Script Translation System for Devnagari To English

Jayshri A.Todase
Electronics & Telecommunication Department
College of Engineering, Pune
Pune, India
jaytodase@gmail.com

Sushama Shelke
Electronics & Telecommunication Department
Dr. D. Y. Patil Institute of Engineering, Management and Research,
Pune, India
shelke.sushama@gmail.com

Abstract—**This paper presents a Machine Translation system for Devnagari to English language translation. Proposed system is able to translate more than one Devnagari sentences to English using rule base approach. The proposed system accepts Marathi language which is derived from Devnagari and translates it to English. This is done by identifying the parts of speech in Marathi sentence, tokenization and identifying English meaning of each word in bilingual dictionary. The words in a sentence are linked by the English grammar rules to have meaningful translation in English.**

Keywords— Machine Translation, Rule Based Translation, Language Translation

I. INTRODUCTION

Every country has its own language and its informative documents are written in their own native languages and structure. Hence, language can be the barrier for readers those who don't have knowledge of foreign languages to read and understand the documents.

Marathi language is derived from Devnagari and it also a native language of Maharashtra, India. Existing documents like books related to the history of Maharashtra, official documents, newspaper which is presently in Marathi language need to be translated to English for their widespread use. But, manual translation is costly, time consuming and this gives rise to the need of an automated translation system which would do the job in an effective way. The proposed Machine translator translates text documents and manuscripts which are written in Marathi language to English. The proposed system is a rule based translation system that translates Marathi to English. This system accepts Marathi text or paragraph as an input. These texts will be separated in words and identify the part of speech (POS) in the sentence according to Marathi grammar. Each Marathi word is then tokenized and tagged as verb, noun, adjective etc. Each tagged Marathi word is searched in bilingual dictionary for the correct meaning before translation into English.

To make meaningful translation from these translated words, some set of rules are pre-decided. These rules are nothing but Basic English grammar rules. All translated words are arranged according to these rules to get meaningful translation in English.

This paper is divided into following sections. Background and previous work in field of Machine Translation is discussed in section II.

The methodology and implementation of Rule based Machine Translation system is discussed in section III. The results and conclusion are finally presented in section IV & V respectively.

II. PREVIOUS WORK

First computer based Machine Translation proposal was been presented by Warren Weaver in 19's. Research began in many universities in the United States. In 1954, the Georgetown-IBM experiment was held in New York. This was the 1st public demonstration on Machine Translation system. It had translated 49 selected Russian sentences into English. The system had only 250 words in database.

After that a lot of research has been done in the field of Machine Translation and it is still going on. Various methodologies have been proposed and developed so far which has been uplifted and downfall also. So many research works in this field has done from English to foreign language like Arabic, Chinese, and Spanish etc. English to some Indian languages like Hindi, Tamil, Bangla, Teluguand Marathi and also for so many other languages Machine Translation has been implemented. Machine Translation is huge area of research. In this section some of the existent work/system of Machine Translation especially on Indian languages has been discussed. A lot of work has been done for Machine translation from English to Indian languages including Marathi as targeted language. However, very few researches are found for converting Marathi script into English. Due to the reduction in the usage of Marathi language day by day, there is a need of such system that provides the knowledge of the language. This paper, present Machine Translator which will translate Marathi (source language) to English (targeted language). The system is also useful as a learning aid.

Machine Translation system can be classified into various categories. Direct translation, Example based, Rule based, Statistical based, Knowledge based, hybrid etc. [1-5].

A. Direct Translation System

It is a simplest approach for Machine Translation. It uses the bilingual dictionary and translates source language to targeted language word to word.

In 1995, "ANUSAARAKAA" project has been developed for translating one Indian language to another Indian language. It support Telugu, Kannada, Bengali, Punjabi, Marathi and translates it to Hindi language and vice versa. The output of this project is understandable but not grammatically correct. [5][6]

B. Example Based Machine Translation

It is knowledge based translator. It uses bilingual corpus and it is basically translation by similarity. The main aim of this type of Machine Translation system is to uses readily available translations.

Ayu et al. [7] in 2011, proposed example based Machine Translation. System is developed to translate Bahasa Indonesia to English by using Moses toolkits. This system is good for short sentences but worst for long sentences.

C. Statistical Based Machine Translation

It is most widely used Machine Translation which is based on knowledge and statistical model. It translates according to PDF (Probability Distribution Function)

D. Hybrid Machine Translation

PramodSalunkhe et al. [1] in 2016, proposed a Hybrid Machine Translation system which translates from English to Marathi. This system uses combined approach of rule based and statistical based Machine Translation. The system can thus optimize the translation output. It gives better results than individual rule & statistical based machine translator. This system can be applicable only for web pages, text documents on Agriculture, Medical reports and tourism related information.

E. Rule Based Machine Translation

It is based on collection of rules (grammar rules) of targeted and source language. It also needs a bilingual dictionary which plays an important role for translation effectively.

AbhayAdapanawaret al. [6] in 2013, proposed a system for English to Marathi language. This system makes use of tokenization, POS tagger, and bilingual dictionary and production rules for translation. This system can translate only Assertive English sentence to Marathi.

Mr. KrushnaBelerao et al. [8] in 2014, proposed a system for English to Marathi translation. This system is able to process the text using parser. It tokenizes and tag each segmented word. Here they are using huge database of English and Marathi wordings. If there is an unfamiliar word which is not in the database, then the system immediately asks to add Marathi word to the vocabulary. The next process is to find production rule from database and translate. Drawback of this system is that it takes only text as input and translates one sentence at one time.

G.V Garjeet al.[9] proposed a rule based Marathi to English Machine Translation system for simple sentences. They too used the same approach which is mentioned above i.e. two rule based approach system. It accepts Marathi sentence which is represented in Shakti Standard format for easier computation. The output is represented as sequence of abbreviated features rules. The proposed system consists of limited number of feature rules which is not able to achieve perfect translation for each sentence.

Kavirajan et al [10] proposed an improved rule based machine translation algorithm using sentence simplification. It works for English to Tamil translation. The biggest advantage of this system is that, it translates long/complex sentences. It follows the same methodology like regular rule based machine translation, except, it uses the sentence simplifier to identify whether the sentence is simple or complex. If the system found the input sentence is complex, then it converts it to simpler one and then translates it.

Machine Translation is very challenging area of research [11][12]. This paper is mainly focused on Rule based Machine Translation for Marathi to English translation. There are very few research has been made for Marathi to English in previous years. Recently, Google launched their Google translate Machine translator.It supports 91 languages. The accuracy of translation is better if we go from English to Marathi translation. But in the case of Marathi to English translation, it can be understood by context but the result is not grammatically correct.

III. PROPOSED METHODOLOGY

In this section, the whole process of Marathi to English Machine Translation by using rule based approach has been discussed. Proposed methodology of our system is shows in Fig.1.

The proposed system accepts Marathi text as input from the user for further translation process.

A. Sentence Separation

The proposed system is able to translate more than one Marathi sentences in targeted English language. For correct translation it is necessary to separate out each sentence for further translation process.

B. Word Separation

Input to word separation block is in form of Marathi sentences. In this block, each word of Marathi sentence get tokenized.

{'मी'}   {'क्रिकेट'}   {'खेळतो'}

C. POS identification & Tokenization process

POS identification and tokenization is very important step in whole Machine Translation system. Syntax pattern of source language and targeted language could be different with respect to grammar [9] [11] [12]. In our case, Marathi sentence pattern is like {SOV} whereas in English it's like {SVO}.

Eg. मी क्रिकेट खेळतो

S O V

I play cricket

S V O

POS identifier identifies noun, pronoun, adjective, verb, helping verb etc. in source language sentence and tokenizes and tags each Marathi word according to POS identifier.

Word separated output: {'मी'} {'क्रिकेट'} {'खेळतो'}

POS identification: pronoun noun verb
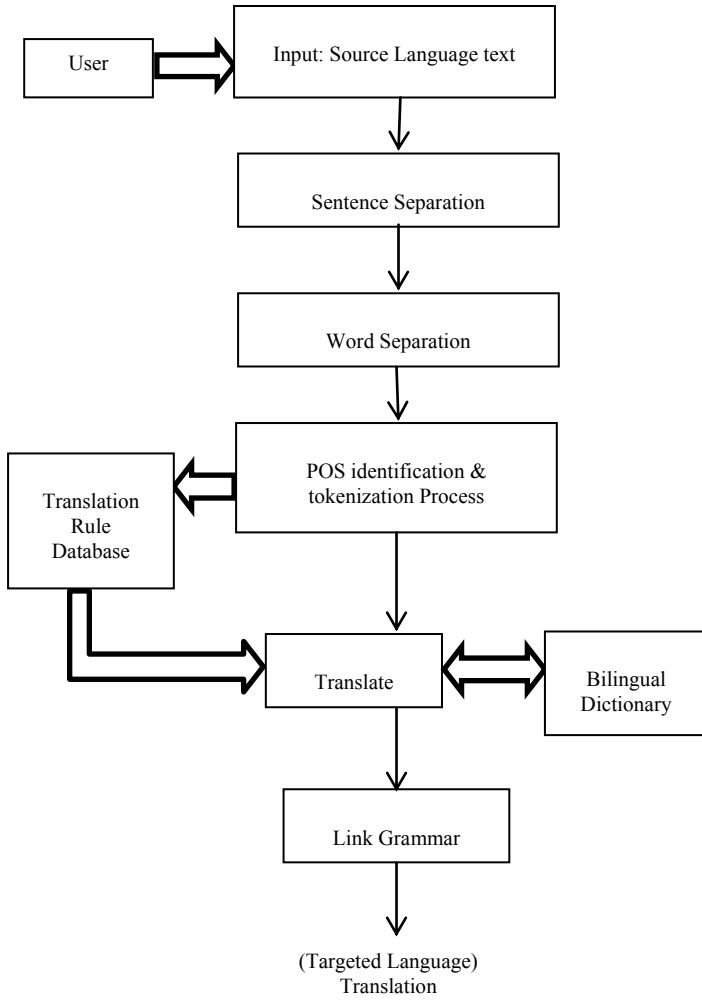
Tokenized output: {'p'} {'n'} {'v'}

Fig. 1. Methodology for Devnagri to English translation using Rule Based Machine Translation

D. Translation Rule Database

Translation Rule database is a collection of targeted language rules of translation with respect to source language structure and rules. Output of tokenization is nothing but the source language structure. This structural rule has been searched in Translation rule database and targeted rule has been

assigned for further translation process. Some of the rules are mention in Table I.

Eg. 'तेपेन महागआहे.'

Input (sentence structure): {'p'} {'n'} {'ad'} {'hv'}

Output (Target Rule): {'p'} {'hv'} {'ad'} {'n'}

E. Translate

Each tagged Marathi word's meaning in English has been search by using bilingual dictionary and gets replaced with it. As an output it produces a literal translation. Token's/tag to the literal translation remain unchanged.

Input: {'ते'} {'पेन} {'महाग'} {'आहे'}

Output: {'It'} {'pen'} {'costly'} {'is'}

The bilingual dictionary which is used in this system can be updated whenever the system gets new Marathi word which does not exist in it. It allows user to update the contained of existent dictionary.

F. Link Grammar

Link grammar block is responsible for generating meaningful translation with proper grammar. Literal translated output is getting arranged according to assigned rule of targeted language.

Input:

Targeted Rule: {'p'} {'hv'} {'ad'} {'n'}

Literal Translation:{'It'} {'pen'} {'costly'} {'is'}

Output: 'It is costly pen.'

TABLE I. TRANSLATION RULES

	Translation Rules	
	Source Structural Rules	*Targeted Structural Rule*
R1	राम चित्रपट पाहत होता.	Ram was watching movie.
	n+n1+v+hv	n+hv+v+n1
R2	हे चित्र आहे.	This is picture.
	p+n+hv	p+hv+n
R3	मी भात खाते.	I eat rice.
	p+n+v	p+v+n
R4	ते पेन महाग आहे.	It is costly pen.
	p+n+ad+hv	p+hv+ad+n

IV. RESULTS

Proposed system accepts source language (Marathi) text as input. User can write it or can be able to copy text in (left-hand side) text box of Graphical Interference Unit as shown in Fig. 2. The system processes on it by considering it as grammatically correct sentences. Just by clicking on Translate button all the process which is mentioned in section III will be done and finally the output of translator in English can be seen in (right hand side) text box of GUI as indicated in Fig.2 and Fig.3.
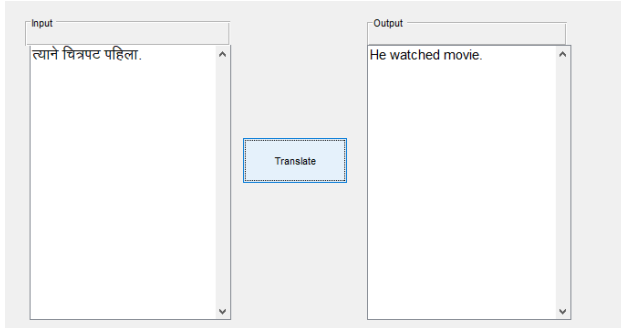

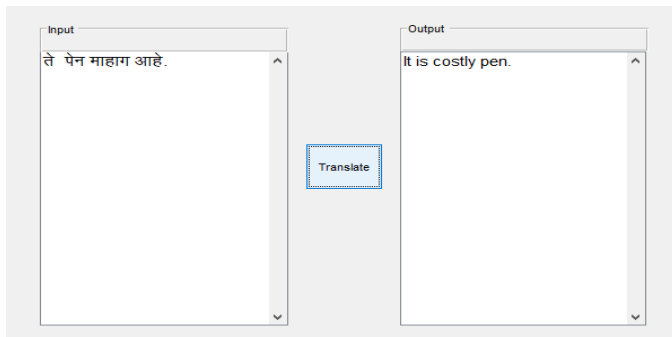Fig. 2. Translation output for simple Marathi sentence to English

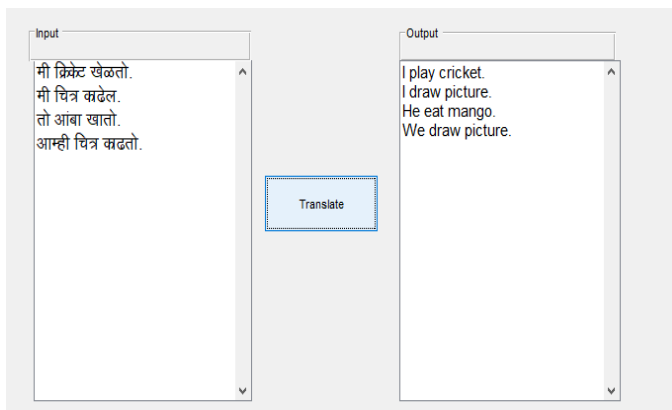Fig. 3. Translation output forMarathi sentence to English

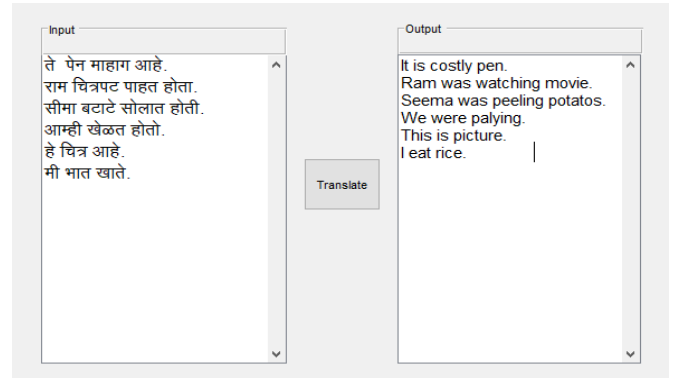Fig. 4. Translation output for multipleMarathi sentence to English

Fig. 5. Translation output for multiple Marathi sentences to English

V. CONCLUSION

This paper deals with Rule based approach of Machine Translation for Marathi to English script translation. The proposed system is able to translate more than one sentence at a time. Identifying the rule of sourcelanguage sentences and translating it according to the targeted language rule is a very challenging task. In the future, this project can be modified to get more accurate results for translation byadding more rules,taking care of article uses andpreposition disambiguation etc.

REFERENCES

[1] Salunkhe, Pramod, Aniket D. Kadam, Shashank Joshi, Shuhas Patil, Devendrasingh Thakore, Shrikant Jadhav. "Hybrid machine translation for English to Marathi: A research evaluation in Machine Translation:(Hybrid translator)." International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 924-931. IEEE, 2016.

[2] Godase, Amruta, Sharvari Govilkar. "Machine Translation development for Indian languages and its approaches." Int. J. Natl. Lang. Comput.(IJNLC) ,vol. 4.2,pp.55-74, April 2015.

[3] Lingam, Keerthi, E. Rama Lakshmi, L. Ravi Theja. "Rule-based Machine Translation from English to Telugu with emphasis on prepositions." , First International Conference on Networks& Soft Computing (ICNSC), pp. 183-187. IEEE, 2014.

[4] Mall, Shachi, Umesh Chandra Jaiswal. "Developing a system for Machine Translation from Hindi language to English language." 4th International Conference on.Computer and Communication Technology (ICCCT), pp. 79-87.IEEE, 2013.

[5] Sindhu, D. V., B. M. Sagar. "Study on Machine Translation approaches for Indian languages and their challenges."International Conference onElectrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT),pp. 262-267. IEEE, 2016.

[6] Adapanawar, Abhay, Anita Garje, Paurnima Thakare, Prajakta Gundawar, Priyanka Kulkarni. "Rule based English to Marathi translation of Assertive sentence." International Journal Of Scientific & Engineering Research,vol. 4, pp. 1754-1756. May 2013.

[7] Ayu, Media A., and Teddy Mantoro. "An Example-Based Machine Translation approach for Bahasa Indonesia to English: An experiment using MOSES."IEEE Symposium on Industrial Electronics and Applications (ISIEA),pp. 570-573. IEEE, 2011.

[8] Mr. Krushna Belerao, V.S. Wadne, "Machine Translation Using Open NLP and Rules Based System English to Marathi Translator", International Journal on Recent and Innovation Trends in Computing and Communication, vol. 2,pp. 1730-1733, June 2014.

[9] G V Garje, A. Gupta, A. Desai, N. Mehta, A. Ravetkar,"Marathi to English Machine Translation for Simple Sentences." International Journal of Science and Research (IJSR) ISSN (Online),vol. 3, pp. 3166-3168, November 2014.

[10] Kavirajan, B. Annand Kumar M, Soman K.P, Rajendran S, Vaitheh S, "Improving the rule based Machine Translation system using sentence simplification (english to tamil)."IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 957-963, 2017.

[11] Garje, G. V., G. K. Kharate, Minal R. Apsangi, Harshad M. Kulkarni, Manasi S. Sant. "Challenges in Rule Based Machine Translation From English To Marathi." In proceedings of 3rdInternational Conference on Recent Trends in Engineering and Technology (ICRET'14), pp. 243-248, 2015.

[12] S.B Kulkarni, P.D Deshmukh, M.M Kazi, K.V Kale, "Linguistic Divergence Patterns in English to Marathi Translation." International Journal of Computer Applications, vol. 87.4, pp. 21-26, February 2014.