DEVELOPING A MORPHOLOGICAL ANALYZER AND GENERATOR FOR AWNGI VERB USING FINITE STATE TRANSDUCER

M.Sc. THESIS

DAWIT KASSAHU ABEBE

HAWASSA UNIVERSITY, HAWASSA, ETHIOPIA

MAY /18/2023

DEVELOPING A MORPHOLOGICAL ANALYZER AND GENERATOR FOR AWNGI VERB USING FINITE STATE TRANSDUCER

DAWIT KASSAHUN ABEBE

A THESIS SUBMITTED TO THE

DEPARTMENT OF COMPUTER SCIENCES,

HAWASSA COLLEGE OF INSTITUTE OF TECHNOLOGY, SCHOOL OF

GRADUATE STUDIES

HAWASSA UNIVERSITY

HAWASSA, ETHIOPIA

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE

DEGREE OF

MASTER OF SCIENCE IN COMPUTER SCIENCES

SPECIALIZATION: COMPUTER SCIENCE

MAY /18/2023

Declaration

I hereby declare that this MSc Specialty or equivalent thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis dissertation have been duly acknowledged.

Name: Dawit Kassahun Abebe

Signature: _____

This MSc Specialty or equivalent thesis has been submitted for examination with my approval as thesis advisor. Name: Tesfaye Bayu (PhD)

Signature: _____

Place and Date of Submission: _____

ADVISORS' APPROVAL SHEET

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY ADVISORS' APPROVAL SHEET

This is to certify that the thesis entitled **"Developing A Morphological Analyzer and Generator for Awngi Verb Using Finite State Transducer"** was submitted in partial fulfillment of the requirements for the degree of Master's with specialization in ***COMPUTER SCIENCE***, the Graduate Program of the Department/School of ***COMPUTER SCIENCE*** and has been carried out by ***DAWIT KASSAHUN*** ID. No***. GPCoSc/0005/12***, under my/our supervision. Therefore I/we recommend that the student has fulfilled the requirements and hence hereby can submit the thesis to the department.

Tesfaye Bayu (PhD) _____ _____

Name of major advisor Signature Date

Girmay Tekle _____ _____

Name of co-advisor Signature Date

SCHOOL OF GRADUATE STUDIES

HAWASSA UNIVERSITY

EXAMINERS' APPROVAL SHEET

We, the undersigned, members of the Board of Examiners of the final open defense by **Dawit Kassahun Abebe** have read and evaluated his/her thesis entitled **"Developing A Morphological Analyzer and Generator for Awngi Verb Using Finite State Transducer"** and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree.

Tesfaye Bayu (PhD)	_____	_____
Name of Major Advisor	Signature	Date
Girmay Tekle (MSc)	_____	_____
Name of co-advisor	Signature	Date
Anitha.V (PhD)	_____	_____
Name of Internal Examiner-I	Signature	Date
Endalkachew Mengesha(MSc)	_____	_____
Name of Internal Examiner-II	Signature	Date
Michael Melese (PhD)	_____	May 16, 2023
Name of External Examiner	Signature	Date
_____	_____	_____
SGS Approval	Signature	Date

Final approval and acceptance of the thesis are contingent upon the submission of the final copy of the thesis to the School of Graduate Studies (SGS) through the Department/School Graduate Committee (DGC/SGC) of the candidate's department.

ACKNOWLEDGMENT

First and foremost, I want to express my heartfelt gratitude to the almighty GOD for giving me the effort and perseverance to complete the work. Then the next my heartfelt gratitude goes to my Advisors **Dr. Tesfaye Bayu** and **Girmay Tekle (MSc)** for their invaluable and outstanding guidance, support, and help throughout my research work.

It is my pleasure to express my deepest gratitude to **Teacher Abebe yihun** and **Engineer Esubalew Agegnehu** for their support regarding the linguistic nature and relevant features of the Awngi language verb structures. Next, I would like to express my sincere thanks to Awngi Journalist **Mr. Solomon Sintayehu** and **Mr. Belyneh kassahun** for providing the relevant data.

My deepest gratitude goes further to my family who is directly or indirectly supported in my life and their advice empowers me to work hard. Particularly my tanks to is to my elder **Sewenet kassahun** and my younger **Getenet kassahun** played a great role in my success. Next to this, I want to express my sincere gratitude and thoughtful thanks to all my dear **friends** without whose help this thesis work could not have been completed. Finally, I would like thanks to **Dill university** for giving me the sponsorship opportunity.

TABLE OF CONTENTS

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS USED

NLP	Natural Language Processing
FST	Finite State Transducer
FSA	finite state automaton
C	consonant
V	vowel
TENSC	tense suffixes class,
PRNSC	Person suffixes class,
NGSC	negative suffixes class
IMPSC	imperative suffixes class,
NUSC	number suffixes class,
GNDSC	gender suffixes class,
CAUSC	causative suffixes class.
+V	verb
+1prs	the first person singular
+2prs	the second person singular
+3prs	the third person singular
+Mas	the masculine gender
+Fem	the feminine gender
+1pl	the first-person plural
+2pl	the second person plural

+3pl	the third person plural
+Adj	Adjective class
+NGE	negative marker
+Impfct	the present and future tense
+Prfct	past tense
+PrsCont	present continuous tense
+PasCout	the past continuous tense
NL	natural language

ABSTRACT

Language is a powerful medium that coordinates day-to-day activities between an entity. The means of interaction between a computer and language structure is studied by natural language processing. Natural language processing has emerged as a means of increasing computers' capability to understand natural languages.

Awngi is one of the natural languages, grouped under the central Cushitic language family and spoken by more the 2.5 million people as native speakers settled in a different part of the Amhara regional state. It is a zona language that started as a medium of instruction in 1989 E.C. and uses the Geez script system. It is the one of under-resourced languages in Ethiopia. This morphological study is the first attempt that is used as a contribution to high-level NLP applications. The study is an implementation of the morphological analyzer and generation based on a finite-state transducer approach. The lexicon formalism is used to design the morphological analyzer and generation of the Awngi verb form because the Awngi language is morphologically very productive in terms of gender, person, number, and tenses. The system was developed using the foma programing language as an implementation toolkit. The foma programing language and the Ubuntu 22.04 Linux-based operating system are used for experimentation. For the experimental thirty-five distinct linguistic rules were developed to form those different word forms. The experiment shows that 76956 words were generated by the system out of this 93.6% of words are correctly generated and analyzed by the system and 6.4% of words were wrongly generated. The factors that the wrongly generated words were the problem of the epenthetic sound vowel ɨ/i/ breaking up consonant clusters of words and the glottal approximant /h/, which is loaned from Amharic appears irregularly in the Awngi language word formation process. So, determining the correct phonological pattern of the glottal approximate /h/ and the alternative techniques for the epenthetic vowel ɨ/i/ management is left for future work.

Keywords: Awngi Language, Morphological system, Natural language processing, Finite stat transducer

CHAPTER ONE: INTRODUCTION

1.1 Background

Language is one of the fundamental aspects of human behavior and it is a powerful medium of communication among humans and coordinating day-to-day activities with human beings. It evolved naturally in humans. The written form of the language serves as means of recording evidence and transmitting what it recorded information from one generation to the next generation [1][2]. Now, a day, technology integrates, different community members from different parts of the world without communication barriers. In this disciple, different actors play their roles to realize the expected communication platform for everybody. So, studying the language and adding technological value to the language is very important.

Linguistics can be defined as the scientific study of language, mainly, human language, and an approach to studying the Linguistic nature of the language is called Computational Linguistics or Natural Language Processing (NLP). Natural Language Processing (NLP) is a field of computer science that studies the interaction between computer and human Language to generate human-understandable information from computers and design and build systems that convert human language into more formal structures that computers can understand [3][4]. The linguistic structure in most sub-Saharan African countries is complex and the majority of speakers belong to Afro-Asiatic language families which also have sub-families Cushitic, Semitic, and Omotic[5].

There are 86 languages spoken in Ethiopia having millions of native speakers, a few of which are Amharic, Afaan Oromo, and Awngi. Awngi (አውኚ) language belongs to the central Cushitic linguistic family, and it has been granted its own Nationality Zonal in the Amhara Region State of Ethiopia, spoken by the sub-group of the Awi people living in central Gojjam in northwestern Ethiopia [6]. It is spoken by at least 2.5 million people as native speakers settled in a wide area of

the northwest part of Ethiopia. The most of speakers of this language are all of the Awi Zone, some areas of the Metekel Zone of the Benishangul-Gumuz National Regional State, and, Alfa and K'wara Woredas of the North-Gonder Zone[7]. Currently, Awngi (**አውኚ**) is the medium of instruction in primary, secondary, and college and degree programs at enjbara University. The Awngi (**አውኚ**) language plays a vital role for the people of the region in public media, social issues, religion, political affairs, economic activities, and technology of the Agew-Awi administrative Zone of /**enjbara /እንጅባራ**/ in the Amhara Regional State of Ethiopia[8].

As pointed out by Tsegaye Misikir [9] Awngi language uses the Geez script system and it has used both inflectional and derivational types of morphological affixation. Each word in the language is morphologically rich and formed by reduplication and affix (prefix, suffix, infix, and circumfixes) and other additional mechanisms. Depending on the morphological difficulty of the Awngi language inflectional and derivational morphologies can produce very large numbers of variants for a single word. The variation of word formation processing is resulting in a strong influence on the effectiveness of morphological analysis tools, information retrieval systems, and general natural language applications[10]. The Awngi language is complex in its morphological structure that needs computerized measures to reduce the size of a lexicon to manage the level and capture the strong relationships existing between different word formations of the language[11]. However, there are no preceding works of any developed morphological system for the Awngi language what we want to do is the first attempt to design and implement a morphological system for the Awngi language.

Natural language processing (NLP) is a branch of modern computational techniques that concerns processing and understanding natural language by the computer system[6]. It permits learning, understanding, and producing human language content. Natural language processing (NLP) is a way of investigating and representing naturally occurring languages at one or more levels of

linguistic study to achieve human-like language processing for a variety of tasks or applications [5]. As pointed out by Yitayal Abate [4][12] Natural language processing (NLP) is recent research and application area related to other information technology methods because mainly lower levels of NLP systems are currently processed or implemented the reason is that lower level work with smaller units of analysis, morphemes, words, a sentence which is rule-governed and higher levels of languages processing needs text and word knowledge which is regularity-governed [4].

Natural language processing (NLP) gives numerous application areas of text processing and the use of natural language processing with a computational system[13]. Computational linguistics focuses on the linguistic structure of the language and developing natural language applications including spilling checker, machine translations, grammar checker, text summarization, word sense disambiguation, morphological analysis, natural language generation and part of speech tagging are some of the natural language application areas[14][4].

Among the application of the natural language processing (NLP) areas, this study is interested in the morphology of the Awngi verb. Morphology is all about the structure of the words. Morphological analysis is a crucial task and it is the process of segmenting the surface word into its lexical components and analyzing the word formation and giving the grammatical functions of morphemes affixed to it [15][3][5]. The morphological system is used as a subcomponent of NLP applications and it is an important mechanism of language engineering applications that regards segmentations of the word into their component morphemes and gives morphological and grammatical information associated with the given word [3][15][16]. Morphological analysis is a special type of software that plays a significant role in the development of all-natural language processing applications because it is a blueprint for all other applications like spelling and grammar checkers, next word-prediction, machine translation, text summarization [13][17].

To develop the morphological system, there are two broad categories of approaches in computational morphology such as rule-based and machine learning in which rule-based requires the domain knowledge of the language that provides highly accurate results [17]. Whereas corpus based is an artificial intelligence concerned with the algorithms that learn from examples [16]. The grammatical structure and morphological properties of the Awngi language are different from other languages to need its morphological system and it is important for the development of higher-level linguistic analysis of the Awngi natural language processing applications.

1.2 Motivation

Natural language processing plays a great role in developing applications that learn, understand, and produce human-like language. Language technology depends on the presence of software for analyzing and generating word forms, therefore most NLP applications no preceding attempted for Awngi language like morphological analyzer, spell checker, grammar checker, word prediction, word sense disambiguation, text summarization, etc. The first reason is informing the way for those who are interested in developing NLP applications on the Awngi language by incorporating the morphological systems in their study because the morphological analyzer plays a vital role or it is a blueprint for the development of any other natural language processing system, especially at higher level linguistic and it is used as a basis or its output is used as an input for future work or development of other NLP applications. The second reason that motivates the researchers is to develop a morphological analyzer for this language because based on the researcher's knowledge there has never been any effort made to develop a morphological analyzer for the Awngi language. This is what motivates us to conduct research work in this area.

1.3 Statement of the problem

Based on their linguistic nature, the languages are classified into different language families. Awngi (አውኚ) is an under-resourced Central Cushitic language family spoken by a sub-group of Awi people settled in the Amhara Region State of Ethiopia[6]. Its Writing system is a form of the Geez script like Amharic and Tigrigna alphabets but it has different sounds and alphabets which makes it not share with Amharic namely, **Q(ቐ), ŋ/ng (ኝ), G (ጐ),** and **ts (ፀ)** because the nature of each language based on its orthography, phonology, and morphology makes unique variations among the languages [7]. The Awngi language has no well-studied linguistic resource that addresses the development of other natural language (NL) applications like morphological analysis, spell checker, grammar checker, machine translation, speech recognition, word prediction, word-sense disambiguation, etc. The absence or limitation of such NLP applications for the Awngi (አውኚ) language may impede the development of language technology and this problem makes it difficult for retrieving electronic documents [7][18].

The study of the given language may contribute something to other languages or may not at all [19]. According to the author, the nature of each language based its orthography, phonology, and morphology makes unique variations among the different languages. Most horns' of Africa languages are classified in to the Afro-Asiatic family and they are morphologicaly complex those languages are from Ethiopia, Eritrea, Djibouti, and Somalia lack well prepared computational resources, such that they are not computationaly developed to the anticipation[5]. Some language in Ethiopia were undertaken morphological analyzer and genetaor by different scholars at different time. For example an automatic morphological analyzer for Amharic text using an unsupervised and auto-segmental approach was designed and developed[20]. The Afaan Oromo morphological analyzer is developed for the application of natural language processing using a corpus-based apprcach[3][17]. The supervised memory-based learing approach was used for Sidaamu Afoo morphological analyzer implementation[2]. Morphological analyzers developed for different

languages in the world using different techniques dose not applicable to the Awngi language. Because of this, each language has its properties, grammatical rules, and the nature of orthographic representations and morphological word formation. The Awngi has its unique word formation structure and nature of orthographic representation and words are represented with appropriate alphabetic symbols that obey all rules required by the language. The Awngi language supports consonant germination and the vowels can be long or short base on constituent sounds for grammatical purposes. Misrepresentations of the required alphabet bring loss of information. For example**, /limi-ts/ ልሚፅ/** meaning in English ‘close’ and **/limi-tsts/ ልሚፅፅ/** meaning in English ‘cause to close’ has completely different meaning. So, to avoid such kind of problem spelling correction, needs a successful morphological analyzer. A morphological generator and analyzer support the spelling correction to help native speakers to convey their thought through the correct script. Awgi is zonal language used most of people living in this area but not working language. The situation leads misinterpratiation of information with working language and Awngi. As a solution machine translation application plays an important role. That is whay a morphological analyzer and generator in right point for Awngi langawag now. In addition to this, the Awngi language is limited to the NLP application studied. For instance, further and higher-level linguistic study needs a morphological analyzer and it has important significance for many language engineering applications like machine translation, question-answering systems, and Internet search engines.

Morphological analyzer plays a vital role in the development of most NLP applications [13][15]. However, as researcher knowledge, there has never been any preceding attempt or previous work to develop a morphological system for the Awngi language. For instance, the researcher has proposed to put a milestone by developing a morphological system for the language and the main goal of this study is to explore the possibility of morphological generation and analyzing the Awngi words. Building a morphological system lays the foundation for the language as it is the

lowest level of text processing technique and provides a base for further development of other NLP applications in the Awngi language.

1.4 Research question of the study

- How to solove the computational aspacts of the verbs and what are the challenges in the language while using FST?
- To what extent FST is significant for Awngi verb morphological analyzer and generator design and implementations?
- How to test and analyze the Awngi morphology system performance?

1.5 Objectives

1.5.1 General objective

The General objective of this study is to design and develop the Morphological analyzer and generator for Awngi verb forms using a Rule-Based Approach.

1.5.2 Specific objectives

To achieve the general objective of the study the following specific objectives are performed.

- To understand the morphological structure of the Awngi verb.
- To collect relevant data in different domain
- To study various techniques employed for morphological analysis and identify the one which is suitable for developing a morphological system for the Awngi language.
- To build a lexicon that is appropriate for demonstrating the experiment.

- To define possible language-specific rules that should modify the morphotactic structure of the word.
- To evaluate the performance of the finite-state transducer (FST) and report the experimental result.

1.6 Scope and Limitation

1.6.1 Scope of the study

The scope of this study is limited to the design and development of a morphological analyzer and generator for Awngi language using rule-based approach implemented by a finite state transducer. This study focuses on the verb part-of-speech (POS) categories of Awngi language and their inflected forms.

Other part-of-speech (POS) categories do not discuss in this study because there are no publicly available NLP tools and another resource the Awngi language that can be integrated with my study. The study should not include compound word formations.

1.6.1 Limitations of the study

Due to the shortage of the time and the lack of well-digitized linguistic materials in the Awngi language all parts of speech are not covered in the study. The affixation is not the same for all word categories and preparing the lexicon is a difficult task and time-consuming to build it manually.

1.7 Significance of the study

The result of the morphological analyzer and generator is one of the starting points for improving the performance of the NLP applications. The morphological system contributes a vital role in the development of many higher-level NLP applications that needs the output of a morphological analyzer as input for their analysis or as a foundation for an application like information retrieval, machine translation, spell checker, grammar checker, etc, [13]. The linguistical significances of this study include the researchers that want to develop an efficient higher-level NLP application for the Awngi language. The study can be applied for:

- It is used as input for other high-level natural languages processing applications like machine translation, information retrieval, speech synthesizer, spelling checker, and grammar checker.
- For increasing the Awngi learning and teaching process.
- It creates a good situation for Awngi language word processing techniques in the area of natural language processing computing environment.

1.8 Methodology

1.8.1 Literature review

There are several scholars primarily focused on the study of morphological analysis and generator in different languages using different techniques in the world. Information from the scholars who have studied of Awngi language traditionally and scientifically is the first basis for this study. To understand and develop the morphological analysis and generation of the Awngi language we have made continuous discussions with Awngi language experts and get background information that is important for the study [7][15]. In addition to this, the researcher reviewed different kinds

of literature including books, journal articles, and the internet to make better understudying of the Awngi grammatical structure. Also, the literature review supports us to identify the research gap between which one is previously done or not and the problem statement which is desirable for further investigation [21][22].

1.8.2 Data collections

So far, the Awngi does not have widely accessible computerized electronics data and tagged corpus text for any NLP applications plus morphological databases. As a result, for the development of a morphological analyzer and generator for the Awngi language, we collect the data from a different source in the form of both hard and soft copies. The domain of the data sources is considered under different categories such as the Awngi newspaper called /**chirbéwa** /ቺርቤዋ/ of the Amhara mass media agency Awngi section, from enjbara university teaching materials, from enjbara teacher education college, from high school Awngi supplementary grammatical textbook, and we discussed the more morphological rule of the Awngi verb form with language experts.

1.9 Research design

The development of the Awngi morphological analyzer and generator uses a rule-based approach that will be experimental using FST (Finite State Transducer). In the era of computational environment, the Awngi language is resources poor. The rule-based approach is selected based on the morphological nature of the Awngi language and data size. The system is designed based on the FST approach. The nature of this research is experimental. To experiment anticipated objective of the study whether achievable the targeted goal or not, different activities are encompassed. The system accepts surface words as input for morphological analysis and a list of lexical words with relevant tags for the generator. The system generates a different word form and information of morphemes according to their morphotactic rule of the language.

1.10 Organization of the thesis

The remaining part of the thesis is organized as follows. Chapter 2 deals with the leiterature review. This chapter presents an overview of language processing regarding to concepts of stat of the morphological analyzer in the computational linguistics and the computational morphology. This is followed by the discussion on the approaches to morphological analyzer and genartoar and discussion on the related work that is highly rlevant and more related to this study. Chapter 3 presents the general linguistic structure to the Awngi language and starts introduction to Awngi language which is followed with the writing system and morphological structure of the Awngi language. Chapter for discusses the detailed description of the graphical design and methodology with its basic components that followed by designing language specific rules. The discussion continues with the evaluation of morphological system is presented. Chapter 5 deals with the experimentation and result of the system. Chapter 6 presents the conclusion our work and shows the fature works that needs to included in the fature in order to enhance the morphological analyzer and generator of the Awngi language.

CHAPTER TWO: LITERATURE REVIEW

2.1 Overview

In this section, we described the basic concepts of natural language processing, particularly based on morphological system, and basic ideas about how morphological generator and analyzer is developed for morphologically rich language to decide better decisions for which tools, and methods, would be discussed, the former studies take place in a different language with different approaches in local and forging language has been reviewed and common morphological analyzer and generator approaches have discussed with their strength and weaknesses. The main objective or goal of the study is to design a morphological generator and analyzer for the Awngi verb forms.

2.2 An overview of language processing

Natural language processing (NLP) is a branch of modern computational techniques and applications that explores how computers can be used to understand and manipulate natural languages (NL) in written text or speech in a way that will be useful for further processing or high levels of (NLP) application[23]. The human language is studied by linguistic and computational linguistics is a subset of both linguistics and computer science[24]. As pointed out [25] Natural language processing (NLP) is concerned with the representation of knowledge and problem-solving algorithms involved in learning, producing, translating text, understanding language, and enabling a computer to analyze a huge amount of written text and understand them.

Computational processing involves different disciplines such as computer science, information science, psychology, mathematics, linguistics, electric and electronic engineering, artificial intelligence, and robotics is the basic foundation of natural language processing (NLP) [24]. The wide range of systems developed with help of NLP applications including spoken language

systems, that integrate speech and natural language, spelling error correction to machine translation, and automatic knowledge extraction from human text, are used in language technology. It plays a special role in computer science because natural language processing applications are important in our communication with machines and many aspects of the computations of linguistic features. Representing naturally occurring texts at one or more levels of linguistic studies to archive human-like language processing is a crucial role of natural language processing (NLP)[26].

As described by [26] linguistics the science of a language behavior requires several levels of language processing which include **phonetics** and **phonology** are knowledge about linguistic sound, morphology is knowledge of meaningful components of words, the syntax knowledge of the structural relationships between words, semantics is knowledge of meaning, pragmatics is a knowledge of the relationship of meaning to the goals and intention of the speaker, and discourse is knowledge about linguistic units larger than a single utterance.

2.3 Morphology in computational linguistics

Morphology in the linguistic discipline examines the word formation and inner structure of words, and how words can be broken down into meaningful pieces it is a branch of computational linguistics that studies words, the internal structure of words, and the way how the words are formed from smaller meaning bearing components[27]. As described by[26] the term morphology refers to the study of various smallest linguistic units called morphemes that cannot be segmented further and the description of their behavior and how they co-occurrence together to form a new word and their grammatical functions. It is the domain of linguistics that study of the structure of words and how the word structure reflects their relation to other words. It is formed by the descriptive grammar of each language that usually describes inflection and word formation, and it is the study, identification, and analysis of the words into individuals and identifying the class

of the morphemes. Generally, there are two categories of mereology in which words can be formed such as derivational morphology and inflectional morphology.

2.3.1 Inflectional morphology

Inflectional morphology is a process of combining a word with a grammatical morpheme, which usually results in a word with a similar word class (part-of-speech) as the original stem and having a more relevant grammatical future[12]. In inflectional morphology, the part-of-speech category does not modify but it is sensitive to the grammatical function (also known as morphosyntactic information) is changed[28]. Inflectional morphology modifies words in terms of number, aspect, case, mood, and so on without creating a new word or affecting grammatical categories. For example, the English word “happy” is a free morpheme if it has written with the negative marker bounded morpheme “un” the resulting inflectional form of the word becomes “unhappy” in both cases the part-of-speech of words are adjective but the meaning is changed. In the word formation process different words are created from stems “work” is an English verb with inflectional forms are ‘works’, by adding 3rd person singular marker /-s/,’ worked’, adding the perfective marker /-ed/ and ‘working’, by adding the present continuous marker /-ing/ are produced by adding different affixation without altering the grammatical category because in all case the part-of-speech category is ‘verb’ the same with stem word work. As pointed out by [29] inflectional morphology is used to show if a word is a plurality or singularity, past tense or not, present continuous or not, and if a word is a comparative or possessive form.

2.3.2 Derivational morphology

Unlike inflectional morphology derivational morphology, is a process of creating or deriving a new word by changing the grammatical category of a word by adding bounded morpheme to the stem word and carrying the semantic information[27]. As described by[28] the main goal of

derivational morphology is deriving a new word by changing the part-of-speech (POS) category of the word (changing addictive to noun) like the addition of /-ness/ bounded morpheme to an English word “quick” the resulting word becomes “quickness” the affix changes both meaning and part-of-speech(POS) category adjective “quick “ become to a noun “quickness”.

2.4 Morphemes

Morphemes are the smallest building block of a linguistic expression that can be identified by the separation of a frequently occurring subtype of morphological structure, of the meaningful sub-word units that cannot be divided into smaller meaningful parts and basic building blocks in morphology[27]. Morphemes are defined as the smallest piece in a language that cannot be decomposed into smaller meaningful units and each unit has grammatical information to add to the overall meaning of the words and it is the ultimate element of morphological analysis [30]. For example, if we see the Awngi word “**empltaq**” “**እምፕልታቕ**” meaning in English is “**society**” or “**ህብረተሰብ**” in Amharic is formed from ‘**እምፕል’, and ‘ታቕ**’ morphemes. Both sub-units ‘**እምፕል’,** and ‘**ታቕ**’ are morpheme which has a meaning assigned or as the minimal units of grammatical analysis. Morphemes can occur in a language composed of different affixes and base words. As pointed out by [10] described that based on their forms, Awngi morphemes like Amharic and English are classified into two major categories as morphemes occur stand-alone (are called free morphemes) and some other morphemes can occur only attached to some other morphemes (are called bounded morphemes), for example **“-ጋው”, ‘gaw’, ”-ፅኾ”,’tsGo’** must always be attached after central meaningful element called root word. A hyphen before bound morpheme**/-ጋው/** and **/-ፅኾ/** tells us that it needs a free morpheme to exist. Example “**ዲስቲኒጋው**” ‘**distini-gaw’**, meaning “I am not happy” or **“ደስተኛ አይደለሁም”**, in Amharic. To understand the individual words bounded morphemes must be attached to some kind of host morphemes and in contrast to free morphemes are those that can stand on their own as individual words like,” **ዲስቲኒ**”, ‘**እምፕል’, ‘ቢሩ.** We call such

words monomorphemic because they consist of a single morpheme and most of the affixes are the cover term for all bound morphemes attached to roots.

On the base of linguistics morphemes in a language is a collection of word formation methods like affixations, reduplication, and compounding that combine with root or stem words to form a different word. Affixes are bounded morphemes that cannot stand alone, but they are commonly used for word formation by attaching to the body of a word. They may alter the grammatical information but they do not carry meaning on their own, rather then they can add or change some aspect of meaning[31]. In a language with productive morphology, affix ordering is an important topic for the development of a morphological analyzer. The affixes can be classified into prefixes, suffixes, circumfixes, and infixes, based on their occurrence in a word and participation in the word formation, by association with the root word in different positions. The prefix is affixed that is attached to the beginning of a lexical item or base morphemes. In English, /un-/, and/in-/ as in "unhappy", and "independent", are prefixes [3]. In the case of the Awngi language, the prior studies pointed out by [18] stated that unlike Amharic and English language Awngi morphology does not experience prefixes rather it will consolidate after suffixes or after base words. For example, when we see the word **"አውኺስኪታ" or "በአዊና",** in Amharic the base word is **"አውኺ",** the suffix is **/ "-ስ"/** and **/ "-ስታ"/.** Affixes that are attached at the end of the free morpheme or lexical item is termed suffix. **/-s/** is the plural marker in English and **/ካ/** is the mostly used plural marker for Awngi nouns. An infix is an affix that does not possess the front and back position of lexical items or base morpheme but is somewhere in the middle of the bass morpheme. Circumfix attached on the left and right end of a lexical item of the words. If the root word consisting both left (prefix) and right (suffix) it is possible to analyze them as consisting of a prefix and suffix that apply to a root word same time[4]. Compounding is a process of forming new words from two or more independent existing words by joining them together to form compound words and individual words can be free morphemes[31].

2.5 Word structure or morphotactic

In every language, there are well-defined ways to sequence the morphemes. In a language, the systematic rule is used to govern pronunciation, word formation, and grammatical pronunciation. These rules or structures of the language are in order in which morphemes are combined to form a word[32]. Typically, there are language-specific word grammars that help to determine the arrangement of morphemes together. For instance, the word grammatical rule is studied by a subfield of morphology called morphotactic[31].

Morphotactics is responsible for governing the grammatical rules in which the combination of morphemes, which are meaning-bearing units, forms a natural word that presents any written text in the language[14]. It implements constraints through rules that state which sequence of the several lexicons can be accessed by the morphological analysis and generator[33]. Complex words are built up from simpler words and morphemes, by using Morphotactics rules and, in the same way, the same rule used that complex words be analyzed into simpler ones that impel that Morphotactics are language specific[34]. To implement the morphological system, it is important to take the morphotactic rule into account. Awngi Morphological is very productive, derivational, and word formation that possesses morphotactic rules. Since the Awngi morphemes **/ሻና-/** meaning "house" and **/ክንቲ/** meaning "Education" different word forms can be formed by combining them like **/ሻናክንቲ/,** and **/ክንቲሻና/,** but the grammatically correct arrangement of morphemes is **/ክንቲሻና/** meaning "to the school". Therefore, the morphological analyzer needs to have a constituent that determines the ground grammatical rule for the formation of words begging suffixed or other methods of word formation.

2.6 Computational morphology

Computational morphology is the distinct subfield of computational linguistics that focuses on how computational analysis, and synthesis of the internal structure of words, and the relationship among words which use for further analysis of natural language processing (NLP) applications[31]. In other words, it is the study of the combination of the smallest meaningful constituents of linguistic expressions to yield words. Computational morphology is the use of computer and computational methods to handle the task of morphological analysis and synthesis automatically to achieve an effective means of storage of words in the lexicon and provide time-efficient lookup capabilities[3]. Tasks involved in computational morphology can be grouped into two parts such as word-form synthesis and analysis and parts-of-speech (POS) or inflectional category. The analysis is a process in which surface word form is evaluated into their lexical representation having its grammatical features whereas synthesis is a reverse process of analysis in which a lexical representation is converted to a surface word form [31].

2.7 Morphological analyzer

A morphological analyzer is a computational process, or linguistic study, of the internal structure and formation of words by splitting the individual words into their smallest meaning-bearing components that cannot be further divided into smaller parts. It describes the way how words are formed by combining these smallest meaning-bearing components[35]. As pointed out by [4] from a computational point of view morphological analyzer is a computer program that analyzes the internal structure given surface of words and identifies the more technical, ingredients of the words. The ultimate goal of the morphological analyzer is to build up a computer system that performs analysis and generation of word form as humans do[1]. In the era of the computational environment, an effective morphological analyzer has a significant role in the development of

several natural language processing (NLP) tasks, or computational linguistic applications, like information extraction (IE), information retrieval (IR), machine translation (MT), spelling checkers, speech recognition, [35], etc. Therefore, developing morphological analyzer tools for languages with highly productive in its morphology has a vital impact on the computational perspective of a language. The capacity of a full morphological analysis of a word form is generally viewed as a segmentation of the word into morphemes, combined with an analysis of the interaction of those morphemes that determine the syntactic classification of the structure of the specific word[36]. To carry out the morphological analysis process some basic knowledge is common to all-natural language to implement applications of morphological analysis, however, the detailed knowledge is specific for each language. It is known that a language contains ambiguities in its morphological composition[18]. So, the disambiguation process needs a different class of linguistic knowledge for fixing those ambiguities in a language such as knowledge of the morphemes of the language, the morphotactic constraints of a language on how the sequence of the morphemes are allowed to be combined, and the orthography or sound changes of the language upon affixation[17].

2.8 Approaches to Morphological Analyzer

2.8.1 Machine Learning Approach

Machine learning is a branch of artificial intelligence that is a fast-growing area of computer science discipline that enables far-reaching applications which are used to instruct computers to use past expertise or recognitions of patterns in historical data and utilize those patterns to solve a given problem[40]. In the machine learning approach, different categories of real word problems have been solved at least for more than five-decade and the ever-increasing amount of digitalized data become available which led to the emergency of the development of machine learning

technology progress[41]. The objectives of machine learning methods are the ability of computers system that computers automatically capture a rule from studying a training set of examples and the machine performs the same task with the data it has not encountered previously or new instances and improves the performance as it gathers more experience. It is a mathematical algorithm that provides computers to learn automatically without or with less intervention from human beings[42][43]. Among the number of different applications natural language processing is becoming the hottest significant application area of machine learning and also machine learning methods are rapidly expanding in a variety of engineering domains including computer games, data mining, security, medical diagnosis, pattern recognition, natural language processing and many more [17][40]. Behind its fundamental advantage, the drawback of the machine learning approach is the need for a huge amount of automated or electronic training data, that requires experts to label ground truth which needs more time, cost, and resource consuming, and the need for massive digital data has led to greater hardware requirements. However, based on the researcher's knowledge Awngi language has no or is very limited to the well-organized electronic resource. Therefore, tackling the problems using a machine learning approach on the Awngi language to develop a morphological system is not feasible because of the requirements of this approach.

2.8.2 Rule-based approach

System developed using rule-based approaches are often efficient and produce better quality outputs because it is based on a theory of morphology laid down by experts those methods let one incorporate sophisticated linguistic theory, such as generative phonology, into computational morphological processes [3][42]. To implement a morphological analyzer and generator using a rule-based approach needs a full knowledge of the linguistic exports and it is based on a set of rules and a dictionary of roots and morphemes. Several systems were developed using the rule-

based approach for both commercial and research purposes and tested quite for a long period till now. In this approach, the rules may contain a large number of syntactical, lexical, or morphological information that may be created either by linguists exporting the language or automatically by computer programs[15][34]. The basic quality of or advantage of systems developed using a rule-based approach over that system developed using a corpus/machine learning approach is the first one is unlike corpus-based approach morphological system developed using a rule-based approach there are no complicated computational challenges, once the ground rules are set that speed up the morphological analysis process. The second factor is the rule-based approach needs less storage requirement and a small amount of data to implement than the morphological system developed using the machine learning approach. The third reason the morphological system developed using a rule-based approach has better accuracy than the system developed using a corpus-based approach is because of predefined rule sets to govern the output of the morphological analyzer. The fourth reason that the morphological system developed using rule-based approaches is more straightforward to correct or modify any handicaps that happen within the morphological system.

2.8.2.1 Two-level morphology

A major revolution in the field of morphological systems started to appear in 1983 by Finnish computer scientist Kimmo Koskenniemi produced his dissertation Two-level morphology as a general computational model for word form [44]. In the history of computational linguistics, two-level morphology was the first general model for the analysis and generation of morphologically complex languages. Two-level morphological model is the linguists' mode traditional distinction between Morphotactic and morphophonemic that accounts for the alternation forms or spelling of morphemes considering the phonological aspects of the language. The main features of the model are the separation and bidirectional application program form of a description (rule and lexicon)

of the language[44]. An important ideas two-level morphological model is that the rules are represented as direct symbol-to-symbol constraints that are applied in their lexical form and surface form in which the constraints represent the lexical context to the surface context, other ways in both contexts at the same time[45]. Representations and rule components are the major constituents of the model that considered lexical (morphophonemic) word form which describes the chain of the morpheme in the word and surface representation which depicts the actual spelling of valid words.

2.8.2.2 Finite State Automata

For the implementations of natural language processing (NLP), finite state technology uses finite state devices as well as automata and transducers. The finite state machine is an abstract computing machine used to recognize or reject an input stream of characters that are found in a particular language[46]. It is the basic model of a computational system and a very efficient finite memory occupation and transition occur because of the inputs at each state that primes the present state to process input and transform the state to the next state[25]. It is a mathematical model that has five states or transitions which are represented by nodes in the graph that govern the system and are more formally defined by the collection of five parameters (Q, ∑, qo, F, δ) and each tuple has unique meanings in the computational model Q: represents a finite set of a state, ∑**:** stands for a finite input alphabet, qo: an element of **Q** that represents the start state, F: a subset of Q stands for the set of accepting states, $\delta: Q \times \sum \rightarrow Q$ represents the transition function or transition matrix between state and for the given state R and input symbol I, the transition function defined δ(R, I) returns the state where state to moves to the next appropriate state[47].

2.8.2.3 Finite state transducer

The transducer is a more general version of the finite state automation model based on the two-level morphological thoughts and a finite state transducer (FST) model defines a relation between a set of strings, even though a finite state automaton (FSA) defines a formal language by defining a set of accepted symbols or strings[48]. The transducer is different from the automata modal in which a finite state transducer reads or with two memory taps of a set of characters on the input tape and produces a set of relations on the output labels. In contrast, a finite state automaton can only be used for recognizing single taps. According to the finite state transducer (FST), transition function T is labeled as T= {b: b, c:c, w: w, h:h……} however in the case of finite state automata, it is represented as T= {b, c, w, h…….}. The finite state approach for morphological analysis is the relation between the surface form of a particular language and its corresponding lemmas which are described as regular relations (Karttunen)[49]. The advantage of finite state technology for processing natural language is pointed to four sources such as the compact representation, which is achieved via minimization, space and time efficiency, reversibility, and modularity of the design that is the properties of relation and regular language[46]. A finite state transducer (FST) is well-defined with 7 tuples denoted as (Q, Σ, Γ, δ, ω, q0, F) where: Q is a finite set of the states, Σ stands for a finite set of input alphabet, Γ stands for a finite set of output alphabet, δ is set of the transition function ($Q \times \Sigma \rightarrow Q$) that map states and an input alphabet to the state, ω stands for transduction function ($Q \times \Sigma \rightarrow \Gamma$) that maps state and input alphabet to output alphabet, q0 is an element of the finite state, and F is $F \subseteq Q$ that is the set of accept or final state.

2.9 Related work

2.9.1 Morphological analysis for foreign language

2.9.1.1 Morphological Analyzer for the Maithili Language

Raza Rahi and Smriti kumar [36] developed a morphological analyzer for the Maithili language by using a rule-based approach implemented using a finite state transducer to model the system. The authors have used the XFST (Xerox Finite State Transducer) tools for generating FST (Finite State Transducer) to analyze morphological fragments of the Maithili language. As the authors studied the work was classified into two phases those was build the lexicon file with help of a lexicon generator, and producing morphological detail using Xerox tools. To produce lexicons of the Maithili language they have used the raw corpus from LDC-IL (linguistic data consortium for the Indian language) and the unique words from the corpus are extracted and sorted for manual processing of the words[35]. After the lexicon file was created the morphological generation phase of the morphological analyzer took as its input a dictionary of the suffix (that contains all lists of suffixes with their morpho-syntactic feature and relevant morphological information of Maithili language), a dictionary of root (that contains all existing root of Maithili language) and FST rules are handwritten and implemented by XFST tool[35]. The authors used around 855430 instance words to test the Maithili morphological analyzer from raw corpus LDC-IL for performance evaluation and any of the instances has no failure of inflectional form as long as the root belongs to the lexicon file. They conducted three experiments (experiment1, experiment2, experiment3) to examine or to check the correctness of the morphological system for the Maithili language, and the experiment1 was verified manually to test with 16572 nouns of Maithili language extracted from the corpus and correctly identified roots by the system was 15246 nouns. The authors identified 634 adjectives in experiment 2 to test and verified manually from those 593 of the

instances were examined correctly by a morphological system of the Maithili language. In experiment 3 they analyzed regular and irregular verbs on 14312 instances of Maithili verb forms which were proved manually. They measured the performance evaluation in different grammatical categories of the Maithili language execute their system with 500 words for each POS category and checked the respected accuracy. The accuracy of Noun inflection words in the Maithili language is 92%, the accuracy of Adjective inflection words in the Maithili language is 93%, the accuracy of Verb inflection words in the Maithili language is 95%, the accuracy of Pronouns inflection words in the Maithili language 98%, and accuracy of Adverb inflection words of Maithili language 98%, was achieved in their study.

2.9.1.2 Morphological analyzer for the Persian Language

Persian morphological finite state system was developed by Karine Megerdoomin [37] at Insight Software based that used Xerox Finite State Technology. He wrote the rules of morphology and lexicon in the format of *lexc* and compiled them into a finite state transducer (FST) and he created FST for each part of the speech category separately and composed them. To perform the phonetic and phonological alternations on the word form has composed phonological rules on the relevant finite state transducer and he used a composed part of speech for the final lexical transducer for morphological analysis of the Persian language[37]. Therefore, the final transducer consisted of a single two-level finite state transducer with surface character at the bottom and the morphological output at the top. He has used twenty phonological rules to capture the various surface word forms and alternations that occur in the language and the rules written in the regular expressions that are represented as continuation paths within lex grammar[37]. The finite-state system he has described to analyze the morphemes in the Persian language was the ability to process multiword tokens within the lex grammar module. To eliminate the ambiguities the authors manipulated the orthographic realization of certain words, instead of the written different rules for plural markers

of the Persian language he wrote a single rule and applied phonological rules. Before the optimization was applied to the finite state, the transducer consists of 178452 states and 928982 arcs. The Persian language currently consists of 43154 lemmas which have different categories and 12000 common proper noun entities listed in the lexicon its coverage was 97.5% on a 7MB corpus that was collected from the online news source and got the resulting accuracy of 95%.

2.9.1.3 Morphological analyzer for the Hebrew language

Shlomo and Wintner [38] developed HAMSAH (Haifa Morphological system for the Hebrew language based on a finite state transducer which was a linguistically motivated rule-based approach and used XFST toolbox for implementation of finite state algorithms that were used to store and process large-scale networks. [38]They developed a publicly available morphological analyzer for modern Hebrew languages and used over 20000 lexicons items. The rule they have developed covers all the morphological, morpho-phonological, and orthographical phenomena that were observed in the Hebrew language. They developed the system consisting of lexicon and finite state rules, represented in Extensible Markup Language (XML) that is used for supported standardization of formats to achieve the readability of both humans and machines and implemented by Xerox Finite State Technology (XFST) respectively[38]. They listed dotted and undotted scripts as well as used one-to-one Latin transliteration. To handle the grammatical category's irregularities, they have used additional or alternative lexicon entries to facilitate adding, replacing, and removing operational elements in the lexicon items. They have general lexicons that contained the lexicon items add, remove and replace elements and are used for building a basic morphological finite state network without affecting the listed in the special lexicons that were built from a set of rules for the add, remove, and replace lexicons[38]. They did different operations to obtain the final network by subtracting the remote network from the general lexicon they used a set difference operator, to add the network they used a set union

operator and applied priority union with the replace network and produced a finite state network with valid inflected forms. The drawback of the Hebrew morphological analyzer was not demonstrated the output system and measured the system quality they were not evaluated systematically because only a small-scale evaluation was conducted. The morphological analyzer they have developed for the Hebrew language has correctly analyzed 91.86%, not included for the analyzer 3.86% and 4.28% tokens were not correctly analyzed for the system.

2.9.1.4 Morphological analyzer for the Turkish language

A morphological analyzer for the Turkish language TRMOR was developed by [39] and used a rule-based approach that was implemented by Stuttgart Finite State Transducer (SFST). They concatenate stem and morphemes in all possible sequences through morphotactic rules and the output string is mapped to correct surface form through morphological rules. The surface level realizes the phonological form of a string via a chain of symbols in the normal inflected form of a word and the lexical level fills the abstract basic form of input with lexical information. Most Turkish word formation has influenced by vowel and consonant harmony that morphophonological rules need to account. The form of the TRMOR morphological analyzer was composed of base stems with derivational participles. Productive and a few unproductive forms of derivational suffixes were recognized by the morphological system for the Turkish language and different morphemes with the same surface forms in the Turkish language were treated differently in the lexicon[39]. The inflectional classes of suffixes differ according to their endings in the TRMOR system and the vowel ending suffixes belong to the VerReg-p inflection subclass and the consonant ending suffixes to the VerbReg-Ym inflectional subclass[39]. In the morphological process of TRMOR, parallel and sequential implementation of the morphological rule takes place on the TRMOR transducer is built. In this system, the compiler reads the lexicon files, and each line is converted to a transducer, and the transducer were combined with the OR

operator. The TRMOR system has used 44 rules to map morpheme sequences to surface forms. They used a list of word forms annotated with correct morphological analysis to evaluate the Turkish morphological system on the gold-standard words to analyze the correct grammatical morpheme. They evaluated the system based on two corpora and compared it to the TRmorph analyzer for the Turkish language[39]. The 1000 word has taken from the Wikipedia list with its compounding frequency range and the 2-million-word corpus was extracted from the lexicon Vikipedi (the Turkish version of Wikipedia). During the execution of the gold standard, they handled both entries of grammatical and orthographically and executed them on this file[39]. They evaluated the system using the evaluation matrixes and achieved the expected result was precision of 94.1%, recall of 79.80%, and F-Measure 86.39%, and the authors quantitative compared coverage of the current system TRMOR and TRmorph prior work with their respective accuracy archived. In the current system, 72% accuracy was achieved which is better compared to the accuracy of prior work which was achieved at 38%.

2.9.2 Morphological Analysis for Local Language

2.9.2.1 Morphological analyzer for Wolaytta language

Free/open-source finite State morphological analyzer and generator for the Wolaytta language was developed by Tewodros A.Gebreselassie, Michael Gasser, Jonathan N.Washingto, and Baye Yimam [14]. They employed a rule-based approach, implemented using an HFST (Helsinki Finite State Transducer) toolkit and tools from Apterium's Svn repository. The authors have studied that the Helsinki Finite State Transducer toolkit was used to implement both the *lex* formalism for defining lexicons and the two-level and *just* formalisms for modeling morphological rules for the Wolaytta language. Those authors used the lexicon of Wolaytta words, Morphotactics, and orthographic rules information of the Wolaytta language to model the morphological analyzer and

they have prepared the Wolaytta language sentence corpus from Wolaytta-English bilingual dictionary, and books written for academic purposes[14]. They have written orthographic rules of Wolaytta language in the HFST architecture in the HFST-TwolC formalism that is used to compile grammar two-level rules into sets of finite-state transducers. They have defined morphemes and morphological rules in the lexicon file and alternation rules of Wolaytta verbs were defined and rules were composed of lexicon files in the HHFST-two files. The total size of tokenized words in the corpus they used was 38479 instances. 16.87% percentage of words were not recognized by the morphological analyzer of the Wolaytta language and 83.13% percentage of words were recognized by the Wolaytta morphological analyzer, the remaining percentage of words fell into out of vocabulary words. They evaluated the accuracy of the system one thousand forms were chosen at random from a corpus of Wolaytta language and they achieved a precision of 94.85% and recall of 94.11% was returned analysis correctly[14].

2.9.2.2 Morphological analyzer for Amharic, Afaan Oromo, and Tigrigna language

Horn Morpho system for morphological processing of Amharic, Oromo, and Tigrinya, developed by Gasser[5] by employing finite state transducer (FST). The system analyzes and generates words in Amharic, Tigrinya, and Oromo. The separate lexical and guesser FST for Amharic and Tigrinya is developed for handling each processing direction and there is a separate finite state transducer for verbs for each language. There are two additional processing functions of Amharic and Oromo which segment input verbs into sequences of morphemes and there are two more functions for Amharic which converts the input orthographic form to a phonetic form[5]. The function of prefixes and suffixes is concatenated into the stem function to create the full verb morphotactic functions. The generation function takes a Stem or root and a set of grammatical features and produces the combinations of possible morphemes. To conduct this research, the authors derived the Amharic lexicon from the Amharic-English dictionary Aklilu which has 1851 verb roots and

6471 noun stems, verb and noun roots of Oromo are extracted from the dictionaries of Gragg and Bitima which has 4112 verb root and 10659 noun stems and for Tigrinya verb roots is derived from Efrem Zacarias online dictionaries that have 602 verb roots. To evaluate the performance of the system words were selected randomly from the word list 200 Tigrinya verbs, 200 Amharic verbs, and 200 Amharic nouns and adjectives were collected. A human reader familiar with the language evaluated the system's functionalities. The resulting accuracy of the system was for Tigrinya verbs 96% of the system output was a correct combination of roots and grammatical structure, for Amharic verbs 99%, for Amharic nouns and adjectives 95.5% of the system output was correct [5]. The morphological generation function was expected to generate 10 to 25 verbs which are randomly selected and the accuracy of the generator for Amharic was 100% and for Tigrinya 93% but the Afaan Oromo words have not been conducted comparable evaluation because its complexity of use of double consonants and vowels words in Oromo.

CHAPTER THREE: LINGUISTIC STRUCTURE TO AWNGI LANGUAGE

3.1 Introduction

In the Amhara Region State of Ethiopia Awngi is grouped under a Central Cushitic language family spoken by more than 2.5 million Awi peoples living in an extensive area of central Gojjam in North-West of Ethiopia[24]. It is one of the thirty (13) Zones in the Amhara regional state of Ethiopia and most speakers of the language live in this Zone and various areas of the Metekel Zone of the Benishangul Gumuz National Regional State. Based on geographical demarcation the Awngi language consists of four main varieties: Awngi, Kimant, Bilen, and Xamtagna, within the Central Cushitic language family. Those people are found in some fragment areas of Amhara and Metekel Zone of Benishangul Gumuz Region.

As described by [6] the Awngi language has been started as a medium of instruction since 1989 for the first time in five selected primary schools in the Awi zone. Present-day Awngi (**አውኚ**) is used as a medium of instruction in grad one (**እምፐላንቲ**) to degree level (primary, secondary, junior, colleges and degree program at enjbara University). This language plays a vital role for the people of the region in public media, social issues, religion, political affairs, economic activities, and technology of the Agew-Awi administrative Zone of **enjbara** (**እንጅባራ**) in the Amhara Regional State of Ethiopia[8].

3.2 writing system of Awngi language

Similar to Amharic and Tigrigna languages, the Ethiopian Fidel script is also used for the Awngi writing system. Which currently used alphabets in the Awngi writing system are based on Geez

alphabets that are permanently implemented from the Amharic writing procedure without changing the structure[6]. The Awngi language consists of 24 basic letters of which five are labialized consonants (**አንቤብሰታንትካ**) and six are vowels (**አንቤብየንትካ**). In the modern Ethiopic Fidel script, each syllable pattern comes in different orders that reflect the seven-vowel sound (**ኧ, ኡ, ኢ, ኣ, ኤ, እ, ኦ**) whereas others represent the consonant sound that is used in Awngi writing system [7]. Awngi writing system does not ensure the basic Amharic alphabets like **'ሐ', 'ቀ', 'ጠ', 'ሠ', 'ኀ', 'ዐ', 'ጸ', 'ፐ', 'ፈ', 'ጨ', 'ጰ'.'ዠ'** because of these characters does not form native words for the Awngi language. However, there are additional two language-specific Fidel or alphabets and four sounds used for the Awngi writing system only which do not apply in Amharic, because of these this alphabet does not form native words for the Amharic language such as (**ቕ**) for the sound **Q**, (**ኽ**) for sound **ŋ/ng**, (**ኸ**) for sound **G,** and (**ዕ**) for sound **ts**[10][50]. inpout

As described by [50] the Awngi phonological sounds show variations in distribution these are ***b, ʤ, x,*** and, **ʒ. [x]** and **[*K*]** are allophones of the same phonemes, which demonstrates the voiceless uvular fricatives, and voiced complement, point of articulation respectively and the variation is on the phonetic level not on phonemic. As described by [51] the occurrence of the voiced **[*K*]**follows after a vowel while the voiceless one **[x],** occurs somewhere else the two sound variants are categorized by the same phoneme. In the same way, the voiced palatal affricate sound [**ʤ],** has the fricative [**ʒ]** which occurs after a vowel, but **ï,** and affricate sound one [**ʤ],** that occurs somewhere else to show the distributional variation of a single sound at a phonemic level[50]. The occurrences of morph **[ʣ],** are rear to the affricate, and the phoneme is realized as **[*z*]**. ***For example,*** we have two words, **siedzdza (ሴዛ)** meaning in English ''**four**'' and **budzd*zi* (ቡዚ)** meaning in English ''**fat**'' shows that the occurrence of affricate, **[*ʣ*]** as long consonant between vowels but in above words while the affricative one **[z]** occurs as a non-long consonant in all environments. ***B*** has two morphs, **[b]** and its variant fricative ***[β],*** and occurrences ensuing a

vowel and **[b]** elsewhere[10]. The Awngi language contains the following consonant listed in table 1 and the pronunciation of several consonants are as they are spelled

Table 3 1:List of Awngi Consonants adapted from [(Wedekind, 2002)

		Labial	Alveolar	Palato-velar		Uvular	
				Plain	Labialized	Plain	Labialized
Plosive	Voiceless	P, ፕ	T, ት	k, ክ	k ʷ, ኩ	Q, ቕ	Qʷ, ቖ
	Voiced	B, ብ	D, ድ	g, ግ	gʷ, ጉ	G, ጕ	Gʷ, ጕ
Affricate	Voiceless		t͡s, ፅ	t͡ʃ, ች			
	Voiced		d͡z, z, ዝ	d͡ʒ/j, ጅ			
Fricative		F, ፍ	S, ስ	ʃሽ			
Post-stopped fricative			ኦ	ሀ			
Nasal		M, ም	N, ን	ʃng (ŋ, ኝ)	ʃŋʷኜ		
Flap			R, ር				
Approximate		W, ው	L, ል	Y, ይ			

3.3 Basic Syllable structure of the Awngi language

The syllabification is language specific. Every language has its structure of syllables. In a language there are four forms of syllable structures (**ኬሌመ ውክር**)**,** these include a CV, CVV, CVC, and CVVC these can be found at word initial, medial and final positions. Syllables (**ኬሌም**) are an interior structure of the linguistic, in this case, Awngi[50].

According to the [52] orthographic word system of the Awngi, language has a basic syllable pattern, and the fact that most cases the Awngi language syllables structure obeys the maximum

acceptable syllable template of **CVC** (**C** standing for a consonant (**አንቤብስታንትካ**) and **V** standing for vowels **(አንቤብየንትካ)**) for the formation of words and preference that there is at most one consonant in each syllable on-set and the rhyme. Example **ድእፊእንት** (***ʤifi'nt***) to represent the cluster consonants of **ድእ**(**ʤi**)=CV **ፊእንት**(**fi'nt**)-CVCC 'fear', **ኽሳንቲእ** (***Gsántî***) CCVC-CV 'big'. Consonant clusters are frequently appearing word-medially. This does not mean that word-medially should be a collection of more than two consonants which is true in most cases in Awngi syllable structure. The most common syllable structure in the Awngi language is CV but this does not mean all words begin with a CV because some words start with VC syllables structure[50].

From these six vowels in the Awngi language, they have a rare occurrence of word initials the only two vowels are acceptable at beaning position a syllable **(ኬሌም)** /**a**/ and/**ɨ**/. Some of the words using vowels at word initially like as in '**asip'**(**አሲፕ**) V-CVC 'think', **'ɨstê'(እስተ**) VC-CV, 'it is called' and **ɨngɨr (እንግእር)** VCC-VC 'back' and vowels in this language are considered as the central, front, and back vowels[51]. This vowel representation of this language is depicted below

Table 3 2:List of Awngi Vowels adapted from (Joswig, July 21, 2003)

	Front(ፍን)	Central(ክቸ)	Back(እንግር)
High(አጉዊ)	ኢ(i)	እ(ɨ)	ኡ(u)
Mid(ልክ)	ኤ (e)		ኦ(o)
Low(ሱቿ)		ኣ(a)	

Vowels in Awngi are sound markers and like another Ethiopian language in Awngi, a single word takes plentiful information due to its very rich morphology that grammatical information is conveyed through affixes that should be attached to the roots **(ሙሲሲኒ).** In the Awngi language, the word never begins with non-central front **(ፍንቴን)** and back (**እንግርቴን**) vowels because of an

issue with word-medial consonant clusters[50][53]. In this language, the high central vowel ɨ(**ɨ**) helps as the default epenthetic vowel and is inserted to break up unwanted combinations of the consonant in which phonological rules are governed. For example, inserting the vowel ɨ/**ɨ**/ like combinations of plosives at different places of articulation as in **kágɨtûGa (ካግእቱኻ)** CV-CV-CV-CV, 'you dried'. Without an epenthetic central vowel (**ɨ**) Awngi language speakers produce a combination of three consonant cluster words these clusters should be easily pronounced as **implá (እምፐላ)** VC-CCV, 'one'. The numbers of consonants are not as important as the ability to pronounce a given combination as determined by the sonority hierarchy [50][54]. To validate, the maximum syllable template CVC in the Awngi language, CCC clusters in the word-medial position are broken by inserting the central vowel **/ɨ/** like as in **impɨlá** which is responsible for a certain interpretation of several sound sequences.

3.4 Morphological system of the Awngi language

As Pinnacle [30] pointed out, morphology **(ምላድ ፅንት)** is the study of the interior structure of the word-formation process and how their structure reproduces their relation to other words. Morphotactic is the process of how the smallest unit is interacting with one another to convey the expected meaning of the words and a set of rules which constitute morphological grammar. The Awngi morphological structure tells a variety of morphological structures that exhibit in the language and the grammar consists of specific rules for each word category [21]. Basic word-formation processes in Awngi are affixation and compounding and words can be one of the following categories roots, suffixes, prefixes, and flexion are the basic morphemes[55].

3.4.1 Awngi Noun morphology

Like another language, the Awngi noun system is a word that is used to name or identify a person, place, or thing. The noun system of the Awngi language is highly inflected for number and gender

[56]. The gender, system exhibits two different genders, “masculine” and “feminine” which are marked by final morphemes /**-i**/ or a zero /**Ø**/ and **/-a/** respectively. For example, muliqisi **(ሙሊቂሲ),** a monk in English or መነኩሴ in Amharic, is marked by final **/-i/** that defines the “masculine” gender and molieqiesa **(ሞሌቂሳ),** a nun in English or **መነኩሲት** in Amharic, which marked by final **/-a/** to define the feminine gender. In the Awngi number system, a set of affixes is very common, singular and plural markers which are used to make Awngi nouns plural or singular. The most plural marker in the Awngi language is /-ka/, (**ካ**) that attached to the last consonant of the stem except in compounding word formation. There is no gender distinction in the plural form of the Awngi language that represents the pluralization of either the masculine or the feminine or both together[57].

Table 3 3:shows plural nouns with /ካ, -ka/

No	Singular form	Plural form
1	ታይ (Tay) ‘’sheep’’ ታያ (Taya) ‘’sheep’’	ታይካ tay-ka ‘sheep’
2	ፊሪሲ (firisi) ‘horse’ ፌረሳ (fieresa) ‘mare’	ፌሬስካ (fieries-ka), ‘horses’
3	ልኩ (lik^{w}), ‘’leg’’	ልኩካ (lik^{w}-ka),‘’legs’’
4	እንፀ (ɨntsu), ‘thin’	እንፀካ (ɨntsu-ka), ‘’Thin ones’’
5	ድኾሪ (dɣori), ‘’donkey’’ ድኾራ (dɣora) , ‘’donkey’’	ድኾርካ (dɣor-ka), ‘’donkeys’’

3.4.2 Verbal morphology of the Awngi language

As pointed to [58] the verbal morphological system of the Awngi language has a wealth of inflectional forms in terms of gender, person, number, and tenses. Word formation in the language

involves several different linguistic features including affixation, reduplication, and compounding because the Awngi language is morphologically very productive. In the Awngi language, verbs occupy the SOV grammatical order in a sentence. The attempt to substitute the position of a verb with another part of speech results in inconvenience for sentence formation in this language[54].

3.4.2.1 Mono-consonantal verbs (እምፐል አንቤብስታንት)

The Awngi verb formation exhibits the mono-consonantal to multisyllabic structure variations as demonstrated below.

Table 3 4:Mono-Consonantal Verb

Verb stem	CV Syllabic structure	Glosses
/ቡ-/	C-	'Carry it
/ቱ-/	C-	'get in'
/ካ-/	C-	'go'
/ኩ-/	C-	'kill'
/ኻ-/	C-	'take'

3.4.2.2 Possession (እግልት) Marking in Awngi language

Possession markers in the Awngi language are morphemes that attach to the possessed noun and pronouns to indicate that one thigs is owned by someone. Hence the possessor noun in this language is placed with the number and gender suffixes for both the possessor and possessed nouns. It indicates the current time possession noun in the Awngi language but does not show the past time in the language[59]. The occurrences of Awngi possessive markers with number suffixes and gender of both the pronouns and possessed nouns are not simple because the possession

marking process number and gender morphemes are also conditioned for grammatical and phenological aspects[7]. The possession markers in this language are **ይው/yi-u**/, **ኩው/ku**, **-kw/, ኺው/ ŋgi-u, ŋgi-t, ŋgi-ti/**, for the singular person and **እኖጂሱ/inojis-u**/, **እንቶጂሱ/intojis-u**/, ጋጂሱ/ **ŋgajis-u**/ for the plural person. The possession suffixes are /-u/, /-ku/, /-kw/, / /-t/, /-ti /, /-su/. The possession markers /yi-u/ and /ŋgi-u/ are for **1st** and **3rd** person pronouns in both gender markers and singular in a number. In the Awngi the suffixes /-u/are marked when the possessed noun is singular masculine and /-t, -ti/ is possessed pronoun is singular feminine. The suffixes /-ku, -kw/ is the possessed noun in the plural and the suffixes /-su/ is used after the plural pronoun but the possessed noun in the plural form. To illustrate it further consider the following examples

Table 3 5:The possession markers of the Awngi language

Awngi	Awngi	English
ይው ሜፃፍ	yi-u metsaf	My book
ኺው ሜፃፍ	ŋgi-w metsaf	her/his book
አቿው እርሳስ	aqa-u ɨrsas	The man's lead
አቓት እሏ	aqqa-t ɨllua or aq-ti-ɨllua	The woman's cow
እኖጂስኩ ቤራ	inojis-ku biera	Our oxen
እኖይሱ ቢሪ	inoyis-u biri	Our ox

3.4.2.3 Negative(አኻለ) marking in Awngi language

In the Awngi language, the negation marker is expressed using both free morpheme and bounded morpheme that is realized with the suffixes/ **-la/-ላ/, /-ga/-ጋ(t-)/, /-ya/-ያ/,** and /- **ti/-ቲ/.** In the Awngi language, the suffix **ti/-ቲ/** is a negative marking morpheme that occurs with an **imperfective** verb before the number and gender markers. For example, /tsafa**ti**w/, /ፃፋቲው/, /tsafa**t**it/, /ፃፋቲት/, /tsafa**ti**nku/, /ፃፋቲንኩ/ in this example **w(ው), t(ት)** and **nku(ንኩ)** are gender and number marker which the **/-t/** occurs before them. In this language the suffix **-ጋ-ga** negates a state of being and

like the past tense copula, it occurs as a word conjugating for both genders. The Awngi negative marker affixes **-la** is always occurring attached to negated verbs to express the present and future actions without showing the gender and number distinctions of the verb and it remains the same for all subjects except first person plural[7][6]. The affixes **-ga** occur without a complement affix which comes following numbers and gender affixes. The **-ti** occurs in clausal verbs following the root and preceding number and gender suffixes. The suffixes **ya/-ያ/** and **ga/ጋ(t-)/**occur with an inflectional verb in the case of number and gander to express the action that did in the past and the percent continuous respectively. To illustrate it further consider the following examples presented in positive and negative pairs of sentences.

A. Alemu **amaKarŋgi-e** kulucif **zumi**-e.
አለሙ አማኻርኜ ኹሉቺፍ ዙሜ።
Alemu Amharic-acc speak-imp every day.

C) Alemu **amaKarŋgi-a** kulucif zume-**la.**
አለሙ አማኻርኜ ኹሉቺፍ ዙመ-ላ።
Alemu do not speak Amharic-acc every day.

B. Abebe **metsa-fo** anviebama-**gi**-yaK.
አበበ መጻፎ አንቬባማ-ጊ-ያኽ።
Abebe is reading a book.

D) Abebe **metsa-fo** anviebama-**gi-ga**-tiwieK.
አበበ መጻፎ አንቬባማጊ-ጋ-ቲዌኽ።
Abebe is not reading a book.

E. lkni gniras dabdabe amaKargni-s **tsafu-Ka**.
ልክኒ ኺራስ ዳብዳቤ አማኻርኺስ ፃፉካ።
Lkni writs the litter for his son.

F) lkni gniras dabdabe amaKargnis tsafa-**ya**.
ልክኒ ኺራስ ዳብዳቤ አማኻርኺስ ፃፉያ
Lkni do not writ the litter for his son.

G. Kabedeqa dedegna-K
ካበደቓ ዴዴኜኽ
The kabede’s wife is short.

H) kabedeqa dedegna **ga-t**-yaK
ካበደቓ ዴዴኜ ጋ-ቲ-ያኽ
Kabede’s wife is not short.

In the above sentence A, B, E, and G are all positive and C, D, F, and H are all negative sentences. The negative marker morpheme **-la** in sentence **/C/is** used with imperfective mood and the negative morpheme in /D/ is used with the state of being a verb.

3.4.2.4 Imperative (አዜዣንቲ) verb system in Awngi

The imperative (አዜዣንቲ) verbs are used to give the command to the second person in both singular and plural form. It is a verb that passes commending information to someone and command of the speaker that influences the someone that we are speaking to perform or to not perform a specific action. Forms of imperatives in the Awngi language exhibit affirmative(positive) and negative forms. If the subject is second person singular most of the time the last letter of the verb should be the **/-ȸ/-ʦ/,/-ቕ/-Q/** or **Ø** type of Awngi letter suffixed and the suffix, **/-ን/** -n/ following fourth letter of Awngi alphabet depicts imperative information about second person plural forms[59]. For example:

Person	imperative verb suffixed	glosses
2nd singular	/ጌፈꭥር-ȸ/gefer-ts/	‘’you make’’
2ndplurar	/ጌፈꭥርዓ-ን/gefertsa-n/	‘’You(pl) make’’
2nd singular	/እንጂዪ-ȸ/ɨnjitsi-ts/	‘’you cook’’
2nd plural	/እንጂዪ-ን/ɨnjitsie-n/	“you(pl) cook’’
2nd singular	/እንኩዋ-ቕ/ɨnkuwa-q/	“you listen”
2nd plural	/እንኩዋቓ-ን/ɨnkuwaqa-n/	“you(pl) listen”

3.4.2.5 Verb of jussive

The jussive verbs are a grammatical mode for issuing orders, commanding, and a mode that expresses an intention to influence the listener's behavior. The jussive verbs are similar to imperative mood but unlike imperative jussive, verbs work with all persons[22][8]. To illustrate it further consider the following examples presented in the table below.

Table 3 6:jussive verb form of Awngi language

Person	Jussive verb with suffix in Awngi	Glosses
1st singular	Sug-is=ሱግ-ስ	let me pound
2nd singular	Sug-t-is=ሱግ-ት-ስ	let you pound
3rd singular.M	Sug-is=ሱግ-ስ	let him pound
3rd singular. F	Sug-t-is=ሱግ-ት-ስ	Let her pound
1st plural	Sug-n-is=ሱግ-ን-ስ	let us pound
2nd plural	Sug-t-in-is=ሱግ-ቲ-ኒ-ስ	let you (pl) pound
3rd plural	Sug-in-is=ሱጊ-ኒ-ስ	let them pound

In the above table, the Awngi jussive verb can be used for all persons with suffixed morpheme **/-is/-ስ/**. In this, the jussive verb exhibits the same suffixes of the first singular and third singular masculines as that of the second singular and third feminism.

3.4.2.6 Negative markers in the imperative's verb of Awngi

The negative marker in the imperative verb is different from the affirmative imperative negative form used in the second person singular and plural form. The suffix which depicts the second person negative imperative form for singular is the last letter of the verb should **/-ቿ/,/-Qie**/ and /-**ቄ**/,/-wie/or **Ø** type of Awngi letter and the suffix /-ን/-n/ following 5th letter of Awngi alphabet depicts imperative information about second person plural forms[7][8]. For example:

Table 3 7:Negative markers in the imperative's verb of Awngi

Person	Negative suffixed of imperative verbs of Awngi	
2nd singular	/ጌፈርዼ/gefer-tsɐ/	''You not make''
2nd plural	/ጌፈርዼ-ን/gefertse-n/	''You(pl) not make''
2nd singular	/እንኩዋቼ/ɨnkuwa-qe/	''you not listen''
2nd plural	/እንኩዋቼ-ን/ɨnkuwaqe-n/	'you (pl) not listen''
2nd singular	/እንጂኩዌ/ɨnku-we/	''you not sit''
2nd plural	/እንጂኩዌ-ን/ɨnkuwe-n /	''you (pl) not sit''

3.4.2.7 Negative tenses form of Awngi language

Tenses are the different kinds of verb forms are used to express an action about the present, past, future, and continuous. In the Awngi language, tenses are not exhibited negative prefix markers to the verb. All negative markers in this language are depicted suffix information such as /-ላ/-la/, /-ያ/-ya/, /ጋ(ቲ)/-ቲ/t/. The suffix which used with percent and future tense, past, continuous, and used with bounded morpheme respectively. The negative morpheme /-la/ does not show the distinction between number and gender. For example:

A. ɨnoji gazite xulucif anvevna-**la**

እኖጂ ጋዚቴ ኹሉቿፍ አንቨቨና-ላ።

we **do not** read the newspapers every day.

B. Lkniji ayŋa gebela kaska-**ya**.

ልክኒጂ አይጋ ጌቤላ ካስካ-ያ።

They did **not** go to town yesterday

C. ɨnoji gazite anvevnanagi-**ga**-ti-wex

እኖጂ ጋዚቴ አንቨቨናናጊ-ጋ-ቲዌኽ።

We are **not** reading newspapers.

D. des-átîⱴá=ዴሳቲኻ

des-nátînⱴá=ዴስናቲንኻ

des-átînkʷî=ዴሳቲንኩዊ

3.4.3. Verbal inflections of Awngi language

In the Awngi grammar, an inflection is a change in the form of a word that shows its grammatical function. The person can be represented as the first, second, or third with number and gender. In the Awngi language, the marker of first-person singular is **/-ኻ/-ሃá**/, and the marker of the plural is/-**ná**/-ና. The second person singular is **/- tá**/-ታ/ and the plural's /**n**/-ን/. The third person singular is /- **áwî**/-ዊ /, /- tátî/**-ቲ** / and plural/- ánk^{w}î/-**ንኩዊ/.** There is no gender distinction in the plural form of the Awngi language represents the pluralization of either the masculine or the feminine or both together. The same phenomena happen with the first- and second-person singular which are not identified by masculine and feminine gender. This depicts the neutrality of gender in the case of plural person and first and second person but unlike this, the third person exhibits the gender variation for masculine gender that represents / **-áwî**/-ዊ / and **/- tátî/-ቲ** / to indicate the feminine gender[22]. To illustrate it further consider the following examples presented in the table below.

Table 3 8:Verbal inflections of Awngi language in the case of gender and number

Person	The inflected form of the Awngi verb	Glosses
1st singular.M	Kasá-ሃáሃ=ካሳኻኽ, kas-a=ካሳ	I go, I had gone
1st singular. F	Kasá-ሃáሃ=ካሳኻኽ, kas-a=ካሳ	I go, I had gone
2nd singular.M	Katá-ሃáሃ=ካታኻኽ, kat-a=ካታ	you go, you had gone
2nd singular. F	Katá-ሃáሃ=ካታኻኽ, kat-a=ካታ	you go, you had gone.
3rd singular.M	kasa-weሃ=ካሳዌኽ, kas-a=ካሳ	He goes, he had gone.
3rd singular. F	Kata-teሃ= ካታቴኽ, kat-a=ካታ	She goes, she had gone.

1st plural	Kan-aɣáɣ=ካናኻ, kan-a=ካና	we go, we had gone.
2nd plural	Kata-nɣáɣ=ካታንኻ, katk-a=ካትካ	You(pl) go, you(pl) had gone
3rd plural	Kasanku-weɣ=ካሳንኩዌኽ, kask-a=ካስካ	they go, they had gone

3.4.3.1 Gender suffix in Awngi verb

The Awngi language has two genders masculine and feminine. So, to point out a certain gender in the word contain a suffix but gender reference suffix is not common in all person. For example, /ት/ቲ/, and **/-t/** indicates feminine gender in the third person. The suffix **/-ው/-ይ/, ዊ/, /-w/, /-y/,/-wi/** represents the masculine gander for person. Example for 3rd person masculine **/ፃፋ-ው/, /tsfa-w/, ‘’ He writes’’** and for **3rd** feminine **/ፃፌ-ት/, /tsfe-t/, ‘’She writes’’.** In the Awngi language, there is no gender distinction in the plural form of all persons.

3.4.3.2 Number suffix in Awngi verb

The number is inflected in two ways in the Awngi a singular and a plural form. A suffix is a word part that is placed at the end of a base word and there are verb suffixes to depict the number in the word. Most of the time the number suffixes show plural numbers rather than singular ones. To explain it further consider the following examples presented in the table below.

Table 3 9:Tabular description of number suffixes in the Awngi verb

Person	Stem word	Glosses	Singular	Plural
1prs.m	/ክንት/, /knt/	‘’Teach’’	/ክንት-**ፄ**/, /knt-**tse**/	/ክንት-**ፅ-ኔ**/, /knt-**ts-ne**/
1prs.f	/ክንት/, /**knt**/	‘’Teach’’	/ክንት-**ፄ**/, /knt-**tse**/	/ክንት-**ፅ-ኔ**/, /knt-**ts-ne**/
2prs.m	/ክንት/, /**knt**/	‘’Teach’’	/ክንት-**ፂ**/, /knt-**tsi**/	/ክንት-**ፄ-ና**/, /knt-**tse-na**/

2prs.f	/ክንት/, /**knt**/	‘’Teach’’	/ክንት-**ቼ**/, /knt-**tsi**/	/ክንት-**ቼ-ና**/, /knt-**tse-na**/
3prs.m	/ክንት/, /**knt**/	‘’Teach’’	/ክንት-**ቼ**/, /knt-**tse**/	/ክንት-**ች-ና**/, /knt-**tsa-na**/
3prs.f	/ክንት/, /**knt**/	‘’Teach’’	/ክንት-**ቼ**/, /knt-**tsi**/	/ክንት-**ች-ና**/, /knt-**tsa-na**/

/-tsne/ፅኔ, /-tsena/ቼና, and /-tsana/ችና in the above example depict the appearance of plurality for the first person, second person, and third person respectively. There is no existence in the singular form of the word and it is possible to conclude that **/-tsne/ and /-tsi/** shows the Singularity of the word in the Awngi language concerning the specified person. The suffix for present tense third person singular she/he exhibits most of the time 5th letter of the Awngi language.

3.4.3.3 Tenses in Awngi

Tenses are different kinds of verb forms, these verb forms are used to express an action whether we may speak about the past, present, future, or continuous actions. Tense can be non-past time and past time which the Awngi language exhibits it. In the Awngi language, there are two identified aspects of a verb form perfective and imperfect [55][58]. For example:

Person	**stem**	**imperfect**	**perfective**	**Negative form**
1prs.m	/make/, /ገፈርቼ/	/geferts-**e**/, /ገፈርቼ-/	/geferts-**ሄ**/, /ገፈርፅ-**ኾ**/	/ገፈርግ-**ላ**/ /ገፈርግ-**ያ**/
1prs.f	/make/, /ገፈርቼ/	/geferts-**e**/, /ገፈርቼ-/	/gefer**ts**-**ሄ**/, /ገፈርፅ-**ኾ**/	/ገፈርግ-**ላ**/ /ገፈርግ-**ያ**/
2prs.m	/make/, /ገፈርግ/	/geferts-**a**/, /ገፈርግ-/	/geferts-**ሄ**/, /ገፈርፅ-**ኾ**/	/ገፈርግ-**ላ**/ /ገፈርጊ-**ያ**/
2prs.f	/make/, /ገፈርግ/	/geferts-**a**/, /ገፈርግ-/	/gefertsi-**ሄ**/, /ገፈርጊ-**ኾ**/	/ገፈርግ-**ላ**/ /ገፈርጊ-**ያ**/
3prs.m	/make/, /ገፈርግ/	/gefertsa-**wi**/, /ገፈርግ-ዊ/	/geferts-**ሄ**/, /ገፈርፅ-**ኾ**/	/ገፈርግ-**ላ**/ /ገፈርግ-**ያ**/

3prs.f /make/, /ገፈርዪ/ /geferts-**i**/, /ገፈርዪ-/ /gefertsi-**ɣʷá /**, /ገፈርዪ-**ʔ**/ /ገፈርዓ-**ላ**/ /ገፈርዪ-**ያ**/

1prpl /make/, /ገፈርዕ/ /geferts-**ne**/, /ገፈርዕ-**ኔ**/ /gefer**ts-naɣʷá**/, /ገፈርዕ-**ናʔ**/ /ገፈርዕ-ናላ/ /ገፈርዓ-**ያ**/

2prpl /make/, /ገፈርዬ/ /gefertse-**n**ɣ**a**/, /ገፈርዬ-ንʔ/ /gefertsi-**ɣʷá**/, /ገፈርዪ-**ʔ**/ /ገፈርዓ-**ላ**/ /ገፈርዪ-ካ**ያ**/

3prpl /make/, /ገፈርዓ/ /gefertsa-**nkʷi**/, /ገፈርዓ-**ንኩዊ**/ /geferts-**ka**/, /ገፈርዕ-**ካ**/ /ገፈርዓ-**ላ**/ /ገፈርዕ-ካ**ያ**/

The illustration shows that the tenses suffix for both genders for the first person and second person indicates the identical suffix concerning its person in the past, percent, and negative form of the verb. In addition to this, the third person masculine and third person feminism is the same tenses suffix with first person and the second person respectively in the past, percent, and negative forms of the verb. In this language, most of the time perfective marker is **-ɣʷá /-ʔ/** and the imperfective marker is the fourth letter of the Awngi language in the above table the -ʦa/-ዓ/ is the imperative marker.

Person stem present continuous past continuous

1prs.m, /make/, /ገፈርዓ-/ /gefertsa-**tagi**-**ya**ɣ/, /ገፈርዓ-**ታጊ-ያ**ʔ/ /gefertsa-**tagi**-ɨxiɣo/, /ገፈርዓ-**ታጊ-እሺ**ʔ/ **Negative** /ገፈርዓ-**ታጊ-ጋ**-ቄʔ/, /ገፈርዓ-**ታጊ**-እሺ**ያ**/

1prs.f /make/, /ገፈርዓ-/ /gefertsa-**tagi**-yaɣ/, /ገፈርዓ-**ታጊ**-ያʔ/ /gefertsa-**tagi**-ɨxiɣo/, /ገፈርዓ-**ታጊ-እሺ**ʔ/ **Negative** /ገፈርዓ-**ታጊ-ጋ**-ቄʔ/, /ገፈርዓ-**ታጊ**-እሺ**ያ**/

2prs.m /make/, /ገፈርዬ-/ /gefertse-**tagi**-yaɣ/, /ገፈርዬ-**ታጊ**-ያʔ/ /gefertse-**tagi**-ɨxiɣo/, /ገፈርዬ-**ታጊ**-እሽʔ/ **Negative** /ገፈርዬ-**ታጊ-ጋ**-ቄʔ/, /ገፈርዬ-**ታጊ**-እሽ**ያ**/

2prs.f /make/, /ገፈርዬ-/ /gefertse-**tagi**-yaɣ/, /ገፈርዬ-**ታጊ**-ያʔ/ /gefertse-**tagi**-ɨxiɣo/, /ገፈርዬ-**ታጊ**-እሽʔ/ **Negative** /ገፈርዬ-**ታጊ-ጋ**-ቄʔ/, /ገፈርዬ-**ታጊ**-እሽ**ያ**/

3prs.m /make/, /ጌፈርፃ-/ /gefertsa-**magi**-yaɣ/, /ጌፈርፃ-**ማጊ**-ያኽ/ /gefertsa-**magi**-ixiɣo/, /ጌፈርፃ-**ማጊ**-እሺኾ/ **Negative** /ጌፈርፃ-**ማጊ**-ጋ-ቄኽ/, /ጌፈርፃ-**ማጊ**-እሺያ/

3prs.f /make/, /ጌፈርጼ-/ /gefertse-**tagi**-yaɣ/, /ጌፈርጼ-**ታጊ**-ያኽ/ /gefertse-**tagi**-ixiɣo/, /ጌፈርጼ-**ታጊ**-እሺኾ/ **Negative** /ጌፈርጼ-**ታጊ**-ጋ-ቄኽ/, /ጌፈርጼ-**ታጊ**-እሺያ/

1prpl /make/, /ጌፈርፅ-/ /geferts-**nanagi**-yaɣ/, /ጌፈርፅ-**ናናጊ**-ያኽ/ /geferts-**nanagi**-ixiɣo/, /ጌፈርፅ-**ናናጊ**-እሺኾ/ **Negative** /ጌፈርፅ-**ናናጊ**-ጋ-ቄኽ/, /ጌፈርፅ-**ናናጊ**-እሺያ/

2prpl /make/, /ጌፈርጊ-/ /gefertsi-**kamagi**-yaɣ/, /ጌፈርጊ-**ካማጊ**-ያኽ/ /gefertsi-**kamagi**-ixiɣo/, /ጌፈርጊ-**ካማጊ**-እሺኾ/ **Negative** /ጌፈርጊ-**ካማጊ**-ጋ-ቄኽ/, /ጌፈርጊ-ካ**ማጊ**-እሺያ/

3prpl /make/, /ጌፈርፅ-/ /geferts-**kamagi**-yaɣ/, /ጌፈርፅ-**ካማጊ**-ያኽ/ /geferts-**kamagi**-ixina/, /ጌፈርፅ-**ካማጊ**-እሺና/ **Negative** /ጌፈርፅ-**ካማጊ**-ጋ-ቄኽ/, /ጌፈርፅ-ካ**ማጊ**-እሺያ/

The tenses suffix in the above illustration shows that masculinism and feminism for the first person and the second person are identical concerning tense in the present continuous, past continuous, and negative form of the verb. In the Awngi language, the verb form remains the same both in affirmative and negative form but '**ያኽ**' in continuous tenses is replaced with '**ጋ-ቄኽ**' irrespective of the subject. In Awngi '**ያኽ**' depicts positivity whereas '**ጋ-ቄኽ**' indicates negative. In the Awngi language, most of the time the percent and future tenses represent the same suffixes. The present continuous marker suffix are /-**tagi** /-**ታጊ/, /-magi/-ማጊ/, /- nanagi** /-ናናጊ/, and **/ -kamagi**/-ካማጊ/**.** The suffix **/-tagi /-ታጊ/,** is used for the first-person singular, second-person singular, and third-person singular feminism, and the suffix **/-magi/-ማጊ/** marks only for third-person singular masculine. The suffix **/- nanagi** /-ናናጊ/ is the first-person plural continuous marker and the suffixes **/ -kamagi**/-ካማጊ/ is the continuous marker for second person plural and third-person plural marker as illustrated above example. Most of the time for all personal classes adding the **/-እሺኾ/ -ixiɣo**/

to the present continuous marker results formation of the past continues the form of the verb except the third person plural which needs unique past continuous markers suffix **/-እሺና/-ixina/.**

3.4.3.4 Passive in Awngi

In passive construction, the object of the active verb becomes the subject. In the Awngi, the suffix /**-st/-ስት/** is the passive marker attached to the active transitive verbs. For example:

The active verb root	**Gloss**	**Passive verb stems**	**Gloss**
gibit/ጊቢት/	build	gibit-ist/ጊቢት-እስት/	be built
ziq/ዝቕ/	drink	ziq-ist/ዝቕ-እስት/	be drunk
dunʦ/ዱንፅ/	break	dunʦ-ist/ዱንፅ-እስት	be broken

In the above example **/-st/-ስት/** suffix in the language is used to attach to active transitive verbs to derive passives verbs by preceding the epenthetic vowel **እ(i)** attached to verbs ending in a consonant to avoid impermissible consonant clusters of this language.

3.4.3.5 Reflexives in Awngi language

The reflexives are a verb where the subject and the object relate to the same entity and a variety of languages use the reflexive forms for suffixes, case endings, word order, and pronouns. This language uses suffixes to express a reflexing meaning. The reflexive markers can be pronounal and verbal reflexives[59]. The former is a marker that exhibits the properties of pronouns and verbal reflexives is a part of the morphology associated with verb suffix which this study is going to in this language. In this language, the verbal reflexive marker is **/-ɣa/**[59]. for example:

Verbal roots	**Gloss**	**verbal reflexives**	**Gloss**
jew-/ጀው-/	buy	jew-t-a-**ɣa**-t-a/ጀውታኻታ	she has bought for herself.

gibit-/ጊቢት-/ build gibit-a-**ɣa**-s-a/ጊቢታኻሳ he has built for himself.

Kew-/ኬው-/ cut kew-a-**ɣa**-s-u-n-a/ኬዋኻሱና they have cut for themselves.

gifir-/ጊፊር-/ make gifir-ts-e-**ɣa**-t-a-n-a/ጊፊርዔኻታና you have made yourself

3.4.3.6 Causatives in Awngi language

As pointed out by [59] the causatives are a situation that involves two factors the cause and its effect. According to morphological connection causative markings in Awngi are composed of a causative morpheme and a base verb. In the Awngi language, the causative verbs are derived by suffixing **/-ʦ/-ɸ/** to the verb base. Causatives can be categorized two as single and double. The addition of **/-ʦ/-ɸ/** to the verb stem is single causative and the doubly causative is the addition of the same suffix **/-ʦʦ/-ɸɸ/** to the base verb as demonstrated in the following example

Table 3 10:Tabular description of Causative forms Awngi verb

Root verb	**Gloss**	**Single causative**	**Gloss**	**Double causative**	**Gloss**
lɨmi-=ልማ	Close	lɨmi-ʦ= ልማɸ	Cause someone to close	lɨmi-ʦʦ= ልማɸɸ	Cause someone to Cause some other to close
kuŋgi-ኩኻጊ	Jump	kuŋgi-ʦ= ኩኻጊɸ	Cause someone to jump	kuŋgi-ʦʦ=ኩኻጊɸɸ	Cause someone to Cause some other to jump

ɣuri-='ʔŕ·ɾ	Sleep	ɣuri-ʦ='ʔŕ·ɾɵ	Cause someone to sleep	ɣuri-ʦʦ='ʔŕ·ɾɵɵ	Cause someone to Cause some other to sleep
giŋi-ɻ'ʔ	Run	giŋi-ʦ=ɻ'ʔɵ	Cause someone to run	giŋi-ʦʦ=ɻ'ʔɵɵ	Cause someone to Cause some other to run

As seen above table single causative verb is further causative to that forms the double causative verbs. The result of the causative suffix does not change part of the speech category nevertheless it changes the subcategorization of the verb forms.

3.4.3.7 Reciprocal in Awngi language

According to the [22][59] the Awngi language has the suffix /- ŋ/ is identified as the reciprocal marker that ɨs suffixed to the verb. For example, **desŋiŋ/ ዴስሻሻ/** ‘to study each other, **desiŋ/ዴስሻ/.** In this example, we understudy that before the reciprocated the base verb should be changed to the infinitival numerals. For more illustrate

dig----------> dig-iŋ ----------> dig-iŋ-iŋ

nearer nearer-inf nearer-inf-recip

‘nearer’ “to nearer” ‘to nearer each other

The second /-ŋ/are markers of reciprocal whereas the first is the markers of infinitival nominal.

3.4.3.8 Causatives reciprocal

In Awngi, the verb causative markers **/-ʦ/** can be combined with the reciprocal markers/-ŋ/ to form verb derivatives. The sequential formation of causative reciprocals is infinitival nominals

are derived from base verbs, causative verbs from infinitive nominals, and the causative reciprocal from the causative verbs. For example**:**

tas-/ታስ-/------->tas-**iŋ**/ታስ-ሻ/-------->tas-**iŋ-its**/ታስ-ሻ-ፅ/-------->tas-**iŋ-its-iŋ**/ታስ-ሻ-ፅ-ሻ/

kick	kick-**inf**	kick-**inf-caus**	kick-**inf-caus-recp**
kick	to kick	cause to kick	to cause them to kick each other

3.4.3.9 copula verb (verb to be)

The link verbs occur together with the main verbs in a clause. In the Awngi language, copular verbs are together with the subject. This verb is a grammatical alternation that exists in the Awngi language. There are suffixes of copula verbs in this language which are the **/-ɣ/-ኽ/, /-yaɣ/ -ያኽ/** this suffix in the verb represents future action, and the/ix-/እሸ-/ suffix is depicted the past tense. They are representing the same for all persons except third person singular masculine and third person plural form. For example

Table 3 11:Example to illustrate the copula of Awngi

Person	Word	copula	Glosses
1st sig	ካሳኻ-ኽ	-ኽ	I will go
2nd sig	ካሳኻ-ኽ	-ኽ	I will go
3rd sig.m	ካሳቄ-ኽ	-ኽ	He will go
3rd sig.f	ካሳቱ-ኽ	-ኽ	She will go
1st Plu	ካናኻ-ኽ	-ኽ	We will go
2nd plu	ካታንኻ-ኽ	-ኽ	You will go
3rd plu	ካሳንኩቄ-ኽ	-ኽ	They will go

CHAPTER FOUR: DESIGN AND METHODOLOGY

4.1 Introduction

This chapter deals with the detailed description of the design and methodology of the morphological generation and analysis, the source where the data was collected, the finite state technology as implementation tool and techniques used, the lexicon definition of the Awngi language, the set of the continuous class definition, graphical representation of the system, the definition of the morphotactic rule, designing language-specific rules that modify the morphotactic rule ordering and morphological system evaluation based on the comparing system generated words with the word collected in different domains were discussed. The finite state technology is the preferred approach that required a special designing technique based on a deep representation of linguistic knowledge of a language and word formation. During the design of the morphological system using the rule-based finite-state approach, the following basic designing issue must be considered to be addressed[46].

The first phase is designing the lexicon/ morphotactic part and the second part is the alternation design phase. However, depending on the nature of the finite state transducer the morphological system is bidirectional. This approach creates a conducive environment for morphological analysis and morphological generator working at the same time by using a regular expression with minor modifications the in the implementation part[48].

4.2 Methodology

There are several related articles and journals primarily the researcher reviewed and focused on the study topic to collect input for designing a morphological system. For this study data was collected from a different source. 574 root words were collected from supplementary grammatical

books and 152 root words were collected from the newspaper. These root words were selected by the language experts from the collected document which is used for implementation. Using this unique lexicon list of data, the researcher designs lexicon and morphotactic rules to experiment with the morphological generator of the Awngi language. To report the implementation of the study the researcher made the experiment that raw test output from a single test with the entire lexicon implementation and the quantitative analysis which deals with the experimental result. The next phase represents an experimental result in terms of a numeric value that can be represented by percent.

4.3 Data sources

Natural language processing is a data-intensive field. For the development of a morphological analyzer and generator for the Awngi language, the researchers collected the data from different sources in the form of both hard and soft copies. The domain of the data sources was considered under different categories such as the Awngi newspaper called cirbewa (ቸርቤዋ), the Injibra university teaching materials, from high school Awngi textbook and we discussed more morphological rules of the Awngi language with experts. The collected data passes through preprocessing step and at the end of the process pure Awngi text is identified. The identified document is tokenized into words by using python programing language by the researcher and tokenized word is given to Awngi language experts to differentiate verbs from other parts of speech category and identify root words from the verb. Lastly, the language expertise identifies distinct 726 root words with input for morphological analysis and generation.

4.4 Implementation Tools and techniques

The modeling and implementation of the Awngi morphological system were conducted based on foma finite-state transducer implementation software. To run the foma finite state machine the researcher selects the Linux-based operating system special Ubuntu 22.04. Foma is a commercial-free and open-source finite-state toolkit that has facilities for a variety of natural language processing applications ranging from morphological analysis to finite-state parsing. Foma finite state can be considered a programming language that has its compiler and it is full of regular expression finite state library intended for multi-purpose[60]. In addition to this, the researcher uses VirtualBox 7.0 to run lexicon files and rules that guide the pattern of linguistics of the Awngi verb word formation. The text editor which has integrated with VirtualBox is used to design the lexicon and to write language-specific rules that govern root-to-morpheme integration. The python programming language is used for tokenizing the documents.

4.5 Lexicon design

The lexicon is a stem word that is used as input to the system. The input word is segmented by the system into its corresponding grammatical structure based on the part of the word input. Morphological generation is revere process of analysis that takes input from the list of every lexicon word of the Awngi language and affixes together with basic information about words. The valid lexeme is followed with an appropriate sequence of tags and the system generated different word forms. For the experimentation different basic information words or affixes are considered.

4.6 Continuous class

The inflectional of the word includes many morphemes that caries basic grammatical information for the given word. The words in the lexicon script is operates under many continuous classes that make the backbone of the morphological analysis. To model this continuous class the legal

sequence of suffixes assigned to the respective class is needed. As stated above the morphological nature of the language has different inflectional affixation. Those fragments of morphemes are grouped into different continuous classes. For the experimentation of this system, we stated the same continuous class. These are TENSC tense suffixes class, PRNSC Person suffixes class, DVSC Derived verb suffixes class, NGSC negative verb suffixes class, IMPSC imperative suffixes class, NUSC number suffixes class, GNDSC gender suffixes class, CAUSC causative suffixes class.

4.7 Tags in Awngi language

Tags are the representation of basic grammatical information of words corresponding to morphemes within the words. To mark each morpheme in the word we have used specific tags including the following.

+V to represent the root verb is the on the class of the verb / ግስ ካል ሜንዴባ ካንትፂ/

+ 1prs to represent the first person singular /እምፐለንቲ ሜንዴቭ ኒፂሊ ቼፍ/

+ 2prs to represent the second person singular/ላቻንቲ ሜንዴቭ ኒፂሊ ቼፍ/

+3prs to represent the third person singular /ሹኻንቲ ሜንዴቭ ኒፂሊ ቼፍ/

+ Mas to represent the gender masculine /ቻሽታርቱ ዖቴ ካንትፂ/

+ Fem to represent the gander is feminine/አንቾቶ ዖቴ ካንትፂ/

+1pl to represent the first-person plural/እምፐለንቲ ሜንዴቭ ሚንቺቶ ቼፍ ካንትፃው/

+2pl to represent the second person plural/ላቻንቲ ሜንዴቭ ሚንቺቶ ቼፍ ካንትፃው/

+3pl to represent the third person plural/ሹኻንቲ ሜንዴቭ ሚንቺቶ ቼፍ ካንትፃው/

+ NGE to represent the negative marker /አኻሊ ካንትፂ/

+Impfct to represent the present and future tense /እንዴ ፍና እስታ ይንታው ጊዝዳ አኮሜችስትኻ ካንትፂ/

+Prfct to represent past tense /ፈይኹደ ጊዝስ አኮሜችስትኛ ካንትቂ/

+PrsCont to represent present continuous tense /ሻሺ አኮሜችስታማጊ አኽኛ ካንትቂ/

+PasCout to represent the past continuous tense

4.8 Alternation rule in Awngi

The output string of the lexicon transducer is modified based on the orthographic, phonological, and morphophonological rules of the language. This is a spelling rule that is used to model the changes in words. In this language, the first person singular masculine form of a verb that has the **‘ቤ’/be** sound changed into the **‘ፔ’/pe** sound in the inflected form of the verb. For example**, በምቤ-->በምፔ/** bambe--->bambe ‘’he swims’’ this is voiced consonant phoneme **ቤ/be** devoicing into **‘ፔ’/pe**. Some verb geminations with the sound **-ፅ/-ʦ**/ consonant when the word ends with the second letter of the language to Convery double causative for example word **ልሚፅ**/lmiʦ/ ‘he closes’ ልሚፅፅ/lmiʦʦ/ “Cause some other to close”.

4.9 Graphical view of Awngi morphological system

Over all graphical representation of the system that shows the internal combination of the lexicon form of the verb with the language-specific rules is reparented as follows.

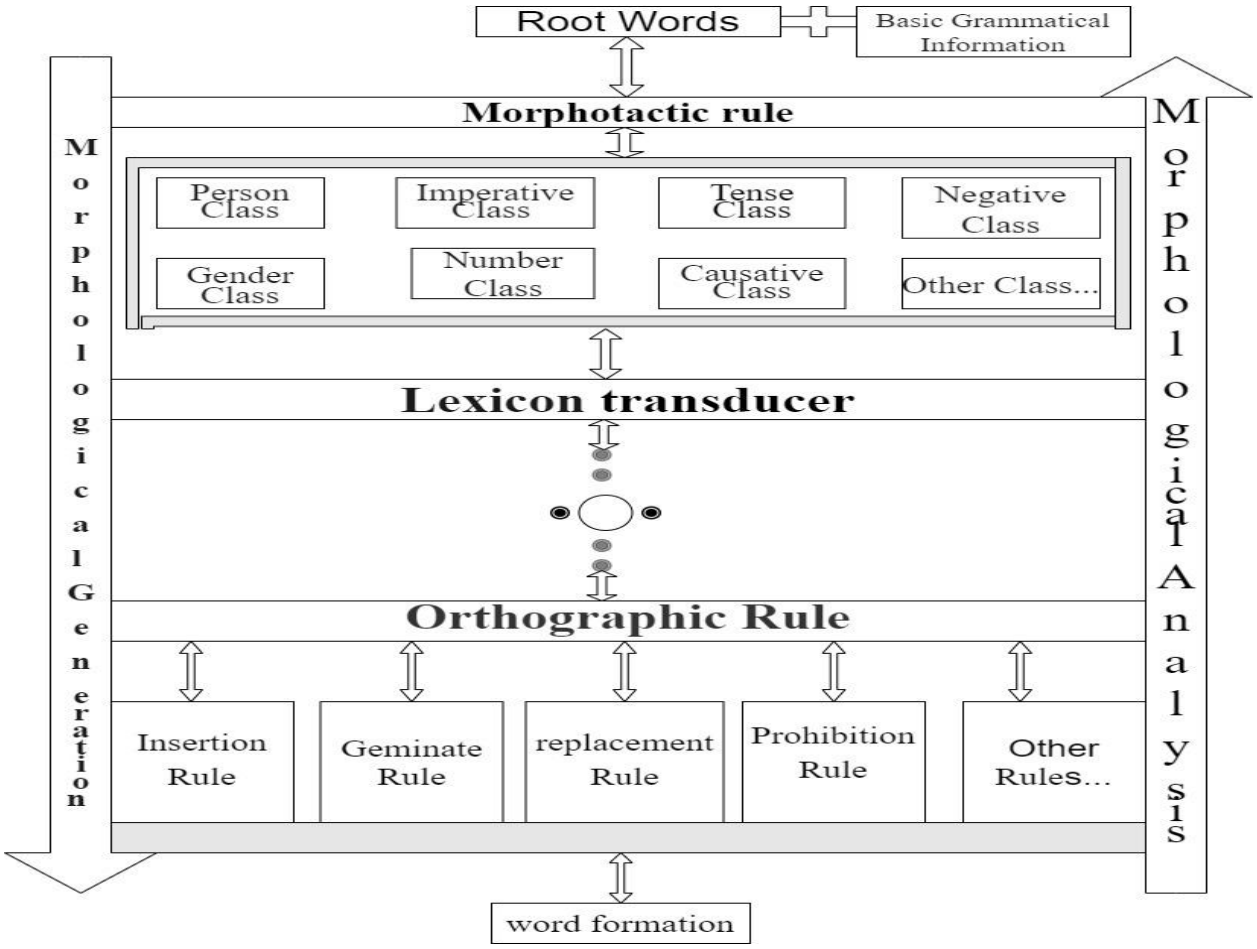

Figure 4. 1:Diagrammatic representation of Morphological Analyzer and Generator

4.10 Morphotactic components

In this section, the researcher discusses the set of continuous classes and the way of scripting them. The language has well-defined ways of sequencing the morphemes. The lexicon part is the list of a valid stem of the language and the Morphotactic rule is the ordering of the morphemes that describes which class of continuous class can follow other classes of other continuous classes inside the implementation. The Morphotactic and lexicon are scripted inside of the text editor under VirtualBox following a legal sequence of the tags. It includes different continuous classes with their morphemes lists. However, when the Morphotactic is compiled in the foma that

produces the Lexicon transducer. This lexicon transducer combines with the rules devolved by the researcher. The components of the Morphotactic are explained as follows.

4.10.1 Person class

1st, 2nd, and 3rd in both singular and plural are encompassed under the person class that contains all suffix that points to a specific person at a certain level. In the person class, most of the suffix for 1st and 3rd singular masculine is pointed to **Ø** which means the root words include these properties in this parson suffix class in the Awngi language. The suffixes **/-ቲ/-ቱ/** points for second-person and third-person singular feminism. Suffix /- ኔ/- ኑ/ for first person plural, /- ና/- ኩ/ for second person plural, and /- ካ/- ኝ/ for third person plural.

4.10.2 Imperatives

As stated in chapter 3 imperative verbs are used to give the command to the second person in both singular and plural form. Most of the time the suffix /-ዕ/-ʦ/, /-ቕ/-Q/ or Ø for second singular and the suffix, **/-ኝ/** -n/ second person plural forms.

4.10.3 Causatives

The causative verbs are derived by suffixing **/-ʦ/-ዕ/** to the verb base. The single suffix **/-ʦ/-ዕ/** represents the single Causatives for the second person singular and doubling the suffix of the single causative can produce the double causative for the second person singular suffix **/-ʦʦ /-ዕዕ/**.

4.10.4 Tense class

Tense can be non-past time and past time which the Awngi language exhibits it. In the Awngi language, there are two identified aspects of a verb form perfective and imperfect. The listed suffixes describe the different kinds of tense forms in this language. Most of the time the imperfect tense is marked with the 5th letter in the Awngi language and the suffixes /-ሄʷá /-ኽ/ for perfective,

/-tagi/-ታጊ/, /- magi/-ማጊ/, **/-nanagi /-ናናጊ/,** and **/-kamagi-/-ካማጊ**/, for present continuous**,** and **/-ɨxiɣo/-ኦሺዮ/** for past continues.

4.10.5 Gender class

In the Awngi language, there is no gender distinction in the plural form of all persons. But the suffix **/-t/-ት/-ti/-ቲ/** indicated the singular person feminism gender in the third person. The suffix **/-ው/-ይ/, ዊ/, /-w/, /-y/,/-wi/** represents the masculine gander for person.

4.10.6 Negative class

The negative marker in imperative verbs is different from that of the tense negative markers. There are four negative markers in the tenses class which are stated he chapter 3 above. The imperative verb negative marker is the same as the plural number marker and the positive imperative marker suffix /-ን/-n/ but following with the 5th letter of the Awngi alphabet depicts imperative negative information about second person plural forms.

4.11 Designing language-specific rules

Orthographic and phonological rules are stated here to govern word formation in a language. This contains an alternation rule that describes which is the spelling changing sound of the language is considered to the formation of words.

4.11.1 Epenthetic rule

The clusters of more than two consonants in this language need the epenthetic vowel phoneme ɨ/ኦ. But the structure of this language states that the central vowel /ɨ/ ኦ / does not appear word-finale. This means the epenthetic vowel in a language allows consonant clusters in this position with no need to insert the epenthetic vowel. The structure of the language allows all other vowels at the word-final but in the word-initial position the only acceptable vowels are /a/ኦ and ɨ /ኦ. In

any potion, the combustion of nasal and geminate consonants is never separated by the epenthetic vowel ɨ/ ʌ. This is the functionality regarding with epenthetic vowel ɨ /ʌ in this language while becoming a word-initial it is the evidence for the same phonetic sequence. When the sound ɨ /ʌ becomes in the middle of more than two consonant clusters to validity more than two consecutive consonants appear in the Awngi language. According to this discussion, the formation of a rule for the sound ɨ /ʌ is proposed with the insertion, replacement, and deletion. If the ɨ /ʌ sound become at the end of the root of the word or the final position of the sound the rule will be designed to remove this sound. The insertion part will be designed in the epenthesis part. The nature of the suffixes for the gemination consonants and nasal combination determines the replacement of this sound for the other.

4.11.2 Replacement rule

Different replacement rule is encompassed to achieve the morpho-phonological processes of the language. The verb is in the third-person singular masculine's ends with voiced sound consonant **b/ብ, d/ድ** will be completely changed to first person singular by voiceless **p/ፕ, t/ት** respectively. when the verb is in any imperative form the root final consonant is **ʤ/ዝ, ʤ/ጅ** changed into voiceless **ʦ/ፅ, ʧ/ች** respectively. The replacement rule is as follows.

Rule 1: ብ → ፕ || con __ "^";

ድ → ት || con __ "^";

ዝ → ፅ || con __ "^";

ጅ → ች || con __ "^";

The **con** is variable regarding the set of components phonemes in the Awngi language.

4.11.3 Nasal Assimilation rule

In the Awngi language nasal occupies only the three places of articulation. This nasal assimilation sound in this language is phoneme n/ንʼ, m/ምʼ, ŋ/ኽʼ. They are inserted in different positions to complete the language-specific rule with different consonants. n/ንʼ comes at the potion before the alveolar (ድድኩ) consonants sound to achieve second and third plural. The phoneme m/ምʼ inserted before the labial(ጊንቲኩ) consonants sound to Convery terminative (ጊዞኺስት). The phoneme ŋ/ኽʼ only appears preceding pala-to-velar or uvular (ትናጊኩ) consonants to Convery infinitive markers in the Awngi language.

Rule 2: [..]→ንʼ || _ Al "+";

[..]→ምʼ || _ La "+";

[..]→ኽʼ || _ Vl "+";

In each of the rules, the variables Al, La, and VI are assigned for alveolar, labial, and palate-velar or uvular consonants.

4.12 Evaluation of morphological system

The system is implemented using foma finite state tools and morphotactic with their expected rules defined in the lexicon file. The language-specific rules are composed with the lexicon file in the foma built function. The morphological system efficiency executed by finite state transducer is measured in different word forms collected from different domains of the data sources stated above in this chapter. To evaluate the efficiency of the system the researcher compares system-generated words with the word collected in different domains. The useful data source for the evaluating word formation of the Awngi morphological system is a word that existed in the collected document. Analysis of expected efficiency is measured based on the predefined tag set

and the legal combination of the root word with the grammatical structure of the given word form that they exhibit. The resultant accuracy of this experiment presented in the term of percentage is calculated as the difference between the total generated words (Tgw1) with the words that were generated wrongly (Tgw2).

CHAPTER FIVE: EXPERIMENTATION AND RESULT

5.1 Experimentation environment

The presentation and discussion of the experimental results are based on the finite state machine. It is specifically based on the foma finite-state toolkit. Foma is a lexicon programming tool adapted for converting regular expressions to finite automata and transducers. It is a very efficient and feature-rich implementation software package that enables the development of various natural language applications. Based on this reason the foma lexicon programing language is selected for implementing the morphological system of the Awngi language. That leads to the of using lexc language to handle large amounts of lexicon with certain simplicity rather than a regular expression of the language. The first phase of experimentation was obtained by running a lexicon and morphotactic on the inbuilt functions of the foma system and the intermediate lexicon transducer was formed which is composed of the alternation rule to generate the words. As described in chapter four linguistic information was collected from the textbook, from Awngi newspapers, and the supplementary grammatical resources of the language. The experiment result was evaluated against the word that occur in the collected documents from a different domain.

5.2 Nasal phoneme m/ም, n/ን/, ŋ/ኽ and place of articulation

As stated in chapter four above these phonemes are inserted in different positions to complete the language-specific rule with the preceding different consonants in the Awngi language. They give grammatical information for words generated. The phoneme n/ን/ inserted before alveolar(ድድኩ) consonants to achieve second and third plural, the phoneme m/ም inserted before labial(ጊንቲኩ) consonants to Convery terminative (ጊዞኺስት) and the phoneme ŋ/ኽ inserted preceding pala-to-velar or uvular consonants (ትናጊኩ) consonants to Convery infinitive markers. Based on this the experimental result of those nasal assimilation phonemes is shown below.

Table 5 1:Nasal phoneme m/ም, n/ን/, ŋ/ኝ inserted to achieve grammatical information of the language

	Total root word	Words generated correctly	Words generated wrongly	Percentage of success	Percentage of failure
Root words take ን before ት	280	1605	75	95.54%	4.46%
Root words take ን before ፅ	188	1082	46	95.9%	4.1%
Root words take ም before ብ	94	504	60	89.4%	10.6%
Root words take ኝ before ቕ	72	397	35	91.9%	8.1%

The information given above shows that the nasal sound participated in the word formation process. The root words take the phoneme n/ን/ that precedes the consonant sound t/ት to generate 1680 words out of these 75 (4.46%) words that were wrongly generated. The 188 root words that take the nasal phoneme n/ን/ before the ʦ/ፅ consonant sound to generate 1128 words out of those 1082 (95.9%) words are correctly generated. The nasal assimilation sound m/ም takes before the last consonant sound b/ብ to generate 564 words and of the words 89.4% results percent success. This nasal phoneme ŋ/ኝ participated in 45 root words before the Q/ቕ to produce 270 words. Out of this generated word, 255 words are categorized correctly generated. General nasal consonants

for the word formation process produce 3804 words Out of these words 216 words are generated wrongly.

Assimilation is the behaviors of sound influenced by neighborhood sounds. In all cases, the error or wrongly generated words were encountered because of the factor that originates from epenthetic sounds in this language. This is the functionality regarding with epenthetic central vowel ɨ /እ in this language while becoming a word-initial it is the evidence for the same phonetic sequence. The epenthetic sound breaks up unwanted consonant clusters in the word middle rather than the at word-initial and finale. This means that in word-medial there should be no cluster of more than two consonants. In the word formation process, the nasal phonemes m/ም, n/ን, and ŋ/ኽ are inserted with other consonant sounds forming more the two-consonant cluster. For example, in the nasal phoneme n/ን/, the root word takes ን before ት like this /ክንትንኩ/ meaning /they learn/ during this time the compiler considered the inserting the epenthetic vowel እ/ɨ/ to break up unwanted four /ክንትን/consonant clusters /ክንእትንኩ/ to bear the grammatical information of words for the Awngi language. In this case, the combination of nasal consonants is never separated by ɨ /እ. This generates phonetically wrong words it is the same reason for all nasal phonemes which take other consonants next to them. For the other three nasal phonemes እ/ɨ/ insertion between is given as /ግንምእብ/, /ሉምንእፅኽ/ and /ዝኽእቕንኽ/ for root word take ም before ብ, root word takes ን before ፅ, root word takes ኽ before ቕ respectively.

5.3 Imperative verb result

Commanding information depends on the tone of the speakers that provide further information. The imperative form of the verb can be categorized as either it is an imperative form or the present tense of the second person singular in the Awngi language. Forms of imperatives in the Awngi language exhibit affirmative(positive) and negative categories. It has suffixes with empty morphemes that states that the form of root words can be imperative based on the nature of the

language. The negative marker in the imperative verb is different from the other forms of verb negative forms used in the second person singular and plural form. For this reason, the negative imperative form suffixes for both second-person singular and plural requires a certain condition to be attached to the free morpheme. The imperative form of the word is described within its negative form in chapter three of this report. Over all experimental result of the imperative verb form is shown below.

Table 5 2:Imperative marker suffixes for word generation

Parson class	Suffix attached to the root word	Numbers Words	Correctly generated words	Wrongly generated words
Second Peron singular	End with -ꣅ	2350	2006	344
	End with -꣄	ꣿ3442	12092	1350
Second-person singular negative	-ꣃ add the end of the fifth letter of Awngi	7708	6592	1116
Second person plural	-ꣃ add end fourth letter Awngi	6674	6219	455
Second-person plural negative	End with -꣆	2068	1972	96
	End with -꣆	1974	1927	47

The imperative form for second-person singular negative and second-person plural in the above experimental takes the same suffixes but the suffixes attachment depends on the nature of the last letter of the root words. When the last root words end with a fifth consonant letter in the language followed by suffixes n/ꣃ for the second person negative and n/ꣃ for the second person plural with the root word end for fourth consonant letters in the language. The factor that emanates error is the glottal sound /h/ on the phonetic level that comes in unpredictable positions within a given

word. It affects the implementation because phonological the exact glottal sound position determination is a challenging task.

5.4 Word generated by geminated the consonants

*Table 5 3:*Word generation by doubling the consonant

Geminated consonant	Total words generated	Percentage out of total Generated words	Correctly geminated ford form	Wrongly geminated word form
ʦ->ʦʦ	15440	20.06%	14315 (92.7%)	1125 (7.3%)
ᶇ->ᶇᶇ	1063	1.4%	925(87.01%)	138(12.98)
ᶇ->ᶇᶇ	1757	2.3%	1729 (98.4%)	28(1.6%)

In this experiment, 23.76 percent of the total generated words were formed by doubling the consonants in the language. From the total words 20.06% of words were generated by ʦ/ʦ->ʦʦ consonant doubling up, 1.4% of words generated by ŋ/ᶇ->ᶇᶇ consonant doubling, and 2.3% of words were produced by n/ᶇ->ᶇᶇ, consonant doubling respectively. 18260 words were generated by consonant doubling out of this word 15440 or more than five times of other two nasal phoneme geminate consonants were generated by ʦ/ʦ->ʦʦ. Out of this word, 92.7% or 14315 words are categorized by the researcher as correctly geminated word forms. In addition to this 1125 or 7.3% of words were categorized as wrongly generated word forms. Because of the ʦ/ʦ->ʦʦ consonant in the Awngi language use in different inflectional morpheme and derivational as defined in the morphotactic part of the rule formation. The reason that encountered the wrong word formation in this experiment is that in the modern Awngi language, there is a glottal approximant [h], a

sound that is a loan from Amharic. This modernly used phoneme has no phonetic and orthographical known pattern for the Awngi word formation process. So that unpredictable pattern phoneme /h/ results in the wrong word forms in this language. Depending on the applicability of the phonemes in different inflectional suffixes like imperative marking, causative marking in both single and double, with its geminated properties to some extent, contradict. In the same way, for two nasal phonemes, 87.01% of words were generated correctly by doubling the ŋ/ሻ, and 98.4% of the correctly geminated words were by nasal phoneme n/ን. The above experiment used consonant gemination to give the basic grammatical information like second person double causative marker. As described in table 5.1 above the wrongly generated words in the geminated word formation of nasal phonemes are encountered from the insertion of the epenthetic vowel ኦ/i/ to break up the unwanted grammatical structure of the language.

5.5 Experimental result of jussive verbs

Unlike the imperative verb jussive verbs generation works with all paranasal classes. It can be generated in positive or negative verb form. The negative form of the jussive verb is using the normal negative markers next to it. The affirmative form of the jussive word generation is going to experiment with as follows.

Table 5 4:Jussive word generation

Person	Suffix	Correctly generated word form	Wrongly generated word form
1st singular and 3rd singular .M	ስ	1640	No
2nd singular and 3rd singular. F	ትስ	1640	No
1st plural	ንስ	820	No

2nd plural and 3rd plural	ኒስ	1640	No

The attachment of suffixes for jussive verb formation is all truthful for each person class because there is no vague rule formation for the jussive verb of the Awngi word generation. The first plural markers of the jussive verb marker are unique other three are the same within their computable form.

Here is the sample graphical view of the word formation process in the Awngi language using the finite state transducer (FST).

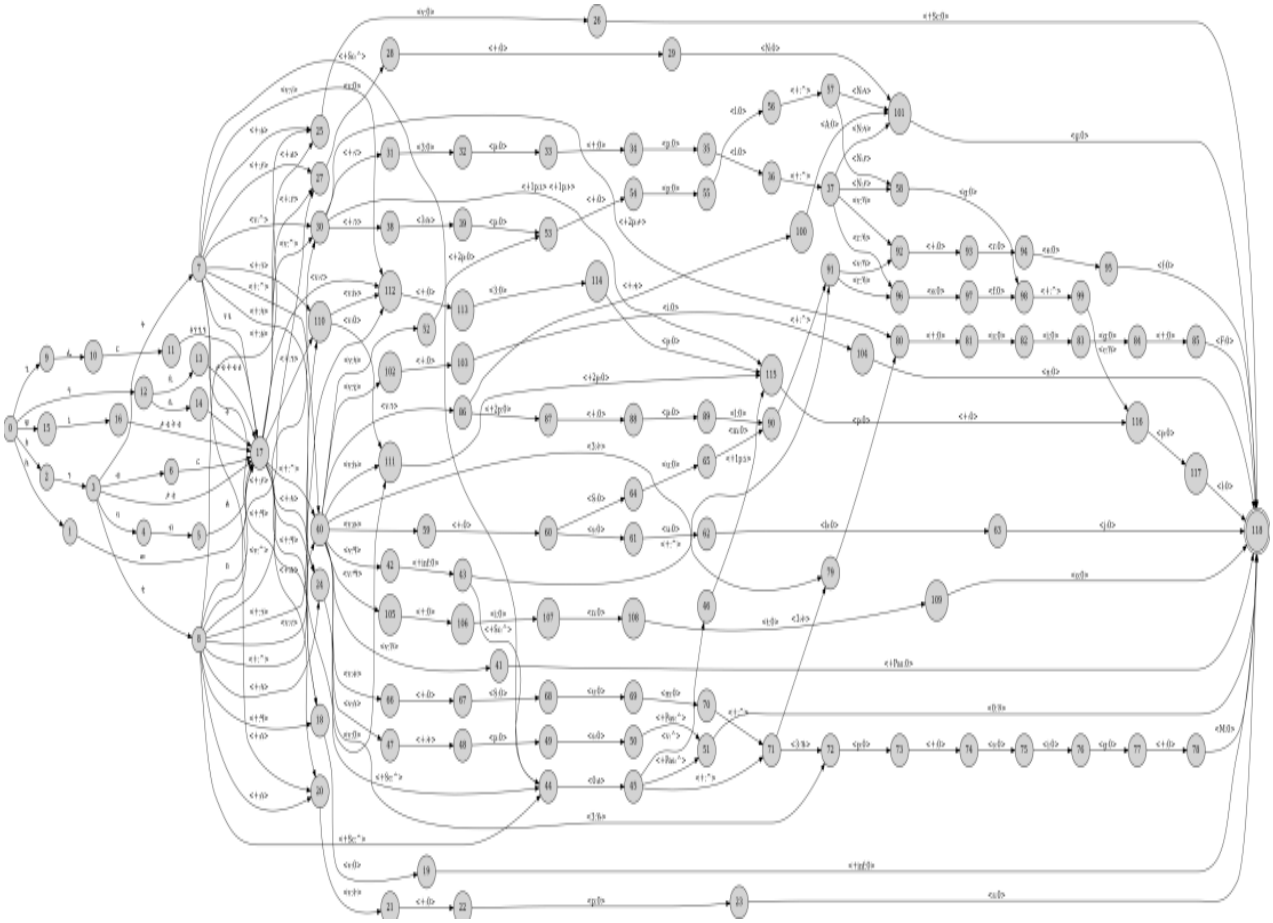

Figure 5 1 Finite stat transducer for Awngi word formation

5.6 Morphological analyzer graphical view

The following image describes the Awngi morphological analyzer output screen on the foma lexicon programming.

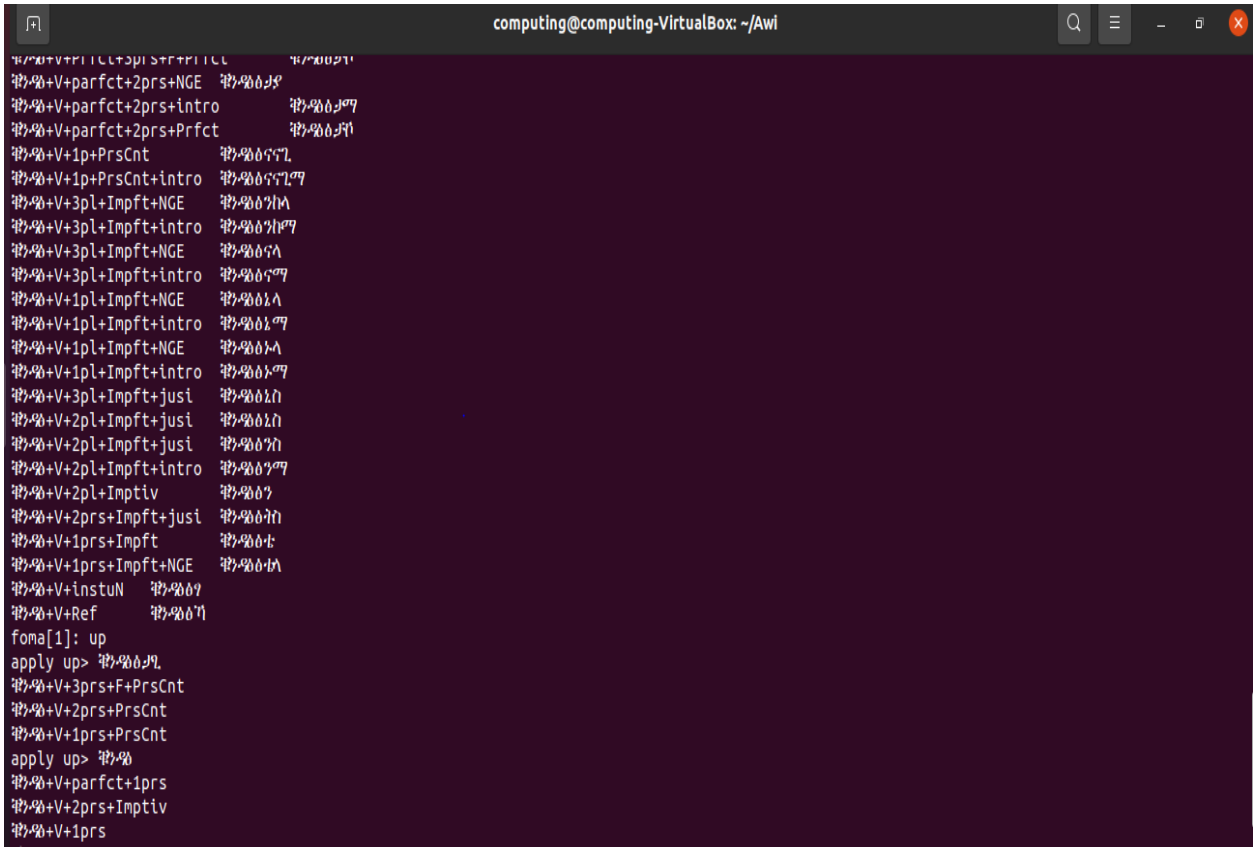

Figure 5 2:Experimentation result of the morphological analyzer for the Awngi language

The output of the system in the figure above shows that for the given two-word *ቹንዳዕዕታኚ* (I am helping the patient) and *ቹንዳዕ* (help patient) each has three grammatical meaning assigned for them. The first analysis of words tells what the activate is doing at the moment for first, second, and, third feminism with the same word form. Because the bounded morpheme (*-ታኚ*) which cares about the contentions meaning of for all person is the same. Second system analysis shows the

word is in the imperative form of the second person singular that passes commending information and at the same time it is the word in the perfective form.

5.7 Morphological generation of Awngi word

All form of the suffix discussed in the above chapter is included for the morphological generation of the words. Awngi language indeed is very predictive in its morphology because, from a single root word, there is a maximum of ninety-two words is generated. Here is the experiment resultant is the morphological system output screen from the single root word described by using foma inbuilt functions.

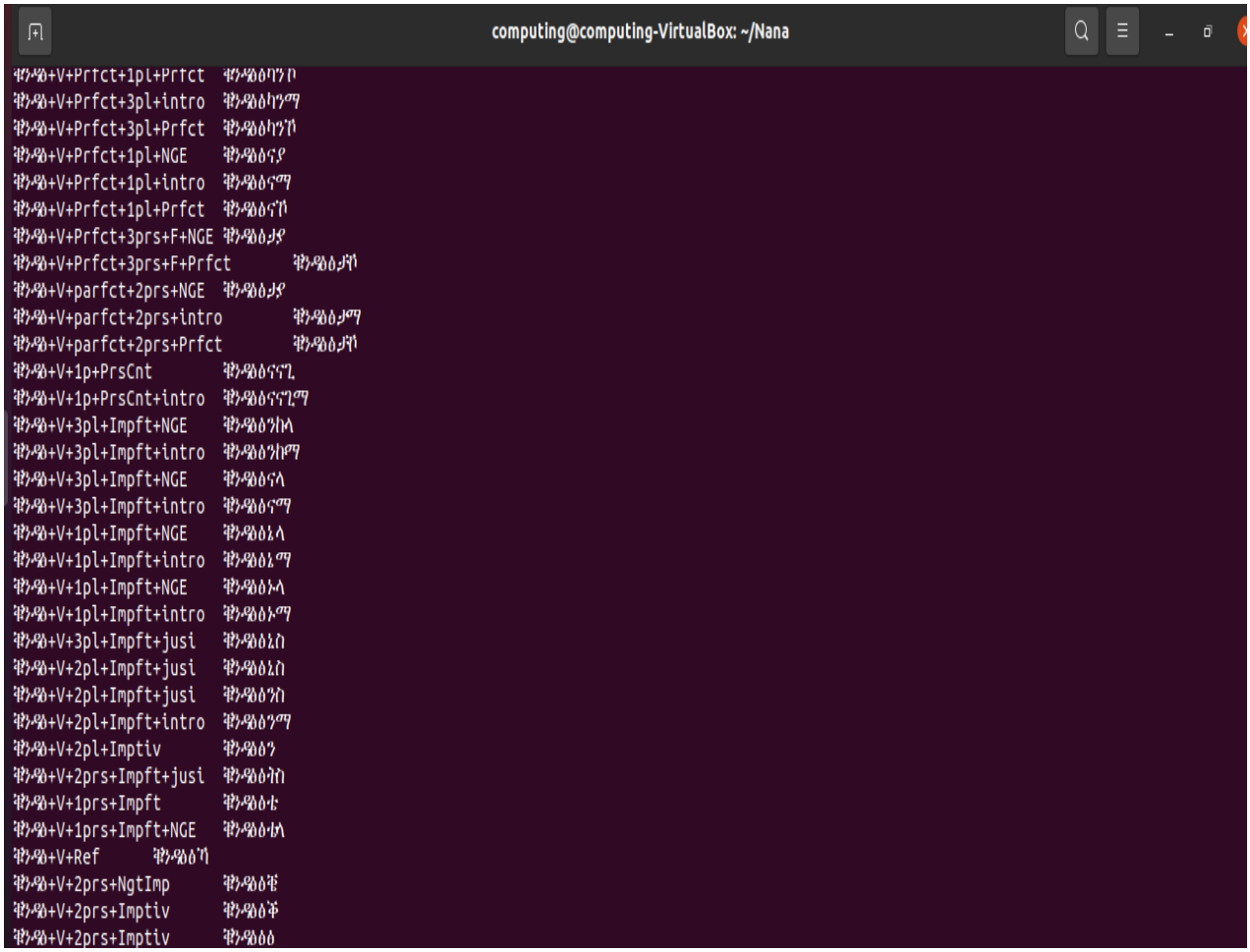

Figure 5 3: word generated from the single root word

5.8 Overall evaluation

Out of seven hundred twenty-six (726) root words seventy-six thousand nine hundred fifty-six (76956) words were generated by the system. The correctness of the experiment has been conducted by researchers and language experts. In the first phase, the generated word is tasted manually by the researchers comparing the structure of generated words with collected word forms from the different domains as stated in the methodology section. To evaluate the morphological system the researcher uses clustering-based techniques based on a regular pattern of endings in this language and the alternation rule components are used as evaluation parameters. At this time researcher faces challenges with the words generated by the system but they do not have them in the collected document. So, this whole system-generated word which are not in the collected also evaluated by the language expertise in addition to the researcher. In this phase, the expertise evaluates each word meaning that the newly system-generated words with which have words in a document word. The performance of the morphological analysis is based on the form of words generated correctly by the system. As evaluated in the two phases argues that the results of the analysis are out of the total generated words (4915) of words encountered out of the word form of this incorrect word formation because of reasons stated in the above experimental.

CHAPTER SIX: CONCLUSION AND FUTURE WORK

6.1 Conclusion

The purpose of the research was to design a morphological analyzer and generation for the Awngi verb form. Language technology plays a great role in developing software for analyzing and generating word forms. The beginning with working with morphological analysis and generation is a vital role in developing different natural language applications. Based on the literature reviewed the Awngi is one of the under-resourced central Cushitic language family has never been any efforts made for morphological studies. The absence of morphological studies in the Awngi will limit the researcher's conduct of high-level NLP research. However, the development of a morphological system for the Awngi language is a very crucial point for further use of the language in many levels of natural language application development. After a deep understanding of the morphological nature of the Awngi, language researchers decide to develop a morphological analyzer and generation system for the Awngi language. To achieve the decided goal the researcher conducted by using a rule-based approach that was implemented by a finite state transducer. During develop the morphological system same challenges were faced the insertion of the epenthetic vowel ɨ/i/ to brake ungrammatical triple /CCC/ consonant cluster that affects nasal consonant insertion and the geminated consonants, word formation, and the loan sound /*h*/ which was not stated clearly in the linguistic structure of Awngi in terms of orthographic or phonological pattern affects the word formation process.

Constrictions of the transducer have two major components. The first component is Morphotactic which is responsible for governing the grammatical rule that combines morphemes with the meaning-bearing unit. This lexical design rule of the language is reshaped or modified by language-specific rules or alternation rules. The second is a language-specific rule of the system

that was designed based on a finite state transducer. In the era of the computational environment, the Awngi language is resource-poor because it has no predefined morphotactic prearrangement of words that are used for morphological analysis. Alternation rules and morphotactic arrangement of words were constructed from scratch to meet the objectives of the study by the researcher. This task takes much time to understand the linguistic nature of the Awngi language. The words in the lexicon script is operates under many continuous classes that make the backbone of the morphological analysis. In the morphotactic part of rule formation, the researcher defines twenty-eight unique continuous classes each of the classes contains a minimum of one too to five inflectional suffixes that participated in different word formations.

To convey the different grammatical information of the Awngi language like plurality, terminative, infinitive marking, imperative marking, causative, and other different functionalists the phonological and orthographic rules were developed to modify the morphotactic parts. To achieve this orthographic and phonological process the different consonants are inserted removed, replaced, and geminated in different positions. Assimilation is the behaviors of sound influenced by neighborhood sounds. The three nasal assimilation sounds are inserted before the different consonant sounds n/ን/ inserted before Alveolar(ድድኩ), m/ም inserted before Labial(ኒንቲኩ) and ŋ/ኻ inserted preceding pala-to-velar or uvular consonants (ትናኒኩ) to generate the words. Secondly, geminated consonants are commonly used word formations process in the Awngi language. There were 726 root words from which 76956 words were constructed using those continuous classes of the morphotactic rule and 35 distinct rules defined. The researcher evaluates this formation of words in different experimental evaluation techniques conducted to achieve the promising performance of the morphological system. The nasal assimilation sound participated in 3804 (4.9%) of words from the total word formation process out of these 216 words was generated incorrectly and 3588 words were generated correctly. Geminated word formation participated in 18260 (23.7%) out of the total word generated and out of these words generated by geminated

consonants 16969 words were generated correctly and 1291 were generated wrongly. Out of the total system-generated words, 4915 words were generated wrongly. The evaluation result shows that 93.6% of words were generated correctly and 6.4% of words were wrongly generated. The factor that emanates the error is that the epenthetic vowel ɦ/ɨ/ insertion breaks nasal and geminated word formation and glottal sound [h], which is not part of the Awngi phonological system.

6.2 Future work

The development of morphological generation and analyzer for the Awngi language is at its initial point. The work that may be needed to improve the morphological system performance is put in the constative studies. Hence the future works that are incomplete or not included in the studies are the following.

Due to the language experts and time constraints, the study was conducted only on the morphological analysis and generation of the Awngi verb forms. So, in the future, another part of the speech category of language requires conducting research and implantation like the Awngi verb form.

Extending this work one can operate further investigation on the Awngi epenthetic vowel ɦ/ɨ/ formulating alternative rules that do not separate the sequence of gemination and nasal assimilation word formation.

The researcher never considered the glottal approximant /h/, which appears irregularly in the Awngi language word formation process. So, discuss with an expert more about the glottal approximant /h/ and conduct experiments to improve the performance of the system.

The experimentation does not include compounding words in the study, the researcher who wants to work with this area can improve word formation by incorporating this future of word formation process.

References

[1] A. WOLDEMARIAM, “Development of Morphological Analyzer for Sidaamu Afoo ADDIS ABABA UNIVERSITY MSc Thesis Report,” 2005.

[2] D. L. Lalego, “Morphological Analyzer for Sidaamu Afoo Desta Legese Lalego. Addis Ababa university MSc Thesis Report,” June, 2020.

[3] A.Diro, “AUTOMATIC MORPHOLOGICAL SYNTHESIZER FOR AFAAN OROMOO.Addis Ababa university MSc Thesis Report,” June, 2010.

[4] Y. Abate and Y. Assabie, “Morphological Analysis of Ge’ez Verbs Using Memory Based Learning,” MSc Thesis Report 2014.

[5] M. Gasser, “HornMorpho : a system for morphological processing of,” *Conf. Hum. Lang. Technol. Dev. Alexandria, Egypt*, pp. 2–5, 2011.

[6] H. Mekonnen, “Amharic-Awngi Machine Translation An Experiment Using Statistical Approach,” *Int. J. Comput. Sci. Eng.*, vol. 7, no. 8, pp. 6–10, 2019, doi: 10.26438/ijcse/v7i8.610.

[7] M. Mihret and M. Atinaf, “Sentiment Analysis Model for Opinionated Awngi Text,” *IEEE AFRICON Conf.*, vol. 2019-Septe, 2019, doi: 10.1109/AFRICON46755.2019.9134016.

[8] W. Barud and W. B. Demilie, “Parts of Speech Tagger for Awngi Language Parts of Speech Tagger for Awngi Language,” no. July, 2010.

[9] T. Misikir, “Developing a Stemming Algorithm for Awngi Text: A Longest match approach,” MSc Thesis Report june 2013.

[10] T. DESTA, “DEVELOPING AWNGI-AMHARIC CROSS-LANGUAGE INFORMATION RETRIEVAL (CLIR): A DICTIONARY BASED QUERY TRANSLATION APPROACH:,” MSc Thesis Report. January, 2015.

[11] J. Micher, “Bootstrapping a Neural Morphological Generator from Morphological Analyzer Output for Inuktitut,” *Proc. Work. Comput. Methods Endanger. Lang.*, vol. 2, no. 1, pp. 37–41, 2019, doi: 10.33011/computel.v2i.455.

[12] K. L. WUDINEH, “DESIGN AND DEVELOPMENT OF AUTOMATIC MORPHOLOGICAL SYNTHESIZER FOR AMHARIC PERFECTIVE VERB FORMS:ADDIS ABABA UNIVERSITY MSc Thesis Report,” no. June, 2012.

[13] D. Kumar, M. Singh, and S. Shukla, “FST Based Morphological Analyzer for Hindi Language,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 4, pp. 349–353, 2012.

[14] T. A. Gebreselassie, J. N. Washington, M. Gasser, and B. Yimam, *A finite-state morphological analyzer for wolaytta*, vol. 244. Springer International Publishing, 2018. doi: 10.1007/978-3-319-95153-9_2.

[15] D. B. Weldegiorgis, "Design and Implementation of Automatic Morphological Analyzer for Ge ' ez Verbs," *Master's Thesis*, pp. 1–144, 2010, [Online]. Available: aau.etd.edu.et

[16] A. van den and W. D. Bosch, *Memory-Based Morphological Analysis*. 1997.

[17] M. D. Mosa, "Morphological Analyzer for Afaan Oromoo Using Machine Learning A Addis Ababa, Ethiopia MSc Thesis Report," 2020.

[18] B. W. GETAHUN, "PART OF SPEECH TAGGING FOR AWNGI LANGUAGE:BAHIR DAR, ETHIOPIA MSc. Thesis Report," 2020.

[19] O. V. Nagel, "Morphological and Evaluation Effect within Russian Syncretic Derivational Model," *Procedia - Soc. Behav. Sci.*, vol. 200, no. April, pp. 318–323, 2015, doi: 10.1016/j.sbspro.2015.08.072.

[20] T. B. Bati, "a Ad Dd Di Is S a Ab Ba Ab Ba a U Un Ni Iv Ve Er Rs Si It Ty Y S Sc Ch Ho Oo Ol L O of F G Gr Ra Ad Du Ua At Te E S St Tu Ud Di Ie Es S S Sc Ch Ho Oo Ol L O of F I in Nf Fo or Rm Ma At Ti Io on N S St Tu Ud Di Ie Es S F Fo or R a Af Fr Ri Ic Ca a Automati," 2002.

[21] B. Asaye Agajie, "Syntactic Structure of Awgni Sentences Based on Form," *Open J. Stud. Linguist.*, vol. 3, no. 1, pp. 13–28, 2020, doi: 10.32591/coas.ojsl.0301.020013a.

[22] D. L. APPLEYARD, "the Morphology of Main and Subordinate Verb Forms in Ethiopian Semitic and," 1976.

[23] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural Language Processing : State of The Art , Current Trends and Challenges Department of Computer Science and Engineering Accendere Knowledge Management Services Pvt . Ltd ., India Abstract," *Sentim. Anal. has become one Most profound Res. areas with increasing growth Soc. media web. Nowadays, millions users Exch. their views, ideas, expressions, Feel. Opin. Soc. media like twitter Facebook. Se*, no. Figure 1, 2017.

[24] A. R. Martinez, "Natural language processing," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 3, pp. 352–357, 2010, doi: 10.1002/wics.76.

[25] M. P. Nelsen *et al.*, *An Introduction to Language Processing with Perl and Prolog An*, vol. 119, no. 4. 2021.

[26] K. Knight and J. May, "Language Processing Ý Ŏü ßï Ý Ŏü ßï," *Inf. Sci. (Ny).*, vol. 349, no. 6245, pp. 3–4, 2015.

[27] Fabiana Meijon Fadul, *Understanding Morphology*. 2019.

[28] D. Jurafsky and J. Martin, “Speech and Language Processing,” *Speech Lang. Process.*, vol. 3, pp. 441–458, 2014.

[29] L. Janda, “Inflectional Morphology,” *Oxford Handb. Cogn. Linguist.*, no. July, 2012, doi: 10.1093/oxfordhb/9780199738632.013.0024.

[30] I. N. G. O. P. L. Ag, “W O R D - F O R M At I O N in English,” *System*, 2003, [Online]. Available: http://books.google.com/books?hl=en&lr=&id=pkqWjrp5oM0C&oi=fnd&pg=PR11&dq=WORD+-+FORMATION+IN+ENGLISH&ots=we3hp68nSR&sig=R23blkt64DAl36brHnAK48DU-NY

[31] A. R. Akmajian, Adrian & Demers and M. R. & Farmer, Ann. K & Harnish, *An Introduction to Language and Communication*. 2001.

[32] A. T. Submitted *et al.*, “A d d i s a b a b a u n i v e r s i t y s c h o o l o f g r a d u a t e s t u d i e s s c h o o l o f i n f o r m a t i o n s t u d i e s f o r a f r i c a,” 2002.

[33] Fabiana Meijon Fadul, *Arabic Computational Morphology Knowledge-based and Empirical Methodse*. 2019.

[34] M. Seiss, “A Rule-based Morphological Analyzer for Murrinh-Patha,” no. 1996, pp. 751–758, 2008.

[35] G. Tolegen, A. Toleu, and R. Mussabayev, “A Finite State Transducer Based Morphological Analyzer for of Maithili Language,” pp. 01–06, 2022, doi: 10.1109/ubmk55850.2022.9919445.

[36] G. Sunilkumar, “Advancing Morphological Analyser for Kannada Using NLP and Machine Learning Approaches,” vol. 29, no. 6, pp. 9135–9148, 2020.

[37] K. Megerdoomian, “Finite-state morphological analysis of Persian,” p. 35, 2004, doi: 10.3115/1621804.1621814.

[38] S. Yona and S. Wintner, “A finite-state morphological grammar of Hebrew,” *Nat. Lang. Eng.*, vol. 14, no. 2, pp. 173–190, 2008, doi: 10.1017/S1351324906004384.

[39] A. Kayabaş, H. Schmid, A. E. Topcu, and Ö. Kiliç, “TRMOr: A finite-state-based morphological analyzer for Turkish,” *Turkish J. Electr. Eng. Comput. Sci.*, vol. 27, no. 5, pp. 3837–3851, 2019, doi: 10.3906/elk-1902-125.

[40] I. Muhammad and Z. Yan, “Supervised Machine Learning Approaches: a Survey,” *ICTACT J. Soft Comput.*, vol. 05, no. 03, pp. 946–952, 2015, doi:

10.21917/ijsc.2015.0133.

[41] S. Ray, “A Quick Review of Machine Learning Algorithms,” *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 35–39, 2019, doi: 10.1109/COMITCon.2019.8862451.

[42] T. Martín Noguerol, F. Paulano-Godino, M. T. Martín-Valdivia, C. O. Menias, and A. Luna, “Strengths, Weaknesses, Opportunities, and Threats Analysis of Artificial Intelligence and Machine Learning Applications in Radiology,” *J. Am. Coll. Radiol.*, vol. 16, no. 9, pp. 1239–1247, 2019, doi: 10.1016/j.jacr.2019.05.047.

[43] D. Chinedu Asogwa, M. Okechukwu Onyesolu, and C. Ijeoma Chukwuneke, “Review of Some Machine Learning Techniques for Big Data, Their Uses, Strengths and Weaknesses,” *Int. J. Trend Res. Dev.*, vol. 6, no. 5, pp. 2394–9333, 2021, [Online]. Available: www.ijtrd.com

[44] K. Koskenniemi, “Two-level Morphology:A general computational model for word-form recognition and production,” *10th Int. Conf. Comput. Linguist. COLING 1984 22nd Annu. Meet. Assoc. Comput. Linguist. ACL 1984*, no. January 1983, pp. 178–181, 1984, doi: 10.3115/980491.980529.

[45] L. Karttunen and K. R. Beesley, “A Short History of Two-Level Morphology,” pp. 1–15, 2012.

[46] S. Wintner, “Strengths and weaknesses of finite-state technology: A case study in morphological grammar development,” *Nat. Lang. Eng.*, vol. 14, no. 4, pp. 457–469, 2008, doi: 10.1017/S1351324907004676.

[47] DhruvinSoni, “Introduction of Finite Automata - GeeksforGeeks.” 2021. [Online]. Available: https://www.geeksforgeeks.org/introduction-of-finite-automata/?ref=lbp

[48] L. Karttunen, “Applications of Finite-State Transducers in Natural-Language Processing.” [Online]. Available: https://web.stanford.edu/~laurik/publications/ciaa-2000/fst-in-nlp/fst-in-nlp.html

[49] K. Lauri, “Applications-of-finite-state Transducers in Natural language processing”, [Online]. Available: https://www.ptonline.com/articles/how-to-get-better-mfi-results

[50] A. Joswig, “The phonology of Awngi,” *SIL Electron. Work. Pap.*, vol. 2010–003, no. September, p. 37, 2010.

[51] T. M. B.A., “THE NOUN PHRASE IN AWNGI,” *Weatherwise*, vol. 37, no. 4, pp. 217–221, 1984, doi: 10.1080/00431672.1984.9933258.

[52] B. A. Joswig, “The Status of the High Central Vowel in Awngi:Paper presented at the 15th International Conference of Ethiopian Studies in Hamburg,” no. 1959, pp. 1–8,

2003.

[53] C. D. Hamm, “Sociolinguistic survey of the Kabalay (Language of Chad),” p. 31, 2002, [Online]. Available: e:%5CArticle%5CHamm(2002-Sociolinguistic survey of Kabalay).pdf

[54] A. Joswig, “Rethinking Awngi Tone,” vol. 4, pp. 1417–1425, 2009, [Online]. Available: http://portal.svt.ntnu.no/sites/ices16/Proceedings/Volume 4/Andreas Joswig - Rethinking Awngi Tone.pdf

[55] D. L. th. M. of M. and S. V. F. in E. S. and APPLEYARD, “the Morphology of Main and Subordinate Verb Forms in Ethiopian Semitic and,” 1976.

[56] A. Studies and A. Studies, “The Nominal System of Awngi (Southern Agaw) Author (s): Robert Hetzron Source : Bulletin of the School of Oriental and African Studies , University of London , Published by : Cambridge University Press on behalf of School of Oriental and African Studi,” vol. 41, no. 1, pp. 121–141, 1978.

[57] S. Anonby, “The phonology of Awngi,” *SIL Electron. Work. Pap.*, vol. 2012–006, pp. 1–20, 2010, [Online]. Available: http://www.sil.org/silewp/2012/silewp2012-006.pdf

[58] R. by: F. R. Palmer, “Review Reviewed Work (s): The Verbal System of Southern Agaw by Robert Hetzron Review by : F . R . Palmer Published by : Linguistic Society of America Stable URL : https://www.jstor.org/stable/412424,” vol. 46, no. 1, pp. 205–208, 1970.

[59] SAMUEL ASFERA W, “WORD FORMATION IN A WNGI Addis Ababa University,” *Cah. d’études africaines*, vol. 46, no. 182, pp. 291–312, 2006.

[60] M. Hulden, “Foma : a finite-state compiler and library,” no. April, pp. 29–32, 2009.

APPENDIX I: ALTERNATION RULE

Vowels/አንቤብፃንትክካ/

define V [ኢ | ኣ | ኤ | እ | ኦ];

#definition of Awngi consonants/አንቤብስታንትካ/

define C [ብ | ክ | ኽ | ቭ | ስ | ሽ | ል | ግ | ኻ | ን | ት | ች | ር | ፍ | ፅ | ው | ቕ | ም | ይ | ድ | ጅ | ዝ | ፕ];

#defining some substrings that determine the nature of suffixes attached to the root word in Awngi language

define schar [ቢ | ኪ | ኺ | ቪ | ሲ | ሺ |ኺ | ቺ | ሪ | ፊ | ሚ | ዪ | ዲ | ጂ | ዚ | ፒ];

#definition

define fro [ባ | ካ | ኻ | ቫ | ሳ | ሻ | ጋ | ጃ | ቻ | ራ | ፋ | ዳ | ጃ | ዛ | ፓ];

#define Alveolar(ድድኩ)

define Al [ት | ድ | ፅ | ዝ | ስ | ን | ል | ር];

#define Labial(ከንፈርኩ)

define La [ፕ | ብ | ም | ው];

define fiv [ቤ | ኬ | ኼ | ቬ | ሴ | ሼ | ጌ | ኼ | ቼ | ሬ | ፌ | ዬ | ጄ | ዜ | ፔ];

#define velar (ትናኒኩ)

define Vl [ክ | ግ | ች | ጅ | ሽ | ኻ | ይ | ቕ | ኽ];

define እግ [ፕ | ብ | ት | ድ | ክ | ግ | ቕ | ች | ጅ | ፅ];

define ሳኒ [ም | ን | ል | ኻ | ፅ];

#To read the lexicon file

read lexc Lexicon.lexc

define Lexicon;

#Rule1: #To Prohibition the place of occurrence of some Vowels to Convery their position in the Awngi language

define ኣdilation ኣ->0 ||_ "^" [..] ;

#Rule2: #insertion of epenthetic vowel ኣ to break up unwanted consonant clusters in the language

define ኣinsertionbtwn [..]->ኣ || CC _ C “+”;

#Rule3:# vowel ኣ insertion before ር, ይ, ው, ፒ, ኹ, ኽ, to full fill the condition of the language

define ኣinsertionbefore [..]->ኣ || _ "^"ር | ይ | ው | ፒ | ኹ | ኽ “+”;

#Rule4: #Nasal Assimilation of phoneme insertion before Alveolar(ድድኩ) consonants to achieve second and third plural.

define ንinsertion [..]->ን || _ Al “+”;

#Rule5: #Nasal Assimilation phoneme insertion before Labial(ኒንቲኩ) consonants to Convery terminative (ኒዞኺስት) .

define ምinsertion [..]->ም || _ La “+”;

#Rule6: #Nasal Assimilation phoneme insertion before pala-to-velar or uvular (ትናኒኩ) consonants to Convery infinitive markers.

define ኻinsertion [..]->ኻ || _ Vl “+”;

Rule7: Add at the end of the word ends with the second letter of Awngi to convey the single causative.

define ፅsinglecaAdd [..]->ፅ|| schar _"^";

Rule8: Add at the end of the word which ends with the fourth letter of Awngi to convey the imperative of the second person plural.

define ንimperativeAdd [..]->ን|| fro_"^";

Rule9: Add when ends with the fives letter of Awngi to convey the imperative negative of the second-person plural.

define ʔimpnegativeAdd [..]->ʔ|| fiv_"^";

#Rule10: #Two consonants come together at the end of the verb insert before the consonant.

define ሳኒinsertion [..]->ሳኒ || _ እግ “+”;

#Rule11: #geminations of the last consonant when the word ends by ፅ to Convery double causative

#Rule12: #replacement of the last consonant of the root to covary the first person singular and imperatives

define ጌreplacement ጌ-> ኬ || _"^";

define ፅconsonentdoubling ፅ->ፅፅ|| _"^";

#Rule13: #geminations of the last consonant when the word ends consonant Convery reciprocal case

define ቻconsonentdoubling ቻ->ቻቻ || [ቻ | ስ ት] _"^";

#Rule14: #geminations of the last consonant when the word ends consonant Convery tense plural makers

define ንconsonentdoubling ን->ንን|| C _"^";

#Rule15: #insertion to indicating first person plural for words containing glottal sound at the end

define ዋinsertion [..]->ዋ ||ኾ_ “+”;

#Rule16: #insertion to indicating first person plural for words containing glottal sound at the end

define ትinsertion [..]->ት || [ስ | ን] _ “+”;

#Rule17: #replacement of the last consonant of the root to covary the first person singular and imperatives

define ቤreplacement ቤ->ፕ, ቧ->ፕ || _"^";

#Rule18: #To Prohibition the place of occurrence of ቻ if it appears initial position from any character remove it.

define ቻdilation ቻ->0 || .#. _ ;

Rule 19: if the word ends ጀ add ስ to it to achieve possession markers.

define ስpossessionAdd [..]->ስ|| ጀ_ "^";

Rule 20: if the word ends ን before the ኩ to it to achieve third person plural markers.

define ንinserbtw [..] -> ን || _ "^"[ኩ] ;

#Rule21: #replacement of the last consonant of the root to covary the first person singular and imperatives

define ዴreplacement ዴ-> ቱ || _"^";

#Rule22: # ስት insertion at the end of the word when the end of the word is by the consonant to achieve the passive forms of the verb.

define ስትinsertion [..]->ስት || C _"^";

Rule23: Add ው at the end of the vowel to convey the male gender marker.

define ውAddaf [..]->ን|| V _"^";

define Cleanup "^" -> 0;

define Cleanup “+” -> 0;

define Cleanuph "-" -> 0;

read lexc Lexicon. lexc

define Lexicon

define Grammar Lexicon .o.

አdilation .o.

አinsertionbtwn .o.

አinsertionbefore .o.

ንinsertion .o.

ምinsertion .o.

ቻinsertion .o.

ፅsinglecaAdd .o.

ኀimperativeAdd .o.

ኀimpnegativeAdd .o.

ሳኂinsertion .o.

ጌreplacement .o.

ፅconsonentdoubling .o.

ቻconsonentdoubling .o.

ኀconsonentdoubling .o.

ዋinsertion .o.

ትinsertion .o.

ቢreplacement .o.

ቻdilation .o.

ስpossessionAdd .o.

ኀinserbtw .o

ዴreplacement .o.

ስትinsertion .o.

ѡ·Addaf .o

Cleanup;

regex Grammar;

APPENDIX II: AWNGI WORD LIST OF VERBS

USED FOR EXPERIMENTATION

ባምፒ	ቤቱና	ካፄ	ኬው	ኮሬም	ስንክ
ባምበ	ቤፅ	ካንት	ኬዜፍ	ኬሼ	ሳሱኻ
ብቭርፄ	ቤኑኻ	ካሳ	ኬፄር	ኪዱኻ	ሴግ
ብቭርባ	ቤሴቾኾ	ካታ	ኬኬር	ኪፀኻ	ሳንክ
ብላፅ	ቤን	ካሴ	ኬሌል	ክም	ሴይ
ብሲ	ቤንኻ	ካፃ	ኬሴ	ኮይ	ስንትሳ
ብራ	ቤፃ	ካስ	ኬቱኻ	ኮፄርኻ	ሶኼን
ብኩ	ቤዬር	ኪስስ	ኬፍ	ኮቾኻ	ሱግታ
ብሬ	ቤዬል	ካፂ	ኬሜ	ኻታ	ሳኩኻ
ብስ	ቦዜን	ካፅ	ኬሌብ	ኾሩኻ	ስር
ብር	ቦዬድ	ካንታ	ኬቱ	ኾኽቱ	ሱጊ
ብኩስ	ቦጌ	ካንቱ	ክንታ	ኹፄ	ስፒ
ብሴ	ቦዜዝ	ኬሴል	ክንቱ	ኽስኾ	ሳቱኻ
ብቾካ	ኪሙኻ	ኬይ	ክንት	ኹሬ	ሳጊ
ብትን	ኪቲ	ኬታ	ክንቲ	ኾኺት	ስንት
ብኽር	ኪሌቓ	ኬሌቼ	ክቺ	ኹራ	ስንቲ
ቤና	ኪስኻ	ኬሽ	ክብ	ኽይ	ሲፍ
ቤዬራ	ኪንክ	ኬሌም	ክራይ	ኹዋ	ሱጋ
ቤሩኻ	ኪንቺ	ኬርት	ክች	ኽይፅ	ሲሺኹ
ቤፉ	ኪስ	ኬንቱባ	ክችኃ	ኹሪ	ሱግ
ቤፍ	ካንቲ	ኬሴት	ኮይፅ	ኾኽታ	ሴብ
ቤዬላ	ኪፅ	ኬኬኽ	ኮሹኻ	ስንትስ	ሳኹኻ

ስንትሴ ልውኻ እንቓሪ እንጁኻ አሬግ እንፁ

ሱግስ ልች እንፃቜ አብን እሻ አንዔሬብ

ሴኔፍ ሊሚፅ እንትፍ አቱም አቺት አሬሱኻ

ሳሴር ሊኪ እንፃቓ አሬከሻ እንዔ እጂ

ሻቂ ሊፉኻ እንጅኩ አሌቜት እሚ እይስ

ሺንኪ ልኻኽ እርዳት አንሼፓ እይታ እሺፅ

ሹሙኻ አሌ አንቺቲ እንኩዋኽ እንጋ ጌቤሩኻ

ሼው አቱ አንቤብ እንፃቖ እሩኻ ጌው

ሻኪዒ አንሺፕ አዜዝኾ እንጂዔ አሴብ ጊኼ

ሻሪ እንጅኩዋ እንቓስ አዊጅ አግፀኻ ግቢት

ሸፍኒ እንኩርሻ እንኩራ እቺስ አንቱ ጌፌፉ

ሻሩኻ እንፂ እንኩርኻ እንኩርስ አኬ ጌሬትት

ሹም እቺት እሲኹዋ እይኹ እይ ጉድም

ሹካት እሚኹ እንከራ እንጂፅ እንፃኽ ጐሩኻ

ሊሙኻ እንኩዋቓ እሚት እሜ አፄ ገሬፍ

ልሹኻ እንኩሩኻ እንኩሬ እሸንኾ አትሸቲ ጊፉርፄ

ላንፅ እሺ እንኩዋቖ አፔስ አገዘ ጋሌብሻ

ልም እንከሬ እንክሩኻ እሉኻ አቡቖ ጌኼንት

ልቺት እንኩዋቜ አንቤፔ እያ እንፃ ገሪ

ሉስ እንጂፂ አቤቱል እስሪ አኪ ግዲት

ልቡ እንከር አሲፕ እምስ አሴ ጋቺ

ልማ እንዜሻሻ እንጂፃ እላቲው አኻ ያጋ

ልሜ እንጂኩ እንክርታ አግስ አቱኻ ይጉዌ

ልቺፅ እንፃቜን እንጅኩዌ እሺኾ እሳን ጌፌርፃ

ሊንቖ አትሸቱኻ እይስታ አሬከ እቱኻ ጌም

ጋቡኻ ግልፃ ኔፌግ ቲሪት ቶኬስ ችሩኻ

ጉዱኻ ጊሊፅ ናኩኻ ታንቢ ቲቹኻ ቹኻ

ጋቹኻ ጌቴምኻ ኔፄል ታፍ ቴር ቺችፍ

ጌሌቤ ጌኔንኻ ኔኬት ቲሪ ቴንኬክ ቻኑኻ

ጌፌርፂ ግቢ ናቺ ትኩም ታስኻ ቻኹኻ

ጉባልት ጉስ ናኼሲ ቲሪፄ ቴሌቾ ፃያ

ጐቤዝ ግችል ናኪኻ ትንኪፍ ታፋ ፄረከ

ግሙኻ ጌጌው ንቹር ታቹኻ ትሸ ፃፁኻ

ጌርኼኽ ጉሸ ቲኪ ትንክፍ ትሺ ፃንኩፃ

ፄቤብ ጉሱኻ ቴኬማ ታጌ ቻቤላ ፒሬካ

ጊኺ ጌሬምፄ ታጋ ቲሪፅ ቺፊ ፃንኩት

ጌሸ ጊኝ ትልኽም ታና ቻቢል ፄረገ

ጋሬድኻ ጌሹ ቴርት ትምብ ቹፍ ፃፊ

ግቢቱ ግሸ ቲፕ ቶኬማ ቻት ፃፍ

ጌል ጋሹኻ ትስኻ ታንኬን ችሪ ፃንቹኻ

ጌሌቤታ ፃኹፅ ትፍ ታኬት ቺፊኻ ፃንኩፄ

ጐዴብ ጉዋር ቲም ቲኩኻ ቹኒቲ ፃንቹዊ

ጌሬፍ ጉልቾ ታቱኽ ቱዌ ቹኒቴ ፃርክት

ጌቤል ኔኬል ታንቡ ትክ ቹኒት ፒኪ

ጊፅ ንካ ታንባ ቱዋ ችርኻ ፃንኩፒ

ጉዋዚዝ ንኻ ቲን ትስ ችግም ፒቾ

ግቢቲ ኑና ታንብ ታምብ ቹኔክ ፄውኻ

ጐፊፅ ኔኪ ቲሪፒ ቲሪፃ ቹኒታ ፃሩኻ

ጋቢ ናያ ታንቱኻ ታንቤ ቺፅኻ ፃፉኻ

ጉዋዚ ናዱኻ ታስ ታትብ ቻቢሌ ፄው

ፃይሻ	ፃምርት	ዊፅ	ቕፂ	ማሊ	ሙሬ
ፂሪክ	ፄንፄኽ	ዊድኹ	ቑፁኻ	ማሲፅ	ማቱኻ
ፄኩኻ	ፃንቑንት	ዎሸ	ቑንዳፅ	ሙቹኻ	ማቔንት
ፂሬኬ	ፁሙኻ	ዊተ	ቑፂ	ሜዜጌብ	ሜዴብ
ጌፌርፄ	ፃሚስ	ዎኬ	ቑፃ	ምርክቲ	ያግፁኻ
ጉልቺ	ፁቕ	ዎካ	ቑፄ	ሞኬር	ይም
ግቢታ	ፂዱኻ	ዎኬል	ቕሸ	ሞጌስ	ይንታ
ጌዌስሻ	ፂኩኻ	ዎይ	ቑሲ	ምሳኽ	ያኹኻ
ጌሬዠ	ፃዴፍ	ዋሌንት	ቑፅ	ማሱኻ	ይስ
ጌፌርፅ	ፃይ	ውባርኹዋ	ቕፒ	ሜቴርሻ	ይጋ
ፄፒ	ፃፌ	ዎሬስ	ቑፉ	ማንቲ	ይፅ
ፃካ	ፃሪ	ዎቴት	ምፅ	ማኹር	ያኻላ
ጊፊሪፅ	ፁቼ	ዊደ	ሚፅ	ሜንዜር	ይው
ፃሚ	ፂቑኻ	ዎኽ	ጅሙኻ	ማኩት	ይፄ
ጌፌር	ፃንኩስ	ወየ	ሞኬርሻ	ማንዲ	ይንቱኻ
ፄፉኻ	ፃንቑስ	ዎሌጅ	ምኬን	ማሉኻ	ይታ
ፂቓ	ፂኒስ	ዎንጀል	ማላ	ማንታ	ያኽሻ
ፃንጌኽ	ፂኑኻ	ዊድ	ማሩኻ	ሜንዴብ	ያጌ
ፃኹኻ	ዋካ	ዎሌል	ሚቑኻ	ማንዱኻ	ይንቡኻ
ፂፁኻ	ዋኽ	ዋቕ	ማንቴ	ሙሪ	ይስሻ
ፁቓ	ውልንቾ	ዋኬ	ሚንዛን	ሚቲስ	ያኻስ
ፃፋ	ዎጌን	ዊዛር	ሜፄዌት	ሜሴል	ይዝኩት
ፃታ	ዋክ	ዎጉኻ	ማልፓ	ማራ	ያኹኺ
ፄጊ	ዎሴን	ወሸቱኻ	ማቼች	ምልኻት	ይና
ፄክ	ውባር	ዎኬስ	ዞሙኻ	ሞፄድ	ያካ

ይጉኻ	ዱንፀኻ	ዴው	ጇሚር	ፋይ	ፍሪ
ይጌ	ዳድ	ዴስታ	ጇፓ	ፈኬ	ዝቻ
ይካ	ድግ	ዱዋ	ጇፔ	ፈኹኻ	ዛቢ
ያኻ	ድው	ዴዌሌ	ጇፒ	ፋቲ	ዝኮ
ያቜኻ	ዴሳ	ዱንፃ	ጅን	ፍቺ	ዛሪ
ያጉኻ	ድኹኻ	ዳንኬ	ጁዌ	ፋን	ዛኹኻ
ይመኻ	ዴዴቻ	ዱዌ	ጀወ	ፍንጂ	ዛሩኻ
ይንቱና	ዴሲ	ዴፅ	ጁፅ	ፉራ	ዙር
ይፃ	ዴዌል	ዴስ	ጁዋ	ፈሬም	ዝቾ
ያኬ	ድኽ	ዴሜከ	ጂፊ	ፍኻ	ዙራ
ይክርቴ	ዳዴኽ	ዱንቱኻ	ጇሜራ	ፉቺ	ዙሜ
ይትካ	ዴሴ	ዴዌላ	ጇው	ፍንጃ	ዝቜ
ዴሌኽ	ዳሱኻ	ዳቜር	ጅል	ፊቻ	ዝክ
ዴሱኻ	ዳሉኻ	ድሚ	ጅፊ	ፍል	ዚዝሊ
ዲግ	ድፅ	ድቾ	ጁኻ	ፍትኽ	ዝኩ
ዴሌል	ዳንጫ	ዳኽት	ጃላ	ፈቴ	ዛራ
ዲዊል	ዱቡኻ	ዳሴቾ	ፈክ	ፉሪ	ዙሚ
ዴሌብ	ዱው	ዴጌን	ፈኩኻ	ፍን	ዝባት
ዲብ	ዱንፅ	ዴክ	ፈያ	ፈቴግ	ዛዙኻ
ዴሬሬብ	ድንቤር	ድኪ	ፊቜ	ፋታ	ዙማ
ዳዴኻ	ዳቡት	ድኹቻ	ፍሸ	ፊቱኻ	ፕንቺ
ዱንት	ዴፈር	ዳሩኻ	ፉሩኻ	ፈን	ፓኽ
ድርብ	ዳሊ	ዴዴክ	ፈኪ	ፍች	ፓኹኻ