



Addis Ababa University College of Natural Sciences

Example Based Amharic Text to Ethiopian Sign Language Machine Translation

Lily Abebe

A Thesis Submitted to the Department of Computer Science in

Partial Fulfillment for the Degree of Master of Science in

Computer Science

Addis Ababa, Ethiopia

June 2018

Addis Ababa University
College of Natural Sciences

Lily Abebe Wubshet

Advisor: Yaregal Assabie (PhD)

This is to certify that the thesis prepared by Lily Abebe, titled:
Example Based Amharic Text to Ethiopian Sign Language Machine Translation and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name _____ Signature _____ Date _____

Advisor: **Yaregal Assabie (PhD)** _____

Examiner: _____

Examiner: _____

Abstract

Ethiopian sign language (EthSL) is an independent language used by the Ethiopian Deaf society. Sign language is a visual gestural language, used by the Deaf for everyday communication. In most societies sign language is not known, and if known, it is not well understood even by their own family. This in turn makes their communication to be dependent on interpreters even in situations like medical treatment that need privacy.

Machine translation for spoken languages is a successful area of research and development. But in case of sign language, especially in case of EthSL it is in its infancy. This is because Sign language research is still a relatively new area when compared to research into spoken languages.

In this study an attempt is made to design and implement an example based Amharic Text to Ethiopian Sign Language Machine Translation system by collecting examples. The example data is processed to produce morphologically analyzed Amharic words, word structures and sentence structures in addition to adopting EthSL dictionary. A prototype performing translation based on knowledge acquired from processed example data is thus developed by involving word feature analysis, matching, alignment and recombination phases. The resulting translation output of EthSL sign is displayed as a sequence of video clips.

The system is assessed in terms of adequacy and fluency with three groups of evaluators for different categories of texts. The evaluation result shows that finger spellings, numbers and words from the adopted dictionary were adequately understood with a rank of 'All meaning' and with 'Good' fluency whereas words out of the adopted dictionary were adequately understood with a rank of 'Most meaning' with 'Acceptable' fluency. In case of sentences 'Much meaning' has been adequately understood with 'Acceptable' fluency level.

The usability of the prototype was ranked as 'Excellent' with a 'Very good' level of ease of use. Deaf members of the test teams have been inspired with the capability of the prototype for future use and they expressed their willingness to contribute for full deployment of the system.

Key words: Ethiopian sign language, Sign language, Example based machine translation

Dedication

This study is dedicated to

- My pre lingual Deaf son Abenezer Solomon and all Ethiopian pre lingual Deaf people.
- My daughters Mahlet Solomon and Hermela Solomon
- My mother Zewde Amare

Acknowledgements

First and foremost, I would like to thank the Almighty God and Virgin Mary for granting me the patience, the strength and most of all the people who have supported me throughout my journey.

Many people have contributed in one way or another to this study. I would like to begin by expressing my gratitude to my advisor Dr. Yaregal Assabie for his guidance and support. His willingness to not only listen to the problems I faced during the study, but for leading me to view and solve them with a new perspective.

Thank you GOD for giving me a family who cares about me more than I do for myself. Words seem so feeble to express my gratitude for my children Abenezer Solomon, Mahlet Solomon and Hermela Solomon for their unwavering support, and above all, for being the greatest and most understanding kids. My special thanks goes to my mother Zewde Amare and my sisters especially Hiyab Hagos. I couldn't have been able to make it without their support.

I wonder how much God cared about me when I think of W/ro Zewditu Tsegaye and her husband Dr. Tibebe Beshah, who was like my co-advisor from the inception to the end of this research, for their support in every direction. Thank you for being part of my life.

Thank you GOD for the friends surrounding me, when I thought of halting this research, YOU speak through them to continue my endeavor. I want to express my gratitude especially for Engidu Kifle, Amare Gebru and Yoseph Belay, for their priceless help in every stage of this study.

I want to express my gratitude to Ato Kifle Gebrehiwot from Kotebe Metropolitan University; Dr. Eyasu Hailu, Ato Andargachew Deneke and Ato Pawlos Kasu from AAU and Meaza Mengistu from Menilik preparatory school. My gratitude also goes to W/t Aseggedech Bekele a sign language professional interpreter and W/ro Tewabech Letibelu a sign language teacher who didn't hesitate to help in every direction and throw words of appreciations and suggestion.

Last but not least, I would like to express my gratitude for family members, friends, colleagues and others that I have not mentioned by name for their support during this study.

Table of Contents

Table of Contents.....	i
List of Tables.....	iv
List of Figures.....	v
List of Algorithms.....	vii
Acronyms and Abbreviations	viii
1. Chapter One: Introduction	1
1.1 Background.....	1
1.2 Motivation.....	4
1.3 Statement of the Problem.....	4
1.4 Objectives	6
1.5 Methods	6
1.6 Scope and Limitations	8
1.7 Application of Results	9
1.8 Organization of the Rest of the Thesis	10
2. Chapter Two: Literature Review	11
2.1 Language Basics	11
2.2 Amharic Language.....	13
2.3 Deafness and Sign Language.....	16
2.3.1 Deafness.....	16
2.3.2 Sign language.....	17
2.4 Ethiopian Sign Language.....	27
2.4.1 EthSL Sign Articulation and Structure	27

2.4.2	Finger Spelling.....	38
2.4.3	EthSL Sentence Structure.....	39
2.4.4	Problems of Translating Amharic to EthSL.....	40
2.5	Machine Translation (MT).....	42
2.5.1	Rule Based MT.....	42
2.5.2	Corpus-based MT.....	45
2.5.3	Hybrid.....	49
2.6	Translation Evaluation Methods.....	49
3.	Chapter Three: Related Works.....	51
3.1	Corpus Based Sign Language Translations.....	51
3.2	EthSL Sign Language Translations.....	52
3.3	Summary.....	54
4.	Chapter Four: Design of Amharic Text to EthSL Translation.....	56
4.1	System Architecture.....	56
4.2	Preprocessing.....	57
4.2.1	Tokenization.....	57
4.2.2	Normalization.....	58
4.3	Word Feature Analysis.....	58
4.4	EBMT Component.....	60
4.4.1	Matching.....	61
4.4.2	Alignment.....	64
4.4.3	Recombination.....	64
4.5	Video Clip production.....	65
5.	Chapter Five: Experimentation.....	66
5.1	Data Acquisition.....	66
5.2	Prototype Implementation.....	68

5.2.1	User Interface Layer	68
5.2.2	Business Logic Layer	69
5.2.3	Data Access Layer	71
5.2.4	The Database	71
5.3	Testing and Evaluation	72
5.4	Results Discussion.....	77
6.	Chapter Six: Conclusion and Recommendations	80
6.1	Conclusion.....	80
6.2	Contribution of the Study	81
6.3	Recommendations.....	82
References.....		83
Annexes		88
Annex A:	Amharic words that are represented by two EthSL signs	88
Annex B:	Days of the week.....	89
Annex C:	EthSL Number signs	90
Annex D:	EthSL Finger Spelling	91
Annex E:	Normalization Algorithms Adopted From Tessema Mindaye	92
Annex F:	Example of Morphologically analyzed Amharic Words.....	94
Annex G:	Details of translations.....	95
Annex H:	Result collection and Result summery forms of system evaluation	99
Annex I:	Words mostly signed by their ASLequivalent other than the EthSL	101

List of Tables

Table 2.1: Seven orders of Amharic letter ‘ለ’	13
Table 2.2: Adequacy and Fluency Scales	50
Table 4.1 Sample morphologically analyzed words.....	59
Table 5.1: Adopted Adequacy Scale	72
Table 5.2: Adopted Fluency Scale.....	73
Table 5.3: Group 1 Test Result.....	74
Table 5.4: Group 2 test result	75
Table 5.5: Group 3 test result	75
Table 5.6: User Acceptance Test.....	76

List of Figures

Figure 2.1: Signing Space.....	21
Figure 2.2: One handed signs without and with body contact.....	23
Figure 2.3: Types of Two handed signs.....	23
Figure 2.4: Sign with a combination of two hand and one hand	24
Figure 2.5: Sign with Non Manual feature	24
Figure 2.6: EthSL Random and mimic sign examples	27
Figure 2.7: EthSL Two handed signs with the same shape and movement	28
Figure 2.8: EthSL Two handed signs showing only the primary hand moves	28
Figure 2.9: Example of showing hand location restriction.....	28
Figure 2.10: Examples of EthSL signs with non-manual features	29
Figure 2.11: EthSL Initialized Signs	30
Figure 2.12: EthSL sign with person ending	30
Figure 2.13 : Synonym signs	30
Figure 2.14: Sign for a single word in Amharic compound in EthSL.....	31
Figure 2.15: Signs for Compound Amharic words and single sign in EthSL	31
Figure 2.16: Signs of compound words in Amharic and compound signs in EthSL.....	32
Figure 2.17: EthSL Possessive pronoun	32
Figure 2.18 EthSL Pronouns.....	33
Figure 2.19: Expressions of spatial relation	33
Figure 2.20: Expressions of the future.....	34
Figure 2.21: Expressions of the present.....	34
Figure 2.22: EthSL sample expressions of the Past.....	34
Figure 2.23: EthSL sign used to represent plural word by adding word ብዙ/bizu/	35
Figure 2.24: EthSL signs indicating plurality by repeating the sign	35
Figure 2.25: Sign whose meaning changed when repeating the sign	36
Figure 2.26: EthSL sign to indicate plurality by moving the sign repeatedly	36
Figure 2.27: Signs for Not (አይደለም).....	37
Figure 2.28: EthSL Negation.....	37
Figure 2.29: EthSL number representation from zero to ten	37
Figure 2.30: EthSL Finger spelling and direction of hand movements	38

Figure 31: Sign for መሄድ 40

Figure 2.32: Ideas expressed with different sign in different documents..... 41

Figure 2.33: Bernard Vauquois ‘pyramid’..... 43

Figure 2.34: The “Vauquois pyramid” adapted for EBMT. 48

Figure 4.1: System Architecture 57

Figure 5.1: Home page 69

Figure 5.2: Amharic letter to EthSL figure spelling translation 70

Figure 5.3: Word translation example for መመገብ and መብላት..... 70

Figure 5.4: Data entry interface 72

List of Algorithms

Algorithm 4.1: Word Feature analysis.....	59
Algorithm 4.2: Structure Matching.....	62
Algorithm 4.3: Feature matching.....	63
Algorithm 4.4: Alignment	64
Algorithm 4.5: Recombination	65

Acronyms and Abbreviations

AAU:	Addis Ababa University
ASL	American Sign Language
ArSL	Arabic Sign Language
Auslan	Australian Sign Language
CBMT	Corpus based MT
DDMT	Data Driven MT
DMT	Direct Machine Translation
EBMT	Example-based MT
ENAD:	Ethiopian National Association of the Deaf
EthSL:	Ethiopian Sign Languages
HoH	Hard-of-Hearing
MoE:	Ministry of Education
MF	Manual Features
MT:	MT
NMF	Non Manual features
NGO	Non-Governmental Organization
POS	Part-of-Speech
RBMT	Rule based MT
SLMT	Sign Language MT
SMT	Statistical MT

SSL	Sweden Sign Language
TL	target language
TM	Translation Memory
USA:	United States of America
WHO:	World Health Organization

Chapter One: Introduction

1.1 Background

Communication is a purposeful activity of information exchange between two or more participants in order to convey or receive the intended meaning by speaking, writing, or using some other medium [1]. Even if Deaf¹ and Hard-of-Hearing (HoH) people, who are completely or partially unable to hear in one or both ear, have a right to get possibly available information from their surroundings, be it knowingly or not, they are deprived of that information as a result of limited communication means. Because of different communication barriers, these people are devoid of easy communication among themselves and with others [2].

Deafness may possibly be congenital in which the hearing loss was already present at birth due to pre-natal disease or genetic defects, or adventitious/ acquired that occur after birth, due to illness, injury and mostly due to exposure to loud noises [3]. There are about 360 million people worldwide that have disabling hearing loss and 32 million of these are children [4] and about one million in Ethiopia with mild to profound hearing loss [2].

People with congenital or acquired hearing impairment at early age and profound are the ones with the most communication problem within a society, as language and communication skill develop most rapidly at early childhood especially before the age of five [5]. This group include pre lingual (native signers) Deaf children, who was born with a hearing loss or whose hearing loss occurred before he/she began to speak, are expected to learn sign language as their primary language [6], which is not actually practiced in Ethiopia where early detection is not practical not only by health centers but also by their hearing families.

People with mild hearing loss and those with age-related hearing loss can be supported by a hearing aid to communicate with spoken language as the hearing aid with different capacity would help them to hear. However a profound Deaf person can't hear any sound and

¹ ¹ It is generally accepted (Leeson, 2003) that 'Deaf' (with a capital D) is used to refer to people who are linguistically and culturally Deaf, meaning they are active in the Deaf community, have a strong sense of a Deaf identity and for whom SL is their preferred language. 'deaf' (with a small d) describes people who have less strong feelings of identity and ownership within the community, who may or may not prefer the local SL as their L1. Hard-of-hearing (HOH) is generally used to describe people who have lost their sense of hearing later in life and have little to no contact with the deaf community or SL usage for various social and cultural reasons.

consequently completely dependent on sign language. Hence, to communicate with a profound Deaf person, sign language must be used. However, it is evident that, in most societies sign language is not known, and if known, it is not well understood even by one's own family.

The themes of International Week of the Deaf 2016 and 2017 “With sign language I am equal” and “Full Inclusion with Sign Language!” respectively emphasize the importance of Sign language in the life of Deaf People. As the themes express Deaf people can fully communicate through sign language.

Sign language is a visual gestural language, used by the Deaf and HoH people for everyday communication. In sign language, information is communicated visually using a combination of manual and non-manual means of expression [2, 7]. The manual parameters are hand shape (configuration), hand posture (orientation), hand location (place of articulation), and hand motion (hand movement). The non-manual parameters include head and body posture, facial expression, gaze, mouth movements and structured gesture. Some signs can be distinguished only by manual parameters, but there are signs which will remain confusing, without the use of non-manual parameters [2, 7].

Researches and studies carried out after the 1960's in the USA and Europe proved that sign languages used by the Deaf are true natural languages with their own grammatical rules [2]. The grammar of sign language is fundamentally different from spoken language. The structure of a sentence in spoken language is linear, one word followed by another; whereas in sign language, a simultaneous structure exists with a parallel temporal and spatial configuration. The configuration of a sign language sentence carries rich information about time, location, person, or predicate [7, 8].

Similar to spoken languages, sign languages vary from country to country. However, Deaf and HoH people from different countries with different sign languages have a unique skill in communicating with each other easily compared to hearing people with different spoken languages. But this doesn't mean a sign language is universal [2].

In Ethiopia, before the introduction of Ethiopian Sign Language (EthSL), American (ASL) and Sweden (SSL) sign languages were the most prominent. The first school for the Deaf in Ethiopia, Mekanissa School for the Deaf (Amha Desta School for the Deaf), was opened in

1961 in Addis Ababa by missionaries from the USA who were using their country's sign language ASL. This in turn had made the educated Ethiopian-Deaf people unaware of their language and culture. To mitigate this challenge, the Department of Special Needs in the Ministry of Education (MoE) organized a committee from different schools and the Ethiopian National Association of the Deaf (ENAD) in order to produce and introduce an Ethiopian sign language (EthSL). Later, the committee published the first EthSL dictionary described in Amharic (ሀ ሞፀሀፍ) in 1975 [2, 9].

In everyday activities, Deaf and HoH people need the support of interpreters. As most interpreters mention, they use both signed Amharic (Amharic on the hand) which is totally different from EthSL and EthSL during translation. Signed Amharic follows the grammar of Amharic whereas EthSL has its own grammar as Signed English follows the grammar of English language and ASL has its own grammar [10].

In addition, the interpreters mention how difficult it is to translate to both pre-lingual and post-lingual Deaf and HoH persons with the same accuracy at the same time. This is so because most post lingual Deaf and HoH persons prefer using signed Amharic for their communication as they are familiar with the speech language (Amharic language) .Whereas the pre-lingual ones prefer their native sign language which is EthSL.

Nowadays, Machine Translation (MT) systems, software to translate text or speech from one language to another or computer translation aids, are preferred over human interpreters because of the following reasons [11]:

- There is too much that needs to be translated,
- Technical materials are too boring for human interpreters,
- Consistency: since computers are consistent, but human interpreters may vary,
- To increase the volume and speed of translation throughput,
- To reduce translation costs.

Although extensive researches have been done on recognition and synthesis of different sign languages to improve communication between Deaf and hearing people, researches on EthSL are very limited. Therefore the main purpose of this study is to develop and introduce *Example Based Amharic Text to Ethiopian Sign Language Machine Translator*.

1.2 Motivation

Even though sign language is used by Deaf and hearing signers fully to communicate among themselves, there is a communication gap between Deaf and hearing people without signing skills. In most situations hearing persons try to communicate with Deaf and HoH by text writing; but, the communication is not satisfactory especially with the pre lingual Deaf where the majority of pre lingual Deaf possesses only poor to moderate reading skills [12]. Since the number of hearing signers is very few, the use of interpreters will be an obligatory requirement in most of Deaf persons' activities which limits their participation.

EthSL is a linguistic medium of communication between Ethiopian people in Deaf communities, allowing them to express their thoughts, knowledge, and needs. There are about one million Deaf people living in Ethiopia with mild to profound hearing loss [2]. Aside from its use among the Deaf, EthSL is used by those who are hard of hearing, and as a second language for hearing people who are in contact with native signers, such as family members and school teachers.

Amharic is the Ethiopian federal government working language and which is widely used as native and learned language in Ethiopia. These facts motivated us to develop Amharic text to EthSL machine translator.

1.3 Statement of the Problem

To ease communication problems of the Deaf and HoH in their day to day activities, interpreters are used even in situations like medical treatment which needs privacy [13]. Finding a qualified sign language interpreter, who is familiar with the ethics and conduct of behavior in the field is difficult. Even if it is assumed that it is possible to get skilled interpreters conceptuality will not be consistent. There are too many concepts to be described in the real world and limited number of standard signs which results misinterpretation [14]. Moreover, small variations in body parts movement represent diverse meanings leading to misconception [15] [14]. Hence, different interpreters may translate a sentence differently having diverse meaning and this brings an undesired effect, which is worse in case of medical and legal issues.

There are different attempts in other countries to mitigate the aforementioned communication problems by the help of Sign Language MT (SLMT) systems, which cannot be applied to Amharic to EthSL Translation directly since they are language specific.

Even though there are various researches done and are ongoing in higher institutes of Ethiopia, on recognition of signs or finger spellings of EthSL and Amharic text to EthSL translators in word level , translation at sentence level is inadequate [16, 17].

Dagnachew [16] developed EthSL synthesizer which translates given Amharic text to the equivalent representation in EthSL at word level with system performance of more than 80%.Daniel [17] also developed a system which translates Amharic text to EthSL at sentence level. Though the research tried to address important aspects it has the following major limitations

- The paper mentioned that it used rule based MT which needs basic grammar rules of EthSL;but the grammar of EthSL has not been soundly studied and documented yet [18]
- It doesn't consider words represented with the same sign in EthSL (one sign in EthSL may represent different words with different meanings). For example መዋል/mewäli/ means spending the day, መቆየት/mek'ojeti/ means to over stay and ማቆየት /mäk'ojeti/ means to keep something for a long time, have same sign.
- It doesn't consider texts containing letters with the same sound
For example, ውሀ, ውሃ, ውኅ, ውኃ, ውሐ, ውሐ /wihε/ represents the same word. 'Water' since the Amharic letters ሀ, ሃ, ሐ, ሓ, ኃ, ኅ /hε/ have same sound.
- It doesn't consider short form of words like “ት/ቤት” for “ትምህርት-ቤት” /timihiriti beti/ to mean School.
- System accuracy is 58.77%, 75.76% and 84% at sentence, letter and number level respectively which is not sufficient for real time applications [17].

It is observed that previous researches are not fully addressing issues towards Amharic text to EthSL SLMT. This study tried to show example based MT can be applied for translating Amharic written text to EthSL. The reason for Example based approach is that EthSL is not

studied well and there are no stated rules, there are only sample statements given to show EthSL grammar is different from Amharic [18].

Thus, this study addresses the following research questions

- Is Example based MT approach applicable for translating Amharic text to EthSL?
- Can we develop an EthSL synthesizer that produces intelligible and natural signs?

1.4 Objectives

General objective

The objective of this study is to design an Example Based Amharic Text to Ethiopian Sign Language Machine Translation System.

Specific objectives

The study has the following specific objectives

- To review Amharic language and EthSL and their interdependence
- To investigate existing Amharic text to EthSL translation systems in order to adopt their best practices and learn from their limitations
- To investigate state of the art of SLMT systems in search of best practices implemented and choose an appropriate one for EthSL
- To organize Amharic and corresponding EthSL text structures at word and sentences level
- To adopt Amharic text to EthSL bilingual dictionaries
- To design Amharic Text to EthSL translation system
- To develop a Prototype System
- To measure the performance of the proposed system

1.5 Methods

Literature review

Critical literature review will be conducted on previous researches in Amharic language, EthSL, Amharic to EthSL translation, MT and SLMT in general. In addition there will be

continuous contact and discussion with individuals and organizations whose work is related to EthSL.

Data collection and management

To achieve the indicated general and specific objectives data will be collected and managed accordingly. The main data collection techniques for the study are document analysis, interview and observations. Documents will be collected and analyzed to extract relevant data for further use.

- EthSL signs will be collected from EthSL dictionary and video recorded sign language tutorials. Videos will be recorded by the help of interpreters and pre lingual Deaf youths as needed in order to adopt Amharic word to EthSL bilingual dictionary containing selected words.
- Text with Amharic and the corresponding EthSL at word and sentences level will be produced by collecting corresponding texts from different documents like Text books (G1-G8) prepared by MoE, tutorials and research papers.
- Facts regarding Deafness, Deaf culture and EthSL will be collected from different stakeholders.

Communication of the Deaf among themselves and with hearing people is observed to be different in sentence construction and even in using different signs for the same concept in different situations and locations.

Apart from document analysis and observations, interview will be conducted with different individuals that have direct relation to the language. Sign language experts working in different colleges and universities, EthSL interpreters and teachers, individuals working in non-governmental organizations (NGO) which work on EthSL and individuals from other organizations who manage issues related to EthSL.

Prototype Development

A prototype with different modules will be developed using various tools to justify if the objective of the study is met and whether its expected functionalities are achieved.

In order to develop the prototype, the tools mentioned below will be used.

- Visual Studio 2013 will be used to develop a web based application with different modules and a graphical user interface.
- Entity Framework 5 will be used as object relational model(ORM)
- SQL Server 2008 will be used for storing and managing our data.

Evaluation and Testing measure

The prototype will be tested to assure that it meets the expected result through qualitative mechanisms. The test will be performed by multiple groups of individuals having different background and experience related to EthSL.

1.6 Scope and Limitations

In machine translation, typical features of source and target languages plays a pivotal role. Managing all words and sentences in a language is a difficult task. Moreover, in the case of Amharic, it's more difficult due to its morphological complexity. In the study, the following cases are not taken into consideration.

- Sentences which needs contextual meaning translations. For Example
 - There are words which have different signs in EthSL according to their context. For instance sentences containing a word መፈለግ /mefeləgɨ/ which has two signs to mean 'search' and 'need' requires contextual translations depending on the meanings in the text. Examples of Amharic words with more than one EthSL corresponding sign for different context is attached in [\(Annex A\)](#)
 - There are also words which are not taken as they are in Amharic Sentence depending on their usage. For example: In Amharic language the word 'ላይ' /läji/ can be used to mean 'above' (the opposite of 'ታች' /tätfɨ/ means 'below') and also to show progress of actions (example: በመክናወን ላይ /bemekenäweni läji/, በመምጣት ላይ /bememiti'äti läji/፣ በመሰራት ላይ /bemesiräti läji/. But in EthSL, this word is used only as opposite of 'above' to mean 'below'; thus, progress of action in EthSL is described by the word 'መቀጠል' /mek'et'eli/ which is equivalent to English word 'continue'. Hence, EthSL equivalent of the Amharic sentence "ሰራው በሂደት ላይ ነው" /siräwi behideti läji newi/ is "ሰራ ሂደት መቀጠል" /sirä hideti mek'et'eli/.

- EthSL Signs which change direction based on subject and object of the Amharic text are not incorporated. For Example መስጠት /mäsit'eti / means to give uses direction from the provider to the receiver.
- Complex sentences are not incorporated
- Words which are spoken and written with different characters due to regional dialect differences. Example ፀፀይ/ts'ehēji/ and ጠፀይ /t'ehēji/ to mean sun and ትላንት/täläniti/ and ትናንት /tinäniti/ to mean yesterday
- The use of ':' in time writing example 10:30 to mean half past ten. ':' is taken as a delimiter character.
- Compound Amharic words which are made of two or more words having a single EthSL sign and Amharic words which have two or more signs when translated to EthSL are not considered.
- It only considers plural words formed by adding the sign ብዙ /bizu/ to mean 'many'.

1.7 Application of Results

As has been discussed in previous sections there are ongoing researches on sign language recognition and synthesis. If these researches become a reality and come of the shelf they will contribute a lot in the lives of the Deaf and community at large.

We believe that this study will:

- contribute to a full fledged implementation of Amharic to EthSL machine translator;
- simplify communication of Deaf persons with their families and the society in general;
- help ENAD and other bodies working on EthSL on preparing dynamic digital dictionary which is more easily understandable than the paper bound dictionaries;
- help ENAD and other bodies working on EthSL on preparing a standardized syntax for EthSL.

1.8 Organization of the Rest of the Thesis

This thesis is divided into six chapters. Chapter Two discusses about basic concepts reviewed from different literatures. Chapter Three incorporates review of related works and the fourth chapter discusses the proposed system and its design. Experimentation, prototype development and evaluation of results are presented in Chapter Five followed by conclusion, contributions of the study and an outlook on possible future work in Chapter Six.

Chapter Two: Literature Review

Concepts that are relevant to convey background information of the study have been discussed in this chapter. Deafness, language basics and system evaluation methods has been discussed briefly to help get the discussion underway. Whereas Amharic language, sign language, MT, types of MT, corpus based MT and aspects of Example based MT (EBMT) have been discussed in depth. EthSL which is the main topic of the study also thoroughly discussed.

2.1 Language Basics

“A language is a complex system of communication with a vocabulary of conventional symbols and grammatical rules that are shared by members of a community and passed on from one generation to the next, that changes across time, and that is used to exchange an open-ended range of ideas, emotions and intentions “ [15].

Natural language has sentences as constituents. A sentence expresses a complete thought (ideas) and is made up of words, which individually or in combination with other words represent ideas. Sentences can in general be broken down into discrete prepositional segments that through their very form and organization contribute to the natural language understanding process [19].The basic atomic unit of meaning is a word. The text you are reading right now has its words clearly separated by spaces, so they stand out as units [20].

Roughly, words are divided into two broad categories. On the one hand there are content words (open-class words) that refer to objects, actions, or proper-content word ties. On the other hand there are function words (closed-class words) that tell us how content words relate to each other. Content words are open-class words, because the class is not a finite set of them, but rather new words are formed and adapted all the time, when new objects (e.g., Google), or types of objects (e.g., web site) enter our mental world. Function words, however, come from a fixed class of words that does not change—at least not more often than every few decades or centuries (the word whom is currently falling out of use). What makes function words tricky for MT is that they fulfill a specific role of how words relate together in one language. This type of role may not exist in another language [20].

An early step of language processing is to divide the input text into units called tokens where each is either a word or something else like a number or a punctuation mark. This process is

referred to as tokenization [21]. For languages that use the Latin alphabet, this is mainly an issue of splitting off punctuation [20].

Part-of-Speech

Linguists group the words of a language into classes (sets) which show similar syntactic behavior, and often a typical semantic type. These word classes are otherwise called syntactic or grammatical categories, but more commonly still by the traditional name parts of speech (POS). Three important POS are noun, verb, and adjective. Nouns typically refer to people, animals, concepts and things. The prototypical verb is used to express the action in a sentence. Adjectives describe properties of nouns [21].

Open-class words fall into three broad categories: words that describe objects (nouns), actions (verbs), and properties. The latter are distinguished into properties that qualify nouns (adjectives), and those that qualify verbs and adjectives (adverbs). Closed-class words are a mixed bag. They include fairly obvious parts of speech such as determiners, pronouns that refer to nouns, prepositions that qualify the role of noun phrases, and coordinating conjunctions such as and. Real text is interspersed with numbers, the possessive marker's, list item markers, various symbols (\$, may be even smiley faces), foreign words, and interjections [20].

One basic task in natural language processing (NLP) is POS tagging. It is the process of assigning a POS tag like noun, verb, pronoun, preposition, adverb, adjective or other lexical class markers to each word in a text. POS tagging is not useful by itself but it is generally accepted to be the first step to understanding a natural language. Most other tasks and applications heavily depend on it [22].

Morphological analysis

A morpheme is defined as the smallest meaningful unit of a language. In linguistics, morphology is the study of words, how they are formed, and their relationship to other words in the same language. It analyzes the structure of words and parts of words, such as stems, root words, prefixes, and suffixes.

Morphological analysis helps to find the minimal units of a word which holds linguistic information for further processing. Morphological analysis plays a critical role in the

development of NLP applications .It is important to be able to handle morphology in English, but it’s absolutely essential when it comes to highly inflecting languages like Amharic [23].

The process of creating words from morphemes is usually known as derivation, while the use of morphemes to modify words is referred to as inflection [15].Stemming reference to a process that strips off affixes and leaves you with a stem. Alternatively, the process may be referred to as Lemmatization where one is attempting to find the lemma (Root) or Lexeme which is citation form of a set of words [21].

2.2 Amharic Language

Amharic is a Semitic language spoken in Ethiopia. It is the second-most spoken Semitic language in the world, after Arabic, and the official working language of the Federal Democratic Republic of Ethiopia. Amharic is also the official or working language of several of the states within the federal system [24].Amharic is written (left-to-right) using ፊደል /fideli/ (Amharic letter) which grew out of the Ge'ez abugida [24].

Amharic language consists of 33 basic and other additional symbols. Each basic symbol has seven different orders. In most languages there are five vowels /a, ä, i, o, u/. In addition to these 5 vowels Amharic has two additional vowels /e, i / the orders are created by combination of each consonant with the seven vowels as indicated in Table 2.1 [25].

Table 2.1: Seven orders of Amharic letter ‘ለ’

	Orders						
	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th
Amharic letter	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ
	/le/	/lu/	/li/	/la/	/lɛ/	/lɪ/	/lo/

Out of the 33 basic consonants, Amharic identifies 28 unique sounds. This implies that, in some cases, more than one consonant is used to represent the same sound. These are:

- ሀ, ሐ and ኀ (hɛ)
- ሰ and ሠ (sɛ)
- ኢ and ዐ (?)
- ኦ and ዑ (ts'ɛ)

The letters in each set represent the same sound. It is important to recognize these letters in NLP tasks. For example, an Amharic word that has the letter ‘ሰ’ can be written in two equivalent ways and must be treated as one for NLP tasks [22]. In case of ‘ሀ’ (hε) there are three basic letters ሀ, ሐ and ሓ and their 4th derived ones ሂ, ላ, ል respectively have also the same sound hence for Amharic word containing this sound there will be six possible combinations for example ሀ-ሂ, ሀ-ሀ, ሐ-ሐ, ሐ-ሐ, ሓ-ሓ, ሓ-ሂ sounds the same /wihe/ to mean water.

Amharic Part of Speech (POS)

Amharic nouns like English are words used to name or identify any of a class of things, people, places or ideas or a particular one of these [19]. Nouns in Amharic can be inflected for gender, number, definiteness, and case.

An important property of the Amharic verbs is that any word that comes at the end of a complete grammatical Amharic sentence is a verb. As a consequence of this property a word at the end of a sentence is expected to be tagged as a verb by an Amharic tagger. Amharic verbs are also known of taking subject markers like ሁ/hu/ to indicate ‘I’ as a subject [19]. Verbs are morphologically the most complex word class in Amharic, with many inflectional forms. A substantial set of words in other word classes are derived primarily from verbs. A single verb may consist time (past, present, and future), gender (male and female), action (command, statement, invitation) and negation (not) [19].

Adjectives in Amharic usually precede the nouns that they modify or describe. ሰነፍ ተማሪ ነው /senefi temäri newi/ in the example the adjective ሰነፍ /senefi/ means ‘lazy’ precedes the noun ተማሪ /temäri/ means ‘student’ which it modifies. But this does not mean that a word is an adjective just because it precedes a noun. For instance, in ይህ በግ /jîhi begi/ means ‘This sheep’, the word ይህ /jîhi/ ‘this’ precedes the noun በግ / begi / ‘Sheep’. Although the word ይህ /jîhi/ functionally shares the feature of an adjective (modifier), it is a pronoun, a demonstrative pronoun [19].

Pronouns, prepositions and conjunctions in Amharic can be individual words like preposition ወደ /wede/ (to) or be bound to other words as affixes. They belong to a closed class. No new words are derived from them [22]. The bound prepositions are mostly prefixes. Conjunctions are suffixes. Common prefixes are ለ /le/ (for), ከ /ke/ (from), በ /be/ (by or with) and የ /je/ (of).

The suffixes are ና/nä/ (and) and ም/mi/ (also). Looking at the first two letters of a word may be enough to identify the prefixes and suffixes. However, there is a trap, some words already have these letters as constituting letters [22].

In Amharic, adverbs can be found in either primitive forms (i.e. as separate words that appear by their own) or in compound forms as combinations of prepositions and some other words but that appear as separate or in rare cases as compound words. In each case they refer to place, time, circumstance etc. [19]. Noun adverbs (or adverbs of position) such as ውስጥ (wisit'i) 'inside' can function alone as an adverb or as a noun. But they often have the locative particle በ /be/ when they are used adverbially. If such noun adverbs are used alone without such prepositions (as በ/be/, ወደ/wedε/, ከ/ke/) they are mostly treated as nouns. On the other hand, noun adverbs prefixed with prepositions are treated as adverbs. That is, ውስጥ /wisit'i/ is a noun while በውስጥ/bewisit'i/ is an adverb. Adverbs can also be formed from adjectives by prefixing በ/be/ as in በሀይል /behejili/ "by force" and በደህና /bedehinä/ "being well" respectively. Adverbs formed in this manner are short adverbial clauses. That is, the resulting words are more of adverbs and are as such considered as adverbs. Days of the week in Amharic language may also be used as either a noun or as an adverb [19]. Adverbs in Amharic are very few. Adverbial functions are often accomplished with noun phrases, prepositional phrases and subordinate clauses. One exception to this is the derivation of adverbs from a few adjectives by adding ና/nä/ [22].

Sentences

Amharic sentences are categorized into simple and complex. A simple sentence is a complete structure that can convey a complete idea. Complex sentences are used to convey complete ideas like simple sentences but unlike simple sentences they are composed of complex phrases. Complex phrases are phrases that have complements and/or modifiers that are sentences themselves [19].

Amharic is a SOV (S (Subject) – O (Object) – V (Verb)) language where words are separated by punctuation marks. [19]. Except in poems, the head verb is usually at the beginning of a sentence. Unlike Arabic and Hebrew, Amharic is written from left to right [22].

Amharic language has its own punctuation marks which demarcate words in a stream of characters which includes ሁለት ነጥብ /huletī net'ibi/ (:), a mark used for separating words. For the modern Amharic, it is left except for hand writing purpose. Its place is almost completely taken over by white space. አራት ነጥብ /rätī net'ibi / (::), it shows the end of a sentence, as a single dot implies end of a sentence in English. ድርብ ሰረዝ /diribi serēzi/ (፣) has the equivalent function of semi-colon of English, ነጠላ ሰረዝ /net'elä serēzi/ (፣) separates lists in Amharic text with equivalent function of comma in English. [22, 26].

In addition to indigenous punctuation marks, some marks have been borrowed from foreign languages. For instance, the exclamation mark ቃለ አጋኖ /k'äle ገgäno/ “!” and the question mark ጥያቄ ምልክት/t'ijäk'e milikiti/ “?” are borrowed from English and used in Amharic language. Unlike English Amharic does not have upper and lower case representations? [26, 27].

2.3 Deafness and Sign Language

2.3.1 Deafness

Deafness is the inability to hear. It can affect one or both ears, totally or partially. Deafness occurring at birth is called congenital deafness. It may also occur later, suddenly or gradually (Acquired). Hearing loss may be mild, moderate, severe, or profound. It can affect one ear or both ears, and leads to difficulty in hearing conversational speech or loud sounds. 'HoH ' refers to people with hearing loss ranging from mild to severe. People who are HoH usually communicate through spoken language and can benefit from hearing aids, cochlear implants, and other assistive devices as well as captioning. 'Deaf' people mostly have profound hearing loss, which implies very little or no hearing. They often use sign language for communication. Over 5% of the world's population 360 million people has disabling hearing loss (328 million adults and 32 million children). Disabling hearing loss refers to hearing loss greater than 40 decibels (dB) in the better hearing ear in adults and a hearing loss greater than 30 dB in the better hearing ear in children. [28].

Human rights can be realized only through one's own language and culture. This is true to all people including Deaf people. The term culture is a very broad concept that encompasses the lifestyle, traditions, knowledge, skills, beliefs, norms and values shared by a group of people.

It also contributes to the artistic expression of members of that group. As many other different human groups, Deaf people have also a culture of their own. They express it through their language, and sign language is a very essential part of the culture [29]. Deaf people themselves use the term “Deaf culture” when referring to the customs, values, world view, habits, and rules of behavior related to their own world [30].

Apart from the places that show inheritance of deafness the Deaf culture is not transmitted from generation to generation. This can be attributed to the fact that only 5-10% of the children in the Deaf society have Deaf parents. More than 90% of the children are born from hearing parents and this results in children inheriting the cultural values from other Deaf people rather than their parents [18].

2.3.2 Sign language

Sign language is at the core of Deaf people’s lives; sign language makes accessibility for Deaf people possible; without accessibility, Deaf people will be isolated. Thus, full enjoyment of human rights for Deaf people is based on the recognition and respect for Deaf culture and identity. Everywhere in the world, language creates culture and vice versa [29].

There are about 70 million Deaf people who use sign language as their first language or mother tongue. It is also the first language and mother tongue to some hearing people and Deaf blind people (tactile sign languages). Each country has one or sometimes two or more sign languages, although different sign languages can share the same linguistic roots in the same way as spoken languages do [31].

There are different definitions given to sign language by various researchers from varying perspective some of the definitions are listed below

“Sign languages are natural languages that were not invented by any single individual. They are shared by the members of a community and passed down from one generation of users to the next. Signed languages do not form a universal language used by Deaf people all over the world, nor are they identical to the types of gesture and mime used by hearing people. They have a similar expressive capacity as spoken languages and are organized around similar grammatical rules. Signed languages have rules for creating new vocabulary and may change

across time, and they are learned by children and appear to be processed by the brain in similar ways to spoken languages” [15].

Oxford Concise Medical Dictionary defines it as, “a form of communication that uses movements of the hands and other parts of the body together with facial expressions instead of sound. There are many different forms of sign language throughout the world. British sign language (BSL) is the form most commonly used in Great Britain. In Northern Ireland, Northern Irish sign language (NISL), and Irish sign language (ISL is the most common form in Ireland) are all used” [32].

ENAD 2003 publication defines it as a language, which is made up of an organized system of signs including gestures, mimes and facial movements, is usually used by the Deaf people, or the hearing who can communicate with Deaf people. Just like spoken language, there are many different sign languages in the world. Like any other languages, sign language can also be elaborated, codified and standardized [14].

Sign languages are the natural languages of Deaf communities and possess phonological systems, morphological systems and syntactic rules, operating within complex grammatical systems [33]. Linguistic work has shown that natural signed languages show all the structural properties of other human languages and that they have evolved independently of the spoken languages which surround them [31]. Although a sign language may be influenced by the spoken language found in the same region, it is distinct, with its own grammar suitable for the efficient use of the body and eyes, and not the vocal/aural system [34].

The acquisition of sign languages by Deaf children with Deaf parents follows the general milestones that characterize acquisition of spoken language for hearing children [33]. Hearing people learn their native spoken language without actually being taught. Language comes naturally. They may have difficulty with the written form of the language, but not in speaking it. Those Deaf people who have Deaf parents (which is less than 10%) also learn their signed languages naturally. However, over 90% of Deaf children have hearing parents who do not know any signed language. Many of these children are not exposed to any language in a natural way during those early critical years of language acquisition. Oral training is not really a substitute for almost effortless natural language acquisition. This lack of early exposure to

any language may be the reason so many Deaf people have difficulty with written language [34].

Trevor Johnston and Adam Schembri points out the myths and misconceptions about sign languages as follows [15] ,

Myths 1: Signed language is a universal language

However signed language is not, a universal language. There are many different signed languages around the world, and many of these have developed independently of each other. Furthermore, not only do signed languages vary from one part of the world to the next, but (like spoken languages), variation can be found in the vocabulary and grammar within particular signed language communities.

Different signers of EthSL may use different signs for the same concept because of their regional origin, educational background and age although signed language is not universal and instead varies from one part of the world to the next, it appears that different signed languages may be more similar to each other than the spoken languages of the world [18].

Myths 2: Signed languages of Deaf communities are based on spoken languages.

Signed languages of Deaf communities are not based on spoken languages, but they may in fact be significantly affected by the spoken language of the surrounding community.

Myths 3: Signed languages are simply pantomime and gesture

Signed languages are not simply pantomime and gesture. Research in linguistics, has demonstrated that the natural signed languages are in fact real human languages, and not simply pantomime and gesture

Myths 4: Signed languages are always iconic

Signed languages are not always iconic the widespread belief that the meaning of all signs comes from their being ‘pictures’ of what they represent.

Myths 5: Signed languages do not have the same expressive capacity as spoken languages.

Signed languages have the same expressive capacity as spoken languages. Contrary to what is sometimes believed, signed languages have the same potential for expressing subtle, technical and complex meanings as spoken languages. Although signed

languages share some properties with gesture and include many iconic signs, this does not mean that they are limited in their expressive capacity

Myths 6: Signed and spoken languages are processed by the brain in different ways

Signed and spoken languages are processed by the brain in similar ways. Signed languages are produced by the hands, face and body, and perceived through vision. This makes them very different from spoken languages that are produced by the speech organs and perceived by hearing. Research has shown, however, that this does not make as great a difference to how signed and spoken languages are processed by the brain as might be expected.

Myths 7: Children learn spoken and signed languages in different ways

Signed languages are acquired by children in the same way as spoken languages. Deaf children with signing parents, signed language acquisition begins at birth and continues through childhood. These children appear to reach all the same developmental milestones at the same age as hearing children learning spoken language.

Myths 8: Signed languages were invented by hearing people

Signed languages were not invented by hearing people. Thus, it can be assumed that signed languages are developed naturally when Deaf people first came together to form Deaf communities.

Myths 9: Signed languages cannot be written down

Signed languages can be written down.

Articulation and Structure

The articulation of a sign language as compared to spoken languages where the primary articulators are the throat, nose and mouth, the main articulators in sign languages are the fingers, hands and arms. The signs themselves are analogous to morphemes in spoken languages and the articulations of the hands and body can be categorized as phonemes like those in spoken languages [12].

However, unlike speech, these phonemes are not linear and sequential, but rather occur simultaneously. There are five categories of phoneme in an articulated sign: manual features

that are the shape, orientation, location and movement of the hand as well as the non-manual features. The simultaneous production of these phonemes allows a signer to communicate different ideas or aspects of a sentence while at the same time exploiting the signing space of their articulation to the full. In this way, complete ideas can be represented in the signing space at any one time. Typical linguistic phenomena used by signers to communicate thoughts and ideas in a way that best uses the visual medium are non-manual features, classifiers, and spatial deictic [12].

Signing Space

Sign languages are gestural languages that are articulated in such a way to make the best use of the space in which articulation can take place, in a similar way that oral speech makes best use of the vocal tract, nose and mouth. This articulation area in SLs is called the signing space. This space extends from just above the head down to the waist and outwards about as far as the arms can extend [12]. A diagram of this space cited in [12] is shown in Figure 2.1

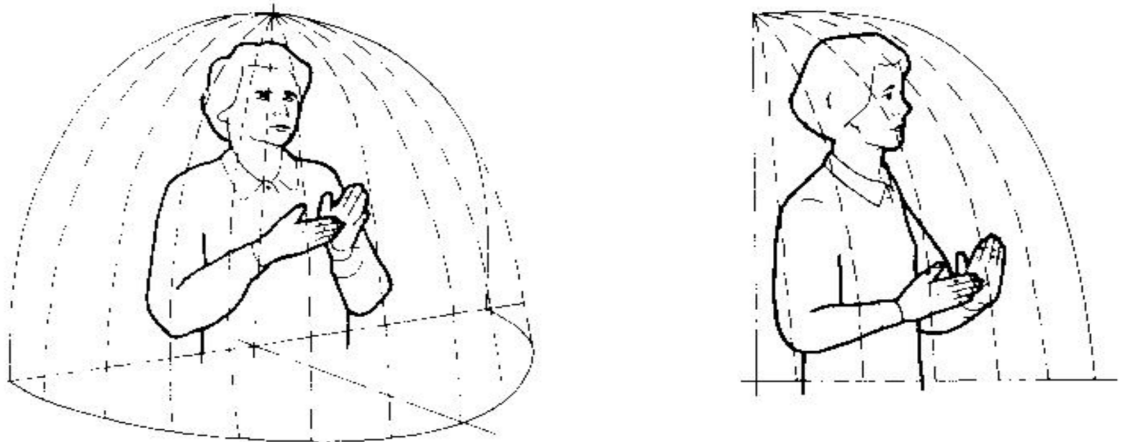


Figure 2.1: Signing Space

Manual signs

Hand shapes are hand configurations which differ non-contrastively from each other, and which signers of a particular signed language may treat as equivalent despite small differences in production [15].

The location of the sign may refer to the hand's actual point of contact on the body, or to the hand simply being significantly near some location on the body. When the sign has no contact

with the body, or when it is not located near some part of the body, it is described as being articulated in neutral space [15].

Movement types in sign phonological structure have been classified into two major categories: primary and secondary movements. Primary movements are sub-classified into path movements (movement from one location to another) and local (or internal) movements (changes in hand shape and orientation) Secondary movements refer to rapidly repeated local movements which can be performed during a path movement or while the hand is stationary [15].

Orientation refers to the direction in which the fingers and palm of the hand are pointing during the production of a sign. The fingers and palm may be oriented into seven basic directions upwards, downwards, right, left, facing each other, towards or away from the signer's body. Diagonal orientations may involve a combination of these elements [15, 18].

There are no signs that are specified as always produced on the right or left hand for all signers. As we will see, some signs are produced with only one hand, and others are produced with one hand acting on the other. In both cases, signers use their dominant hand as the main articulator, depending on whether a signer is right-handed or left-handed [15].

Manual signs are divided into three broad classes, based upon the involvement of one or both hands. One-handed, two-handed and a combinations of one-handed and two-handed elements. Signs within these classes may be further categorized into subclasses based on how they reflect particular patterns of combination [15].

One-handed signs may be subdivided into one-handed signs that are produced in the signing space, and do not involve contact on the body, such as እንጅራ /ʔinidzērä/ (an Ethiopian traditional pita) and one-handed signs that make contact with the body, such as ለምን /ləmini / means 'Why ' [15] as shown in Figure 2.2

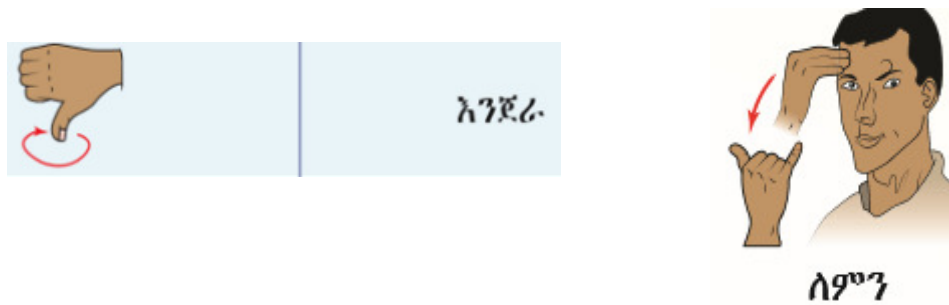


Figure 2.2: One handed signs without and with body contact

Two-handed signs may be subdivided into four types, examples are shown in Figure 2.3. Two-handed signs which both have the same hand shape and in which both hands move in the signing space without making contact with the body such as ክፍል/kifili/ means class. Two-handed signs which both have the same hand shape and, make contact with the body, such as ጤና ይስጥልኝ/t'enä jisit'ilijni/ means Hello. Two-handed signs which both have the same hand shape and one hand acts on the other, such as ስም/simi/ means name. The above three sub classes are referred as double-handed [15]. The other subclass is two handed signs in which one hand acts on the other, and the two hands have different hand shapes, as ቫይ /fäji/ means tea (these are also referred to as two-handed signs) [15].

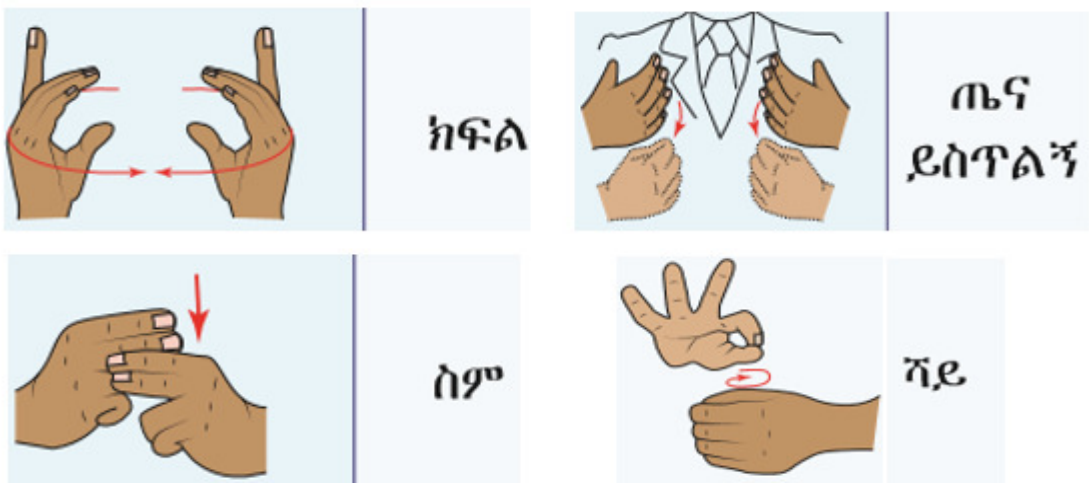


Figure 2.3: Types of Two handed signs

Combinations of one-handed and two-handed signs are signs that involve some combination of the above types, as in the signs **ዘንቢል/zenbīli/** means shopping bag as shown in Figure 2.4

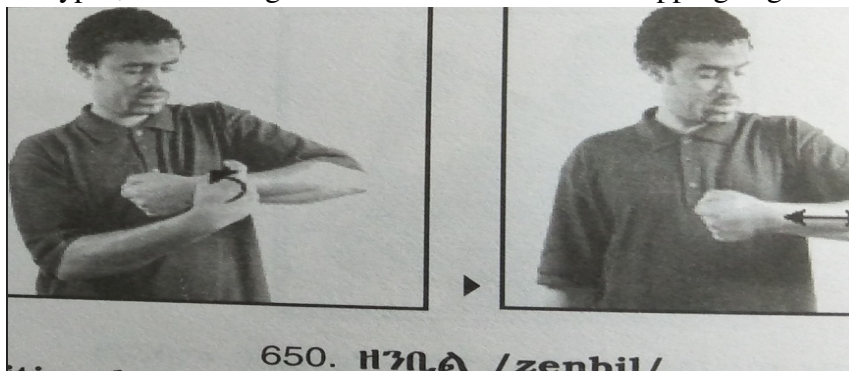


Figure 2.4: Sign with a combination of two hand and one hand

Non-manual signs

Non-manual features (NMFs) are predominantly concomitant with manual signs, they consist of movements or expressions of parts of the body other than the hands that can express emotion, intensity or act as morphological and syntactic markers. These consist of “eyebrow movement, movement of the eyes/cheeks, mouth patterns, tilting of the head, movement of the upper body and shoulder movements” [12].The inclusion of an NMF with a manual sign in discourse can alter the meaning of a sign and its absence can render a sign meaningless [12].Example **ሎሚ/lomi/** means lemon, without the NMF it will not be understood as shown in Figure 2.5



Figure 2.5: Sign with Non Manual feature

NMFs play an important role in composing signs in conjunction with MFs. NMFs can be classified into three types of roles. The first is essential: If an NMF is absent, the sign can have a completely different meaning. The second type of NMF is a qualifier or emotion. In spoken language, intonations, or changes in pitch, can express emotions, such as happiness and sadness; likewise, in Sign language, NMFs are used to express emotion; the third type of NMF

actually plays no role in the sign. In some cases, NMFs remain from a previous sign and are meaningless. Native signers naturally discard any meaningless NMFs based on their knowledge of SL [32].

A classifier is a specific hand shape that can represent a particular person or object. The classifier is then put into motion in an iconic way to make a sentence or part of a sentence. For example, one can say “My car raced down the street, careened to the left, then to the right, then turned over” by setting up the hand shape classifier for “vehicle” as my car, then putting the classifier in motion, showing spatially what the sentence says. In fact, the ASL version would contain even more information about the timing and severity of the motions. There is a grammatically correct way to use classifiers; the description of a rollover is not simply iconic gesturing. Because of the complexity of the ASL grammar and its essentially infinite way to modulate signs to change their meaning, there is no universally accepted written form of ASL or any other sign language [34].

Finger Spelling

In some cases, words may not have corresponding sign; this forces the Deaf to use fingerspelling. Finger spelling is a method of representing words using the signs of each letter the word comprises. If the word has seven characters, it will be represented by the combination of seven signs, one for each letter [25].

Sign alphabets are used by the Deaf to connect to spoken language alphabets, and since the Deaf cannot use their voices to pronounce spoken letters and words, these alphabets are used by Deaf people who know a spoken language as a second language to illustrate the spoken alphabet [32]. Finger spelling is the use of hand configurations to represent the letters of a writing system, with different systems used in different parts of the world. When one fingerspells fully, one more or less ‘writes in the air’, spelling words out manually, letter by letter. Finger spelling may be regarded as an artificial sign system [15].

Finger spelling is a manual code for representing the letters of the English alphabet and is thus not a signed language in and of itself. Finger spelling is generally mixed in with signing and is especially used for spelling nouns (place names, people’s names, names of everyday objects, etc.), or for spelling English words that have no direct signed equivalent [15].

Unlike written language, where a text expresses ideas in a linear sequence, sign language employs the space around the signer for communication, and the signer can use a combination of MFs and NMFs. These are called 'multichannel signs'. The relationship between multichannel signs could be parallel or they could overlap during sign language performance [32].

Sign Language Grammar

Researches indicate that sign languages, as any natural language have their own grammar. [35]. A sign language has its own morphology (the way signs are put together to form a sentence), phonology (the way in which phonemes are combined) and syntax (the grammatical arrangement of signs in a sentence).

To clearly describe these elements of a sign language research must be done regarding its grammatical structures. In the case of EthSL, this facet of the language is not well studied. [18, 35].

Notation Systems

There have been many attempts at creating writing systems for Sign languages, but most are not usable by the general public as they consist of numeric codes or symbols to encapsulate the phonetics or phonology of signs and are not easily learned, written nor is there a standardized accepted form [12].

Signed language writing systems come in two forms: glossing and notation. Glossing refers to the practice of using spoken language translations of signs, together with special symbols to represent the use of space and facial expression. Notation, in contrast, involves the use of special symbols to represent the physical features of signed language itself. The most well-known examples would be Stokoe Notation, first created by William Stokoe, and HamNoSys or the Hamburg Notation System from the Institute for German Sign Language in Hamburg, Germany. None of these systems, however, are intended as practical ways of communicating in a written form of a signed language: they are intended to represent signs and signed utterances for linguistic analysis [15].

2.4 Ethiopian Sign Language

EthSL is an independent language used by the Ethiopian Deaf society which satisfies characteristics of sign languages. All languages have a behavior of inheritance this leads to most languages being influenced by others likewise EthSL also influenced by ASL, SSL and Amharic, but this doesn't mean that EthSL is ASL, SSL or Amharic language [18].

EthSL is composed of random and mimic signs. The random signs has no resemblance with the things they represent like ስም/simi/ to mean name while the mimic signs have some resemblance with the things they represent [18] like ገገፎ /geñifo/ means 'Porridge' and ስጋ /sigä/ means 'Meat' as shown in Figure 2.6



Figure 2.6: EthSL Random and mimic sign examples

2.4.1 EthSL Sign Articulation and Structure

EthSL has its own linguistic behavior, [18] and behaviors shared with other sign languages. EthSL encompass different manual and non-manual signs.

Manual Features

There are one handed and two-handed manual signs in EthSL. Hands are recognized as primary hand and secondary hand. The right hand is primary for right handed person whereas left hand is primary for left handed person [18].

Pawlos indicates some guidelines to articulate EthSL signs as follows [18]

- One-handed signs should be articulated using primary hand
- For two-handed signs if both hands move, then both hands should have the same hand shape (similarity rule) Example መጻሕፍ/ mets'ihəfi/ means book , ክፍል/kifili/ means class as shown in Figure 2.7

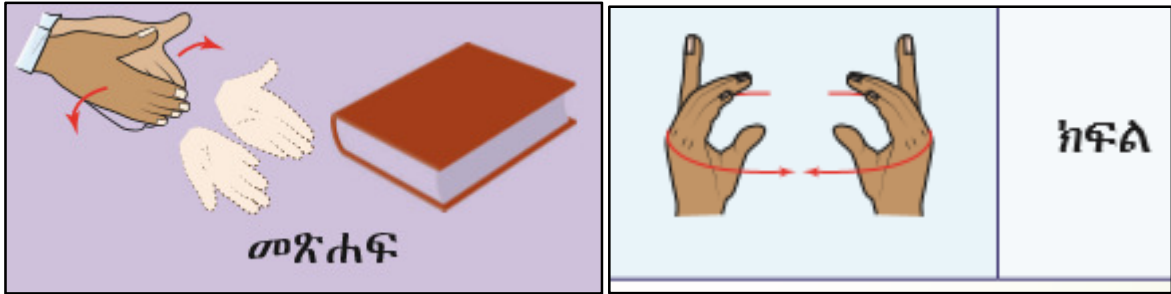


Figure 2.7: EthSL Two handed signs with the same shape and movement

- For two handed signs if hands have different shapes then only the primary hand moves, (Primary የንቁነት/ jɛnik'unɛti/ rule) ቁጣ/k'it'ä/ to mean Ethiopian traditional flat bread



Figure 2.8: EthSL Two handed signs showing only the primary hand moves

- There is hand location restriction in representing the name of males and females. It is not allowed to touch above or by the side of your eyes to represent a female name and below your eyes to represent a male name [25]. The relationship between hand location and gender can be shown clearly with signs used to articulate corresponding words used for male and female like father/ mother (አባት/äbäti/ እናት/äinäti/) and brother/sister (ወንድም/wɛnidimi/ እህት/äihiti/) as shown in Figure 2.9, having the same hand shape, orientation and movement and difference in location

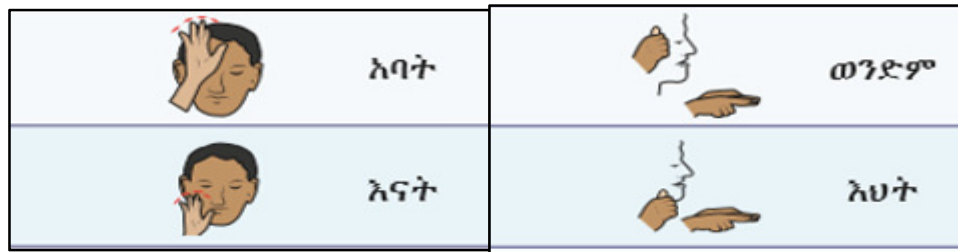


Figure 2.9: Example of showing hand location restriction

- Direction is very important on signing since it will change meaning. It will provide information about the subject and object ,example ሞስጠት to say 'give' use the give

sign from you to the other body and to say ‘take’ use the same sign with the opposite direction [25].

Non Manual features

EthSL also incorporates the behavior of using non manual features including facial expressions and other movements of the body in different signs. It is important to remember that a signer can only be said to have used signs proficiently when he or she has a full understanding of the non-manual features and can use them fully when communicating [36]. Signs that express happiness, sadness, pain and other emotions must include these non-manual features as indicated in Figure 2.10

Example in EthSL መጫወት/ጨዋታ /metʃ'äweti/tʃ'ɛwätä/ to play, ስቃይ/sik'äji/ to mean torment.

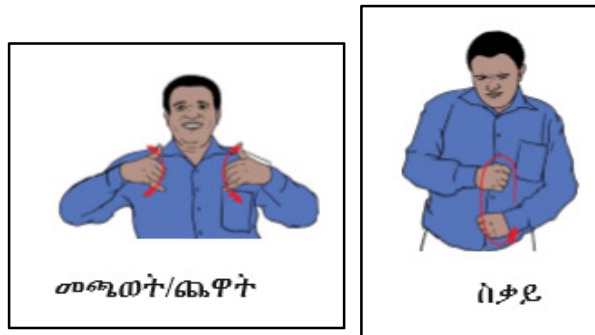


Figure 2.10: Examples of EthSL signs with non-manual features

Intensity: In spoken Amharic intensity is shown by adding words and by changing the pitch of the sound. In EthSL intensity is expressed by increasing the speed of signing for the given sign or by adding facial expressions. For example by increasing speed while signing the word ‘Walk’ we can indicate fast walking.

Initialized signs: There are some EthSL Initialized signs. Initialized signs are thought to be influenced by the introduction of oralism into Deaf education. The hand shape used in these signs is the first, or initial, letter of the word in the spoken language [12]. In case of EthSL the hand shape used in these signs is the base of the first, or initial, letter of the word in Amharic. For example, እንጀራ/ ?inidzerä/ (hand shape of ኦ/?/) and ወሃ/Wihä/ (hand shape of /ወ/) as shown in Figure 2.11 and signs of days of the week ([Annex B](#))



Figure 2.11: EthSL Initialized Signs

Person Ending: Person ending signs are those signs that are used to show person's occupation or nationality. Person ending is performed by combining the sign and pointing to in front of the chest and moving down መምህር/ məmihiri/ to mean teacher it includes the sign of ማስተማር /mäsitəməri/ teaching with person ending sign as shown in Figure 2.12

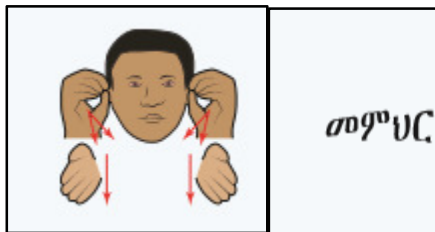


Figure 2.12: EthSL sign with person ending

Synonyms: There are signs which are represented by the same sign like መዋል/mewäli/ means spending the day መቆየት/mek'ojeti/ means to over stay ማቆየት / mäk'ojeti/to keep something for long time as shown in Figure 2.13

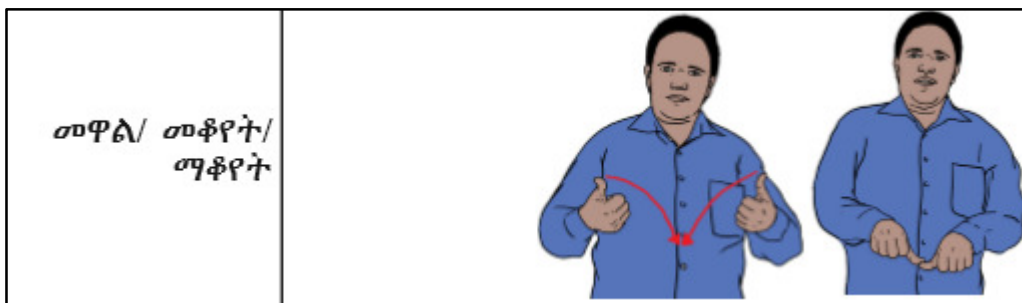


Figure 2.13 : Synonym signs

Compounding: As two words can combine to form a compound word in the spoken language, two signs can be combined to represent a single item in EthSL [18] .

Single word in Amharic compound in EthSL

Example ሸበት/ḥibeti/ and ጠመኔ/ṭ'emene/ are shown in Figure 2.14

ነጭ/netḥ'i/+ጸጉር/p'eguri/ = ሸበት/ḥibeti/ (White + hair = white hair)

ነጭ/netḥ'i/+መጻፍ/mets'äfi/= ጠመኔ/ṭ'emene/ (white +write=chalk)

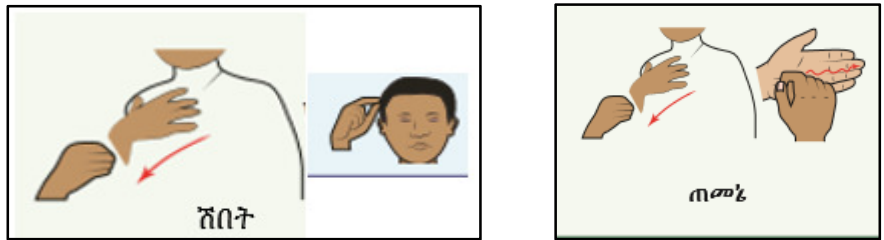


Figure 2.14: Sign for a single word in Amharic compound in EthSL

Compound words in Amharic single sign in EthSL

Example ትምህርት ቤት/timihiriti beti/ means School, የምልክት ቋንቋ/jemilikiti k'wanik'wa/ means sign language has a single sign in EthSL as shown in Figure 2.15 respectively



Figure 2.15: Signs for Compound Amharic words and single sign in EthSL

Compound words in Amharic compound signs in EthSL

Example ጥቁር ሰሌዳ/t'ik'uri seledä/= ጥቁር/t'ik'uri/+ አራት መዓዘን/ገጾትi meጎzeni/

ቀይ/k'ejji/ + ስር/Siri/ = ቀይስር/k'ejjisiri/ are shown in Figure 2.16



Figure 2.16: Signs of compound words in Amharic and compound signs in EthSL

Pronouns and Possessive Pronouns

Pronoun እኔ/ጎine/ to mean I or me is articulated with hand shape ' ሃ' /nε/ and tilted upward neck and የእኔ/ጎεጎine/ mine is articulated with hand shape 'ሸ'/ጎε/and neck tilted forward with the same movement (touching the chest once) and orientation (in ward)as shown in Figure 2.17. My self is articulated with hand shape 'አ'/ጎ/with side orientation and touching the chest repeatedly



Figure 2.17: EthSL Possessive pronoun

Other possessive pronounce are articulated with hand shape ሸ'/ጎε/ and neck tilted forward with the same movement and orientation as their corresponding pronouns (አንተ /ጎnite/) you for male አንቺ/ጎnitji/ you for female እናንተ/ጎinänite/ you for plural እነርሱ /ጎinerisu/ they and እኛ/ጎijä/ we) shown in Figure 2.18 [36].

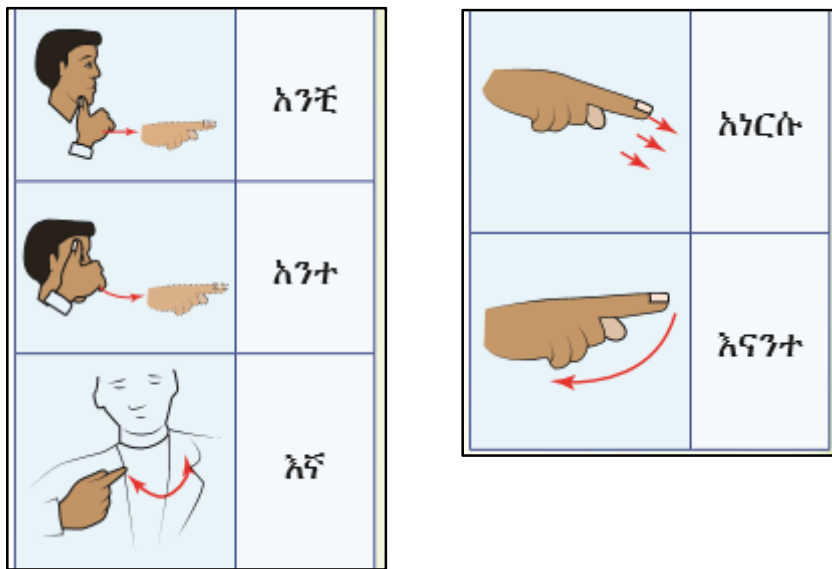


Figure 2.18 EthSL Pronouns

Facial expressions and hand movements have an important grammatical role in EthSL when describing Space, amount, and time.

Place: When describing spatial relationships with signs, it is possible to simply point in the direction in which the object or person being discussed is found. Objects in our vicinity or a short distance away can be indicated by placing our hands parallel to our shoulders, using slight facial expressions and pointing to the side. Things further afield can be expressed with emphasized facial expressions and hand gestures above the shoulders pointing in their directions for example እሩቅ /ʔiruk'i/ means far away as shown in Figure 2.19 [36].

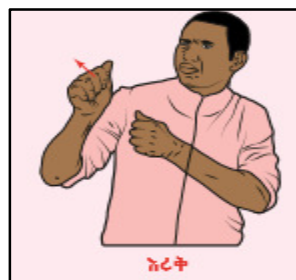


Figure 2.19: Expressions of spatial relation

Time (Tense): In EthSL signs for present are signed in front of the body, past moving the body backward and future by moving body forward in addition to including tense indicator signs [36].

Prophecies or the future are expressed by using the space directly in front the signer. For example ነገ /nɛgɛ/ means tomorrow, ከነገ ወዲያ /kɛnɛgɛ wɛdijä/ means the day after tomorrow and ወደ ፊት/Wɛdɛ fiti/ means in the future are shown in Figure 2.20.

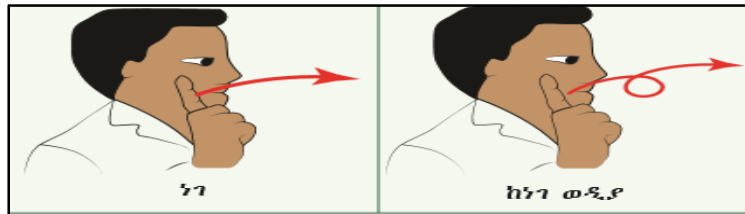


Figure 2.20: Expressions of the future

The present time can be expressed by using the space in front of the abdominal region of the signer. For example አሁን/?huni/means now, ዛሬ/zäre/ means today are shown in Figure 2.21.

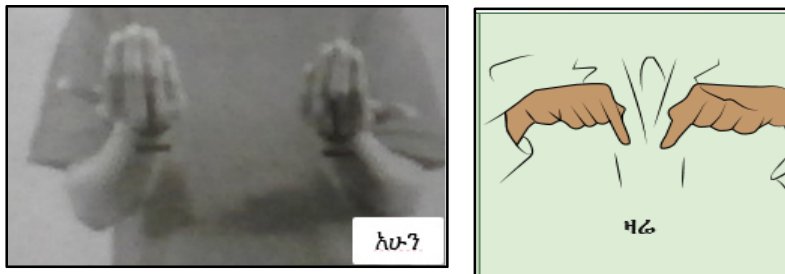


Figure 2.21: Expressions of the present

The past is expressed in relation to the space behind the signer or by using the hand movements backwards. For example ትናንት/tinäñiti/ means yesterday, ከትናንት ወዲያ/kɛ tinäñiti wɛdijä/ means the day before yesterday) are shown in Figure 2.22.

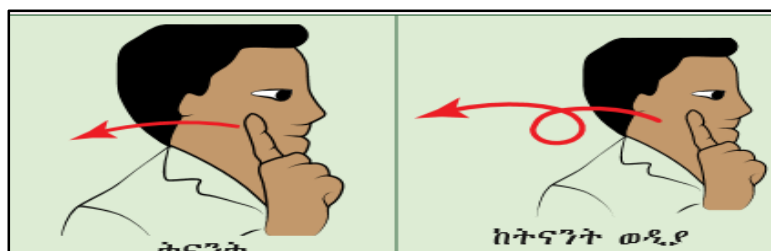


Figure 2.22: EthSL sample expressions of the Past

Amount: There are different methods in EthSL to indicate amount. One method is to add the sign for many (**ብዙ/bizu/**) after the noun. For example **መፅሀፍቶች/mets'ihəfotʃi/** means books can be articulated by the sign for **መፅሀፍ/mets'ihəfi/** book followed by the sign **ብዙ/bizu/** to mean many as shown in Figure 2.23 [36].

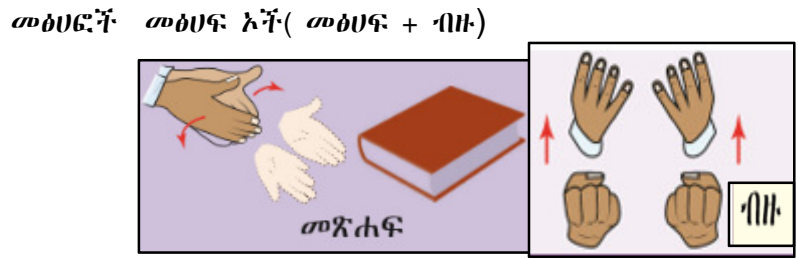


Figure 2.23: EthSL sign used to represent plural word by adding word **ብዙ/bizu/**

Another method is to repeatedly use a sign. For example the sign for **ቅጠል/k'it'eli/** means leaf when it repeated results **ቅጠል ቅጠል/k'it'elä k'it'eli/** means leaves, the same way when the sign for iron **ብረት/bireti/** repeated it results a sign for **ብረታ ብረት/biretä bireti/** as shown in Figure 2.24



Figure 2.24: EthSL signs indicating plurality by repeating the sign

But it doesn't mean that repeated signs are always results the plural form for example when the sign **ውይይት /wijijiti/** means discussion repeated it results **ክርክር/kiṛikiṛi/** to mean argument not **ውይይቶች/wijijitotʃi/** not discussions as shown in Figure 2.25



Figure 2.25: Sign whose meaning changed when repeating the sign

The other is there are signs when the sign moves repeatedly to the right or to the left results the plural form for example ዛፍ/Zäfi/ means Tree repeatedly from it results the sign for forest ጫካ/tj'äkä/ as shown in Figure 2.26



Figure 2.26: EthSL sign to indicate plurality by moving the sign repeatedly

Negation

Negation is expressed by concatenating አይደለም/?jidelemi/ (one of the two signs shown in Figure 2.27) to mean not to the sign of base word. For example አልወድም/?liwededimi/ mean 'I don't like' will be articulated with the sign of መውደድ/?mewidedi/ mean like followed by the sign of አይደለም/?jidelemi/ (not) as shown in Figure 2.28



Figure 2.27: Signs for Not (አይደለም)

እውዳለሁ Vs አልወድም = መውደድ + አይደለም

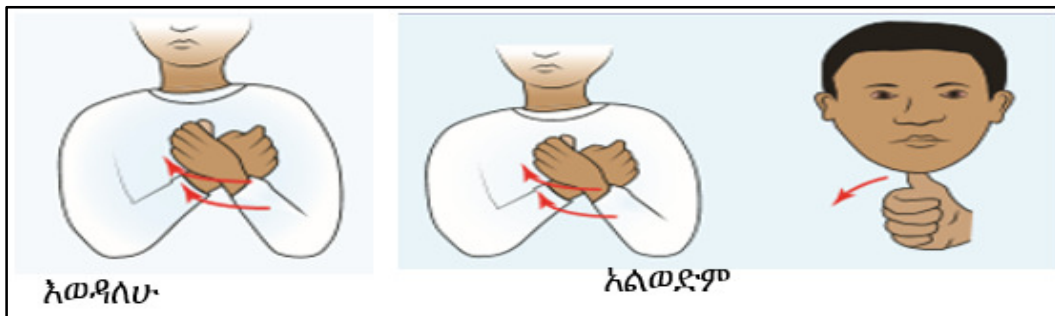


Figure 2.28: EthSL Negation

Number Representation

There are basic signs for numbers 0, 1 to 10 as shown in Figure 2.29 and there are also basic signs for numbers 20, 30, 40, 50, 60, 70, 80, 90, hundred, thousand, million and billion ([Annex C](#)). Any other number is represented as a combination of these basic numbers.

	0 ዜሮ		4 አራት		8 ስምንት
	1 አንድ		5 አምስት		9 ዘጠኝ
	2 ሁለት		6 ስድስት		10 አሥር
	3 ሦስት		7 ሰባት		

Figure 2.29: EthSL number representation from zero to ten

2.4.2 Finger Spelling

Finger spellings are representations of the spoken language letters used to articulated objects that have no signs in the sign language vocabulary. For example names of places and individuals as well as ideas [36].

EthSL has a key in using Amharic letters (fidels), different hand configuration to represent basic letters and different hand movements to represent their relatives as shown with example in Figure 2.30. ([Annex D](#))



Figure 2.30: EthSL Finger spelling and direction of hand movements

As we discuss in section 2.3.2, finger spellings are not part of any sign languages, the same way የጣት ፊደላት/jēt'ätì fideli/ finger spellings used to represent Amharic letters are not part of EthSL and used only in the following conditions

- used to express names of persons and places
- words for which no signs exist
- for emphasis
- to communicate with someone with inferior sign language skills

It is essential to keep in mind the following guidelines when we articulate ideas using finger spelling [18, 36].

- Finger spelling should be articulated with the primary hand
- Letters in a word should be articulated separately
- During finger spelling hand should be below the chine and in front of shoulder
- Eye contact should be maintained with the audience
- Palm should always be facing away from the signer
- We have to insert hyphen (-) between every letter of a word which is going to be finger spelled during inscription
- The signer must be clearly visible

- There should be optimum distance between the signer and the persons being communicated with.
- There should be optimum speed during articulation
- We have to indicate the end of one word with a four fingers closed hand with side orientation by moving from right to left before starting the next word

2.4.3 EthSL Sentence Structure

In spoken languages word is the base of meaning but in case of sign language the base of meaning is a sign, but sign language is not the exact copy of spoken language. This is to mean that sign language is a natural language with its own grammar [36].

Intensely studied sign languages like ASL and Auslan have identified Grammar, in case of EthSL even if it is not intensively studied and has a stated rule it has its own grammar which is different from Amharic and other languages When we articulate EthSL signs based on Amharic or Affan Oromo Grammar we articulate Signed Amharic or Signed Afann Oromo not EthSL [18] .

Here are some examples of Amharic and EthSL equivalent sentences to show some guidelines of EthSL sentences. Since Amharic has no capital and small letter variation it is not possible to show the SL equivalent like in ASL, by writing the equivalent sentence with capital letter, hence we follow the strategy followed by [18] underline the EthSL sentences to differentiate from their Amharic Equivalent

Example 1. አማርኛ ወደ ትምህርት ቤት እሄዳለሁ /wɛdɛ tɪmihiriti beti ʔihedälehu/

አምቋ እኔ ትምህርት ቤት መሄድ /ʔine tɪmihiriti beti məhedɪ/

EthSL has limited signs hence it has single sign for the base word and the morphological derivations are not indicated except that indicate pronouns in the above example እሄዳለሁ/ʔihedälehu/ to mean ‘ I will go’ is replaced by the main verb መሄድ/məhedɪ/ and the suffix ‘ሁ’/hu/ indicating me is replaced by እኔ/ʔine/’ I ’. Since the sign for መሄድ/məhedɪ/ indicate the direction there is no need to show the sign for the preposition word ወደ/wɛdɛ/ as shown in Figure 2.31.



Figure 31: Sign for መሄድ

Example 2. አማርኛ ስምህ ማን ይባላል /simihi mäni jibäläli/

አ.ምቋ አንተ ስም ማ /?niite simi mä

In EthSL words like ይባላል/jibäläli/, ነው/nəwi/, used for questioning are expressed by pointing towards the object and like in the above example ስምህ/ simihi/ since ህ/hi/ indicate the pronoun አንተ/ ?niite/ means you expressed with the sign of the base word ስም/ simi / preceded by the pronoun አንተ/?niit/

2.4.4 Problems of Translating Amharic to EthSL

Like spoken languages variation can be found in the vocabulary and grammar within particular signed language communities. Different signers of Auslan may use different signs for the same concept because of their regional origin, educational background and age [15].

The same way different signers of EthSL use different signs for the same concept and also in case of EthSL it is more serious because even materials prepared by stakeholders like ENAD MoE and NGOs uses different signs for the same concept example of words which are articulated with different signs in different books, schools and locations as shown in Figure 2.32. Example እንቁላል/?inik'uläli/ means egg, ተፈጥሮ/tefət'iro/means nature አይደለም/?jidelemi/ means not.

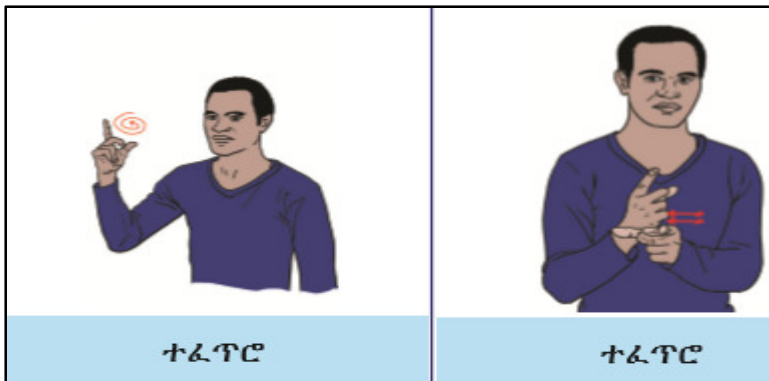
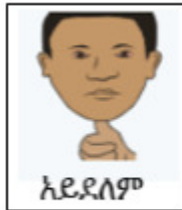


Figure 2.32: Ideas expressed with different sign in different documents

And also there are sentence translation examples which differ

Amharic ስምህ ማን ነው? /simihɨ māni newi/ meaning ‘What is your name?’

EthSL አንተ ስም ማን? /nɨtɛ simi mā/ [18] or ስም ማን/simi mā/ [36]

The other problem is since there is shortage of sign language vocabulary Deaf society mostly create their own signs used locally like at home, classroom, church and so on . Even professional interpreters have difficulty translating in different situations.

The process of teaching people with hearing impairment using sign language in Ethiopia is difficult and this is because of two principal reasons. The first is lack of proper sign language for technical and scientific terminologies. The second reason is in relation to the immaturity of level of the language and the shortage of sign language vocabulary, which forces teachers or interpreters to create their own sign (symbol) to conduct the teaching Process for hearing impaired students that may differ from school to school [14].

2.5 Machine Translation (MT)

MT as a branch of computational linguistics which is an automatic process by computerized systems that convert a piece of text (written or spoken) from one natural language referred to as a source language to another natural language called the target language with human intervention or not, and with the objective of restoring the meaning of the original text in the translated text [37].

The term MT is a standard name for computerized systems responsible for the production of translations from one natural language into another with or without human assistance. It is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another [38].

There are three main approaches to build a machine translator. Namely knowledge driven approach also known as Rule based MT(RBMT), Data Driven Machine translation (DDMT) approaches which is also know an corpus based MT. Finally, hybrid MT which combines the advantages of the above mentioned methods. The classes are based on underlying theory, for RBMT uses linguistic theory while DDMT uses data theory [37].

2.5.1 Rule Based MT

Rule-Based MT (RBMT), is a general term that denotes MT systems based on linguistic information about source and target languages basically retrieved from (bilingual) dictionaries and grammars covering the main semantic, morphological, and syntactic regularities of each language respectively.

The following are the shortcomings that are associated with RBMT approach:

- Insufficient amount of really good dictionaries.
- Building new dictionaries is expensive.

- The need to manually set for some linguistic information.
- Difficulty in dealing with rule interactions in big systems,
- Ambiguity and idiomatic expressions.
- Failure to adapt to new domains.
- Although RBMT systems usually provide a mechanism to create new rules and extend and adapt the lexicon, changes are usually very costly and the results, frequently, do not pay off [39].

There are three different approaches under the rule-based MT Approach. They are Direct, Transfer-Based and Interlingua MT Approaches respectively.

Though they all belong to the RBMT, they differ in the depth of analysis of the source language and the extent to which they attempt to reach a language-independent representation of meaning or intent between the source and target languages.

Their dissimilarities can be obviously observed through the Vauquois Triangle which illustrates these levels of analysis. The Vauquois Triangle shows comparative depths of intermediary representation, Interlingua MT at the peak, followed by transfer-based, then direct translation. As shown in Figure 2.33, the length of the arrows signifies the amount of ‘effort’ required for translation [39].

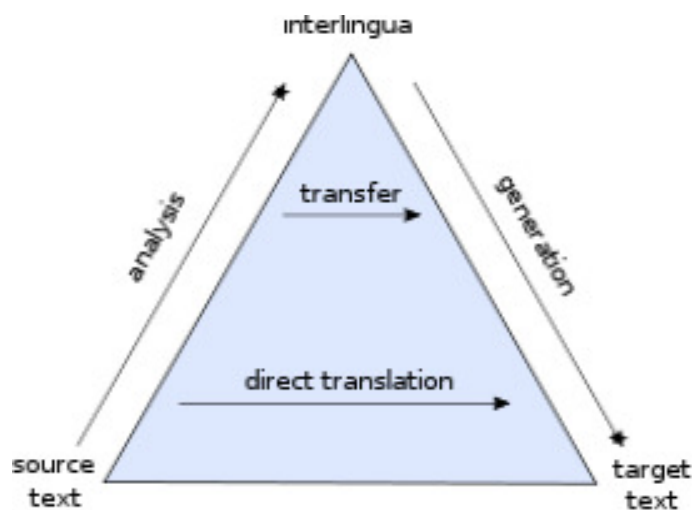


Figure 2.33: Bernard Vauquois ‘pyramid

Direct MT (DMT) Approach

DMT approach is the oldest and less popular approach. Direct translation is made at the word level. MT systems that use this approach are capable of translating a language, called source language directly to another language, called target language (TL). Words of the source language are translated without passing through an additional/intermediary representation. The analysis of source language texts is oriented to only one TL. Direct translation systems are basically bilingual and uni-directional. Direct translation approach needs only a little syntactic and semantic analysis [39].

Interlingua MT Approach

Interlingua MT is one instance of rule-based machine-translation approaches. In this approach, the source language, i.e. the text to be translated, is transformed into an Interlingua language, i.e. a —language neutral representation that is independent of any language. The target language is then generated out of the Interlingua. One of the major advantages of this system is that the Interlingua becomes more valuable as the amount of target languages it can be turned into rises [39].

Transfer-based MT Approach

Because of the disadvantage of the Interlingua approach, a better rule-based translation approach was discovered, called the Transfer-based Approach.

Transfer-based MT is similar to Interlingua MT in that it creates a translation from an intermediate representation that simulates the meaning of the original sentence. Unlike Interlingua MT, it depends partially on the language pair involved in the translation. On the basis of the structural differences between the source and target language, a transfer system can be broken down into three different stages: i) Analysis, ii) Transfer and iii) Generation.

In the first stage, the Source language parser is used to produce the syntactic representation of a Source language sentence. In the next stage, the result of the first stage is converted into equivalent Target language oriented representations. In the final step of this translation approach, a target language morphological analyzer issued to generate the final target language texts [39].

One of the limitations with transfer Based MT approach is that rules must be applied at every step of translation. There are rules for source language analysis (syntactic/semantic), rules for source-to-target transfer and rules for target language generation. It is also difficult to do as much work as possible in reusable modules of analysis and synthesis as well as keep transfer modules as simple as possible [39].

2.5.2 Corpus-based MT

A corpus (plural *corpora*) can be simply defined as no more than a collection of language data, composed of written texts, transcriptions of speech or a combination of recorded speech and transcriptions. It is a special collection of textual material collected according to a certain set of criteria the general issue is whether the corpus is a representative sample of the population of interest. A sample is representative if what we find for the sample also holds for the general population [21].

R. Kibble categorizes corpora into three broad categories as follows [40]

Monolingual corpora consist, as the name suggests, of data from a single language.

Comparable corpora include a range of monolingual corpora in different languages, preferably with a similar level of balance and representativeness, and can be used for contrastive studies of those languages.

Parallel corpora include original texts in one language with translations of those texts in one or more different languages. Parallel corpora can be used to train statistical translation systems

A further distinction is between corpora consisting solely of the original or 'raw' text and those that have been marked up with various annotations.

Finally, corpora can be further classified according to their structure: – an unorganized collection of individual texts (Isolated)–organized texts by categories (Categorized)– some organized with overlapping categories (Overlapping).texts organized to indicate language use over time(Temporal) [40].

Corpus based MT (CBMT) (also referred as data driven MT) is an alternative approach for MT to overcome the problem of knowledge acquisition problem of rule based MT. CBMT uses, as its name points, a bilingual parallel corpus to obtain knowledge for new incoming translation. This approach uses a large amount of raw data in the form of parallel corpora.

This raw data contains text and their translations. These corpora are used for acquiring translation knowledge. CBMT approach is further classified into Statistical MT and Example-based MT (EBMT) approach [39].

In CBMT a distinction is made between, on the one hand, statistical MT(SMT), based primarily on word frequency and word combinations, and on the other hand, example-based MT(EBMT), based on the extraction and combination of phrases (or other short segments of texts). In both cases the corpora comprise bilingual texts (originals and their translations) [41].

Statistical MT Approach

Statistical MT (SMT) is generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora. The initial model of SMT, based on Bayes Theorem, proposed by Brown et al. takes the view that every sentence in one language is a possible translation of any sentence in the other and the most appropriate is the translation that is assigned the highest probability by the system [39].

M. D. Okpor points out the challenges of Statistical MT Approach in "MT Approaches: Issues and Challenges" as follows [39].

- Corpus creation can be costly for users with limited resources.
- The results are unexpected. Superficial fluency can be deceiving.
- Statistical MT does not work well between languages that have significantly different word orders (e.g. Japanese and European languages).

Example-based MT Approach

The idea for EBMT dates from about the same time, though the paper presented by Makoto Nagao at a 1981 conference was not published until three years later (Nagao 1984). The essence of EBMT, called "MT by example guided inference, or MT by the analogy principle" by Nagao, is succinctly captured by his much quoted statement stated below [42, 43]:

Nagao 1984 mentioned translation

"Man does not translate a simple sentence by doing deep linguistic analysis, rather, Man does translation, first, by properly decomposing an input sentence into certain fragmental phrases then by translating these phrases into other language phrases, and finally by

properly composing these fragmental translations into one long sentence. The translation of each fragmental phrase will be done by the analogy translation principle with proper examples as its reference.”

Example Based MT (EBMT) is an alternative model for rule based systems in MT world. In rule based systems linguistic knowledge is established by rules, but in EBMT, the linguistic knowledge is extracted from previous examples of translations [43].

Example-based MT (EBMT) is characterized by its use of bilingual corpus with parallel texts as its main knowledge, in which translation by analogy is the main idea. An EBMT system uses a set of sentences in the source language and corresponding translations of each sentence in the target language with point to point mapping, and uses those examples to translate other, similar source-language sentences into the target language [39].

The basic premise is that, if a previously translated sentence occurs again, the same translation is likely to be correct again. EBMT systems are preferable in that they require a minimum of prior knowledge; therefore, they are quickly adaptable to many language pairs [38].

EBMT process as case-based reasoning. This paradigm has been evolved as an alternative to rule based systems. In case based reasoning, the experience is derived from past 'cases', and the problem is solved using this experience taken as knowledge. A new translation (which corresponds to "new problem" in Case-Based Reasoning) is achieved by finding the most appropriate example from the translation database; then the selected example is used as a model for translating to the required target language [43].

EBMT is often linked with the related technique of "Translation Memory" (TM). This link is strengthened by the fact that the two gained wide publicity at roughly the same time, and also by the use of the term "memory based translation" as a synonym for EBMT. Although they have in common the idea of reuse of examples of already existing translations, they differ in that TM is an interactive tool for the human interpreter, while EBMT is an essentially automatic translation technique or methodology. They share the common problems of storing and accessing a large corpus of examples, and of matching an input phrase or sentence against this corpus; but having located a (set of) relevant example(s), the TM leaves it to the human to decide what, if anything, to do next, whereas this is only the start of the process for EBMT [42].

Nagao correctly identified the three main components of EBMT: matching fragments against a database of real examples, identifying the corresponding translation fragments, and then recombining these to give the target text [42].

It is perhaps instructive to take the familiar pyramid diagram, probably first used by Vauquois (1968), and superimpose the tasks of EBMT as shown in Figure 2.34. The source text analysis in conventional MT is replaced by the matching of the input against the example set (MATCHING: finding the example closest to the input sentence [32]). Once the relevant example or examples have been selected, the corresponding fragments in the target text must be selected. This has been termed “alignment” or “adaptation” and, like transfer in conventional MT, involves contrastive comparison of both languages (ALIGNMENT). Once the appropriate fragments have been selected, they must be combined to form a legal target text, just as the generation stage of conventional MT puts the finishing touches to the output (RECOMINATION) as shown in Figure 2.34.

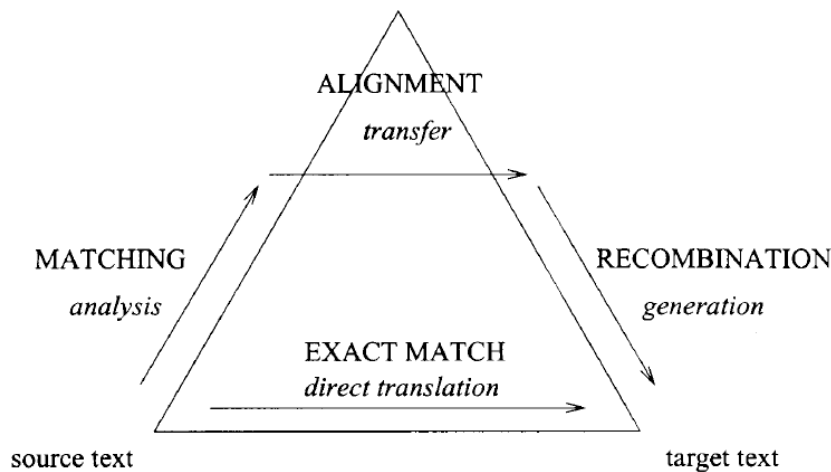


Figure 2.34: The “Vauquois pyramid” adapted for EBMT. .

One argument for exploring EBMT approaches is that since it is based on actual texts, output translations should be more readable and more sensitive to contexts than RBMT systems, i.e. of higher quality in appropriateness and idiomaticity. A second argument is that EBMT systems can be more easily improved, by the addition of more examples from bilingual corpora; whereas the improvement of RBMT systems involves the modification and addition of complex rules and lexical entries. A third is that EBMT does not involve the complexities

of lexical and structural transfer found in (most) recent RBMT systems, i.e. that the basic architecture of EBMT is simpler and less prone to failure than RBMT. As a fourth point, it is argued that EBMT can deal with cases of translation involving complex structural differences and subtle lexical choices that RBMT often fails at. In general, the argument in favor of EBMT is its potential to improve the generation of TL sentences [44].

When one of the following conditions holds true for a linguistic phenomenon, rule-based MT is less suitable than EBMT [42].

- Translation rule formation is difficult.
- The general rule cannot accurately describe [the] phenomenon because it represents a special case.
- Translation cannot be made in a compositional way from target words.

EBMT is an attractive approach to translation because it avoids the need for manually derived rules. However, it requires analysis and generation modules to produce the dependency trees needed for the examples database and for analyzing the sentence. Another problem with EBMT is computational efficiency, especially for large databases, although parallel computation techniques can be applied [39]. Furthermore as the system is example based, gathering examples, number of examples, suitability of them and lastly storage of examples builds up the main problems and divergence point of example based MT systems [43].

2.5.3 Hybrid

Hybrid MT is a method of MT that is characterized by the use of multiple MT approaches within a single MT system, which has proven to have better efficiency in the area of MT systems. Combining some of the best features of some approaches to form a hybrid approach helps in taking care of the challenges posed by many approaches [39].

2.6 Translation Evaluation Methods

Translation of a system can be evaluated manually by human evaluators or using automatic evaluators.

Automatic evaluation metrics were invented within the last decade and are frequently used to determine the accuracy of MT outputs. All of these metrics aim to show a good correlation with human evaluation. The problem with these metrics, however, is that they are designed to

evaluate natural language, which has a different representation. Natural language representation is linear, while SL is multi-channel [32, 12].

Human (manual) evaluation, translation system is tested by stakeholders and their judgments will be collected to be summarized as the system's strength and weakness. This method is considered the most reliable way to evaluate any translation system [32] even if it is expensive, time-consuming and it is subjective because the resulting evaluation scores typically are not identical to other evaluators' scores. Many factors can affect the score judgment, such as evaluator knowledge.

MT techniques can be evaluated by Adequacy and Fluency evaluation methods. Adequacy tests the amount of meaning expressed and conveyed in the translation, whereas Fluency tests the well formedness and fluency of the translation. Two five-point scales for adequacy and fluency used in human evaluation as shown in Table 2.2 [32].

Table 2.2: Adequacy and Fluency Scales

Adequacy		Fluency
5	All meaning	Flawless
4	Most meaning	Good
3	Much meaning	Acceptable
2	Little meaning	Flawed
1	None	Not acceptable

Chapter Three: Related Works

There are different researches ongoing locally and globally in sign language recognition and synthesis to improve communication between Deaf and hearing people. We reviewed various documents related to the study, and referred useful ideas in different chapters especially in literature review of this document. In this chapter we only present researches conducted in EthSL synthesis and two Example Based sign language synthesis of Irish and Arabic sign languages which we think are more related to our study.

3.1 Corpus Based Sign Language Translations

Many corpus-based studies on SL translation have been conducted recently [32]. In this section, three related works on Example based which gave lesson to this paper are discussed.

A research work by Sara Morrissey and Andy Way, [45] was conducted with the objective of developing SLMT for translating English text into Irish Sign Language (ISL). Data-driven methods which can learn from natural language data are followed and a corpus of text was developed in the source and target language with identified domain of airport information announcements. The research results in more than two thirds of the words produced are correct and almost 60% of the time the word order is also correct.

The other research reviewed is a work on the application of data-driven MT to sign Languages. The experiment demonstrated that despite the difficulty faced in Source sign language data suitability to data-driven MT, translation was possible with dataset as small as 561 sentences. The researchers also projected that given a larger dataset in a more restricted domain and using a more sophisticated system, there are grounds for the development of a successful Sign language data-driven MT system [12]. The work also pointed out the data-driven approach used, examined the MT system utilized and mentioned modification made for the treatment of SLs and their annotation. Through sets of automatically evaluated experiments in both language directions, they considered the merits of data-driven MT for SLs and outlined the mainstream evaluation metrics used. To complete the translation into SLs, the addition and manual evaluation of a signing avatar for real SL output is discussed.

The dissertation study conducted in [32] presented a corpus-based system for translation from Arabic text into Arabic Sign Language (ArSL). The research used corpora in two domains where the first one being restricted to instructional language used in schools for deaf students; for this domain 203 Arabic sentences that comprise the most common instructions given by teachers to deaf students and 710 signs are taken. The second domain was constructed from children's story written in Arabic called "Rabbit and Fox"; the signed version of this story contains more than 2,478 signs. The main strength of the research is that it shows multiple ways to implement EBMT approach by developing different versions of Arabic text-to-ArSL EBMT systems. Finally, on the basis of the evaluation results, two methods, phrase-based and similarity based methods, were combined to construct a system that can deliver translations with higher accuracy. It also used concatenated sign video clips as output rather than a signing avatar for simplicity and to distinguish translation or sign synthesis errors easily.

3.2 EthSL Sign Language Translations

Endale [35] developed a self-learning web-based hypermedia learning tool for EthSL. The researcher developed a prototype of hypermedia-learning tool called a Medical Ethiopian Sign Language Learning Tool (MESL2T) by taking health environment as a case which helps hearing people capable of reading and writing Amharic language to learn EthSL. Even if it is not a sign language synthesis research, it is pioneer in being the first study of EthSL in relation to information technology.

Minilik [46] carried out a research with the general objective of developing an avatar based translation system by adopting a machine translation approach that can take an Amharic text as an input and generates an EthSL equivalent of the input text by Ethiopian Finger alphabet. This work is the first in using avatars to animate EthSL finger spelling which can be taken as its strength. Its main drawback is that it uses finger spellings to spell the Amharic word rather than using the corresponding signs of the word in EthSL. But, as per the discussion in about sign languages Finger spelling is a manual code for representing the letters of the spoken language alphabet and mostly used to illustrate words which have no corresponding sign. Since the researcher used only Finger spelling in the translation, it cannot be taken as a translation from Amharic to EthSL.

Masresha [47] conducted a research with the objective of exploring the possibility of developing an automatic translation system that allows users to translate written Amharic text into EthSL using a virtual human. The intended users of this system are those who can read and write Amharic language. The research focuses on Amharic personal and possessive pronouns in that the system takes Amharic word, letter and number as an input and the output is avatar animation of EthSL. The performance of the study scored average accuracy of identification 51% for words, 76% for letters and 80% for numbers. Using the system, it is possible to translate Amharic text into gestures that a hearing impaired can understand and utilize it to develop written language skills in addition to learning sign language. But this study works on word level only with lower accuracy of 51%; it also doesn't include sentence level translation.

Dagnachew [16] also studied about developing Amharic text to EthSL conversion MT system at word level. The system encompasses Amharic morphological analyzer, Amharic morpheme to EthSL mapping bilingual dictionary, and SiGML transcription of selected EthSL words. RBMT approach was used for limited vocabulary and it resulted in a system which maps a given Amharic word to EthSL equivalent sign whenever the word exists in the bilingual dictionary or finger spell; but if the word doesn't exist, it performs morphological analysis and synthesis. The system performance was found to be more than 80%. The strengths of this study are the use of morphological analyzer which enables to translate more words and the use of EthSL finger spellings when there is no equivalent sign in the bilingual dictionary. This study thus informs that using morphological analyzer to analyze Amharic words is mandatory for understanding the words of the language.

Daniel [17] developed Amharic sentence to EthSL translator using RBMT approach. He uses rule based approach and supports simple Amharic sentences, letters, or numbers, and outputs 3D animation of EthSL with system accuracy of 58.77%, 75.76% and 84% at sentence, letter and number level respectively. The main strength of this study is that it consider translation of Amharic text into EthSL at sentence level. EthSL experts discussed that EthSL grammar is not studied yet even if it has its own grammar. Hence the main drawback of the research is using rule based approach for MT to a language having no stated rules which is not trustworthy.

The other shortcoming observed is the use of animated avatars for output in which the Hamnosy code needs to be written manually. This manual writing may result errors in code which in turn results error in the output both in manual and non-manual features of the sign. Errors caused by incorrect Hamnosys code can be one of the reason for low accuracy of sentence translation.

3.3 Summary

From the review on related works, a lot of information and knowledge has been acquired on sign languages, EthSL, and the different approaches for machine translation. One of the major lessons learnt was the existence of different approaches to develop a machine translation and the potential to develop various architectures for a selected approach. The criticality of understanding the source and target language prior to selecting appropriate approach for machine translation is also understood from the review; hence, domain experts of Amharic and EthSL languages are involved in this work to understand the languages' characteristics to be an input for the proposed solution. Moreover applying translation techniques suggested on researches for other languages is found to be difficult due to peculiar characteristics and complexity of both Amharic language and EthSL; thus, algorithms which can be adopted to fit for this study are taken from the reviewed related work. In general, reading on different literatures helped us identify gaps of the researches done on EthSL and suggest possible approaches to handle them.

Even if Dagnachew tried to include a morphological analyzer in his study to understand words of Amharic language he did not identify and take all important features of Amharic words during translation to EthSL. The other researches on EthSL do not consider the use of morphological analyzer. Important features of Amharic words to be extracted from the result of morphological - analysis during translating to EthSL will be identified.

In this study we also consider words represented with the same sign in EthSL (one sign in EthSL may represent different words with different meanings), texts containing letters with the same sound and short form representation of Amharic words.

In using avatars as a means to display signs, translation errors and errors committed by the avatar cannot be distinguished easily [32]. Hence using sequential video clips instead of avatars is proposed for this study. The need for analyzing Amharic and EthSL parallel examples to identify required structures and features for translation purpose is recognized

since EthSL has no standardized rule. Thus, this study will try to address identified gaps and proposed suggestions to achieve improved translation.

Chapter Four: Design of Amharic Text to EthSL Translation

This chapter discusses the design of an Example Based Amharic Text to Ethiopian Sign Language Machine Translator. It states what typical features of Amharic language are included in the design process. It also explains the details of components on the architecture and their algorithms.

4.1 System Architecture

The system takes Amharic text as an input and gives its equivalent EthSL translation as a sequence of video clips. The system architecture shown in Figure 4.1 is composed of three essential components that are Preprocessing, Word Feature analysis and EBMT.

The text is preprocessed to make it suitable for further steps. Feature analysis is then performed on the preprocessed sequence of words to identify required features and to use it as an input to the main EBMT component. The EBMT component uses examples from the database for translation. The detailed explanation of the components is given in the next sections.

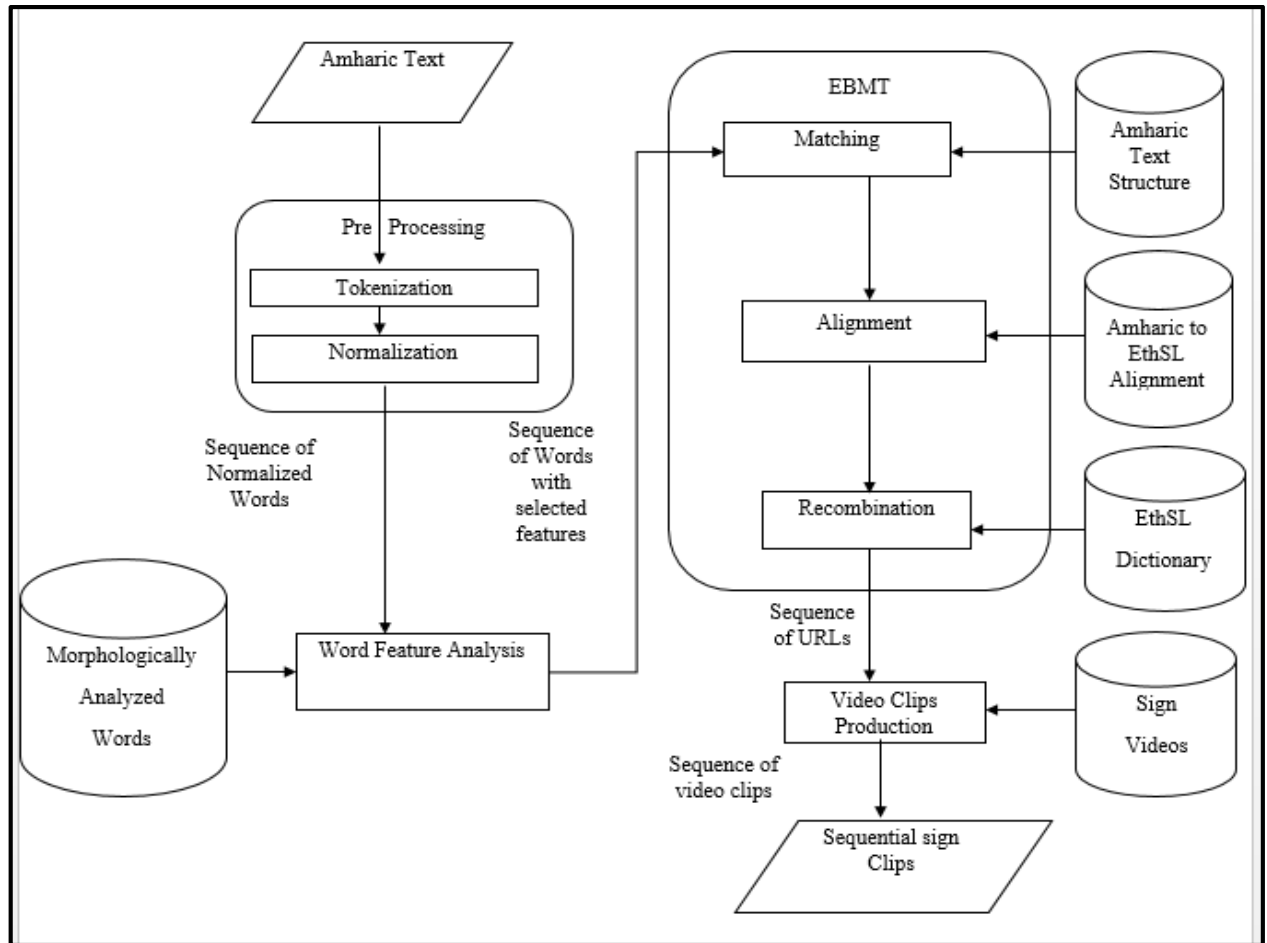


Figure 4.1: System Architecture

4.2 Preprocessing

Pre-processing is involved in preparing the input sentence into a format that is suitable for further analysis. The pre-processing stage in the architecture consists of tokenization and normalization.

4.2.1 Tokenization

The first procedure in this preprocessing component is tokenization. The tokenization takes Amharic text entered by the user and tokenizes it into a sequence of Amharic words. The tokenization of words is performed using Amharic punctuation marks as delimiter characters which are white space, ' '(hulet netib), '⋮' (arat netib), '፣' (netela serez), '፤' (dereb serez), '!' (kaleagano) and '?' and (Question mark). Also it takes '?', '⋮', '፣' and '!' as indicators of end of a sentence and use to differentiate sentence types.

4.2.2 Normalization

Normalization is performed on the word tokens that result from Tokenization. As Tessema discusses in [48], two types of normalization issues arise in the Amharic language. The first one is the identification and replacement of short hand representation of a word that is written using forward slash “/” or period “.”. An example is the replacement of “ት/ቤት” by “ትምህርት ቤት” /timihiriti beti/. The second normalization issue is the identification and replacement of Amharic alphabets that have the same pronunciation. The replacement is made using a representative alphabet from a set of similar alphabets. For example, identifying words like ውሃ, ውሀ, ውሐ, ውሐ, ውኅ, ውኃ/ wihä/ which are pronounced the same to mean “Water” and replace by ውሀ/wihä/. We used Tesema’s [48] algorithms attached in ([Annex E](#))

4.3 Word Feature Analysis

Morphological analysis is important for morphologically complex languages like Amharic because it is impossible to store all possible words in the database. Hence it is used to reduce various forms of a word to a single root word and to analyze different features of the word. Even if, the use of morphological analyzer is mandatory, it is not integrated in this component. In this study morphological analysis for Amharic words is performed by Amharic language experts based on important and mostly used features needed during Amharic to EthSL translation. These features are identified with the help of Amharic language experts, EthSL experts and EthSL interpreters since we could not find documented information. As a result POS, root, Negation, plurality, pronouns (subjective, objective and possessive), and tense are identified as required features for translation. The result of the analysis is organized as database to be used for feature extraction.

Samples of analyzed words are indicated in Table 4.1. It can be seen from the table that for Amharic word አልተጫወተም/ʔlitetj’äwetemi/ means ‘he didn’t play’, the morphological analysis results the words POS is verb with root ‘ጭወት’/tj’iwiti/ and subjective pronoun እሱ/ʔisu/ means he .The value for Negation = 1 indicates the word is negative which has equivalent meaning in EthSL as ‘አይደለም’/ʔjidelemi/ indicating not . If value is 0, it means that the word has positive sense. These features are extracted in feature extraction phase.

Table 4.1 Sample morphologically analyzed words

Amharic Word	Amharic Word Root	Amharic Word POS	AttachedPreposition	is it Negated	is it Plural	Subjective Pronoun	Objective Pronoun	possessive Pronoun	Amharic Word Tense
መብላት	ብልእ	ስም		0	0				
በላ	ብልእ	ግስ		0	0	እሱ			ሀላፊ
በላች	ብልእ	ግስ		0	0	እሷ			ሀላፊ
መጽሀፎች	መጽሀፍ	ስም		0	1				
አልሰጠሁትም	ስጥእ	ግስ		1	0	እኔ	እሱ		
አልተጮደሁትም	ጮድት	ግስ		1	0	እሱ			

This component accepts a sequence of normalized words and for each word it extracts important features from morphologically analyzed words database and results in sequence of words with their selected features as described in Algorithm 4.1. The sequence of words with their selected features are thus input to the EBMT component.

High Level Description of Word Feature Analysis Algorithm

Input: a sequence of normalized words

Output: a sequence of words with their extracted features

For each word in the text

search the word from morphologically analyzed words

If it exists then extract word features

else

quit the process

end if

end for

if all words of the sequence found

then return the sequences of words with their

features

endif

Algorithm 4.1: Word Feature analysis

4.4 EBMT Component

As it was discussed in literature review, EthSL is not a well-studied language and has no stated rule for its grammar; thus, in this work corpus based approach is selected rather than RBMT. In addition, there are no sufficient EthSL documented sentences that can be collected and EBMT on the other hand requires a far smaller corpus size than SMT. Therefore, Example based approach is selected rather than statistic based approach. The benefits of selecting EBMT approach as discussed in [32] are:

- The accuracy of the approach is easily extended by simply adding extra examples to the corpus.
- The approach does not require linguistic rules, relying purely on example-guided suggestions.
- Unlike SMT approach, EBMT can translate using a limited corpus,
- Its accuracy depends only on the quality of the examples and their degree of similarity to the input text.
- There are many ways to implement EBMT systems

A prerequisite for a data-driven approach is a large bilingual corpus aligned at sentence-level from which training and testing data are extracted. As indicated in [12] the manual modality of Sign languages and the lack of a standardized writing system contributes to the limited availability of Sign language data both in terms of desired quantity and quality for use in a data-driven SLMT system. Similarly limited materials about EthSL are available and most of the materials contains word level sign equivalents.

A database for this component is designed to encompass Amharic text structures with their EthSL equivalent structure at word and sentence level, and an adopted EthSL dictionary of letters, words, and synonym words.

The EBMT component consists of three stages Matching, Alignment and Recombination both function according to the data stored.

4.4.1 Matching

This stage is used to find the closest match for the input from Amharic sentence structure examples in the knowledge base. The closest match is obtained by performing two phase matching, structure matching and features matching. Both phases use the result of the previous stage which is a sequence of words with their selected features.

Structure matching

In this matching phase the structure of the input sentence will be matched with structure of example cases in the knowledge base (Amharic Text Structures). Example cases having the same structure (mainly based on POS) with the input sentence will be selected as a candidate sentence and will be ready for the next selected features matching phase as a sequence of candidate sentences. We adopted and modified the algorithm used in [32] Algorithm 4.2.

Note: in the algorithms: IS for Input Sentence, EC for Example Cases and CS for Candidate Sentence are used

Algorithm: structureMatching ()

Input: Structure of Input Sentence (IS), Example Cases (EC)

Output: Candidate sentence structures (CS) and no of candidates (m)

```
m ← 0
Foreach (EC)
  simScore ← 0
  if |IS| > |ES|
    then n ← |EC|
    else n ← |IS|
  end if
  for i ← 1 to n do
    if IS[i,pos]=EC[i,pos]
      then if IS[i,pos]='N' or IS[i,pos]='V'
            or IS[i,pos]='adj'
            then simScore ← simScore + 2
            else simScore ← simScore + 1
            end if
      end if
    end for
  if simScore >= n + 2
    CS[m] ← EC[i, id],
    m ← m+1
  end if
end foreach
No of selected cases ← m
```

Algorithm 4.2: Structure Matching

Feature Matching

Here, word features of the input sentence will be matched against word features of the candidate sentences for searching a sentence case with maximum similarity. The EthSL sentence equivalent of the finally selected sentence is taken to the Alignment stage of EBMT Algorithm 4.3.

```
Algorithm :featureMatching()
Input: IS words with their features ,Candidate sentences(CS)
Output:Id of a selected sentence with maximum similarity
(selectedSenCaseId)
maxsimfeature← 0|
For i← 0 to No of selected cases
  do Simfeature ← 0
  For j← 0 to no of words in IS
    do For k← 0 to no features
      do If ISwordfeat[i,j,k ]--CSwordfeat[i,j,k]
        then simfeature ← simfeature+1
        end if
      end for
    end for
    If (simfeature > maxsimfeature)
      then maxsimfeature← simfeature
      selectedSenCaseId← CS[id]
    end if
  end for
```

Algorithm 4.3: Feature matching

4.4.2 Alignment

Alignment is used to identify which parts of the corresponding translation are to be reused. EthSL sentence structures for the corresponding Amharic Sentence structures are aligned with the help of experts based on the features in the Amharic words and it is stored in the knowledge base. During the alignment task, the system thus takes the corresponding EthSL Structure and align the input word with its feature based on the data it retrieves from the database (Algorithm 4.4).

High Level Description of Alignment Algorithm

Input: sequence of words with their features and
selected example sentence structure id

Output: aligned sequence of words and their features

For the selected structure id

search its corresponding EthSL Structure

Align the sequence of words and their features according
to the corresponding example EthSL structure

end for

Algorithm 4.4: Alignment

4.4.3 Recombination

Recombination is the last stage next to alignment which manages recombination of the aligned EthSL words with their corresponding signs URLs. Hence the result of the interpretation will be displayed as series of signs in video format Algorithm 4.5.

High Level Description of Recombination Algorithm

Input: aligned sequence of words and their features

Output: sequence of URLs

For every word and feature in the sequence

 Search the word or features URL from basic word Table
 of EthSL Dictionary

 If it exists return the URL

 else search it in the synonym table

 if it exists return the URL

 else

 quit

 end if

end if

Algorithm 4.5: Recombination

4.5 Video Clip production

The sequence of URLs received from previous stage is used to retrieve the sign video clip from the sign videos database and play the video clips one after another sequentially.

Chapter Five: Experimentation

This chapter discusses about Prototype Implementation, Testing and evaluation of the proposed design of the Example Based Amharic Text to Ethiopian Sign Language Machine Translation.

5.1 Data Acquisition

The main sources of data in this study are

- Experiences of pre lingual and post lingual Deaf signers
- Sentences constructed by pre lingual signers taken from Kotebe university graduate program paper,
- Amharic to EthSL dictionary,
- Text books prepared by MoE for Grades 1-8,
- Video records prepared by AAU and NGOs,
- Experience of EthSL Experts from AAU, Kotebe University, MoE and NGOs,
- Experience of EthSL interpreters
- And also experiences of families of Deaf.

A database for this study is built initially by collecting about 160 parallel example sentences of Amharic and EthSL. Important and mostly used features of Amharic words during Amharic to EthSL translation are identified with the help of Amharic language experts, EthSL experts and EthSL interpreters. Based on these identified features and technical support from experts, preprocessing, word feature analysis and EBMT alignment stages of the designed system discussed in chapter four, are performed on the collected Amharic sentences.

A total of 1700 words, 600 words taken from the example sentences and 1100 more additional words, are morphologically analyzed as per the selected features by the help of Amharic language experts. The result of the analysis is organized as knowledge base for later reference during translation. Sample of analyzed words is attached in [\(Annex F\)](#).

As a result of the Morphological analysis, features are extracted and alignment is performed. Accordingly, 38 distinct word structures with their EthSL equivalent are identified. Using these distinct word structures and adding sentence type information (statement, interrogative, etc), the example sentences are also analyzed and aligned; hence, 91 distinct sentence

structures with their EthSL equivalent are also identified. Then, the identified word structures and sentence structures with their EthSL equivalent are added to the knowledge base. Therefore, the knowledge base is taken as a primary knowledge to be referenced during Example based translation at word level and sentence level.

Secondly Amharic to EthSL dictionary of about 245 EthSL finger spellings (fidels), and about 280 words from the dictionary prepared by ENAD and some words which are not included in the dictionary were adopted by adding the root forms of the words. This is due to the fact that Amharic words in the existing Amharic to EthSL dictionary are not in their root form. Root forms were obtained from morphological analysis in word feature analysis phase. The root forms may not be the most used forms of the word especially in case of verbs. The root form of Amharic words are generated from sequence of corresponding Amharic letters in the sixth sound (ሳድስ ፊደላት/ sädisi fideläti). For example ብልፅ/ biliṣi/, ግጥም/ ገግጥም/ are root forms of መብላት/ mebiläti/ means to eat and መጫወት/ metj'äwetä/ means to play respectively. These roots are not found in the referenced dictionary but are added in the adopted dictionary. The adoption is done by aligning the words with the URL of the corresponding sign's video record by the help of a sign language interpreter and a pre lingual signer.

EthSL words which are represented by the same sign are organized as synonym words by aligning them to their corresponding word in the adopted dictionary. An equivalent long form representation for words written in their short form like “ት/ቤት” for “ትምህርት ቤት”/timihiriti beti /are also included in our knowledge base and whenever these cases appear the system refers to the knowledge base.

More word examples, sentence examples and analyzed words can be added to the knowledge base by experts using the system interface. Similarly the adopted dictionary can be enriched by adding more words with their corresponding signs.

5.2 Prototype Implementation

This prototype is an Asp .Net web based application which accepts Amharic text from the user and translates it to the equivalent EthSL. Translation is performed based on the knowledge base and other computations.

Developing a full-fledged functional machine translator is a resource intensive task. It needs involvement of language and IT experts and requires many time and also needs a storage and a processor because of its complexity.

The prototype was developed, tested and evaluated using a laptop with Intel® Core™ i3 CPUM 370 @ 2.40 GHZ speed, 4.00 GB (3.8 GB usable) of Installed memory (RAM), 400 GB of hard Disk capacity, with 64 bit Windows 10 Operating System.

C# programming and Entity frame work as object relational mapper using Visual studio 2013 integrated development environment (IDE) are used as a development tools.

The prototype is a three tier application

- User interface layer responsible to accept input Amharic text from the user and display the output EthSL video clips.
- Business logic layer is the one responsible to handle every logic behind translation
- Data access layer is the back end component which integrates the Business logic and the interface with the knowledge base using entity Ado data.

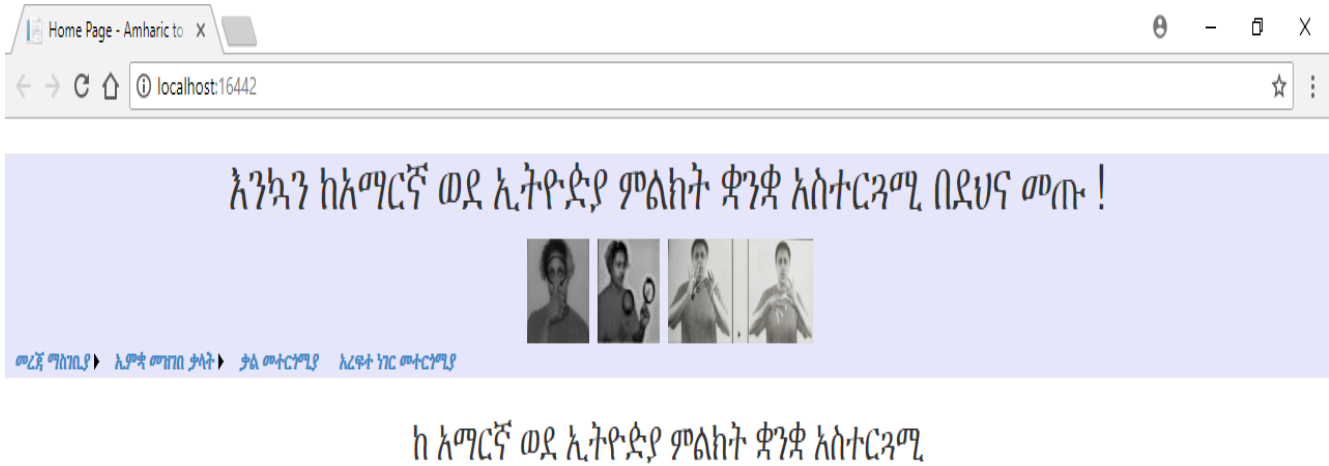
The database is SQL 2008 Database which incorporates adopted dictionary, synonym words, and words with extracted features, word structures, and sentence structures.

5.2.1 User Interface Layer

The User Interface component is responsible for accepting Amharic text, trimming preceding and exceeding white spaces and pass it to the Business logic component. The result of the translation is also displayed as a sequence of videos. Figure 5.1 shows the home page of the user interface. In general it has four menus which are:

- Adopted dictionary (አምቋ መዝገበ ቃላት) with two sub menus find word (ቃል መፈለጊያ) and Find Finger spelling (ፊንገር መፈለጊያ) from the adopted dictionary
- Option for word translation (ቃል መተርጎሚያ)

- Option for sentence translation (አረፍተ ነገር መተርጎሚያ)
- Option for data entry (መረጃ ማስገቢያ) which enables to add more examples and words to the system.



© - AMETHSLT

Figure 5.1: Home page

Every menu link traverse in to the corresponding page which holds different controls like labels, textboxes, dropdown lists, buttons to handle event driven procedures and a video control to display the output video clips.

5.2.2 Business Logic Layer

Business logic layer is a layer which is responsible to handle every logic behind translation. Every back end components discussed in design phase are implemented here except the data access module which is handled by data access layer in order to get a translation result as shown in the examples. Details of translations are attached at [Annex G](#)

Example1. Figure 5.2 shows screen shoots of Amharic letter to EthSL figure spelling translation



Figure 5.2: Amharic letter to EthSL figure spelling translation

Example 2. Figure 5.3 shows screen shots of a word translation for መመገብ /memegebi / መመገብ /memegebi/ is a synonym word for መብላት /mēbilāti/ to mean to eat. When we try to translate መመገብ /memegebi/ the system cannot find the word in the adopted dictionary; instead መመገብ /memegebi/ is found in the knowledge base as a synonym word for መብላት /mēbilāti/. Hence, the system displays the sign for መብላት /mēbilāti/



Figure 5.3: Word translation example for መመገብ and መብላት

5.2.3 Data Access Layer

Database First Approach is used to develop the prototype and create the entity data model from an existing database. ADO.NET Entity Data Model is added into this layer which eases creating model codes (classes, properties, DbContext etc.) from the database and those codes become the link between the database and the controller.

5.2.4 The Database

The database is denoted using SQL server 2008. Basic information of Amharic texts like letters, words and features of the words are stored as a sequence of Amharic characters (Unicode characters) using visual geez. Latin1_General_100 windows collation and binary sorting is set to be able to search and access Amharic characters and text.

The database encompasses tables of

- Amharic words and their EthSL equivalent sign's URL
- Synonym words which are represented by the same sign with a word found in the adopted dictionary
- Words with their features which is the result of morphological analysis performed by Amharic language experts
- Word and sentence structures resulted from the analysis of the pairs of Amharic to EthSL example sentences collected for the study.
- Alignment of Amharic text structure to its EthSL equivalent structure at word and sentence level
- Short form and long form representation of Amharic words

To enrich the database, experts can use the data entry interface. The system checks if the data to be newly added already exists or not; so, it control duplication. Sample data entry interface is shown in Figure 5.4

የቃላት ትንታኔ ማስገቢያ

ቃል አስገባ የቃል ክፍል ምረጥ

ስርወ ቃሉን አስገባ መስተዋድድ (Preposition)

የባለቤት ተውላጠ ስም (Subjective pronoun) ተገኝቷል ተውላጠ ስም (Objective Pronoun)

የባለቤትነት አመልካች (Possessive Pronoun) ጊዜ አመልካች (Tense)

አሉታ ያመለክታል? (Is negated?) Yes NO

ብዙ ቁጥር ያመለክታል? (Is Plural?) Yes NO

Figure 5.4: Data entry interface

5.3 Testing and Evaluation

The prototype developed was evaluated in order to justify whether the objective of the study is met and its expected functionalities are achieved. The evaluation also assures whether the design is achievable and convenient.

Human assessment results of Adequacy and Fluency are used for evaluation of the prototype. Adequacy tests the amount of meaning expressed and conveyed in the translation whereas Fluency tests the well formedness and fluency of the translation. In this study adequacy and fluency scales are adopted as shown in Table 5.1 and Table 5.2.

Table 5.1: Adopted Adequacy Scale

Adequacy		
5	All meaning	ሙሉ በሙሉ
4	Most meaning	በአብዛኛው
3	Much meaning	በመጠኑ
2	Little meaning	በጥቂቱ
1	none	ምንም አልተላለፍም

Table 5.2: Adopted Fluency Scale

Fluency		
5	Flawless	እንክን የለሽ
4	Good	ጥሩ
3	Acceptable	ተቀባይነት ያለው
2	Flawed	ጉድለት ያለበት
1	Not acceptable	ተቀባይነት የሌለው

The prototype was evaluated by test cases of five categories at word level and sentence level. These are finger spellings, words from the adopted dictionary, words out of the adopted dictionary, numbers and sentences.

Three groups of testers were selected for the test process by considering societies who are frequent or permanent users of EthSL for communication or due to their profession. Thus, the testers are nominated using convenience sampling. This method is a type of non-probability or non-random sampling where members of the target population that meet certain practical criteria such as easy accessibility, geographical proximity, availability at a given time, or the willingness to participate are included for the purpose of the study . It is also referred to the researching subjects of the population that are easily accessible to the researcher [49].

The test groups are from different background in terms of their expertise, age and exposure.

- The first group (G1) embraces a group of 15 Grade 4-8 Deaf students, 3 students from each grade level.
- The second group (G2) embraces three hearing signers, three native signers, three post lingual signers whose age is above 20. The third group (G3) embraces one EthSL expert, two interpreters and three EthSL teachers.

The test results were collected using result collection and result summery forms attached in ([Annex H](#)).

Testers from G2 and G3 performed the test at word level and sentence level after forming test cases by their own. To create test cases, they selected words from the morphologically analyzed ones and constructed sentences from them. Finally, they summarized the result in the result summery form.

For group G1 which is composed of Deaf students, words are selected and sentences are constructed by the help of their teachers. Similarly, the test was taken place and the results were summarized with the help of the teachers.

The test result for all the three group of testers are shown in Table 5-3, Table 5-4 and Table 5-5 which are the English equivalents of the Amharic result collection and summery forms attached in [Annex H](#).

Table 5.3: Group 1 Test Result

Test cases category	No of test cases	Adequacy							Fluency					Avg.	Rank
		5	4	3	2	1	Avg .	Rank	5	4	3	2	1		
Finger spellings	30	28	0	1	1	0	4.83	5	24	0	2	4	0	4.47	4
Words from the Adopted Dictionary	30	24	3	2	1	0	4.67	5	17	8	4	1	0	4.37	4
Numbers	30	27	0	2	1	0	4.77	5	12	15	2	1	0	4.27	4
Words out of the Adopted Dictionary	30	14	10	5	1	0	4.23	4	7	9	6	7	1	3.47	3
Sentences	15	2	5	5	3	0	3.40	3	4	4	3	3	1	3.47	3

Table 5.4: Group 2 test result

Test cases category	No of test cases	Adequacy							Fluency					Avg .	Rank
		5	4	3	2	1	Avg .	Rank	5	4	3	2	1		
Finger spellings	27	27	0	0	0	0	5.00	5	7	11	7	2	0	3.85	4
Words from the Adopted Dictionary	27	27	0	0	0	0	5.00	5	5	12	8	2	0	3.74	4
Numbers	27	27	0	0	0	0	5.00	5	13	6	8	0	0	4.19	4
Words out of the Adopted Dictionary	27	19	8	0	0	0	4.70	5	4	9	9	5	0	3.44	3
Sentences	27	4	7	10	3	3	3.22	3	3	7	10	4	3	3.11	3

Table 5.5: Group 3 test result

Test cases category	No of test cases	Adequacy							Fluency					Avg .	Rank
		5	4	3	2	1	Avg .	Rank	5	4	3	2	1		
Finger spellings	18	18	0	0	0	0	5.00	5	9	6	3	0	0	4.33	4
Words From the adopted Dictionary	18	18	0	0	0	0	5.00	5	8	6	4	0	0	4.22	4
Numbers	18	18	0	0	0	0	5.00	5	10	6	2	0	0	4.44	4
Words out of the adopted Dictionary	18	10	8	0	0	0	4.56	5	7	6	4	1	0	4.06	4
Sentences	18	3	7	5	1	2	3.44	3	1	6	7	1	3	3.06	3

The evaluation result of the three test groups G1, G2 and G3 for finger spellings, numbers and words from the adopted dictionary were adequately understood with approximated result of 5 representing ‘All meaning’ with good fluency value 4. Thus, the average value still lies in the same category even if it varies by small amount.

The test result of ‘words out of the adopted dictionary’ test cases by G1, G2 and G3 for (adequacy, fluency) pair is found to be in the rank (4, 3), (5, 3) and (5, 4) respectively. G3 has scored better result relative to G1 & G2. The reason is G3 consists of interpreters, experts and teachers who follow the dictionary and written knowledge properly which is somewhat in line with the background knowledge.

The rank for both adequacy and fluency of evaluating sentences is found to be 3 for all test groups. This indicates much meaning has been adequately understood and the fluency is to the acceptable level.

The other type of test shown in Table 5.6 is conducted to assure the usability and accessibility of the developed prototype with the questions and result mentioned in table. The questions are attached in [Annex H](#) in Amharic format. Ranking was scaled in five categories as Poor = 1, Fair = 2, Good = 3, Very good = 4 and Excellent = 5.

Table 5.6: User Acceptance Test

Questions	No of testers	5	4	3	2	1
The prototype is usable for the Deaf if it is fully implemented	30	30				
The prototype is easy to use	30	11	17	2		

The usability of the prototype was ranked as excellent and the test teams have were aspired while mentioning their future hope by enriching the system for full deployment. On the other hand, ease of use was ranked as ‘very good’ which is a very encouraging result suggesting the prototype is suitable for translating Amharic text to EthSL.

5.4 Results Discussion

The developed prototype is tested with individuals of different background and expertise being in categories. The result has shown that it is a promising study for the purpose it is intended to. An excellent system usability level is achieved with full score since the sign videos are more understandable than their corresponding pictures with explanations found in books.

The result for the three sets of test cases (finger spelling, number and words from the adopted dictionary) in terms of the level of adequacy and fluency by all test teams fall in the same rank of 'All meaning' and 'Good fluency' level respectively. This is because, the nature of the test sets are single letter or word or number which doesn't have grammar issue. Whereas the result of the rest test sets (words out of the adopted dictionary and sentences) has been influenced by grammatical structure clearly revealed while each team was debating to evaluate the fluency.

Generally the reason for lower fluency result caused by the factors listed below as per the test teams' comment.

- Quality of the video recording was affected since the recording was not done by professionals. Moreover it was recorded by webcam.
- The noticeable time gap between successive playing videos of a single user input created confusion for the testers and affected the fluency.
- The test teams understood adequately some EthSL signs which change direction based on subject and object of the Amharic text; but, their articulation was not in the proper way leading to value deduction on the fluency.
- There are cases where the signs of some numbers are represented differently from the normal one, informal but common especially for the tens digit.

For example 220 two+ hundreds+ twenty whereas 225 two +hundreds+ two+ five

- The other fluency issue specially arose by G2 evaluators was that the videos should be taken by a professional with a single interpreter (signer) and within the same environment considering the background and even the clothes of the signer. This issue arose videos of signs are recorded in different environments with different signers.

The signs are then stored at word level and the signs will get displayed depending on the input Amharic text.

Additional observations

- During evaluation with G1, commonly used words like egg and shoes are not recognized in the first test trial. Then after discussing with their teachers, the signs were changed according to the previous dictionary “ሀ ማፅሀፍ” where some signs are borrowed from ASL. Doing so made G1 team identify the signs. This indicates that there are some words still use signs from the previous dictionary. Sample for such words are attached in [Annex I](#).
- The other observation is that there are modifications on some signs and they commonly communicate with the modified one instead of the one from the dictionary. For example:
 - Using Amharic hand configuration mixed with ASL movement is used for signing days of the week.
 - Two handed signs are modified in a way to be described by one hand signs only like to show sister and brother.
- The sign to show past actions in EthSL has the sign of ማጠቃለያ means finish and ማጠቃለያ is articulated differently in the previous dictionary “ሀ ማፅሀፍ” and the new dictionary which affects the fluency result given by different groups.
- Words with and without attached prepositions like ለ /lɛ/ (for), ከ /kɛ/ (from), በ /bɛ/ (by or with) and የ /jɛ/ (of) have similar EthSL signs which indicates that these prepositions do not have an effect during translation.
- In case of sentences similar pronouns might exist in more than one word. In this case, the repeated pronouns (Subjective and Objective) are removed leaving only one during alignment.

For example እነሱ አይጫወቱም → እነሱ ጭውጉ እነሱ አይደለም

Suggestions from EthSL Language users

After completing the formal test, there was a casual discussion with test teams for their suggestion and feeling about the study. They provided valuable ideas of which the major ones are listed.

- Developing the adopted dictionary to a full-fledged level contributes a lot to the Deaf society. This is because the video signs are easily understandable than the previous dictionaries. In the ENAD dictionary, signs are described with lots of arrows, directions and other symbols whereas in the previous dictionary “ሀ ሞፅሀፍ” descriptions are given by Amharic text which is not easy to realize.
- There are Amharic words having two or more signs registered in different dictionaries or having modified signs or having locally created signs from specific culture and religion aspects. Enriching the adopted dictionary in this study to a higher standardized level by incorporating all situations will help to capture EthSL to a better extent.
- The video signs need to occupy most part of the screen to be recognized simply.
- In the recorded video signs, the signers have to include mouth movement to enable lip reading for improved understanding.

Chapter Six: Conclusion and Recommendations

This Chapter summarizes the overall aim of the study and the implemented prototype features. It also mentions about evaluation result and recommendations for future work.

6.1 Conclusion

In this study an attempt is made to design and implement Amharic Text to Ethiopian Sign Language Machine Translation system. Investigation was performed into the existing translation approaches and an EBMT approach has been chosen. EBMT was selected due to lack of linguistic studies on EthSL grammar and insufficient parallel Amharic text and EthSL examples collected. As a result this study came up with a design and a prototype of Example based machine translation system which analyzes Amharic text and synthesizes equivalent representation in EthSL.

There have been a number of researches conducted on translation of Amharic text to EthSL. Even if the researches tried to develop translators using different approaches that can translate Amharic text at word, number and even at sentence level, there is still a gap to address the requirement. Hence, this study is initiated with a general objective of designing an Example Based Amharic Text to Ethiopian Sign Language Machine Translation System.

The study is conducted to explore whether Example based MT approach is applicable for translating Amharic text to EthSL with intelligible and natural signs. The design has a word feature analysis component to analyze and get the important features of the input text. It also has a component with matching, alignment and recombination phase which performs the translation based on the knowledge acquired from example cases so that the resulting output sign is displayed as a sequence of videos. The database encompasses about morphologically analyzed Amharic words, distinct word structures and sentence structures with their EthSL equivalent. Amharic to EthSL dictionary of about 245 EthSL Finger spellings (fidels), and about 280 words is also included.

The evaluation result for finger spellings, numbers and words from the adopted dictionary were adequately understood with approximated result of 5 representing 'All meaning' with 'Good' fluency value of 4. Whereas words out of the adopted dictionary were adequately understood with approximate result of 4 representing 'Most meaning' with 'Acceptable

'fluency value of 3. Adequacy and fluency of sentences is found to be 3 indicating 'Much meaning' has been adequately understood and the fluency is to the 'Acceptable' level. The usability of the prototype was ranked as 'Excellent' with 'very good' ease of use.

The overall result is found to be encouraging and promising as per the evaluation result and feedback gathered. Accordingly, Deaf members of the test teams have been inspired with the capability of the prototype for future use and they expressed their willingness to contribute for full deployment of the system.

From the general observation, experts and individuals involved in the study recommended that developing a full functional and efficient machine translator is a team effort that needs a lot of coordination. Those to be coordinated can be pre and post lingual Deaf (who are the main stakeholders), Amharic and EthSL language experts (linguists), system analyzers, designers, developers, interpreters, video recorders and editors. In addition, ENAD, MoE, and Universities especially AAU have to give attention to the problem at hand and work in harmony so that a standardized EthSL will be in use. Besides which creating awareness that EthSL is different from Signed Amharic is the other role to be played by ENAD and other stakeholders.

6.2 Contribution of the Study

The main contributions of the study are:

- An Amharic to EthSL translator has been designed by incorporating morphologically analyzed Amharic's words.
- An adopted digital dictionary of Amharic to EthSL has been developed by integrating Amharic root word forms
- Important features required for translation of Amharic text to EthSL has been identified.
- Unique structures have been generated and incorporated in the system for Amharic – EthSL at word and sentence level.
- Intelligible and natural signs are used to display the output by means of sequential sign video clips rather than avatars.

6.3 Recommendations

An effort has been exerted in this study to develop EBMT based Amharic to EthSL translation system which can be upgraded by adding more examples. Additional features can be added or the existing components of the translator can be modified in order to increase the performance of the translator. Even if this study has dealt with many issues, there are a number of potential future works for the attainment of a full-fledged system. Therefore, we recommend the following future works:

- Adopting all available EthSL Dictionaries
- Building a larger size parallel corpora of Amharic to EthSL sentences
- Building an Amharic morphological analyzer specially for sign language synthesis considering critical features
- Automating the process of aligning words in an Amharic sentence into their EthSL equivalents for collected examples
- Adding disambiguation module to facilitate conceptual translation of Amharic words that have different signs in EthSL based on the context.
- Preparing automatic evaluation metrics applicable for EthSL
- Developing bi-directional translator from EthSL to Amharic text and vice versa
- Incorporating Amharic speech recognizer to translate EthSL by including finger spellings that enables writing on the air to represent Amharic words having no corresponding EthSL sign.
- Incorporates system notifications with EthSL apart from Amharic text to be understood by system users, the hearing impaired.
- Integrating the system with a signing avatar

Finally, as a mother of a pre lingual Deaf, I wish to see a full-fledged bi-directional Amharic to EthSL translation system implemented and used in public organizations so that pre lingual Deaf can access public services easily. I am confident that the help of my advisor and others will continue in building a wider network to support the aforementioned future works. I believe this study will attract more attention from the local research centers and center of disabilities opening the door for more collaboration and funding support to the achievement of full-fledged Translator.

References

- [1] Wikipedia, October 23 2015. [Online]. Available: <https://en.wikipedia.org/wiki/Communication>.
- [2] E. N. A. o. t. Deaf, Ethiopian Sign language Dictionary, 2008.
- [3] W. H. Organization, October Web Accessed August 19,2015. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs300/en>.
- [4] "WHO media center," WHO, [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs300/en/>. [Accessed 18 May 2017].
- [5] M. A. S. Nodoushan, "Language and Literacy Development in Prelingually-Deaf Children," Iran.
- [6] Rafi Shemesh, "Hearing Impairment: Definitions, Assessment and Management," [Online]. Available: <http://cirrie.buffalo.edu/encyclopedia/en/article/272>.
- [7] T. J. a. A. Schembri, An Introduction to Sign Language Linguistics, New York: Cambridge University Press, 2006.
- [8] B. B. Æ. K.-F. K. Ulrich von Agris Æ Jo`rg Zieren Æ Ulrich Canzler Æ, "Recent developments in visual sign language recognition," 2007.
- [9] በ. ሚ. ሙ. የ. በ. በ. ጽ. ት. በ. ሙ. የ. ማ. በ. የተዘጋጀ, የአማርኛ የምልክት ቋንቋ መስማትና መናገር ለተሳናቸው 1ኛ መጽሐፍ, A.A, 1968.
- [10] E. E. Schneider, September 4,2015. [Online]. Available: <http://www.lesstutor.com/eesASLIntro.html>.
- [11] J. Hutchins, October 23 2015. [Online]. Available: <http://ourworld.compuserve.com/homepages/WJHutchins>.
- [12] S. Morrissey, "Data-Driven Machine Translation for Sign Languages," Dublin , 2008.

- [13] JENNIFER HARRIS & CLAIRE BAMFORD, "The Uphill Struggle: services for Deaf and hard of hearing people," *Disability & Society*, Vols. Vol.16, No.7, no. issues of equality, participation and access, p. pp.969–979, 2001.
- [14] R. Girmay, "The Assessment of Ethiopian Sign Language Educational Interpreters among Teachers In five Regional States of Ethiopia," Addis Ababa, 2012.
- [15] T. J. A. A. Schembri, *Australian Sign Language (Auslan) An Introduction to Sign Language Linguistics*, New York: Cambridge University Press, 2007.
- [16] D. F. Wolde, "Machine Translation System for Amharic Text to Ethiopian Sign Language," Addis Ababa, 2011.
- [17] D. Z. Zeleke, "Amharic Sentence to Ethiopian Sign Language Translator," Addis Ababa, 2014.
- [18] ዱ. ካሱ, የኢትዮጵያ የምልክት ማስተማሪያ መመሪያ, አዲስ አበባ ዩኒቨርሲቲ ፕሬስ, 2015.
- [19] A. A. ARGAW, "AUTOMATIC SENTENCE PARSING FOR AMHARIC TEXT," Addis Ababa, 2002.
- [20] P. KOEHN, *Statistical Machine Translation*, Edinburgh: CAMBRIDGE UNIVERSITY PRESS, 2010.
- [21] C. D. Manning, *Foundations of Statistical Natural Language Processing*, London, England: The MIT Press, 2000.
- [22] B. G. Gebre, "Part of Speech Tagging for Amharic," UNITED KINGDOM, 2010.
- [23] M. A. a. Y. Assabie, "Development of Amharic Morphological Analyzer Using Memory-Based Learning," [Online].
- [24] "Amharic," wikipedia, [Online]. Available: <https://en.wikipedia.org/wiki/Amharic>. [Accessed 23 September 2015].
- [25] D. Assefa, "Amharic Speech Training For The Deaf," 2006 .

- [26] S. H. Yesuf, "AMHARIC WORD SENSE DISAMBIGUATION USING WORDNET," AAU, Addis Ababa, March 2015.
- [27] M. S. a. T. K. Tessema Mindaye, "The Need for Amharic WordNet," in *5th International Conference of the Global WordNet Association*, Mumbai, 2010.
- [28] "Facts," WHO, [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs300/en>. [Accessed 18 May 2017].
- [29] "Human Rights," WFD, [Online]. Available: <https://wfdeaf.org/human-rights/> . [Accessed 18 May 2017].
- [30] "Deaf as a Linguistic and Cultural Group," WFD, [Online]. Available: <https://wfdeaf.org/human-rights/crpd/deaf-as-a-linguistic-and-cultural-group>. [Accessed 18 May 2017].
- [31] "Sign Language," WFD, [Online]. Available: <https://wfdeaf.org/human-rights/crpd/sign-language/> . [Accessed 18 May 2017].
- [32] A. Almohimeed, "Arabic Text to Arabic Sign Language Example-Based Translation System," Southampton, 2012.
- [33] A. I. Hilger, "Production fluency in spoken language users acquiring a sign language as an L2 in adulthood," Illinois, 2013.
- [34] A. C. a. R. E. Ladner, "Hearing Impairments," Washington, 2008.
- [35] E. ASEFA, A Hypermedia Learning Tool for Ethiopian Sign Language, Addis Ababa, 2005.
- [36] ጥ. ኢ. ኢ. ድርጅት, የኢትዮጵያ ምልክት ቋንቋ የተማሪ ሙጽሐፍ 1-8 ክፍል, የኢፌዲሪ ትምህርት ሚኒስቴር.
- [37] L. M. W. N. Benson Kituku*1, "A Review on Machine Translation Approaches," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, Vols. Vol. 17, No. 1., pp. 182 - 190 , 2016.

- [38] A. P. J., "Machine Translation Approaches and Survey for Indian Languages," *Computational Linguistics and Chinese Language Processing*, Vols. Vol. 18, No. 1, 2013.
- [39] M. D. Okpor, "MT Approaches: Issues and Challenges," *IJCSI International Journal of Computer Science Issues*, Vols. Vol. 11, No 2,, no. Issue 5,, 2014.
- [40] R. Kibble, Introduction to natural language processing, London, 2013.
- [41] J. HUTCHINS, "Recent advances in example-based machine translation," *Text, Speech and Language Technology*, vol. vol. 21, 2003.
- [42] H. SOMERS, Machine Translation, Netherlands.: Kluwer Academic Publishers, 1999.
- [43] H. DOĞAN, "EXAMPLE BASED MACHINE TRANSLATION WITH TYPE ASSOCIATED TRANSLATION EXAMPLES," Bilkent, 2007.
- [44] J. HUTCHINS, "Towards a definition of example-based machine translation," [Online]. [Accessed may 2016].
- [45] S. M. a. A. Way, "Developing a Sign Language Machine Translation," in *Assistive Technologies for People with Vision & Hearing Impairments*, 2007.
- [46] M. Tesfaye, "A MT Approach to Translate Amharic Text to Ethiopian Sign Language," AAU, Addis Ababa , 2010.
- [47] M. T. Erenso, "Automatic translation of Amharic text to Ethiopian sign language,," Addis Ababa, 2010.
- [48] T. M. Mengistu, "DESIGN AND IMPLEMENTATION OF AMHARIC SEARCH ENGINE," AAU, Adiss Ababa, July, 2007.
- [49] S. A. M. R. S. A. Ilker Etikan, "Comparison of Convenience Sampling and Purposive Sampling," *American Journal of Theoretical and Applied Statistics* , Vols. Vol. 5, No. 1,, pp. pp. 1-4, 2016.
- [50] W. F. o. t. Deaf, October 2015. [Online]. Available: <http://wfdeaf.org/whoarewe>.

- [51] J. B. a. H. N. Daniel Stein, "Morpho-Syntax Based Statistical Methods for Automatic Sign Language Translation," Aachen, Germany, Web Accessed August 19, 2015.
- [52] J. Hutchins, "Current commercial machine translation systems and computer-based translation tools: system types and their uses," 2005. [Online]. Available: <http://ourworld.compuserve.com/homepages/WJHutchins>. [Accessed 2016].

Annexes

Annex A: Amharic words that are represented by two EthSL signs

For Example the following signs shows መጠባት /met'et'äti/ to drink nonalcoholic drink and alcoholic drinks respectively.


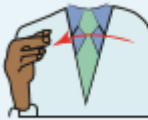
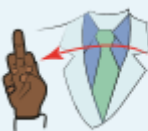






መፈለግ /mefelēgi/ which has two signs to mean 'search' and 'need' requires

762. መፈለግ¹ (ከምኞት ጋር ለተያያዘ ነገር ብቻ)
to seek, desire, want, need (It is for something that is related to desire/wish only.)






763. መፈለግ²: ፍለጋ (ሰው/ዕቃን ከመፈለግ ጋር ለተያያዘ ነገር ብቻ)
to search, to look for (It is for in search of a person/ thing only.)






Annex B: Days of the week

	እሁድ		ማክሰኛ		ሐሙስ		ቅዳሜ
	ሰኞ		ረቡዕ		ዓርብ		

Annex C: EthSL Number signs

	0 ዜሮ
	1 አንድ
	2 ሁለት
	3 ሦስት
	4 አራት
	5 አምስት
	6 ስድስት
	7 ሰባት
	8 ስምንት
	9 ዘጠኝ
	10 አሥር

	10	አሥር
	20	ሃያ
	30	ስላሳ
	40	አርባ
	50	ሃምሳ

	60	ሥልሳ
	70	ስባ
	80	ስምንያ
	90	ዘጠኛ
	100	መቶ

Annex D: EthSL Finger Spelling

ታላቅ ትላትቦ

ታላቅ ትላትቦ ለመገኘት ለሁሉም ሰዎች ማሳሰቢያ

ታላቅ ትላትቦ ለመገኘት ለሁሉም ሰዎች ማሳሰቢያ

ታላቅ ትላትቦ ለመገኘት ለሁሉም ሰዎች ማሳሰቢያ

Annex E: Normalization Algorithms Adopted From Tessema Mindaye

The algorithm for replacing the alphabets (fidels) is shown in *Figure 4.5* below.

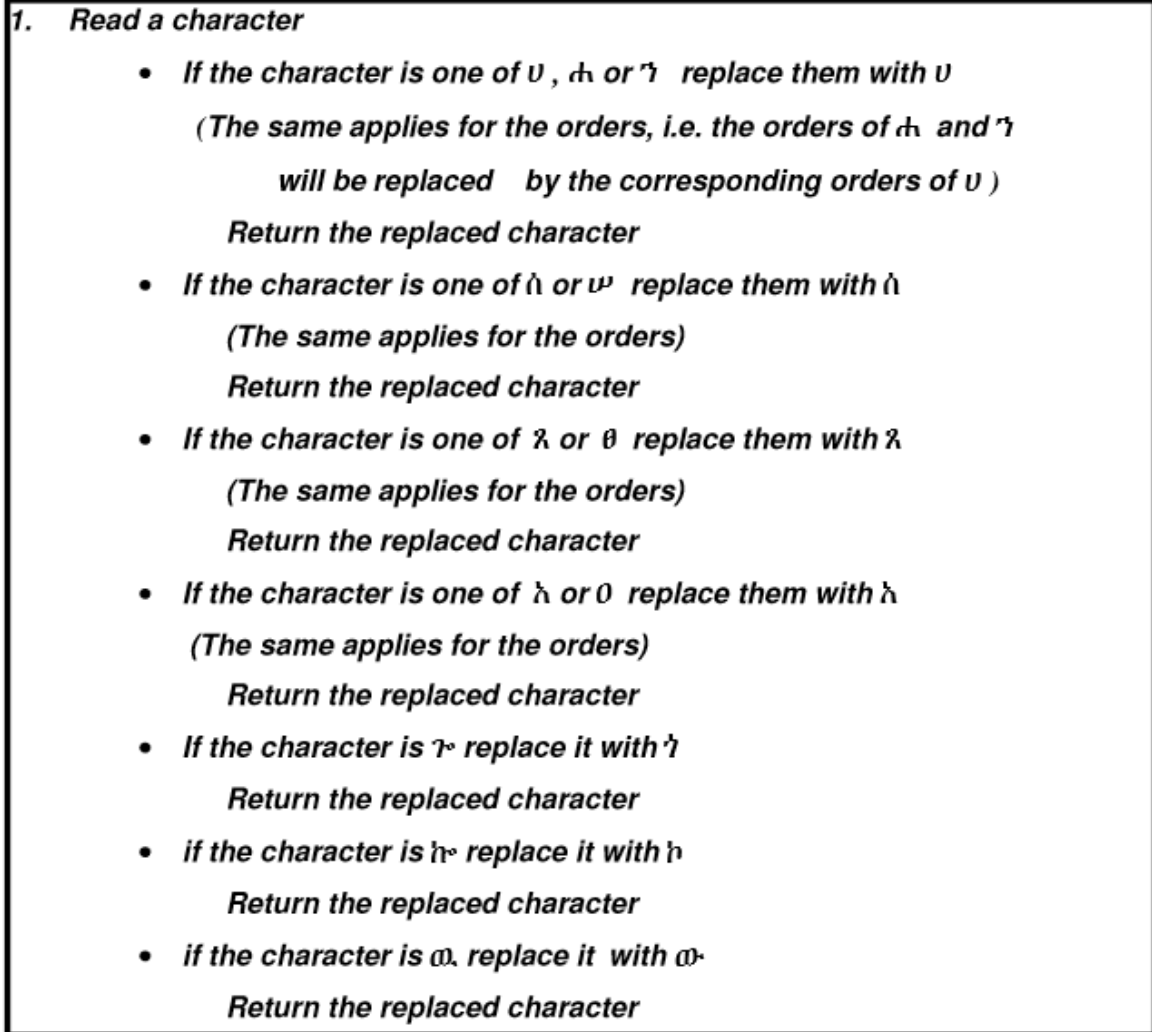


Figure 4-5 Algorithm of Character Replacer

```
1. Read a character before "/" or "."
2. Search the storage for that character
   If found
       Return the corresponding expanded word
   Else
       Return the original word
```

Figure 4-6 Algorithm of Word Expander

Annex F: Example of Morphologically analyzed Amharic Words

Amharic Word	Amharic Word Root	Amharic Word POS	Attached Preposition	is it Negated	is it Plural	Subjective Pronoun	Objective Pronoun	possessive Pronoun	Amharic Word Tense
ሊበሉ	ብልእ	ግስ		0	0	እነሱ			ትንቢት
ልበላ	ብልእ	ግስ		0	0	እኔ			ትንቢት
ልትበላ	ብልእ	ግስ		0	0	እሷ			ትንቢት
መብላት	ብልእ	ስም		0	0				
በሉ	ብልእ	ግስ		0	0	እነሱ			ሀላፊ
በላ	ብልእ	ግስ		0	0	እሱ			ሀላፊ
በላሁ	ብልእ	ግስ		0	0	እኔ			ሀላፊ
በላህ	ብልእ	ግስ		0	0	አንተ			ሀላፊ
በላች	ብልእ	ግስ		0	0	እስዋ			ሀላፊ
በላን	ብልእ	ግስ		0	0	እኛ			ሀላፊ
አልበላሁም	ብልእ	ግስ		1	0	እኔ			
አበላላሁ	ብልእ	ግስ		0	0	እኔ			
እየበላ	ብልእ	ግስ		0	0	እሱ			አሁን
እየበላሁ	ብልእ	ግስ		0	0	እኔ			አሁን
የበላሁት	ብልእ	ግስ	የ	0	0	አኔ			
መስጠት	ስጥእ	ስም		0	0				
ሰጠ	ስጥእ	ግስ		0	0	እሱ			
ሰጠች	ስጥእ	ግስ		0	0	እሷ			
ሰጠኝ	ስጥእ	ግስ		0	0	እሱ	እኔ		
ሰጡ	ስጥእ	ግስ		0	0	እነሱ			
አልሰጠሁትም	ስጥእ	ግስ		1	0	እኔ	እሱ		
መጽሀፍ	መጽሀፍ	ስም		0	0				
መጽሀፎች	መጽሀፍ	ስም		0	1				
መጫወት	ጫወት	ስም		0	0				
ሲጫወቱ	ጫወት	ግስ		0	0	እነሱ			
ሲጫወት	ጫወት	ግስ		0	0	እሱ			
ስንጫወት	ጫወት	ግስ		0	0	እኛ			
ተጫወተ	ጫወት	ግስ		0	0	እሱ			ሀላፊ
ተጫወተች	ጫወት	ግስ		0	0	አሷ			ሀላፊ
ተጫወቱ	ጫወት	ግስ		0	0	እነሱ			ሀላፊ
ተጫወትን	ጫወት	ግስ		0	0	እኛ			ሀላፊ
ትጫወታለች	ጫወት	ግስ		0	0	አሷ			
አልተጫወተም	ጫወት	ግስ		1	0	እሱ			

Annex G: Details of translations

i. Adopted dictionary (ኢምቋ መዝገበ ቃላት)

This option enables a user to get EthSL equivalent of an Amharic letter and word from the adopted dictionary.

For Amharic letter translation, the submenu “ፊደል መፈለጊያ” accepts an Amharic letter and displays its equivalent EthSL finger spelling as shown in the figure below



Amharic letter to EthSL figure spelling translation

For Amharic word translation, the sub menu “ቃል መፈለጊያ” accepts an Amharic word and performs the following:

- It checks for the existence of relatives of characters ሐ, ኀ, ሠ, ዐ, ፀ and normalizes them if it finds any.
- The normalized word is searched from the adopted dictionary.
 - If it exists, it displays the corresponding sign clip.
 - If not, it searches for the existence of synonyms. If it finds a synonym, it displays the corresponding synonym sign.

- If it doesn't have a synonym, it will display a message informing the user to use the word translation option as shown in Figure 5.3.



Message display for missing word search

Example መመገብ /memegēbi/ is a synonym word for መብላት /mēbiläti/ to mean to eat. When we try to translate መመገብ /memegēbi/ the system cannot find the word in the adopted dictionary; instead መመገብ /memegēbi/ is found in the knowledge base as a synonym word for መብላት /mēbiläti/. Hence, the system displays the sign for መብላት /mēbiläti/ as shown in the Figure 5.4.

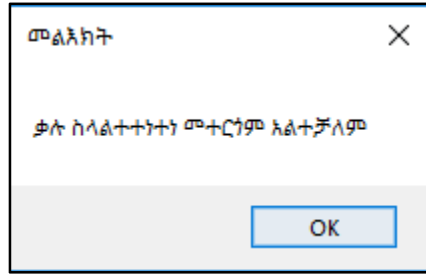


Word translation example for መመገብ and መብላት

ii. Word translator (ቃል መተርጎሚያ)

This option is used to perform Amharic word to EthSL translation; it is also capable of translating numbers written in digits by treating them as a word. Unlike the adopted dictionary word translation submenu, it can translate words whose roots found in the adopted dictionary too. Thus, the translation process follows the following steps.

- It first performs normalization.
 - It checks for the existence of relatives of characters ሐ, ኀ, ሠ, ዐ, ቀ and normalizes them if it finds any.
 - It checks whether the word is in short hand form and replaces it with its expanded form.
- The normalized word is searched among morphologically analyzed words in the knowledge base.
 - If it exists, its features will be extracted and similarity analysis is performed to identify the best matching word structure.
 - EthSL equivalent word structure of the selected Amharic word structure is identified.
 - Alignments of the Amharic word features will be performed based on the identified EthSL word structure
 - EthSL equivalent of the aligned structure is recombined and displayed to the user as sequence of sign videos
 - But if the word is not found in the knowledge base, it displays message informing the word could not be translated as shown in Figure 6.6. In addition, it is recorded in the system as an attempted word with frequency for search attempt. This information is used by experts to know if there are frequently requested words and hence to add them to the system.



Message display for missing word

iii. Sentence Translator (አረፍተ ነገር መተርጎሚያ)

This option works for sentences translation as per the steps described below.

- Text accepted from the user will be tokenized with punctuation and white spaces used as delimiters. The result is sequence of words.

Word level processing is performed initially

- Each word will then be processed in a similar way to **Word translator (ቃል መተርጎሚያ)** menu as mentioned in (ii) except the final alignment and recombination step. It continues to next step if all words can be found in the system; otherwise it displays message informing the user that translation cannot be performed.
- EBMT component computation
 - Candidate sentence structures are identified
 - Similarity analysis is performed to identify the best matching sentence structure.
 - EthSL equivalent sentence structure of the selected Amharic sentence structure is identified.
 - Alignments of the Amharic words with their features will be performed based on the identified EthSL sentence structure
 - EthSL equivalent of the aligned structure is recombined and displayed to the user as sequence of sign videos

ከ አማርኛ ወደ ኢትዮጵያ ምልክት ቋንቋ መተርጎሚያ ሲስተም ላይ

ለተደረገው ሙከራ አስተያየት መሰብሰቢያ

የገመገሙትን መተርጎሚያ የግምገማ ውጤት ማጠቃለያ ከዚህ በታች ባለው ቅፅ ላይ እንዲሞሉልን በትኩረት እንጠይቃለን ።

የመስማት ችሎታ :- መስማት የተሳነው መስማት የሚችል ኢምቋ ተጠቃሚ

1. በመረጃ መሰብሰቢያው ቅፅ መሰረት መልእክት በማስተላለፍ(Adequacy) በየደረጃው የተለዩ ፅሁፎች ብዛት ማጥቃለያ ሰንጠረዥ

ሙሉ በሙሉ(5) በአብዛኛው (4) በመጠኑ (3) በጥቂቱ(2) ምንም እልተላለፍም(1)

የሙከራው አይነት	ጠቅላላ ብዛት	የተጠቀሰው ደረጃ የተሰጣቸው ፅሁፎች				
		5	4	3	2	1
ለተሞክሩ ቁጥሮች						
ከመዝገበ ቃላት ለተሞክሩ ቃላት						
ከመዝገበ ቃላት ውጭ ለተሞክሩ ቃላት						
ለተሞክሩ አረፈተ ነገሮች						

2. በመረጃ መሰብሰቢያው ቅፅ መሰረት በተገቢው መንገድ(Fluency) ለመተላለፋቸው በተሰጣቸው መስፍርት መሰረት የተለዩ ፅሁፎች ብዛት ማጥቃለያ ሰንጠረዥ

እንከን የለሽ (5) ጥሩ(4) ተቀባይነት ያለው(3) ጉድለት ያለበት (2) ተቀባይነት የሌለው (1)

የሙከራው አይነት	ጠቅላላ ብዛት	የተጠቀሰው ደረጃ የተሰጣቸው ፅሁፎች ብዛት				
		5	4	3	2	1
ለተሞክሩ ፊደላት						
ለተሞክሩ ቁጥሮች						
ከመዝገበ ቃላት ለተሞክሩ ቃላት						
ከመዝገበ ቃላት ውጭ ለተሞክሩ ቃላት						
ለተሞክሩ አረፈተ ነገሮች						

3. ሲስተሙ አገለግሎት ላይ ቢውል መስማት ለተሳናቸው ሊኖረው የሚችለው ጠቀሜታ በጣም ከፍተኛ ከፍተኛ መካከለኛ ዝቅተኛ ጠቀሜታ አይኖረውም

4. ሲስተሙ ለአጠቃቀም ያለው ምቹነት(ለአጠቃቀም ቀላል መሆን) በጣም ከፍተኛ ከፍተኛ መካከለኛ ዝቅተኛ ምቹ አይደለም

Annex I: Words mostly signed by their ASLequivalent other than the EthSL

እንቁላል	Egg
ክብር	Honor
ሀሳብ	Idea
መጨረስ	Finish
ቦታ	place
እሺ	Ok
በኋላ	After
እይደለም	Not
ብርቱካን	Orange
ማዘን	Sad
ካልሲ	Socks
ህዝብ	People
መስኮት	Window
ጫማ	Shoes
እኛ	We

Signed Declaration Sheet

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

Declared by:

Name: Lily Abebe

Signature: _____

Date: _____

Confirmed by advisor:

Name: Yaregal Assabie(PhD)

Signature: _____

Date: _____