

Machine Translation: Need of the Time

P.R. Devale^{1*}, S.H. Patil²

^{1*}Dept. of Information Technology, College of Engineering, Bharati Vidyapeeth Deemed University, Pune, INDIA

²Dept. of Computer Engineering, College of Engineering, Bharati Vidyapeeth Deemed University, Pune, INDIA

*Corresponding Author: prdevale@bvucoep.edu.in, Tel.: +91-9158052231

Available online at: www.ijcseonline.org

Received: 26/Dec/2017, Revised: 05/Jan/2018, Accepted: 20/Jan/2018, Published: 31/Jan/2018

ABSTRACT- Machine Translation provides a solution in breaking the language barrier so that humans can transform information; it is important application of Natural Language Processing. There are many different languages spoken in this world. Among those language English is the global language. Though it is global language many people can't understand English especially in rural areas of India. From the many years machine translation has been a topic of research. There are many methods and techniques for achieving the machine translation. In India many regional languages are spoken. The mother of all these native languages in India is Sanskrit. A great storage of knowledge with subjects like medicine, mathematics, Geography, Geology, Astronomy, philosophy and many others is kept alive and fresh in Sanskrit for thousands of years. Hence here we have chosen Sanskrit as the target language. Most of the literature, political documents etc. are available in English. So we have chosen English as a source language. This paper illustrates about the language translation mechanism which converts English text to Sanskrit text using Rule Based approach. Presently work on machine translation in India is performed at various locations like JNU, IIT Kanpur, CDAC Pune and many more.

Keywords: Rule based dictionary approach, Parser, Bilingual dictionary, Formant Synthesizer. Morphological analyzer, Corpus based translation, English-to-Sanskrit Transliteration.

I. NEED FOR MACHINE TRANSLATION

Machine translation is very helpful to the people from different places to understand an unknown language without the aid of a human translator. The module presented concerns with the Machine Translation domain of Natural Language Processing. The Source Language (SL) is the language which is to be translated and the Target Language (TL) is the language in which it is translated. During the translation, the syntactic structure and semantics structure of both source language and target language should be considered. We decode the meaning of the source input text in its entirety. The translator must interpret and analyse the text. This process requires deep knowledge of the grammar, semantics, syntax, idioms, etc., of the source language and Target Language.

II. MACHINE TRANSLATION APPROACHES

There are total four approaches of Machine Translation [6]. They are Direct, Rule Based, Corpus Based and Knowledge based.

In Direct Machine Translation [6] there is no intermediate representation of codes. It is only word by word translation with help of dictionary followed by some syntactic rearrangement. It requires little analysis of text and without parsing.

In Rule based Machine Machine Translation (RBMT) [6], parse tree is produced as intermediate representation. It depends on rules for morphology, syntax, lexical analysis, semantic analysis and generation therefore known as Rule based. Rule based can be of two types as Transfer based and Interlingua. In Transfer based, we transfer without intermediate representation from Source Language to Target Language while in Interlingua some intermediate code representation is produced through which input language is translated to Target Language by inter language codes.

Three techniques of machine translation – Direct, Interlingua [1] and Transfer based are applicable to rule based machine translation system.. Other MT systems like SMT and EBMT requires huge parallel corpus for training, which is not readily available for Indian languages. The source of parallel corpus is internet and texts. The rule based system which has been developed follows the transfer based approach of reordering rules.

In Corpus based Machine Translation it requires Corpora for language pair. . It is further classified into Statistical Machine Translation and Example Based Translation

According to the statistical machine translation (SMT) [8][9][10] is a machine translation where translations are generated on the basis of statistical models whose parameters

are derived from the analysis of bilingual text corpora. The SMT is a corpus based approach, where a massive parallel corpus is required for training the SMT systems. The SMT systems are built based on two probabilistic models: language model and translation model. The advantage of SMT system is that linguistic knowledge is not required for building them. The example based machine translation (EBMT) is the corpus based approach without any statistical models. The example based systems are trained with the parallel corpus of example sentences, similar to SMT systems. In Knowledge based MT systems [7], It requires large knowledge base that includes both ontological and lexical knowledge and also it is semantic based approach for language analysis as introduced by Artificial Intelligence researchers.

III. COMPARING ENGLISH WITH SANSKRIT

If we compare the Grammar for both English and Sanskrit [8] then English sentences always in order of subject-verb-object format while Sanskrit has free word order. i.e it can have the form SVO, SOV,

IV. PROPOSED SYSTEM

In our proposed system we used the rule based approach [5][7]. In this method, firstly, the words from the source sentence are taken separately.

In the parsing step the POS (Part Of Speech) tag information and dependency information of these words is obtained. Using this, each word is assigned by source morph features. Then with help of transfer link rule file the corresponding target structure is generated. In the machine translation system English-Sanskrit bilingual dictionary [1] is used for translating the English words to equivalent Sanskrit words.

Parser [2][4][8] is an algorithm which produces a syntactic structure for a given input. The parser is the first component of the rule based machine translation system and it is used on the source (English) side. The Parser is used for four main purposes in the machine translation system. The parser is used for syntactic analysis of the English sentence in order to give the parse tree structure of the English sentence by context free grammar.

This tree structure [2] is required for re-ordering the source (English) sentence with respect to the target (Sanskrit) sentence by transfer rules. The parser is used for Parts of Speech (POS) tagging of the English sentence to give English words and their corresponding POS tags.

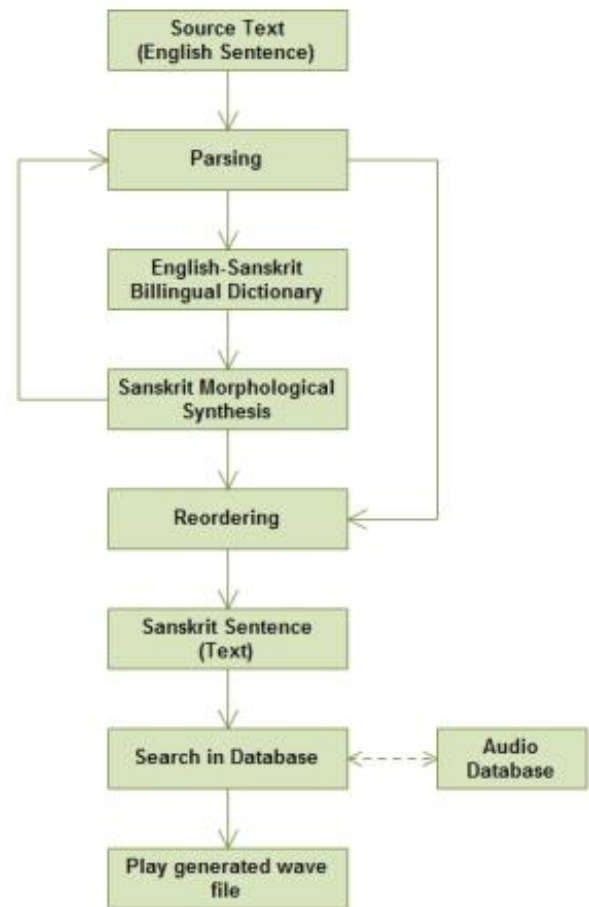


Figure 1: The process of RBMT approach with TTS

The parser is used for stemming the words of English sentence, to get their corresponding root words. The root words of English obtained after stemming are used to find the equivalents of Sanskrit words from bilingual dictionary. The parser is used for the morphological analysis of words in the English sentence, to get the morphology of English words. The morphology information of English is used in the morphological synthesizing for equivalent Sanskrit words.

V. BILINGUAL DICTIONARY

In the machine translation system English-Sanskrit bilingual dictionary [2] is used for translating the English words to equivalent Sanskrit words. Here the pre-processed bilingual dictionary is loaded into the database and MySQL server is used. The dictionary is separated into seven different databases based on the POS categories: Noun, Verb, Adjective, Adverb, Pronoun, Preposition and General.

VI. TRANSLITERATION

In this process, the English text is replaced with the Sanskrit text by preserving the spell. The proper nouns like person

names and place names, named entities, may not have the equivalent Sanskrit words in the bilingual dictionary

In such cases, the system will not produce good output. Such words could not be translated but these words had to be transliterated. After parsing the English sentence the transliteration is invoked. Because only after the parsing, the proper nouns could be identified, by POS tagging them with NNP (proper noun singular) or NNPS (proper noun plural). The word with 'NNP' or 'NNPS' POS category can be directly transliterated without entering into other translation modules.

VII. MORPHOLOGICAL GENERATOR

It adds morphology to the words. A bi-directional morphological Generator cum Morphological Analyzer has been developed for Sanskrit, to synthesize morph to the Sanskrit words. The morph synthesizer is used in the rule based machine translation system for helping in English to Sanskrit translation. The morphological information of English words will be transferred to Sanskrit words.

VIII. REORDERING BY TRANSFER RULES

The reordering denotes the change in syntactic structure of source text with respect to the target text. The syntactic information of English sentence which is come from Parser is checked for the match in the database of reordering rules. When the syntactic pattern of English sentence matches with the source rule, then the corresponding Sanskrit rule is taken and the source tree structure of parser is modified with respect to the target rule. If the pattern matches, then transfer rule is applied to the child nodes of all branches in parse tree.

IX. CONCLUSION

It has become very important, the process of machine translation. Amongst various methods of machine translation, the rule based machine translation method is discussed here, which has its own advantages and limitations also. But still it provides good result.

REFERENCES

- [1] V.M.Barkade, "English to Sanskrit machine translation semantic mapper", International Journal of Engineering Science and Technology Vol. 2(10), 2010, 5313-5318, 2010
- [2] V.M Barkade, "English to Sanskrit Machine Translator Lexical Parser and Semantic Mapper", National Conference On "Information and Communication Technology", NCICT-IOJ, 2010.
- [3] Vimal Mishra, "Approach of English to Sanskrit machine translation based on case-based reasoning, artificial neural networks and translation rules", International Journal of Knowledge Engineering and Soft Data Paradigms, 2010-12-01
- [4] P.R.Devale,V.M..Barkade "English to Sanskrit Machine Translator: Lexical Parser", International Journal on Computer Science and Engineering, ISSN-0975-3397,Vol.02, N0.06, 2010

- [5] P.R.Devale,V.M..Barkade "English to Sanskrit Machine Translator: Semantic Mapper", International Journal of Engineering Science and Technology, ISSN-0975-5462,Vol.02, N0.10, 2010
- [6] P.R.Devale, D.T.Mane, "Rule Based Approach for English-to-Sanskrit Machine Translation and Synthesizer System", Oriental Journal of Computer Science & Technology : An International Journal, Vol.3, No.2, ISSN : 0974-6471, 2010
- [7] P.R.Devale,V.M..Barkade "English to Sanskrit Machine Translator: Lexical Parser and Semantic Mapper",International Journal on Computer Science and Engineering , ISSN-0975-3397, 2010
- [8] P.R.Devale Sandeep Warhade, "Statistical Machine Translation Approach for English-to-Sanskrit Translation in Ubiquitous Environment", International Journal of Engineering and Innovative Technology, ISSN:2277-3754,Vol 1, Issue 6, 2012
- [9] P.R.Devale Sandeep Warhade, "Design of Phrase Based Decoder for English-to-Sanskrit Translation", Journal of Global Research in Computer Science, ISSN:2229-371X, Vol 3, Issue 1, 2012
- [10] P.R.Devale Sandeep Warhade "English-to-Sanskrit Statistical Machine Translation with Ubiquitous Application", International Journal of Computer Applications, ISSN:0975-8887, Vol.51, Issue.01, 2012

Authors Profile

Mr. P.R. Devale pursued Bachelor of Engineering in Computer Engineering from Walchand College of Engineering, Sangli (India) in 1992 and he pursued his Master of Engineering in Computer Engineering from Bharati Vidyapeeth Deemed University College of Engineering, Pune (India) in 2004. Currently he is pursuing his Ph.D in Computer Engineering from Bharati Vidyapeeth Deemed University, Pune (India).

Mr. S.H. Patil pursued Bachelor of Engineering in Computer Engineering from Walchand Institute of Technology, Solapur (India) in 1989 and he pursued his Master of Engineering in Computer Engineering from University of Pune (India), in 1992. He has completed his Ph.D in Computer Engineering from Bharati Vidyapeeth Deemed University, Pune (India) in 2009. Currently he is working as Professor in the Department of Computer Engineering in Bharati Vidyapeeth Deemed University College of Engineering, Pune (India).