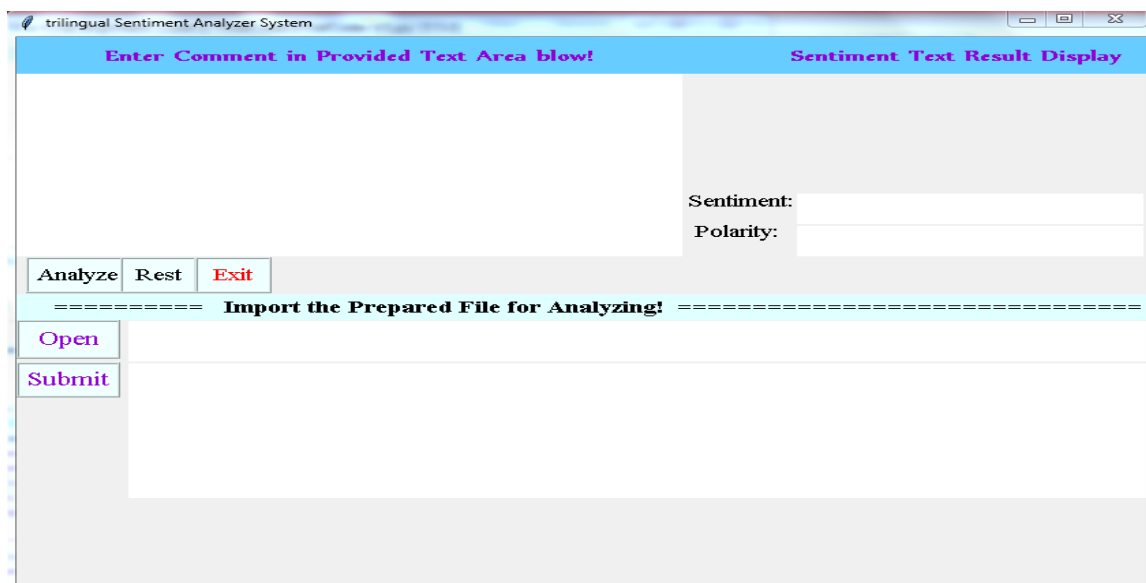


Figure 5-1: *Trilingual Sentiment Analysis System UI*

<sup>5</sup> <https://notepad-plus-plus.org/>

<sup>6</sup> <https://pypi.python.org/pypi/PyMySQL/>

Figure 5.2 shows the system provides summarized sentiment information.

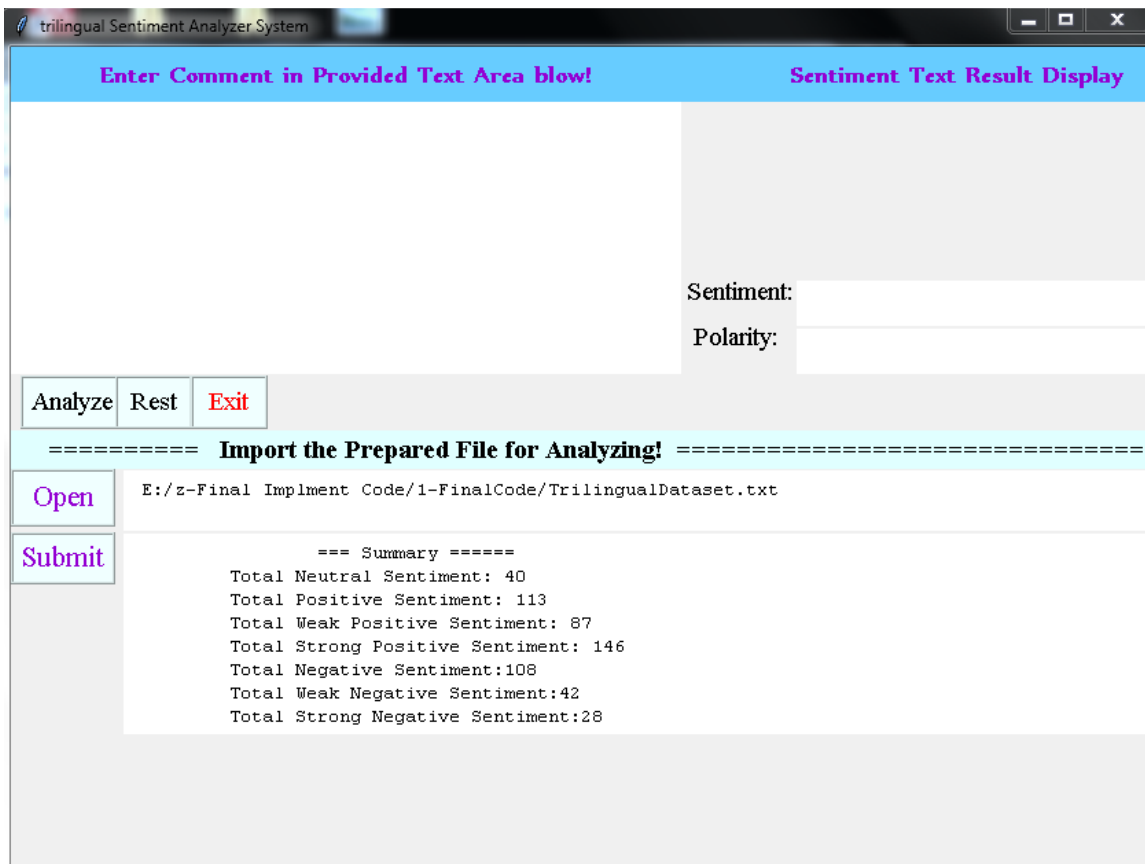


Figure 5-2: Analyzed Sentiment Information Result

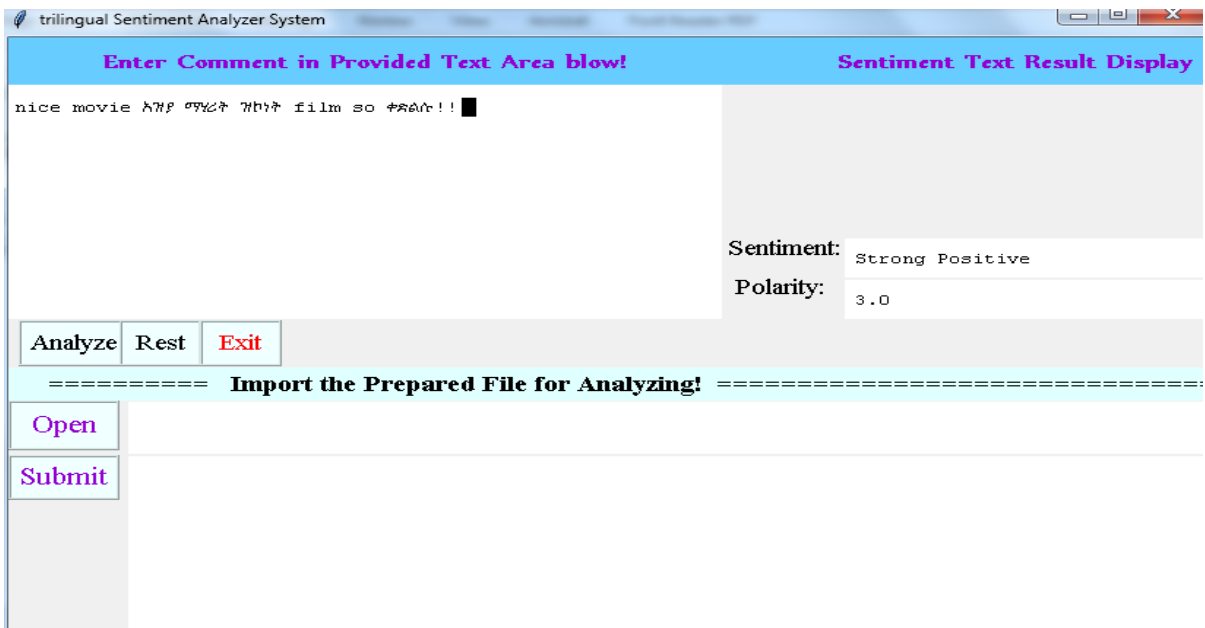


Figure 5-3: Result of Input Sentiment Text with its Sentiment and Polarity Values

## 5.4 Data Collection

In this research work, different datasets were collected for language identification, to build trilingual sentiment lexicons and to evaluate the performances of the trilingual sentiment analysis system.

**Language identification:** English, Amharic and Tigrigna languages were detected in word level using lexicon based approaches. To build the lexicon, English, Amharic and Tigrigna wordlists were used. English wordlists are collected from [69] which have a total of 109,584 English words for English language identification. Similarly, Tigrigna and Amharic wordlists were collected from [70]. Manual preprocessing activities were made for Amharic and Tigrigna wordlists in order to avoid unimportant words and phrases. After the preprocessing activity is completed, a total of 341,669 Amharic words and 384,205 Tigrigna words were used for Amharic and Tigrigna language identification.

**Trilingual sentiment lexicon:** the trilingual sentiment lexicon is built manually by collecting sentiment words from various resources. For English sentiment word detections, English sentiment words were collected from [71] and for Amharic and Tigrigna mentioned in Section 4.2.2. In general, the trilingual sentiment lexicon contains English, Amharic and Tigrigna sentiment words and the sentiment words are validated by linguistic expertise. Some lists of the sentiment words are provided on the Appendix part. Table 5-1 shows the total number of sentiment words used to build the trilingual sentiment lexicons.

Table 5-1: *Sentiment Words in Trilingual Sentiment Lexicons*

| Language | Sentiment words |          |                          |
|----------|-----------------|----------|--------------------------|
|          | Positive        | Negative | Context valence shifters |
| English  | 2006            | 4783     | 25                       |
| Amharic  | 1009            | 1264     | 18                       |
| Tigrigna | 1663            | 1122     | 26                       |

Finally, for performance evaluation of trilingual sentiment analysis system, 564 sentiment sentences have been manually collected from Facebook and YouTube on the domain of movie, music, politics and hotel. These domains were collected because of absence of any prepared available written sentiment sentences in Amharic and Tigrigna languages. At the time of collecting the data, the sentiments on the posted topics are selected randomly. After collecting the sentiment data, we made manual sentiment classification with help of linguistic expertise. The manual sentiment classification helps to compare the results generated by the system. The manual sentiment classification is used to assign the polarity of the sentiment sentences and helps to compare the results of the system after conducting the experiment. In this research work, sentiment sentences are categorized into three basic classes based on their semantic orientations: positive, negative and neutral sentiment classes. While, the positive and negative sentiment classes are classified further into positive, strong positive, weak positive, negative, strong negative and weak negative sentiment classes. The sentiment sentences can contain English, Amharic, Tigrigna and mixed language of words since different opinion holders can express their feelings for the same topic in different perspectives; it can use different languages or can combine languages for their expressions. For example, consider this sentence “nice movie አዝያ ማሃሪት ዝኮነት film so ቀጽልሉ!!!” or “too much happy አይደለሁም በዚህ music” In this example, the first sentiment sentence contains Tigrigna and English words in the movie topic, while the second sentiment sentence contains English and Amharic words in the music topic. In this study, such kinds of sentences are considered and the sentiment is determined based on the sentiment words in the sentence. However, there are also pure Amharic, Tigrigna and English sentiment sentences on the topic from opinion holders. Finally, the collected sentiment sentences were categorized into predefined categories: positive, strong positive, weak positive, negative, strong negative, weak negative and neutral sentiment classes.

## 5.5 Results and Discussion

### 5.5.1 Evaluation Procedures

The experiment is done to measure the overall performance of the developed trilingual sentiment analysis system. In this research work, a total of 564 sentiment sentences were used to test the accuracy of the system. The test results achieved were presented in Section 5.5.3 in Table 5-2.

### 5.5.2 Evaluation Results

The role of this activity is to describe the evaluation metrics of the designed system and followed by its test results. We have used precision, recall and F-measure evaluation metrics to measure the effectiveness of the selected approach. In this study, the experiment is done based on the sentiment polarity classes and evaluates each evaluation metrics corresponding to each sentiment polarity classes. The sentiment polarity classes are classified into seven categories: positive, weak positive, strong positive, negative, weak negative, strong negative and neutral classes. At last, the precision, recall and F-measure were calculated in the experiment for each polarity classes.

**Precision** measures the correctness of the classifier. It is the ratio of the correctly classified sentiment categories over the sentiment classified into correct sentiment categories. In other words, it measures how many of the sentiment sentences that the system classified in the correct class. A high precision indicates the system achieves the better correctness on sentiment sentence classification.

$$P = \frac{TP}{TP + FP} \dots\dots\dots (3)$$

Where, P stands for precision, TP stands for true positive and it denotes the number of sentiment sentences classified in correct sentiment class and FP stands for false positive, it represents the number of incorrectly classified sentiment sentences as correct class.

**Recall** is the ratio of the total number of correctly classified sentiment class to the total number of correct sentiment in the actual class. High recall refers the system is better and not miss correct classes.

$$R = \frac{TP}{TP + FN} \dots\dots\dots (4)$$

R stands for recall, TP stands for true positive and it donates the number of sentiment sentences classified in a correct class, FN stands for false negative and it donates the number of sentiment sentences classified incorrectly in other class.

**F-score** is the harmonic mean of precision and recall. That is:

$$F\text{-measure} = \frac{2 * P * R}{P + R} \dots\dots\dots (5)$$

### 5.5.3 Test Results

In this research work, to evaluate the performance of the system and its effectiveness on sentiment polarity classifications, the experiment is conducted using the collected sentiment datasets. In this experiment, we considered the effects of contextual valence shifter terms and taking an account of negation terms in sentiment sentences. The negation and contextual valance shifter terms include the amplifiers, deminishers and conjunction terms in sentiment sentences, due to this, the sentiment polarity classes are classified into seven categories which are positive, strong positive, weak positive, negative, strong negative, weak negative and neutral classes. Table 5-2 shows the evaluation results of the experiment on each sentiment polarity classes.

Table 5-2: *Evaluation Result of the Experiment*

| Polarity class  | Evaluation Metrics |            |             |
|-----------------|--------------------|------------|-------------|
|                 | Precision (%)      | Recall (%) | F-score (%) |
| Weak positive   | 89.31              | 72.67      | 80.13       |
| Positive        | 95                 | 85.95      | 92.40       |
| Strong positive | 94.54              | 90.29      | 92.36       |
| Negative        | 91.61              | 84         | 87.64       |
| Weak negative   | 83.79              | 85.28      | 84.53       |
| Strong negative | 85.14              | 78.14      | 81.49       |
| Neutral         | 73.07              | 97.16      | 83.41       |
| Average         | 87.49              | 84.78      | 85.99       |

### 5.5.4 Discussion

As shown from Table 5-2 above, the experimental results are different for each polarity classes. In this Subsenction, we discuss the results of each polarity classes and the reasons of the variation of the results on each polarity classes.

In Table 5-2, the precision of positive and strong positive polarity classes are higher than the remaining polarity classes, because much of the test datasets contain the positive

sentiment sentences and the sentiment sentences belong in the positive and strong positive sentiment classes. Therefore, the size of the sentiment sentences belongs in positive and strong positive sentiment classes help to achieve better performance results in the polarity classification of the system. In negative and strong negative polarity classes, the precision is higher than their recalls. This indicates the classification approach predicts the right sentiment classification in negative and strong negative sentiment polarity classes and the system achieves better correctness in the classification. In neutral polarity class, the system achieves the lowest precision, but the highest recall from the other polarity classes. The reason is the effect of misspelling sentiment words in sentiment sentences, the equal occurrence of sentiment terms in the sentence, the size of the sentiment words in the trilingual sentiment lexicons and the presence of sarcastic sentence in sentiment sentences. For example, the sentence ‘I appreciate the way how he expresses about narrow minded politicians’, this is a sentiment sentence and it has positive polarity, however due to equal number of sentiment term occurrence (i.e., one positive term (appreciate) and one negative term (narrow)) in the sentence, the system assigns the sentence wrongly into neutral polarity class. Other example also, ‘this is a beautiful cultural wedding song!!!!!!’ in this example, the sentiment sentence contains a wrong spelling sentiment word (i.e., beautiful) and the system considers all the terms in the sentence are neutral and the sentence has neutral polarity values. If the sentiment sentence contains misspelled sentiment terms, equal number of sentiment terms, or not correctly identified the language of the sentiment terms or the sentiment term not exists in the trilingual sentiment lexicons, then the system automatically assigns the polarity of the sentence into neutral polarity classes. Therefore, an incorrect polarity classification of the sentiment sentence into neutral polarity class has an influence to the performance results of neutral polarity classes. In general, the absence of spelling checker, the existence of sarcastic sentences and ambiguous words in sentiment sentences have their own negative impacts/influences in overall performance of the system in sentiment classifications.

In general, the values of the evaluation metrics are encouraging, large size of positive and negative sentiment terms in the lexicon, appropriate language identification, concerning the effects of valance shifters and negation terms in sentiment sentences are assumed to improve the performance of the sentiment polarity classifications.

## Chapter Six: Conclusion and Future Work

### 6.1 Conclusion

There are several social media sites that use different languages around the world. It is obvious that the contents of the information available on the social media can be in different languages. With the high availabilities of multilingual text data on social media, it needs to handle the social media textual contents written in multiple languages and needs to convert the multilingual data into understandable and easily useable formats. Regarding this, multilingual sentiment analysis tries to analyze and classify sentiment sentences based on the sentiment polarities using various approaches.

Accordingly, this study addressed the knowledge gap or problem of sentiment analysis on trilingual sentiment sentences, typically available on social media. In order to address this problem, this study investigated an open source tools which provide various services. The levels of sentiment analysis and different techniques used for sentiment analysis and classifications like lexicon based approach, machine learning approach and hybrid approach were studied. This research work, investigated the linguistic behaviors of English, Amharic and Tigrigna languages especially on their morphological characteristics. The comprehensive reviews on monolingual and multilingual sentiment analysis on social media and related areas were undertaken. An understanding of the techniques and linguistic behaviors allowed the formulation of the requirements for each component of trilingual sentiment analysis system. The trilingual sentiment analysis system consists of preprocessor, language identifier, morphological analyzer, sentence constructor using root words, sentiment term detector, sentiment term polarity weight determiner and sentiment classifier components.

In developing the trilingual sentiment analysis system, this study employed language identification component for the purpose of morphological analysis and detecting sentiment terms of the corresponding languages in the inputted texts. The languages are from different families with their own writing scripts. In order to maintain the data uniformity and consistency, the Ethiopic scripts were converted into Latin scripts since Latin script representations are less complex and contain small numbers of letters relative to Ethiopic scripts. Analyzing of trilingual sentiment sentences are based on sentiment terms in the trilingual sentiment lexicons and some rules for detecting and

determining the polarity weights of sentiment terms were built in the sentiment sentences.

Finally, in order to verify the objective, evaluation of trilingual sentiment analysis system was performed using social media data. The evaluation takes from the performance test of the system. The success of the demonstration and performance test clearly showed that the feasibility of providing the system for analysis of sentiment texts on social media. To evaluate the performance of the system on sentiment analysis, 564 sentiment sentences were collected from different topics that were posted on Facebook and YouTube. To demonstrate the effectiveness of the system precision, recall and F-measure evaluation metrics were conducted and an average precision of 87.49%, average recall of 84.78% and average F-measure of 85.99% were obtained from the experiment. This study showed promising results, but more comprehensive future works make this more findings improved.

## 6.2 Contribution of the Thesis

The findings of this study can have considerable contributions as explained here under.

- Studied the trilingual sentiment analysis on social media using English, Amharic and Tigrigna languages help as bridge for other related works on trilingual sentiment analysis.
- Developed an algorithm that identifies the language of input texts from trilingual sentiment sentences.
- Developed an algorithm that detects and determines polarity values of sentiment terms from trilingual sentiment sentences.
- Built the trilingual sentiment lexicons for sentiment term detection and polarity weight determination in trilingual sentiment sentences.
- Developed a trilingual sentiment analysis system that analyzes and classifies sentiment polarities of trilingual sentiment sentences.

## 6.3 Future Work

The primary purpose of this study was to design and develop a trilingual sentiment analysis system on social media. In order to develop a full-fledged trilingual sentiment analysis system on social media, it needs coordinated team work from linguistic expertise and computer science expertise. Even if the system had already demonstrated good performance on realistic setting, still it needs further improvements. Future work may address the following in order to have a full-fledge of trilingual sentiment analysis system.

- Building more comprehensive list of positive and negative sentiment words to enhance the performance of the system.
- Identifying the language in phrase level and text level helps to include sentiment classification using phrase patterns.
- Developing sentiment terms in domain specific provides to take account of domain dependent sentiment terms and enhance system performance.
- Modeling ambiguous sentiment words helps to perform an appropriate sentiment classification.
- Modeling sarcastic or ironic sentences helps to reduce an incorrect classification of sentiment polarities.
- Analyzing sentiments in feature level helps to determine the sentiments on specific features/aspects of the objects.

## References

- [1] A. M. Kaplan and M. Haenlein, “Users of the World, Unite! The Challenges and Opportunities of Social Media”, *Business Horizons*, Vol. 53, No. 1, pp. 59–68, 2010.
- [2] E. Tromp, “Multilingual Sentiment Analysis on Social Media”, Unpublished Masters Thesis, Department of Mathematics and Computer Science, Eindhoven University of Technology, 2011.
- [3] Bo Pang and Lillian Lee, “Opinion Mining and Sentiment Analysis”, *Foundations and Trends® in Information Retrieval*, Vol. 2, No. 1-2, pp. 1–135, 2008.
- [4] Sentiment Analysis, Available at: [https://en.wikipedia.org/wiki/Sentiment\\_analysis](https://en.wikipedia.org/wiki/Sentiment_analysis), Last accessed on November 21, 2017.
- [5] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan and Claypool Publishers, May 2012.
- [6] M. Hu and B. Liu, “Mining and Summarizing Customer Reviews”, in *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp. 168-177, 2004.
- [7] P. D. Turney, “Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews”, in *Proceeding of the 40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 417–424, July 2002.
- [8] Selama G/Meskel, “Sentiment Mining Model for Opinionated Amharic Texts”, Unpublished Masters Thesis, Department of Computer Science, Addis Ababa University, 2010.
- [9] Tulu Tilahun, “Opinion Mining from Amharic Blog”, Unpublished Masters Thesis, Department of Computer Science, Addis Ababa University, 2013.
- [10] Wondwossen Mulugeta, “A Machine Learning Approach to Multi-Scale Sentiment Analysis of Amharic Online Posts”, *HiLCoE Journal of Computer Science and Technology*, Vol. 2, No. 2, pp. 80-87.

- [11] M. Al-ayyoub, S. B. Essa, and I. Alsmadi, “Lexicon-based Sentiment Analysis of Arabic Tweets”, *International Journal of Social Network Mining*, Vol. 2, No. 2, pp. 101–114, 2015.
- [12] R. M. Duwairi and I. Qarqaz, “Arabic Sentiment Analysis Using Supervised Classification”, in *Proceeding of the International Conference on Future Internet of Things and Cloud*, pp. 579–583, August 2014.
- [13] M. Abdul-Mageed, S. Kuebler, and M. Diab, “SAMAR: A System for Subjectivity and Sentiment Analysis of Arabic Social Media”, *Journal of Computer Speech and Language*, Vol. 28, No. 1, pp. 20-37, 2014.
- [14] E. Boiy and M. F. Moens, “A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts”, *Journal of Information Retrieval*, Vol. 12, No. 5, pp. 526–558, 2009.
- [15] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A.Y.A Hawalah, A. Gelbukh, and Q. Zhou, “Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques”, *Cognitive Computation*, Vol. 8, No. 4, pp. 757-771, August 2016.
- [16] G. Chowdhury, “Natural Language Processing”, *Annual Review of Information Science and Technology*, Vol. 37, pp. 51–89, 2003.
- [17] Natural Language Processing, Available at: [https://en.wikipedia.org/wiki/Natural\\_language\\_processing](https://en.wikipedia.org/wiki/Natural_language_processing), Last Accessed on October 28, 2017.
- [18] A. Jurek, M. D. Mulvenna, and Y. Bi, “Improved Lexicon - based Sentiment Analysis for Social Media Analytics”, pp. 1–13, 2015.
- [19] S. Sun, C. Luo and J. Chen, “A Review of Natural Language Processing Techniques for Opinion Mining Systems”, *Information Fusion*, Vol. 36, pp. 10–25, 2017.
- [20] Y. Liu, X. Huang, A. An, and X. Yu, “ARSA : A Sentiment-aware Model for Predicting Sales Performance Using Blogs”, in *Proceeding of the 30<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 607–614, 2007.

- [21] W. Zhang and S. Skiena, “Trading Strategies to Exploit Blog and News Sentiment”, in *Proceeding of the 4<sup>th</sup> International AAAI Conference on Weblogs and Social Media*, pp. 375–378, 2010.
- [22] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welp, “Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment”, in *Proceeding of the 4<sup>th</sup> International AAAI Conference on Weblogs and Social Media*, pp. 178–185, 2010.
- [23] B. O’Connor, R. Balasubramanyan, B. R. Routledge, and N. Smith, “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series”, in *Proceeding of the 4<sup>th</sup> International AAAI Conference on Weblogs and Social Media*, pp. 122–129, 2010.
- [24] N. Chambers, V. Bown, E. Geno, X. Tian, E. Young, and E. Yang, “Identifying Political Sentiment Between Nation States with Social Media”, in *Proceeding of the Conference on Empirical Method in Natural Language Processing*, pp. 65–75, September 2015.
- [25] C. Nopp and A. Hanbury, “Detecting Risks in the Banking System by Sentiment Analysis”, in *Proceeding of the Conference on Empirical Methods in Natural Language Processing*, pp. 591–600, September 2015.
- [26] T. Nasukawa and J. Yi, “Sentiment Analysis: Capturing Favorability Using Natural Language Processing”, in *Proceeding of the 2<sup>nd</sup> International Conference on Knowledge Capture*, pp. 70-77, 2003.
- [27] J. Wiebe, T. Wilson and C. Cardie, “Annotating Expressions of Opinions and Emotions in Language”, *Language Resource and Evaluation*, pp. 165-210, 2005.
- [28] B. Liu, *Sentiment Analysis and Subjectivity*, Handbook of Natural Language Processing, 2<sup>nd</sup> Edition, 2010.
- [29] N. Jindal and B. Liu, “Mining Comparative Sentences and Relations”, in *Proceeding of the 21<sup>st</sup> National Conference on Artificial Intelligence*, Vol. 2, pp. 1331–1336, 2006.
- [30] N. Jindal and B. Liu, “Identifying Comparative Sentences in Text Documents”, in *Proceeding of the 29<sup>th</sup> Annual International ACM SIGR Conference on Research and Development in Information Retrieval*, pp. 244-251, 2006.

- [31] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques”, in *Proceeding of the Conference on Empirical Methods in Natural Language Processing*, pp. 79–86, 2002.
- [32] G. Somprasertsri and P. Lalitrojwong, “Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization”, *Journal of Universal Computer Science*, Vol. 16, No. 6, pp. 938–955, 2010.
- [33] M. Abdul-Mageed and M. Diab, “AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis”, pp. 3907–3914, 2012.
- [34] A. Moreno-Ortiz and C. P. Hernández, “Lexicon-Based Sentiment Analysis of Twitter Messages in Spanish”, *Natural Language Processing*, Vol. 50, pp. 93–100, 2013.
- [35] K. Denecke, “Using SentiWordNet for Multilingual Sentiment Analysis”, *IEEE 24<sup>th</sup> International Conference on Data Engineering Workshop*, 2008.
- [36] A. Balahur and M. Turchi, “Multilingual Sentiment Analysis Using Machine Translation?”, in *Proceeding of the 3<sup>rd</sup> Workshop on Computational Approach to Subjectivity and Sentiment Analysis*, pp. 52–60, 2012.
- [37] G. Yan, W. He, J. Shen and C. Tang, “A Bilingual Approach for Conducting Chinese and English Social Media Sentiment Analysis”, *International Journal of Computer and Telecommunications Networking*, Vol. 75, pp. 491–503, 2014.
- [38] P. Palanisamy, V. Yadav and H. Elchuri, “Serendio : Simple and Practical Lexicon Based Approach to Sentiment Analysis”, *Second Joint Conference on Lexical and Computational Semantics*, Vol. 2, pp. 543–548, 2013.
- [39] R. Rajput and A. K. Solanki, “Review of Sentimental Analysis Methods Using Lexicon Based Approach”, *International Journal of Computer Science and Mobile Computing*, Vol. 5, No. 2, pp. 159–166, 2016.
- [40] Machine Learning, Available at: [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning), Last accessed on November 02, 2017.
- [41] P. K. Singh and M. S. Husain, “Methodological Study of Opinion Mining and Sentiment Analysis Techniques”, *International Journal on Soft Computing*, Vol. 5, No. 1, pp. 11–21, 2014.

- [42] N. Mehra, S. Khandelwal, and P. Patel, “Sentiment Identification Using Maximum Entropy Analysis of Movie Reviews”, 2002.
- [43] G. Vaitheeswaran, “Combining Lexicon and Machine Learning Method to Enhance the Accuracy of Sentiment Analysis on Big Data”, *International Journal of Computer Science and Information Technologies*, Vol. 7, No. 1, pp. 306–311, 2016.
- [44] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller, “Introduction to WordNet : An On-line Lexical Database”, *International Journal of Lexicography*, Vol. 3, No. 4, pp. 235-244, 1990.
- [45] A. Goyal and H. Daum, “Generating Semantic Orientation Lexicon Using Large Data and Thesaurus”, in *Proceeding of the 2<sup>nd</sup> Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 37- 43, 2011.
- [46] S. Mohammad, C. Dunne, and B. Dorr, “Generating High-Coverage Semantic Orientation Lexicons From Overtly Marked Words and a Thesaurus”, in *Proceeding of the Conference on Empirical Methods in Natural Language Processing*, pp. 599–608, August 2009.
- [47] W. Medhat, A. Hassan, and H. Korashy, “Sentiment Analysis Algorithms and Applications : A Survey”, *Journal of Ain Engineering*, Vol. 5, No. 4, pp. 1093–1113, 2014.
- [48] S. M. Kim and E. Hovy, “Determining the Sentiment of Opinions”, in *Proceeding of the COLING Conference*, 2004.
- [49] Language, Available at: <https://en.wikipedia.org/wiki/Language>, Last Accessed on August 19, 2017.
- [50] Baye Yimam, *የአግሮኛ ሰዋሰድ*, 2<sup>nd</sup> Edition, Mega Printing and Enterprise, Addis Ababa, 2002 (EC).
- [51] Daniel Teklu, *ዘበናዊ ሰዋሰድ ቋንቋ ጉግሮኛ*, 4<sup>th</sup> Edition, Mega Printing and Enterprise, Addis Ababa, 2002 (EC).
- [52] A. Carstairs-Mccarthy, *An Introduction to English Morphology*, Edinburgh University Press, Endinburgh, 2002.

- [53] Amharic Studies, Available at: <https://languages.ufl.edu/academics/llc-languages/amharic-studies/>, Last Accessed on December 10, 2017.
- [54] Martha Yifiru Tachbelie, “Morphology-Based Language Modeling for Amharic”, Dissertation to Obtain the Degree of Doctor of Science, Department of Computer Science, Hamburg University, August 2010.
- [55] Wolf Leslau, *Reference Grammar of Amharic*, Harrassowitz Wiesbaden, 1995.
- [56] Tigrigna, Available at: <http://aboutworldlanguages.com/tigrinya>, Last Accessed on December 10, 2017.
- [57] Origin of Amharic, Tigrinya and Tigre Languages, Available at: <http://orvillejenkins.com/languages/tigretigrinyageez.html>, Last Accessed on December 10, 2017.
- [58] English Language, Available at: [https://simple.wikipedia.org/wiki/English\\_language](https://simple.wikipedia.org/wiki/English_language), Last Accessed on November 06, 2017.
- [59] A. Fabricius, *The Handbook of English Linguistics*, Blackwell Publishing Ltd, UK, 2006.
- [60] F. M. Kundi, A. Khan, S. Ahmad, and M.Z. Ashhar, “Lexicon-Based Sentiment Analysis in the Social Web”, *Journal of Basic and Applied Scientific Research*, Vol. 4, No. 6, 2014.
- [61] X. Wan, “Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis”, in *Proceeding of the Conference on Empirical Methods in Natural Language Processing*, pp. 553–561, 2008.
- [62] S. Volkova, T. Wilson, and D. Yarowsky, “Exploring Sentiment in Social Media : Bootstrapping Subjectivity Clues from Multilingual Twitter Streams”, in *Proceeding of the 51<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*, pp. 505–510, 2013.
- [63] M. Porter, “The Porter Stemming Algorithm, An ALgorithm for Suffix Stripping”, Vol. 14, No. 3, pp. 130-137, July 1980.
- [64] የኢትዮጵያ ቋንቋዎች ጥናትና ምርምር ማእከል, *አማርኛ መዝገበ ቃላት*, ኦርትዕቲክ ማተሚያ ቤት, Addis Ababa, 1993.

- [65] Amharic Dictionary, Available at: <http://www.amharicdictionary.com/>, Last Accessed on July 19, 2017.
- [66] J. Mason, *ሰዋሰው ትግርኛ*, the Read Sea Press, 1996.
- [67] Tigrinya-English Dictionary, Available at: <https://www.geezexperience.com/>, Last Accessed on June 10, 2017.
- [68] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, 'Reilly Media, Cambridge, 2009.
- [69] English Wordlist Available at: <http://www-01.sil.org/linguistics/wordlists/english/>, Last Accessed on September 23, 2017.
- [70] Biniam Gebremichael, Wordlist and Spell checking for Amharic and Tigrigna, Available at: <https://www.cs.ru.nl/~biniam/geez/crawl.php>, Last Accessed on September 23, 2017.
- [71] Opinion Lexicon or Sentiment Lexicon, Available at: <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>, Last Accessed on June 10, 2017.

# Appendix-1: Approval Letter for Linguistic Expertise

**አዲስ አበባ ዩኒቨርሲቲ**  
**በተፈጥሮ ሳይንስ ኮሌጅ**  
**የኮምፒዩተር ሳይንስ ትምህርት ክፍል**

**የመጠይቅ አላማ:**




በአዲስ አበባ ዩኒቨርሲቲ በኮምፒዩተር ሳይንስ ትምህርት ክፍል ለሁለተኛ ዲግሪ (ማስተርስ) ምርምር ማግኒያ የሚሆን በአንግሊዘኛ፣ በአማርኛ እና በትግርኛ ቋንቋ ላይ መሰረት ያደረገ ኮምፒዩተራዊ ዘዴ የሰዎች አስተዳደር መለያ ሲስተም ለመሰራት ግብአት የሚሆኑ ለአንግሊዘኛ 2006 አዎንታ እና 4783 አሉታ፣ ለአማርኛ 1009 አዎንታ እና 1264 አሉታ፣ ለትግርኛ 1663 አዎንታ እና 1122 አሉታ በአጠቃላይ ከ10,000 በላይ አዎንታ እና አሉታ የሚገልጹ ቃላት ተሰብስበዋል።

በቅድሚያ ለሚያደርጉልን ትብብር ከልብ አያመሰግንን፤ የእነዚህን ቃላት ትክክለኛነት እንዲያረጋግጡልን በትህትና እንጠይቃለን።

**የመጠየቅ አዘጋጅ:**

መብራህቱ ታደሰ በኮምፒዩተር ሳይንስ ትምህርት ክፍል የሁለተኛ ዲግሪ(ማስተርስ) ተማሪ

**ያረጋገጠው:**

1. አለሙድህን ተገ ብሔራዊ ቋንቋ የTEFL ንጥረት ክፍል የPh.D ተማሪ 
2. ደርጋ ለገሰ ገሰ ገሰ የአማርኛ ንጥረት ክፍል የPh.D ተማሪ 
3. ለደብረ ለደብረ የቋንቋ ማናን ፊደል ደብረ ለደብረ ገሰ ገሰ 

## Appendix-2: Sample English Positive Sentiment Words

|                |                |             |               |             |              |
|----------------|----------------|-------------|---------------|-------------|--------------|
| Abound         | appreciate     | Fair        | lovable       | Nice        | Success      |
| Abounds        | appreciated    | Fairly      | lovably       | Nicely      | Suffice      |
| Abundance      | appreciates    | fairness    | love          | Nicer       | Sufficed     |
| Abundant       | appreciative   | Faith       | loved         | Nicest      | Suffices     |
| Accessible     | appreciatively | Faithful    | lovely        | Nifty       | sufficient   |
| Acclaim        | appropriate    | Fame        | lover         | Nimble      | sufficiently |
| Acclaimed      | approval       | Famed       | loves         | Noble       | Suitable     |
| acclamation    | approve        | Famous      | loving        | Nobly       | sumptuous    |
| Accolade       | beautiful      | famously    | loyal         | Noiseless   | Super        |
| Accolades      | beautifully    | Fancier     | loyalty       | Notably     | Superior     |
| accommodative  | beautify       | Fancy       | lucid         | Nourish     | superiority  |
| accomplish     | beauty         | Fanfare     | lucidly       | Nourishing  | Supple       |
| accomplished   | beckon         | Fans        | luck          | Nourishment | Support      |
| accomplishment | beckoned       | fantastic   | luckier       | Novelty     | supportive   |
| Accurate       | beckoning      | fascination | luckiest      | Nurturing   | supports     |
| Accurately     | beckons        | fashionable | luckiness     | Obtainable  | supremacy    |
| Achievable     | believable     | fashionably | lucky         | Openly      | supreme      |
| achievement    | beloved        | Fast        | lucrative     | Openness    | supremely    |
| achievements   | benefactor     | Faster      | luminous      | Optimal     | surmount     |
| Acumen         | beneficent     | Fastest     | lush          | Optimism    | Surpass      |
| Adaptable      | beneficial     | faultless   | luster        | Optimistic  | Surreal      |
| Adaptive       | beneficially   | Favor       | lustrous      | Opulent     | Survival     |
| Adequate       | beneficiary    | favorable   | luxuriant     | Orderly     | Swank        |
| Adjustable     | benefit        | Favored     | luxuriate     | Originality | swankier     |
| Admirable      | benefits       | Favorite    | luxurious     | Outdone     | swankiest    |
| Admirably      | benevolence    | fearless    | luxuriously   | Outshine    | Swanky       |
| Admiration     | benevolent     | fearlessly  | luxury        | Outsmart    | sweeping     |
| Admire         | best           | feasible    | lyrical       | Outstrip    | Sweet        |
| Admirer        | better         | feasibly    | magic         | Outwit      | Sweetly      |
| Admiring       | blameless      | fecilitous  | magical       | Ovation     | sweetness    |
| Admiringly     | bless          | Feisty      | magnificent   | Overjoyed   | Swift        |
| Adorable       | blessing       | felicitate  | magnificently | Overtake    | swiftness    |
| Adore          | bliss          | Felicity    | majestic      | Overtaking  | Talent       |
| Adored         | blissful       | Fertile     | majesty       | Overtook    | Talented     |
| Adorer         | blissfully     | Fervent     | manageable    | Painless    | Talents      |
| Adoring        | blithe         | fervently   | marvel        | Painlessly  | Tantalize    |
| Adoringly      | blockbuster    | Fervid      | marveled      | Palatial    | tenacious    |
| Adroit         | bloom          | Fervidly    | marvelous     | Pamper      | tenaciously  |
| Adroitly       | blossom        | Fervor      | marvelously   | Pampered    | Tenacity     |
| Adulate        | bolster        | Festive     | marvels       | Pampers     | Tender       |
| Adulation      | bonny          | Fidelity    | master        | Panoramic   | Tenderly     |
| Adulatory      | bonus          | Fiery       | masterful     | Paradise    | Terrific     |
| Advanced       | bonuses        | Fine        | masterfully   | Paramount   | terrifically |
| Advantage      | boom           | Finely      | masters       | Passion     | Thank        |

|                |             |              |             |              |             |
|----------------|-------------|--------------|-------------|--------------|-------------|
| advantageous   | booming     | Finer        | mastery     | Passionate   | Thankful    |
| advantageously | boost       | Finest       | matchless   | Passionately | Thinner     |
| advantages     | boundless   | Firmer       | mature      | Patience     | Thrilled    |
| adventurous    | bountiful   | Flashy       | maturely    | Patient      | Thrilling   |
| Advocate       | brainiest   | Flatter      | maturity    | Patiently    | thrillingly |
| Advocated      | brainy      | flattering   | meaningful  | Patriot      | Thrills     |
| Advocates      | brand-new   | flatteringly | memorable   | Patriotic    | Tidy        |
| Affability     | brave       | flawless     | merciful    | Peace        | Timely      |
| Affable        | bravery     | flawlessly   | mercifully  | Peaceable    | Tingle      |
| Affably        | breeze      | flexibility  | mercy       | Peaceful     | Titillate   |
| Affectation    | bright      | Flexible     | merit       | Peacefully   | titillating |
| Affection      | brighten    | Flourish     | meritorious | Peach        | tolerable   |
| affectionate   | brighter    | flourishing  | merrily     | Peerless     | Tough       |
| Affinity       | brightest   | Fluent       | merriment   | Peps         | Tougher     |
| Affirm         | brilliance  | Flutter      | merriness   | Perfect      | toughest    |
| Affirmation    | brilliances | Fond         | merry       | Perfection   | Traction    |
| Affirmative    | brilliant   | Fondly       | mesmerize   | Perfectly    | Tranquil    |
| Affluence      | brilliantly | fondness     | mesmerized  | Permissible  | tranquility |
| Affluent       | brotherly   | foolproof    | mesmerizes  | Perseverance | transparent |
| Afford         | bullish     | foremost     | mesmerizing | Persevere    | Treasure    |
| Affordable     | buoyant     | foresight    | miracle     | Personages   | Trendy      |
| Affordably     | cajole      | formidable   | miracles    | Personalized | Triumph     |
| Agile          | calm        | fortitude    | miraculous  | Phenomenal   | triumphal   |
| Agilely        | calming     | fortuitous   | modern      | phenomenally | triumphant  |
| Agility        | calmness    | fortuitously | modest      | Picturesque  | Trivially   |
| Agreeable      | capability  | fortunate    | modesty     | Piety        | Trophy      |
| agreeableness  | capable     | fortunately  | momentous   | Playful      | Trump       |
| Agreeably      | capably     | Fortune      | morality    | Pleasant     | Trumpet     |
| Alluring       | captivate   | fragrant     | motivated   | Pleasantly   | Trust       |
| Alluringly     | captivating | Free         | neat        | Pleased      | Trusted     |
| Altruistic     | catchy      | freedom      | unity       | Pleases      | Trusting    |
| altruistically | celebrate   | freedoms     | unlimited   | Pleasure     | trustingly  |
| Amaze          | celebrated  | Fresh        | useful      | Plentiful    | Trusty      |
| Amazed         | celebration | friendliness | valuable    | Pluses       | Truthful    |
| amazement      | concise     | friendly     | victory     | Polite       | truthfully  |
| Amazing        | confidence  | Fun          | wealthy     | Verifiable   | versatile   |

### Appendix-3: Sample English Negative Sentiment Words

|             |           |             |             |             |              |
|-------------|-----------|-------------|-------------|-------------|--------------|
| Abnormal    | babble    | Careless    | damage      | elimination | Negate       |
| Abolish     | backache  | Caricature  | damn        | emaciated   | Neglect      |
| Abort       | backbite  | Carnage     | damnable    | emasculate  | Negligence   |
| Aborted     | backward  | Carp        | damnation   | embarrass   | Negligent    |
| Aborts      | Bad       | Castigate   | damned      | embattled   | Nemesis      |
| Abrade      | baffle    | Casualty    | dampner     | embroil     | Nepotism     |
| Abrasive    | Bait      | Cataclysm   | danger      | emergency   | Nervous      |
| Abrupt      | Balk      | Cave        | dark        | emphatic    | Nettle       |
| Abscond     | banal     | Censure     | dastard     | emptiness   | Neurotic     |
| Absence     | Bane      | Chafe       | daunt       | encroach    | Neurotically |
| Absurd      | banish    | Chaff       | dawdle      | endanger    | Niggle       |
| Abuse       | barren    | challenging | daze        | enemies     | Niggles      |
| Abusive     | baseless  | Chaos       | dead        | enemy       | Nightmare    |
| Accidental  | Bash      | Chaotic     | deaf        | enervate    | Nitpick      |
| Accost      | bashful   | Chasten     | dearth      | enfeeble    | Nitpicking   |
| Accursed    | bashing   | Chastise    | death       | enflame     | Noise        |
| Accusation  | bastard   | Chatter     | debacle     | engulf      | Noises       |
| Accuse      | bearish   | Cheap       | debase      | enjoin      | Noisier      |
| Accusing    | bestly    | Cheat       | debaser     | enmity      | Noisy        |
| Acerbate    | bedlam    | Cheesy      | debauch     | enrage      | Nonsense     |
| Acerbic     | befoul    | Chide       | debt        | enslave     | Nosey        |
| Ache        | Beg       | Childish    | debts       | entangle    | Notoriety    |
| Acrimony    | beggar    | Chill       | decay       | entrap      | Notorious    |
| Adamant     | beguile   | Chintzy     | decayed     | envious     | Noxious      |
| Addict      | belabor   | Choke       | deceit      | erase       | Nuisance     |
| Addicted    | belated   | Choleric    | deceitful   | erode       | Numb         |
| Addicting   | Belie     | Choppy      | deceive     | error       | Obese        |
| Addicts     | belittle  | Chore       | declaim     | evade       | Object       |
| Admonish    | bellicose | Chronic     | decline     | evil        | Oblique      |
| Adulterate  | bemoaning | Chunky      | decrement   | eviscerate  | Obliterate   |
| adversarial | bemused   | Clamor      | decrepit    | exacerbate  | Obliterated  |
| Adversary   | Bent      | Clash       | decry       | excessive   | Oblivious    |
| Adverse     | berate    | Clique      | defamation  | exclusion   | Obnoxious    |
| Adversity   | bereave   | Clog        | defamations | excoriate   | Obscene      |
| Afflict     | bereft    | Cloud       | defamatory  | excuse      | Obscenely    |
| Affliction  | berserk   | Clueless    | defame      | exhaustion  | Obscenity    |
| Afflictive  | beseech   | Clumsy      | defect      | exhausts    | Obscure      |
| Affront     | beset     | Clunky      | defensive   | exhort      | Obscured     |
| Afraid      | besiege   | Coarse      | defiance    | exile       | Obscures     |
| Aggravate   | besmirch  | Cocky       | defiant     | expensive   | Obscurity    |
| aggravating | bestial   | Coerce      | deficient   | expire      | Obsess       |
| Aggression  | betray    | Coercion    | defile      | explode     | Obsessive    |
| Aggressive  | betrayal  | Coercive    | defiler     | exploit     | Obsolete     |
| Aggressor   | betrayer  | Cold        | deform      | expulse     | Obstacle     |

|             |           |             |             |             |             |
|-------------|-----------|-------------|-------------|-------------|-------------|
| Aggrieve    | betrays   | Collapse    | defunct     | expunge     | Obstinate   |
| Aggrieved   | bewail    | Collude     | defy        | exterminate | Obstinately |
| Aghast      | beware    | Collusion   | degenerate  | extinguish  | Obstruct    |
| Agonies     | bewitch   | Combative   | degrade     | extort      | Obstructs   |
| Agonize     | Bias      | Combust     | deject      | fabricate   | Obtrusive   |
| Agonizing   | bicker    | Comical     | delay       | fabrication | Obtuse      |
| Agony       | bigotries | Complain    | delinquency | facetious   | Occlude     |
| Aground     | bigotry   | Complex     | delinquent  | facetiously | Occluded    |
| Ail         | Bitch     | Complicit   | delirious   | fail        | Occludes    |
| Ailing      | biting    | Conceit     | delirium    | fake        | Occluding   |
| Ailment     | bitter    | Concern     | delude      | fall        | Odd         |
| Aimless     | Blab      | Concerns    | deluge      | famine      | Offence     |
| Alarmed     | blabber   | Concession  | delusion    | fanatic     | Offend      |
| Alarming    | Blah      | Condemn     | demean      | farce       | Offender    |
| Alarmingly  | blame     | condemned   | demise      | farcical    | Offending   |
| Alienate    | Bland     | Condemns    | demolish    | fat         | Offenses    |
| Alienated   | blandish  | condescend  | demon       | fatal       | Offensive   |
| Alienation  | blasted   | Confess     | demonic     | fatigue     | Officious   |
| Allege      | blatant   | Confession  | demonize    | fault       | Ominous     |
| Allergic    | blather   | Conflict    | denial      | faze        | Ominously   |
| Allergies   | Bleak     | Conflicted  | denied      | fear        | Omission    |
| Allergy     | bleed     | Conflicting | denies      | fell        | Omit        |
| Aloof       | blemish   | Conflicts   | denigrate   | fever       | Onerous     |
| Altercation | Blind     | Confront    | denounce    | fibber      | Onerously   |
| Ambiguity   | blister   | Confuse     | dense       | fickle      | Onslaught   |
| Ambiguous   | bloated   | consequence | dent        | fiend       | Opinionated |
| ambivalence | blockage  | Contagious  | dented      | fissures    | Opponent    |
| ambivalent  | bloody    | contaminate | dents       | fist        | Oppose      |
| Ambush      | blotchy   | Contend     | deny        | flair       | Opposition  |
| Amiss       | Blow      | Contort     | deplore     | flak        | Oppositions |
| Amputate    | blunder   | contortions | deprave     | flake       | Oppress     |
| Anarchism   | Blunt     | Contradict  | depress     | flakey      | Oppression  |
| Anarchist   | Blur      | Contrive    | deprive     | flaking     | Oppressive  |
| Anarchy     | Blurt     | Corrode     | desert      | flaky       | Oppressors  |
| Anemic      | Bore      | Corrosion   | desolate    | flare       | Ordeal      |
| Anger       | break     | Corrosive   | despise     | flares      | Orphan      |
| Angry       | bribery   | Corrupt     | despised    | flaunt      | Ostracize   |
| Anguish     | bruise    | Costly      | despoil     | flaw        | Outcry      |
| Animosity   | brutal    | Crack       | despoiler   | overkill    | Outlaw      |
| Annihilate  | bumps     | Crash       | despondence | irritate    | Overact     |
| Annoy       | Bunk      | Craze       | despondency | isolate     | Overacted   |
| Anomaly     | burden    | Crazily     | despondent  | jam         | Overawe     |

### Appendix-4: Sample Amharic Positive Sentiment Words

|        |       |          |         |          |         |         |
|--------|-------|----------|---------|----------|---------|---------|
| ሀላፊነት  | መሰረት  | ሲማርኩ     | ቆንጆዎች   | ብልሆች     | ንፁሃት    | ብልህ     |
| ሀመልማል  | መሰናዳት | ሲማርክ     | በሀቅ     | ብልጠተኛ    | ንፁህ     | መለገስ    |
| ሀቀኛ    | መሳካት  | ሲሳይ      | በሀይለኛ   | ብልጠተኛዎች  | ንፁህ     | መላእክት   |
| ሀቀኛዎች  | መረዳዳት | ሲራክ      | በሀይል    | ብልጠት     | ንፁህነት   | መልማት    |
| ሀቅ     | መረጋጋት | ሲጣፍጡ     | በሀይማኖት  | ብልጣብልጥ   | ንፁሆች    | መልአክ    |
| ሀብታም   | መርካት  | ሲጣፍጥ     | በሀጋዊ    | ብልጥ      | ንፅህና    | መልከመልካም |
| ሀብታሞች  | መሸባረቅ | ሳቂታ      | በህግ     | ብልጦች     | አህዛብ    | መልካም    |
| ሀብት    | መበለህ  | ሳያመነታ    | በለወለተኞች | ብልፅግና    | አላማ     | መልካምነት  |
| ሀዋሪያ   | መበረከት | ስልት      | በልማት    | ብሩክ      | አላማችን   | መልካሞች   |
| ሀያል    | መበራከት | ስልጡን     | በመልካም   | ብራሾ      | አላማዎች   | መመስገን   |
| ሀይለኛ   | መበርታት | ስልጡኖች    | በመሰካት   | ብርሀን     | አልማዝ    | መማረክ    |
| ሃይለኛ   | መተጋገዝ | ስልጣኔ     | በማስተዋል  | ብርቅ      | አመስጋኝ   | መሰረታዊ   |
| ሀይል    | መታዘዝ  | ስልጣን     | በማእረግ   | ብርቅዮ     | አመስጋኞች  | ሰልጣኔ    |
| ሀይማኖተኛ | መታገስ  | ስመጥር     | በምህረት   | ብርቱ      | አመቺ     | ሰብአዊ    |
| ሀይማኖት  | መነቃቅያ | ስምም ነት   | በምርጥ    | ብርቱዎች    | አመነ     | ሰናይ     |
| ሁነኛ    | መንቃት  | ስርዓት ያለው | በምቶት    | ብርታት     | አመነች    | ሰኬታመኞች  |
| ህሊና    | መንፈሳዊ | ስርዓትያለው  | በሰላም    | ብርታቶኞች   | አመኔታ    | ሰኬታሞቹ   |
| ህዝባዊ   | መዋደድ  | ሰብቅል     | በሰፊው    | ብፁእ      | አማልክት   | ሰኬታሞች   |
| ህያው    | መድሀኒት | ሰብእና     | በሰፋፊ    | ተሀድሶ     | አማረ     | ሰኬቶች    |
| ህጋዊ    | መግባባት | ሰተና      | በሰኬት    | ተላልቅ     | አማኖች    | ሰወዳቹ    |
| ህግ     | መጠብብ  | ሰንዱ      | በረከተ    | ተልእኮ     | አሜን     | ሰጪ      |
| ሆደ ቡቡ  | መፅናት  | ሰንዱዎች    | በረከተች   | ተመራጭ     | አምላክ    | ሰፊ      |
| ሆደ ባሻ  | መፅናናት | ሰኬተኛ     | በረከተኛ   | ተመጣጣኝ    | አሰልጣኝ   | ሰፋፊ     |
| ለሀብታሞች | መፈቃቀር | ሰኬታማ     | በረከቱ    | ተሰጥአ     | አሰተማሪ   | ሰፋፊዎች   |
| ለቁምነገር | መፋጠን  | ሰኬታማዎች   | በረከታማ   | ተሳክቶልናል  | አሰተማሪዎች | ቅንነት    |
| ለበረታ   | መሁራን  | ሰኬታም     | በረከታም   | ተስማሚ     | አሰደናቂዎች | ቅኖች     |
| ለበጎ    | መሁራዊ  | ሰኬት      | በረከት    | ተስማሚ የሆነ | አሳመረ    | ቅዱሳን    |
| ለባም    | መሁሮች  | ስወዲሽ     | በረከቶች   | ተስማሚ የሆነ | አሳማኝ    | ቅዱስ     |
| ለባሞች   | መሉ    | ስወድክ     | በርህራሄ   | ተስፋ      | አሳኩ     | ቅዱሶች    |
| ለብመዋል  | መሩቃን  | ስያምር     | በሽጋዎች   | ተቀባይነት   | አሳካች    | ቅድስት    |
| ለተከበሩ  | መሩቃኖች | ስጦታ      | በቁምነገር  | ተባራከተ    | አስቂኝ    | ቅድስና    |
| ለተከበራ  | መቼዎች  | ረቂቅ      | በቂ      | ተነሳሽነት   | አስተማማኝ  | ቆራጥ     |
| ለተከበራቹ | መያተኛ  | ረኪቻለሁ    | በቅልጡፍ   | ተአማኒነት   | አስተምሮኛል | ቆንጆኛት   |
| ለተዋቡ   | መያዊ   | ረከተናል    | በበጎ     | ተአምረኛ    | አስተካካይ  | ቆንጆ     |
| ለትልቅ   | ሚና    | ረዥኸም     | በበጎነት   | ተአምር     | አስተዋይ   | ከምርጥ    |
| ለእውነት  | ሚዛን   | ረዥም      | በትልቅ    | ተአምራዊ    | አስተዋይነት | ከምቹ     |
| ለክቡራን  | ማለፍያ  | ረዥሞች     | በትክክል   | ተአምሮች    | አስተዋዮች  | ከሰላም    |
| ለወዳጅ   | ማመስገን | ረጋ       | በትጋት    | ተከላካይ    | አስታዋሽ   | ከሽጋዎች   |
| ለወዳጆች  | ማማሩ   | ሩህሩህ     | በቸሩ     | ተወዳጅ     | አስደማሚ   | ከቆንጆዎች  |
| ለዘላቂ   | ማማር   | ራእይ      | በቸርነቱ   | ተደስተናል   | አስደሳች   | ከበሬታ    |
| ለዛ     | ማስተዋል | ራእዮች     | በነጻነት   | ተድላ      | አስደናቂ   | ከቸሩ     |

|           |         |         |         |         |         |         |
|-----------|---------|---------|---------|---------|---------|---------|
| ለደግ       | ማስደነቁ   | ርህሩህ    | በንቃት    | ተገቢ     | አስገራሚ   | ከአሪፍ    |
| ለደጎች      | ማራኪ     | ርህሩህነት  | በንፁህ    | ተጋሰች    | አስገራሚዎች | ከእውነት   |
| ለገሰ       | ማራኪዎች   | ርህራሄ    | በአለማ    | ተግባቢ    | አስፈላጊ   | ከወዳጅነት  |
| ለገሰች      | ማርካት    | ሸጋ      | በአሪፍ    | ተጠቃሚ    | አሪፍ     | ከወዳጆች   |
| ለገናና      | ማርኮኛል   | ሸጋው     | በእምነት   | ተጨባጭ    | አሪፎች    | ከደህና    |
| ለጋሰ       | ማሸነፍ    | ሸግቱ     | በእውቀት   | ተጫዋች    | አራዳ     | ከጀግናዎች  |
| ለጋሰነት     | ማእረገኛ   | ሸግዩ     | በእውነት   | ተፈላጊ    | አርበኛ    | ከገናና    |
| ለጋሶች      | ማእረግ    | ሻጋዎች    | በእድል    | ታላቅ     | አርአያ    | ከጎበዝ    |
| ለጋሽ       | ማእረጎች   | ሽሙንሙን   | በእድገት   | ታላቆች    | አርኪ     | ከጎበዞች   |
| ለግሰናል     | ማእረጎኞች  | ቀላል     | በኩራት    | ታማኝ     | አርኪዎች   | ከጠበቡ    |
| ለጎሉ       | ማደግ     | ቀልጣፋ    | በወዳጅ    | ታታሪ     | አሸበረቀ   | ከጠናካራዎች |
| ለጎላ       | ምሁራዊ    | ቀራጭ     | በወዳጅነት  | ታዋቂ     | አሸብራቂ   | ከጥሩ     |
| ለጎበዞች     | ምሁር     | ቀሽት     | በወዳጆች   | ታዛዥ     | አሸብራቂዎች | ከፍተኛ    |
| ለጠንካራ     | ምህረተኛ   | ቀብራራ    | በውለታ    | ታድያስ    | አሸነፈ    | ኩሩ      |
| ለጠንካሮች    | ምህረት    | ቀኝ      | በዘላቂነት  | ታጋሽ     | አሸናፊ    | ኩራተኛ    |
| ለጥሩ       | ምሰጉን    | ቀያማ     | በደህና    | ታጋች     | አሸናፊዎች  | ኩራታም    |
| ለፀኑ       | ምሳሌ     | ቀያይ     | በደስተኞች  | ትሁት     | አቅል     | ኩራት     |
| ለፍቅር      | ምስገና    | ቀይ      | በደስታ    | ትህተኞች   | አቅም     | ክቡራን    |
| ለፍትህ      | ምስጋና    | ቀዮች     | በደግነት   | ትህትና    | አበርቺ    | ክቡር     |
| ሉባ        | ምሩቅ     | ቀጥታ     | በድፍረት   | ትልቅ     | አብነት    | ክብረታማ   |
| ልሙጥ       | ምርጫምርጥ  | ቁልጡፍ    | በጎ      | ትልቆች    | አናብስት   | ክብረታም   |
| ልማታዊ      | ምርጥ     | ቁልጣፋ    | በጎኞች    | ትብብር    | አንበሳ    | ክብረት    |
| ልማት       | ምርጦች    | ቁም ነገረኛ | በጎዎች    | ትእግስቶች  | አንበሳ    | ክብረኞች   |
| ልማቶች      | ምቸተኛ    | ቁምነገረኛ  | በጎፍቃድ   | ትእግስተኛ  | አንበሶች   | ክብር     |
| ልምላሜ      | ምቹ      | ቁምነገራም  | በጠቀመ    | ትእግስተኞች | አንድነት   | ወላድ     |
| ልምድ       | ምቶት     | ቁምነገር   | በጥሩ     | ትእግስት   | አንጠልጣይ  | ወሰክ     |
| ልባዊ       | ምቶት ያለው | ቁምነገሮች  | በፅዳት    | ትክክለኛ   | አእምሮ    | ወሳኝ     |
| ልብ        | ምክር     | ቁርጠኛ    | በፍቅር    | ትክክል    | አወቀ     | ወርቃማ    |
| ልብ የሚያደርስ | ምያ      | ቁርጠኛዎች  | በፍትህ    | ትጉ      | አወንታ    | ወርቃወርቅ  |
| ልብ የሚያደርስ | ምያተኛ    | ቁርጠኝነት  | በፍጥነት   | ትጉህ     | አወንታዊ   | ወርቅ     |
| ልእልና      | ምያተኛዎች  | ቁድሳኖች   | በፍፁም    | ትጋተኛ    | አዋቂዎች   | ወሮታ     |
| ልክ        | ምያተኞች   | ቁጥብ     | ባለሙያ    | ትጋት     | አዋኪ     | ወደል     |
| ልገሳ       | ምገስ     | ቁጥብነት   | ባለራእይ   | ትግል     | አዘኔታ    | ወዳጄ     |
| መለበሙ      | ሰላማዊ    | ቃል      | ባለራእዮች  | ቸር      | አዛኝ     | ወዳጅ     |
| መለበማቸው    | ሰላም     | ቅልጥፍና   | ባለቁምነገር | ቸርነት    | አዝናኝ    | ወዳጅነት   |
| መለበም      | ሰላምተኛ   | ቅነኛ     | ባለእውቀት  | ቻይ      | አይብ     | ወዳጆቹ    |
| መለበምዋ     | ሰልጡን    | ቅን      | ባለውለታ   | ችሎታ     | አይነተኛ   | ወዳጆች    |

## Appendix-5: Sample Amharic Negative Sentiment Words

|        |        |         |        |         |        |       |
|--------|--------|---------|--------|---------|--------|-------|
| ሃላቀር   | ሌብነት   | ሙሰኞች    | ስስታም   | ቅሌታም    | ተቃራኒ   | ከይሲ   |
| ሃኬተኛ   | ሌቦች    | ሙሰና     | ስስታም   | ቅሌጣን    | ተቃዋሚ   | ከዳተኛ  |
| ባልተኛ   | ልሞሽ    | ሙንኩራት   | ስስታም   | ቅማታም    | ተቃዋሚ   | ከፊፊይ  |
| ችኮ     | ልቅ     | ሙጥኝ     | ስራ አጥ  | ቅማታምነት  | ተቃዋሚዎች | ኩምትር  |
| አሻፊ    | ልቅሶ    | ሙጥኝ     | ስራ ፈት  | ቅራቅንባም  | ተቃውሞ   | ኩራተኛ  |
| ዝብርቅርቅ | ልክስክስ  | ሙጭጭ     | ስራአጥ   | ቅራኔ     | ተቅማጣም  | ኪሳራ   |
| ፊዘኛ    | ልክፍታም  | ማሰልቸት   | ስራፈት   | ቅሬታ     | ተቅማጣም  | ካሌች   |
| ሃላቀር   | ልዘብተኛ  | ማሳዘን    | ስርቅታ   | ቅር      | ተብታባ   | ክህደት  |
| ኃላቀሮች  | ልግመኛ   | ማሳፈር    | ስቃይ    | ቅቅታም    | ተተናኮለ  | ክልክል  |
| ሀሜተኛ   | ልግዝግዝ  | ማስገደድ   | ስቃይ    | ቅናታም    | ተተናኮለች | ክስ    |
| ሀሜተኞች  | ልግዝግዝ  | ማስጠንቀቅያ | ስብቅል   | ቅናት     | ተቺ     | ክፈ    |
| ሀሜታም   | መሀያሞች  | ማስጨነቅ   | ስንፍነት  | ቅንዝረኛ   | ተናደደ   | ክፉ    |
| ሀሜታሞች  | መሀይም   | ማርተኛ    | ስንፍና   | ቅንጣት    | ተናዳጅ   | ክፋተኛ  |
| ሀሜት    | መሀይምነት | ማሸንክ    | ስያስጠሉ  | ቅዠታም    | ተንካሽ   | ክፍ    |
| ሐሜት    | መለመን   | ማበድ     | ስያስጠላ  | ቅጡ      | ተንኮለኛ  | ኮስታራ  |
| ሀሜት    | መልክ ቢስ | ማንባት    | ስደተኛ   | ቅጣት     | ተንኮለኛ  | ኮርማታ  |
| ሃሞተ ቢስ | መልክ ጥፋ | ማጣት     | ስድ     | ቅፅበታዊ   | ተንኮለኞች | ኮተታም  |
| ሀሰተኛ   | መልክቢስ  | ማጭበርበር  | ስድብ    | ቅፅበት    | ተንኮል   | ኮተት   |
| ሀሰቱኛ   | መልክጥፋ  | ማፈርያ    | ስድነት   | ቆማጣ     | ተደባዳቢ  | ወሊዋይ  |
| ሀሰት    | መመናከር  | ማፈርያዎች  | ስዶቸ    | ቆራጣ     | ተጠራጣሪ  | ወላቃ   |
| ሀሰቶች   | መማቀቅ   | ምላሰኛ    | ስጋት    | ቆሻሻ     | ተጠቂ    | ወላዋይ  |
| ሀራም    | መሰናክል  | ምስቅልቅል  | ስግብግብ  | ቆሻሻም    | ተግራሪ   | ወልካፋ  |
| ሀራም    | መሰከር   | ምስቅልቅል  | ረሃብተኛ  | ቆሻሻነት   | ተፅእነኛ  | ወልጋዳ  |
| ሃብተ ቢስ | መሳሳት   | ምስኪን    | ረሃብተኞች | ቆሻሻዎች   | ቱባ     | ወሰላታ  |
| ሃብተቢስ  | መሳብብ   | ምቀኛ     | ረባሽ    | ቆንቃን    | ታማሚ    | ወረርሽኝ |
| ሃካይ    | መስሪ    | ምቅኝነት   | ረብሸኛ   | ቆፈን     | ታማሚዎች  | ወረተኛ  |
| ሃካይ    | መስነፍ   | ምቅኝነት   | ረብሻ    | ቆፍ      | ታራ     | ወረኛ   |
| ሃካዮች   | መስጠላት  | ምፀት     | ርሃብ    | ቋጣሪ     | ትምክህተኛ | ወረደ   |
| ሃኬተኛ   | መረን    | ሞላጫ     | ርሃብተኞች | ቦኃላቀር   | ትምክህት  | ወራዳ   |
| ሀዘን    | መሪር    | ሞልቃቃ    | ርካሽ    | ቦኃላቀርነት | ትረባ    | ወሬ    |
| ሀዘንተኛ  | መራራ    | ሞሽላቃ    | ሸረኛ    | ቦኃላቀሮች  | ትርምስ   | ወሬኛ   |
| ሃዘንተኛ  | መራር    | ሞሽላቃ    | ሸራፊ    | ቦሀሜት    | ትርፍራፊ  | ወሬኛ   |
| ሀዘንተኞች | መርዘኛ   | ሞኛሞኝ    | ሸርሙጣ   | ቦሀሰተኛ   | ትርፍራፊ  | ወቀሳ   |
| ሃጢአት   | መርዘኞች  | ሞኝ      | ሸርካታ   | ቦሀሰት    | ትችት    | ወንጀለኛ |
| ሁከት    | መርዛም   | ሞኝነት    | ሸቃባ    | ቦሀዘን    | ትችት    | ወንጀል  |
| ህሊና ቢስ | መርዛምነት | ሞኞች     | ሸባ     | ቦሀገወጥ   | ትንሽ    | ወከባ   |
| ህሊናቢስ  | መርዛሞች  | ሞዛዛ     | ሸንከኛ   | ቦሀገወጥነት | ትእቢተኛ  | ወገናዊ  |
| ህመም    | መርዝ    | ሞጃጃ     | ሸንከኛ   | ቦለጌ     | ትእቢተኛ  | ወፊፌ   |
| ህመምተኛ  | መርገም   | ሞጋጋ     | ሸከኛ    | ቦሉታ     | ትእቢተኞች | ዋልጌ   |
| ህመማን   | መሸወድ   | ሰለቸኝ    | ሸካካ    | ቦሌባ     | ትእቢታም  | ዋሽ    |
| ህቅታ    | መቀለድ   | ሰልቺ     | ሸክም    | ቦሌቦች    | ትእቢት   | ዋሸች   |
| ህገወጠኛ  | መቅኝነት  | ሰልቺ     | ሸፊፊ    | ቦልቃጣ    | ትእቢትነት | ውስብስብ |
| ህገወጠኞች | መቀኞች   | ሰስትጨክን  | ሸፋጥ    | ቦመርዝ    | ቸሌተኛ   | ውስን   |
| ህገወጥ   | መቃብር   | ሰቃጢ     | ሹጣም    | ቦመተናኮል  | ቸሌተኛ   | ውርደተኛ |

|        |        |         |         |        |       |       |
|--------|--------|---------|---------|--------|-------|-------|
| ህገወጦች  | መቃወም   | ሰቃጢ     | ሻገተ     | በመከራ   | ቸልተኛ  | ውርደታም |
| ሀፌረት   | መቅሰፊት  | ሰቅጣጭ    | ሻጋታ     | በሙሰና   | ቸጋራም  | ውርደት  |
| ህፍረት   | መቅዘፍት  | ሰቅጣጭ    | ሻጥረኛ    | በሙሰኞች  | ቸግረኛ  | ውርደቶች |
| ሆዳም    | መቅጫ    | ሰበራዎች   | ሻፈዳ     | በሞኝነት  | ቸላ ባይ | ውርጅብኝ |
| ሆዳሞች   | መባለግ   | ሰበበኛ    | ሽማግሌ    | በሰነፍ   | ቸላባይ  | ውሽታም  |
| ኋላቀር   | መተተኛ   | ሰበበኝነት  | ሽረኛ     | በሴራ    | ቸሰታ   | ውሽታም  |
| ለኋላቀር  | መተናኮል  | ሰበበኞች   | ሽረኛ     | በሰህተት  | ቸኩሌ   | ውሽታሞች |
| ለኋላቀሮች | መታመም   | ሰበብ     | ሽቁምቁም   | በሽብር   | ቸኩል   | ውሽት   |
| ለሀሜተኛ  | መታበት   | ሰበብተኞች  | ሽበት     | በሽተኛ   | ቸክቻካ  | ውሻ    |
| ለሃጫም   | መቸገር   | ሰበብነት   | ሽባ      | በሽተኞች  | ቸክቻካ  | ውሻምብር |
| ለመቀኛ   | መና     | ሰባራ     | ሽብር     | በሽታ    | ቸኮ    | ውድመት  |
| ለመቀኝነት | መናኛ    | ሰባቂ     | ሽብርተኛ   | በሽታም   | ቸኮሊ   | ዉሽታም  |
| ለመከራ   | መናጢ    | ሰነፍ     | ሽብርተኝነት | በቀለኛ   | ቸገራም  | ንቃሽ   |
| ለሙሰኞች  | መናጢዎች  | ሰነፎች    | ሽጎታም    | በቀሌ    | ቸጋራም  | ንትርክ  |
| ለሙሰና   | መናፈግ   | ሰከረች    | ሽፌታ     | በቂም    | ቸግረኛ  | ንዝንዝ  |
| ለማኝ    | መንቻክ   | ሰካራም    | ሸካካ     | በተጎከል  | ቸግረኞች | ንደት   |
| ለማኝነት  | መንቻክ   | ሰካራሞች   | ሸጣጣ     | በትአቢት  | ቸግር   | ንጭንጭ  |
| ለማኞች   | መንከፍ   | ሰያቱ     | ቀ ሽም    | በቸግር   | ነሁሊሊ  | ንጭጭ   |
| ለማጅ    | መንዛዛት  | ሰይጣን    | ቀላል     | በአደጋ   | ነሁላላ  | ንፈግ   |
| ለሞኝ    | መአት    | ሰይጣኖች   | ቀላጅ     | በውሸት   | ነቀፊ   | ንፋግ   |
| ለሞኞች   | መከረኛ   | ሰጨክኑ    | ቀላጅ     | በደሌ    | ነቃፊ   | ንፌግ   |
| ለሴራ    | መከራ    | ሰፍጠኛ    | ቀልደኛ    | በደል    | ነውረኛ  | ንፍግ   |
| ለቂም    | መከራዎች  | ሰፍጠኛ    | ቀልደኞች   | በድሀነት  | ነውረኞች | ንፍግ   |
| ለቃቃሚ   | መካናት   | ሱሰኛ     | ቀልድ     | በጎሪጥ   | ነውራም  | ንፍግነት |
| ለቃቃሚ   | መካን    | ሱሰኞች    | ቀማተኛ    | በጥላት   | ነውር   | ንፍጎች  |
| ለብላቢ   | መካንነት  | ሱሳም     | ቀማተኞች   | በጥባጭ   | ነውጠኛ  | አለመሆን |
| ለተቸገረ  | መካኖች   | ሱሳሞች    | ቀማኛ     | በጨለማ   | ነውጥ   | አለመሆን |
| ለተቸገረች | መዋረድ   | ሲተናኮሉ   | ቀማኞች    | በጨለማዎች | ነዝነዝ  | አለባሌታ |
| ለተቸገሩ  | መዋሸት   | ሲተናኮል   | ቀሳጢ     | በጭባጭ   | ነዝናዝ  | አለታዊ  |
| ለካፊ    | መዘዘኛ   | ሲያጨበርባራ | ቀሳፊ     | በጭባጭ   | ነዝናዝ  | አለክላኪ |
| ለገዝጋዛ  | መዘዝ    | ሲያጨበርብሩ | ቀርፋፋ    | በጭካኔ   | ነዳይ   | አላጋጭ  |
| ለግጠኛ   | መደብደብያ | ሲያጨበርብር | ቀርፋፋ    | ቡከን    | ነገረኛ  | አላግባብ |
| ለፈዳዳ   | መደናገር  | ሲጨክን    | ቀሽም     | ቡካን    | ነገረኛ  | አሳሳች  |
| ለፍላፊ   | መደንዘዣ  | ሳይሳካ    | ቀበጥ     | ባለጌ    | ነገረኛ  | አሳበበ  |
| ለፍሰፍሰ  | መደደብ   | ሴሰኛ     | ቀባጣሪ    | ባለጌዎች  | ነጣቂ   | አሳበበች |
| ለፍዳዳ   | መጋኛ    | ሴሰኛ     | ቀብቃባ    | ባላገር   | ነጭናጫ  | አሳበቡ  |
| ለፍጫቃ   | መጣጣ    | ሴሰኛ     | ቀነተኛ    | ባልተኛ   | ነፃ    | አሳባቂ  |
| ለቃሳም   | መጥፈዎች  | ሴረኛ     | ቀናተኛ    | ባልገዋል  | ነፈዝ   | አሳባቂ  |
| ለጢ     | መጥፋት   | ሴረኞች    | ቀውስ     | ባልጋለች  | ነፋጊ   | አሳዛኝ  |
| ላጠፉ    | መጥፎ    | ሴራ      | ቂጣ      | ባርነት   | ነፌናፊ  | አደገኛ  |
| ሌባ     | መጥፎነት  | ሴራነት    | ባእድ     | ባርያ    | ነፌጠኛ  | አደጋ   |

## Appendix-6: Sample Tigrigna Positive Sentiment Words

|        |         |         |         |         |        |        |
|--------|---------|---------|---------|---------|--------|--------|
| ጽቡቕ    | ለጋስነት   | ምዕቡል    | ርህሩህ    | ዕምበባ    | ወሓለ    | ዘደንቕ   |
| ጽቡቕቲ   | ለጋስ     | ማዕበለ    | ራህሪህ    | ዓምበበ    | ዉሒልና   | ተደነቕ   |
| ብጽቡቕ   | ለገሰ     | ማዕቢልና   | ራህሪህና   | ዓምቢብና   | ወሓለታት  | ተደነቕና  |
| ጽቡቃት   | ለጊስና    | ምዕቡላት   | ርህሩሃት   | ዓምቡብቲ   | ወሒለን   | ተደነቕቲ  |
| ጸቢቀ    | ለጋሳት    | ማዕቢለን   | ራህሪህን   | ዓምቢበን   | ወሒልኩም  | ተደነቕን  |
| ጸቢቕና   | ለጊስን    | ማዕቢልኩም  | ራህሪህኩም  | ዓምቢብኩም  | ወሒሎም   | ተደነቕኩም |
| ጸቢቕን   | ለጊስኩም   | ማዕቢሎም   | ራህሪህም   | ዓምቢበም   | ወሒላቶም  | ተደነቕም  |
| ጸቢቕኩም  | ለጊስክን   | ማዕቢላቶም  | ራህሪሃቶም  | ዓምቢባቶም  | ወሒላተን  | ተደነቕኩም |
| ጸቢቆም   | ለገስቲ    | ማዕቢላተን  | ራህሪሃተን  | ዓምቢባተን  | ወሒልክን  | ተደነቕተን |
| ጸቢቃቶም  | ክልግስ    | ማዕቢልክን  | ራህሪህክን  | ዓምቢብክን  | ወሕለይቲ  | ተደነቕክን |
| ጸቢቃተን  | ትልግስ    | ምዕብልቲ   | ርህርህቲ   | ዕምበባታት  | ወሓለት   | ተደነቆም  |
| ጸቢቕክን  | ክልግሱ    | ምምዕባሎም  | ረህርህነት  | ዓምቢበም   | ክውሕል   | ተደነቕ   |
| ክጽብቕ   | ክትልግስ   | ምምዕባላተን | ክርህርህ   | ዓምበበት   | ትውሕል   | ተደነቕት  |
| ትጽብቕ   | ክልግሳ    | ምምዕባላቶም | ትርህርህ   | ክዕምብብ   | ክውሕሉ   | ክተደነቕ  |
| ክትጽብቕ  | ክትልግሱ   | ምዕባለ    | ክርህርሁ   | ትዕምብብ   | ክትውሕል  | ከደንቕ   |
| ጽብቕነት  | ምልጋስኩም  | ምዕባልነት  | ክትርህርህ  | ክዕምብብ   | ክውሕላ   | ክተደነቕ  |
| ክጽብቁ   | ምልጋሳቶም  | ክምዕብል   | ክርህርሃ   | ክትዕምብብ  | ምውሓሎም  | ከደንቓ   |
| ክትጽብቁ  | ምልጋሳተን  | ክምዕብሉ   | ክትርህርህ  | ክዕምብባ   | ምውሓላተን | ድንቕ    |
| ክጽብቃ   | ምልጋሶም   | ክትምዕብል  | ምርህራህም  | ምዕምባበም  | ምውሓሎም  | ምድናቕም  |
| ምጽባቆም  | ምልጋሳተን  | ክምዕብላ   | ምርህራሃተን | ምዕምባበተን | ወሓለ    | ምድናቓተን |
| ምጽባቃተን | ምልጋሳቶም  | ክምዕብሉ   | ምርህራሃቶም | ዝዓምበበ   | ዝውሓለ   | ዘደነቕ   |
| ምጽባቃቶም | ዝለገሰ    | ትምዕብል   | ዝራህሪህ   | ዝዓምበበት  | ዝውሓለት  | ዘደነቕት  |
| ዝጸበቀ   | ዝለገሰት   | ዝማዕበለ   | ዝራህሪህት  | ዝዓምብብካ  | ዝውሓልካ  | ዘደነቕካ  |
| ዝጸበቀት  | ዝለገሰካ   | ዝማዕበለት  | ዝራህሪህካ  | ዝዓምብብኩም | ዝውሓልኩም | ዘደነቕኩም |
| ዝጸበቕካ  | ዝለገሰኩም  | ዝማዕበልካ  | ዝራህሪህኩም | ዝዓምብብክን | ዝውሓልክን | ዘደነቕክን |
| ዝጸበቕኩም | ዝለገሰክን  | ዝማዕበልኩም | ዝራህሪህክን | ዝዓምብብና  | ዝውሓልና  | ዘደነቕና  |
| ዝጸበቕክን | ዝለገሰና   | ዝማዕበልክን | ዝራህሪህና  | ዕውት     | ዋሓለ    | ዘገርም   |
| ዝጸበቕና  | መልክዕኛ   | ዝማዕበልና  | ቀሲና     | ተዓወተ    | ውሕልና   | ተገረመ   |
| ትሕጎስ   | መልክዐ    | ሰልጡን    | ቅሱን     | ተዓወትና   | ዋሕለለ   | ተገሪምና  |
| ተሓጎስ   | መልኪዕና   | ሰልጠነ    | ቀሰነ     | ተዓወትቲ   | ውሕሉል   | ተገረምቲ  |
| ሕጉስ    | መልክዕኛታት | ሰልጢና    | ቁስናት    | ትዕወት    | ውሕልነት  | ትገርም   |
| ሕጉሳት   | መልኪዕን   | ሰልጡናት   | ቀሲነን    | ተዓዊተን   | ዉዕወዕ   | ተገሪመን  |
| ሕጉስቲ   | መልኪዕኩም  | ሰልጢነን   | ቀሲነኩም   | ተዓዊትኩም  | ዋዕወዐ   | ተገሪምኩም |
| ተሓጊስ   | መልክዕ    | ሰልጢነኩም  | ቀሲኖም    | ተዓዊቶም   | ዋዕዊዕና  | ተገሪምም  |
| ተሓጊስና  | ክምልክዕ   | ሰልጢኖም   | ቀሲናቶም   | ተዓዊታቶም  | ዉዕዉዓት  | ተገሪማቶም |
| ተሓጊስን  | ትምልክዕ   | ሰልጢናቶም  | ቀሲናተን   | ተዓዊታተን  | ዋዕዊዐን  | ተገሪማተን |
| ተሓጊስኩም | ክምልክዕ   | ሰልጢናተን  | ቀሲነክን   | ተዓዊትክን  | ዋዕዊዕኩም | ተገሪምክን |
| ተሓጊሶም  | ክትምልክዕ  | ሰልጢነክን  | ቅሰንቲ    | ተዓወተ    | ዋዕዊዐም  | ተገረመት  |
| ተሓጊሳቶም | ክምልክዓ   | ሰልጥንቲ   | ቅሳነት    | ተዓወተት   | ዋዕዊዓቶም | ክተገርም  |
| ተሓጊሳተን | ክምልክዐ   | ሰልጥንነት  | ክቀሰን    | ክዕወት    | ዋዕዊዓተን | ክግረመ   |
| ተሓጊስክን | ምምልካዕኩም | ክሰልጥን   | ክቀሰኑ    | ክዕወቱ    | ዋዕወዕክን | ክትግረም  |
| ክሕጎስ   | ምምልካዓቶም | ትሰልጥን   | ክትቀሰን   | ክትዕወት   | ውዕወዕቲ  | ክግረማ   |
| ክሕጎሱ   | ምምልካዓተን | ክሰልጥኑ   | ክቀሰና    | ክዕወታ    | ውዕዋዐ   | ምግራሞም  |
| ክትሕጎስ  | መልኪዕክን  | ክትሰልጥን  | ክትቀሰን   | ዕውቲ     | ዋዕወዐት  | ምግራማተን |

|         |         |         |        |         |         |          |
|---------|---------|---------|--------|---------|---------|----------|
| ክሕጎሳ    | ምልኩዕቲ   | ክሰልጥና   | ምቅሳኖም  | ዕውታት    | ትውዕውዕ   | ምግራሞም    |
| ምሕጋሶም   | ምምልካዖም  | ክትሰልጥኑ  | ምቅሳናተን | ምዕዋቶም   | ክውዕዉዕ   | ዘገረመ     |
| ምሕጋሳተን  | ምምልካዓተን | ምስልጣኖም  | ምቅሳናቶም | ምዕዋታተን  | ክውዕውዑ   | ዘገረመት    |
| ምሕጋሳቶም  | ምምልካዓቶም | ምስልጣናተን | ዝቀሰነ   | ተዓዊቶም   | ክትውዕውዕ  | ዘገረምካ    |
| ታሕጋስ    | ዝመልክዐ   | ምስልጣናቶም | ዝቀሰነት  | ዝተዓወተ   | ክውዕውዓ   | ዘገረምኩም   |
| ታሕጋስነት  | ዝመልክዐት  | ዝሰልጠነ   | ዝቀሰነካ  | ዝተዓወተት  | ምውዕዋዖም  | ዘገረምክን   |
| ተሓጎሰ    | ዝመልካዕካ  | ዝሰልጠነት  | ዝቀሰነኩም | ዝተዓወትካ  | ምውዕዋዐን  | ዘገረመና    |
| ዝተሓጎሰ   | ዝመልካዕኩም | ዝሰልጠነካ  | ዝቀሰነክን | ዝተዓወትኩም | ምውዕዋዖም  | ዝሳነ      |
| ዝተሓጎሰት  | ዝመልክዕክን | ዝሰልጠነኩም | ዝቀሰና   | ዝተዓወትክን | ዝዋዕወዐ   | ተሳነዩ     |
| ዝትሓጎስካ  | ዝመልክዕና  | ዝሰልጠነክን | በሊጽና   | ዝተዓወትና  | ዝዋዕወዐት  | ተሳነና     |
| ዝትሓጎስኩም | ማራኺ     | ዝሰልጠና   | ብሉጽ    | እዱብ     | ዝዋዕወዕካ  | ተሳነኩም    |
| ዝትሓጎስክን | ማረኽ     | ስምረት    | በለጸ    | አደበ     | ዝዋዕወዕኩም | ተሳነዩን    |
| ዝትሓጎስና  | ማሪኽና    | ሰመረ     | በሊጽና   | አዲብና    | ዝዋዕወዕክን | ተሳነዩም    |
| ልብምነት   | ማራኺት    | ሰሚርና    | ብሉጽት   | አዲብኩም   | ዝዋዕወዕና  | ተሳነዩቶም   |
| ለባም     | ማሪኽን    | ሰሙራት    | በሊጸን   | እድብቲ    | ውዕዋዐ    | ተሳነዩተን   |
| ለባማት    | ማሪኽም    | ሰሚረን    | በሊጽኩም  | እዱባት    | ዉቁብ     | ተሳነክን    |
| ትልብም    | ማሪኽን    | ሰሚርኩም   | በሊጸም   | ክእደቡ    | ውቃብ     | ተሳናይት    |
| ክልብም    | ማራኽነት   | ሰሚርም    | በሊጸቶም  | ክእደብ    | ውቁባት    | ተሳነዩ     |
| ክልብሙ    | ክማርኽ    | ሰሚራቶም   | በሊጸተን  | ክትእደብ   | ዉቅብቲ    | ተሳነዩት    |
| ክትልብም   | ክማርኽ    | ሰሚራተን   | በሊጽክን  | ክእደባ    | ወቂብን    | ትሳነ      |
| ክልብሙ    | ክትማርኽ   | ሰሚርክን   | ብልጽቲ   | ተአዲብን   | ወቁብኩም   | ክሳነ      |
| ክልብማ    | ክማርኽ    | ሰምርቲ    | ክበልጽ   | ተአዲብኩም  | ወቂብም    | ክሳነዩ     |
| ለቢመ     | ክማርኽ    | ሰምርነት   | ትበልጽ   | ተአዲብም   | ወቁባቶም   | ክትሳነይ    |
| ለቢምና    | ምምራኽ    | ክሰምር    | ክበልጹ   | ተአዲባቶም  | ወቂባተን   | ክሳነዩ     |
| ለቢመን    | ምምራኽቶም  | ክሰምሩ    | ክትበልጸ  | ተአዲባተን  | ወቁብክን   | ክሳነዩ     |
| ለቢምኩም   | ምምራኽተን  | ክትሰምር   | ክበልጸ   | ተአዲብክን  | ውቅብቲ    | ምስናዩም    |
| ለቢምም    | ማራኺት    | ክሰምራ    | ክበልጹ   | ትእደብ    | ክውቅቡ    | ምስናዩተን   |
| ለቢማቶም   | ምምራኽም   | ክትሰምሩ   | ምብላጸም  | ተአዲብም   | ዉቃብ     | ምስናዩም    |
| ለቢማተን   | ምምራኽን   | ምስማርም   | ምብላጸተን | ተአደበ    | ውቅብና    | ዘስማዕማዕ   |
| ለቢምክን   | ምምራኽቶም  | ምስማራተን  | ምብላጸቶም | ተአደበት   | ክውቁብ    | ተስማዕሚዐ   |
| ለበምቲ    | ዝማረኽ    | ምስማራቶም  | ብልጹነት  | ምእደባም   | ትውቅብ    | ተስማዕሚዕና  |
| ምልባሞም   | ዝማረኽት   | ዝሰመረ    | ዝበለጸ   | ምእደባተን  | ክውቁቡ    | ተስማዕኩም   |
| ምልባማተን  | ዝማረኽካ   | ዝሰመረት   | ዝበለጸት  | ዝተአደበ   | ክትውቁብ   | ትስማዕማዕ   |
| ምልባማቶም  | ዝማረኽኩም  | ዝሰመርካ   | ዝበለጸካ  | ዝተአደበት  | ክውቅባ    | ተስማዕሚዐን  |
| ዝለበመ    | ዝማረኽክን  | ዝሰመርኩም  | ዝበለጸኩም | ዝተአደብካ  | ምውቃብም   | ተስማዕሚዕኩም |
| ዝለበመት   | ዝማረኽና   | ዝሰመርክን  | ዝበለጸክን | ዝተአደብኩም | ምውቃብን   | ተስማዕሚዐም  |
| ዝለበምካ   | ምቹእ     | ዝሰመርና   | ዝበለጸና  | ዝተአደብክን | ምውቃብም   | ተስማዕሚዕኩም |
| ዝለበምኩም  | መቹእ     | ርቡሕ     | በሪህና   | ዝተአደብና  | ዝወቀበ    | ተስማዕሚዓተን |
| ዝለበምክን  | መቹእና    | ትረብሕ    | ብሩህ    | ኩርዕ     | ዝወቀበት   | ተስማዕሚዕክን |
| ዝለበምና   | ምቹእት    | ረብሐ     | በርህ    | ከርዐ     | ዝወቀብካ   | ክስማዕማዕ   |
| ለዋህ     | መቹእን    | ረብሐ     | በሪህና   | ኮሪዕና    | ዝወቀብኩም  | ክስማዕመዑ   |
| ለወሀ     | መቹእኩም   | ረብሐና    | ብሩሃት   | ኩርዓት    | ዝወቀብክን  | ክትስማዕማዕ  |
| ለዊህና    | መቹእም    | ርቡሐት    | በሪህን   | ክኮርዕ    | ዝወቀብና   | ክስማዕመዓ   |
| ለዋሃት    | መቹእቶም   | ረቢሐን    | በሪህኩም  | ክኮርዑ    | ዘዕግብ    | ክስማዕመዑ   |
| ክልውህ    | መቹእተን   | ረቢሕኩም   | በሪህም   | ክትኮርዕ   | አዕገበ    | ምስምማዖም   |

## Appendix-7: Sample Tigrigna Negative Sentiment Words

|         |        |         |         |        |        |        |
|---------|--------|---------|---------|--------|--------|--------|
| ኪሕዶም    | ኮሪዬ    | መሕዘኒ    | መከራ     | ሰረዘ    | ደንቆሮ   | ጽሉልቲ   |
| ኪሕደን    | ኮሪይት   | መሕዘኒት   | መከራታት   | ሰረዘት   | ደንቆረ   | ፈዛዝ    |
| ካሕዳም    | ኮሪይና   | ከሕዝኑ    | መውት     | ሰረዙ    | ደንቆረት  | ፈዘዘ    |
| ክሕዳቶም   | ኮሪያቶም  | ከተሕዝን   | ዝሞተ     | ሰሪዘን   | ደንቁረን  | ፈዚዞም   |
| ክሕዳተን   | ኮሪያተን  | ከሕዝና    | ዝሞተት    | ሰሪዞም   | ዘደንቆረ  | ፈዛዛት   |
| ካሕዳማት   | ክኩሪይ   | ከሕዝኑ    | ዝሞቱ     | ምስራዝ   | ዝደንቆረት | ፈዚዘን   |
| ክፉእ     | ኮረየት   | አሕዘንቲ   | ዝሞታ     | ሰቓይ    | ዝደንቆሩ  | ዝፈዘዘ   |
| ክፍአ     | ክትኩሪይ  | መሕዘኖ    | መድረቢዳ   | አሳቐየ   | ዝደንቆርና | ዝፈዘዘት  |
| ከፊአም    | ክክርያ   | አሕዚኖም   | መድረባዳታት | አሳቐየት  | ደንቆረታት | ዝፈዘዙ   |
| ከፊአን    | ኮሪይቲ   | አሕዚነን   | መጻጽ     | ሰቐያት   | ድንቁርና  | ዝፈዘዘ   |
| ከፊአቶም   | ምክራየን  | አሕዚነና   | መጻጽቲ    | ዝተሳቐየ  | ዲያብሎስ  | ፋሕሻዋት  |
| ከፊአተን   | ምክራዮም  | አሕዚናተን  | ክመጽጽ    | ዝተሳቐየት | ዲያብሎሳት | ፋሕሻው   |
| ከፊአና    | ምክራያቶም | ሕዙን     | መጻጻ     | ተሳቐየ   | ድንዘዝ   | ፋሕሸው   |
| ዝከፍአ    | ምክራያተን | ሕዝንቲ    | ዝመጻጸ    | ሰነፍ    | ደንዘዘ   | ፋሕሸዎም  |
| ዝከፍአት   | ዝኮረየ   | አሕዘነ    | ማእሰርቲ   | ሰነፈ    | ደንዘዘት  | ፋሕሸውን  |
| ዝከፍአ    | ዝኮረያ   | ዘሕዝን    | ተአሰረ    | ሰነፈት   | ደንዚዞም  | ዝፋሕሸው  |
| ዝከፍአ    | ዝኮረዩ   | ሓደገኛ    | ተአሰረት   | ሰነፋት   | ደንዚዘን  | ዝፋሕሸውት |
| ክፋአት    | ዝኮረይና  | ሓደጋ     | ዝተአሰረ   | ሰነፍና   | ዱንዘዛት  | ፎካሰ    |
| ትከፍአ    | ዝኮረይኩም | ብሓደጋ    | ዝትአሰረት  | ሰነፎም   | ደንዚዛተን | ፎኪሰ    |
| ክትከፍአ   | ዝኮረይክን | ሕልኸልኸ   | ዝተአሰሩ   | ሰነፈን   | ደንዚዛቶም | ፎከሰ    |
| ክፉአት    | ሃሰየ    | መሕለኸለኸ  | ዝተአሰራ   | ዝሰነፈ   | ዝደንዘዘ  | ፎከሰቲ   |
| መክፈኢ    | ሃሰየት   | ዘሕለኸልኸ  | መእሰርቲ   | ዝሰነፈት  | ዝደንዘዘት | ፎካሳት   |
| ዘክፍአ    | ሃሰይና   | ሕልኸልኸነት | እሱራት    | ዝሰነፉ   | ዝደንዘዙ  | ዝፎከሰ   |
| ክፉአት    | ሃሰይኩም  | ሕማቕ     | እሱር     | ሰነፎም   | ዝደንዘዛ  | ዝፎከሱ   |
| ወንጀለኛ   | ሃሰይክን  | ሓመቐ     | ምምራር    | ሰነፈን   | ደነዝ    | ፎኪሰን   |
| ወንጀል    | ሃሰያተን  | ሓመቐት    | አማረረ    | ሰነፋቶም  | ድንገት   | ዝፎከሳ   |
| ወንጀለኛታት | ሃሰያቶም  | ሓሚቕም    | አማረረት   | ሰነፋተን  | ድንገተኛ  | ጠሊማቶም  |
| ወደቐ     | ክሃሰያ   | ሓሚቕን    | አማረርና   | ዝሰነፈ   | ብድንገት  | ዝጠለሙ   |
| ወዲቐ     | ክሃሰዩ   | ሕማቓት    | ከማርር    | ዝሰነፈት  | ገበን    | ዝጠለማ   |
| ወዲቕና    | መሀሰይቲ  | ዝሓመቐ    | ከተማርር   | ሰአነ    | ገበነኛ   | ጥልመት   |
| ወደቐት    | ሃሰይቲ   | ዝሓመቐት   | መማረርቲ   | ሰኡን    | ገበነኛታት | ጠላሚ    |
| ወደቐ     | ዝሃሰየ   | ዝሓመቐ    | ዘማርሩ    | ሰኡናት   | ጉሁይ    | ጠለምቲ   |
| ወዲቕም    | ዝሃሰየት  | ዝሓመቐና   | ዘማርራ    | ዝሰአነ   | ጉሁያት   | ጠሊመ    |
| ወዲቓቶም   | ዝሃሰይኩም | ዝሓመቐክን  | አማረርም   | ዝሰአነት  | ጎሃየ    | ጠሊማ    |
| ወዲቓተን   | ዝሃሰይክን | ዝሓመቐኩም  | አማረርን   | ሰከራም   | ጎሃየት   | ጠላም    |
| ዝወደቐ    | ዝሃሰይና  | ብሕማቕ    | ከማርራ    | ሰከረ    | ጉሒላ    | ጥምየት   |
| ዝወደቐት   | ሃሰዩ    | ሕሱም     | ምርባጽ    | ሰከረት   | ጉሒላታት  | ጠመየ    |
| ዝወደቐ    | ሃሰየት   | ሕሱማት    | ተረበጸ    | ሰካራማት  | ዝጎሐለ   | ጠመየት   |
| ዝወደቐ    | ሃሳስ    | ሕሱምቲ    | ተረበጸት   | ዝሰከረ   | ዝጎሐለት  | ጥሙያት   |
| ውደቐት    | ሃሳሳት   | ክትሓስም   | ተረቢጽና   | ዝሰከረት  | ዝጎሐሉ   | ጠሚይና   |
| ውግእ     | ክሃሱ    | ክሓስማ    | ክትርበጽ   | ዝሰከር   | ዝጎሐላ   | ጠሚየን   |
| ተዋግእ    | ካሃሳ    | ሓሳማት    | ክርበጸ    | ሰይጣን   | ጉሒሎም   | ጠሚዮም   |
| ተዋጊአም   | ምሀሳሰን  | ክሓስም    | ርቡጻት    | ሰይጣናት  | ጉሒሊን   | በደለኛ   |
| ተዋጊአን   | ምሀሳሰም  | ክሓስሙ    | ርብጽቲ    | ሰያጥን   | ጉድላት   | በደለ    |

|         |        |        |        |        |        |        |
|---------|--------|--------|--------|--------|--------|--------|
| ተዋጊኦቶም  | ምህሳሳቶም | ዝሓሰመ   | ዝተረበጸ  | ረሳሕ    | ጎድሎ    | በደል    |
| ተዋጊኦተን  | ምህሳሳተን | ዝሓሰመት  | ዝተረበጸት | ረስሐ    | ጎደሎታት  | ተበደለ   |
| አዋገአ    | አህሱስ   | ዝሓሰምና  | ክርበጽ   | ረስሐት   | ዝጎደለ   | በዲለን   |
| አዋገኦት   | ዝሃሰሰት  | ምሕሳምም  | ክርበጹ   | ረሳሓት   | ዝጎደለት  | በዲላቶም  |
| አዋገእና   | ዝሃሰሰ   | ምሕሳመን  | ተረቢጸም  | ዝረስሐ   | ዝጎደሎ   | በዲላተን  |
| አዋገኦን   | ዝሃሰሱ   | ምሕሳማና  | ተረቢጸን  | ዝረስሐ   | ጉድኦት   | ዝበደለ   |
| አዋገኦቲ   | ዝሃሰሰና  | ሕሱር    | ምሽቓል   | ክረስሐ   | ጎድኦ    | ዝበደለት  |
| ዘሕፍር    | ዝሃሰስክን | ሕሳር    | ተሻቐለ   | ክረስሐ   | ተጎደኦ   | ተቓወመ   |
| አሕፊረ    | ሓራቕ    | ሕሰረት   | ተሻቐለት  | ረበሸ    | ተጎድኦት  | ተቓዋሚ   |
| አሕፊረት   | ከሕርቐ   | ዝሓሰረ   | ተሻቐልና  | ረባሺ    | ተጎዲኦም  | ተቓወሞ   |
| መሕፊረ    | ከሕርቓ   | ዝሓሰረት  | ተሻቓሊት  | ረበሸቲ   | ተጎዲኦን  | ተቓወምቲ  |
| ዘሕፍሩ    | መሕረቐቲ  | ዝሓሰርና  | ክትሻቐል  | ረበሸት   | ተጎዲኦቶም | ተካል    |
| ዘሕፊራ    | ዝሓረቐ   | ሓሳር    | መሻቐሊ   | ረቢሽና   | ተጎዲኦተን | ተካላት   |
| ዘሰቅቐ    | ዝሓረቐና  | ሕሳራት   | ዝተሻቐለ  | ረቢሽን   | ዝተጎድኦ  | ክትክላ   |
| መሰቀቐ    | ዝሓረቐክን | ዝሓሰረ   | ዝተሻቐለት | ረባሺት   | ዝተጎድኦት | ተጸላኢ   |
| ሰቀቀን    | ዝሓረቐኩም | ዝሓሰረት  | ዘሻቐል   | ረባሾ    | ዝተጎድኦ  | ጸልኦ    |
| ሰቐቐ     | ዘሕርቐ   | ብዝሓሰረ  | ምዕማጽ   | ራብሻ    | ዝተጎድኦ  | ጸላኢ    |
| ዘስካሕክሕ  | ዘሕርቐ   | ሕሱራት   | ዓመጽ    | ዝረበሸ   | ጌጋ     | ጸልኢ    |
| መስካሕክሒ  | ሓርፋፍ   | ሕርቃን   | ዓመጸ    | ተረበሸ   | ጌጋታት   | ጸላኦቲ   |
| አስክሕክሒ  | ሓርፋፋት  | ሓረቀ    | ዓመጸት   | ተረብሸና  | ግጉይ    | ተጸራሪ   |
| ዘስካሕክሓ  | ክሕርፍፍ  | ሓረቀት   | ዓሚጽና   | ርጉም    | ጎሓፍ    | ተጸረረ   |
| መስካሕክሕቲ | ክትሕርፍፍ | ሓረቁ    | ተዓመጸ   | ረገመ    | ጎሓፋት   | ተጸረርቲ  |
| ዘቃጽል    | ምሕርፋፎም | ሓሪቆም   | ተዓመጸት  | ረገመት   | ጎዳኢ    | ተጸሪሮም  |
| ቃጽሎ     | ምሕርፋፊን | ሓሪቃተን  | ዝትዓመጸ  | ረገምና   | ጎድኦ    | መርዚ    |
| መቃጸሊ    | ምሕርፋፍና | ከሕርቐ   | ዝትዓመጸት | ዝተረገመ  | ጎዲኦም   | መርዛም   |
| መቃጸልቲ   | ሓርፊፊ   | ክተሕርቐ  | ዝተዓመጹ  | ዝተረገመት | ጎድኦት   | ሃካይት   |
| አቃጸላት   | ሓርፊፊት  | ከሓርቁ   | ዝተዓመጸ  | ርጉማት   | ጎዲኦን   | ታህኪይት  |
| አቃጸላ    | ዝሓርፊፊ  | ከሓርቃ   | ዝተዓመጽና | ረገምም   | ጎዲኦቶም  | ተሃንጢዮም |
| አቃጸሎም   | ዝሓርፊፊት | አሕሪቆም  | ዓማጺ    | ረገመን   | ጎዲኦተን  | ተሃንጢዮን |
| አቃጸላን   | ሃንደበታዊ | አሕሪቀን  | ቀናኦ    | ሸፋጢ    | ጎዲኦና   | በዓለገ   |
| ዝንጉዕ    | ሃንደበት  | አሕረቀ   | ቀንኦ    | ሸፊጠ    | ዝጎድኦ   | ባርያ    |
| ዘንገዐ    | ብሃንደበት | ሓሪቃቶም  | ቀንኦት   | ሸፊጢት   | ዝጎድኦት  | ሸጉራት   |
| ዘንገዐት   | ሃካይ    | አሕረቀት  | ቀንኢ    | ሸፋጠ    | ዝጎድኦ   | መሸገርቲ  |
| ዝንጉዓት   | ተሃኪይና  | አሕረቐቲ  | ዘቅንኦ   | ሸፊጥ    | ዝጎድኦ   | ጠሊመን   |
| ደምሰሰ    | ተሃኪዮም  | ዘሕርቐ   | ቀናኦት   | ሸጉር    | ዝጎድኦና  | ጠሊማተን  |
| ደምሰሰት   | ተሃኪዮን  | መሕረቐቲ  | ቅንኦት   | ተሸገረ   | ጎዳኦቲ   | ከሓደት   |
| ለመነ     | ህኩያት   | ህንጡይ   | ቀጢን    | ተሸገረት  | ጉድኦት   | ኪሕደን   |
| ለማኒ     | ዝተሃከዩ  | ተሃንጠዩ  | ቀጣን    | ተሸገረና  | ጠለመ    | ክሕደት   |
| ለማኒት    | ዝተሃከዩት | ተሃንጠዩት | ቐጡዕ    | ተሸገረም  | ጠሊምና   | ዝከሓዱ   |
| ለመንቲ    | ዝተሃከዩ  | ተሃንጢይና | ቐጣዐ    | ሸግር    | ጠሊምም   | ዝከሓዱ   |

## Appendix-8: Inflection of Tigrigna Sentiment Words

| Word | Inflection | Analysis  | Latin       | Prefix | Stem  | Suffix | Root | Root  |
|------|------------|-----------|-------------|--------|-------|--------|------|-------|
| ጸቡቕ  | ጸቡቕ        | ጸቡቕ       | Sbuq        | -      | Sbuq  | -      | Sbq  | ጸ-ቡ-ቕ |
|      | ጸቡቕቲ       | ጸቡቕ-ቲ     | Sbuq-ti     | -      |       | -ti    |      |       |
|      | ብጸቡቕ       | ብ-ጸቡቕ     | b-Sbuq      | b-     |       | -      |      |       |
|      | ጸቡቕታት      | ጸቡቕ-አት    | Sbuq-at     | -      |       | -at    |      |       |
|      | ጸቢቀ        | ጸቢቕ-አ     | Sebiq-e     | -      | Sebiq | -e     |      |       |
|      | ጸቢቕና       | ጸቢቕ-ና     | Sebiq-na    | -      |       | -na    |      |       |
|      | ጸቢቀን       | ጸቢቕ-አን    | Sebiq-en    | -      |       | -en    |      |       |
|      | ጸቢቕኩም      | ጸቢቕ-ኩም    | Sebiq-kum   | -      |       | -kum   |      |       |
|      | ጸቢቆም       | ጸቢቕ-አም    | Sebiq-om    | -      |       | -om    |      |       |
|      | ጸቢቃቶም      | ጸቢቕ-አቶም   | Sebiq-atom  | -      |       | -atom  |      |       |
|      | ጸቢቃተን      | ጸቢቕ-አተን   | Sebiq-aten  | -      |       | -aten  |      |       |
|      | ጸቢቕክን      | ጸቢቕ-ክን    | Sebiq-kn    | -      |       | -kn    |      |       |
|      | ክጸብቕ       | ክ-ጸብቕ     | k-Sbq       | k-     | Sbq   | -      |      |       |
|      | ትጸብቕ       | ት-ጸብቕ     | t-Sbq       | t-     |       | -      |      |       |
|      | ክትጸብቕ      | ክት-ጸብቕ    | kt-Sbq      | kt-    |       | -      |      |       |
|      | ጸብቕነት      | ጸብቕ-ነት    | Sbq-net     | -      |       | -net   |      |       |
|      | ክጸብቁ       | ክ-ጸብቕ-አ   | k-Sbq-u     | k-     |       | -u     |      |       |
|      | ክትጸብቁ      | ክት-ጸብቕ-አ  | kt-Sbq-u    | kt-    |       | -u     |      |       |
|      | ክጸብቃ       | ክ-ጸብቕ-አ   | k-Sbq-a     | k-     |       | -a     |      |       |
|      | ምጽባቆም      | ም-ጽባቕ-አም  | m-Sbaq-om   | m-     | Sbaq  | -om    |      |       |
|      | ምጽባቃተን     | ም-ጽባቕ-አተን | m-Sbaq-aten | m-     |       | -aten  |      |       |
|      | ምጽባቃቶም     | ም-ጽባቕ-አቶም | m-Sbaq-atom | m-     |       | -atom  |      |       |
|      | ዝጸበቀ       | ዝ-ጸበቕ-አ   | z-Sebeq-e   | z-     | Sebeq | -e     |      |       |
|      | ዝጸበቀት      | ዝ-ጸበቕ-አት  | z-Sebeq-et  | z-     |       | -et    |      |       |
|      | ዝጸበቅካ      | ዝ-ጸበቕ-ካ   | z-Sebeq-ka  | z-     |       | -ka    |      |       |
|      | ዝጸበቅኩም     | ዝ-ጸበቕ-ኩም  | z-Sebeq-kum | z-     |       | -kum   |      |       |
|      | ዝጸበቅክን     | ዝ-ጸበቕ-ክን  | z-Sebeq-kn  | z-     |       | -kn    |      |       |
|      | ዝጸበቅና      | ዝ-ጸበቕ-ና   | z-Sebeq-na  | z-     |       | -na    |      |       |

## Appendix-9: Inflection of Amharic Sentiment Words

| Word | Inflection | Analysis   | Latin          | Prefix | Stem  | Suffix | Root  | Root |
|------|------------|------------|----------------|--------|-------|--------|-------|------|
| ቅዱስ  | ቅዱስ        | ቅዱስ        | Qdus           | -      | qdus  | -      | ቅ-ድ-ስ | qds  |
|      | ቅዱሳን       | ቅዱስ-አን     | qdus-an        | -      |       | -an    |       |      |
|      | ቁድሳኖች      | ቁድስ-አኖች    | quds-anoc      | -      | quds  | -anoc  |       |      |
|      | ቅድስና       | ቅድስ-ና      | qds-na         | -      | qds   | -a     |       |      |
|      | ቅድስት       | ቅድስ-ት      | qds-t          | -      |       | -t     |       |      |
|      | ቅዱሶች       | ቅዱስ-አች     | qdus-oc        | -      |       | -oc    |       |      |
|      | የቅድስት      | የ-ቅድስ-ት    | ye-qds-t       | ye-    |       | -      |       |      |
|      | የተቀደሰ      | የተ-ቀደሰ-አ   | yete-qedes-e   | yete-  | qedes | -e     |       |      |
|      | የተቀደሰች     | የተ-ቀደሰ-አች  | yete-qedes-ec  | yete-  |       | -ec    |       |      |
| ሰነፍ  | ሰነፍ        | ሰነፍ        | Senef          | -      | senef | -      | ሰ-ን-ፍ | snf  |
|      | ሰነፎች       | ሰነፍ-አች     | senef-oc       | -      |       | -oc    |       |      |
|      | መሰነፍ       | መ-ሰነፍ      | me-senef       | me-    | snef  | -      |       |      |
|      | በሰነፍ       | በ-ሰነፍ      | be-senef       | be-    |       | -      |       |      |
|      | እንደሰነፍ     | እንደ-ሰነፍ    | `Ende-senef    | `Ende- |       | -      |       |      |
|      | እንደሰነፎች    | እንደ-ሰነፍ-አች | `Ende-senef-oc | `Ende- |       | -oc    |       |      |
|      | ከሰነፍ       | ከ-ሰነፍ      | ke-senef       | ke-    |       | -      |       |      |
|      | ከሰነፎች      | ከ-ሰነፍ-አች   | ke-senef-oc    | ke-    |       | -oc    |       |      |
|      | ሰንፍነት      | ሰንፍ-ነት     | snf-net        | -      | snf   | -net   |       |      |
|      | ሰንፍና       | ሰንፍ-ና      | Snf-na         | -      |       | -na    |       |      |
| ፍቅር  | ፍቅር        | ፍቅር        | Fqr            | -      | fqr   | -      | ፍ-ቅ-ር | fqr  |
|      | ፍቅረኛ       | ፍቅር-አኛ     | fqr-eNa        | -      |       | -eNa   |       |      |
|      | ፍቅረኞች      | ፍቅር-አኞች    | fqr-Noc        | -      |       | -Noc   |       |      |
|      | ፍቅረኛዎች     | ፍቅር-አኛዎች   | fqr-eNamoc     | -      |       | eNamoc |       |      |
|      | የፍቅር       | የ-ፍቅር      | ye-fqr         | ye-    |       | -      |       |      |
|      | በፍቅር       | በ-ፍቅር      | be-fqr         | be-    |       | -      |       |      |
|      | ለፍቅር       | ለ-ፍቅር      | le-fqr         | le-    |       | -      |       |      |
|      | አፍቃሪ       | አ-ፍቃር-ኢ    | `o-fqar-i      | `o-    | fqar  | -i     |       |      |

## Appendix-10: Sample Data of Trilingual Sentiment Sentences

1. ከጠበኩት በላይ the best of film i like more
2. ካብ ትግራይ ጎበዝ ዘሓፍተይ በጣም ያምራል ማዚቻው
3. Really raya music great በጣም ደስ ይላል በርታ!
4. እዛ ደርፊ ሰሚዐ ናይ ብሓቂ እየ ዝብለካ ነቢዐ ብጣዕሚ ደስ ትብል ወለድኻን ዓድኻን ተዘክረካ እያ i like these song...
5. ዳኑ ሸኮር nice song አጃኪ በርትዒ
6. ዋው nice music ትክክለኛ ናይ ትግራይ ሙዚቃ!!
7. በቀልድ የተዋዛ፣ ትልቅ ቁም ነገር አዘል ፊልም - የወንዶች ጉዳይ 2 is excellent.
8. ደጋግመው ቢያዩት የማይሰለች ምርጥ movie
9. Nice movie አዝያ መሃሪት ዘኮነት Film so ቀፅልዎ
10. wowww keep up my bro so very nice song ደቂ ዓዲይ ደስ በሃሎ
11. so best music keep it up
12. wow we have a beautiful culture i am so glad am one of zs ppl
13. The food was terrible in the hotel.
14. The rooms are very clean and well equipped and the staffs are super nice.
15. ብጠዓሚ መሃሪት ዘኮነት ፊልም እያ::
16. ድምዒ ወያነ ብጣዕሚ እየ ዝፈትወኩም ቀፁሉሉ
17. ድምዒ ወያነ ዘይኮነስ ድምዒ ጸማማት ጥራሕ እዩ
18. እቶም መራሕትና ጸቢብ ኣተሓሳስባ እዩ ዘለዎም::
19. በጣም የሚደነቅ የአገራችንን የፊልም ደረጃ የሚያሳድግ ፊልም ነው:: በርቱ ተበራቱ እንላለን
20. ዋው በጣም ቆንጆ ፊልም ነው
21. ህዝባችን የተሰማውን ሀዘን እሚገልፅበት ጊዜ ቅርብ ነው::
22. ጠቅላይ ሚኒስትር መለስ ዜናዊ የዓለማችን ምርጥ ሰው ናቸው::
23. የቆየ ነው ግን ያምራል
24. አሪፍ ፊልም አይደለም::
25. ምንታ ክፍል በጣም ንጹህ አልነበረም ::
26. አስጠሊ ሆቴል ነው ::
27. እስካሁን ካየሁቸው ፊልሞች በጣም የሚያስጠላ አላማ የሌለው? ገንዘብ ለመስብሰብ ብቻ የተደረገ ስራ ፊልም::
28. The most beautiful movie ever thanks you!!
29. በዚህ ዘፈን too much happy አይደለሁም

## Appendix-11: Ethiopic Scripts/Fidel Vs Latin Scripts

| First order |      | Second order |      | Third order |      | Fourth order |      | Fifth order |     | Sixth order |     | Seventh order |     |
|-------------|------|--------------|------|-------------|------|--------------|------|-------------|-----|-------------|-----|---------------|-----|
| ሀ           | he   | ሁ            | Hu   | ሂ           | hi   | ሃ            | ha   | ሄ           | hE  | ሀ           | h   | ሀ             | ho  |
| ለ           | le   | ሉ            | Lu   | ሊ           | li   | ላ            | La   | ሌ           | lE  | ለ           | l   | ለ             | lo  |
| ሐ           | He   | ሐ            | Hu   | ሐ           | Hi   | ሐ            | Ha   | ሐ           | HE  | ሐ           | H   | ሐ             | Ho  |
| መ           | me   | ሙ            | Mu   | ሚ           | mi   | ማ            | ma   | ሜ           | mE  | ሞ           | m   | ሞ             | mo  |
| ሠ           | ^se  | ሠ            | ^su  | ሠ           | ^si  | ሠ            | ^sa  | ሠ           | ^sE | ሠ           | ^s  | ሠ             | ^so |
| ረ           | re   | ሩ            | Ru   | ሪ           | ri   | ራ            | Ra   | ራ           | rE  | ር           | r   | ር             | ro  |
| ሰ           | se   | ሱ            | Su   | ሲ           | si   | ሳ            | Sa   | ሴ           | sE  | ሰ           | s   | ሰ             | so  |
| ሸ           | xe   | ሸ            | Xu   | ሸ           | xi   | ሸ            | Xa   | ሸ           | xE  | ሸ           | x   | ሸ             | xo  |
| ቀ           | qe   | ቁ            | Qu   | ቂ           | qi   | ቃ            | Qa   | ቄ           | qE  | ቅ           | q   | ቆ             | qo  |
| ቆ           | Qe   | ቆ            | Qu   | ቆ           | Qi   | ቆ            | Qa   | ቆ           | QE  | ቆ           | Q   | ቆ             | 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   | ቲ            | Tu   | ቲ           | ti   | ታ            | Ta   | ቲ           | tE  | ት           | t   | ቲ             | to  |
| ቸ           | ce   | ቸ            | Cu   | ቸ           | ci   | ቸ            | Ca   | ቸ           | cE  | ቸ           | c   | ቸ             | co  |
| ኸ           | Ke   | ኸ            | Ku   | ኸ           | Ki   | ኸ            | Ka   | ኸ           | KE  | ኸ           | K   | ኸ             | Ko  |
| ኅ           | ^he  | ኅ            | ^hu  | ኅ           | ^hi  | ኅ            | ^ha  | ኅ           | ^hE | ኅ           | ^h  | ኅ             | ^ho |
| ነ           | ne   | ኑ            | Nu   | ኒ           | ni   | ና            | Na   | ኔ           | nE  | ነ           | n   | ነ             | no  |
| ኘ           | Ne   | ኘ            | Nu   | ኘ           | Ni   | ኘ            | Na   | ኘ           | NE  | ኘ           | N   | ኘ             | No  |
| አ           | ^e   | አ            | ^u   | አ           | ^i   | አ            | ^a   | አ           | ^E  | አ           | ^   | አ             | ^o  |
| ከ           | Ke   | ከ            | Ku   | ከ           | ki   | ከ            | Ka   | ከ           | kE  | ከ           | k   | ከ             | ko  |
| ወ           | We   | ወ            | Wu   | ወ           | wi   | ወ            | Wa   | ወ           | wE  | ወ           | w   | ወ             | wo  |
| ዐ           | 'e   | ዐ            | 'u   | ዐ           | 'i   | ዐ            | 'a   | ዐ           | 'E  | ዐ           | '   | ዐ             | 'o  |
| ዘ           | Ze   | ዘ            | Zu   | ዘ           | zi   | ዘ            | Za   | ዘ           | zE  | ዘ           | z   | ዘ             | zo  |
| ዠ           | Ze   | ዠ            | Zu   | ዠ           | Zi   | ዠ            | Za   | ዠ           | ZE  | ዠ           | Z   | ዠ             | Zo  |
| የ           | Ye   | የ            | Yu   | የ           | yi   | የ            | Ya   | የ           | yE  | የ           | y   | የ             | yo  |
| ደ           | De   | ደ            | Du   | ደ           | di   | ደ            | Da   | ደ           | dE  | ደ           | d   | ደ             | do  |
| ጆ           | Je   | ጆ            | Ju   | ጆ           | ji   | ጆ            | Ja   | ጆ           | jE  | ጆ           | j   | ጆ             | jo  |
| ገ           | Ge   | ገ            | Gu   | ገ           | gi   | ገ            | Ga   | ገ           | gE  | ገ           | g   | ገ             | go  |
| ጠ           | Te   | ጠ            | Tu   | ጠ           | Ti   | ጠ            | Ta   | ጠ           | TE  | ጠ           | T   | ጠ             | To  |
| ጨ           | Ce   | ጨ            | Cu   | ጨ           | Ci   | ጨ            | Ca   | ጨ           | CE  | ጨ           | C   | ጨ             | Co  |
| ጸ           | Pe   | ጸ            | Pu   | ጸ           | Pi   | ጸ            | Pa   | ጸ           | PE  | ጸ           | P   | ጸ             | Po  |
| ፀ           | ^Se  | ፀ            | ^Su  | ፀ           | ^Si  | ፀ            | ^Sa  | ፀ           | ^SE | ፀ           | ^S  | ፀ             | ^So |
| ጸ           | Se   | ጸ            | Su   | ጸ           | Si   | ጸ            | Sa   | ጸ           | SE  | ጸ           | S   | ጸ             | So  |
| ፈ           | Fe   | ፈ            | Fu   | ፈ           | fi   | ፈ            | Fa   | ፈ           | fE  | ፈ           | f   | ፈ             | fo  |
| ፐ           | Pe   | ፐ            | Pu   | ፐ           | pi   | ፐ            | Pa   | ፐ           | pE  | ፐ           | p   | ፐ             | po  |
| ሷ           | lWa  | ሷ            | mWa  | ሷ           | ^sWa | ሷ            | sWa  | ሷ           | xWa | ቁ           | qWe | ቁ             | qWi |
| ቁ           | qWa  | ቁ            | qWE  | ቁ           | qW   | ቁ            | Qwa  | ቁ           | QWE | ቁ           | QW  | ቁ             | vWa |
| ቸ           | cWa  | ቸ            | ^hWi | ቸ           | ^hWa | ቸ            | ^hWE | ቸ           | ^hW | ጧ           | TWa | ቸ             | pWa |
| ከ           | kWi  | ከ            | kWa  | ከ           | kWE  | ከ            | kW   | ከ           | Kwi | ከ           | KWa | ከ             | KWE |
| ኸ           | KW   | ኸ            | zWa  | ኸ           | dWa  | ኸ            | jWa  | ኸ           | gWe | ኸ           | gWi | ኸ             | gWa |
| ጌ           | gWE  | ጌ            | gW   | ጌ           | HWa  | ጌ            | rWa  | ጌ           | Qwe | ጌ           | bWa | ጌ             | tWa |
| ኸ           | ^hWe | ኸ            | nWa  | ኸ           | NWa  | ኸ            | kWe  | ኸ           | Kwe | ኸ           | ZWa | ጌ             | CWa |
| ጸ           | PWa  | ጸ            | Swa  | ጸ           | fWa  | ጸ            | .    | ጸ           | .   | ጸ           | .   | ጸ             | .   |

## 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 dully acknowledged.

Declared by:

Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

Confirmed Advisor:

Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

Place and date submission: Addis Ababa, March 2018