Bidirectional Dictionary Based Machine Translation for Wolaytegna-Amharic by Java

Temesgen Mengistu Helana
Department of Computer Science
School of Informatics
Wolaita Sodo University, WSU
Wolaita Sodo, Ethiopia

Abstract:- In this paper, Wolaytegna to Amharic machine translation were conducted using dictionary based machine translation approach. Machine translation system one of a key purpose Natural Language processing and it is a process of translating from one language to another. In this study the researcher were translated two Ethiopian languages one local language (Wolaytegna) and the other one is official language of the country (Amharic) by using dictionary based approach. This research is very important for the development of the Wolaytegna language which is spoken by around 7 million people in Wolaytta zone and other part of the Ethiopia. For this research we used Java, MYSQL database and 5400 word entries in dictionary were created in the database to create accurate translation. For all words of source language we defined meaning in target language in bilingual dictionary. The proposed methodology uses dictionary for translating word by word without much because this kind of approach is very advisable for linguistically less resourced language like Wolaytegna.

Keyword:- Wolaytegna, Machine translation, Dictionary, Bilingual, Multilingual, Natural language processing.

I. INTRODUCTION

Translation systems plays a vital role in narrowing the communication barrier between human race from different corner of the world. Natural Language Processing (NLP) is a core discipline in machine translation and it is field of computer science devoted to the improvement of models and technologies empowering computers to use human languages both as input and output [3]. One of the aim of NLP is to develop computational models that can have equal performance like in the task of reading, writing, learning, speaking and understanding. Computational models are useful to explore the nature of linguistic communication as well as for enabling effective human-machine interaction.

This speedy growth of data on internet was encouragement for the MT researchers to develop more profitable MT systems to deliver a worldwide communication.

In this research work the Bilingual dictionary which is used in the Wolaytegna to Amharic translation and vice versa is the core components of a machine translation of these two languages. There are many approaches for developing the MT systems, each approach has their own advantages and disadvantages. Out of these approaches dictionary based machine translation the most recommended for linguistically less resourced language like Wolaytegna. In Ethiopia there are about 80 different languages are available from which Wolaytegna is the 7th most spoken language which is spoken by around 7 million people in the country specially by Wolaytta people in SNNPR and one of a language with few resource published electronically in internet and other different media. Oppositely Amharic is historically advantageous language in Ethiopia because different regime at different period in Ethiopia used the language as official language of the country so that it is one of linguistically well resourced languages compared to other Ethiopian languages.

So this research work will supports Amharic speakers to use Wolaytegna and Vice-versa by using dictionary based machine translation.

The biggest challenge for Statistical Machine Translation is to get the high quality corpus because of insufficient sources of the data for the language like Wolaytegna. Dictionary Based Machine Translation (DBMT) approach is used when less number of linguistic resources is available for the languages. In the dictionary based translation, a system is defined which contains set of source language word and corresponding target language words. During the run time, dictionary based translation use bilingual corpus as its database which is defined in the form of dictionary. This database is stored in the translation memory. Since the two languages, Wolaytegna and Amharic have the same grammatical sentence structure, so that when the system encounters the any sentence the system does not require any rearrange in the sentence rather it translates directly by retrieving from the translation memory.

II. CHALLENGES IN DICTIONARY BASED MACHINE TRANSLATION SYSTEM

In this research some of few challenge while translating Wolaytegna to Amharic are in Wolaytegna some words have same spelling and pronunciation but different meaning based on sentence context. So that the dictionary based machine translation translates one language to the other by using word based dictionary stored in database.

Amharic language has no proposition and it takes both prefix and postfix but Wolaytegna takes only postfix. This difference made to two language translation challenging because form one root Wolaytegna word we may numerous words with postfix which are may or may not from same Amharic word. The following example illustrates sample words:

Example:
Naaga - □□□ -   wait (he)
Naagu- □□□   wait (she)
Naagikke- □□□□□   don't wait (he)
Naagiis- □□□□ waited
Naago   □□□□   let him to wait
Naagoo- □□□□□   can I wait?   …… (he)

III. PROPOSED SYSTEM MODEL

Most of the Wolaytegna to Amharic translation looks similar to the following sample example described below. The sentence formation [Subject, Object , Verb] is similar in both languages.
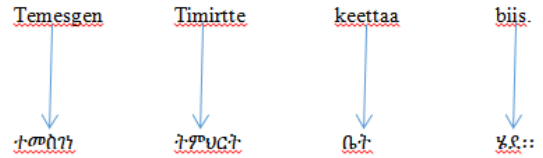Example: => Temesgen went to school.

Fig 1:- The diagram illustrates grammatically how the two language translation works

As described in the above diagram translation cannot have *Fig* happen similarly for all Wolaytegna sentence to Amharic because while we are working with tense, gender, pronounce and other aspects in special case sometimes the sentence reverses and words may be reshuffled. Within this translation system we considered all these aspects.
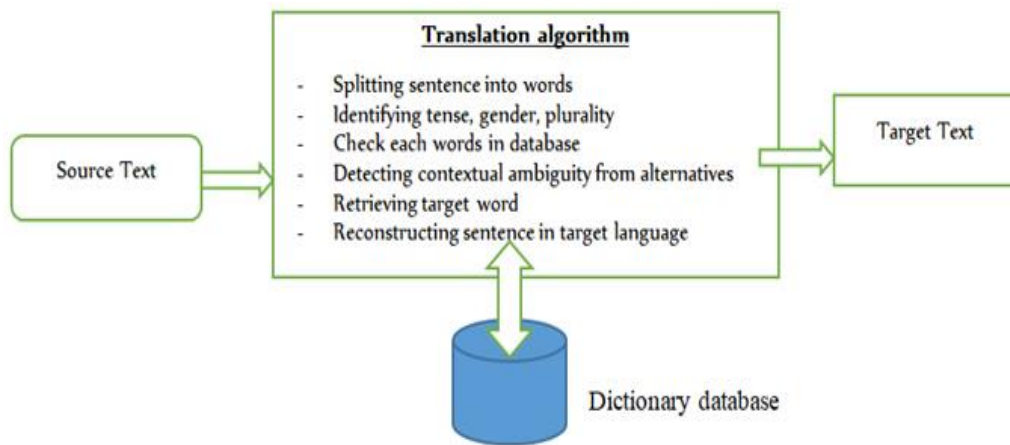

Fig 2:- Framework of dictionary based machine translation of Wolaytegna to Amharic

We will explain basic details of those steps in architectural model in the following content.

A. Splitting sentence into words
In this research translation can be word based, phrase based or sentence based; so that if the input is phrase or sentence it must be break into words because the entry in database is only word based in dictionary.

B. Identifying tense, gender, plurality
Since both language uses prefix and postfix to identify tense, gender and plurality basically it is better to detect the root word and prefix and postfix attached to the word. In this research work we faced challenge with considering the three key morphological contents.

C. Detecting contextual ambiguity from alternatives
A word may have more than one meaning in different sentence with same spelling and pronunciation.

D. Retrieving target word
In this stage of machine translation checking for the availability of each words of source language in given sentence and storing to defined array.

E. Reconstructing sentence in target language
After extracting equivalent meaning from dictionary for each words in source text reconstructing process takes place.

IV.   THE ALGORITHM

A. Pseudo code
Step1 : Start

Step2 : read source text

Step 3: split source_text into word

Step 4: for word in words
        If Check in database
                Retrieve matching meaning value
                        target_text←   value
                Else Break, not found in db
Step 5: display target_text
Step 6: End

B. Flowchart
       The following flowchart elucidates how the machine translation algorithm works. In this case we considered if the input text is a single word or a sentence it can translates the input into the target text using dictionary database.
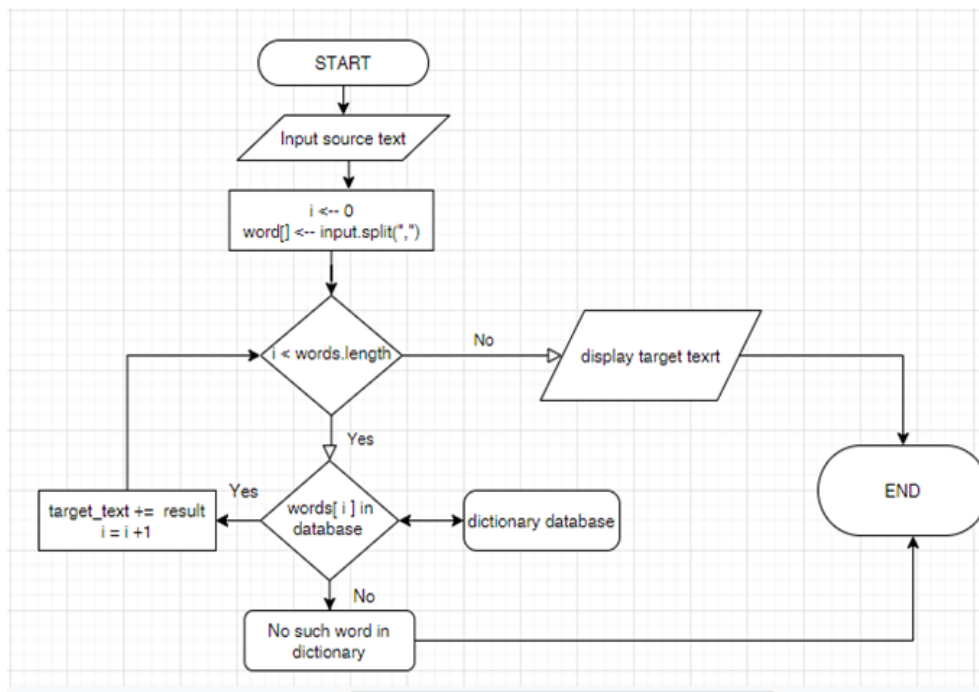

Fig 3:- Dictionary based machine translation algorithm of Wolaytegna to Amharic

V.   EXPERMENTAL RESULTS

       In the following section some of the screenshot outputs of the experimental reports and the sample codes related with bidirectional dictionary based machine translation system were illustrated.

```
BufferedReader br;
BufferedWriter bw;
public DataIO() {
    try {
        br = new BufferedReader(new FileReader("C:\\Users\\TMH\\Documents\\NetBeansProjects\\mori.txt"));
        String line = "test";
        String sql;
        while (line != null) {
            line = br.readLine();
            String st[] = line.split(",");
            sql = "insert into bilangu values (" + st[0] + ",'" + st[1]+ "','" + st[2] + "');";
            dbconn cc = new dbconn();
            cc.stm = cc.con.createStatement();
            cc.stm.executeUpdate(sql);
        }
    } catch (Exception io) {
        System.err.println("What???:" + io.getMessage());
    }
}
```

Fig 4:- Inserting created text dictionary into database

```
public void Translate2(){
    String sql;
    String stt = meaning.getText().toString();
    String []iput = stt.split(" ");
    int k=iput.length;
    String data = "";
    try {
        sql = "select walayttato,amaratto from bilangu";
        dbconn cc = new dbconn();
        cc.stm = cc.con.createStatement();
        cc.rs = cc.stm.executeQuery(sql);
        for (int i=0 ; i < k ;i++){
            while(cc.rs.next()){
                if(cc.rs.getString(2).equalsIgnoreCase(iput[i])){
                    data= data +" " + cc.rs.getString(1);
                    //System.out.println(""+data);
                    break;
                }
            }
            cc.rs.first();
        }wolayttatto.setText(data);
    } catch (SQLException ex) {        }
}
```

Fig 5:- Retrieving the MT entry from the dictionary

id	walayttato	amaratto
641	achchiis	ጠየቀ
642	achchan	አቅራብያ
643	adaba	በረከት
644	adabiis	በረከተ
645	aduqissiis	አስረዘመ
646	aduqqiis	ረዘመ
647	adussa	ረጎርም
648	afuttaa	እምባ
649	afuutiya	ጉንፋን
650	aguntta	እሹህ
651	aguwa	መተው
653	ammanettenna	ማይታመን
654	ammanettiis	ተማመነ
655	anjjuwa	ምረቃት
656	anqqariis	ፋፋ
657	anuwa	ጎንዴ

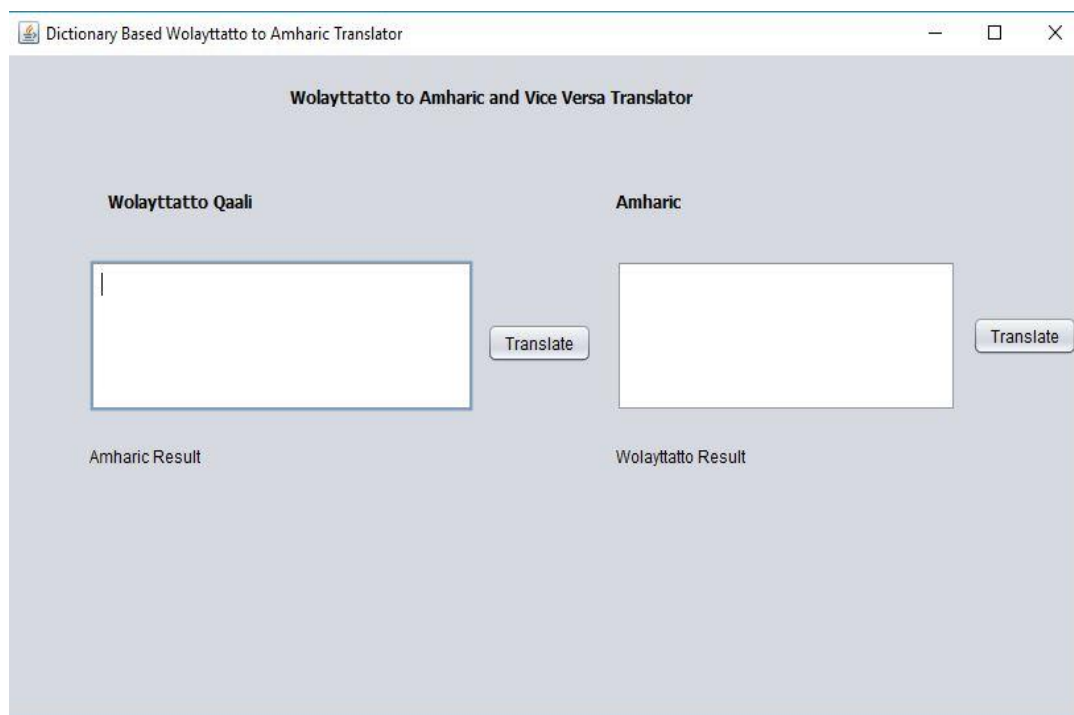Fig 6:- Store entries for dictionary in wamp database.

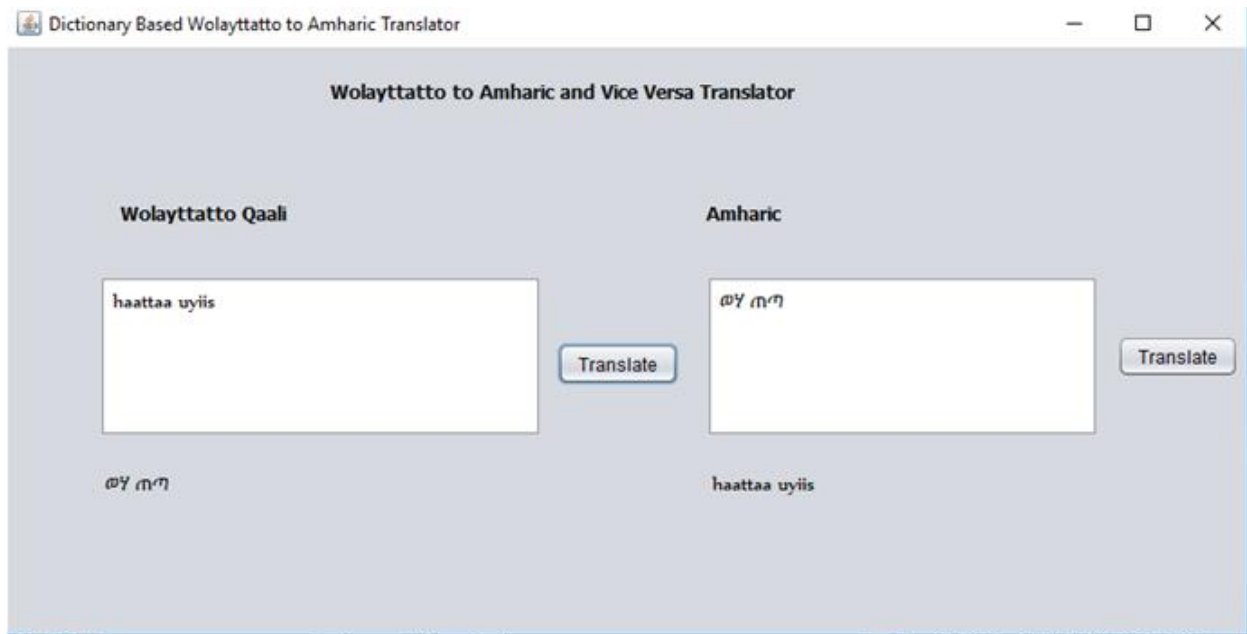Fig 7:- The main interface before feeding texts of source language.

Fig 8:- The figure illustrated how bilingual machine translator works

VI.   CONCLUSION

Machine Translation plays a vital role in breaking the language barrier and promoting the Interlingua communication in a multilingual country like Ethiopia. In this paper, dictionary based machine translation approach is used for developing the MT system for Wolaytegna and Amharic. The dictionary based approach is well suited for the languages which have the minimal linguistic resources and for the languages with the similar structure. For dictionary based approach bilingual dictionary is the crucial resource. Here bilingual dictionary with 4500 entries is developed and stored in Mysql and the user interface were designed with java. The postfix and prefix of Amharic language words are translated to Wolaytegna language word with only postfix by machine translation algorithm. Dictionary based approach can be further improved by adding more corpus and contextualize grammatical translation for both the languages. Words with multiple meaning were challenge in Bilingual dictionary based machine translation in Wolaytegna to Amharic.

REFERENCES

[1]. D V Sindhu and B M Sagar 2017 IOP Conf. Ser.: Mater. Sci. Eng. 225 012182
[2]. W.John Hutchins and Halord L. Somers, "An Introduction To Machine Translation", Academic Press Ltd.,1992, pp 1-9.
[3]. Daniel Jurafsky and James H. Martin, Speech and Language Processing, Pearson Education Inc, 2005.
[4]. Remya Rajan , , "Rule Based Machine Translation from English to Malayalam" .2009
[5]. K. Narayana Murthy "A Machine Assisted Translation System"   ,Department   of Computer and Information Sciences, University of Hyderabad, Hyderabad, INDIA.
[6]. Tariku Tsegaye, "ENGLISH -TIGRIGNA FACTORED STATISTICAL MACHINE TRANSLATION" 2014
[7]. S. Kereto, C. Wongchaisuwat, Y. Poovarawan, "Machine translation research and development´. In proceedings of the Symposium on Natural Language processing in Thailand, pages 167-195, March 1993.
[8]. K. Narayana Murthy "A Machine Assisted Translation System" ,Department of Computer and Information Sciences, University of Hyderabad, Hyderabad, INDIA.
[9]. Murthy. K,"MAT: A Machine Assisted Translation system", In Proceedings of Symposium on Translation Support System(STRANS-2002), IIT Kanpur.2002, pp. 134-139.
[10]. Balajapally, P., Bandaru, P., Ganapathiraju, M., Balakrishnan, N., & Reddy, R. (2006). Multilingual Book Reader: Transliteration, Word-to-Word Translation and Full-text Translation. In VAVA 2006.
[11]. Jingying Zhao, Hai Guo, Zhenhong Zheng, Nan Jiang, "The Implemention of Chinese-Tai Lue Electronic Dictionray Based on C#", Department of Computer Science and Engineering University of Dalian Nationalities CHINA, 2010