Rule Based English To Marathi Translation Of Assertive Sentence

ABHAY ADAPANAWAR[1], ANITA GARJE[2], PAURNIMA THAKARE[3], PRAJAKTA GUNDAWAR[4], PRIYANKA KULKARNI[5]

Abstract— In proposed system we are dealing with the rule based English to Marathi translation of assertive sentences. This is basically a machine translation. In this system we are going through various processes such as tokenization, part of speech tagging etc. Database of production rules is maintained which plays important role in translation. English to Marathi bilingual dictionary has been formed for the purpose of language translation.

Index Terms— Artificial Intelligence, Language Translation, Lexical Analysis, Machine Translation, Natural Language Processing,

Rule based translation, POS tagging
.

1 Introduction

Marathi is one of the richest languages among all the languages exist in the world and one of the largely spoken languages in the world. More than 72 million people speak in Marathi as their native language. It is ranked 19th, based on the number of speakers .Marathi is the mother language of India and also a large number of people in southern area of India (Maharashtra) speak and write in Marathi.

Marathi is a member of the Indo-Aryan languages. It is derived from Sanskrit. It is written left-to-right, top-to-bottom of page (same as English). Its vocabulary is akin to Sanskrit. Though the vocabularies are quite difficult at first, but to some extent there are similarities with English as exemplified by the following words in Table 1.

TABLE 1
COMPARISON OF THE SIMILARITIES BETWEEN DIFFERENT LANGUAGES MAJOR HEADINGS

Language	Word				
English	month	mother	new	night	nose
Sanskrit	mãs	matar	nava	nakt	nãs
Marathi	mahina	mata	navin	ratra	naak

2 Need Of Translation

People of different linguistic background could not able to interact with each other. This concept of translation will help people to communicate comfortably. Also it will help to fill communication gap between two linguistically different backgrounds. It will help to the people in the villages, who have taken education of English.

3 Problem Statement

There are four types of sentences 1.Assertive sentence, 2.Interrogative sentence, 3.Exlamatory sentence.4.Imperative sentence. Any sentence will belong to one of this type. We have taken assertive sentences, to restrict scope of the project.

Purpose of the Natural Language Processing is to convert English sentence to Marathi (Assertive). Firstly the user enters the English sentence the perquisite is user must enter grammatically correct then it undergoes different process such as tokenization, dictionary lookup, POS tagging, rule matching etc. In the end we get the output in the human readable format.

In this system meaning is taken into consideration while translating sentences. It's not just word to word mapping.

4 Solution Prerequisite

To provide solution to above problem, the database of set of rules should be maintained for mapping English sentence to Marathi. These rules are called as production rules. English to Marathi dictionary database is required for fetching Marathi words for specified English words .Also we should have the deep knowledge of grammar of source language and target language.

4.1 Grammar of Source Language and Target Language:

Here source language is English and Target language is Marathi. Every language has parts of speech i.e. Verb, noun preposition, etc.

Structure of language changes depending on the arrangement of parts of speech. For e.g.-"I am going to school". This is one English sentence. Here "I" is a subject; "am going" is verb phrase. Verb phrase means "auxiliary verb+ subsequent verb" and "to school "is an object. So structure of sentence is "Subject+Verb+Object".Translation of this sentence in Marathi is "Mi shalet jaat ahe".'I' is translated as 'Mi'in Marathi, 'am' becomes 'ahe','going' becomes 'jaat' and 'to school' becomes 'shalet'in Marathi. Here "Mi" is Subject, "shalet" is an object and "jaat ahe" is a verb. So structure of sentence in Marathi is

"Subject+Object+Verb".

For proper language translation, it is necessary to understand the grammar of both languages.

4.2 English to Marathi Bilingual Dictionary

It is necessary to have dictionary. Because with the help of dictionary we get the corresponding Marathi word which plays important role in transaltion.Dictionary database is endless.Therfore we can extend the database according to need.

In dictionary we store English word, corresponding Marathi word. And transliteration of that word.

4.3 Adding Production Rules To Database

We have shown the production rules in table2 for both English and the Marathi sentences side by side. 'r' represents the rule in English and r' represents corresponding rule in Marathi. There are individual sentence patterns for English and Marathi sentences. These rules are in pair wise. Because a sentence pattern in English must have a corresponding sentence pattern in Marathi which is used for language translation. These rules are predefined and must be precisely given in the language translation system. For the language translation purpose, an English sentence pattern will change to a Marathi sentence pattern according to a particular rule. This rule is given in the production rule table. In this table there are very few rules represented to give the idea that how the production rule works.

TABLE 2:PRODUCTION RULES

English Pattern(r)		Marathi Pattern(r')	
r1	S→ n + v + n'	r1'	S → n + n' + v
	Seema + was peeling + potatoes		Seema batate Solat hoti.
r2	S→ n+v+n'+n''	r2'	S→ n+n''+n'+v
	Knowledge + lights + the way + to heaven		Dnyan Swargakade janyacha rasta ahe.
r3	S→ p+art+adj+n	r3'	S→ p+n+adj+art
	It + is a + costly + pen		Te pen mahag ahe
r4	S→ P+v	r4'	S→ P+v
	We+were playing		Amhi khelat hoto
r5	S→ n+v	r5'	S→ n+v
	The moon+shines		Chndra chamakto
r6	S→ P+d+v	r6'	S→ P+d+v
	We+all+breathe		Apan sagle shwas gheto
r7	S→ d+art+n	r7'	S→ d+n+v
	This+is a+picture		He chitra aahe
r8a	S→ n+v+(p+n')	r8a'	S→ n+(p+n')+v
	Karim+cut+(his+finger)		Karim ne tyache bot kapale
r8b	S→ n+v+(p+art+adj+n')	r8b'	S→ n+(p+art+adj+n')+v
	Grandfather+told+(us+a+funny+story)		ajobane gamtidar gosht sangitli
r8c	S→ n+v+(n'+adv)	r8c'	S→ n+(n'+adv)+v
	Habib+goes+(to college+regularly)		Habib collegela roz jato
r8d	S→ n+v+(d+n'+c+n'')	r8d'	S→ n+(d+n'c+n'')+v
	I+washed+(my+hands+and+face)		mi maze hat ani chehra dhutla
r9a	S→ p+v+n	r9a'	S→ P+n+v
	I+eat+rice		Mi bhat khato/khate
r9b	S→ p+v+(n+p')	r9b'	S→ p+(p'+n)+v
	He+told+the news+to everyone		tyane saglyana batmi sangitli
r9c	S→ p+v+(n+d+n')	r9c'	S→ p+(d+n'+n)+v
	We+visited+(Moynamoti+last+year)		amhi magil varshi maynamotila gelo hoto
r9d	S→ p+v+adv	r9d'	S→ p+v+adv
	He+was sleeping+then		to zopla nantar
r10	S→ p+n+v+pre+d+p'+n'	r10'	S→ p+n+p'+n'+pre+v
	each+Muslim+is brother+of+every+other+Muslim		ek muslim dusrya muslim cha bhau asato
r11	S→ v+p+d+adj+n		S→ p+d+adj+n+v
	Give+him+some+crisp+potao-chips		tyala thode kurkurit batate chips de

5 Translation Process

5.1 Tokenization

Input is the assertive sentence, which should be grammatically correct. Then it converts the sentence into tokens i.e. words. We have used "open-nlp" in programming .Open-nlp is the open source tool, provided for performing different processes, which are required in translation. For tokenization have used "tokenize"method from "tokenizer" class.

Input: - Sentence

Output: -Word level Token

5.2 POS tagging:

Part of speech tagging is the process of assigning a part of speech to each word in the sentence. Identification of the parts of speech such as nouns, verbs, adjectives, adverbs for each word of the sentence helps in analyzing the role of each constituent in a sentence.

For this process, we need "tag" method from "tagger"class of open-nlp.

Input:-tokens

Output:-tag to each token

5.3 Search tokens into Dictionary

English to Marathi bilingual dictionary is maintained. Tokens which we got from 1st step are searched into the dictionary and given to translator.

Input:-token

Output:-corresponding Marathi word for each token

5.4 Search Rule into Database

As we have stated above, we are going to store the production rules in database. So the given sentence will be translated according to rule. For this, after pos tagging and getting appropriate Marathi word from dictionary, those Marathi words are arranged according to rule and corresponding Marathi translation is shown to user.

Input:-Source language sentence on which Pos tagging and tokenization is performed.

Output:-Rule matching and corresponding Marathi sentence

6. Translation Process With Example

Let us take following example and see the translation process:

E.g.-He gives me a pen.

1. First requirement is these words must be present in dic-

tionary.
If they are not present then enter them in dictionary.
2. To add the production rule for this sentence. We must tokenize it.then we get 5 words as 1.He, 2.gives, 3.me, 4.a, 5. Pen
3. Each word will be assigned one tag and index as follows:
He: [0] PRB (means Pronoun)
Gives: [0] VBZ (means Verb)
Me: [1] PRB (means Pronoun)
A: [0] DT (mean determiner/Article)
PEN: [0] NN (Means Noun)
Index indicates how many items are present of particular type. Here in this example two pronouns are present so for "He" index is [0] and for "Me"index is [1].
4. Then we add corresponding structure of target language.
If we translate the given sentence manually to Marathi then sentence in Marathi is: "To mala pen deto"
So we need to add corresponding Marathi rule as–'He me a pen gives'
Again we need to tokenize the target language sentence.
So we get tokens as follows:
He: [0] PRB (means Pronoun)
Me: [1] PRB (means Pronoun)
A: [0] DT (mean determiner/Article)
PEN: [0] NN (Means Noun)
Gives: [0] VBZ (means Verb)
5. So if we add rule to database it is stored as follows:
PRB-VBZ-PRB-DT-NN|PRB-PRB-DT-NN-VRB
Left part shows structure of English sentence and right part shows corresponding rule in Marathi.
6. Thus we have words in dictionary and production rule to database. Now when user will give input to translator as"He gives me a pen". This will match with above rule and it will show output as"To mala pen deto"

7 Conclusion

In this paper, we have shown a totally new approach for language translation. In India, there is very little work on English to Marathi language translation done. Among them this research is totally a different one. The language translation architecture that is represented here is not developed before. The task that we have done in this paper can be extended more. A lot research is possible in this field. We have tried to keep variation among the English sentences that we have translated into Marathi sentences. But we have not completed all the variety of sentences. Since it is Natural Language Processing (NLP) the number of variation is almost unlimited. It is because the language is changeable according the time. Many words are expired and not used nowadays. On the other hand, many new words are added in the language. This is a Human Language Technology (HLT) that is people are making new words of languages. So there is unlimited opportunity to upgrade the current research.

8 References

[1] Bharati,Akshar,Vineet Chaitanya, Rajeev Sangal, Natural Language Processing: A Paninian Perspective, Prentice-Hall of India,1995
[2] Sangal, Rajeev,Akshar Bharati, Dipti Misra Sharma, Lakshmi Bai, Guidelines For POS And Chunk Annotation For Indian Languages, December
[3] Sangal, Rajeev,Dipti Misra Sharma, Lakshmi Bai, Karunesh Arora, Developing Indian languages corpora: Standards and practice, November
[4] Sangal, Rajeev, Shakti Standard Format: **SSF,** January 2007
[5] Bonnie J. Dorr, Pamela W. Jordan, John W. Benoit, 'A Survey of Current Paradigms in Machine Translation', LAMP TR-027, Dec. 1998
[6] Bonnie J. Dorr, 'Interlingual Machine Translation: A Parameterized Approach',IEEE transaction on Artificial Intelligence, Volume 63, Issue1-2 (October 1993)
[7] Dr. Shridhar Shanvare, 'Abhinav Marathi Vyakaran, Marathi Lekhan', Vidya Vikas Mandal, Nagpur.