UNIVERSAL NETWORKING LANGUAGE BASED MULTILINGUAL MACHINE TRANSLATION SYSTEM FOR AMHARIC LANGUAGE

HENOK MESKELE

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

School of Electrical Engineering and Computing

Department of Computer Science Engineering

Office of Graduate Studies

Adama Science and Technology University

ADAMA, ETHIOPIA

June, 2019

UNIVERSAL NETWORKING LANGUAGE BASED MULTILINGUAL MACHINE TRANSLATION SYSTEM FOR AMHARIC LANGUAGE

HENOK MESKELE

ADVISOR: MESIFEN ABEBE (PHD)

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

School of Electrical Engineering and Computing and Department of Computer Science Engineering

Office of Graduate Studies

Adama Science and Technology University

ADAMA, ETHIOPIA

June, 2019

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF COMPUTING

APPROVAL SHEET

This research thesis entitled "**Universal Networking Language Based Multilingual Machine Translation System For Amharic Language**" has been read and approved as meeting the preliminary research requirements of the Department of Computer Science and Engineering in partial fulfillment for the award of the degree of Master Science in Information Systems Engineering, Adama Science and Technology University.

Adama, Ethiopia

_____	_____	_____
Name of the Student	Signature	Date
_____	_____	_____
Advisor	Signature	Date
_____	_____	_____
Head of the Department	Signature	Date
_____	_____	_____
Head of the School Dean	Signature	Date
_____	_____	_____
PG Office	Signature	Date

Declaration

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

Name: Henok Meskele ID:GSR/0250/09

Email:Henokmeskele2007@gmail.com

Signature:_____ Date: _____

Acknowledgements

First of all I really thank for God from inside my heart because of continuously God Spirit makes great favor things in our day today up to end of our life. Next to these special thanks for the research and science community, to show us the right direction how to do research and they give us a great inspiration to the universe, not only these, to share their new ideas and innovations advancement in the technology as well as, in the field of engineering “thanks for technology” it brings modern global world.

I would like to thank, “Adama science and Technology University”, the school of Electrical Engineering already they offer me a chance to engaged MSC program, and almost all Computer Science Program staff who are helping on these study to be completed and finalized, l respect to all the ASTU community to their special treatment during my study time.

My special appreciations and thank extend to my respectful advisor Dr. Mesfin Abebe for his support, guidance and patience throughout the preparation of this thesis. He always gave me helpful advice and comments, and provided me with useful resources. Having worked with him was a fruitful experience.

Finally, l would like to thanks for all my family members; last but not least my father Meskele Boset and my mother Genet Bekele for their endless precious support throughout my life. Expressly for dear my father lost before six months, I learned a lot of things from his long-life experience; I need to say thanks a lot my father: “rest in peace to you”. Lastly, this thesis work will be memorable to Dear my father “Meskele Boset”.

Table of Contents

Lists of Tables

Lists of Figures

Abstract

In this technologically integrated world, most people can easily access a huge volume of electronic information in the common platforms like internet, but the problem comes with the languages by which the information is represented. The machine translation is the name for computerized methods that automate the translations of documents from one language to another; it is a multi-disciplinary research area from linguistics, computer science, artificial intelligence, statistics, mathematics, philosophy and other.

*Universal Networking Language framework is an Interlingua platform for handling multilanguages over Inter. The UNL framework is used for various natural language processing tasks such as: information retrieval, text simplification and semantic reasoning, machine translation, representing, describing, summarizing, and storing information in an independent natural language (NL) format. There are two basic movements: Unlization is the process of representing the NL into UNL and Nlization is the process of generating a NL out of a UNL expression. UNL is far better than other methods, by using the UNL we only need [2*n] modules, while other methods uses [n*(n-1)] modules where n is the number of NL that are going to be translated.*

In Ethiopia, there are more than 80 nations and nationalities are speaking in their own native languages. Amharic is one of the members of the Semitic branch of the Afro-Asiatic family. The majority of the speakers of Amharic can be found in different parts of the country. It has five dialectical variations across different parts of the country. These include dialects like Addis Ababa, Gojam, Gonder, Wollo and Menz. But there are also Amharic speakers in other countries, Outside Ethiopia; it is used in Egypt, Israel, Canada, Sweden, Eritrea, and the US. Thus, it has official status is spoken by many people as their native and second language. In addition, it is a language with many literatures.

There are relatively fewer contributions of development of Amharic language under UNL framework. The previous conducted research was done mostly bilingual pair language, doesn't support multilanguage environments over net, old-fashioned to integrate centrally and outdated to manage as a common platform for all NL representations, they are very complex to develop corpuses and lingustics resources in quantity and quality due to they follow diffent techniqies to

repreesent a given languages. Tthat's why people are not satisfied with an internet usage due to linguistic diversity over Internet. Currently, these results are create incompatibilities among the various methodologies are used. Then, it is not possible to avoid language barriers all over the world, even if we get together all the results in one system.

The methodology is used for the conducted research is going to create UNlization and Nlization for Amharic language by using Interactive Analyser tools that takes Amharic language sentences as input and delivers UNL, it includes a dictionary and grammar for Amharic language analysis. It represents the sentences as semantic networks in the UNL format and dep-to-surface natural language generator tools is fully automatic natural language generation system that takes a UNL input and delivers an output in Amharic language without any human intervention. It generates the sentences out of semantic networks represented in the UNL format. Both of these tools are in its current release, it is a web application developed in Java and available at the UNLdev. The data collection technique is used for comparison UNL framework with others machine translations and support for prepared the corpus, dictionary and grammar for Amharic Language. Some of the sample corpus are very few in numbers, not able to handle complex sentences and few literatures are reviewed in Ethiopian local language contexts as the references and not covered all regional languages under this conducted research study.

The Evaluation was done F-1 score; it is a precision and recall for calculating the grammar accuracy, the final outcome is 0.640 and 0.0340 in the process of Unlization and Nlization respectively. Lastly, the evaluation of F-1score measure accuracy of the systems can be further improvement needs depending on the specific dictionary and grammar syntax structures; to come up handling for very large complex Amharic sentences at paragraph level and this study in the future can be integrated into other Ethiopian local languages.

Keywords:— *Universal Networking Language; Machine Translation; Source Language; Target Language; Knowledge Base; Natural Language Generation Surface; Interactive Analyze* .

List of Acronyms

ACL	Association Of Computational Linguistics
AI	Artificial Intelligence
ALPAC	Automatic Language Processing Advisory Committee
ANN	artificial Neural Network
AU	Africa Union
EU	European Union
CEC	Commission Of The European Communities
CETA	Centre D'etudes Pour La Traduction Automatique (Grenoble)
CRDSD	Current Research And Development In Scientific Documentation
DECO	Deconverter
DRULE	Disambiguation Rule
ELT	Electro-Technical Laboratory (Tokyo)
ENCO	Enconverter
EU	European Union
EUGEAN	Natural Language Generation Surface
EURATOM	European Atomic Energy Authority
EBMT	Example-based MT
IAN	Interactive Analyzer
IAS	Institute Of Advanced Studies
IBM	International Business Machines
IL	Interlingua, Intermediary Language
KB	Knowledge Base
LDC	Linguistics Data Consortium
LM	Language Monthly
LRC	Linguistics Research Center (Texas)
MIT	Massachusetts Institute Of Technology
MT	Machine Translation
NBS	National Bureau Of Standards
NL	Natural Language

NLP	Natural Language Processing
NLG	Natural Language Generation
NN	Neural Network
NMT	Neural Machine Translation
DNMT	Deep Neural Machine Translation
NP	Noun Phrase
NPL	National Physical Laboratory (Teddington, Middlesex)
NSF	National Science Foundation
OCR	Optical Character Recognition
PP	Prepositional Phrase
PBSMT	Phrase-based SMT
RADC	Rome Air Development Center
SL	Source Language
SMT	Statistical Machine Translation
SRS	Semantically Relatable Sequence
TA	Traduction Automatique
TAUM	Traduction Automatique De l'Université De Montréal
TL	Target Language
TRULE	Transformation Rule
UCLA	University Of California At Lossangeles
UMIST	University Of Manchester Institute Of Science And Technology
UN	United Nation
UNDL	Universal Networking Digital Language
UNL	Universal Networking Language
UNU	United Nation University
USAF	United States Air Force
UWS	Universal Words

Chapter One

1. Introduction and Background

This chapter presents comprising the introductory part of the machine translation and universal networking language is one of supporting Multilanguage framework environment over Internet, background of the study, statement of the problem, basic research questions, objectives, significance, limitation and scope of the study.

1.1. Introduction

Machine Translation (MT) is one of oldest dreams of humanity's to come up a dream as the solution to represented a given all natural languages (NLs) with various researchers' for human kinds among in the world . It is the use of computers to automate interpretations through computerized development methods that translating all or part of documents from one source language (SL) to another target language (TL) [1]. It is a multi-disciplinary field of research from linguistics, computer science, artificial intelligence, mathematics, philosophy and other [1].

Universal Networking Language (UNL) framework is a new direction of Interlingua representation platform for handling Multi-language over internet. It is an artificial language it has been exploited for several different tasks in NL engineering, such as for representing, describing, summarizing and storing information for multilingual document generation, summarization, text simplification, information retrieval and semantic reasoning. There are two basic processes: Unlization is the process of representing and analyzing the information taken by Natural Language (NL) into UNL and Nlization is the process of generating and mapping a NL out of a UNL graph [2]. UNL represents the information in three different types of semantic units, namely, Universal words (UWs), Relations and Attributes [3]. UWs are form of vocabulary of UNL, It has all the corresponding components of a NL, includes word concept, these words that linked together to generate a UNL expression of a sentence called Relations. There are 56 relations that are used to connect two or more UWs used to build semantic network, which define the role of each word in a sentence. Attributes: there are 87 attributes to clarify the semantics of a sentence to show the speaker's point of view is expressed information about time or aspect of the event, number, polarity, modality [4].

In the field of natural language processing (NLP), UNL has been used by various researchers as an Interlingua for automated MT system [1]. It has two main types namely, Enconverter (Enco) used for converting the text from a SL to UNL and Deconverter (Deco) used for converting from UNL to a TL [5]. Currently, researchers are studying to apply UNL in different languages in other countries, English language is mostly used language on the Internet, even though most Ethiopian people have difficulty of understanding the skill of this language [4] a lot of information is available in different language which is used by few people. MT system is the only means to distribute and transmission information but it needs much more efforts involving direct and indirect costs as per the number of languages that are going to be translated. Currently, Maximum translation works are done involving two pairs of languages. An effective solution is automatic translation system that users can easily read and understand. However, there are more than 5,000 languages in the entire world, the development of automated MT software for each language pair is extremely difficult, it is impossible for human to manually translate large no of documents. Especially for the languages that have less range of users [6]. But in Ethiopia UNL system have not been implemented until now. Amharic is the official language of Ethiopia which is near to international languages, so this system has its own importance to the local as well as international community. In this study, to introduce the UNL's applications of encoding and decoding process into Amharic language. This raises a MT it is wide research area in AI field as a challenge in the diversified multilingual environment over internet.

To avoid the linguistics barrier as a solution is to provide multilingual system software or website as the standard managed centrally. To translate a text from the SL into the TL is as intermediator, from any NL into the intermediate. Thus, with N language instead of having [n*(n-1)] translation pairs, we just need [2*n] translation pairs, this is because now only 2 conversions need to be done for that particular NL, from the NL to UNL and then from UNL to the NL. The development of automatic MT software for all languages is very challenging. Therefore, it needs to study another solution is mandatory to develop the translation system. UNL is one of the intermediate translation, presently through UNL has been developed currently for around 50 different languages [7]. While UNL has been studied and widely used in the world but in Ethiopia till now is going on under evaluation, it is not properly applied in a successfully way. This research has been carried out the investigation of the UNL deployment of Enco and Deco functions, initially, prepared Amharic corpus, towards create dictionary and Grammar localization. Lastly, evaluate the systems using F-1 to calculating precision and recall for grammar's accuracy, the aim of this study, is the feasibility of application of

linguistic resources of UNL system are evaluated through experimental result by test case to be accepted. In addition, these tools for supporting the independent format encoding and decoding process of language [7]. It is one of the intermediary languages that presently translated any NL to others, until now, has been developed around 50 different languages. For instance, research results for Russian and English [6] UNL Platform tools for 48 languages [8], [9] now Amharic is next.

1.1. Background

In this technologically integrated world, most people can easily access a huge volume of electronic information in the common platforms like internet, but the problem comes with the dissimilarity of languages. So, to solve this problem there is a requirement of some translating medium, which may be human translators, but they are limited and expensive resource in this competitive world. Thus, fully automated MT is emerged as a possible tool to solve language translation problems and comes in the focus of linguists, computer scientists and engineers to propose improvements in the translation [10]. The two major drivers of translation technology are assumed as war and commerce. Merchants have carried out trade with nations far removed from their own land. MT naturally became a necessity in such circumstances. Peoples are much better connected each other. But language barriers remain, and pose a challenge to communication [11].

In Ethiopia it has more than 80 nations and nationalities are speaking in their own languages. Amharic (አማርኛ) is the official language of Ethiopia as a member of the Semitic branch of the Afro-Asiatic family next to Arabic, it has many ethnic group of peoples are speaking as their own native languages in Ethiopia and other countries. Automatic MT system to Amharic language is very important for more than million users can communicate by their native language to others internationally and locally. This study tries to attempt integration Amharic language into UNL intermediate.

The UNL Program started in 1996, it was proposed by Dr. Hiroshi Uchida in Senior Research in Institute of Advanced study (IAS) as an initiative of the of the United Nation University (UNU) in January 2001 in Japan. [8]. The universal networking digital language (UNDL) Foundation, to be responsible for the development and management of the UNL Programme, it is a non-profit international organization, its headquarters are found in Geneva, Switzerland. From the very beginning, a group of university, round the world engaged in developing the UNL program, that's the UNL Society, involving many specialists in applied science and linguistics, who are at work

creating the linguistic resources developing the web structure of the UNL System. It provides technological support co-ordinates the implementation of the Programme it has already crossed important milestones [7]. The overall architecture of the UNL System has been developed with a set of basic software tools necessary for its functioning. More information about the earlier system can be found; these are being tested and improved [12], [13], [14], [15], [16].

1.2. Statement of the Problem

All languages have not enough automated resources over internet. This reveals that, all needed information may not be accessed by a desired language. Moreover, many institutions are produced large amount of documents from day to day activities, it is difficult towards human translators handle all these amounts of documentation, because, if the number of documents are too much **bulk**, the translation will not perfect and it takes more efforts to finish on a given fixed interval time [17]. The exponential growth of the Internet has made its content increasingly difficult to find, access, present and maintain for a wide variety of users [18]. In addition, the current Internet's content was designed for humans to read, not for computer to manipulate meanings [19]. The existing functions of computer are limited in just retrieving information and displaying to people without processing the semantics. Therefore, computers give little support in finding, accessing and maintaining the Internet's content. Language variations losses the enhancement of linguistic resources effectiveness,

Apart from sociocultural issues, it has as economic and political dimension [17]. Institutions like the united nation (UN), Federal States Based on language like Ethiopia and India or the Africa union (AU) have to face every day with the linguistic barriers. It is assured that the huge amount of documentations these institutions produce every day: 6 for the UN, about 80 in our country, 26 for the EU. As Ethiopian context, federal government and regional states may use two languages for their daily activity. There are relatively less contributions to integrate for Amharic language under UNL framework.

The conducted research is going to create Unlization and Nlization for Amharic language by using Interactive Analyser (IAN) tools that takes Amharic language sentences as input and delivers UNL, it includes a dictionary and grammar for Amharic language input analysis. It represents the sentences as semantic networks in the UNL format and dep-to-surface natural language generator (EUGENE) tools is fully automatic natural language generation (NLG) system that takes a UNL

input and delivers an output in Amharic language without any human intervention. It generates the sentences out of semantic networks represented in the UNL format. Both of these tools are in its current release, it is a web application developed in Java and available at the UNLdev. Both IAN and EUGEAN have been accepted globally as a standardized tool for Unlization and Nlization process. Meanwhile there is no central automatic UNL based Enco and Deco function to Amharic language [7]. Unlike previous tools which are old-fashioned and outdated, complex which are not easily integrated centrally, UNL is easily integrated Interlingua representation for all language resources.

Many researchers have come up with different systems in different approaches. Each has their special representation of a given language. This results in incompatibilities among the approaches. Then, it is not possible to avoid language barriers all over the world, even if we get together all the results in one system. Even though, different models were used as a solution, for N number of languages, in all methods, it requires [n*(n-1)] modules to be developed , which needs large amount of computational devices and it takes more time to developing translation rules [9], [17] . Direct method for MT is most suitable for closely related languages. The transfer and statistical approach is suitable for languages with different structures. The statistical approach is a promising in the field of MT, but its strength depends largely on the amount of resources required in terms of corpus [20] [21]. But, since the Amharic language is still one of the less resourced languages it is very difficult to find an Amharic corpus thousands of sentences under UNL. The UNL framework Interlingua required [n*2] Enco and Deco modules, it has the maximum highest grammar accuracy on precision and recall for F-1 score measure, which is a validation of rule based dictionary entry and limited grammar syntax structures through these Interlingua model. To solve problems with other models, the thought of UNL framework as a standard for all systems was born. With the results of the past work can be applied to the present and the future research and infrastructure development [22]. To deal with the above problems, UNL framework was defined as an artificial language that is able to represent information and knowledge described in NL [23]. Thus, it enables computers to process information in the form of knowledge across the language barrier [24], [25].

1.3. Motivation

The increasing needs and use of Internet for cultural, educational and commercial services are improved, but many of the people are not satisfied with Internet practice due to linguistic diversity, which is reduces the possible of Internet as a vehicle of knowledge for everybody [25]. Since the

advent of computers, researchers around the world have worked towards developing a system that would overcome the language barriers over net [25]. As the information on Internet is rising and updated daily, it is becoming more and more difficult for the people to find the information they are looking for on the Internet. The Internet changes the world very fast. Now we will notice large quantity of information on net. But most of this information is in English which means, still now it is language dominant. There is also a demand of Internet based online translation services, Request of online versions of electronic dictionaries as 'translation systems' to help human translators for translation. The Internet also suggests other roles for such as searching data and information is a challenging obstacle jobs in any languages "Cross Language Information Retrieval Systems" are considered, Google translation is one of the interesting tools for any user.

The UNL framework could be a unique artificial semantic language that captures the meaning of written documents on the illustration of ideas and its relations. Therefore, the researcher is interested to study applying integration for Amharic language MT system under the UNL framework. The researcher is a native Amharic language speaker eager to associate Amharic language via other languages using UNL framework Interlingua based MT Interlingua intermedium. Additionally, the researcher has good attitude and focuses to challenges of multi-lingua based system for Amharic local NL to assimilate any others language.

1.4. Research Questions

This study to create the awareness of various types of MT systems and evaluation technique in the multilanguages environments. Lastly, compared different types of MT models to find the best framework for deployment for Amharic local language by answering the following basic research questions are:

1. What are the most challenges and suitable framework used for multilingual system?
    - What are the existing MT models and evaluation techniques?
    - Which best models are not yet tested currently for Amharic language?
2. How to integrate effectively Amharic language to UNL framework can be applied?
    - Which Amharic language parts of speech is translated effectively?
    - Which errors occur most frequently during experimentation study?

1.5. Objective of the Study

1.5.1. General Objective

The general objective of this study is to design and develop UNL based Automatic multilingual Machine Translation system for Amharic language.

1.5.2. Specific Objective

The following are the specific objectives of the research that help to achieve the general objective of the study.

- To review various literatures in order to have an understanding on concepts, principles, Terminologies, Technologies of studying UNlization with IAN, Nlization with EUGENE frameworks and forward research ideas.
- To build the UNL MT model that represented the acquired domain knowledge using suitable UNL knowledge representation technique (UNLKB).
- To implement Enco and deco function on corpus UC-A1 provided by UNDL Foundation towards localization for Amharic language.
- To create Analysis and generation Dictionary, grammar (Trules and Drules) for both Unlization and Nlization to all major part-of-speeches respectively.
- To evaluate Unlization and Nlization by calculate F-1 of grammar accuracy of precision, recall of UC-A1 sentences.

1.6. Significance of the Study

From this study, institutions which are using more than two different languages are the immediate beneficiaries from the developed systems. It has great significance for local and international researchers in the field of NLP tasks will be considered governmental and regional offices of Ethiopia to provide good services for their customers without language constraint. It is good opportunity for accessing high quality MT system with additional application improvements increasing speed of access linguistics resources via the user that allowing them to communicate with other people in their own languages. It reduces costs of developing contents in an independent way and Allows distributing, receiving, exchange and share textural information on the Web. The users can share information and knowledge across native languages in a high speed MT .Citizens, governments, international organizations, and enterprises are beneficiary from this study and offers

opportunities for information sharing, education, and e-business. The ultimate goal is to promote sustainable development and dialogue among civilizations, economic prosperity for all nations and nationalities as well as peace among them. This study greatly reduces developmental cost necessary to make knowledge processing among various languages.

1.7. Scope, Limitations and Delimitation

The scope of this study is to design and develop a Multilanguage MT system for only Amharic language, excluding other Ethiopian local languages. The development of Enco and Deco are primary research towards UNL for Amharic Language using common Enco & Deco tools through IAN & EUGEANE respectively, provided by UNDL foundation. The corpus uc-a1 manually translate into Amharic from the English version, very few in numbers a set of 100 simple type of Amharic sentences, are not handle very large & complex sentences and used for prepared corpus, the dictionary for mapping the appeared words in the corpus and grammatical rules done linguistic inflections in order to sense language structure per limited instructions for handling semantic network. The evaluation of this study where done by F-1 score which are accuracy of precision, recall for Enco & Deco process for Amharic language.

The researcher of this study is encountered the following limitations, moreover, No one conduct research without some obstacles, some of these were the following: Financial and time constraint impassable to fully cover all parts of grammatical units, the sample size is the set of 100 sentences and the system is not able to handle very large complex sentences and very few literature survey for a reference. It can be difficult to generalize the findings of the study at all regional NLs of Ethiopia. It is not accessed easily available document that related to the UNL application.

Since UNL system is being developed by multiple interested communities from different backgrounds it amounts to different dialects that may need translation from source to target. This created the difference in terminologies in different centers [7]. Moreover, the difference in language structures such as subject object verb (SOV) and subject verb object (SVO) and so on, UNL specification such as dictionary and grammar become a challenge to make central system for all languages. However, the system is expected equally correct results for all languages if transformed by different persons [7], [26], otherwise, the impact of transformation of any given NLs are a little

bit difficult to achieved successfully integrate to UNL framework Interlingua representation for a human kind

1.8. Organization of the Thesis

This thesis is divided into seven chapters these are: introduction, review of literature; research methodology, UNL framework and the Amharic language characteristic, design and development of UNL, experimentation, results and discussions; conclusion and future work:

- ***First chapter***: contains introduction to MT and Amharic language, background of study, motivation and research questions, objectives of the research, the main objective of this study is to design and develop a UNL MT system for Amharic language Enco and Deco function.
- ***Second chapter:*** deals with literature review and related work, types of MT, challenges and evaluation techniques before the invention of computers that beginning of automated MT until now and describe the related work of the research.
- ***Third chapter***: deals with the research methodology, a step by step procedure that finalized Amharic Enco and Deco function in IAN and EUGEAN tools for Unlization and Nlization respectively. The Enco accept input Amharic sentences into UNL and The Deco getting Amharic sentences from UNL, these modules has special tasks with in several tabs as components to do their jobs from one SL to other TL through attributes and relations.
- ***Fourth chapter***: explain UNL framework and the Amharic language characteristic in details of the UNL framework architecture sample specifications and to identify their associations how to work and integrated to the UNL components like IAN and EUGEAN, and detail definitions about Amharic local language characteristics with the help of sample sentences.
- ***Fifth chapter****:* that comprises the design and development of Unlization and Nlization in details of the Enco and Deco Design with based on sample specifications how to prepared corpus, dictionary, and grammar and to calculated the F-1 score for grammar accuracy on precision and recall, the evaluation of proposed system tested on with the help of sample corpus.
- ***Sixth chapter****:* presents the experimentation, result and discussion of the thesis work in order to implement the development of Enco and Deco functions for Amharic language. The actual output automatically generated by IAN and EUGENE with the help of

grammars and dictionaries, the expected output is the same set of sentences provided from the UNDL foundation English sentences version translated equivalent Amharic sentences. The UNL is compared with the expected output and the actual output in the system is evaluated based on F-1 precision and recall of grammar accuracy that displays the result with in detailed explanations.

- ***Seventh chapter***: presents the conclusion and future work of the thesis work in order to improve the proposed system under UNL framework development of Enco and Deco functions for Amharic language and other Ethiopian local languages.

Chapter Two

2. Literature Review and Related work

This chapter presents the review of literatures comprising of major Translation categories namely Human and Advanced MT, historical development of MT, paradigms of MT, Challenges and Evaluation techniques

2.1. Literature review

Relevant literature such as books, journal articles, conference papers and different resources from internet that are related to MT system was done. Thus, the System of development tools and techniques that were how to use for the system for utilization also reviewed. Moreover, to assess what others have been done in the area to understand the problem in a better way. Related research works and projects were assessed what kind of methodology they were used in the study and explain the result discussions. Finally, we talk about the recommended solutions. In general, comprehensive investigations of available literature on different techniques of MT system and Amharic language structure where made carefully, the details are in chapter four [27] [28] [3] [12], [54], [55] , [53], [52].

2.2. Translation

Translation is a complex activity that requires an enormous effect and a series investigation for researchers. Since, the issues of heterogeneous languages are a key international treaty to small restaurant welcome posters. Translation tasks can be categorized broadly into human or manual translation and Automatic Machine Translation or simply MT [29].

2.3. Expert Human Translators

Expert human translators use their background knowledge, mainly to resolve grammar and semantical ambiguities of a given source language to the target language that can have to be forced to compelled by some translation rules [30].These consultants appear to agree that the 3 requisites in an exceedingly translator are (1) smart information of the target language, (2) understanding of the subject matter, and, (3) adequate knowledge of the source language. Furthermore, whereas smart translations are created by some translators have a general appreciation of knowledge domain, the best technical translations are generally made by experts in the technical field covered [29]. One

of the problems is that there are too few human translators, and that there is a limit on how far their productivity can be increased without automation. In short, it appears as if automation of translation may be a social and political necessity for contemporary societies that don't want to impose a standard language on their members [31].

2.4. Machine Translation Techniques

Humans and machines may play different roles in MT methods. These are given below:

Table 1: Techniques of MTs

MT methods	*Descriptions*
Fully automatic MT	The whole process is carried out by the machine, without any intervention of the human user.
Human-aided MT	The process is carried mainly by the machine, with some intervention of the human user, either as a pre-editor or as a post-editor, or during translation process itself.
Fully human MT	The whole process is carried by the human user, without any intervention of the machine.
Machine-aided human MT	Machine-aided human MT The process is carried mainly by the human user, with some help of the machine.

2.5. Machine Translations

MT has become an interesting commodity from the earliest days people tried to find out methods to translate external languages into their own. The stone of Rosetta found in Egypt is one of these ancient witnesses. Since the WWII and the following Cold war automatic translation became a very hot research item. MT means going by algorithm from machine-readable source text to useful target text, without operate to human translation or editing [8]. It was a long-term scientific dream of enormous social, political, and scientific importance. It was one of the earliest applications suggested for digital computers, but turning this dream into reality has turned out to be a much harder, and in many ways a much more interesting task than at first appeared. Nevertheless, though there remain many outstanding problems, some degree of automatic translation is now a daily reality [32]. However, the area of MT is still relatively unidentified throughout the world and in our country too, it is even unknown to the linguists, the informing specialists, and the members of the translation community. This unfamiliarity is mainly the outcome of ignoring the advanced technologies that have been used in the area, and is sometimes the result of refusing the pressing

need for using it in cultural and scientific exchange, considering all its branches and trends, from one language to another [18].

2.6. Historical Development of MT

Here, an overview on MT's historical development summary from the MT Archive as a library on itself since the early 1933's up to early 1989.

Table 2: Historical Development of MT

Period	***Source***	***Major Developments***
1933-1946	Hutchins, 2000 [9]	• General-purpose machine which could also function as a mechanical multilingual dictionary. • A mechanical dictionary with detailed proposals for coding and interpreting grammatical functions using universal symbols in a multilingual translation device • Ideas for using the newly invented computers for translating NLs.
1951-1955	Hutchins, 2003 [10]	• the beginning of MT research at MIT and the first MT conference at MIT • the first demonstration of a MT system by IBM and • the first MT book appeared • The first journal called Mechanical Translation was founded.

1956-1965	Hutchins, 2000 [10]	• the beginnings of MT-direct translation model, 'interlingua' model and the 'transfer approach'(described in the next section) • practical development at the IBM, and a system installed for the US Air Force produced (crude but apparently usable) translations • David Hays developed the first syntactic parser based on dependency grammar. • The Georgetown Automatic Translation (GAT) system was developed-had three levels of analysis: morphological (including identification of idioms), syntagmatic (agreement of nouns and adjectives, government of verbs, modification of adjectives, etc.), and syntactic (subjects and predicates, clause relationships, etc.) • Translation was seen as a series of decoding and encoding processes, via a series of strata (graphemic, morphemic, lexemic and sememic). • Fifty universal 'lexical functions' were identified at the deep syntactic stratum covering paradigmatic relations (e.g. synonyms, antonyms, verbs and their corresponding agentive nouns, etc.) and a great variety of syntagmatic relations (e.g. inceptive verbs related to given nouns, conference: open, war: break out; idiomatic causatives, compile: dictionary, lay: foundations, etc.) • MT research groups had been established in many countries throughout the world.
1966	Hutchins, 2000 [10]	• In 1964 the government sponsors of MT in the United States (mainly military and intelligence agencies) asked the National Science Foundation to set up the Automatic Language Processing Advisory Committee (ALPAC) to examine the situation. • In its famous 1966 report it concluded that MT was slower, less accurate and twice as expensive as human translation and that "there is no immediate or predictable prospect of useful MT translation" • It saw no need for further investment in MT research; instead it recommended the development of machine aids for translators, such as automatic dictionaries, and the continued support of basic research in computational linguistics. • The ALPAC report ended the previous perception of MT as the leading area of research in the investigation of computers and NL. Computational linguistics became an independent field of research.

1967-1976	Hutchins, 2000 [10]	• Considered the quite decade and there was a change of direction • Where "first generation" research of the pre-ALPAC period (1956-1966) had been dominated by mainly 'direct translation' approaches, the "second generation" post-ALPAC was to be dominated by 'indirect' models, and both are Interlingua and transfer based. • The TAUM project (Traduction Automatique de l'Université de Montréal) had two major achievements: firstly, the Q-system formalism for manipulating linguistic strings and trees (later developed as the Prolog programming language), and second, the Météo system for translating weather forecasts.
1976-1989	Hutchins, 2000 [10]	• Translation Systems such as Systran, Logos and METAL were in designed for general application • most of the computer companies (Fujitsu, Hitachi, NEC, Sharp, Toshiba) developed software for computer-aided translation, mainly for the Japanese-English and English-Japanese markets • Investigations of AI methods in MT began in the mid-1970s • Translators were becoming familiar with the benefits of computers for their work – word processing, creation of individual glossaries, facilities for on-line access and transmission of documents. • Characterized by the MT revival of research
Since 1989	Hutchins, 2000 [10]	• Emergence of Corpus-based approaches • statistical methods have been a major focus of many research groups • rule-based research continued in both transfer and interlingua systems • Institute of Advanced Studies of the United Nations University , japan, (Tokyo) began its multinational interlingua based MT project – based on UNL (Universal Networking Language), for initially the six official languages of the United Nations and six other widely spoken languages (Arabic, Chinese, English, French, German, Hindi, Indonesian, Italian, Japanese, Portuguese, Russian, and Spanish) – involving groups in some 15 countries. • The most significant development of the decade has been the growing interest in spoken language translation. • development of systems for specific subject domains and users has also expanded rapidly

2.7. Machine Translations and Internet

Translation is a complex activity which involves knowledge of many fields, no matter what type of translation belongs to. The translator must know the basic background knowledge, but the translator's personal knowledge and information are limited. Now, the network provides a huge shared data bank for the translator, which includes encyclopedia, the dictionaries and some special resources through search engines. Thus the translation quality can be improved by reducing understanding errors [5].

The emergence of Internet influenced the MT development practices these includes are [2]:

- The appearance MT software products specifically for translating Web pages and electronic mail messages.
- Many MT vendors have been providing Internet-based online translation services for on-demand translation. Some of the online services (most of them free of charge), e.g. Softissimo with online versions of its Reverso systems, LogoMedia with online versions of LogoVista and PARS. Some services offer post-editing by human translators (revisers), at extra cost, but in most cases the results are presented untouched in any way.
- The Internet has also encouraged somewhat less reliable companies to offer online versions of electronic dictionaries as 'translation systems'. Anyone using such products for translating full sentences is bound to get unsatisfactory results – although if they do not know the Target languages (TLs) they will be unaware of the extent of the incomprehensibility of the results.
- The Internet suggests other roles for MT, as components of other language-based operations. Searching information and data in languages unknown or poorly notable by the user could be a formidable obstacle. Translation of individual keywords is not true translation – just as electronic dictionaries are not 'translation systems' – and further research is being conducted. Still further off is the possibility of multilingual summarization; there are monolingual systems for producing automatic summaries of documents and articles, but combining summarization with automatic translation is a major challenge for researchers [2].

2.8. Machine Translation Challenges

Translating written text requires understanding the context of the language and can pose several challenges. These challenges were identified in several literatures which includes [1]:

- Language Structure: Languages can be differentiated in the way they order words in sentences; some languages follow SOV (Subject-Object-Verb) structure like Hindi, Turkish, and Japanese etc.; while some follow SVO like English, Malay, and Germanic etc. Some languages follow VSO structure like Arabic, Hebrew etc. and very few languages follow VOS structure. Translation may become difficult when a sentence in SL/TL follows one type of structure and in TL/SL follows another [6].
- Translate or Transliterate : This is the issue of how to decide whether a word in SL is to be translated or transliterated, where transliteration means representing word given in SL into a word in TL, in which the character set of the TL only need to be used, keeping the pronunciation of the original word intact. Some words in a sentence are required to be translated, whereas some words do not need to be translated but rather transliterated like name of countries, states, people, places, organizations etc. In some cases it becomes difficult to make such a differentiation among words [17].
- Tokenization: The analysis of the sentence so as to divide it into tokens or individual words. In most of the languages tokenization is easy as white space or punctuation is used as word boundary, but there are some languages where it is difficult to find tokens to differentiate words where words are not delimited by common delimiters: white space or punctuations. In such cases additional algorithms are required first to tokenize sentence and then to translate, which makes the translation procedure more difficult [29].
- Homograph: Some languages support homographs, which are words having same spellings but entirely different meanings like English, Chinese, Japanese To choose correct meaning of such words one needs to know context information of that word, which is difficult to judge sometimes, especially by a machine [11].
- Speaker's Attitude: Attitude of the speaker can change the meaning of a word and in return that of the sentence. In a written text, it is difficult to judge the speaker's intentions and therefore might result in incorrect translation. Sometimes a positive sentence, but said in derisive manner can mean something entirely different. For example: "Do you actually suppose he's brilliant?" can mean, the person is asking whether someone is brilliant or not, but

if this sentence is said in sarcastic manner then it can mean that the speaker is doubting person's brilliance [33].

- Languages with multiple scripts: Some languages support multiple scripts and these scripts make translation all the more difficult, because one needs to form a lot of rules for translation for all types of scripts [19].
- Language Writing Systems: Languages all over the world can be categorized into three categories based on the writing systems they use, which are Alphabetic systems, Syllabic and Logographic languages. Alphabetic system uses a small set of letters or symbols called alphabet representing phoneme of language as in English, Hindi, Amharic language etc.; Syllabic uses a set of written symbols called syllabary representing syllables, which are used to constitute words in that language as in Korean, Japanese language; Logographic uses logograms, where a logogram is a single character representing a grammatical word in that language as in Chinese language. Logographs pose difficult challenges when needs to be translated especially [30].
- Grammatical Variations: Languages have some common features like Gender, Number, and Prepositions. But some variations are there and these can create ambiguity. For example, Gender: Some languages allow two genders like Hindi, French, Spanish etc., while some languages allow three, neutral being the third, e.g. English .Moreover, gender affects Pronouns and verbs. Number: In some languages, number system is of two types (Singular/Plural) e.g. English, Hindi etc., while a few languages have three types, dual being the third, e.g. Arabic and Hebrew. In such languages, plural is used for clusters with more than two items. Number affects noun, pronoun, verb and adjective in a sentence. Similarly, modifiers are placed before noun in some languages, while in others these are placed after nouns. Such variations in languages pose challenges while translating text, especially when SL follows one type of features and TL another [10].
- Morphology: There are many difficult problems of the NLP branches in MT, it's hard to get breakthrough before solving some basic problems because of the ambiguous meanings in morphology analysis, the part-of-speech tagging, the syntax analysis and the semantic analysis. The main problems in the MT exist in the following aspects [29]:
- Polysemy: In MT process, sometimes at first you need to determine the specific meanings of a word, as well as the correspondence in the TL, but the relationship between any two different languages is not one-to-one correspondence, because the words can derive extended meaning in its own evolution in order to meet the need of the presentation, therefore polysemy

appeared. The solution to polysemy is the word sense disambiguation, also called the semantic tagger, which means giving every word in the text its semantic encoding according to its respective context. At present, polysemy typography study is still in its beginning stage; mainly include artificial intelligence methods, dictionary-based methods and corpus-based approaches [9].

- Word Concurrency: Theoretically it refers to the syntax distribution of one word with two or even more kinds of parts of speech, these words are called the concurrent word. Therefore, although the part of speech of a word in a sentence is definite, because it belongs to a certain context, sometimes the part-of-speech tagging cannot guarantee the correctness completely. The so-called part-of-speech tagging determines grammar- category of each word, thus determines its part of speech. Part-of-speech tagging automatically is the automatic elimination of ambiguous meanings of word concurrency.
- Ambiguous meanings of phrase structure: Different adhesive relations of nouns, prepositions and so on can result different reasonable meanings. Distinguishing the structure ambiguity of the MT is a wide spread problem, because it involves a deep analysis and comprehension of NL. At present, we cannot do MT on the basis of the comprehension of NL, what we can do now is just to study it as a kind of applied technology. Using the strategy of ambiguity to ambiguity is only an expedient way before the system has the capacity to identify structural ambiguity [34].
- The quality of translation cannot be guaranteed: The difficulties to distinguish semantic ambiguity and choice of semantic meaning degrade the quality of translation, which was caused mainly by the lag behind of the theory research on the development basis of MT. At present, the entire translation system was established mainly on the syntactic level, the limited semantic which information system is used for the syntactic level, the TL relies solely on simple syntactic structure conversion from the SL, but this is not enough. In addition, the language rules which the system used are still based on traditional grammar, they cannot reflect the complex characteristics of NL, the quality of the translation has not made substantial progress, a lot of problems existed 50 years ago has not resolved now. The existing translation progress relied more on the computer-resource development, but not on the translation method. Since the 1990s, the computer and network technology developed a lot, therefore the range of available technologies are further expanded, which includes full-text search technology, voice input, network information gathering and the query engine, and so on [35].

2.9. Machine Translation Paradigms

MT approaches have been grouped into a number of categories. Automated MT systems can be classified according to their main paradigm as shown in Fig. 2.1 mainly as rule-based and empirical systems. The distinctive characteristic is the resources they use [11].

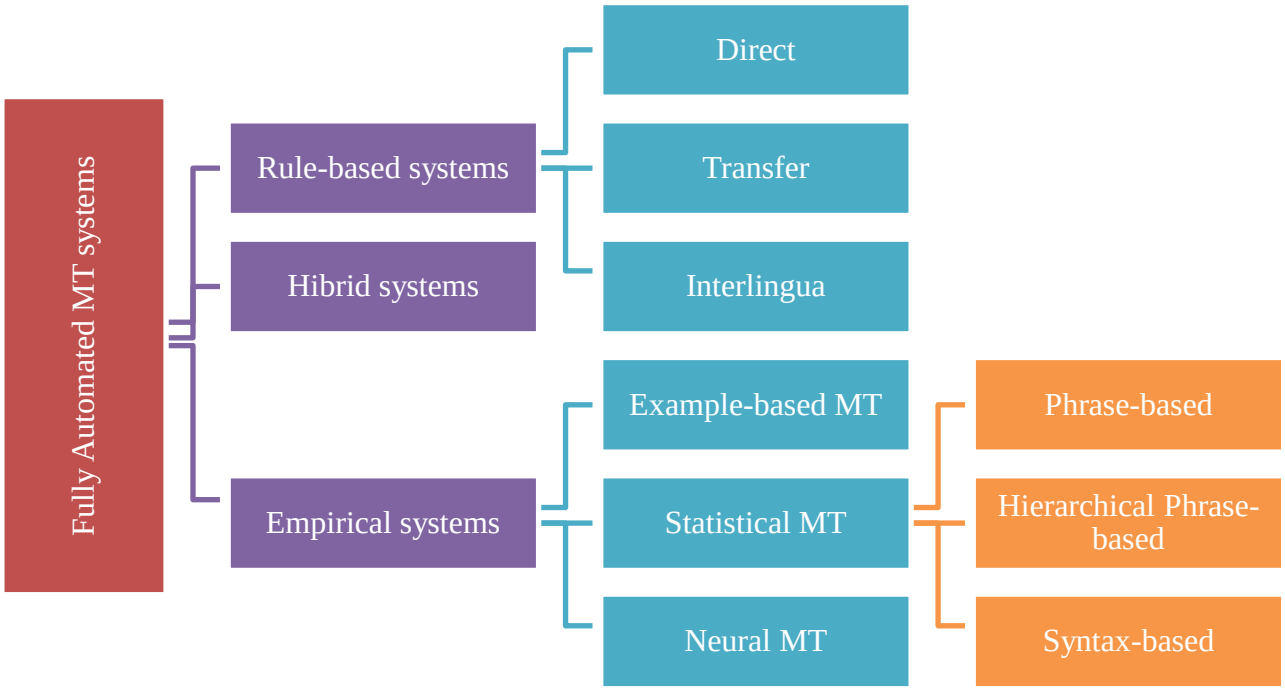

Figure 1: Taxonomy of fully automated MT

- **Rule-Based Systems**: The amount of the linguistic techniques used in a translation system distinguishes in direct, transfer and Interlingua rule based MT approaches [6]. The linguistic approach utilized in this technique is split into 3 phase's source [17], [36], [37].

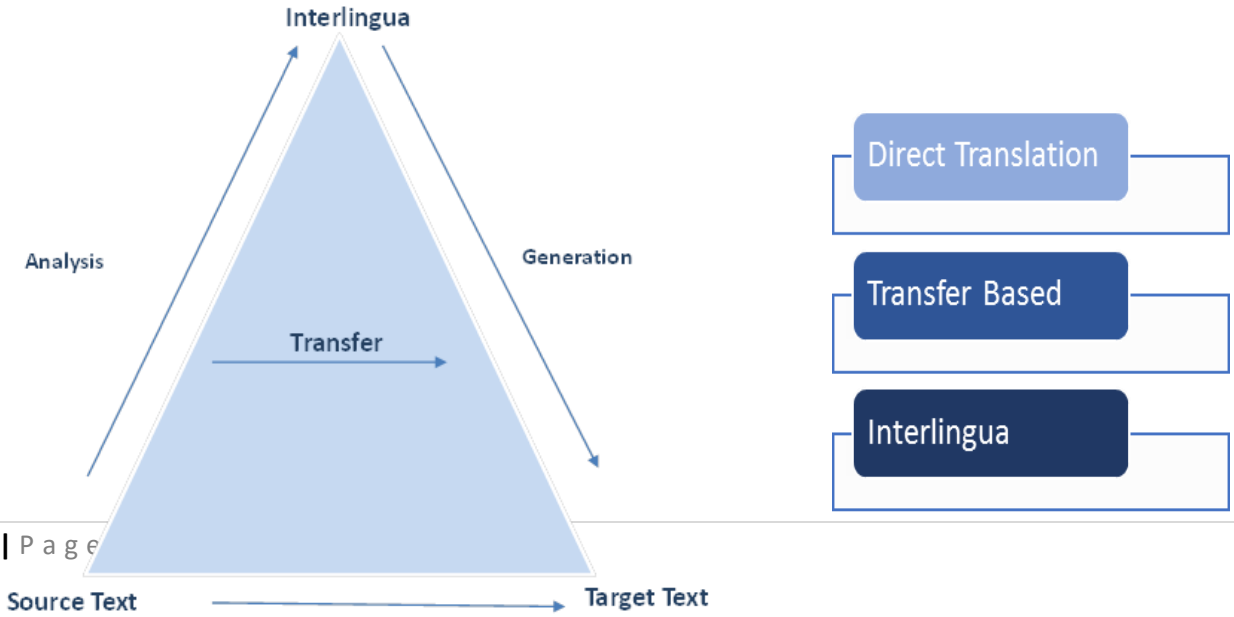

Figure 2: Vaquees Triangle.

1. **Analysis**: Analyzing the source text in the intermediate representations.
2. **Transfer**: Transferring intermediate representations into intermediate representations in the target language.
3. **Generation**: Generating the new text in the target language from the intermediate representations in the target language.

Table 3: Rule-based approaches

Rule-based Approches			
Type of translation	***Direct translation***	***Transfer approach***	***Interlingua approach (UNL)***
Description	• The source sentence is segmented into words. • The transfer step uses a bilingual table that associates for each word. • Translation rules and rearrangement for translating and reordering the words in the target sentence.	• To transform the SL parse structure into a TL parse structure the source sentence is analyzed with a parser and grammar. • A bilingual table, containing the transfer rules between the SL and TL.	• The translation of SL to TL is the process of Enco the SL and deco the TL. • It is able to use multilingual translations among Enco and deco modules [2*N]
Prerequisites	Dictionary entries and linguistic rules		
Advantages	The quality of the translation is very good, particularly for specialized translation.		
Limitations	The development of linguistic rules is expensive in terms of time and effort.		

- ***Hybrid Systems****:* It is a method of MT that is characterized by the use of multiple MT approaches within a single MT system. The motivation for developing hybrid systems protect from the failure of any single technique to achieve a satisfactory level of accuracy [11].
- ***Empirical Systems***: The translation system that involves example-based, statistical and neural approaches.
- ***Example-based MT*****:** The system is trained from bilingual parallel corpora containing sentence pairs. Sentence pairs contain sentences in one language with their translations into another. The particular example shows associate example of a tokenize combine, that means that the sentences vary by only 1 part. These sentences make it simple to learn translations of portions of a sentence [11].
- ***Statistical MT***: The system estimate probabilistic models by frequency counts in corpora. The best translation in this system is that maximizing the product of models. This system requires two models [26]:
  1. Translation model: which provides for each source segment the translation candidates in the target language?

2. Language model: This is responsible for taking the constraints executed by the syntax of the target language using probabilistic functions.

Table 4: Statistical approaches.

Statistical Approach			
Type of translation	Phrase-based	Hierarchical Phrase-based	Syntax-based
Description	• Translate phrases as atomic units. • NL input is segmented in phrases each phrase is translated into TL are reordered. • To reduce the restrictions of word-based translation by translating whole sequences of words.	• It consists of words, sub-phrases and this hierarchy is intended to capture reordering among phrases. • The hierarchical phrase pairs are obtained from a synchronous context-free grammar (CFG) learned from parallel corpora without syntactic information.	• The basic idea is to use synchronous grammars which are able to generate the source and the target simultaneously. • Synchronous grammars are learned from parallel corpus and that makes the approach very slow in comparison to the PBSMT systems.
Prerequisite	Large parallel corpora		
Advantages	• dealing with lexical ambiguity and requires minimal human effort • can be created for any language pair that has enough training data		
Limitations	Difficulty of building sufficient linguistic resources in quantity and quality		

- **Neural MT**: This approach uses a large artificial Neural Network (ANN) to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model. Deep Neural MT (DNMT) is an extension of NMT. Both use a large neural network with the difference that DNMT processes multiple NN layers instead of just one [26]. The key profit to the approach is that one system will be trained directly on supply and target text, no longer requiring the pipeline of specialized systems used in statistical machine learning. Unlike the standard (PBMT) system that consists of the many little sub-components that area unit tuned individually, (NMT) tries to make and train one, massive (NN) that reads a sentence and outputs a correct translation [37].

2.10. MT Evaluation Techniques

2.10.1. Human Evaluation Methods

There are new types of various developments of human evaluation methods of MT outputs. These are summarized the table below source [38].

Table 5: Human Evaluation Methods

Human Evaluation Methods	
Traditional Human Assessment	• Intelligibility and Fidelity • Fluency, Adequacy, Comprehension • Further development
Advanced Human Assessment	• Task Oriented • Extended Criteria • Utilizing Post-editing • Segment ranking

- **Traditional Human Assessment**

1ntelligibility & Fidelity: The requirement that a translation be intelligible means that, as far as possible, the translation should read like normal, well-edited prose and be readily understandable in the same way that such a sentence would be understandable if originally composed in the translation language. The requirement that a translation is of high fidelity or accuracy includes that the translation should, as little as possible, twist, distort, or controvert the meaning intended by the original [38].

***Adequacy*:** The evaluator is asked to look at each fragment, delimited by syntactic constituent and containing sufficient information, and judge the adequacy on a scale 1-5. The results are computed by averaging the judgments over all of the decisions in the translation set.

***Fluency*:** The fluency evaluation is compiled with the same manner as that for the adequacy except for that the evaluator is to make intuitive judgments on a sentence by sentence basis for each translation. The evaluators are asked to determine whether the translation is good English without reference to the correct translation. The fluency evaluation is to determine whether the sentence is well-formed and fluent in context [38].

***Comprehension*:** The modified comprehension develops into the "Informativeness", whose objective is to measure a system's ability to produce a translation that conveys sufficient information, such that people can gain necessary information from it. Developed from the reference

set of expert translations, six questions have six possible answers respectively including, "none of above" and "cannot be determined" [38].

***Further development*:** The "Linguistics Data Consortium" (LDC) develops two five-point scales representing fluency and adequacy for the annual MT evaluation workshop. The developed scales become the widely used methodology when manually evaluating MT is to assign values. The five point scale for adequacy indicates how much of the meaning expressed in the reference translation is also expressed in a hypothesis translation; the second five point scale indicates how fluent the translation is, involving both grammatical correctness and idiomatic word choices.

- **Advanced Human Assessment**

***Task Oriented*:** The principal steps include identifying the user-performed text-handling tasks, discovering the order of text handling task tolerance, analyzing the linguistic and non-linguistic translation problems in the corpus used in determining task tolerance, and developing a set of SL patterns which correspond to diagnostic target phenomena [38].

Extended Criteria:

- ***Suitability***: whether results are suitable in the particular context in the system is to be used.
- ***Interoperability***: whether with other software or with hardware platforms.
- ***Reliability***: it doesn't break down all the time or take long time to get running again after breaking down.
- ***Usability***: easy to get the interfaces, easy to learn and operate, and looks pretty.
- ***Efficiency***: when needed, keep up with the flow of allocated documents.
- ***Maintainability***: being able to modify the system in order to adapt it to particular users.
- ***Portability***: one version of a system can be replaced by a new version, because MT systems are rarely static and they tend to be improved over time as resources grow and bugs are fixed.

***Utilizing Post-editing*:** A measure of quality is to compare translation from scratch and post-edited result of an automatic translation. This type of evaluation is however time consuming and depends on the skills of the translator and post-editor [38].

***Segment ranking*:** Each time, the source segment and the reference translation are presented to the judges together with the candidate translations of five systems. The judges will rank the systems from 1 to 5, allowing tie scores [38].

2.10.2. Automatic Evaluation Methods

There are new types of various developments of Automatic evaluation methods of MT outputs. These are summarized the table below source: [38].

Table 6: Automatic Evaluation Methods

Automatic Evaluation Methods	
Traditional Automatic Assessment	• Lexical similarity- Word order, Precision and Recall, and Edit distance • Linguistic features- Syntactic similarity and Semantic similarity
Advanced Automatic assessment	• Deep Learning Models

- **Traditional Automatic Assessment**

Edit Distance: it's a natural living of the assessment of errors in MT output, (number of errors, divided by range of reference words), with every edit viewed as fixing miscalculation. We define edit distance to be the number of insertions, deletions and substitutions that are required in order to make a system translation equivalent in meaning to that of a reference translation [38].

***Precision and Recall*:** Machine translation can be evaluated using precision, recall, and the F-measure. These standard measures have significantly higher correlation with human judgments than recently proposed alternatives. More importantly, the standard measures have a natural interpretation, which can facilitate insights into how MT systems might be improved [35]. The evaluation is performed by using F-1 it is the measure of a grammar's accuracy rates through grammars and Dictionary written for the given sentences on the scale of (0.1 to 1.0).

Word order: The order of the word in MT is very important to ensure the quality of translation. However, the language diversity also allows different structures of the sentence. Successfully achieving the penalty on wrong ordination rather than on the properly completely different order, the candidate sentence that has completely different ordination with the reference is well structured, attracts a lot of interests from researchers [35].

2.11. Related Works

In this section discussed a review of MT research works that are studied previously in Domestic and overseas on dissimilar languages, what has been done on different languages by different persons in

different countries and what was the results and the gaps, finally UNL based research is considered as one types of Interlingua for MT approaches.

- **Research activities in UNL**

MT in Ethiopia is relatively new idea it has been only a few decades old which is comparing than other countries; it has a long way to go further steps this can be very advantageous for Ethiopia researchers to learn from others experience. Various MT systems introduces in many researchers with different approaches, Earlier, in Ethiopia, different dictionary-based CLIR for Amharic to other language experiments were conducted at different place by different study, the first Amharic-UNL based research was conducted at 2016 "Implement Amharic sentence generates from a given input unl expression", with the generation of NLG impact of the UNL of the SL and TL. There were a UNL dictionary and grammar was used to translate UNL expression to Amharic sentences. NLG from Interlingua representation Amharic sentence automatically generate from a given UNL input with in the java program excluding the web portal and Enco function by Kibrewossen yitbarek, 2016, at Addis Abeba University. [39]

Another similar UNL based study is "Information Retrieval System using UNL for Multilingual Question Answering" experiment was conducted by Kanu Goel at June 2016. Presently, question answering systems do not provide support for multiple languages, the experiment was conducted at a very compelling use case where a problem is solved in the Interlingua approach. This system acts as bridges used like Interlingua to avoid the language barrier among dissimilar languages. The system can be integrated with the Eugene framework as a whole ensure that the system finally gives the output in the NL rather than in the UNL format. [40]. On the other hand, a lot of study was done; the best study has been undertaken to CLIR systems for many Ethiopian local languages each other or with a foreign Languages that are available on the internet and university archive [40].

- **Research Activities in Enco and Deco**

Research on UNL has different approaches; these are development of Enco and Deco modules applications in other contexts such as KB representation, KB management, multilingual search engines, MT and language-independent Universal Digital Library etc.

- A prototype system for converting Brazilian Portuguese into UNL and reconverting from UNL into Brazilian Portuguese with 'Enco' and 'Deco' tools, respectively proposed by [41]. This system consists of three important modules, the lexical, the syntactic and the semantic.

- They are observed the 'Enco' and Universal Parser tools provided by UNDL require inputs from a human expert who is rarely available and the performance is not quite adequate, they have proposed new 'HERMETO' system which converts English and Brazilian Portuguese into UNL [38]. It has an interface with debugging and editing facilities along with its high level syntactic and semantic grammar that make it easier user friendly.
- UNL system evolving a Semantically Relatable Sequence (SRS) based approach has been proposed a conversion system to convert Punjabi language to UNL by [42], [17].
- The integration of 'Ariane-G5' system proposed for French Enco and Deco, it is a generator of MT systems. In this system, Enco takes place in two steps; it is analysis of the French text to produce the representation of its meaning in the form of a dependency tree, and is lexical structural transfer from the dependency tree to an equivalent UNL graph [43].
- A multifunctional linguistic processor, 'ETAP-3', as an extension of 'ETAP' MT system to a UNL based Multi-lingua MT system has been proposed by [44].
- The computational analysis of complex case structure of Bengali for a UNL based MT. They provided the details of the rule theory of 'Enco' and 'Deco' tools which are driven by analysis rules and generation rules respectively [45]. And, have been proposed a framework for converting Bangla to UNL and have also proposed a procedure to construct Bangla to UNL dictionary. The system developed by [46] uses ant colony algorithm for semantic analysis and fuzzy UNL graphs for conversion process [45].

• Related works

The following table it shows the recent and much closed papers related to MT are labeled here: Table.1. Define the Contributions and gaps of the related work.

Table 7: Related Work

Author, year and Title	***Finding***	***Method***	***Gaps***
Taghbalout et.al 2015. [Amazigh Noun Inflection in the Universal Networking Language]	-To build the first Amazigh dictionary for the UNL project. - Successfully integrated into the UNL framework.	UNL	-Now treated only nouns, verbs and particles categories are not included. -future plan is to focus on inflectional paradigms of verbs and to undertake the mapping between items in

			corresponding UW
Md. Zakir Hossain et.al, 2010. [Some Proposed Standard Models for Bangla Dictionary Entries of Bangla Morphemes for UNL]	-to Models for Dictionary, Grammatical Attributes of Bangla words Roots, Krit Prottoy and Kria Bivokti. -to make a converting the Bangla to UNL and vice versa drop barrier for Bangla to other.	UNL	-limited number of words and rules is considered in this paper. -it shows the designed model works perfectly for Bangla words. -All the Bangla words and rules are thought-about in future
Muhammad Firoz et.al, 2010. [Generation of Attributes for Bangla Words for (UNL)]	-it contributes morphology of Bangla roots words, Primary suffixes and developing grammars. -make Enco and Deco for Bangla. -it contributes morphology of Bangla roots words, Primary suffixes and developing grammars. -make Enco and Deco.	UNL	-future plan is to build a Bangla language server that will contain a complete Bangla Word Dictionary.
Kibrewossen yitbarek.2016 [Amharic sentence generation from interlingua representation]	-Implement Amharic sentence generates from a given input unl expression.	java platform code	-only done Amharic sentence generate or unl Deconverter without the UNL portal web framework.
Kanu Goel June 2016 [Information Retrieval System using UNL for Multilingual Question Answering]	-previously question answering systems do not support for multiple languages. -it gives a compelling use case where a problem is solved in UNL.	UNL	-The system can be integrated with the Eugene framework as a whole ensure that the system finally gives the output in the NL rather than in the UNL format.

- ## Summary

The requirement for construction a SMT system for Amharic language is needed the availability of large parallel corpora for a training set and a test set. Amharic doesn't own this type of corpus over UNL archive, that's why we need to have chosen the UNL Interlingua approach for this study. The most challenging task is to design a multilingual environment where many languages are to be translated, instead of using SMT approach an Interlingua is more preferable. UNL is an electronic language for computers to specific and exchange all types of knowledge. Researchers around the world have towards developing a system to overcome language barriers. While many systems developed by various organizations each has its special representation of a given language. This results incompatibilities between systems still now impossible to break language barriers all over the world, even if all the results are combined in one system. Against this backdrop, the concept of UNL as a common language for all computer systems emerged to avoid the language barriers on internet and web. It is an intermediate language the concept of relations among words associated with in attribute to understands a sentence rather than translating into another, IAN framework is used to analyze a query and gives us a freedom to check the results more than one language with a single interface at the same time and the system can be integrated with the Eugene framework finally gives the output in the NL rather than in the UNL format.

Chapter Three

3. Research Methodology

In this chapter the methodology of the conducted research is presented. It starts by presenting the research design, followed by the methods and procedures of data collection.

3.1. Overview

To achieve the objectives has been followed to implement the proposed system. These procedures are implemented the main objective of research work is to create UNlization , Nlization by using IAN and EUGEAN tools for Amharic local language detail of this is given below:-

- Relevant document were reviewed related to UNL and MT systems
- Taking online Certificate of Proficiency CUP500 and CLEA1000: in UNL through VALERIE is to understand deeply the concepts of UNlization and Nlization
- To perform the analysis of framework: It is a document certifying that has reached a standard of knowledge of UNL concern.
- Data collection and analysis used for comparing tools and for preparing corpus, dictionary and grammar.
- Localization of Grammar and Dictionary to Amharic Local Language under UNL.
- Detail understanding of IAN tools for Enco of Amharic language into Unl expressions and EUGEAN tools for Deco of UNL expressions to Amharic language and design the model.
- Translation (manually) of UC-A1 corpus into Amharic text is an input for UNL, the words appearing in the resource was made dictionary and grammar have been done for using UNlization and Nlization of Amharic Sentence.
- By analyzing and generation features of Amharic language with Trules paradigms have been done, the generated actual output compared to expected output and
- To display the final evaluation of F1 -measure it is used to calculate the accuracy of the grammars which is precision and recall.

3.2. Review of Related Literature

A lot of manuals, books, papers, Slides and examples can be accessed from the official website of UNL [17]. From UNL data in sequential order can be obtained in the form of UNL Documents.

These constitute various UNL converted from NL topics like Biographies, Wikipedia Articles, UNESCO Documents etc. These can be used for various research related activities.

3.3. Data Collection Methods

The researcher used to collect data for comparison support among UNL with other MT approaches and to document analysis conclusion. Moreover, interview with concerned bodies; experts from NLP areas accordingly to identify the data type. Finally, prepared the corpus AC-U1 data a set of 100 sentences provided by UNDL foundation it was adopted localization from English text into Amharic text. Frequently, visiting an official website of UNL for collecting useful information that used on different UNL platforms resources [47]. Extremely, so many papers reviewed on related to the deployment of UNL components like IAN, EUGEAN and others tools, useful data was gathered for further enhancement of this study about Amharic native language from previously conducted researches locally and abroad archive seriously.

3.4. Data Analysis Techniques

The UNL framework is the packaging tools for studying various UNL applications used to integrate any natural language to UNL framework. The Enco and Deco are the core software of the UNL framework for MT and other purposes; these tools are responsible for developing the coaching corpus of uc-a1 data, mapping the dictionary words that appear in the training corpus and developing a transformation grammar for completion of implementation of the proposed system. This study useful in different specific tasks to perform data analysis, search and generation of any natural language for end user's needs. For instance, in building the resources such as free Lingware dictionaries and grammars can be shared other users. This Lingware are later used by professional linguists, computational linguists and non-professional users most of these tools are shareware, in the sense that the source code can be customize or develop through experts. Accordingly, experts are used freely and provided to other users in order to have access to these software, experts are supposed to sign the UNL Development Set agreement [7].

The researcher used to implement the tools that are UNL's applications of Unlization to reuse it for the process of encoding Amharic language into UNL expression through "IAN tools and Nlization to decoding UNL expression into an Amharic through "EUGEAN tools. The corpus data were used UC-A1 contains a set of 100 Amharic sentences provided by UNDL foundation manually translate into Amharic from the English version, it make a dictionary entries mapping based on the UWs

found in the corpus and done grammatical inflections in order to sense language structure T-rules and D-rules per limited instructions. Finally, tried to evaluate the F-1 score measure the grammar accuracy on precision and recall results and display the comparison of Enco and Deco process of Amharic language the output results are shown in the output consol.

3.5. Localization

Localization is that the method of adapting dictionaries and grammars to a particular language, which is referred to as "locale". In the UNL framework, localization of existing resources is one of the strategies that can be adopted in order to create basic dictionaries and grammars required to process uc-a1 corpora [12]. It is the process of familiarizing dictionaries and grammars from the English version to Amharic language In the UNL; it is one of the techniques avoiding doing resources from the scratch, in order to create basic Enco and Deco process.

3.6. Overview of UNL Tools and Materials

Different free software's and designing tools that are used online an open source UNL Framework Implementation, these tools are used the UNL application framework operation it consists of:

Table 8: Overview of UNL System

Types of software	***Name and Descriptions***
Language Resources	• UNLKB–Linguistic Knowledge on concepts that is common to every language. • Universal Word Dictionary, analysis and generation rules.
Language Servers	• DeConverter - automatically Deconvert UNL into NL. • EnConverter - interactively Enconverter NL text into UNL.
Software tools	• UNL Editors - used to make UNL documents. • UNL Explorers - used to view UNL documents. • UNL Verifiers – verifies UNL expression for correctness. • UNL Proxy servers – Provides with language servers. • Concept Definitions –concepts is connected to other concepts. • UNL Documents - UNL term is described each sentence into NL.

3.7. Evaluation Techniques and Testing

This study had been evaluated for ensuring the performance of the system through experts with test case. This processes tells how proposed system accurate and significant to the Amharic language end users. The experimental testing is concerned with how system addresses the needs of the Amharic Language users for searching and retrieved data from different languages. This was done by selected domain expert evaluators along with questionnaires and based on experimental results analysis in detailed. The testing tools have been done in a set of 50 Amharic sentences with dictionary and limited grammatical rules applying encoding from an Amharic into UNL and decoding Amharic outputs out of UNL. The final result will be share in the online community the dictionary, T-rules and D-rules grammar to expand for checking such as language specialists demonstrate semantic and grammatical structure of sentences. Finally, tried to compare F-1 measure which is calculated the grammar accuracy of precision and recall between 0 and 1 respectively, based on the example sentences during in the process of Enco and deco functions the rate of sentences is checked when compare at the source and target sentences translation [12].

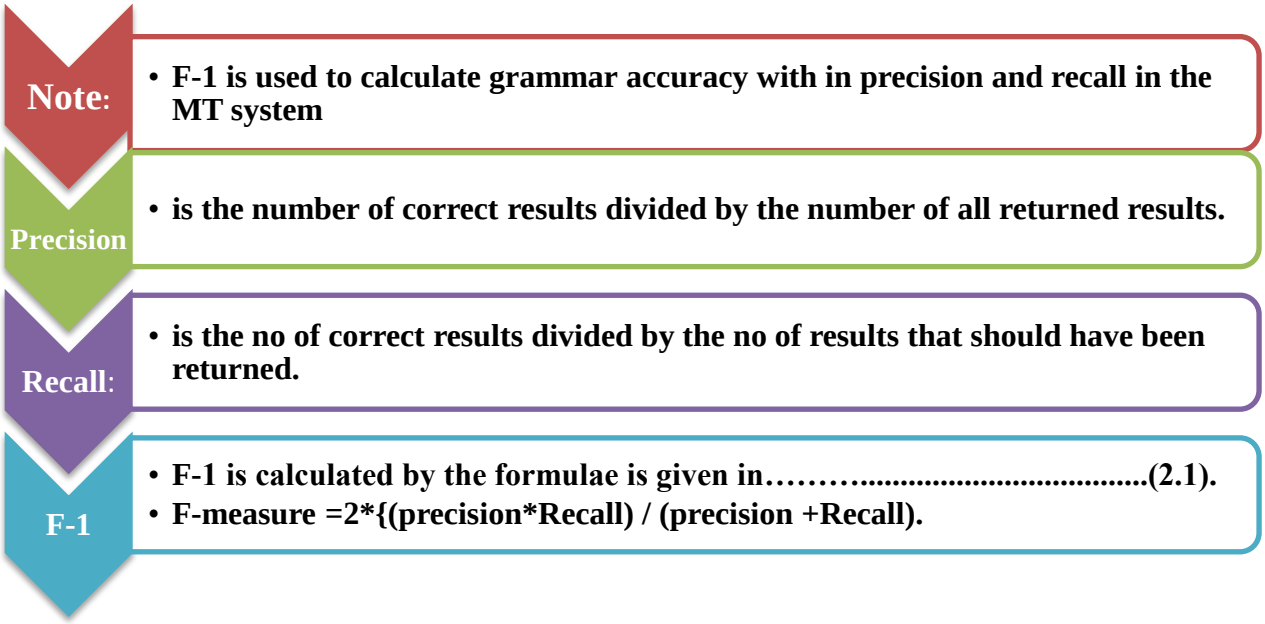

Figure 3: F-1 measure

Chapter Four

4. UNL Framework and Amharic Language Structure

In this chapter discussed about the general UNL framework architecture of the IAN and EUGENE tools and Amharic language Structures and unique characteristics is explained.

4.1. UNL Platform

The UNL Platform it is an Interlingua system development in web application. The system provides an integrated environment that users can produce UNL Documents from NL and vice versa. To producing UNL expressions it involves enconverting process from a given NL. Based on the user needs, to meet the requirements different levels of functions will be expected. Even, users can directly enconvert sentences of any language from a HTML formatted document into UNL and create a TL HTML document by deconverting the UNL into the TL. In the simple mode of enconverting process, word selection can be manually done by users in a simple way. In the skilled mode, modifications on dictionary entries of words or morphemes are possible in word selection. With this operate users will turn out correct UNL Expressions. UNL Expressions is altered victimization UNL graphic editor. The UNL graphic editor consults the UNL Ontology to verify the results of modifications source [48], [49].

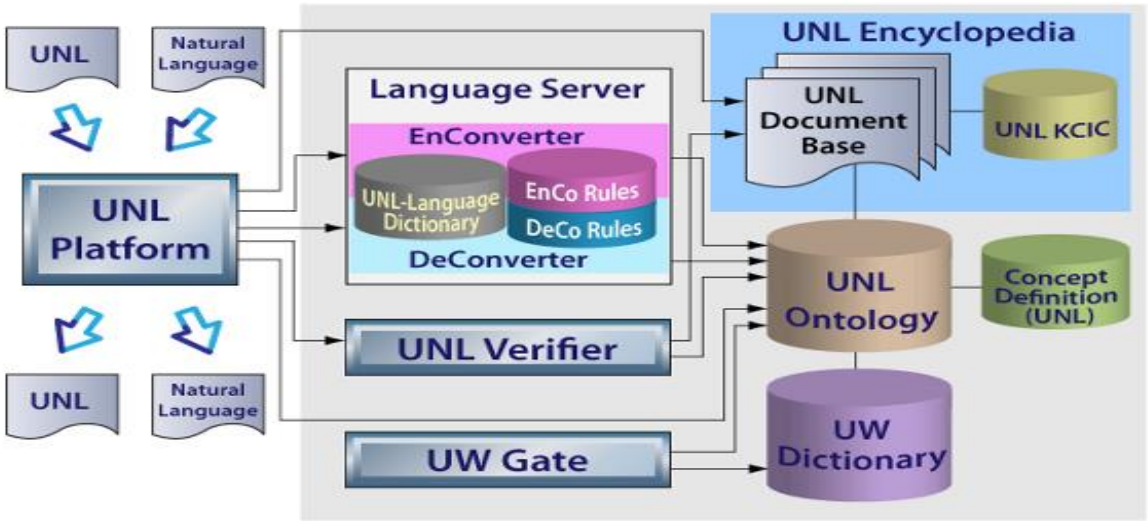

Figure 4: UNL Platform General Structure source [48], [49].

4.2. Mechanism of Conversion Between Languages and UNL

The EnConverter and DeConverter are the core software in the UNL System. The EnConverter converts NL sentences into UNL expressions. The Universal Parser (UP) is a specialized version of the EnConverter. It generates UNL Expressions from annotated sentences with concerning the UW wordbook while not exploitation grammatical options. All UNL Expressions are verified by the UNL Verifier, and then to be stored in the format of UNL Document. The DeConverter converts UNL Expressions to NL sentences. Both the EnConverter and DeConverter perform their functions based on a set of grammar rules and a word dictionary of a TL. Whether consulting the UNL Ontology and/or a co-occurrence dictionary in EnConverter or DeConverter is optional. Arrows in solid line show dataflow and arrows in broken line show access. Fig.4.2. shows the mechanism how the conversion between a NL and a UNL Document is carried out in the UNL System source [48], [49].

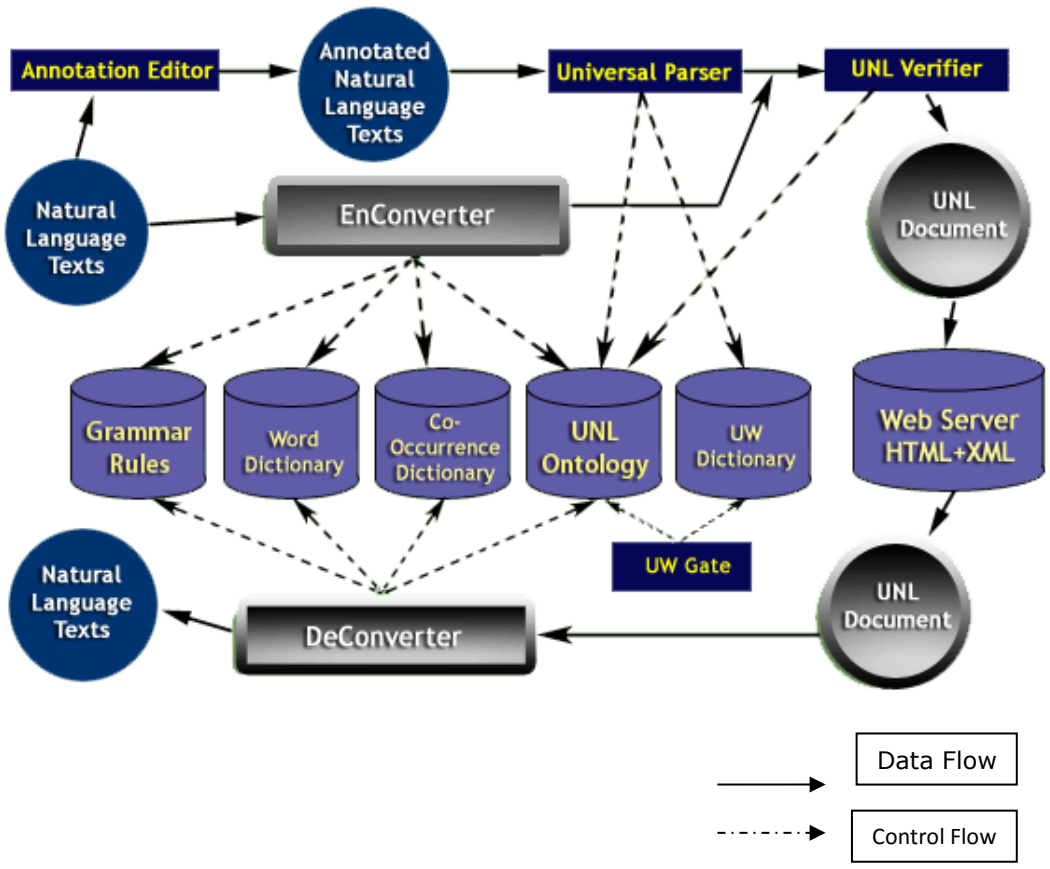

Figure 5: Conversion mechanism of the UNL System source [48], [49].

UNL web

It is a web portal system that is created and maintained by the UNDL foundation to facilitate the research accepted in the field of NLP development and reduce the language barriers. It is a property of United Nation (UN) it serves as an asset to bring together the developers and researchers onto one integrated platform for all NL. It consists of five basic systems as following [12].

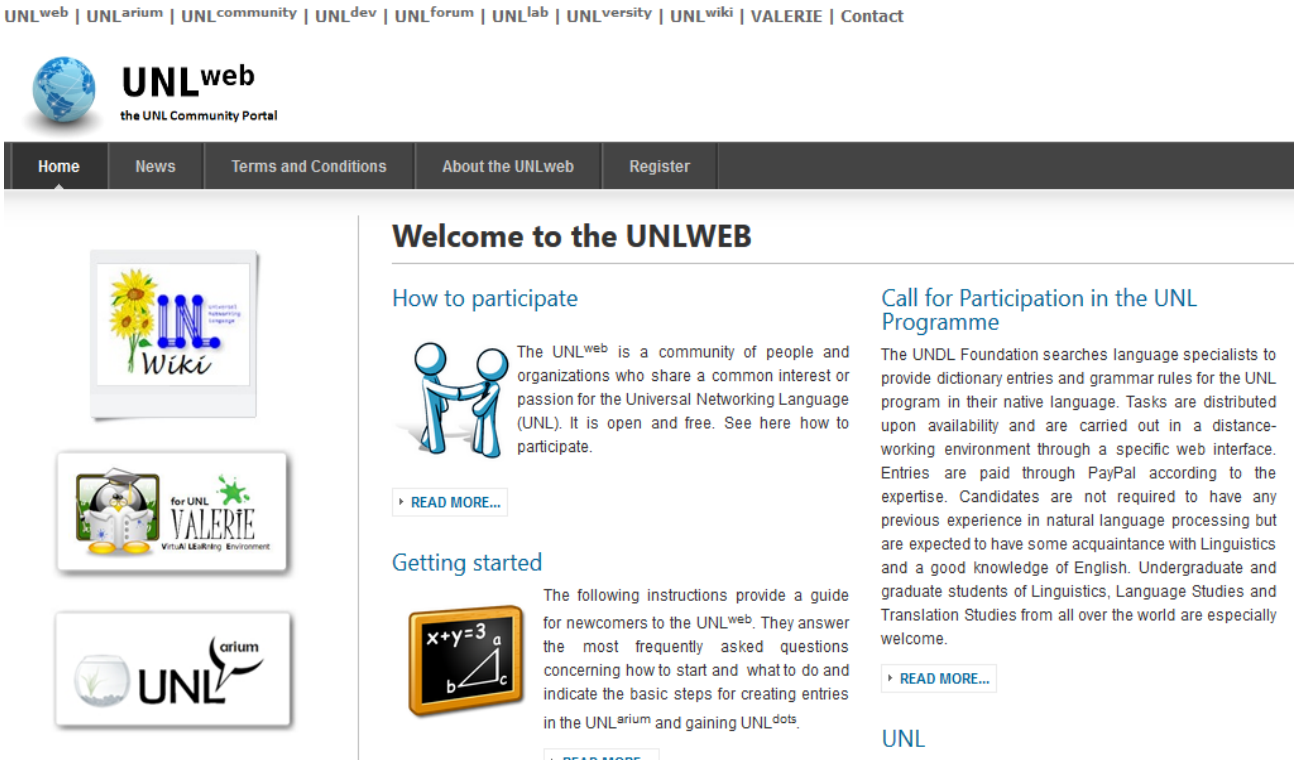

Figure 6: UNL web

- The UNLarium, it provides UNL development environment to create and edit resources.
- The UNLdev, it is facilitate the computer programs applications related to the framework.
- The UNLwiki, it providing the necessary documentation and instructions for carrying out UNL related activities.
- The UNLforum, it is a social portal to discuss the issues relating to UNL framework and spread the information.
- VALERIE, a virtual learning environment for teaching and Training for research developers in UNL platform.

4.3. UNL Editor

It is a UNL writing tool provides the graphic interface to allow users manipulate high-level graphs instead of low-level UNL statements. However, the language specialists are required to upload the texts to be analysed, the writer selecting the corresponding UWs creating semantic relations among nodes and assigning attributes to the nodes in the graph [12].

4.4. Enco and Deco modules

According to 3rd workshop, , the first variance of UNL from existing Interlingua is described well but, it was not the only application of MT, is used to serve various NLP purposes as a language for meaning representation [28]. Another difference is to producing a UNL text is not supposed to be fully automatic it should be an interactive process divided the computer and a human expert "writer" [37]. Due to the interactive Enco tools which serves as input text can be made as good as user's need, the UNL writer will correct results if there were an error and eliminate the remaining ambiguities. The users can show outputs by using a Deco tools into his owns language to test the validity of the UNL expression and then improve it again until fully satisfied with the final result. Along with the various applications of UNL like UNlization and Nlization are attracting many researchers towards UNL this module: Using common Enco and Deco tools provided by the UNDL foundation, Integration UNL into different pre-existing MT approaches and Creating new universal language architecture from scratch [12].

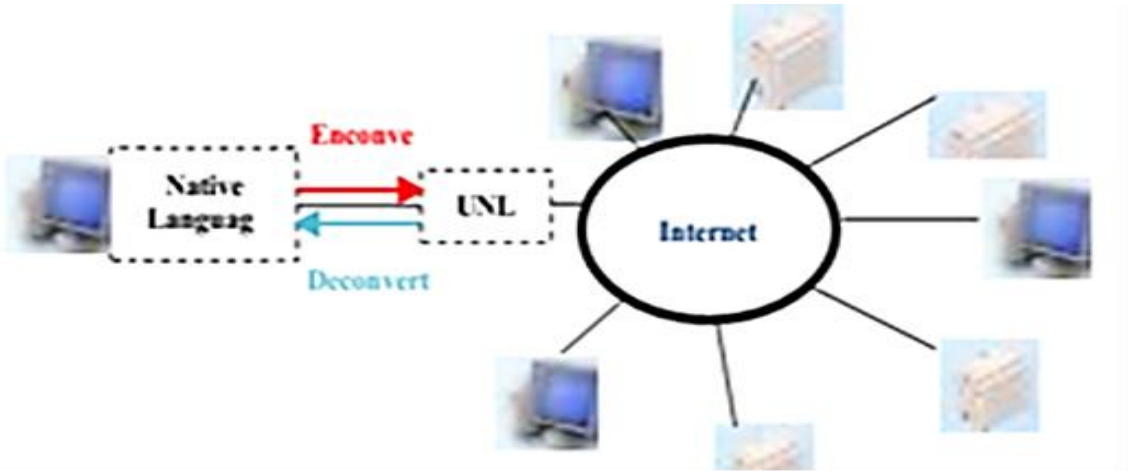

Figure 7: Enco and Deco modules

4.5. Amharic Language Characteristics

Amharic language: Amharic language is second largest most-spoken Semitic language in the world next to Arabic [13]. It is the official working language of government of Ethiopia, among 89 languages which are registered in the country. It uses a unique script, which has originated from ancient language, the Ge‟ez alphabet; it is the liturgical language of the Ethiopian Orthodox Church. It has been the working language of the government, the military, and of the Ethiopian Orthodox Tewahedo Church throughout modern times. [50]. The majority of the speakers of Amharic can be found in different parts of the country, It has five dialectical variations across different parts of the country [14]. These include dialects like Addis Ababa, Gojam, Gonder, Wollo and Menz [15]. But there are also Amharic speakers in other countries, Outside Ethiopia; it is used in Egypt, Israel, Canada, Sweden, Eritrea, and the US. Thus, it has official status is spoken by many people as their native and second language. In addition, it is a language with many literatures [50].

Amharic writing system : Unlike other Semitic languages, such as Arabic and Hebrew, Amharic /ʾämarəña/ script uses a grapheme based writing system called fidel /fidälə/ which is written and read from left to right [16]. Modern Amharic has inherited its writing system from Ge'ez /gəʾəzə/, which is still the classical and orthodox church language of Ethiopia [51] [15].

Amharic language structures: Amharic has a complex morphology which combines consonantal roots and vowels words. Amharic and English differ largely in their morphology, syntax and the writing system they use. Therefore, the search engines that square measure primarily developed for English cannot expeditiously be accustomed retrieve Amharic documents. Amharic language has the following characteristics that should be considered in IR and MT systems [50].

- Morphological variations: is morphologically complex words are inflected with prefixes, suffixes and infixes.
- Character Redundancy: it has some redundant symbols with the same sound in its alphabet. Some of the characters can be used interchangeably without any meaning difference in the language.
- Short Words: it is common to shorten some words using the forward slash („/‟) and the English full stop (ነ ጥቦ (.)).
- Variations due to Pronunciations: Some words have multiple spelling variants. The main reason for such problems is due to the multiple regional dialects of the spoken language. In

addition, most of the words which are adopted from English or other languages are written in different formats there is no standardization in spelling for such words which in turn result in huge variation in transliteration. These variations in spelling are difficult for Web IR applications such as Google, which retrieves documents based on exact term matching.

- Stop-words: Most search engines and query MT systems do not consider extremely common words in order to save disk space or to speed up search results. Although Amharic does not have standard stop word lists are considered to be stop words.
- Amharic Punctuation Marks: there are different punctuation marks and among these punctuation marks ። , ፣ , ፤ and? Are the most commonly used ones. They are used in query or document tokenization as word delimiters in addition to white spaces [50].

4.5.1. Amharic Morphology

The Amharic language has been declared to have word categories as **ስም**(noun), **ግስ** (verb), **ቅፅል** (adjective), **ተውሳከ ግስ** (Adverb), **መስተዋድድ**(preposition), and **ተውላጠ ስም (**pronoun)[[52].

Noun: a word will be categorized as a noun, if it can be pluralized by adding the suffix **ኦች/ዎች** and used as nominating something like person, animal, and so on.

Verb: Any word which can be placed at the end of a sentence and which can accept suffixes as/**ህ/,/ሁ/,/ሽ/**, etc. which is used to indicate masculine, feminine, and plurality is classified as a verb.

Adjective: any word that qualify a noun or an adverb, which actually comes before a noun**(e.g. ጎበዝ ተማሪ)** and after an adverb**(በጣም ጎበዝ).** Other specific property of adjectives is, when pluralized, it will repeat the previous letter of the last letter for the word **(e.g. ረዥም--› ረዣዥም).**

Adverb: it will be used to qualify the verb by adding extra idea on the sentence. The Amharic adverbs are limited in number and include **ትናንት, ገና, ዛሬ, ቶሎ, ምንኛ, ክፉኛ,እንደገና** and **ልጅኛ.**

Preposition: preposition is a word which can be placed before a noun and perform adverbial operations related to place, time, cause and so on; which can't accept any suffix or prefix; and which is never used to create a new word. It includes **ከ፣ለ፣ወደ፣ስለ፣እንደ…**

Pronoun: this category further can be divided as deictic specifier, which includes **ይህ, ያ, እሱ, እስዋ, እኔ, አንተ, አንች…;** quantitative specifier, which includes **አንድ, አንዳንድ, ብዙ, ጥቂት, በጣም…;** and possession specifier such as **የእኔ, የአንተ, የእሱ…** [53], [52].

- Especially in its verbs, Amharic uses the root-pattern morphological phenomenon. The words in the language consist of stems and affixes. Note that the stem forms can take on various affixes [54]. For example:
  1. The subject is marked on the verb using subject suffix pronouns, as in **ሰበርኩ**-' I broke'; **ሰበርን** - 'we broke'; **ሰበርክ** - 'you(second person masculine) broke'; **ሰበርሽ** - 'you(second person feminine) broke'; **ሰበሩ** - 'you(second person respected) broke'; **ሰበራችሁ** - 'you(second person plural) broke'; **ሰበረ** - 'he broke'; **ሰበረች** - 'she broke'; **ሰበሩ** - 'they broke',
  2. The direct object is marked on the verb, as in **ሰበረኝ** - 'he broke me'; **ሰበረን** - 'he broke us'; **ሰበረህ**- 'he broke you(second person masculine)'; **ሰበረሽ** - 'he broke you(second person feminine)'; **ሰበሮት** - 'he broke you(second person respected)';**ሰበራችሁ**-'he broke you(second person plural)'; **ሰበረዉ** - 'he broke him'; **ሰበራት** - 'he broke it'; **ሰበራቸዉ** - 'he broke them',
  3. Some prepositional phrase complements are optionally marked on the verb as in **እስክትመጣ ጠብቃለሁ** - 'I wait until she comes' or 'I wait until you (second person masculine) come'.
  4. functional elements like negation maker, conjunction and some auxiliary verbs are also bound morphemes that are attached to the verb:
     - **አልሰበርም** - 'I will not be broken'
     - **ሲሰብር** - 'while he was breaking'
     - **ተሰብሬአለሁ** - 'I have been broken'

Verb arguments may be indicated by suffix pronouns and a word may stand alone as a sentence. Illustration: **ተሰብሯል**- 'It is broken'

The definite article in Amharic is a bound morpheme which is attached to a noun or to the first inflected element in a noun phrase. Illustration: **የተሰበረዉ** - 'The one that is broken'

4.5.2. Amharic Punctuation Marks and Numerals

In Amharic, there are different punctuation marks used for different purposes. In the old scripture, a colon (**ሁለት ነጥብ**፡) has been used to separate two words. These days the two dots are replaced with whitespace. An end of a statement is marked with four dots **(አራትነጥብ**።) while **ነጠላሰረዝ**(፣ or ፥) is used to separate lists or ideas just like the comma in English.

In Amharic, numbers can be represented using Arabic symbols. It has also its own number representations, Ethiopic number representations. Similarly numbers can be represented in a word alphanumerically. Table 3.1 shows the Arabic, Amharic and alphanumeric representation of numbers from [32].

Table 9: Number Representations in Amharic adapted from [48].

Arabic	Ethiopic	Alphanumeric	Arabic	Ethiopic	Alphanumeric
1	፩	**አንድ**	20	፳	**ሃያ**
2	፪	**ሁለት**	30	፴	**ሰላሳ**
3	፫	**ሦስት**	40	፵	**አርባ**
4	፬	**አራት**	50	፶	**አምሳ/ሀምሳ**
5	፭	**አምስት**	60	፷	**ስልሳ/ስድሳ**
6	፮	**ስድስት**	70	፸	**ሰባ**
7	፯	**ሰባት**	80	፹	**ሰማኒያ**
8	፰	**ስምንት**	90	፺	**ዘጠና**
9	፱	**ዘጠኝ**	100	፻	**መቶ**
10	፲	**አስር**	1000	፲፻	**ሺ/ሽህ**

Table 10: Amharic fraction and Ordinal representation adapted from [48].

Fraction	Amharci representation	Ordinals	Representation
½	**ግማሽ**	1st	**አንደኛ/ቀዳማዊ**
1/3	**ሲሶ**	2nd	**ሁለተኛ/ዳግማዊ**
¼	**ሩብ/እርቦ**	3rd	**ሶስተኛ/ሳልስ**
2/3	**ሁለትሲሶ / ሁለትሶስተኛ**	4th	**አራተኛ/ራብዕ**
¾	**ሶስትአራተኛ**		
1/10	**አስራት**		
2X	**እጥፍ**	9th	**ዘጠንኛ/ዘጠነኛ**
2.X	**ሁለት ነጥብ X**	10th	**አስረኛ**

On the above table we can easily observe that in the collected text if there exist fraction and ordinal representations are found it is normalised into the equivalent Amharic representation.

4.5.3. Amharic Syntax

In the main and subordinate clauses of Amharic the normal word order is SOV, S = Subject, O = direct Object and V = Verb. This is unlike English which is SVO or VSO in Classical Arabic. Modifiers generally precede the words they modify. For example, the difference between Amharic and English, in this aspect, can be seen in the construction of relative clauses such as: **ሰለሞን ደብዳቤውን ፃፈ** which means, 'Solomon wrote the letter'. Its word-to-word translation is "Solomon the letter wrote". The other characteristics of Amharic syntax, as mentioned by [54] are:

1. **Postpositions:** Amharic has a set of postpositions like, **ዉሰጥ** 'inside', **ለይ** 'on', **ገር** 'with', **አጠገብ** 'near'. Nouns followed by these postpositions usually have at the same time the prefix **በ** which bears the meaning of *in, at, by* or the prefix **ከ** which means *from*. For example: **ከጠረጴዛዉ ላይ ዉሰድ** - 'take from the table'; **ጠረጴዛዉ ላይ አስቀምጠዉ**- 'put it on the table'.
2. Limited use of the plural in nouns: Although Amharic has a morphological category of plural in nouns, the plural form is not usually used with numerals or words indicating quantity. In languages such as English, if a numeral other than 'one' is used with a noun, then the noun must be in the plural, e.g. five houses. Amharic allows phrases like *five house* - ***አምስት ቤት***, *many house* - ***ብዙ ቤት*** to mean five houses and many houses, respectively. Note that the plural form of ***ቤት*** is ***ቤቶች***.
3. There is a tendency for verbs in subordinate clauses to lack the full specification of tense, person, and so on, to be made explicit in the main clause. The Amharic verb has a system of five tenses used in the main clauses:

    a) Present-future - **ይሰብራል** 'he breaks, is breaking, will break'
    b) Past - **ሰበረ** 'he broke'
    c) Past Continuous - **የሰብር ነበር** 'he was breaking'
    d) Perfect - **ሰብረል** 'he has broken'
    e) Past Perfect - **ሰብሮ ነበር** 'he had broken'

Normally, in subordinate clauses the system is reduced so that only tenses (a) and (b) are used and tense (a) has a special 'short' form [54].

- The Amharic basic sentence is constructed from noun phrase and verb phrase **(ስማዊ ሀረግ+ ግሳዊ ሀረግ)** [53]**.**

Sentence = noun_phrase + Verb_Phrase.

For example, the sentence: **"ሁለት ትልልቅ ልጆች ትናንት በሞኪና ወደ አዳማ ሄዱ።"** have noun phrase **"ሁለት ትልልቅ ልጆች"** and verb phrase **"ትናንት በሞኪና ወደ አዳማ ሄዱ"**.

The statement **(አረፍተ-ነገር)** will have the noun phrase and verb phrase combinations. The noun phrase and the verb phrase further will be divided to different particles such as other sub noun phrase and verb phrase, noun, adjectives, specifier and so on… [54], [55] , [53], [52].

Chapter Five

5. Design and Development of Unlization and Nlization for Amharic

This chapter presents experimental analysis of the implementation of Unlization and Nlization, preparing the corpus; makes it dictionary and grammar for Amharic local language into UNL framework.

5.1. Design of UNlization and Nlization for Amharic Language

Each of the integrated languages is expected to make use of the UNL system components and resources to fit into the UNL system. The Amharic UNL system is responsible for developing the Amharic resources necessary to encode and decode of Amharic utterances into and from UNL. The resources required corpus, include Amharic analysis and generation dictionaries as well as Amharic analysis and generation grammars. Now, it had been built for Amharic analysis and generation for both dictionaries and grammar was done [12].

IAN and EUGENE tool: Both IAN and Eugean tools is fully automatic tool it takes NL document as input and delivers an output in UNL and it takes UNL document as input and delivers an output in NL without any human intervention separately. IAN has 6 tabs and Eugean

Table 11: IAN's and Eugean tools

<table>
<tr><td rowspan="3">IAN</td><td colspan="6">Quick start</td></tr>
<tr><td colspan="6">• As part of the UNLdev, user must be registered in order to login.
• Interactive it takes NL as input and delivers an output in UNL without human.</td></tr>
<tr><td colspan="6">• IAN has 6 tabs</td></tr>
<tr><td>Welcome:</td><td>NLInput:</td><td>Dictionaries:</td><td>T-Rules:</td><td>DRules:</td><td>IAN console:</td></tr>
<tr><td rowspan="4">EUGEAN</td><td colspan="6">Quick start</td></tr>
<tr><td colspan="6">• As part of the UNLdev, user must be registered in order to log in.
• Automatic takes UNL as input and delivers an output in NL without human</td></tr>
<tr><td colspan="6">• Eugean has 5 tabs</td></tr>
<tr><td colspan="2">UNL input:</td><td>Dictionary:</td><td>T-Rules:</td><td>D-Rules:</td><td>EUGENE console:</td></tr>
</table>

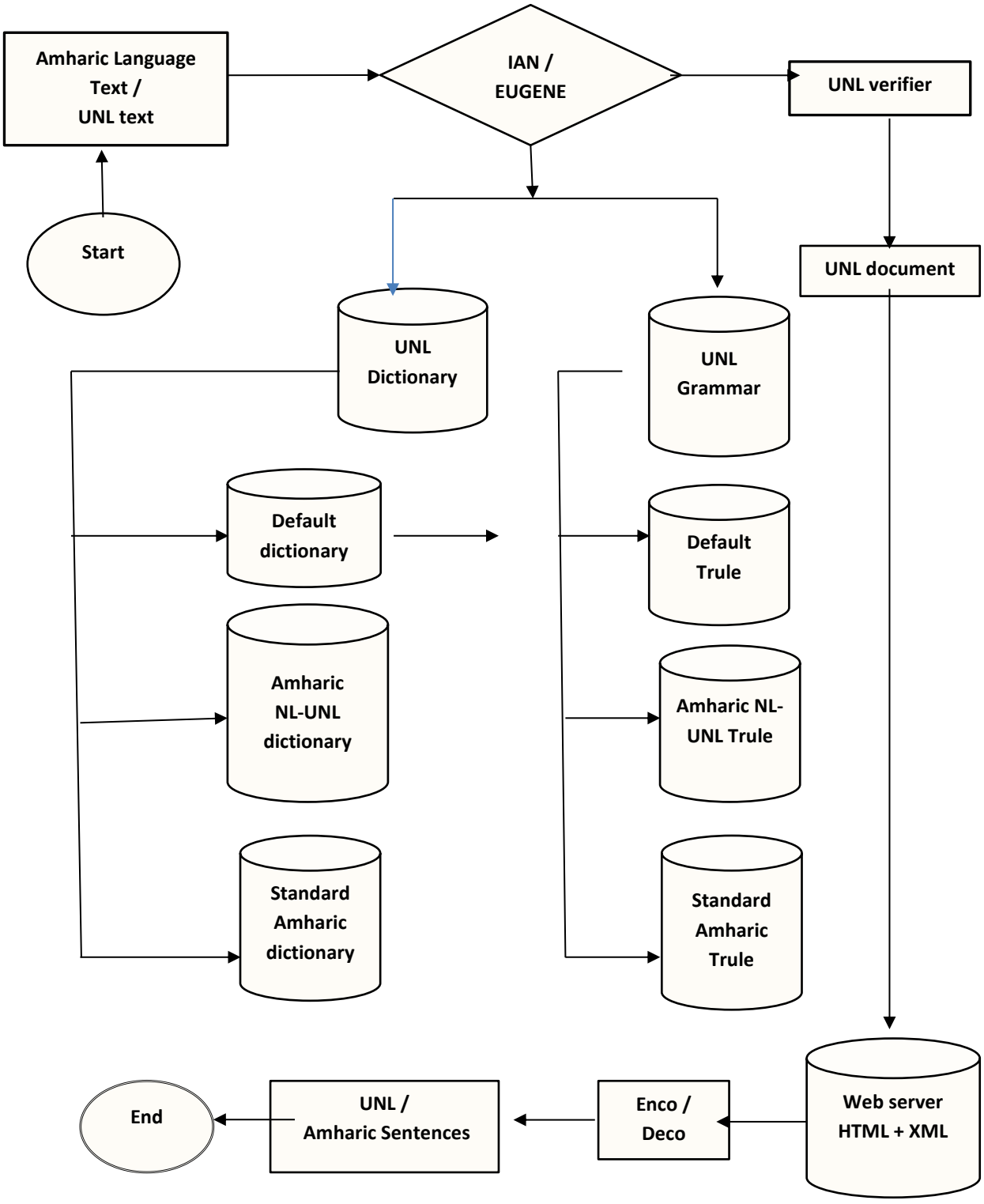

Figure 8: Enco and Deco components

IAN and EUGENE Structures and Techniques: IAN is the acronyms for Interactive Analyser systems takes NL input and delivers UNL, it includes a dictionary and grammar for NL analysis system. It represents NL sentences as semantic networks in the UNL format and EUGENE is the dep-to-surface natural language Generator it is fully automatic NLG system it takes a UNL input and delivers an output in NL without any human intervention. It generates NL sentences out of semantic networks represented in the UNL format. Both of these tools are in its current release, it is a web application developed in Java and available at the UNLdev. IAN and Eugean Interface***:*** Both are as a universal engine, IAN must be used to the SL and EUGENE must be used to the TL to be provided through IAN and EUGENE's interface separately [12].

Table 12: IAN's and Eugean Interface

IAN Interface	*EUGENE Interface*
IAN is used to the SL interface.	Eugean is used to the TL interface.
The input is NL document to be UNLized.	The input is document to be NLized
The NL-UNL analysis dictionary: UWs mapped into NL with corresponding features.	The UNL-NL generation dictionary: UWs mapped into NL with corresponding features.
The NL-UNL analysis TGrammar: A set of Trules convert NL sentences into UNL.	The UNL-NL generation TGrammar: a set of Trules convert the UNL into NL.
The NL-UNL analysis Dgrammar: A set of Drules improve the results of tokenization of the transformation.	The UNL-NL generation Dgrammar: a set of Drules improve the results of tokenization of the transformation.

5.2. UNLization for Amharic Language

EnConverter: To perform the encoding process is an online tool is called IAN (Interactive Analyser for UNLization): it is software developed in Java web environment to perform encoding application. It takes a NL input and provides UNL form and each NL can be integrated and stored on the server. The syntactic processing is done automatically through the grammar and dictionary rules provided for the specific NL. Therefore, data sources can be exploited anytime, anywhere, without being dependent on geographic distance. IAN analyzes the input with the assistance of T-Rules and dictionaries. [12]

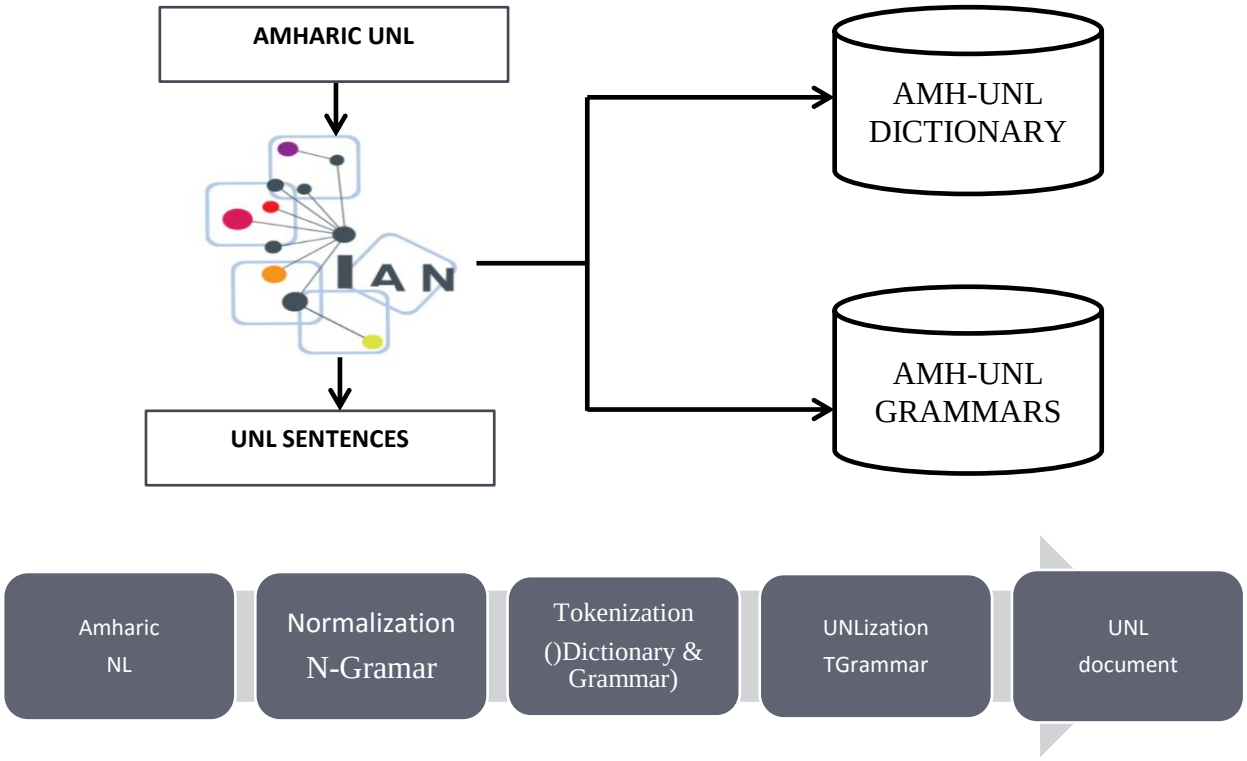

Figure 9: Mechanism for the processing of the NL sentence

5.3. NLization for Amharic

Deconverter: To perform the decoding process is an online tool is called EUGENE (from dep-to-surface generator for NLization): is an application developed in the Web environment to create sentences in NL from UNL expressions. I t is fully automatic NLG, which was developed by the UNDL organization like IAN it takes the UNL as the input and outputs the NL that is desired without any human intervention and it is language-independent and has to be parameterized to the NL input through a dictionary and a grammar, provided as separate interpretable files [12].

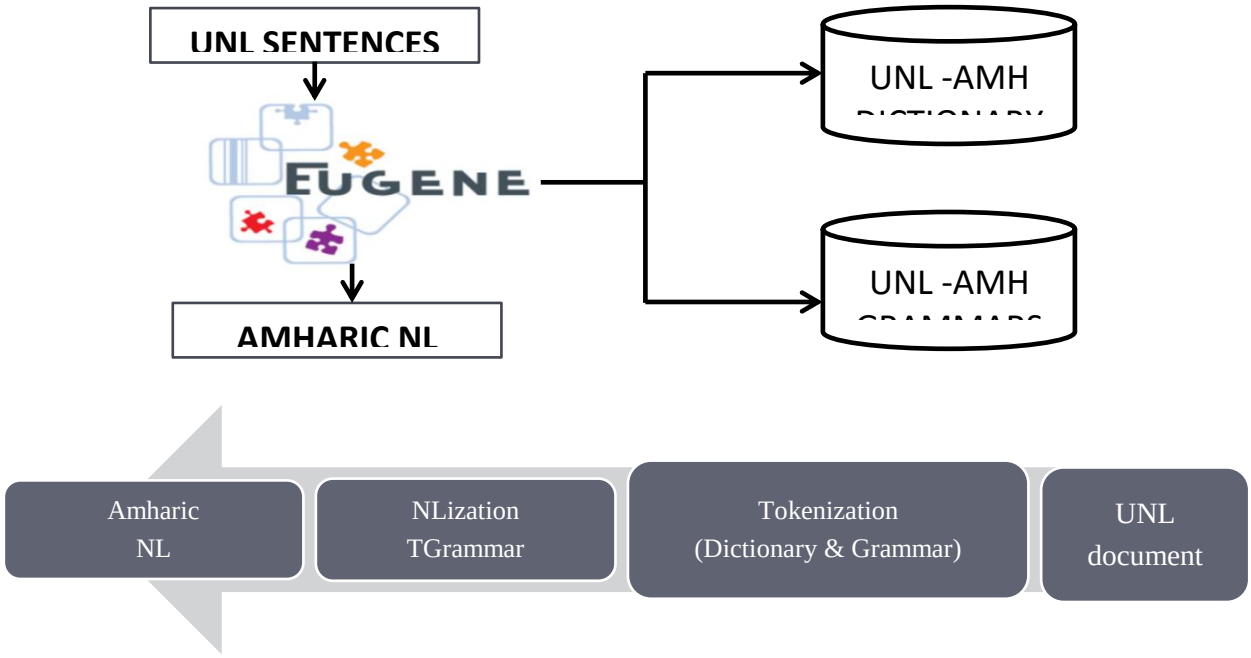

Figure 10: Mechanism for the processing of the NL sentence

5.4. Amharic Training corpus used for UNL

The Amharic Language corpus specification: The corpus uc-a1 was extracted from a simplified and translated version of "The Hare and the Tortoise", by Aesop [17]. It is Associate in Nursing experimental corpus accustomed prepare the initial versions of the Amharic descriptive linguistics for sentence-based UNLization and NLization, exploitation IAN and EUGENE [12].

- Uc-a1 in English, to be translated manually in order to be used as the input for the Unlization process (with IAN).
- Uc-a1 in UNL, to be used "as it is", as the input for the Nlization process (with EUGENE).

English	***Amharic***
hare	ጥንቸል
the hare	ጥንቸሉ
the tortoise	ኤሊው
a race	ውድድር
the slow pace	ቀርፋፋው ጉዞ
assented to the proposal	በጥያቄው ተስማማ

5.5. Amharic Dictionary used for UNL

The dictionaries, where NL-UNL dictionaries to be used in NL analysis are provided. There can be several different dictionaries, and can be loaded to process the same corpus, but be sure that they are loaded in the correct order (because the order of the entries within the lexicon will matter for tokenization). Reordering the dictionaries through the option "reorder dictionaries" at the top menu can be done. **Note**: Dictionaries area unit plain text files with one entry per line within the following format [12], [3].

[NLW] ”UW” (ATTR) < FLG, FRE, PRI >; COMMENTS

Here, NLW is the NL word; UW is the corresponding Universal Word, ATTR, the list of features of the NLW, FLG, The three-character language code according to ISO 639-3, FRE, and the frequency of NLW in natural texts. Used for NL analysis (NL-UNL). It can range from 0 (less frequent) to 255 (most frequent).PRI, the priority of the NLW used for NL generation (UNL-NL). It can range from 0 to 255.

Table 13: UNL Experiment drive system

UNLization and NLization		
Items	***Data type Enco***	***Data type Deco***
NL input: uploaded NL input.	UCA1_Amharic.txt (corpus UCA1-Amh).	UCA1_unl_Amharic.txt
Dictionaries: uploaded Dictionaries.	-amh_unl_dic.txt (entries appearing in UCA1) -Default Dictionary (blank space, punctuation signs and other generic entries)	-unl_amh_dic.txt (entries appearing in UCA1) -Default Dictionary (blank space, punctuation signs and other generic entries)
N-grammar: uploaded "N-rules".	-Normalization Grammar (used to normalize the input text)	- Normalization Grammar (used to normalize the input text)
T-grammars: uploaded T-rules.	- Standardization Grammar (used to standardize the dictionary) -AMH-UNL T-Grammar	-Standardization Grammar (used to standardize the dictionary entries) -UNL-AMH T-Grammar

	(language-specific rules) -	(language-specif-rules)
	-Default T-Grammar (generic rules)	-Default T-Grammar (generic rules)
D-grammar: uploaded D-rules.	amh_unl_dgrammar.txt (d rules)	UNL-AMH D-Grammar (DRules).

SCENARIO 1: NOMINAL CONCEPTS

[ፉከራ]{}"assertion"(LEX=N,POS=NOU,NUM=SNG)<amh,0,0>;

[ፉከራው]{}"assertion"(LEX=N,POS=NOU,NUM=SNG,att=@def)<amh,0,0>;

[መስመሩን]{}"course"(LEX=N,POS=NOU,NUM=SNG,att=@def)<amh,0,0

[መስመሩ]{}"course"(LEX=N,POS=NOU,NUM=SNG,att=@def)<amh,0,0>;

[ቀን]{}"day"(LEX=N,POS=NOU,NUM=SNG)<amh,0,0>;

[መጨረሻዉ]{}"end"(LEX=N,POS=NOU,NUM=SNG,att=@def)<amh,0,0>;

[እግሮች]{}"foot"(LEX=N,POS=NOU,NUM=PLR)<amh,0,0>;

[እግሮቹ]{}"foot"(LEX=N,POS=NOU,NUM=PLR,att=@def)<amh,0,0>;

SCENARIO 2: ADJECTIVE CONCEPTS

[አይሳካም]{}"impossible"(LEX=J,POS=ADJ)<amh,0,0>;

[አጭር]{}"short"(LEX=J,POS=ADJ)<amh,0,0>;

[አጫጭር]{}"short"(LEX=J,POS=ADJ,att=@def)<amh,0,0>;

[ቀርፋፋ]{}"slow"(LEX=J,POS=ADJ,att=@indef)<amh,0,0>;

[ወጥ]{}"steady"(LEX=J,POS=ADJ)<amh,0,0>;

[ፈጣን]{}"swift"(LEX=J, POS=ADJ)<amh,0,0>;

SCENARIO 3: ADVERBIAL CONCEPTS

[አስቀድሞ]{}"already"(LEX=A,POS,AAV,SEM=TME,att=@perfective)<amh,0,0>

[ዘግይቶ]{}"late"(LEX=A,POS=AAV)<amh,0,0>;

[አንድ ቀን]{}"one day"(LEX=A,POS=AAV,SEM=TME)<amh,0,0>;

[እንደዋዛ]{}"simply"(LEX=A,POS=SAV)<amh,0,0>;

[ቀጥታ]{}"straight"(LEX=A,POS=AAV)<amh,0,0>;

[አብረዉ]{}"together"(LEX=A,POS=AAV)<amh,0,0>;

SCENARIO 4: VERBAL CONCEPTS

[ተስማሙ]{}"agree"(LEX=V,POS=VER,TRA=TST,ATE=PAS,att=@past)<amh,0,0>

[ተስማማ]{}"assent"(LEX=V,POS=VER,TRA=TST,ATE=PAS,att=@past)<amh,0,0>

[ያምናል]{}"believe"(LEX=V,POS=VER,TRA=TST,ATE=PAS,att=@past)<amh,0,0>

[መወሰን]{}"fix"(LEX=V,POS=VER,TRA=TST,VBL=INF)<amh,0,0>

[መምረጥ]{}"choose"(LEX=V,POS=VER,TRA=TST,VBL=INF,att=@decision)<amh,0,0>

[ሄደ]{}"go on"(LEX=V,POS=VER,TRA=NTST,ATE=PAS,att=@past)<amh,0,0>

[ጋደምአለ]{}"lay own"(LEX=V,POS=VER,TRA=NTST,ATE=PAS,att=@past)<amh,0,0>

[መለሰ]{}"reply"(LEX=V,POS=VER,TRA=TST,ATE=PAS,att=@past)<amh,0,0>

[አላገጠ]{}"ridicule"(LEX=V,POS=VER,TRA=TST,ATE=PAS,att=@past)<amh,0,0>

[ሮጠ]{}"run"(LEX=V,POS=VER,TRA=NTST,ATE=PAS,att=@past)<amh,0,0>

[ጀመሩ]{}"start"(LEX=V,POS=VER,TRA=TST,ATE=PAS,att=@past)<amh,0,0>

[ቆመ]{}"stop"(LEX=V,POS=VER,TRA=NTST,VBL=INF)<amh,0,0>

[አልቆመም]{}"stop"(LEX=V,POS=VER,TRA=NTST,VBL=INF,ATE=PAS,att=@not)<amh,0,0>;

[አንቀላፋ]{}"takeanap"(LEX=V,POS=VER,TRA=NTST,ATE=PAS,att=@past)<amh,0,0>

[ነቃ]{}"wake up"(LEX=V,POS=VER,TRA=NTST,ATE=PAS,att=@past)<amh,0,0>

[አሸንፏል]{}"win"(LEX=V,POS=VER,TRA=TST,ATE=PAS,att=@past)<amh,0,0>

[ያሸንፋል]{}"win"(LEX=V,POS=VER,TRA=TST)<amh,0,0>

II. CLOSED-CLASS CATEGORIES

Words those are not associated to ordinary Universal Words (UW), but to attributes and/or relations, or to the pro-form UW.

SCENARIO 1: DETERMINERS (LEX=D)

DEMONSTRATIVE DETERMINERS (POS=DEM)

[ሌላ]{}"" (LEX=D,POS=DEM,att=@other)<amh,255,0>

[ብቻ]{}"" (LEX=D,POS=DEM,att=@only)<amh,255,0>

[ተመሳሳይ]{}"" (LEX=D,POS=DEM,att=@same)<amh,255,0>

[ያ]{}"" (LEX=D,POS=DEM,NUM=SNG,att=@distal)<amh,255,0>

[ይህ]{}"" (LEX=D,POS=DEM,NUM=SNG,att=@proximal)<amh,255,0>

[እነዛ]{}"" (LEX=D,POS=DEM,NUM=PLR,att=@distal)<amh,255,0>

SCENARIO 2: QUANTIFIERS (POS=QUA)

[ጥንድ]{}"" (LEX=D,POS=QUA,att=@paucal)<amh,255,0>

[ጥቂት]{}"" (LEX=D,POS=QUA,att=@paucal)<amh,255,0>

[ትንሽ]{}"" (LEX=D,POS=QUA,att=@paucal)<amh,255,0>

[ብዙ]{}"" (LEX=D,POS=QUA,att=@multal)<amh,255,0>

[ሁሉም]{}"" (LEX=D,POS=QUA,att=@all)<amh,255,0>

[እያንዳንዱ]{}"" (LEX=D,POS=QUA,att=@each)<amh,255,0>

[በቂ]{}"" (LEX=D,POS=QUA,DIS=AFT,att=@enough)<amh,255,0>

[ማንኛውም]{}"" (LEX=D,POS=QUA,att=@every)<amh,255,0>

[ጥቂት]{}"" (LEX=D,POS=QUA,att=@paucal)<amh,255,0>

[አብዛኛው]{}"" (LEX=D,POS=QUA,att=@most)<amh,255,0>

SCENARIO 3: INTERROGATIVE DETERMINERS (POS=QUA or POS=DEM)

[ምን]{}"" (LEX=D,POS=DEM,att=@wh)<amh,255,0>

[የትኛዉ]{}"" (LEX=D,POS=DEM,att=@wh)<amh,255,0>

SCENARIO 4: NUMERALS (LEX=U) and CARDINALS (POS=CDN)

[ዜሮ]{}"0" (LEX=U,POS=CDN,DIGIT)<amh,255,0>;

[አንድ]{}"1" (LEX=U,POS=CDN,DIGIT)<amh,255,0>;

[ሁለት]{}"2" (LEX=U,POS=CDN,DIGIT)<amh,255,0>;

[ሶስት]{}"3" (LEX=U,POS=CDN,DIGIT)<amh,255,0>;

[አስር]{}"10" (LEX=U,POS=CDN,DIGIT)<amh,255,0>;

[አስራአንድ]{}"11" (LEX=U,POS=CDN,DIGIT)<amh,255,0>;

[አስራሁለት]{}"12" (LEX=U,POS=CDN,DIGIT)<amh,255,0>;

[አስራሶስት]{}"13" (LEX=U,POS=CDN,DIGIT)<amh,255,0>

SCENARIO 4: FRACTIONS (POS=PTN) (only canned expressions are represented)

[እሩብ]{}"1/4"(LEX=U,POS=PTN)<amh,255,0>

SCENARIO 5: VERBS (auxiliary verbs and copula)

[ነኝ]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PRS,PER=1PS,att=@present)<amh,255,0>

[ነን]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PRS,PER=2PS,PER=1PP,PER=2PP,PER=3PP,att=@present)<amh,255,0>

[ነዉ]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PRS,PER=3PS,att=@present)<amh,255,0>

[ነህ]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PRS,PER=3PS,att=@present)<amh,255,0>

[አይደለም]{}""(LEX=I,LEX=V,LST=CTT,POS=COP,POS=AUX,ATE=PRS,PER=3PS,att=@present;@not)<amh,255,0>

[ነበር]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PAS,PER=3PS,att=@past)<amh,255,0>

[ነበርሁኝ]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PAS,PER=1PS,att=@past)<amh,255,0>

[ነበረች]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PAS,GEN=FEM,PER=3PS,att=@past)<amh,255,0>

[ነበረ]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PAS,GEN=MCL,PER=3PS,att=@past)<amh,255,0>

[ነበሩ]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PAS,PER=3PP,att=@past)<amh,255,0>

[ነበርን]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PAS,PER=1PP,att=@past)<amh,255,0>

[ነበራቹ]{}""(LEX=I,LEX=V,POS=COP,POS=AUX,ATE=PAS,PER=2PP,att=@past)<amh,255,0>

CAN (modal)

[ይችላል]{}"" (LEX=I,POS=AUX,POS=MOV,att=@ability)<amh,255,0>;

[አይችልም]{}"" (LEX=I,POS=AUX,POS=MOV,att=@ability;@not)<amh,255,0>

LOOK (copula) e.g. he looks hungry

[ይመስላል]{}""(LEX=V,POS=COP,PER=3PS,ATE=PRS,GEN=MCL,att=@speculation;@present)<amh,0,0>

[ትመስላለች]{}""(LEX=V,POS=COP,PER=3PS,ATE=PRS,GEN=FEM,att=@speculation;@present)<amh,0,0>

MAY (modal) he may arrive

[ምናልባት]{}"" (LEX=I,POS=AUX,POS=MOV,att=@possibility)<amh,255,0>

MUST (modal) he must arrive

[ግዴታ]{}"" (LEX=I,POS=AUX,POS=MOV,att=@obligation)<amh,255,0>

OUGHT (modal) he ought to arrive

[አለበት]{}"" (LEX=I,POS=AUX,POS=MOV,att=@probability)<amh,255,0>

SCENARIO 6: PRONOUNS (LEX=R)

PERSONAL PRONOUNS (POS=PPR)

[እኔ]{}"00.@1" (LEX=R,POS=PPR,CAS=NOM,PER=1PS)<amh,255,0>

[አንተ]{}"00.@2.@male" (LEX=R,POS=PPR,CAS=NOM,CAS=OBL,PER=2PS)<amh,255,0>

[አንቺ]{}"00.@2.@female"(LEX=R,POS=PPR,CAS=NOM,CAS=OBL,PER=2PS)<amh,255,0>

[እነንተ]{}"00.@2" (LEX=R,POS=PPR,CAS=NOM,CAS=OBL,PER=2PP)<amh,255,0>

[እሱ]{}"00.@3.@male" (LEX=R,POS=PPR,CAS=NOM,PER=3PS)<amh,255,0>

[እሱዋ]{}"00.@3.@female" (LEX=R,POS=PPR,CAS=NOM,PER=3PS)<amh,255,0>

[እኛ]{}"00.@1.@pl" (LEX=R,POS=PPR,CAS=NOM,PER=1PP)<amh,255,0>

[እነሱ]{}"00.@3.@pl" (LEX=R,POS=PPR,CAS=NOM,PER=3PP)<amh,255,0>

SCENARIO 7: POSSESSIVE PRONOUNS (POS=SPR)

[የእኔ]{}"00.@1" (LEX=R,POS=SPR)<amh,255,0>; mine is beautiful

[የአንተ]{}"00.@2" (LEX=R,POS=SPR)<amh,255,0>; yours is beautiful

[የእነንተ]{}"00.@2.@pl" (LEX=R,POS=SPR)<amh,255,0>; yours is beautiful

[የእሱ]{}"00.@3.@male" (LEX=R,POS=SPR)<amh,255,0>; his is beautiful

[የእሱዋ]{}"00.@3.@female" (LEX=R,POS=SPR)<amh,255,0>; hers is beautiful

[የእኛ]{}"00.@1.@pl" (LEX=R,POS=SPR)<amh,255,0>; ours is beautiful

[የእነሱ]{}"00.@3.@pl" (LEX=R,POS=SPR)<amh,255,0>; theirs is beautiful

SCENARIO 8: REFLEXIVE PRONOUNS (POS=FPR)

[እራሴ]{}"00.@1" (LEX=R,POS=FPR,PER=1PS)<amh,255,0>

[እራስህ]{}"00.@2.@male" (LEX=R,POS=FPR,PER=2PS)<amh,255,0>

[እራስሽ]{}"00.@2.@female" (LEX=R,POS=FPR,PER=2PS)<amh,255,0>

[እራሱ]{}"00.@3.@male" (LEX=R,POS=FPR,PER=3PS)<amh,255,0>

[እራሱ]{}"00.@3" (LEX=R,POS=FPR,PER=3PS)<amh,255,0>

[እራሷ]{}"00.@3.@female" (LEX=R,POS=FPR,PER=3PS)<amh,255,0>

[እራሳችን]{}"00.@1.@pl" (LEX=R,POS=FPR,PER=1PP)<amh,255,0>

[እራሳቹ]{}"00.@2.@pl" (LEX=R,POS=FPR,PER=2PP)<amh,255,0>

[እራሳቸዉ]{}"00.@3.@pl" (LEX=R,POS=FPR,PER=3PP)<amh,255,0>

SCENARIO 9: INDEFINITE PRONOUNS (POS=NPR)

[ትንሽ]{}"00.@paucal" (LEX=R,POS=NPR)<amh,255,0>; both have a little

[ሁሉንም]{}"00.@all" (LEX=R,POS=NPR)<amh,255,0>; neither has all

[ሌላ]{}"00.@other" (LEX=R,POS=NPR)<amh,255,0>; anybody has another

[ማንም]{}"00.@any" (LEX=R,POS=NPR)<amh,255,0>; either has any

[ማንም]{}"00.@any.@person" (LEX=R,POS=NPR)<amh,255,0>; anybody has another

[ምንም]{}"00.@any.@thing" (LEX=R,POS=NPR)<amh,255,0>; anyone has anything

[ሁለቱም]{}"00.@both" (LEX=R,POS=NPR)<amh,255,0>; both have little

[ማንኛዉም]{}"00.@each" (LEX=R,POS=NPR)<amh,255,0>; each has nothing

[አንዱንም]{}"00.@either" (LEX=R,POS=NPR)<amh,255,0>; either has any

[በቂ]{}"00.@enough" (LEX=R,POS=NPR)<amh,255,0>; few have enough

[ጥቂት]{}"00.@paucal" (LEX=R,POS=NPR)<amh,255,0>; few have everything

[ብዙ]{}"00.@multal" (LEX=R,POS=NPR)<amh,255,0>; many have much

[ተጨማሪ]{}"00.@more" (LEX=R,POS=NPR)<amh,255,0>; less is more

[ብዙ]{}"00.@multal" (LEX=R,POS=NPR)<amh,255,0>; many have much

[ማንም]{}"00.@no.@person" (LEX=R,POS=NPR)<amh,255,0>; no one has others

[ምንም]{}"00.@no" (LEX=R,POS=NPR)<amh,255,0>; none has some

[ሌላ]{}"00.@other" (LEX=R,POS=NPR)<amh,255,0>; other has more

[ሌሎች]{}"00.@other.@pl" (LEX=R,POS=NPR)<amh,255,0>; no one has others

[የሆነ ሰዉ]{}"00.@person" (LEX=R,POS=NPR)<amh,255,0>; somebody has something

[አንድ ሰዉ]{}"00.@person" (LEX=R,POS=NPR)<amh,255,0>; nobody has someone

[አንድ ነገር]{}"00.@thing" (LEX=R,POS=NPR)<amh,255,0>; somebody has something

SCENARIO 10: INTERROGATIVE PRONOUNS (POS=IPR)

[እንዴት]{}"00.@wh" (LEX=R,POS=IPR)<amh,255,0>

[መቼ]{}"00.@wh" (LEX=R,POS=IPR)<amh,255,0>

[የት]{}"00.@wh" (LEX=R,POS=IPR)<amh,255,0>

[ማን]{}"00.@wh" (LEX=R,POS=IPR)<amh,255,0>

[ማንን]{}"00.@wh" (LEX=R,POS=IPR)<amh,255,0>

[የማን]{}"00.@wh" (LEX=R,POS=IPR)<amh,255,0>

[ለምን]{}"00.@wh" (LEX=R,POS=IPR)<amh,255,0>

SCENARIO 11: DEMONSTRATIVE PRONOUNS (POS=DEP)

[እነዚህ]{}"00.@proximal.@pl" (LEX=R,POS=DEP,NUM=PLR)<amh,255,0>

[ይህ]{}"00.@proximal" (LEX=R,POS=DEP,NUM=SNG)<amh,255,0>

[እነዚያ]{}"00.@distal.@pl" (LEX=R,POS=DEP,NUM=PLR)<amh,255,0>

SCENARIO 12: PREPOSITIONS (LEX=P,POS=PRE)

[ስለ]{}"" (LEX=P,POS=PRE,rel=cnt,att=@about)<amh,255,0>

[ላይ]{}"" (LEX=P,POS=PRE,rel=plc,att=@above)<amh,255,0>

[ላይ]{}"" (LEX=P,POS=PRE,rel=man,att=@against)<amh,255,0>

[መካከል]{}"" (LEX=P,POS=PRE,rel=plc,att=@among)<amh,255,0>

[ዙሪያ]{}"" (LEX=P,POS=PRE,rel=plc,att=@around)<amh,255,0>

[በፊት]{}"" (LEX=P,POS=PRE,rel=tim,att=@before)<amh,255,0>

[ከታች]{}"" (LEX=P,POS=PRE,rel=plc,att=@below)<amh,255,0>

[ከስር]{}"" (LEX=P,POS=PRE,rel=plc,att=@beneath)<amh,255,0>

[ከጎን]{}"" (LEX=P,POS=PRE,rel=plc,att=@beside)<amh,255,0>

[ሩቅ]{}"" (LEX=P,POS=PRE,rel=plc,att=@far_from)<amh,255,0>

[ለ]{}"" (LEX=P,POS=PRE,rel=ben,att=@for)<amh,255,0>

[ከ]{}"" (LEX=P,POS=PRE,rel=plc,att=@from)<amh,255,0>

[ላይ]{}"" (LEX=P,POS=PRE,rel=plc,att=@on)<amh,255,0>;

[ጨጪ]{}"" (LEX=P,POS=PRE,rel=plc,att=@out_of)<amh,255,0>;

[ጀምሮ]{}"" (LEX=P,POS=PRE,rel=tim,att=@since)<amh,255,0>;

[ስር]{}"" (LEX=P,POS=PRE,rel=plc,att=@under)<amh,255,0>;

[እስከዝያ]{}"" (LEX=P,POS=PRE,rel=tim,att=@until)<amh,255,0>

SCENARIO 13: CONJUNCTIONS (LEX=C) and COORDINATING (POS=CCJ)

[እና]{}"" (LEX=C,POS=CCJ,rel=and)<amh,255,0>; John and Mary

[ወይም]{}"" (LEX=C,POS=CCJ,rel=or)<amh,255,0>; John or Mary

SCENARIO 14: SUBORDINATING (POS=SCJ)

[ቢሆንም]{}"" (LEX=C,POS=SCJ,POS=AVR,att=@although,rel=seq)<amh,255,0>;

[እንደ]{}"" (LEX=C,POS=SCJ,POS=AVR,att=@as,rel=man)<amh,255,0>

[ግን]{}"" (LEX=C,POS=SCJ,POS=AVR,att=@but,rel=and)<amh,255,0>

5.6. Amharic Grammars used for UNL

To build the Amharic grammar within the UNL it requires Valerie certificates; CLEA 700 and 1000. Grammars are the formalizations required to transform Amharic sentences into the UNL and vice versa. The Amharic grammar is composed of the analysis and the generation grammar. Moreover, both analysis and generation grammars are divided into two main types namely: TRules are used to transform NL into UNL Expression and DRules are used to prevent wrong lexical choices, provoke best matches and check the consistency of words. These grammars are the foremost tough and also the actual goal of the complete analysis task [49].

Grammar Analysis: It is composed of rules that analyses Amharic input sentences morphologically, syntactically and semantically in order to reach its abstract meaning in UNL format. It requires an accurate understanding of Amharic language. Here, the UNL system provides the tools and methods capable of decomposing the sentence into its basic constituents, understanding and encoding the intended meaning behind each constituent and the meaning reflected by its artificial grammatical form as well as its position in the sentence. Moreover, the linguistic relation between every constituent and {also the} others ought to also perceive and encoded. The process of analyzing NL into semantic networks is carried out in a modular manner. The following figure illustrates the design of the analysis grammar is showing the five stages of analysis [49].

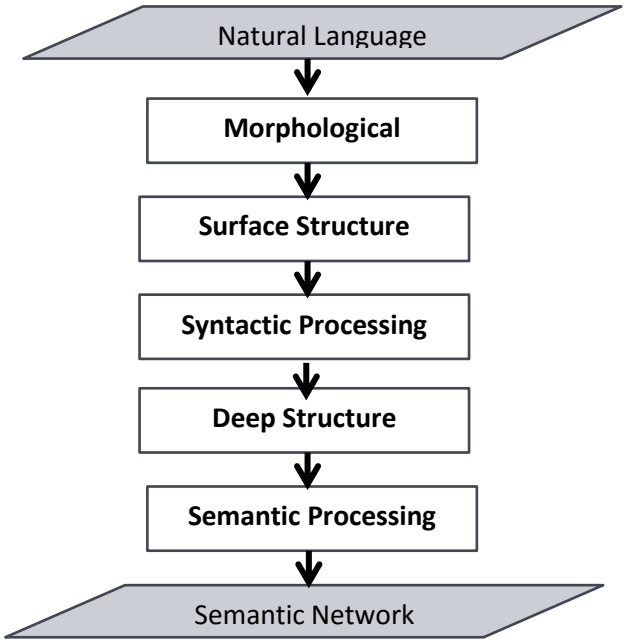

Figure 11: Design of the analysis Grammar

Note: Fourth tab is T-Rules where, NL-UNL transformation grammar (the grammar to be used to process the natural language input) is provided. T-rules may be classified according to the type of modification that they promote. LL, or list-to-list, where the initial state and the final state are list structures. There are 5 different sub types of LL rules [3].

ACTION	*RULE*	*DESCRIPTION*
ADD	(%x):=(%x)(%y);	The node %y is added to the right of the Node %x
	(%x):=(%y)(%x);	The node %y is added to the left of the Node %x
DELETE	(%x):=-(%x); (%x):=;	The node %x is deleted.
REPLACE	(%x):=(%y);	All the instances of the node %x will be replaced by the node %y
MERGE	(%x)(%y):=(%x and %y);	The nodes %x and %y will be merged Where %x and %y are nodes.

TT, or tree-to-tree, where the initial state and the final state are tree structures. There are 3 Different subtypes of TT rules.

ACTION	*RULE*	*DESCRIPTION*
ADD RELATION	SYN1(%x;%y):=+SYN2(%w;%z);	The relation SYN2 between the Nodes %w and %z will be added to the graph containing the relation SYN1 between the Nodes %x and %y.
DELETE RELATION	SYN(%x;%y):=-SYN(%x;%y); SYN(%x;%y)=;	The relation SYN between the Nodes %x and %y will be deleted (the nodes %x and %y will not be deleted).
REPLACE RELATION	SYN1(%x;%y):=SYN2(%w;%z);	The relation SYN1 between the Nodes %x and %y will be replaced by the relation SYN2 between the nodes %w and %z Where SYN is a syntactic Relation, and %x, %y, %z and % w are nodes.

NN, or network-to-network, where the initial state and the final state are network Structures. There are 3 different subtypes of NN rules.

ACTION	*RULE*	*DESCRIPTION*
ADD RELATION	SEM1(%x;%y):=+SEM2(%w;%z);	The relation SEM2 between the Nodes %w and %z will be added to the graph containing the relation SEM1 between the nodes %x and %y.
DELETE RELATION	SEM(%x;%y):=-SEM(%x;%y);	The relation SEM between the Nodes %x and %y will be deleted (the nodes %x and

	(%x):=;	%y Will not be deleted).
REPLACE RELATION	SEM1(%x;%y):=SEM2(%w;%z);	The relation SEM1 between the Nodes %x and %y will be replaced by the relation SEM2 between the nodes %w and %z Where SEM is any of the existing UNL relations, and %x,%y, %z and %w are nodes.

LT, or list-to-tree, converts lists into trees. There are 2 different subtypes of LT.

ACTION	***RULE***	***DESCRIPTION***
ADD	(%x)(%y):=+SYN(%x;%y);	The relation SYN is created between the nodes %x and %y if there is a linear relation between them (the linear relation is not deleted).
REPLACE	(%x)(%y):=SYN(%x;%y);	The linear relation between %x and %y is replaced by the relation SYN between the same nodes (i.e., the linear relation is deleted) Where SYN is a syntactic relation, and %x and %y are nodes.

TN, or tree-to-networks, converts trees into networks. There are 2 types of TN rules.

ACTION	***RULE***	***DESCRIPTION***
ADD	SYN(%x;%y):=+SEM(%w;%x);	The semantic relation SEM between the nodes %w and %x is created if there is a syntactic relation SYN between the nodes %x and %y.
REPLACE	SYN(%x;%y):=SEM(%x;%y);	The syntactic relation SYN between the nodes %x and %y is replaced by the semantic relation SEM between the same nodes. Where SYN is a syntactic relation, SEM is a semantic relation, and %x, %y, %w and %z are nodes.

Grammar Generation: UNL should generate NL in order to interact with the user for the purposes of translation, information retrieval, etc. It is an efficient and robust means for NLG, the process of generating NLG may be realized to some extent as the reflect image of the analysis process; the abstract meaning reached after the analysis process is transformed into a NL sentence.

Natural Language Generation System: In this case UNL graph is pre-processed by the NN rules in order to become a more easily traceable semantic network. As a result the network structure is born-again, by the NT rules, into a syntactic structure, which is still distant from the surface structure, as

it is directly derived from the semantic arrangement. Then the deep syntactic structure is transformed into a surface syntactic structure by the TT rules. The surface syntactical structure undergoes several different changes in keeping with the TL rules that generate a NL-like list structure. This list structure is finally realized as a NL sentence by the LL rules. The process of NLG has been shown in Figure 12.

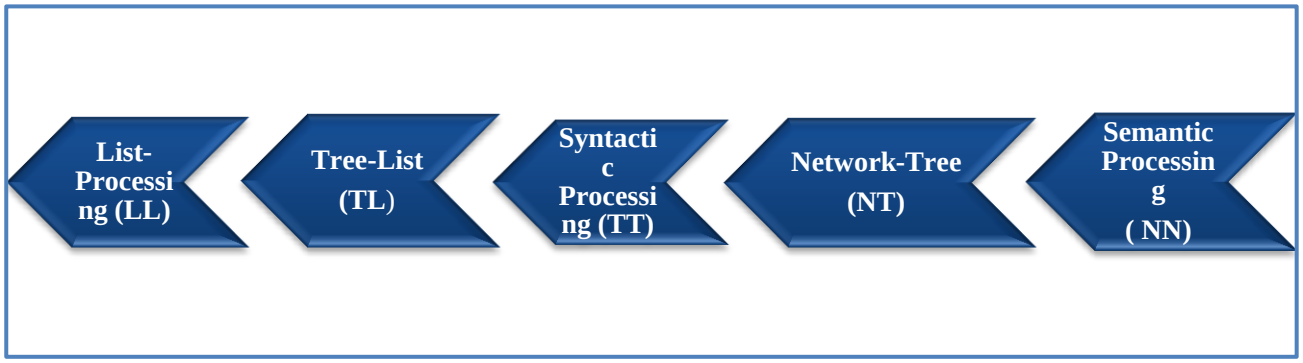

Figure 12: Process of NLG

As sentences area unit complicated structures which will contain nested or embedded phrases, each the analysis and therefore the generation processes could also be interleaved instead of pipelined. This means that the natural flow represented higher than is simply traditional and not necessary. During NLG, a LL rule may apply prior to a TT rule, or a NN rule may be applied after a TL rule. Rules area unit algorithmic and should be applied within the order outlined within the descriptive linguistics as long as their conditions area unit true, no matter the state. The description of rules used during NLG has been given in next sections [12] [3].

- ✓ **Grammar in NLization**: In the UNL framework, a grammar is a set of rules that are used to generate UNL out of NL, and NL out of UNL. Along with the UNLNL dictionaries, they organize the basic resource for UNLization and NLization. NL sentences and UNL graphs are supposed to convey the same amount of information in different structures: whereas the former arranges data as an ordered list of words, the latter organizes it as a network. In that sense, going from NL into UNL and from UNL into NL is ultimately a matter of transforming lists into networks and vice-versa. The UNL framework assumes that such transformation can be carried out progressively, i.e., through a transitional data structure: the tree, which could be used as an interface between lists and networks. Accordingly, there are five different types of rules [47].

Table 14: Different types of generation rules

Generation Rules	Description
NN	Semantic Processing (network-to-network)
NT	Deep-Structure Formation (network-to-tree)
TT	Syntactic Processing (tree-to-tree)
TL	Surface-Structure Formation (tree-to-list)
LL	List processing(list-to-list)

1. Amharic Morphology

Scenario 1: Nouns

Number

(N,PLR,^@pl):=(-PLR,+att=@pl)

Scenario 2: Adjectives

Degree

([በጣም],%x)(J,%y):=(%y,+att=@more); more expensive > expensive.@more

([እጅግ][በጣም],%x)(J,%y):=(%y,+att=@most); most expensive > expensive.@most

Scenario 3: Verbs

Tense and aspect

({@present|@past|@future},^TNS):=(+TNS);

(AUX,%x)(V,GER,%y):=(+att=@progressive,+att=%x,%y); is/was/be/been killing

;([አስቀድሞ])(V,%x):=(+att=@perfective,%x); already killed

;(V,%x)([አስቀድሞ]):=(+att=@perfective,%x); killed already

(^AUX,^BLK,^[not],^[n't])(V,{^AUX|COP},PRS,^TNS):=()(+att=@present,+TNS); kill

(^AUX,^BLK,^[not],^[n't])(V,{^AUX|COP},PAS,^TNS):=()(+att=@past,+TNS); killed

({^AUX,^BLK,^[not]|SHEAD|CHEAD},%x)([ለ])(V,%y):=(%x)(-ATE,+ATE=INF,%y); to kill

Modality

(MOV,%x)(V,%y):=(%y,+att=%x); can.@ability kill > kill.@ability

;([አለበት],%x)(V,%y):=(%y,+att=@probability)

;({[have]|[has]|[had]},%x)([to])(V,%y):=(%y,+att=@obligation)

Negation

([not])([to],%x)(V,%y):=(%x)(+att=@not,%y)

({[not]|[n't]})({V,^AUX|J|N|A|D},%x):=(+att=@not,%x)

Reflexivity and reciprocity

(VB,%x)(FPR):=(%x,+att=@reflexive)

(VB,%x)(CPR):=(%x,+att=@reciprocal)

Scenario 4: Numbers and Numerals

Normalization

({SHEAD|CHEAD|^BLK,^DIGIT},%x)([አንድ],%z)({"መቶ"|"ሺህ"|"ሚሊየን"|"ቢሊየን"|"ቲሪሊየን"},%y):=(%x)("1",[1],[[1]],U,CDN,DIGIT,%w)(%y); አንድ ሺህ > 1 ሺህ

({SHEAD|CHEAD|^BLK,^DIGIT,^[አንድ]},%x)({"መቶ"|"ሺህ"|"ሚሊየን"|"ቢሊየን"|"ቲሪሊየን"},%y):=(%x)("1",[1],[[1]],U,CDN,DIGIT,%z)(%y); ሺህ > 1 ሺህ

(DIGIT,%x)(COMMA,%y)(DIGIT,%z):=(%xand%z); 1,234 > 1234

(DIGIT,%x)(PUT,PERCENTAGE,%y):=(%xand%y,-PUT); 1 % > 1%

([ሲደመር],%x)(DIGIT,%y):=([+],[[+]],PUT,PLUS)(%y); ሲደመር 7 > + 7

([ሲቀነስ],%x)(DIGIT,%y):=([-],[[-]],PUT,HYPHEN)(%y); ሲቀነስ 7 > - 7

(DIGIT,%x)([ነጥብ],%y)(DIGIT,%z):=(DIGIT,%x)(".",[[.]],PERIOD,%w)(DIGIT,%z);

1ነጥብ5>1.5

Scenario 5: Cardinals

Round numbers (fill in the missing zeros)

(DIGIT,TRILLION,%x)({^DIGIT,^BLK,^PUT,{^[እና]}|STAIL|CTAIL},%y):=(%x)([[0]],[0],"0",BILLION,DIGIT)(%y);…….

(DIGIT,THOUSAND,%x)({^DIGIT,^BLK,^PUT,{^[እና]}|STAIL|CTAIL},%y):=(%x)([[0]],[0],"0",HUNDRED,DIGIT)(%y);

(DIGIT,THOUSAND,%x)(DIGIT,^HUNDRED,%y)({^[መቶ],^BLK|STAIL|CTAIL},%z):=(%x)([[0]],[0],"0",HUNDRED,DIGIT)(%y)(%z);

({SHEAD|CHEAD|^BLK},%z)(DIGIT,DOZEN,%x)({^DIGIT,^BLK,^PUT,{^[እና]|^[ና]}|"ሺህ"|"ሚሊየን"|"ቢሊየን"|"ቲሪሊየን"|STAIL|CTAIL},%y):=(%z)(%x)([[0]],[0],"0",DIGIT)(%y);

Non-round numbers (merge digits)

([0],0,%x)(DIGIT,^00,^000,%y):=(%xand%y,-0,+00);

([0],0,%x)(DIGIT,00,^000,%y):=(%xand%y,-0,-00,+000);

({SHEAD|CHEAD|^BLK,^DIGIT},%x)(DIGIT,TRILLION,%y)(DIGIT,BILLION,000,%z):=(%x)(%yand%z,-TRILLION);

({SHEAD|CHEAD|^BLK,^DIGIT},%x)(DIGIT,BILLION,%y)(DIGIT,MILLION,000,%z):=(%x)(%yand%z,-BILLION);

({SHEAD|CHEAD|^BLK,^DIGIT},%x)(DIGIT,MILLION,%y)(DIGIT,THOUSAND,000,%z):=(%x)(%yand%z,-MILLION);

({SHEAD|CHEAD|^BLK,^DIGIT},%z)(DIGIT,THOUSAND,%x)(DIGIT,HUNDRED,%y):=(%z)(%xand%y,-THOUSAND);

({SHEAD|CHEAD|^BLK},%z)(DIGIT,HUNDRED,%x)(DIGIT,00,%y):=(%z)(%xand%y,-HUNDRED,-00,+000);

({SHEAD|CHEAD|^BLK},%z)(DIGIT,DOZEN,%x)(DIGIT,^SECOND,%y):=(%z)(%xand%y,-DOZEN,+00);

Assigning values to digits

(DIGIT,%x)("ቲሪሊየን"):=(%x,+TRILLION);

(DIGIT,%x)("ቢሊየን"):=(%x,+BILLION);

(DIGIT,%x)("ሚሊየን"):=(%x,+MILLION);

(DIGIT,%x)("ሺህ"):=(%x,+THOUSAND);

(DIGIT,%x)("ሞቶ"):=(%x,+HUNDRED);

Ordinals

({DIGIT|CDN},%x)([ኛ]):=(%x,ORD,att=@ordinal)

Fractions

([አንድ],%a)(DIGIT,ORD,%x):=("1",[[1]],DIGIT,CDN,%b)(%x); a fifth > አንድ አምስተኛ

([ግማሽ],%x)(N,%y):=("1/2",[[1/2]],[1/2],U,DIGIT,PTN,W,%z)(%y); half book > 1/2 book

(DIGIT,CDN,%x)({[ሩብ]|[እሩብ]},%y):=(%x)("4",[4],[[4]],U,DIGIT,ORD,%z); one quarter / 1/4

Others

(DIGIT,%x)("ጊዜ"):=(%x,+@times,+LEX=A,+POS=AAV); two times > 2.@times

Scenario 6: Expressions of Time

Hours :- Converting hours from words into the format HH:MM:SS

(DIGIT,%x)(BLK)({"ሴ"|"ሴኮንድ"|"ሴኮንዶች"}):=(%x,+SECOND);

(DIGIT,%x)(BLK)({"ደ"|"ደቂቃ"|"ደቂቃዎች"}):=(%x,+MINUTE);

(DIGIT,%x)(BLK)({"ሰ"|"ሰዓት"|"ሰዓቶች"}):=(%x,+HOUR);

(HOUR,%x)(MINUTE,%y)(SECOND,%z):=(%x)(":")(%y)(":")(%z);

Special cases

({[ግማሽ]|[ተኩል]|[1/2]})([ሰዓት]):=([[30]],+DIGIT,+MINUTE); half an hour

(DIGIT,%y)(BLK)([ከሩብ]):=(%y,+HOUR)([[15]],+DIGIT,+MINUTE); a quarter past ten

([ለ])(BLK)(DIGIT,%y)(BLK)([እሩብ])(BLK)([ጉዳይ]):=(%y,+HOUR,+REDUCE1)([[45]],+DIGIT,+MINUTE);

(DIGIT,%x)([ከ])(DIGIT,%y):=(%x,+HOUR)(%y,+MINUTE);

Dates

(YEAR,%x)({ዓም|ዓዓ},%y):=(%xand%y)

(MONTH,%x){(COMMA)|("")}("/([01][0-9])?[0-9][0-9]/",^YEAR,%y):=(%x)(%y,+YEAR)

({"መስከረም"})(DIGIT,%x)(""):=(%x,+DAY)([[1]],+DIGIT,+MONTH);

({"ጥቅምት"})(DIGIT,%x)(""):=(%x,+DAY)([[2]],+DIGIT,+MONTH);…..

({"ጳጉሜ"})(DIGIT,%x)(""):=(%x,+DAY)([[13]],+DIGIT,+MONTH);

Scenario 7: Amharic Syntax

2.1 NP (Noun Phrase) is A word or group of words that function in a sentence as subject, object or prepositional object.

Determiners (D)

(D,att,%x)(NB,%y):=(%y,+att=%x); the.@def book > book.@def

(D,att,%x)(NP,%y):=(%y,+att=%x); the.@def book of John > book of John.@def

(N,%x)(D,att,AFT,%y):=(%x,+att=%y); books enough.@enough > books.@enough

(D,att,%x)([of])(NP,%y):=(%y,+att=%x); all.@all of them > them.@all

Ordinals (ORD)

(DIGIT,ORD,%y)(NB,%x):=(XB(%x;%y,+spec,+mod),+LEX=N,+XB=NB,%z); first book > XB(book,first)

Quantifiers (DIGIT)

(DIGIT,^ORD,%y)(NB,%x):=(XB(%x;%y,+spec,+qua),+LEX=N,+XB=NB,%z); two books > XB(books,two)

Posessive determiners (POD)

(POD,%x)(NB,%y):=(XP(%y;%x,+spec,+pos),+LEX=N,+XP=NP,%z); my book > XP(book,I)

Possessive pronouns (SPR)

(N,%x)(SPR,%y):=(XB(%x;%y,+spec,+pos),+LEX=N,+XB=NB,%z); book of mine > XP(book,I)

Genitive (GNT)

(N,%x)(PTC,GNT)(N,%y):=(XB(%y;%x,+spec,+pos),+LEX=N,+XB=NB,%z); John's book > XB(book,John)

Indefinite pronouns (NPR) is a pronoun that does not refer to any person or thing in particular (i.e anything, something, anyone, everyone).

(NPR,%x)(NP,%y):=(XB(%y;%x,+spec,+qua),+LEX=N,+XB=NB,%z); none of them > XB(them, none)

Interrogative determiners

([የማን],%x)(N,%y):=(XB(%y;%x,+spec,+pos),+LEX=N,+XB=NB,%z);

JP (Adjective Phrase)

Intensifiers (SAV)

(SAV,att,%x)({AB|JB|DB},%y):=(%y,+att=%x); very.@plus beautiful > beautiful.@plus

({AB|JB},%x)(SAV,att,AFT,%y):=(%x,+att=%y);

(SAV,^att,%x)({AB|JB|DB},%y):=(XB(%y;%x,+spec,+man),+LEX=%y,+XB=%y);

PP (Prepositional Phrase)

([ከ])(%x,XP)([ጀምሮ]):=(%x,-LEX,-XP,+LEX=P,+XP=PP,+adjt,+rel=man,+att=@from_on)

CP (Complementizer Phrase)

Complementizers

(%x,V)(CMR)(VP,%y):=(XB(%x;%y,+comp,+rel=obj);

Adverbializers

(AVR,%x)(XP,%y):=(%y,-LEX,-XP,+LEX=C,+XP=CP,+adjt,+att=%x,+rel=%x);

Coordination

(VP,%x)({[እና]|COMMA})(VP,%y):=(and(%y;%x),+LEX=V,+XP=VP,%xy);

(NP,%x)({[እና]|COMMA})(NP,%y):=(and(%y;%x),+LEX=N,+XP=NP,%xy);

(AP,%x)({[እና]|COMMA})(AP,%y):=(and(%y;%x),+LEX=A,+XP=AP,%xy);

(JP,%x)({[እና]|COMMA})(JP,%y):=(and(%y;%x),+LEX=J,+XP=JP,%xy);

(PP,%x)({[እና]|COMMA})(PP,%y):=(and(%y;%x),+LEX=P,+XP=PP,%xy);…..

(VP,%x)([ወይም])(VP,%y):=(or(%y;%x),+LEX=V,+XP=VP,%xy);

(NP,%x)([ወይም])(NP,%y):=(or(%y;%x),+LEX=N,+XP=NP,%xy);

(AP,%x)([ወይም])(AP,%y):=(or(%y;%x),+LEX=A,+XP=AP,%xy);

(JP,%x)([ወይም])(JP,%y):=(or(%y;%x),+LEX=J,+XP=JP,%xy);

(PP,%x)([ወይም])(PP,%y):=(or(%y;%x),+LEX=P,+XP=PP,%xy)

Scenario 8: Other rules

Punctuation

(BLK):=; deletes the blank space

(XP,%x)(QMARK):=(%x,att=@interrogative);

(XP,%x)(EMARK):=(%x,att=@exclamative);

Scenario 9: Dgrammar

This is used to improve the performance of the tokenization and of the transformation.

(PTC)(^BLK)=0;

(XP)({","|C},%z)(^XP)=0;

(NB)(N,GNT,^NP)=0;

(VP)(^STAIL,^CTAIL,^PUT,^C,^XP)=0;

(VB)(^STAIL,^CTAIL,^PUT,^C,^XP)=0;

(JB)(P)=0;

(J)(CCJ)(N)=0;

(C)(and (;att=@but),XP)=0;

✓ **The Amharic Language Evaluation Technique and Specification**

F-1 measure for this study is the evaluation techniques for calculating grammar accuracy which is precision and recall in the process of Enco and Deco functions, it require two files. The First one is the expected output file and the second one is the actual output file. The actual and the expected results must have the same number of sentences and must be aligned; the first sentence in the actual result will be compared to the first sentence in the expected result, the second with the second, and so on. The documents must be provided in plain text format (.txt) with utf-8 encoding and two parameters required to the calculation of F-1 [12].

- Actual output file contains the result that is generated by IAN and EUGENE.
- Expected output file contains those sentences which were manually translated.

Table 15: Assessment of the UNL

Type of Process	***UNL Evaluation Requirements***
Unlization	• Actual output: the output provided by IAN tool in Amharic language from the translated version of Amharic UC-A1. • Expected output: UC-A1 in UNL
Nlization	• Actual output: the output provided by EUGENE in Amharic language from the input file UC-A1 in UNL • Expected output: the human-translated version of UC-A1 used as the input for the UNLization.

Chapter Six

6. Experimentation, Result and Discussion

In this chapter the Experimentation, Results and Discussion of the conducted research is presented. It starts by presenting the research experimental analysis and followed by evaluation and testing tool for experimental Results and finally explained an open discussion in detailed.

6.1.Development of Unlization and Nlization for Amharic Language

The development of various applications of UNL framework based system For Amharic Language, Enco and Deco are primary research towards UNL system. On this study, the researchers are used to achieve Unlization and Nlization for Amharic Language Using common Enco and Deco tools through IAN and EUGEANE respectively provided by the UNDL foundation.

6.2.Experimental Investigation

In order to make UNlization and Nlization on UNL system through IAN and EUGEAN tools respectively, localization of prepared the corpus, dictionaries and grammars deal with the follow steps below:

- ✓ ***Localization:*** Localization is process of localization of existing UNL framework resources into the Amharic language Enco and Deco activities. That can be local uc-a1 corpus, dictionaries and grammars required to process the primary research of the UNL framework. Some instructions concerning the localization of the English dictionary and the English grammar into Amharic NL. Notes here there are the steps to be localized [12].

    1. Instead of creating a whole new dictionary from scratch, you may try localizing the English dictionary. In any case, the resulting dictionary should reflect your translated version of the corpus. i.e., all entries appearing in your translated version of the corpus should appear in the dictionary.
    2. In order to prepare the grammar, study the Grammar Specs. In many cases, it is simpler just to localize the English grammars to Amharic locale rather than creating a whole grammar from the scratch.
    3. Tokenization is not necessary, because the corpus is already segmented.

4. D-rules used in tokenization are used either to prevent wrong lexical matches or to provoke right lexical matches. Drules utilized in transformation are wont to either stop or provoke the applying of sure rules.
5. The F-measure compares the actual output exported from IAN with the expected output available at uca1_unl.txtof the UNDL page.

✓ ***Preparing the Corpus:*** Prepare the coaching corpus: Translate manually the fifty sentences of the corpus uca-1 from English into Amharic language as terribly shut on the first, and supply one single translation for each sentence**.**

1. Translate manually the sentences of uc-a1 from English into Amharic language. Very close to the original and provide one single translation for each sentence it is input file to IAN.
2. Save the translated texts in a plain text (.txt) file with utf-8 encoding and upload it to unlweb>unldev>IAN>NL files.
3. Upload the UNL files corresponding to the texts of uc-a1 to unlweb > unldev > EUGENE > UNL document.
4. Save the translated Amharic text in a plain text (.txt) file with utf-8 decoding and upload it to unlweb > unldev > IAN > NL files.

✓ ***Preparing the Dictionary:*** The analysis and generation dictionary matching to the word forms appearing in translated version of the corpus. The dictionary must be created in UNL-NL dictionary in a plain text (.txt) file with utf-8 encoding, decoding with the Dictionary Specs. Creating an Amharic dictionary for training corpus. The dictionary must contain all and only the word forms appearing in the training corpus must comply with the Dictionary Specs.

1. Create the analysis dictionary and generation dictionary matching to the word forms appearing in translated version of the corpus. The wordbook should be created in compliance with the wordbook Specs and will contain solely tags accessible at the tagset.
2. Save the NL-UNL and UNL-NL dictionary in a plain text (.txt) file with utf-8 encoding, decoding and upload it to unlweb > unldev > IAN > dictionaries and unlweb > unldev > projects > EUGENE > dictionaries.
3. Upload the Default Dictionary to unlweb >unldev > IAN > dictionaries Unlization and unlweb >unldev > generation > dictionaries Nlization.

4. Save the NL-UNL dictionary in a plain text (.txt) file with utf-8 encoding and upload it to ulnweb > unldev > IAN > dictionaries.

✓ ***Preparing the Grammar:*** The analysis and generation grammar necessary to transform the text into Unl expression and to generate Amharic text of corpus uc-a1 in UNL. This grammar is the most difficult and the actual goal of the whole analysis and generation tasks of the whole UNL system.

1. Create the NL-UNL analysis transformation grammar necessary to analyze the Amharic sentences of the translated corpus in UNL. This descriptive linguistics is that the most tough and also the actual goal of the entire analysis task.
2. Provide the UNL-NL generation TGrammar and Dgrammar necessary to generate Amharic language sentences from the UNL corpus to Amharic language.
3. Create the Amharic inflectional paradigms only for the inflected forms appearing in the UNL-NL Amharic dictionary.
4. Export the inflectional grammar of Amharic language from unlarium > grammar > [your locale] > export.
5. Save it in a plain text (.txt) file with utf-8 encoding and upload it to unlweb > unldev>IAN>t-rule and unlweb>unldev>projects>EUGENE> rules.
6. Create the NL-UNL analysis disambiguation grammar in order to improve the results of the transformation and to control the process of tokenization. Again, this grammar is normally developed inside IAN, because we have to test it every time.
7. Provide the UNL -NL generation disambiguation grammar in order to improve the results of the transformation and to control the process of tokenization. Again, this grammar is normally developed inside EUGENE, because we have to test it every time.
8. Test the grammar against the corpus and provide the necessary changes.

✓ ***Preparing the evaluation:*** The actual outputs must be evaluated against the expected outputs using the F-1 calculated at unlweb > unlarium > grammar > [locale] > F-1 >.

1. Export the actual output of IAN range=1-100, trace-level=none and
2. Export the actual output of EUGENE range=1-100, trace-level=Minimal.
3. Check the F-measure of the actual output at unlweb > unlarium >tools>F-1.
4. If the F-measure > 0.9, upload the corpus, dictionary and grammars to Valerie (unlweb>Valerie>clea1500) in order to have evaluated.

6.3. Experimentation of Unlization and Nlization for Amharic

This experiment was done on corpus uc-a1 is a document consist of total 100 sentences has been categories all the major part-of-speech provided by UNDL foundation, this corpuses are sample English sentences has been represented from NL form into UNL. First manually converted into Amharic language using IAN for input to UNL framework and generates Amharic sentences out of UNL framework using EUGEAN tools, UNlization and Nlization of uc-a1 was done with the help of total [502] TRules and [494] dictionary entries are included for both process.

- ✓ **Amharic Training corpus**

For prepare the training corpus we have translated from English into Amharic language. We have tried to be as close as possible to the original and provided one single translation for each sentence. This is used as input document file and our goal is to provide automatically, through IAN tools as input for each translated or full translated corpus sentence are given below, see Appendix 1.

- ✓ **Amharic UNL Dictionary**

To build the dictionary within the UNL it requires an Amharic dictionary for our training corpus it contains all and the word forms appearing in our training corpus and the Amharic analysis dictionary lists all Amharic words and enumerates all the possible word forms of each word it includes about entries and Amharic generation dictionary only lists a base form to represent each lexical item. Along with each base form is a set of morphological rules that enable the generation of the different word forms required in different context. The Amharic dictionaries are used, along with the Amharic grammars, in the process of Unlization and Nlization They follow the UNL dictionary Specs and use the tags provided in the Tagset are provided in two different formats:

1. The Amh-UNL Analysis dictionary: it is presented in the enumerative format; it brings all word forms and not only base forms:

    - Amh-UNL dictionary used for the Unlization.
    - Default dictionary used to handle blank spaces, punctuation signs and other language-independent information.
    - Amh-UNL Transformation grammar used for the Unlization.
    - Default transformation grammar for Unlization.
    - Disambiguation grammar used for the Unlization.

2. The UNL-amh Generation dictionary: it is presented in the generative format; it brings only base forms with the orders to generate the corresponding inflections:

   - UNL-amh Amharic dictionary used for the Nlization.
   - Default dictionary used to handle blank spaces, punctuation signs and other language-independent information.
   - UNL-amh Transformation grammar used for the Nlization.
   - Default transformation grammar for Nlization.

✓ **Amharic grammars used for UNL**

The Grammar is a set of rules that is used to generate UNL out of NL, and NL out of UNL. Along with dictionaries, they constitute the basic resource for Unlization and Nlization. Amharic NL-UNL is analysis grammars necessary that are necessary to Unlized the NL sentences of the translated corpus are created. These grammars are the foremost hard and the particular goal of the complete analysis task. Three different types of grammars are prepared namely: N-grammar, TGrammar and Dgrammar-grammar: These grammars prepared and used to transform the list structure of the natural language sentence into a graph structure of the UNL graph.

Actual output results for Amharic Language

Enco Example for Amharic language

```
[/P]
[P:1]
[S:1]
  {org}
    ጥንቸሉ እና ኤሊው
  {/org}
  {unl}
     and(hare:01,tortoise:05)
  {/unl}
[/S]
[/P]
[P:2]
[S:14]
  {org}
    ኤሊው መለሰ
  {/org}
  {unl}
     agt(reply:03.@past,tortoise:01)
  {/unl}
[/S]
```

Deco Example for Amharic language

```
{UNL}
  and(hare:01,tortoise:05)
{/UNL}
```

The Amharic text output: ጥንቸሉ እና ኤሊው

```
{UNL}
  agt(reply:03.@past,tortoise:01)
{/UNL}
```

The Amharic text output: ኤሊው መለሰ

Here, we included the actual and expected the 1st ten sentences which is generated automatically from the UNL framework system.

==========

Results Automatically generated from the UNL framework system

==========

>>SENTENCE = 1

>>SCORE = 1.00

>>STATUS = CORRECT

>>INPUT

ጥንቸል

>>ACTUAL OUTPUT

[W]

hare:01

[/W]

>>EXPECTED OUTPUT

[W]

hare

[/W]

==========

>>SENTENCE = 2

>>SCORE = 1.00

>>STATUS = CORRECT

>>INPUT

ጥንቸሉ

>>ACTUAL OUTPUT

[W]

hare:01.@def

[/W]

>>EXPECTED OUTPUT

[W]

hare.@def

[/W]

==========

>>SENTENCE = 3

>>SCORE = 1.00

>>STATUS = CORRECT

>>INPUT

ኤሊው

>>ACTUAL OUTPUT

[W]

tortoise:01.@def

[/W]

>>EXPECTED OUTPUT

[W]

tortoise.@def

[/W]

==========

>>SENTENCE = 4

>>SCORE = 1.00

>>STATUS = CORRECT

>>INPUT

ውድድር

>>ACTUAL OUTPUT

[W]

race:01.@indef

[/W]

>>EXPECTED OUTPUT

[W]

race.@indef

[/W]

==========

>>SENTENCE = 5

>>SCORE = 1.00

>>STATUS = CORRECT

>>INPUT

ቀርፋፋው ጉዞ

>>ACTUAL OUTPUT

mod(pace:03.@indef,slow:01)

>>EXPECTED OUTPUT

mod(pace.@def, slow)

==========

>>SENTENCE = 6

>>SCORE = 1.00

>>STATUS = CORRECT

>>INPUT

አጫጭር እግሮች

>>ACTUAL OUTPUT

mod(foot:03.@idef.@pl,short:01.@def)

>>EXPECTED OUTPUT

mod(foot.@def.@pl, short)

==========

>>SENTENCE = 7

>>SCORE = 0.57

>>STATUS = WRONG (The discrepancy of UWs is above the limit)

>>INPUT

የኤሊው ቀርፋፋው ጉዞው

>>ACTUAL OUTPUT

mod(pace:05.@def,:01)

mod:01(tortoise:01.@def,slow:03)

>>EXPECTED OUTPUT

mod(pace.@def, slow)

mod(pace.@def, tortoise.@def)

==========

>>SENTENCE = 8

>>SCORE = 0.50

>>STATUS = WRONG (The discrepancy of relations is above the limit)

>>INPUT

የኤሊው አጫጭር እግሮች

>>ACTUAL OUTPUT

mod(foot:05.@idef.@pl,:01)

mod:01(tortoise:01.@def,short:03.@def)

>>EXPECTED OUTPUT

mod(foot.@def.@pl, short)

pof(tortoise.@def, foot.@def.@pl)

==========

>>SENTENCE = 9

>>SCORE = 1.00

>>STATUS = CORRECT

>>INPUT

የመስመሩ መጨረሻዉ

>>ACTUAL OUTPUT

mod(end:03.@def,course:01.@def)

>>EXPECTED OUTPUT

mod(end.@def, course.@def)

==========

>>SENTENCE = 10

>>SCORE = 0.00

>>STATUS = WRONG (The discrepancy of UWs is above the limit)

>>INPUT

ጥንቸሉ እና ኤሊው

>>ACTUAL OUTPUT

and(tortoise:05.@def,hare:01.@def)

>>EXPECTED OUTPUT

and(hare.@def, tortoise.@def)

==========

6.4. Implementation Summary

This chapter explained that how with the help of [502] TRules and [494] dictionary entries Unlization and Nlization of corpus uc-a1 has been done successfully for all the major Part-of-Speech for Amharic Language. Further improved the uc-a1 has been done using UNL approach with the help of Additional TRules and dictionary entries to enhance the accuracy of the output results. Unlization and Nlization have been carried out with the help of IAN and EUGENE tools respectively developed by UNDL foundation.

This study, as mentioned earlier has focused on the Unlization and Nlization of Amharic Sentence Structures. Apart from this process that was followed, the outcome of results has also been discussed in this section. Evaluation is done by using F-1 is the measure of a grammar's accuracy of the T-rules and Dictionary for the given sentences between (0/1). It considers both the precision and the recall of the grammar to calculate the score, according to the formula given in chapter two (2.1). Experimental Analysis in UNL is a NL-representation which can be used for refining, describing, semantic searching, generating and mapping the NL an independent format. IAN and EUGENE are an effective online tools developed by UNDL foundation available at UNL-arium page, which is used for Unlization and Nlization process respectively. The help of grammars and Dictionary entries, the proposed system has been done on corpus uc-a1 Amharic sentences, and the system tested through F1-score which is calculated precision and recall, two parameters required to the calculation of F-1 measure for testing the System the actual output file and the expected output file will be compared. The testing process is an important aspect in order to make sure the usefulness of the system. The evaluation of Amharic language Sentence Structures have been performed on Corpus uc-a1. The results of the system are very promising; the outputs of system are very good and satisfactory level as a standard in a MT field of study.

6.5. Experimental Results

***Testing of UNlization* :** In the process of enconverter the values of precision, recall, number of processed sentence is 100, returned sentence is 100 and correct sentences is 64 for uc-a1 is given

in Table 15. TRules and dictionary entries were created for Unlization of uc-a1 corpus F-1 is comes out to be 0.640 as shown in figure 13: Testing details of uc-a1.

```
============================================================
F-MEASURE VERIFICATION REPORT
Documentation available at http://www.unlweb.net/wiki/F-Measure
============================================================
```

Table 15: F-1 for Unlization of uc-a1.

Actual output	=actual-all.txt
Sentences processed	= 100
Sentences returned	= 100
Sentences correct	=64
Precision	= 0.640
Expected output	= expected-all.txt
Recall	= 0.640
F-Measure	**= 0.640**

```
============================================================
```

RESULTS

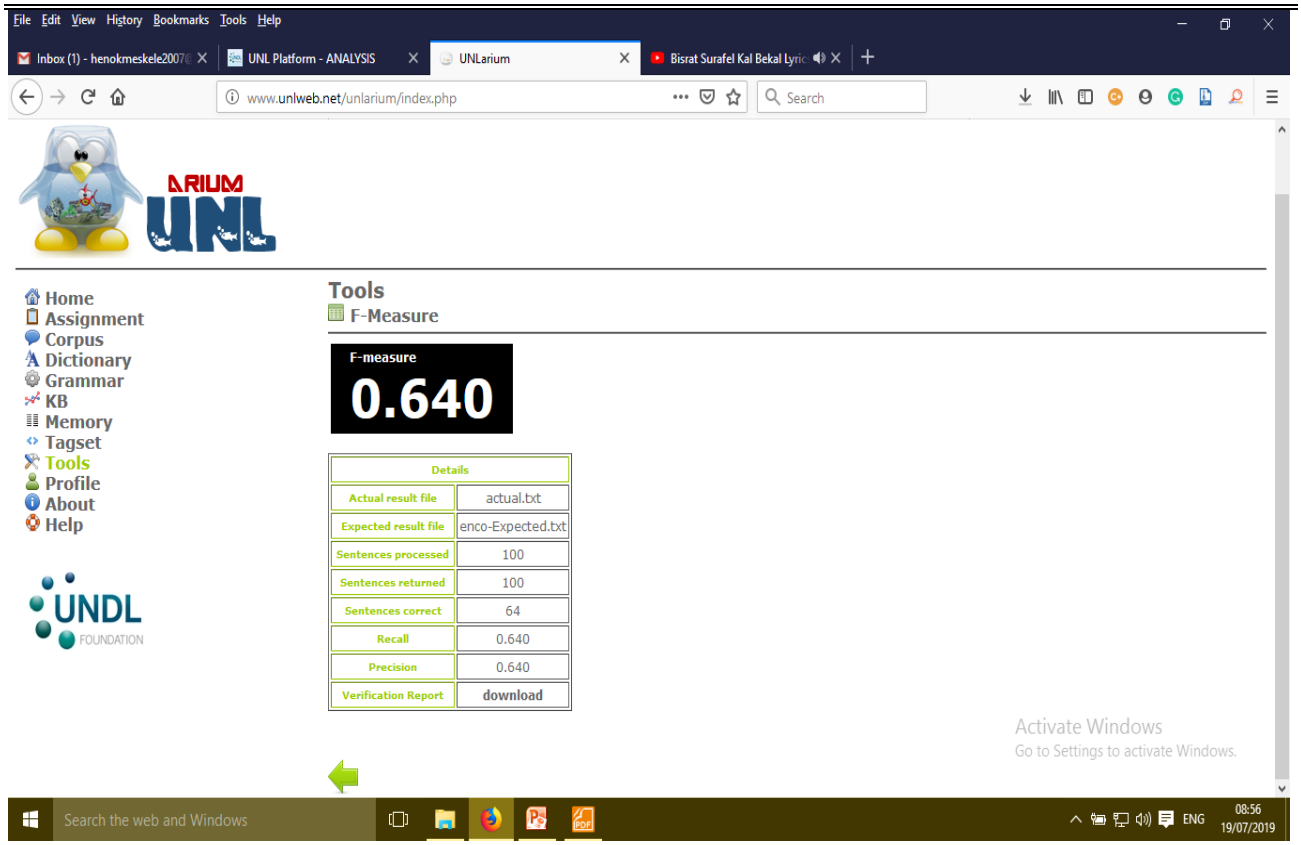

Figure 13: for Unlization of uc-a1.

Testing of Nlization **:** In the process of deconverter the values of precision, recall, number of processed sentence is 50, returned sentences is 29 and correct sentences is 16 for uc-a1 is given in Table 16. TRules and dictionary entries were created for Unlization of uc-a1 corpus F-1 is comes out to be 0.405 as shown in figure 14: Testing details of uc-a1.

==========

F-MEASURE VERIFICATION REPORT

Documentation available at http://www.unlweb.net/wiki/F-Measure

Table 166: F-1 for Nlization of uc-a1.

Actual output	=actual-all.txt
Sentences processed	= 50
Sentences returned	= 29
Sentences correct	=16
Precision	= 0.320
Expected output	= expected-all.txt
Recall	= 0.552
F-Measure	**= 0.405**

==========

RESULTS

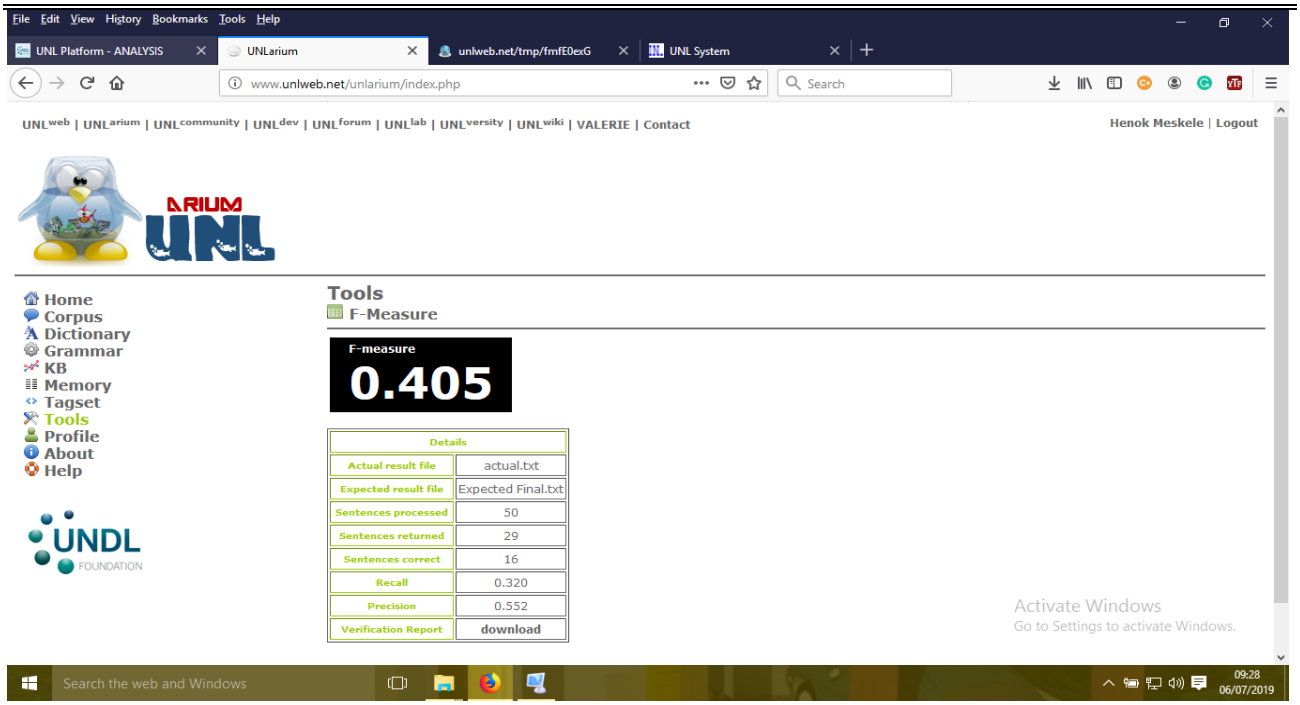

Figure 14: for Nlization of uc-a1.

6.6. Experimental Discussions

***Open Discussion*:** The sample corpus that is used for as training and test set as an input analysis and NLG system, IAN tool and EUGENE tool is an effective online tool for Unlization and Nlization released in 2012. This study explains how the given Amharic text can be converted into UNL its corresponding grammar using IAN tools and to Generation of Amharic NL from UNL its matching grammar and dictionary using EUGENE has been done. The dataset are used UC-A1 corpus provided by UNDL Foundation, it was originate in English language to all the participant languages. Even if, this corpus were downloaded and manually converted to Amharic language it contains a set of 100 NL sentences it is covered all the major part-of speeches, the famous story of 'The Tortoise and the Hare' from Aesop's Fables. The grammar and Dictionary entries has been UNlization and Nlization process of this corpus has been done, via a word dictionary containing the information of about words that correspond to UWs included in the input UNL expressions with their grammatical attributes that describe the behaviors of the words, and about T-rules for Describing how to construct a sentence using the information from the input UNL expressions and defined word dictionary. For testing matter, these corpus was uploaded on UNLweb was done in separately for each corpus go using IAN. The output UNL of the entire was copied from IAN console and saved in UTF-8 format. This is the actual output file. Unl web also provides the expected output file in UNL format. These actual and expected UNL files are then uploaded at UNLarium and the F-1 system is described in chapter two..2.1. The accuracy of the proposed system has been calculated using F-1 score to calculate precision and recall. This tool is very efficient resources to evaluate MT system accuracy of UNlization and Nlization of uc-a1 corpus, final out came to be 0.640 and 0.340 it can be shared in online community with other users.

Chapter Seven

7. Conclusion and Recommendation

This chapter presents the conclusions drawn from the study and recommended solutions.

7.1. Conclusion

This study provided a good experience the researcher in UNL framework based Application. It is promising contributions for Automatique MT system for Amharic language that integrated to Unlization and Nlization. This system used for language engineering tasks such as MT, AI, semantic search engine, CLIR, multilingual question and answering systems.

- As a MT trend, there are around more than 50 languages which can communicate, exchange data, and help users interact each other from one NL to others based on UNL Interlingua. This research output is also one of these achievements.
- As UNL is an artificial language that includes information retrieval concept that search better quality information to satisfy the user information need through accessing multiple resources in various languages using multi lingual search engine which includes Amharic language too.
- Amharic language based multilingual question and answering system can be extended now. Previously, in Ethiopian local languages concern, a lot of research activities was done in question and answering system to different local languages, but through UNL now it is possible to make Amharic language based Multilingual question and answering system it will be work on providing best answers from different resources and different language written documents if the answer is valid.
- Currently, Artificial intelligence and smart devices are expected to accept some commands and instructions from various multiple languages. That is why UNL is a portable standardization Interlingua to use for various purposes, as a back end components for different applications. UNL has the Potential to bridge the gap across all the languages of the world. It was launched to erase linguistic barriers and is used to access, transfer, and process information on the Internet [17].
- The UNL contains many components that are the language resources of a package in any NL inputted to the desired language. It is far better than other statistical approaches, where N

number of languages are to be translated between one to another, it would require a separate system for every language, i.e., n*(n-1). So, Interlingua is better than other approaches since it requires only 2*n language systems.

- UNL is a medium for representing information in a language independent format. It is recognized as the only living Interlingua capable enough for representing the general purpose contents. With the help of DRules, TRules and Dictionary entries, the system has been tested on UC-A1 for Amharic language the outcomes out to be 0.640 and 0.340.
- The UNL can be used not only for NL translation but also for other NLP tasks such as information retrieval, text simplification and semantic reasoning, multilingual information service, question and answering, UNL based search engine. Moreover the system can be improved to achieve F-Measure score for Amharic language and interested researchers can extend the study to develop UNL module for others Ethiopian languages.

7.2. Future work

Currently, Unlization and Nlization of Amharic language had been done on sentence level using IAN and EUGEAN tools. Some of the work in this study can be extending by including some components in future: Can be improve by expanding the dictionary and with more semantic information for any NL semantic based searching will be develop for Amharic language. It can be improve to handle very large and complex sentences will be at paragraph level with ambiguous words can be handling by disambiguation module. It can be integrate in to Afan Oromo, Tigrigna, Afar and Somali Ethiopian local languages which are recognized by under UNL. To achieve F-1 measure score of grammar accuracy for Amharic Language for precision and recall to be improve. These finding extended for other UNL applications like LILY (i.e., Language-to-Inter language-to-Language system), TUT (i.e., Text-to-Text through UNL), and KEYS (i.e., Knowledge Extraction system). These are the applications which are under development by UNDL Foundation.

References

[1]. “History of UNL", Available: [Online]. http://www.unlweb.net/wiki.[Accessed 3 June 2019].

[2]. A. Chawla, "UNL Punjabi Deconverter," 20. UNL Punjabi Deconverter Anjuman Chawla (Roll No 8043102) 20. UNL Punjabi Deconverter Anjuman Chawla (Roll No 8043102) 2019.

[3]. www.undl.org, [Online]. Available: http://www.unlweb.net/wiki / [accesed jan20 1919].

[4]. N. N. a. A. M. T. Amelina, ""Cross-lingual Information Retrieval."," 31. Amelina, Nasharuddin Nurul, and Abdullah Muhamad Taufik. "Cross-lingual In" Electronic Journal of Computer Science and Information Technology", Vols. 31. Amelina, Nasharuddin Nurul, and Abdullah Muhamad Taufik. "Cross-lingual Information Retrieva2, no. (e JCSIT), 2010..

[5]. V. P.T.L.Thuyền, " “Expansion of UNL –Vietnamese dictionary on UNL explorer”," Journal of Science and technology, Vols. N.11(84), 2014. , no. P.T.L.Thuyền, V.T.Hung, “Expansion of UNL –Vietnamese dictionary on UNL explorer”, Journal of Science and technology, The University of Danan N.11(84), 2014. , p. , 2014. .

[6]. B. K. D. D. V. S. "Translation Challenges and Universal Networking Language," International Journal of Computer Applications , no. (0975 – 8887) , 32. Baljeet Kaur Dhindsa and Dharam Veer Sharma Translation Challenges and Universal Networking Language International Journal of Computer Applications (0975 – 8887) Volume 133 – No.15, January 2016. 32. Baljeet Kaur Dhindsa and Dharam Veer Sharma Translation Challenges and UNL International Journal of Computer Applications (0975 – 8887) Volume 133 – No.15, January 2016. 15, January 2016..

[7]. A. I. f. R. "concept of UNL," American Institutes for Research with Juárez and The Natoma Group, American Institutes for Research with Juárez and The Natoma Group, 2011..

[8]. ALPAC, "Language & Machines : Computers in Translation & Linguistics," National Academy of Sciences, Washington.D.C., 1966.

[9]. W. &. L. E. Hutchins, "Petr Petrovich Troyanskii (1894-1950): a forgotten pioneer of machine translation," Machine Translation, vol. 15 , no. 3, pp. 187-221, 2000.

[10] W. J. Hutchins, "Machine translation: half a century of research and use," http://ourworld.compuserve.com/homepages/WJHutchins, 2003.

[11] M. . R. Costa-jussa, Hybrid Approaches to Machine Translation, Switzerland: Springer, 2016.

[12] www. unlweb.net, "“Scope and goals of UNL”," [Online]. Available:http://www.unlweb.net. [Accessed [Accessed 3 June 2019].

[13] C. S. Agency, "Addis Ababa, Ethiopia: Central Statistics Agency," Central Statistics Agency, 2007. [Online]. Available: http://www.csa.gov.et . [Accessed june 19 june 2019].

[14] L. Paul, "Language of the World. Sixteenth edition.," SIL International Publications, Dallas, Texas:, 2009.

[15] Michael Melese Woldeyohannis, "Amharic Speech Recognition for Speech Translation," Actes de la conférence conjointe JEP-TALN-RECITAL , vol. volume 11 , no. : TALAF, p. 4, 2016,.

[16] G. Hudson, The World’s Major Languages : Amharic. In The World’s Major Languages., Oxon and New York:: Routledge, 2009.

[17] P. Bhattacharyya, Machine Translation, London: CRC Press, 2015.

[18] H. Shamsabadi, A. M. Nafchi and F. Afzali, "A Study of Computational Linguistics and Machine Translation," International Journal of Applied Linguistics and Translation, vol. 1, no. 5, pp. 68-72, 2015.

[19] C. Goutte, N. Cancedda, M. Dymetman and G. Foster, Learning Machine Translation, Massachusetts: The MIT Press, 2009.

[20] Y. Wilks, Machine Translation: its Scope and Limits, Sheffield: Springer, 2009.

[21] W. Hutchins, "From first conception to first demonstration: the nascent years of machine translation," Machine Translation, vol. 12, no. 3, pp. 195-252, 1997.

[22] L. Han, "Machine Translation Evaluation Resources and Methods: A Survey," Ireland, 2018.

[23] S. Saini and V. Sahula, "A Survey of Machine Translation Techniques and Systems for Indian Languages," Jaipur, 2015.

[24] N. T. Alsohybe, N. A. Dahan and F. M. Ba-Alwi, "Machine-Translation History and Evolution: Survey for Arabic-English Translations," Current Journal of Applied Science and Technology, vol. 23, no. 4, pp. 1-19, 2017.

[25] L. Qingjun, L. Lina and G. Caiyun, "The Application of Internet Technology in Translation," Advances in Intelligent and, vol. 139, pp. 127-132, 2012.

[26] B. K. Dhindsa and D. V. Sharma, "Translation Challenges and Universal Networking Language," International Journal of Computer Applications, vol. 133, no. 15, pp. 36-40, 2016.

[27] "UNL:UNL Sentence, "[Online].Available http://www.unlweb.net/wiki/UNL_sentence.. [Accessed [Accessed 7 Jan 2019]:

[28] "the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III)," Mumbai, December 2012..

[29] M. A. Chéragui, "Theoretical Overview of Machine translation," 2012.

[30] J. Hutchins, "Multiple Uses of Machine Translation and Computerised Translation Tools," 2009.

[31] J. Cardeñosa, A. Gelbukh and E. Tovar, Universal Networking Language: Advances in Theory and Applications, Mexico: Centre for Computing Research, 2005.

[32] D. Arnold, L. Balkan, S. Meijer, R. L. Humphreys and L. Sadler, Machine Translation: An Introductory Guide, USA: NCC Blackwell, 2001.

[33] S. Dubey, "Survey of Machine Translation Techniques," International Journal of Advance Research in Computer Science and Management Studies, vol. 5, no. 2, pp. 39-51, 2017.

[34] S. R. "Machine Translation:," New Delhi, 1994. 2002..

[35] V. T. H. "“Reuse of Free Online MT Engines to Develop a Metasystem of Multilingual Machine Translation”," EsTAL , EsTAL , 2004,.

[36] I. B. A. c. "Information Based Approach," Mumba, 2000..

[37] P. K. R. S. "“Punjabi to UNL EnConversion System”, “DeConverter Specifications.," Springer: Sadhna, Academy Proceedings in Engineering UNL centre,, vol. Version 2, 2012. .

[38] R. M. "“Semantically relatable sets: building blocks for representing semantics,”," Proc. 10th MT Summit, Phuket, 2005.

[39] K. Yitbarek, "Implement Amharic Sentence generates from Interlingua Representation," Addis Abeba, 2016.

[40] D. P. B. Kanu Goel, "Information Retrieval System using UNL for Multilingual Question Answering," COMPUTER SCIENCE AND ENGINEERING DEPARTMENTTHAPAR UNIVERSITY, PATIALA, June 2016.

[41] . R. M. e. ""“Hermeto: A NL–UNL Enconverting Environment,”," in Universal Network Language: Advances in Theory and Applications,, Vols. vol. 12,, no. J. Cardeñosa , A. Gelbukh , and E. Tovar, Ed(s). México, Rese, pp. pp. 254-260., 2005.

[42] K. D. . P. B. ""“Universal Networking Language based analysis and generation of Bengali case structure constructs,”," México, Research on Computing Science,, Vols. vol. 12,, no. J. Cardeñosa , A. Gelbukh , and E., pp. pp. 215-229., 2005.

[43] E. B. ""“About and around the French Enconverter and the French Deconverter,”," México, Research on Computing Science,, Vols. vol. 12,, no. J. Cardeñosa , A. GelX-Axis:- Part-ofspeeches Y-Axis:- F-Measurebukh , and E. Tovar, Ed(s)., pp. pp. 157-166., 2005,.

[44] E. I.M. Boguslavsky, ""“Interactive EnConversion by means of the ETAP-3 system,”," México, Research on Computing Science,, Vols. vol. 12,, no. J. Cardeñosa , A. Gelbukh , and E. Tovar, Ed(s). , pp. pp. 230-240., 2005,.

[45] E. M. Choudhury, ""“Bridging Bangla to Universal Networking Language- a human language neutral meta-language,”," Proc. 8th Int. Conf. on Computer and Information Technology,, vol. 2, no. Dhaka, Bangladesh,, pp. 104-109., 2005, .

[46] M.Lafourcade, ""“Semantic analysis through ant algorithms, conceptual vectors and fuzzy UNL graphs,," México, Research on Computing Science, , vol. vol. 12, no. J. Cardeñosa , A. Gelbukh , and E. Tovar, Ed(s). México,, pp. pp. 125-137., 2005,.

[47] http://www.unl.unu.edu, "unls30.doc," [Online]. Available: [accesed jan 20 1918].

[48] B. K. D. a. D. V. Sharma, "Translation Challenges and Universal Networking Language International," International Journal of Computer Applications, vol. Volume 133 – No., no. (0975 – 8887), 2016..

[49] T. undl.org, [Online]. Available: http://www.unlweb.net/wiki /. [accesed jan, 20 1919].

[50] e. Mequannint Munye, "Amharic-English Bilingual Web Search Engine," Addis Ababa, Ethiopia., 2012,.

[51] Abyssinica, "Amharic, the Official Language of Ethiopia," 2015. [Online]. Available: available at http://dictionary.abyssinica.com/amharic.aspx. . [Accessed june 19 june 2019].

[52] Tadesse Beyene, “The Ethiopian Writing System”,, Paper presented at the 12,Conference of Ethiopian Studies: Michigan State University, 1994.

[53] B. Yimam., Amharic sewasew, Addis Ababa: TMMAD, ethiopian, 1986..

[54] R. e. a. .. Cowley, The Amharic Language-Description. In Language in Ethiopia., London:: Oxford University Press., 1976.

[55] T. E. W. S. i. B. M. e. a. (. ,. Bender L.M. and Ferguson C., Languages in Ethiopia,, London: Oxford University Press,, 1976.

[56] D. Bahdanau, K. Cho and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," 2015.

[57] A. Lifeng Han, " An Augmented Machine Translation Evaluation Metric," Faculty of Science and Technology , University of Macau, 2014 .

[58] J. Ollie, "Machine translation [Online]. Available: http://en.wikipedia.org," [Accessed 26 jan 2019].

[59] B. K. D. a. D. V. Sharma, "Translation Challenges and Universal Networking Language International," International Journal of Computer Applications, vol. Volume 133 – No., no. (0975 – 8887), 2016..

[60] B. H. "Perspectives of DBMT for monolingualauthors on the basis of LIDIA-1, an implemented mockup.," COLING-94,, 1994..

[61] B. É. G. P. "Developing MT lingware through Internet : ARIANE and the CASH interface.," Ohme, Japon, 1/1, pp. 15-22., 1997..

[62] Q.-A. M. . G. P. Boitet C., "an integrated environment for automated translation and human revision. Proc.," ARIANE-78,, ARIANE-78,, 1982..

[63] e. a. R. T. Martin, "“Specification of the UNL-Portuguese enconverter-deconverter prototype,”," Relatórios Técnicos do ICMC-USP,, Vols. , pp., no. pp.1-10., pp. 1-10., 1997.

[64] "the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III)," Mumbai, December 2012..

APPENDIX I – Corpus UC-AI

Table 17: Corpus UC-A1

ጥንቸል	ቀበሮው መስመሩን መምረጥ አለበት
ጥንቸሉ	ቀበሮው መዳረሻውን መወሰን አለበት
ኤሊው	ቀበሮው መስመሩን መምረጥ እና መዳረሻውን መወሰን አለበት
ውድድር	እኔ በውድድር ውስጥ አንተን አሸንፍሃለሁ
ቀርፋፋው ጉዞ	ጥንቸሉ በጥያቄው ተስማማ
አጫጭር እግሮች	ጥንቸሉ ከመንገዱ ዳር ጋደም አለ
የኤሊው ቀርፋፋው ጉዞ	ጥንቸሉ ከዛፍ ስር አንቀላፋ
የኤሊው አጫጭር እግሮች	ቀርፋፋ ነገር ግን ወጥ ጥረት ያሸንፋል
የመስመሩ መጨረሻዉ	ኤሊው ውድድሩን አስቀድሞ አሸንፏል
ጥንቸሉ እና ኤሊው	ኤሊው ቀጥታ ወደ መስመሩ መጨረሻው ሄደ
ቀርፋፋ ነገር ግን ወጥ ጥረት	ኤሊው በቀርፋፋ ነገር ግን ወጥ ጉዞ ሄደ
ቀርፋፋ ነገር ግን ወጥ ጉዞ	ኤሊው በቀርፋፋ ነገር ግን ወጥ ጉዞ ቀጥታ ወደ ውድድሩ መጨረሻው ሄደ
የኤሊው አጫጭር እግሮች እና ቀርፋፋው ጉዞ	ሁለቱም አብረዉ ጀመሩ
ኤሊው መለሰ	አንተ ፈጣን ነህ
ኤሊው አልቆመም	አንተ እንደ ንፋሱ ፈጣን ነህ
ጥንቸሉ ነቃ	ፉከራው አይሳካም
ኤሊው ሄደ	ፉከራው እንደዋዛ አይሳካም
ጥንቸሉ አንቀላፋ	ጥንቸሉ በሚችለው በፍጥነት ሮጠ
እኔ አሸንፍሃለሁ	እሱ በጣም ዘግይቶ ነበር
ኤሊው ውድድሩን አሸንፏል	ጥንቸሉ ፉከራው እንደዋዛ እንደማይሳካ ያምናል

ጥረት ውድድርን ያሸንፋል	አንተ እንደ ንፋሱ ፈጣን ብትሆንም እኔ በውድድር ውስጥ አሸንፍሃለሁ
ጥንቸሉ በአጫጭር እግሮቹ አላገጠ	ጥንቸሉ ከመንገዱ ዳር ጋደም አለ እና ከዛፍ ስር አንቀላፋ
ጥንቸሉ በኤሊው አጫጭር እግሮቹ አላገጠ	ቀበሮው መስመሩን መምረጥ እና መዳረሻውን እንዲወስን ተስማሙ
ጥንቸሉ በኤሊው አጫጭር እግሮቹ እና ቀርፋፋው ጉዞ አላገጠ	በመጨረሻም ጭንቸሉ ነቃ እና በፍጥነት ሮጠ ነገር ግን በጣም ዘግይቶ ነበር
ጥንቸሉ አንድ ቀን በኤሊው አጫጭር እግሮቹ እና ቀርፋፋ ጉዞ አላገጠ	ጥንቸሉ ፉከራው እንደዋዛ እንደማይሳካ ያምናል እና በጥያቄው ተስማማ

APPENDIX II – Relations and Attributes

Table18: UNL Relations 1

Agt	Agent	a thing which initiates an action
And	conjunction	a conjunctive relation between concepts
Aoj	thing with attribute	a thing which is in a state or has an attribute
Bas	Basis	a thing used as the basis(standard) for expressing degree
ben	Beneficiary	a not directly related beneficiary or victim of an event or state
cag	co-agent	a thing not in focus which initiates an implicit event which is done in parallel
cao	co-thing with attribute	a thing not in focus is in a state in parallel
cnt	Content	an equivalent concept
cob	affected co-thing	a thing which is directly affected by an implicit event done in parallel or an implicit state in parallel
con	Condition	a non-focused event or state which conditioned a focused event or state
coo	co-occurrence	a co-occurred event or state for a focused event or state
dur	Duration	a period of time during an event occurs or a state exists
fmt	Range	a range between two things
frm	Origin	an origin of a thing
gol	goal/final state	the final state of object or the thing finally associated with object of an event

ins	Instrument	the instrument to carry out an event
man	Manner	the way to carry out event or characteristics of a state
met	Method	a means to carry out an event
mod	modification	a thing which restrict a focused thing
nam	Name	a name of a thing
obj	affected thing	a thing in focus which is directly affected by an event or state
opl	affected place	a place in focus where an event affects
Or	Disjunction	disjunctive relation between two concepts
per	proportion, rate or distribution	a basis or unit of proportion, rate or distribution
plc	Place	the place an event occurs or a state is true or a thing exists
plf	initial place	the place an event begins or a state becomes true
plt	final place	the place an event ends or a state becomes false
pof	part-of	a concept of which a focused thing is a part
pos	possessor	the possessor of a thing
ptn	Partner	an indispensable non-focused initiator of an action
pur	purpose or objective	the purpose or an objective of an agent of an event or a purpose of a thing which exist
qua	Quantity	a quantity of a thing or unit
Rsn	Reason	a reason that an event or a state happens

Scn	Scene	a virtual world where an event occurs or state is true or a thing exists
Seq	Sequence	a prior event or state of a focused event or state
Src	source/initial state	the initial state of an object or thing initially associated with the object of an event
Tim	Time	the time an event occurs or a state is true
Tmf	initial time	the time an event starts or a state becomes true
Tmt	final time	the time an event ends or a state becomes false
To	Destination	a destination of a thing
Via	intermediate place or state	an intermediate place or state of an event

Table 19: UNL Attributes 1

Describing logicality	@transitive, @symmetric, @identifiable, @disjointed
Describing times	@future, @past, @present
Describing aspects	@begin, @complete, @ continue, @end, @progress, @state...
Describing genericity & specificity	@generic, @def, @indef, @not, @ordinal
Describing emphasis, focus & topic	@emphasis, @entry, @focus, @topic...
Describing attitudes	@affirmative, @imperative, @interrogative, @ request.
Describing feelings & judgments	@ability, @grant, @wish, @will, @obligation, @possible, @regret,...

APPENDIX III – Certifications

CERTIFICATIONS

UNDL FOUNDATION

Universal Networking Language

CLEA1000

Certificate of Language Engineering Aptitude

The UNDL Foundation certifies that

henok meskele

has successfully completed the exercises leading to the Certificate of Language Engineering Aptitude (CLEA1000), offered by the Virtual Learning Environment for UNL (VALERIE).

Geneva, 29 March 2018

Tarcisio G. Della Senta, PhD
President
UNDL Foundation

Ronaldo T. Martins, PhD
Language Resources Manager
UNDL Foundation