"A SURVEY PAPER FOR HISTORY AND EVALUATION SYSTEM OF MT IN NLP TASKS FOR AMHARIC LANGUAGE"

Prof (Dr). Jayant Shekhar

Professor, Adama Science and Technology University, Adama

Mr. HENOK MESKELE

PG Student, Adama Science and Technology University, Adama

Dr. Mesfin Abebe Haile

CSE-PG-Coordinator, Adama Science and Technology University, Adama

I. ABSTRACT

In the field of Natural Language Processing society, Universal Networking Language has been worked an area of new multi-lingual MT approached for many researchers during last couple of years. UNL is an artificial language used for representing and generation information in a natural-language-independent format and provides unambiguity meaning across multiple languages. This study performs the UNLization and NLization processes of all the major Part-of-Speech for Amharic language, in both cases it is going to be being enhancement of the performances of the systems will have an input Analysis as well as output generation Dictionary entries and limited grammatical syntax structure: like Transformation Rules with disambiguate rules.

This study focused on UNLization, is a process of converting Amharic sentences to UNL using IAN, It is an online tool processes input sentences into UNL and NLization is a process of generating an Amharic sentences out of a UNL using EUGEAN, It is an online tool processes input UNL into Amharic sentences. A detailed description of IAN tool and EUGEAN tool has also been explained in chapter 3 and 4 later. The researcher, proposed a system that has also been encoding corpus UC-A1 which is 50 sentences manually translated into Amharic and try to input with IAN per including the dictionary entry and grammar syntax which is used to make Enconverter functions, Decoding corpus UC-A1 from UNL 50 sentences automatically try to generate or translated into Amharic sentences using EUGEAN per including the dictionary entry and grammar syntax which is used to make Deconverter functions for Amharic language under UNL.

The system is tested on corpuses UC-A1 of 13 Amharic sentences provided by UNDL Foundation and finally calculates Grammar accuracy through F-1 score that is available on the developmental page of UNDL foundation, which is considers both the

precision and recall of the grammar to compute the score its value lays on 0 to 1.

II. INTRODUCTION

The automation of translation has been one of oldest humanity's dreams and various researchers' are trying an attempt to fulfill it. It has been an area of enormous interest among researchers during last couple of decades and also known as "automatic translation" or "mechanical translation," is the name for computerized methods that automate all or part of the process of translating from one human language to another ("History of UNL, 2013.). It is a multi-disciplinary field of research as it incorporates ideas from linguistics, computer science, artificial intelligence, statistics, mathematics, philosophy and many other fields ("History of UNL, 2013.).

UNL based MT is appeared as a new approach and also in this direction one type of intermediate Interlingua representation platform for handling Multi-language environment over internet. It is an artificial language for representing, describing, summarizing, and storing information in a

natural-language-independent format. In the UNL approach, there are two basic movements: UNLization and NLization. UNLization is the process of representing and analyzing the information conveyed by NL statements into UNL and NLization is the process of generating and mapping a NL document out of a UNL graph.[(Chawla, 2019)] These processes are completely independent. Currently, the main goal of the UNLization process has been to convert the given information text into the intermediate language and vice versa i.e., UNL which is of language-independent and machine-tractable systems that have three components UWs database, with the help of many relations, and attributes.

In the field of NLP, UNL has been used by various researchers as an Interlingua approach for automated MT system ("History of UNL, 2013.). It has consists of two main components namely, Enconverter used for converting the text from a source language to UNL and Deconverter used for converting from UNL to a target language (P.T.L.Thuyền, 2014.). Currently, many scholars are studying to apply UNL in different languages in other countries, but till now Multi lingual UNL based Amharic MT system has not been considered in a well manner. Amharic is the official language of

Ethiopia which is near to international languages. So the translation of Amharic language has its own importance to the local as well as international community. In this study, the researcher try to introduced the tools that are UNL's applications of encoding process for Amharic sentence into UNL expression and decoding the UNL expression into Amharic sentence. It is impossible for human to manually translate huge number of documents.

✓ **Building Blocks of UNL**

UNL represents the information in three different types of semantic units, namely, Universal Words (UWs), Relations and Attributes (J. Ollie, 2018.). UNL represents information sentence by sentence. Each sentence is converted into a hyper-graph (also known as UNL graph) having concepts represented as nodes and relations as directed arcs. The concepts are represented by UWs and UNL relations are used to specify the role of each word in a sentence. The subjective meanings intended by the writer are expressed through UNL attributes. UNDL Foundation has formally defined the specifications of UNL

➢ **Universal Words**: Although UWs mainly used natural language words, it has some words of other languages and the semantic information to

define the concept of natural language. In this way, it allows to limit the inherent ambiguity of words in natural language.

- **Relations**: There are 56 relations in UNL, which are used to connect two or more UWs used to build semantic network of a UNL expression. These relations are the edges of the UNL graph or binary relations which directly generate UNL expressions.

- **Attributes: -** There are 87 attributes to clarify the semantics of a sentence (Amelina, 2010.). It is used for the purpose of describing the subjective information of the sentences. They show speaker's point of view. The attributes include information about time or aspect of the event, number, polarity, modality, etc.

III. BACKGROUND

In this technologically integrated world, most people can easily access a huge volume of multi-lingual electronic information in the common platforms like internet, but the problem comes when the content is written in the language which is not recognizable by a user. 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 translation is emerged as a viable tool to solve language translation problems and comes in the focus of linguists, computer scientists and engineers to propose improvements in the translation (Dubey, 2017).

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. Translation naturally became a necessity in such circumstances. Nowadays, navigations through physical landscape have been augmented at a much larger scale with explorations in the virtual world. People are much better connected. But language barriers remain, and pose a challenge to communication (Bhattacharyya, 2015).

In Ethiopia it has more than 80 nations and peoples are speaking in their own native languages. Amharic (አማርኛ) is the official language of Ethiopia and also one of a member of the Semitic branch of the Afro-Asiatic language family next to Arabic. It has many ethnic groups and peoples are speaking as their own native languages in Ethiopia. Automatic multilingual MT system to Amharic language must certainly be very useful for more than million Amharic

language users by their native language to communicate for others languages internationally and locally. This study tries to integrate Amharic local language into UNL intermediate MT environment.

History: The UNL Programme started in 1996, as an initiative of the Institute of Advanced Studies of the United Nations University in Tokyo, Japan in January 2001 (J. Ollie, 2018.). The UNU set up an autonomous organization, the UNDL Foundation, to be responsible for the development and management of the UNL Programme. The Foundation, a non-profit international organization, has an independent identity from the UNU, although it has special links with the UN. It inherited from the UNU/IAS the mandate of implementing the UNL Programme so that it can fulfill its mission. Its headquarters are based in Geneva, Switzerland. From the very beginning, a consortium of university departments from all regions of the world has been engaged in developing the UNL. That's the UNL Society, a global-scale network of Research and Development teams, involving several specialists in computer science and linguistics, who are at work creating the linguistic resources and developing the web structure of the UNL System. The UNDL Foundation provides technological support and co-ordinates the implementation of the Programme ("History of UNL, 2013.). It has already crossed important milestones. The overall architecture of the UNL System has been developed with a set of basic software and tools necessary for its functioning (More information about the earlier system can be found, (Sharma, 2016.) (the UNL 3rd Workshop, , December 2012.), (Sinha, R.M.K.,, 1994. 2002.) (Information Based Approach conference, 2000.) & (http://www.unl.unu.edu, 1918.) These are being tested and improved.

IV. STATEMENT OF THE PROBLEM

Amharic language is the most widely used spoken language in Ethiopia. There are relatively fewer contributions in the field of computerization and development of this language under UNL system. Other tools that were used previously are old-fashioned and outdated, since they were very complex and failured to being integrated on a central site. But, IAN and EUGEAN have been accepted globally as a standardized tool for UNLization and NLization. So, meanwhile there is no central automatic UNL based MT system for Enco and Deco to Amharic language, it makes it difficult to translate the language to other local, international

languages and vice versa, it is much further difficult for humans to translate vast amount of documents. Even though different multi-lingual approaches were used as a solution to MT system, for **N** number of languages, in all approaches, it requires **n*(n-1)** components to be developed ("Universal Networking Language:UNL Sentence, 2018.) (Baljeet, 15, January 2016.), which needs large amount of computational devices and it takes more time to developing translation rules.

Furthermore, the previous approaches lack to centrally managing the system due to in compatibilities frameworks in the case of varied of systems with each other. On the other side, all languages have no enough resources over internet which is language barrier. So, any needed information cannot be accessed by a desired language. The Internet has emerged as the most powerful networking infrastructure for communication. By using the Internet, people all over the world can exchange any information to each other at anytime and anywhere. However, there is still a language barrier which prevents people in different countries from communicating by their own language (P. Kumar , 2012.) To this end a number of strategies have been proposed including the production of MT tool.

Certainly, many researchers, research group have come up with different translation systems in different approaches. It is concluded that direct approach for MT is most suitable for closely related languages i.e. the languages with similar structure. The transfer approach 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 (unlweb.net, 2013.), (undl.org, 1918.). But, since, the Amharic language is still one of the less resourced languages; it is very difficult to find an Amharic corpus whose size could be larger than thousands of sentences under UNL. The best approach is the UNL Interlingua approach with in n*2 approach Enco and Deco functions, 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 UNL Interlingua approach.

Linguistic barriers damage with the enhancement of linguistic diversity over internet and the value that native language as one of the main means of communication to express one's cultural identity. Apart from sociocultural issues, linguistic diversity also

knows as economic and political dimension (Cardeñosa, et al., 2005). Institutions like the United Nations, Federal States Based on language like Ethiopia and India or the African Union have to face every day with the barriers that linguistic diversity enacts. It is well known that the huge amount of documentations these institutions produce every day, which have to be produced in all their official languages: 6 for the UN, About 80 in our country 26 for the European Union (UNL:, 2005). As Ethiopian context federal government and regional states use more than two languages for their day today activity. It is simply unfeasible to rely on human translators for the production of all these amounts of documentation. (Cardeñosa, et al., 2005).

The exponential growth of the Internet has made its content increasingly difficult to find, access, present and maintain for a wide variety of users (P.T.L.Thuyền, 2014.). In addition, the current Internet's content was designed for humans to read, not for computer to manipulate meaningfully (V. T. Hung, , 2004,). It means the available functions of computer are limited in just retrieving information and displaying to people without processing the semantics. Therefore, computers give little support for people in finding, accessing and maintaining the Internet's content. To deal with the above problems, UNL was defined as an artificial language that is able to represent information and knowledge described in NL (Boitet C., 1982.). Thus, it enables computers to process information in form of knowledge across the language barrier (Blanchon H., 1994.) (Blanc É. , 1997.). While lots of different systems have been developed by various organizations; each has their special representation of a given language. This results in incompatibilities between systems. Then, it is impossible to break language barriers in all over the world, even if we get together all the results in one system.

Against this backdrop, the concept of UNL as a common language for all computer systems was born, With the approach of UNL, the results of the past research and development can be applied to the present development, and make the infrastructure of future research and development (American Institutes for Research, 2011.).This study greatly reduces developmental cost of knowledge or contents necessary to make knowledge processing by sharing knowledge and contents.

V. MOTIVATION

The increasing needs and use of Internet for cultural, educational & commercial services are improved, but many of the people are not satisfied with an internet usage due to linguistic diversity, which indicate that reduces the possible of Internet as a vehicle of knowledge for everybody (Blanc É. , 1997.). Since the advent of computers, researchers around the world have worked towards developing a system that would overcome the language barriers over net (Blanc É. , 1997.).

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. A lot of research was done in the field of MT local and abroad with various approaches, but still now will not resolve multi-linguistic barriers on over net. The Internet changes the world very fast. Now we can find vast amount of knowledge on Internet. 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 MT such as searching data and information in any languages unknown or poorly known by the user is a challenging obstacle. "Cross Language Information Retrieval Systems" are designed in this direction. The UNL system revolves around a unique artificial language that captures the meaning of written documents. This language is based on the representation of concepts and its relations. Therefore, the researcher is interested to study the UNL based multi-lingua MT for Amharic language because the researcher is native Amharic speaker and is eager to familiarize Amharic language to other languages using UNL based translation system. Additionally, the researcher has good attitude to face challenges of multi-lingua based MT system for Amharic local language.

VI. RESEARCH QUESTIONS

This study contributed to make the awareness of various suitable MT systems for Amharic language that covered in all major part of speech and find out which one is the best performances by answering the following questions.

1. What are the main parts of speech is translated and which errors occur most frequently in our corpus?

2. What are the types of existing MT system appropriate for multilingual approach and what is the most common multilingual MT is suitable for Amharic language?

3. What solutions are best those have not yet been tested for Amharic language, which one is the most effective?

VII. SIGNIFICANCE OF THE STUDY

From this study, institutions which are using more than two different languages are the immediate beneficiaries from the developed systems. The prototype also has great significance for researchers in the field of NLP, in governmental and regional offices to provided good services for their customers without language drawback. There is also a new opportunity for accessing high quality Amharic language MT system with additional improvements to increasing speed of access multiple linguistics resources via UNL web by Internet technologies which would increase usage as well as other application services.

VIII. LIMITATIONS OF THE STUDY

The researcher of this study also faced the following limitations which might reduce the quality of the under taken research. Moreover, No one conduct research without some obstacles. The most challenges were the following:

o It will be better if the study had been included many samples. But, the study's sample size is not being large enough due to financial and time constraints; it can be difficult to generalize the findings of the study at all regional languages of Ethiopia.

o It is very important if it was conducted in all regional states official languages that found in Ethiopia, but the study will become developed for currently Ethiopian federal government official language Amharic only.

o Very limited availability access documents related to the UNL implementation.

o The system at this time is not able to handle very large and complex sentences.

- o Very few literature researches on UNL in Ethiopians language as a reference.
- o Financial & time constraint blocked to fully cover all parts of grammatical units of Amharic language.

IX. LITERATURE REVIEW

✓ TRANSLATION

Translation is a complex phenomenon that has a huge effect on everyday life. This can range from the translation of a key international treaty to the following multilingual poster that welcomes to a small restaurant. Translation tasks can be categorized broadly into human or manual translation and Automatic Machine Translation or simply Machine Translation (MT).

✓ EXPERT HUMAN TRANSLATORS

Expert human translators use their background knowledge, mostly subconsciously, in order to resolve syntactical and semantical ambiguities which machines will either have to leave unresolved, or resolve by some 'mechanical' rule which will ever so often result in a wrong

translation (Wilks, 2009).These experts seem to agree that the three requisites in a

translator, in order of importance, are (1) good knowledge of the target language, (2) comprehension of the subject matter, and, (3) adequate knowledge of the source language. Furthermore, while good translations are made by some translators who have a general appreciation of scientific knowledge, the best technical translations are generally made by experts in the technical field covered. It also seems clear that a restricted competence in the source language is adequate when the translator is expert in the subject matter (ALPAC, 1966). 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 seems as though automation of translation is a social and political necessity for modern societies which do not wish to impose a common language on their members (Arnold, et al., 2001).

✓ MACHINE TRANSLATIONS

MT presumably means going by algorithm from machine- readable source text to useful target text, without recourse to human translation or editing (ALPAC, 1966).The whole chain of human language relates to two instances. The first one is when the speaker gives utterance to words and the second is when these words reach the

listener's ears and understands them. The processes of written text MT form the main and basic links of this chain, because it is absolutely necessary to face them when speech MT is being conducted. It is mandatory to analyze the translation and substitution of speech from vocal signs to written text in the source language for it to be usable by a machine. At this step, the translated texts will be processed for creating and compounding corresponding and fitting vocal signs in the source language. Performing all these productive, substitutional, and analytical processes between speaking and writing are on the responsibilities of electronics experts (Shamsabadi, et al., 2015).

MT 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 (Arnold, et al., 2001). 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 (Shamsabadi, et al., 2015).

✓ **INTERNET & MACHINE TRANSLATION**
Translation is a complex activity which involves knowledge of many fields, no matter what type of translation belongs to, there are two phases: comprehension and presentation. Usually the difficulties of translation are not caused by the incomprehensible words, but due to the lack of background knowledge, especially when translating unfamiliar materials, the translator must know the basic background knowledge, but the translator's personal knowledge and information are limited, so translators need to have a large number of reference materials, such as encyclopedia, dictionaries, and so on. But it's very difficult for the translators to own these traditional references due to the objective conditions. Now, the network provides a huge shared data bank for the translator, which includes encyclopedia, the dictionaries and some special resources (Qingjun, et al., 2012).

Moreover, with the widespread and growing use of mobile telephones, there are an increasing number of manufacturers providing translation software for these devices (Hutchins, 2009). All of these provide many conveniences to the resolution of translation difficulties. With the network technology, the translator can find related information to get a general understanding of the related subject through search engines, online encyclopedia, electronic dictionary, online terminology and online newspapers and magazines, thus the translation quality can be improved by reducing understanding errors. As time passes, the translator will use these knowledge in the later translation practice, thus the translator's ability to understand will be expanded and enriched (Qingjun, et al., 2012).

The emergence of Internet influenced the MT development. These developments include (Hutchins, 2003):

- o The appearance MT software products specifically for translating Web pages and electronic mail messages
- o 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.

- o The Internet has also encouraged somewhat less reliable companies to offer online versions of electronic dictionaries (or phrase books) as 'translation systems'. Anyone using such products for translating full sentences (and text) is bound to get unsatisfactory results – although if they do not know the target languages they will be unaware of the extent of the incomprehensibility of the results.
- o The Internet suggests other roles for MT, as adjuncts (or components) of other language-based operations. Searching data and information in languages unknown or poorly known by the user is 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 (Hutchins, 2003).

X. 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 (Costa-jussa, 2016).
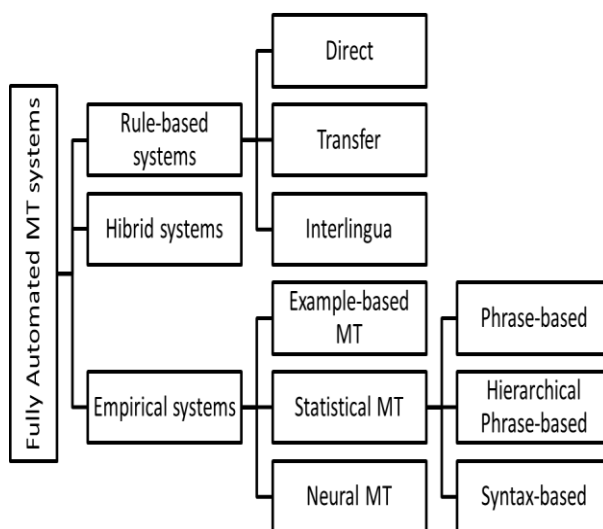

Fig.1. Taxonomy of fully automated machine translation systems.

XI. RULE-BASED SYSTEMS

The amount and the linguistic techniques used in a translation system distinguishing direct, transfer and Interlingua rule based MT approaches (Costa-jussa, 2016).
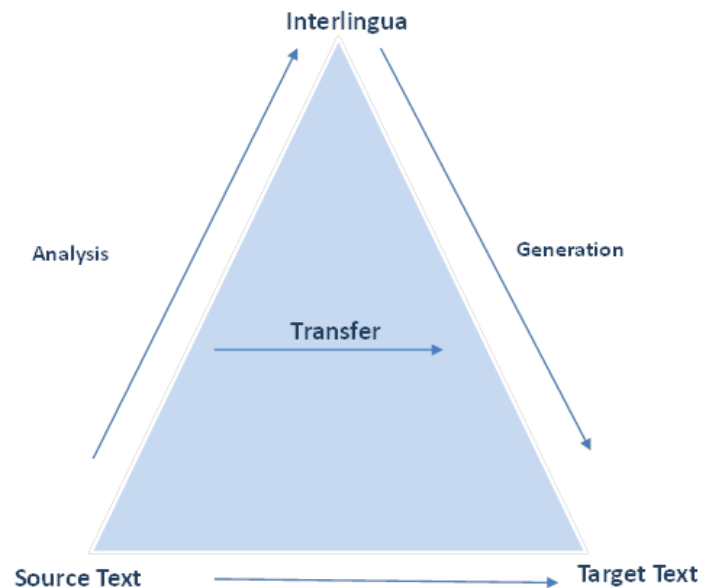

Fig.2. Vaquees triangle.

✓ Direct

In this type of system the source language is given as input and the output is known as the target language. Typically, the approach is unidirectional and only takes one language pair into consideration at a time. To get a target translation of any source sentence one needs (Saini & Sahula, 2015):

- A dictionary that will map each source word to an appropriate target word.
- Rules representing regular source sentence structure.
- Rules representing regular target sentence structure.

✓ Transfer

In the transfer approach, the source text is analyzed into an abstract representation that still has many of the characteristics of the source, but not the target language. This representation can range from purely syntactic to highly semantic. In the syntactic transfer, some type of tree manipulation into a target language tree converts the parse tree of the source input. This can be guided by associating feature structures with the tree. Whatever representation is used, transfer to the target language is done using rules that map the source language structures into their target language equivalents. Then in the generation stage, the mapped target structure is altered as required by the constraints of the target language and the final translation is produced (Chéragui, 2012).

✓ Interlingua

Interlinguas are semantic languages designed to represent the meaning of any given text, ideally

they are language neutral and are precise, unambiguous, and formal languages. Interlinguas are intimately tied up with ideas about the representation of meaning, being meaning the most abstract and deepest level of linguistic analysis (that should be common to all languages, far enough from surface representation of languages).An Interlingua is "another language" in the sense that it has autonomy and thus its components need to be defined: vocabulary and "relations" mainly. Besides, and Interlingua is an artificial language that should be as expressive as natural languages (Cardeñosa, et al., 2005).

✓ Universal Networking Language (UNL)

UNL is an Interlingua that was proposed in 1996 by the UNU, to access, transfer, and process information on the Internet in the natural languages of the world. UNL represents information sentence by sentence. Each sentence is converted into a hypergraph having concepts as nodes and relations as directed arcs (Bhattacharyya, 2015). UNL was launched to erase linguistic barriers. Linguistic barriers limit with the enhancement of linguistic diversity and the value of native languages as one of the main means to express one's cultural identity. Apart from sociocultural issues, linguistic diversity also

knows an economic and political dimension. Institutions like the United Nations or the European Union have to face every day with the barriers that linguistic diversity imposes (Cardeñosa, et al., 2005).

✓ 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 MT systems stems from the failure of any single technique to achieve a satisfactory level of accuracy (Bhattacharyya, 2015).

XII. EMPIRICAL SYSTEMS

✓ Example-based MT

Example-based MT systems are trained from bilingual parallel corpora containing sentence pairs. Sentence pairs contain sentences in one language with their translations into another. The particular example shows an example of a minimal pair, meaning that the sentences vary by just one element. These sentences make it simple to learn translations of portions of a sentence (Bhattacharyya, 2015).

✓ Statistical MT

SMT systems estimate probabilistic models (language and translation models) by frequency counts in corpora. For a language model, monolingual data in the target language is needed; translation probabilities are extracted from relations in both the target and the source language and therefore parallel data are used to build the translation model. At the end, the best translation in a statistical system is that maximizing the product of models. Differences in SMT systems mainly arise from the definition of the minimal unit for translation (word vs. phrase vs. constituent). Dictionaries and/or constituents obtained with RBMT systems can help in the translation as a complement to the parallel corpus, and rules can be used for reordering the input or the output as a pre-process or post-process respectively (Costa-jussa, 2016).

✓ Phrase-based

Phrase-based translation (PBSMT) is the natural evolution of word-to-word translation. When considering a phrase instead of a word, the (small) context and the local reordering of every word is taken into account. A phrase here is just a sequence of words but it is not necessarily a linguistic element consistent with a word alignment between the source and the target. This is positive because the number of phrases that can be extracted from a corpus is larger than the number of syntactic elements and it allows

more freedom in the translation. But on the other hand, the extraction of noisy phrases leads to non-grammatical translations. (Costa-jussa, 2016).

✓ Hierarchical Phrase-based

It consists of words and 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 (Costa-jussa, 2016).

✓ Syntax-based

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 (Costa-jussa, 2016).

✓ Neural MT

Neural machine translation (NMT) is an approach to machine translation that uses a large artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model. Deep neural machine translation is an extension of neural machine

translation. Both use a large neural network with the difference that deep neural machine translation processes multiple neural network layers instead of just one (Costa-jussa, 2016). The key benefit to the approach is that a single system can be trained directly on source and target text, no longer requiring the pipeline of specialized systems used in statistical machine learning. Unlike the traditional phrase-based translation system which consists of many small sub-components that are tuned separately, neural machine translation attempts to build and train a single, large neural network that reads a sentence and outputs a correct translation (Bahdanau, et al., 2015).

XIII. RESEARCH ACTIVITIES IN UNL.

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

✓ ENCO AND DECO MODULES
A prototype system for converting Brazilian Portuguese into UNL and Deconverting UNL expressions into Brazilian Portuguese with 'Enco' and 'Deco' tools, respectively have been proposed by (R. T. Martin,

1997). Their system consists of three important sub-modules, namely, the lexical, the syntactic and the semantic modules.

(R.T. Martins, et.al, 2005) have noted the 'Enco' and Universal Parser tools provided by UNDL foundation require inputs from a human expert who is seldom available and as such their performance is not quite adequate, They have proposed the 'HERMETO' system which converts English and Brazilian Portuguese into UNL. This system has an interface with debugging and editing facilities along with its high level syntactic and semantic grammar that make it more user friendly.

For developing a UNL based MT system Semantically Relatable Sequence (SRS) based approach has been used by (R. Mohanty,et.al, 2005). (P. Kumar , 2012.) Has been proposed an Enconversion system to convert Punjabi language to UNL.

(E. Blanc,et.al, 2005,) has performed the integration of 'Ariane-G5' to the proposed French Enco and French Deco; it is a generator of MT systems. In the proposed system, Enco takes place in two steps; first step is analysis of the French text to produce the representation of its meaning in the form of a dependency tree and second step is lexical and structural transfer from the dependency tree to an equivalent UNL graph.

(I.M. Boguslavsky, 2005,) Has been proposed a multifunctional linguistic processor, 'ETAP-3', as an extension of 'ETAP' machine translation system to a UNL based machine translation system.

(K. Dey, , 2005) Has presented the computational analysis of complex case structure of Bengali for a UNL based MT System. They provided the details of the rule theory of 'Enco' and 'Deco' tools which are driven by analysis rules and generation rules respectively. . (M. Choudhury, 2005,) have proposed a framework for converting Bangla to UNL and have also proposed a procedure to construct Bangla to UNL dictionary. The system developed by (M.Lafourcade, 2005,) uses ant colony algorithm for semantic analysis and fuzzy UNL graphs for Enconversion process.

✓ **THE 3RD WORKSHOP (DECEMBER 2012)**

According to the 3rd Workshop on Cognitive Aspects of the Lexicon (the UNL 3rd Workshop, , December 2012.), the first variance of UNL from existing Interlingua is described well but, it was not the only application of MT it is used to serve various NLP purposes as a language for meaning

representation. Another difference from others is to producing a UNL text is not supposed to be fully automatic it should be an interactive process with the labor divided the computer and a human expert "writer". One way of achievements (P. Kumar , 2012.). Due to the interactive Enco tools which serves as input text can be made as good as users 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.

✓ **MT IN ETHIOPIA**

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 that learn from others experience. Various MT systems introduces in many researchers come up with different approaches in the world it is concluded that direct approach is most suitable for closely related languages with similar structure. The transfer and SMT approach is suitable for languages with different structure both of these are required large corpora. The best method for multilingual environment is the UNL

Interlingua model it works n*2 Enco and Deco function within the evaluation methods used F-1 measure highest grammar accuracy through some dictionary entry rules on limited grammar syntax structures.

More recently, different dictionary-based Amharic English and Amharic-French CLIR experiments were conducted at a series of CLEF ad hoc tracks [Alemu et al., 2004; 2005; 2006]. The first Amharic-English information retrieval was conducted at CLEF 2004 with emphasis on analyzing the impact of the stop word lists of the source and target languages. While a bilingual dictionary was used to translate Amharic queries to English bags-of-words unmatched (untranslatable) source language terms were translated and added into the dictionary manually. Another similar dictionary-based Amharic-English CLIR experiment was conducted at CLEF 2006 by employing Amharic morphological analyzer and part of speech tagging to facilitate more accurate query translation. On the other hand, to the best study has been undertaken to Afaan Oromo CLIR or a number of other Ethiopian Languages. Thus, this is an initial research work that is intended to investigate and develop an Oromo-English CLIR system. (Kibrewossen yitbarek, 2016,) the first thesis work done on Amharic sentence generation from interlingua

representation UNL based Implement Amharic sentence generator that automatically generates Amharic sentence from a given UNL input with in the java program excluding the web portal and Enco function.

✓ **AMHARIC LANGUAGE IN MT**

Amharic language is the second most-spoken Semitic language in the world, next to Arabic, and the official working language of the Federal Democratic Republic of Ethiopia. Amharic uses a unique script, which has originated from ancient language, the Ge''ez alphabet, which 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. Outside Ethiopia, Amharic is used in Egypt, Israel, Sweden, Eritrea, and the United States. Thus, it has official status and is spoken by many people as their native and second language. In addition, it is a language with many literatures. Of these who speak Amharic, a significant number of them (usually the educated class) can understand and speak English as well (Mequannint Munye, 2012,).

Amharic has a complex morphology which combines consonantal roots and vowel intercalation. Amharic and English differ substantially in their morphology, syntax and the writing system they use. Therefore, the search engines which are mainly developed for English cannot efficiently be used to retrieve Amharic documents. Amharic language has the following characteristics that should be considered in IR and translation systems. (Mequannint Munye, 2012,).

Morphological variations: is morphologically complex words are inflected with prefixes, suffixes and infixes. For example, the Amharic word "በላ" has morphological variations such as: "በሉ" ፣ "በሊን"፣ "በሊች" ፣ "በሊሁ" ፣ "በሊችሁ"፣ "አስበሊ", etc.

Character Redundancy: it has some redundant symbols with the same sound in its alphabet. For example, ኣ, አ, ዓ and ዐ, ሰ and ሠ, and ጸ and ፀ have the same sound. These 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 ጥብ(.)). For example the Amharic words ወታደር, ዓመተ ምህረት,

ወልደ ስሊሴ can be shortened as ወ/ር, ዓ.ም, ወ/ስሊሴ, respectively. These shortened words will be expanded to their normal forms for dictionary lookup or IR systems.

Variations due to Pronunciations: Some words have multiple spelling variants like ጣት፣ ጥዋት፣ ጠዋት፣ ጠዋት፣ etc. 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 like ሚሊዮን፣ ሚልዮን፣ ሚልየን, etc. 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 information retrieval applications such as Google, which retrieves documents based on exact term matching.

Stop-words: Most search engines and query translation 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, ሁሉ, ሁኔታ, ሆነ, በሒሊ, በጣም, ብቻ, ወደ, ናቸው; etc are considered to be stop words.

Amharic Punctuation Marks : there are different punctuation marks and among these punctuation marks ።(አራት ነጥብ), ፣ (ነ ጠሊስረዝ), ፤ (ድርብ ስረዝ),?(የ ጥያቄ ምኸክት) are the most commonly used ones. They are used in query or document tokenization as word delimiters in addition to white spaces (Mequannint Munye, 2012,)

XIV. MT EVALUATION TECHNIQUES

✓ HUMAN EVALUATION METHODS

There are various types & new developments of human evaluation methods of machine translation outputs. These are summarized the table below.

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

Table 3. Human Evaluation Methods (Han, 2018)

✓ **INTELLIGIBILITY & 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.

✓ **FLUENCY, ADEQUACY AND COMPREHENSION**

✓ **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.

2.1..1. 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.

2.1..2. 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".

✓ **FURTHER DEVELOPMENT**

The "Linguistics Data Consortium" (LDC) develops two five-point scales representing fluency and adequacy for the annual machine translation 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.

2.1..1. 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 source language patterns which correspond to diagnostic target phenomena.

2.1..2. EXTENDED CRITERIA

- **suitability**, whether results are suitable in the particular context in which the system is to be used;

- **interoperability**, whether with other software or with hardware platforms;

- **reliability**, i.e., don'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 dealt documents;

- **maintainability**, being able to modify the system in order to adapt it to particular users; and

- **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.

2.1..3. 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.

2.1..4. 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.

✓ **AUTOMATIC EVALUATION METHODS**

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

Table 4. Automatic Evaluation Methods (Han, 2018)

✓ **Edit Distance:** Edit distance, as a Machine Translation (MT) metric, is an intuitive measure of the rate of errors in MT output, (number of errors, divided by number of reference words), with each edit viewed as fixing an error. 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.

✓ **Precision and Recall:** Precision is a widely used criterion in the MT evaluation tasks. For instance, if we use the ♯ correct to specify the number of correct words in the output sentence and the ♯ output as the total number of the output sentence, then the precision score

of this sentence can be calculated by their quotient value. (Lifeng Han, 2014).

Both precision and recall are defined in terms of the maximum match size, which is the weighted sum of the lengths of the longest matching text blocks between candidate and reference sentences. Precision is the maximum match size divided by the length of the candidate sentence; recall is the maximum match size divided by the length of the reference sentence. The maximum match size can be adjusted to weight longer matches more or less heavily by using a different exponent. On the matching value a Recall and Precision is defined, where

Recall (Candidate| Reference) = MMS (Candidate, Reference)/ |Reference|

Precision (Candidate| Reference) = MMS (Candidate, Reference)/ |Candidate|

NOTE: Evaluation of these sentences is performed by using F-measure. In the UNL System, the F-measure (or F1-score) is the measure of a grammar's accuracy. F-measure rates the T-rules grammars and Dictionary written for the given sentences on the scale of (0.1 to 1.0). It considers both the precision and the recall of the grammar to compute the score, according to the formula given in (2.1).

F-measure= 2 x ((precision x recall) / (precision + recall)) ... (2.1)

Precision: -Is the number of correct results divided by the number of all returned results.

Recall: - Is the number of correct results divided by the number of results that should have been returned.

✓**Word order:** The right word order places an important role to ensure a high quality translation output. However, the language diversity also allows different appearances or structures of the sentence. How to successfully achieve the penalty on really wrong word order instead of on the correctly different order, the candidate sentence that has different word order with the reference is well structured, attracts a lot of interests from researchers in the NLP literature.

✓ **SUMMARY**

Therefore, the requirement for construction a SMT system for Amharic language is needed the availability of large parallel corpora for training 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 express and exchange every kind of information. 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 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.

XV. THE EXPECTED OUTPUT?

The researcher had been provide for every users that enabling them to communicate with other people in their own languages. It reduces costs of developing contents in an independent way, It Allows distributing, receiving, exchange and share textural information on the Web. Determine the exact Representation & Grammatical Pattern with

in concept; the users can share information and knowledge across native languages in a high-speed Machine translation. Citizens, governments, international organizations, and enterprises are beneficiary from this study & offers opportunity for information sharing, education, and e-business. The ultimate goal is to promote sustainable development & dialogue among civilizations, economic prosperity for all nations as well as peace among them.

IIIV. Literature Review

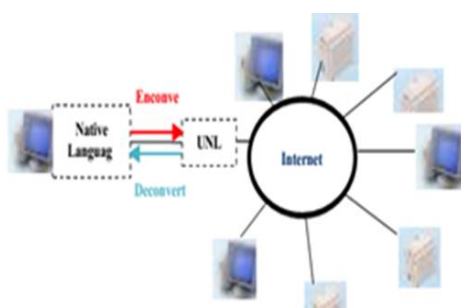

Fig.3. UNL multilingual translation system (source Hung, 2014)

✓ How the UNL express information?

The UNL represents information, i.e. meaning, sentence by sentence. Sentence information is represented as a hyper-graph having Universal Words (UWs) as nodes and relations as arcs. This hyper-graph is also represented a set of directed binary relations, each between two of the UWs present in the sentence. The UNL expresses information classifying objectivity and subjectivity. Objectivity is expressed using UWs and relations. Subjectivity is expressed using attributes by attaching them to UWs (Guillaume, 2007).

✓ Need of UNL MT system

There is an infinite need to develop a MT system for Amharic language on developing stages but all are limited only pair of languages, corpus based or dictionary based i.e. the performance results always depend on the size of sample that conducted research done. The researcher has fulfilled the gaps in the proposed study using UNL approach doing EnCo and deco for Amharic language. Final result for Amharic users can access large data retrieval from 15 languages already implemented by UNL representation Interlingua and vice versa. Such type of systems for Amharic language will be important for more than 90 million of Amharic language users to satisfy the information need in a better way with in short period of access-time.

XVI. RESULTS AND DISCUSSION

This chapter presents the results and discussions on the work carried out in this thesis. This work, as mentioned earlier has focused on the UNLization and NLization

Amharic Sentence Structures. Apart from this process that was followed, the outcome of both UNLization and NLization has also been discussed.

The testing of process is an important aspect in order to establish the usefulness of the system. The evaluation of Amharic language Sentence Structures have been performed for CORPUS UC-A1. The results of the system are very encouraging and the outputs of system are very good.

Evaluation of these sentences is performed by using F-measure. In the UNL System, the F-measure (or F1-score) is the measure of a grammar's accuracy. F-measure rates the T-rules and Dictionary written for the given sentences on the scale of (0.1 to 1.0). It considers both the precision and the recall of the grammar to compute the score, according to the formula given in (5.1).

XVII. EXPERIMENTAL RESULTS AND DISCUSSION

Universal Networking Language is a natural-language independent language which can be used for refining, describing, and semantic searching. Interactive Analyser tool is an effective online tool developed by UNDL foundation used for UNLization of any natural language. With the help of

TRules and Dictionary entries, the proposed system has been tested on corpus for Amharic Language, and their F-measure / F1-score is calculated with the help of online tool developed by UNDL foundation available at UNL-arium, as shown in Figure 5.1. Same tool is used for calculating F-Measure of UC-A1. F-measure is the measure of a grammar's accuracy. Two parameters required for the calculation of F-measure. These are:-

Precision

Precision is the number of correct results divided by the number of all returned results.

Tools F-Measure	
F-measure	**0.202**

Details	
Actual result file	actual.txt
Expected result file	CLEA1500-UCA1-UNL-AMH.txt
Sentences processed	50
Sentences returned	49
Sentences correct	10
Recall	0.200
Precision	0.204
Verification Report	download

Recall

Recall is the number of correct results divided by the number of results that should have been returned.

F-measure is calculated by the formulae given in (5.1).

F-measure =2*{(precision*Recall) / (precision +Recall)}

NOTE: For calculating F-measure we require two files. First is the expected output file and second is the actual output file. Actual output file contains the result that is generated by IAN and EUGENE and expected output file contains those sentences which were manually translated earlier from English to Amharic. The documents must be provided in plain text format (.txt) with UTF-8 encoding. 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. F-measure for CORPUS UC-A1 is given in next sub-sections.

5.2.1. Testing of UC-A1 Enco

The values of Precision, Recall, number of processed, returned and correct sentences for UC-A1 is given in Table 5.

Table 5.1: Testing details of UC-A1

S.	Parameters	Value
1.	Sentences processed	
2.	Sentences returned	
3.	Sentences correct	
4.	Precision	
5.	Recall	

TRules and Dictionary entries were created for UNLization of UC-A1 and UC-A2. F-measure of UC-A1 comes out to be 0.970 as shown in Figure 5.1.

Figure 5.1: Snapshot showing F-measure of UC-A1

5.2.3. Testing of UC-A1 Deco

The values of Precision, Recall, number of processed, returned and correct sentences for UC-A1 is given in Table 5.1.

Table 6. Testing details of UC-A1

S. No	Parameters	Value
1.	Sentences processed	
2.	Sentences returned	
3.	Sentences correct	
4.	Precision	
5.	Recall	

TRules and Dictionary entries were created for UNLization of UC-A1 and UC-A2. F-measure of UC-A1 comes out to be 0.970 as shown in Figure 5.1.

Tools
F-Measure

F-measure
0.202

Details	
Actual result file	actual.txt
Expected result file	CLEA1500-UCA1-UNL-AMH.txt
Sentences processed	50
Sentences returned	49
Sentences correct	10
Recall	0.200
Precision	0.204
Verification Report	download

Fig.5. Snapshot showing F-measure of UC-A1.

✓ **Result Summary**

The sample corpus that are used for training and test set. The final comes out F-1 score measure is to be 0.202 and 0.204 for UNLization and NLization respectively though the given UC_A1 data as shown in Figure 5.1.

6. CONCLUSION & RECOMMENDATION

XVIII. CONCLUSION

UNL is an internationally standardized Interlingua that has the Potential to bridge the gap across all the languages of the world. The UNL Program was Comes in existence as by efforts of the Institute of Advanced Studies of the United Nations University in Tokyo, Japan in 1996.The United Nations University set up an autonomous organization in January 2001, named as UNDL Foundation, to be responsible for UNL Programme, for its development and management. In multi-lingual machine translation environment, where large number of languages are to be translated between one-another, it would require a separate system for every language, i.e., n*(n-1) language systems will be required for n number of languages, So, interlingua approach is far more better approach than other approach, because here, only 2*n language systems are required. UNL is a NL independent language which can be used not only for NL translation but also for other NLP tasks like information retrieval, text simplification and semantic reasoning *etc*. As far as multilingual translation of the NL is concerned, the UNL approach is far more better than other statistical approaches, because by using the approach of UNL we only need *2n* components while other methods/approaches uses *n*(n-1)* components where *n* is the number of natural languages.

IAN tool is an effective online tool for UNLization provided by UNDL Foundation and released in 2012. The corpuses UC-A1 are provided by UNDL foundation for testing the system. This study explains how the given NL Amharic text can be converted to

UNL with the help of IAN using Trules and its corresponding grammar. This tool is very efficient resources like TRules, Analysis Grammar *etc.,* can be shared with other users. T h i s CORPUS UC-A1 where provides UNDL foundation, it was manually converted to Amharic language for input to IAN. After that with the help of TRules and Dictionary entries has been UNLized. The accuracy of the proposed system has been calculated using F-1 for UC-A1 the help of an online tool developmental page by UNDL Foundation. The F-1 of UC-A1 came out to be 0.024 for UNLization respectively.

EUGENE tool is an effective online tool for NLization provided by UNDL Foundation and released in 2012. The corpuses UC-A1 are provided by UNDL foundation for testing the system. Generation of Amharic NL sentences from interlingua like UNL, with EUGENE using T-rules and UNL-NL dictionary, has been described in this thesis. NLization of these corpuses has been done, using a word dictionary containing the information of about words that correspond to UWs included in the input UNL expressions with their grammatical attributes (features) 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. F-1 for UC-A1 came out to be 0.024 for NLization respectively.

XIX. RECOMMENDATION

UNLization and NLization, of Amharic language had been prepared on sentence level using IAN and EUGEAN tools. Some of the work that can be carried out in future includes:

- o UNLization and NLization will be on real Corpuses at paragraph level. UNDL Foundation has provided a corpus which contains different stories like ant and the grasshopper *etc.*
- o Since UNL captures semantics of the NL so semantic based searching system can be developed for Amharic.
- o Work can be extended to carry out UNLization and NLization of Afan Oromo, Tigrigna, Afar and Somali Ethiopian local languages which are recognized by Enesco and unl.
- o UNLization and NLization of this system need to be improved so as to achieve F-1 of 1.000.
- o The proposed system can be tested for other NLs spoken peoples international.

o The work can be 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). LILY, TUT, and KEYS are the applications which are under development by UNDL Foundation.

XX. FUTURE SCOPE OF THE STUDY

In future, this study work can be enriched by incorporating some components into the system. Some of the works that can be carried out in the future are following below here:

o The accuracy of the system can further improved by expanding the Amharic –UW dictionary and enriching it with more semantic information.

o The rules can be further improved linguistically to increase the accuracy of the system and to handle very large and complex sentences.

o The ambiguous words can be handled by adding a word sense disambiguation module to the proposed Amharic Enco and Deco.

REFERENCE

1. American Institutes for Research (2011 In partnership with Juárez and The Natoma Group (June 2011).

2. Blanc É. & Guillaume P. (1997) Developing MT lingware through Internet : ARIANE and the CASH interface. Proc. of Pacific Association for Computational Linguistics 1997 Conference (PACLING'97), Ohme, Japon, 2-5 September 1997, 1/1, pp. 15-22.

3. Blanchon H. (1994) Perspectives of DBMT for monolingualauthors on the basis of LIDIA-1, an implemented mockup. Proc. of 15th International Conference on Computational Linguistics, COLING-94, 5-9 Aug. 1994, 1/2, pp. 115—119.

4. Boitet C., Guillaume P. & Quézel-Ambrunaz M. (1982) ARIANE-78, an integrated environment for automated translation and human revision. Proc.

5. 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.

6. V. T. Hung, "Reuse of Free Online MT Engines to Develop a Metasystem of Multilingual Machine Translation", EsTAL 2004,

7. P. Kumar and R.K. Sharma, "Punjabi to UNL EnConversion System", Springer: Sadhna, Academy Proceedings in Engineering UNL centre, "EnConverter Specifications. Version 3.3", http://www.undl.org/, 2012. UNL centre, "DeConverter Specifications. Version 2.7"

8. http://www.unl.unu.edu/unlsys/unl/unls30.doc

9. Information Based Approach,International Conference on Knowledge Based Computer Systems, Mumbai, 2000.

10. Sinha, R.M.K., Machine Translation: The Indian Context, in International Conference on Applications of Information Technology in South Asian Languages,AKSHARA'94, New Delhi, 1994. 2002.

11. Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III), pages 137–146, COLING 2012, Mumbai, December 2012.

12. 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.

13. Amelina, Nasharuddin Nurul, and Abdullah Muhamad Taufik. "Cross-lingual Information Retrieval." Electronic Journal of Computer Science and Information Technology (e JCSIT)2 (2010).

14. UNL Based Machine Translation System for Punjabi Language Parteek Kumar(Registration No. 90603501).

15. UNL-ization of Punjabi with IAN(Vaibhav Agarwal (801131028)

16. http://www.unl.ru.

17. www.undl.org/unlpf , http://www.unl.ru,

18. www.undl.org/unlexp, http://www.eolss.net,

19. http://www.undl.org/ , http://www.unlweb.net/wiki /

20. "Scope and goals of UNL" [Online]. Available:http://www.unlweb.net/unlweb/index.php?option=com_content&view=article&id=58:unl&catid=54: unlweb[Accessed 3 June 2013].

21. "History of UNL,"[Online]. Available: http://www.unlweb.net/wiki/Introduc

tion_to_UNL[Accessed 3 June 2013].

22. "Universal Networking Language:UNL Sentence," [Online]. Available: http://www.unlweb.net/wiki/UNL_sentence. [Accessed 7 June2013].

23. J. Ollie, Machine translation [Online]. Available: http://en.wikipedia.org/wiki/Machine_translation[Accessed 26 March 2013].

24. UNL Punjabi Deconverter Anjuman Chawla (Roll No 8043102)

25. M.S. Maučec, Z. Kačič, D. Verdonik, "Statistical machine translation of subtitles for highly inflected language pair", Pattern Recognition Letters, Vol. 46, p. 96-103, September, 2014. Elsevier Science Inc. New York, NY, USA.

26. I. Taghbalout, F. Ataa Allah, M. Elmarraki, "Towards an Amazigh UNL dictionary", Proceedings of the 2nd WSEAS International Conference on Systems, Control and Informatics, p. 129-132, November 28-30, 2014, Athenes, Greece, WSEAS Press.

27. V. Stoykova, "Lexical Information for Bulgarian in Universal Networking Language",

Proceedings of the 8th WSEAS International Conference on Computer Engineering and Applications, p. 414-417, January 9-11, 2014, Tenerife, Spain. WSEAS Press.

28. P. Kumar, "UNL Based Machine Translation System for Punjabi Language", Ph.D. thesis, School of Mathematics and Computer Applications, Thapar University, February 2012, Patiala, India.

29. J.H. Greenberg, "The Languages of Africa". The Hague, 1963.

30. H. Uchida, M. Zhu. "The Universal Networking Language (UNL) Specification Version 3.0". Edition 3, Technical Report, UNU, (2005/6-UNDL Foundation, International Environment House, Tokyo, 2004).

31. S. Tripathi and J. K. Sarkhel, "Approaches to machine translation", Annals of Library and Information Studies, vol. 57, 2010, pages 388-393. Article (CrossRef Link).

32. Automatic translation for vietnamese based on unl language.

33. H. Uchida, M. Zhu and T. Della Senta, UNL: A gift for a millennium,

Institute of Advanced Studies, United Nations University, Tokyo.1999.

34. H. Uchida, UNL: Universal Networking Language – An Electronic Language for Communication, Understanding, and Collaboration. UNU/IAS/UNL Center. Tokyo, Japan, 1996.

35. I. Boguslavsky, J. Cardeñosa, C. Gallardo, L. Iraola, "The UNL Initiative: An Overview", Lecture Notes in Computer. Science, Volume 3406: 377– 87, 2005 .

36. J. Cardeñosa, A. Gelbukh and E. Tovar (eds.), "Universal Networking Language: advances in theory and applications", Mexico City, National

37. S. Alansary, M. Nagi and N. Adly, "Machine Translation Using the Universal Networking Language (UNL)", in proceedings of the 8thInternational Conference on Language Engineering, Ain Shams University, 2008.

38. H. Uchida, The Universal Networking Language Beyond Machine Translation, International Symposium on Language in Cyberspace, Seoul, Korea, September 2001.

39. T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American, May 2001.

40. O. Lassila et. al, The Semantic Web and its languages, IEEE Intelligent Systems, pages 67-73, November-December 2000.

41. UNL Center of UNDL Foundation, Universal Networking Language (UNL) Specifications Version 2005, Available at http://www.undl.org/unlsys/unl/unl2005/

42. H. Uchida, M. Zhu, UNL2005 from Language Infrastructure toward Knowledge Infrastructure, special speech at PACLING, August 2005.

43. Universal Networking Language: Advances in Theory and Applications, ISBN: 970-36-0226-6 , ISSN: 1665-9899.

44. The Universal Networking Language beyond Machne Translation Hiroshi Uchida, Meiying Zhu UNDL Foundation-September 26, 2001.

45. [1] R. T. Martins, L.H.M. Rino, O.N. Osvaldo, R. Hasegawa, and M.G.V. Nunes, "Specification of the UNL-Portuguese enconverter-deconverter prototype," Relatórios

Técnicos do ICMC-USP, 1997, pp.1-10.

46. [2] R.T. Martins, R. Hasegawa, M. Graças, and V. Nunes, "Hermeto: A NL–UNL Enconverting Environment," in Universal Network Language: Advances in Theory and Applications, vol. 12, J. Cardeñosa , A. Gelbukh , and E. Tovar, Ed(s). México, Research on Computing Science, 2005, pp. 254-260.

47. [3] R. Mohanty, A. Dutta , and P. Bhattacharyya,"Semantically relatable sets: building blocks for representing semantics," Proc. 10th MT Summit, Phuket, 2005, pp. 1-8.

48. [4] P. Kumar, and R.K. Sharma , "Punjabi Enconversion System" Sadhana,Part 2, April 2012, pp. 299–318.

49. [5] K. Dey, and P. Bhattacharyya, "Universal Networking Language based analysis and generation of Bengali case structure constructs," in Universal Network Language: Advances in Theory and Applications, vol. 12, J. Cardeñosa , A. Gelbukh , and E. Tovar, Ed(s). México, Research on Computing Science, 2005, pp. 215-229.

50. [6] E. Blanc, "About and around the French Enconverter and the French Deconverter," in Universal Network Language: Advances in Theory and Applications, vol. 12, J. Cardeñosa , A. GelX-Axis:- Part-ofspeeches Y-Axis:- F-Measurebukh , and E. Tovar, Ed(s). México, Research on Computing Science, 2005, pp. 157-166.

51. [7] I.M. Boguslavsky, L. Iomdin, and V.G. Sizov, "Interactive EnConversion by means of the ETAP-3 system," in Universal Network Language: Advances in Theory and Applications, vol. 12, J. Cardeñosa , A. Gelbukh , and E. Tovar, Ed(s). México, Research on Computing Science, 2005, pp. 230-240.

52. [8] M. Choudhury, H. Ershadul, Y.A. Nawab, Z.H.S. Mohammad, and R.M. Ahsan, "Bridging Bangla to Universal Networking Language- a human language neutral meta-language," Proc. 8th Int. Conf. on Computer and Information Technology, Dhaka, Bangladesh, 2005, 104-109.

53. [9] M.Lafourcade , "Semantic analysis through ant algorithms,

conceptual vectors and fuzzy UNL graphs," in Universal Network Language: Advances in Theory and Applications, vol. 12, J. Cardeñosa , A. Gelbukh , and E. Tovar, Ed(s). México, Research on Computing Science, 2005, pp. 125-137.

54. [10] Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III), pages 137–146, COLING 2012, Mumbai, December 2012.

55. [11] P. Kumar and R.K. Sharma, "Punjabi to UNL EnConversion System", Springer: Sadhna, Academy Proceedings in Engineering UNL centre, "EnConverter Specifications. Version 3.3", http://www.undl.org/, 2012. UNL centre, "DeConverter Specifications. Version 2.7".

56. [12] Information Retrieval System using UNL for Multilingual Question Answering Kanu Goel (801432007) THAPAR UNIVERSITY PATIALA–147004 June 2016.

57. [13] Amharic-English Bilingual Web Search Engine Mequannint Munye et.al October 28-31, 2012, Addis Ababa, Ethiopia.

© Associated Asia Research Foundation (AARF)

A Monthly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories.

Page | 49