ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 1, July 2018

English to Sanskrit rule based machine translation

P.R. Devale[1], S.H. Patil[2]
1. Dept.of Information Technology, College of Engineering, Bharati Vidyapeeth Deemed University, Pune, INDIA
2. Dept.of Computer Engineering, College of Engineering, Bharati Vidyapeeth Deemed University, Pune, INDIA

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 Also Sanskrit is chosen as a target language as it has gained wide acceptance not only in India but across the globe. Here we propose to develop a converter which converts English Statement to Sanskrit statement using Rule based approach of Machine Translation.***

Keywords: **Machine translation; Rule based; Lexical Parser, Semantic mapper, Transliteration.**

I. INTRODUCTION

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.

II. NEED FOR MACHINE TRANSLATION

Machine translation is very helpful to the people from different places to understand an unknown language without the aid of an human translator. The module presented concerns with the Machine Translation domain of Natural Language Processing.

Manuscript received: 25 June 2018
Manuscript received in revised form: 20 July 2018
Manuscript accepted: 07 August 2018
Manuscript Available online: 10 August 2018

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 analyze the text. This process requires deep knowledge of the grammar, semantics, syntax, idioms, etc., of the source language and Target Language.

III. MACHINE TRANSLATION APPROACHES

There are total four approaches of Machine Translation [7]. They are Direct, Rule Based, Corpus Based and Knowledge based.

In Direct Machine Translation [7] there is no intermediate representation of codes. It is only word by word translation with help of dictionary followed by some syntactic rearrangement.

In Rule based Machine Translation (RBMT) [7], 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.

The rule based machine translation system translates the source text into target text by a set of linguistic rules. Three techniques of machine translation – Direct, Interlingua [7] and Transfer based are applicable to rule based machine translation system. The rule based machine translation system is developed by hand coded rules for translation. The system requires good linguistic knowledge to write the rules and a bilingual dictionary is also needed. 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 parallel texts are not widely available in internet and in multi-lingual text books, the alignment of sentences vastly vary. On the other hand, the rule based systems are highly suited for translation of English to Indian Languages because the bilingual dictionary could be collected easily compared to parallel corpus and the rules could also be written well with the help of linguists..

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 1, July 2018

In Corpus based Machine Translation [7] 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) [11],[12],[13],[14] 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. The example based systems generally don't learn from the corpus. They store the parallel corpus and uses matching algorithms to search and retrieve the sentences.

In Knowledge based MT systems [9], 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. The basic Artificial Intelligence approaches include semantic parsing, lexical decomposition into semantic networks and resolving ambiguities.

IV. COMPARING ENGLISH AND 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. For e.g .the order of English sentence (ES) and its equivalent translation in Sanskrit sentence (SS) is given below.

ES: I am going to home
SS :Aham Sadanam Gamishyami .(SOV) **OR**
SADANAM Aham Gamishyami. (OSV) **OR**
Gamishyami Aham Sadanam (VSO)
Thus Sanskrit sentence can be written using SVO, SOV and VSO order.

V. PROCESS OF LEXICAL PARSING

The semantic standard representation was designed to provide a simple description of the grammatical relationships in a sentence that can easily be understood and effectively used by people without linguistic expertise who want to extract textual relations. The sentence relationships are represented uniformly as semantic standard relations between pairs of words. A parser breaks data into smaller elements, according to a set of rules that describe its structure.

Following are the Steps to generate a Parse Tree:

Step I: Input is a English sentence.

Step 2: Lexical Analyzer: Creates Tokens

Step3: Tokens generated acts as an input to Semantic analyzer

Step 4: Semantic analyzer: Creates a parse tree

Step 5: Output is a parse tree

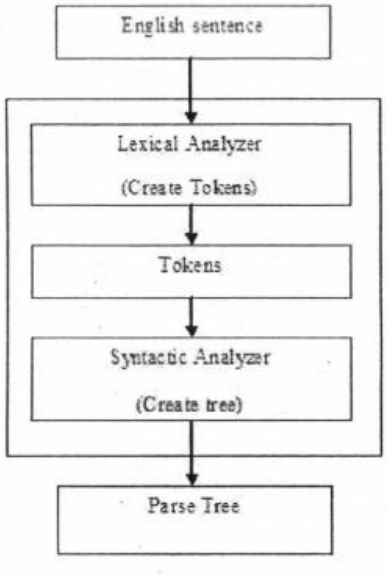

Fig 1: Generation of a Parse Tree

Table 1: Tokenization

Noun (nsubj)	Token1	Bell
Participial modifier (partmod)	Token 2	based
Preposition (prep)	Token 3	in
Noun (nn)	Token 4	Los
Noun (nn)	Token 5	Angeles
Verb	Token 6	makes
Conjunction (conj)	Token 7	and
Verb	Token 8	distributes
Adjective (amod)	Token 9	electronic
Adjective(amod)	Token 10	computer
Conjunction (conj)	Token 11	and
Adjective (amod)	Token 12	building
Directobject (dobj)	Token 13	products

For Example, the sentence:
Bell, based in Los Angeles, makes and distributes electronic, computer and building products is broken into following tokens as shown in Table 1.

VI. APPROACH USED FOR RULE BASED MACHINE TRANSLATION

Major approaches of Machine Translation are rule-based machine translation (RBMT, also known as the rational approach). Rule based translation consists of:

I. Process of analyzing input sentence of a source language syntactically and or semantically

2. Process of generating output sentence of a target language based on internal structure each process is the

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 1, July 2018

strength of the rule based method is that the information can be obtained through introspection and analysis

3. Word based Matching

The matching is not only word by word but it will be semantic (meaningful) matching based on the relationship been established.

(I) (a) ES: A man eats vegetables.
SS: Narah shaakarn khaadati
(A) (Man) (vegetables) (eats)
. (b) ES: Acids eats metal.
SS: Aambat dhaatum nashyati.
(Acids) (Metal) (Eats)
(2) (a) ES: He eats potatoes.
SS: Sah sukantham khaadati.
(He) (Potatoes) (Eats)
(b) ES: Sulphuric acid eats iron.
SS: Gandhak lauham nashyati
(Sulphuricacid) (iron)

VII. PROCESS OF SEMANTIC MAPPING

In this we have to look up in Sanskrit dictionary we are matching English semantic word with the dictionary Sanskrit word i.e. Sanskrit word in English .This matching is not word by word but it will be semantic (meaningful) matching based on the relationship been established in the first module i.e. Lexical Parser. Different Rules are considered for Mapping:

I. Equality Rule

English word directly mapping to Sanskrit word
e.g.
ES : A man eats vegetables.
SS : Narah shaakam khaadati.
(A) (man) (vegetables) (eats)

2. Synonyms Rule (word having same meaning)

English words mapping to Sanskrit word
e.g.
ES : He is a good lfine l excellent man
SS: Sah sajjana

3. Antonyms Rule (word having opposite meaning)

English word not directly mapping to Sanskrit word
e.g.
ES : He is **not** good man
He is a **bad** man
SS : Sah Durjana

Step 1: The output from the first module i.e. lexical parser acts as input to the semantic Mapper

Step 2: The tokens generated from the first module is stored in Data Structure. These tokens have grammatical relations which are represented with various Symbols e.g. conj , nn ,nsubj , det, dobj etc.

Step 3. Look up in Sanskrit dictionary we are matching English semantic word with the dictionary Sanskrit word. This matching is not word by word but it will be semantic (meaningful) matching based on the relationship been established.

Step 4: After matching the selected words from the Sanskrit dictionary are kept as another data structure.

Step 5: Identify the relationships among the various Sanskrit words from these Data Structures.

Step 6. Basic Writing Rules

I. Keep sentences short
2. Make sure sentences are grammatical
3. Avoid complicated grammatical constructions
4. Avoid words which has several meaning

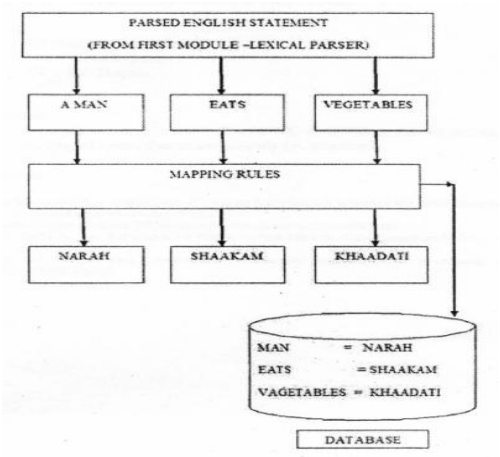

Fig 2: Semantic Mapping

VIII. CONCLUSION

Today looking at the case study of multi-lingual nation, 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] Choe, M.H (2017). Interdisciplinary ICT-Ocea Shachi Dave etal, "Interlingua-based English-Hindi Machine Translation and Language Divergence", Machine Translation, Vol. 16, Issue 4, pp: 251 - 304, 2001.

[2] Promila Bahadur, A.K Jain,D.S Chauhan, "ETranS-English to Sanskrit Machine Translation" ICWET 2012, Bombay, ACM 2012.

[3] V.M.Barkade, P.R.Devale "English to Sanskrit machine translation semantic mapper", International Journal of Engineering Science and Technology Vol. 2(10), 2010, 5313-5318, 2010.

[4] V.M Barkade, P.R.Devale "English to Sanskrit Machine Translator Lexical Parser and Semantic Mapper", National Conference on "Information and Communication Technology", NCICT-IOJ, 2010.

[5] 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.

[6] P.R.Devale, V.M.Barkade "English to Sanskrit Machine Translator: Lexical Parser", International Journal on

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 8, Issue 1, July 2018

Computer Science and Engineering, ISSN-0975-3397, Vol.02, N0.06, 2010.

[7] P.R.Devale, D.T.Mane, " A Design Towards English to Sanskrit Machine Translation and Synthesizer System using Rule Based Approach", , International Journal of Multi-Disciplinary Research and Advances in Engineering , ISSN : 0975-7074, Vol 2, No.II, 2010.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] P.R.Devale Sandeep Warhade "Language Learning and Translation with Ubiquitous Application through Statistical Machine Translation Approach", International Journal of Advances in Engineering & Technology, ISSN: 2231-1963, Vol.4, Issue.02, 2012.

AUTHOR'S 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). His area of Specialization is Microprocessor and Microcontrollers, Computer Organization and Parallel Processing. He has a teaching experience of 25 years.

Dr. 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 Vidyappeth Deemed University College of Engineering, Pune (India). His area of Specialization is Operating System, Distributed Systems and Advanced Computer Architecture He has a teaching experience of 28 years.