

# Marathi to English Sentence Translator for Simple Assertive and Interrogative Sentences

G.V. Garje, PhD  
HOD, Department of  
Computer Engineering  
Pune Vidyarthi Griha's  
College of Engg.& Tech.  
Pune, India

Akshay Bansode  
Department of  
Computer Engineering  
Pune Vidyarthi Griha's  
College of Engg.& Tech.  
Pune, India

Suyog Gandhi  
Department of  
Computer Engineering  
Pune Vidyarthi Griha's  
College of Engg. & Tech.  
Pune, India

Adita Kulkarni  
Department of  
Computer Engineering  
Pune Vidyarthi Griha's  
College of Engg. & Tech.  
Pune, India

## ABSTRACT

Due to globalization English has become the official language of the world. About 71 million people speak Marathi as their native tongue. The major goal of proposed system is to develop software system which would translate Marathi Simple Assertive and Interrogative Sentences to corresponding English sentences. The quality of translation of existing system is very coarse. Since, there exist no fully functional Marathi to English Translation Systems; using rule-based approach we intend to develop one such system to produce translation with better quality.

## Keywords

Grammar, Marathi, Natural Language Processing, Parser, Rule-based Machine Translation

## 1. INTRODUCTION

Communication has been a vital part of the life of humans from the beginning of time. With about 71 million Marathi speaking people and varied works in Marathi literature and novels calls for translation [4]. Languages are the tools for effective communication. Marathi is one of the top 22 official languages of India [7]. Research and documents these days are usually in the English language that are universally recognized and accepted. Existing documents that are currently in the Marathi language need to be translated to English for their widespread use. 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. Also, there is not much work done so far for translation of Indian languages. English is a Subject-Verb-Object language while Marathi language is Subject-Object-Verb and is relatively of free word order. Hence its translation is a challenging task. The major goal of proposed system is to develop a system which would translate Marathi Simple Assertive and Interrogative Sentences to corresponding English sentences. The system takes Marathi sentence as an input and its lexical analysis is performed for tokenization. Every token produced by lexical analysis is searched in the Marathi lexicon. If the token is found in the lexicon, its morphological information is retrieved. If all such tokens corresponding to Marathi tokens are found, then English sentence is produced using English grammar rules.

## 2. RELATED WORK

### 2.1 Google Translate

It is a free translation service available to translate text, speech, etc. from one natural language to another. It offers a web interface, mobile interface for android and iOS. It uses Statistical Machine Translation i.e. machine translation in which translation is generated using statistical translation

models, parameters of which are derived from the analysis of bilingual text corpora. If corresponding word is not found in the text corpora, accurate translation is not obtained. Moreover the Google translate does not check the syntax of the given sentence.

### 2.2 Existing Morphological System:

The morphological system being used is developed by consortium of Institutions in India which is maintained by IIT Bombay. It is funded by TDIL (Technology Development for Indian Language), Department of IT, Government of India [8]. The system accepts Marathi sentence/paragraph as input in UTF-8 or WX format and gives a morphological analysis of sentence/paragraph. This helps in identifying the context of sentence/paragraph. It gives morphological information such as category, gender, number, person, suffix and root of each word in sentence.

## 3. PROPOSED SYSTEM

The proposed system is a translation system for translating simple assertive and interrogative Marathi sentences into corresponding English sentences using rule based approach.

### 3.1 Rule Based Translation approach

It is a machine translation approach based on linguistic information of source and target languages which are retrieved from dictionary and grammars covering the main morphological, semantic and syntactic regularities of both languages. The Rule Based Machine Translation is based on linking the structure of given input sentence with the structure of demanded output sentence, necessarily preserving their unique meaning.

For such translation one needs:

- 1) A bilingual dictionary for mapping the words from source language to target language.
- 2) Grammar rules representing regular source and target language sentence structure.

## 4. SYSTEM ARCHITECTURE

Architecture consists of following components:

- 4.1 Parsing
- 4.2 Bilingual lexicon/ Dictionary
- 4.3 Target language generator

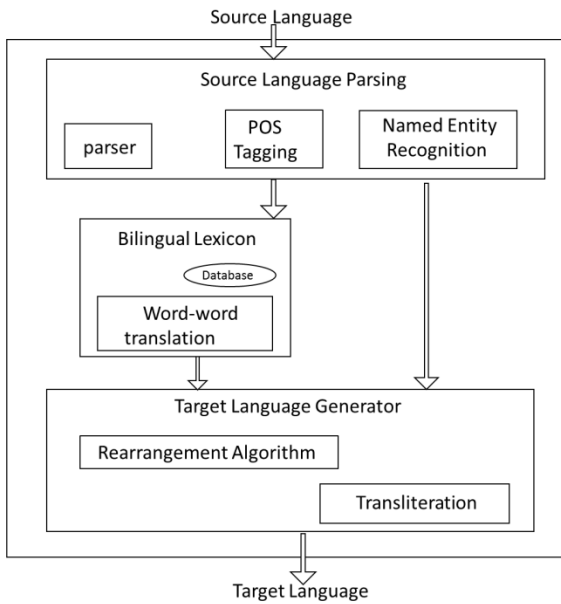


Figure 1 System Architecture

#### 4.1 Parsing

Source language parsing is implemented by using following three components:

##### 4.1.1 Parser

##### 4.1.2 Named Entity Recognizer

##### 4.1.3 Parts of Speech (POS) Tagger

The parser processes the given input sentence and separates each word. Named Entity Recognizer associates each word with its root word. This makes it easier to match the translation and target language word. Parts of Speech tagger tags each word in the sentence with its role, e.g. a word maybe a noun, verb, adjective, etc.

The output of parser is the root word, its tense, gender, and multiplicity in a given input sentence. The output of parser is passed to the Target Language Generator.

#### 4.2 Bilingual lexicon/Dictionary

A bilingual lexicon is used for storing words of source language along with the words of target language. The source language words are searched in the bilingual lexicon based on the root words provided by the parser and then the inflection forms of these words in the target language are generated based on POS tag of words. The lexicon will return a source language word and corresponding target language word.

#### 4.3 Target language generator

Target language generator is implemented using two components: Transliteration and Rearrangement Algorithm. In transliteration phase these Target Language words are transliterated in the Target Language script. In rearrangement algorithm the tokens of source language are rearranged according to the structure of target language using target language grammar rules. Here rule based approach will be followed [2]. The output is displayed in target language script.

Example:

Input sentence: तो पहिला आला.

This sentence is passed to the Marathi shallow parser. The analysis of the input Marathi sentence obtained from parser is represented in the Shakti Standard Format (SSF) [6], which

makes it easier for computation and also gives a fixed representation of the analysis.

Output of the parser is shown below:

Table 1. Output of the Parser

<Sentence id="1">			
1	((	NP	<fsaf='तो,conj,,,,,' poslcat="NM" head="तो">
1.1	तो	PRP	<fsaf='तो,conj,,,,,' poslcat="NM" name="तो">
	)		
2	((	NP	<fsaf='पहिला,adj,m,sg,,d,,' head="पहिला">
2.1	प हि ला	QO	<fsaf='पहिला,adj,m,sg,,d,,' name="पहिला">
	)		
3	((	VGF	<fsaf='ये,v,m,sg,,d,ला,ला' kridanta_type=la aspect=p tense= mood= kridanta_cm=ला head="आला">
3.1	आ ला	VM	<fsaf='ये,v,m,sg,,d,ला,ला' kridanta_type=la aspect=p tense= mood= kridanta_cm=ला name="आला">
3.2	.	SYM	<fsaf='.,pun,,,,,' poslcat="NM">
	)		
</Sentence>			

By using Bilingual lexicon, corresponding English root word is mapped to the Marathi root words.

तो→he

पहिला→ first

आला → come

Now, these words are arranged by using different rearrangement rules. For this sentence following rule is applied.

PRP + QO + VM → PRP + VM + QO

He + first + come → He + come + first

The abbreviations can be understood with the help of the following description:

**Table 2: Tags for Parts of Speech of Parser [3]**

Sr. No.	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential <i>there</i>
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP\$	Possessive pronoun
20.	QO	Ordinals
21.	RB	Adverb
22.	RBR	Adverb, comparative
23.	RBS	Adverb, superlative
24.	RP	Particle
25.	SYM	Symbol

26.	TO	To
27.	UH	Interjection
28.	VB	Verb, base form
29.	VBD	Verb, past tense
30.	VBG	Verb, gerund or present participle
31.	VBN	Verb, past participle
32.	VBP	Verb, non-3rd person singular present
33.	VBZ	Verb, 3rd person singular present
34.	VM	Verb Main
35.	WDT	Wh-determiner
36.	WP	Wh-pronoun
37.	WP\$	Possessive wh-pronoun
38.	WRB	Wh-adverb

After that different grammar rules are applied for checking suffix, prefix, tense, etc. to generate target language sentence. The generated sentence is –

“He came first.”

## 5. CONCLUSION

It has been observed that rule based machine translation involves generating a lot of rules and handling of exceptions as well and can produce better quality translation. The system will make use of Shallow parser, Bilingual Lexicon and Rearrangement algorithms to generate better quality translations.

This system can be extended in many ways. The system is intended for simple assertive and interrogative sentences. It can be extended for other types of simple sentences such as exclamatory, imperative, etc as well as complex and compound sentences. The system can be also used as a module for a universal system. Apart from these extensions disambiguation of nouns and verbs will be a major improvement to the system.

## 6. ACKNOWLEDGMENT

We thank Mr. Manish Patil (Persistent Systems Ltd, Pune) for his support, help and guidance without which this system would not be what it is.

## 7. REFERENCES

- [1] G V Garje, Adesh Gupta, Aishwarya Desai, Nikhil Mehta, Apurva Ravetkar, “ Marathi to English Machine Translation for Simple Sentences”, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358
- [2] Abhay Adapanawar, Anita Garje, Purnima Thakare, Prajakta Gundawar, Priyanka Kulkarni, “Rule Based English to Marathi Translation of Assertive Sentence”, International Journal of Scientific & Engineering

Research, Volume 4, Issue 5, May-2013 1754 ISSN 2229-5518

- [3] “AnnCorra : Annotating Corpora, Guidelines For POS And Chunk Annotation For Indian Languages”, Akshar Bharati, DiptiMisra Sharma, Lakshmi Bai, Rajeev Sangal, Language Technologies Research Centre, IIT, Hyderabad
- [4] <http://www.censusindia.gov.in/2011-common/censusdataonline.html>
- [5] Prof. Goraksh V. Garje, Manisha Marathe, Urmila Adsule, “Translation of Simple English Interrogative Sentences to Marathi Sentences”, in proceedings of ICWET-10, Mumbai, Maharashtra.
- [6] Akshar Bharati, Rajeev Sangal, Dipti M Sharma, “SSF: Shakti Standard Format Guide”
- [7] G.V. Garje, 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 International Conference on Recent Trends in Engineering and Technology (ICRET’14), published in Elsevier digital laboratory.
- [8] Website of the shallow parser : <http://ltrc.iiit.ac.in/analyzer/marathi/>
- [9] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu “Bleu: a Method for Automatic Evaluation of Machine Translation”, IBM Research Report, September 17, 2001
- [10] Ananthkrishnan Ramanathan, ”Statistical Machine Translation”, Department of Computer Science and Engineering, Indian Institute of Technology, Bombay