ABSTRACT
This paper presents a proposed system for machine translation of English interrogative sentences to their Marathi counterpart. The system takes simple interrogative English sentence as an input and performs its lexical analysis. Every token produced by lexical analysis is searched in the English lexicon. If the token is found in the lexicon, its morphological information is retrieved. The local word groups are formed depending on the morphological information of a token, if required. This local word groups are checked against grammar rules of English language for syntactic validity of input sentence. This is done by using bottom up parsing technique. If the input sentence is syntactically correct, the corresponding Marathi token is searched in Marathi Lexicon. If all such tokens corresponding to English tokens are found, then Marathi sentence is produced using Marathi grammar rules. The paper emphasizes on developing production rules for simple English sentences.

Categories and Subject Descriptors

Keywords

1. INTRODUCTION
The Language is meant for communicating world. By far the largest part of human linguistic communication occurs as speech. Written language is a fairly recent invention and still plays a less central role than speech in most activities. But processing written language is easier, in some ways, than processing speech. Machine translation has a much greater significance in breaking the language barrier within the sociological structure. As English continues to be the link language, the machine translation system uses English as the source language and Marathi as the target language which is one of the Indian regional languages. Since it is difficult to use human translators for volume, machine-aided translation would be the solution. Machine translation has multiple approaches:
  1. Direct Architecture
  2. Transfer Architecture
  3. Interlingua Architecture

Direct Architecture is the most primitive form of translation replacing source language word with the target language word. Interlingua approach needs an in-depth semantic analysis and the target language is generated in this approach. Transfer approach lies in between the two extremes; it works on the syntactic level and involves semantic in some places. Our system follows the transfer approach with rule-based translation and emphasizes on interrogative sentences.

2. RELATED WORK
The field of Natural Language Processing has emerged in its own right and a large number of research groups around the world are working on it. Japanese National Project (Mu) successfully completed industrial prototype of language translation. Eurotra Project in Europe has hundreds of researchers working on it. AksharBharati Group at IIT Kanpur has developed a tool ANUSARAKA to translate simple Telugu sentences into corresponding Hindi sentences. IIT Mumbai, INCST (National Center for Software Technology), IIIT Hyderabad are working on language translation. Different approaches are described by Dorr in 1998[2]. Rule based machine translation from English to Urdu using transfer approach is done by Naila Ata[5]. They handled case phrases and verb postpositions through concept of Pannian Grammar. The system, AnglaHindi, translates English to Hindi which is version of AnglaBharti proposed by R.M.K. Sinha[7]. Dilshan D. Silva has a developed Sinhala to English translator with various inbuilt tools like grammar tool, dictionary, Unicode fonts, debugging tool, add word tool[3].
3. SYSTEM OVERVIEW

English is a SVO language while Marathi language is SOV and is relatively of free word order. The block schematic diagram of our system is depicted in Fig. 1. The input to the system is simple interrogative English sentence like ‘What is your name?’. Lexical Analyzer splits the input stream into tokens separated by delimiters which is more manageable by the parser.

Lexicon is the set of known words with complete morphological information such as its root, category, case, gender and its number etc. Morphological Analyzer accepts tokens and checks whether tokens are present in the lexicon or not. If the token is present, system will retrieve complete morphological information about it.

Example:

Input: What is your name?

Output of morphological analyzer:

What [ Root : what ]
   Category : pronoun ]

Is [ Root : is ]
   Category : auxiliary ]

Your [ Root : your ]
   Category : pronoun
   Gender: male
   Number : singular

In Local Word Grouper (LWG), grouping of all the tokens after assigning morphological information to them will be carried out. For the above example Local word Grouping is as shown below:-

The Pronoun ‘your’ and Noun ‘name’ becomes Noun Phrase (NP) according to the rules given below. The NP and VP-PPS becomes NP-VP-PPS. Finally the Pronoun ‘what’, Auxiliary verb ‘is’ and NP-VP-PPS forms a sentence S. This way LWG will take place using bottom-up parsing technique.

The rules used in LWG are as follows:

1) PRONOUN := what | your | NULL
2) AUX := is | NULL
3) NP := name | NULL
4) NP := PRONOUN NP
5) VP := VERB
6) VERB := NULL
7) PP := PREP NP
8) PREP := NULL
9) VP-PPS := VP PP
10) NP-VP-PPS := NP VP-PPS
11) S := PRONOUN AUX NP-VP-PPS

The structure of the valid sentence represented as parse tree is as shown in Fig. 2.

Figure 1: Block schematic of the system

Figure 2. Parse tree generated by syntax analysis

Syntactic analysis exploits the result of morphological analysis to build a structural description of the sentence. A parser will be designed for given grammar and the lexicon to check whether given sentence is constructed according to the rules of the
grammar. The rules mentioned above shows a simple context free phrase structure grammar for English.

Local Word Separator separates tokens from the statement generated in LWG to search corresponding token in target language dictionary.

Mapping Block maps the tokens of source language (English) to corresponding tokens of target language (Marathi).

Marathi is relatively free word order language still there are units which occur in fixed order. Noun phrase (NP) and verb phrase (VP) can be formed using only local information and more importantly they provide sufficient information for further processing of the sentence. So the LWG provides all the necessary information with minimum computational efforts.

Production rules for Marathi words are as given below.

1. Noun:
   a) "karant " to Noun:
      In the sentence , if the preposition is present in noun phrase (NP), then “ karant prayaya” is applied to NP as per Table
      e.g.: Did you come for me?
      For me := preposition-noun
      “maza” becomes “mazya”.

   b) “Shasti prayaya “ to noun :
      If noun in the sentence has (‘s) then depending on the gender of noun following it , “pratyaya(cha, chi, che etc.)” are applied to it.
      e.g.: Is that Goraksha’s book?
      “Garaksha’s” become “Gorakshache”.

   c) “karant” to Proper Noun:
      No “karant” is applied to proper noun.
      e.g.: Did you come for Atul?
      Here Atul is proper noun hence no “karant” is applied to it.
      Likewise for all the cases rules are written for nouns, pronouns, adjectives and paticiples. Some of the rules are shown in table1.

2. Verb and Auxiliary:
   The “pratyaya” of verb depends on tense of the sentence and gender, number and the person. As shown in table 2 & table 3.

   The output will be generated in Roman script. It can be converted to Marathi language script by using existing software available.

<table>
<thead>
<tr>
<th>Table 1: Karant Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
</tr>
<tr>
<td>The Noun of ‘a’ karant changes to ‘aa’ karant</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Adjective/ Participle/ Pronoun Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Singular</td>
</tr>
<tr>
<td>Plural</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3: Verb table for present tense for Marathi root “ja”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1st Person</td>
</tr>
<tr>
<td>2nd Person</td>
</tr>
<tr>
<td>3rd Person</td>
</tr>
</tbody>
</table>

4. FUTURE WORK
The major challenge is to resolve the ambiguity in the meaning of words in the sentence. E.g. - ‘Are you standing near the bank?’ two possible contexts of the word ‘bank’- bank of river or the money bank.

Grammar of the English Language sometimes allows the change in the sequence of words without changing the meaning of the sentence. E.g. - ‘Should Ram have gone to the store?’ can be written as “Should have Ram gone to the store?”. The former sentence is translated by our system correctly but the latter is not. To allow such flexibility, there is a need to make rules more generalized.
There are three types of questions: 1. Yes/No questions (starting with auxiliary verbs) 2. Information questions (Wh questions) 3. Tag questions e.g. ‘You are going to school, aren’t you?’. Our system supports first two types of questions. The system can be extended for tag questions.

In this paper, translation of simple interrogative sentences is discussed which can be extended for different types of sentences such as imperative, declarative and exclamatory sentences. The scope can be further expanded for compound and complex sentences.

CONCLUSION

This paper presents a proposed system for machine translation of English interrogative sentences to their Marathi counterpart. Our system follows the transfer approach with rule-based translation and emphasizes on interrogative sentences. It is difficult to frame the generalized rules for Marathi because grammar of English and Marathi are out of line. The system is successfully tested on majority of simple interrogative sentences using our production rules.

REFERENCES


