International Journal of Future Generation Communication and Networking
Vol. 13, No. 2, (2020), pp. 113 - 120

A Review on various approaches in Machine Translation for Sanskrit Language

Dr.Santosh Deshpande.
Director, MES's, IMCC.

Ms.Neha Kulkarni.
Research Student MES's, IMCC.

Abstract

Machine Translation is an emerging field in computer science. It is one of the most significant applications of Natural Language Processing. Aim to focus on Sanskrit in Machine Translation is to come across the language suitability, its morphology and employ appropriate Machine Translation techniques. A review has been conducted on various approaches in Machine Translation in this paper. It begins with introduction to Natural Language Processing and its applications. Different types of ambiguities are discussed. Silent features of Sanskrit language are discussed. Then focuses on Sanskrit is used for Machine Translation and highlights the language features for Machine Translation. Different approaches of Machine Translation are given like Rule based, Statistical based, Direct etc. A survey of the work done on various machine translation systems either developed or under the development. General structure for Sanskrit Machine Translation system (SMTS) is discussed.

Keywords: *Machine Translation, Natural Language Processing, Sanskrit, Morphology, Lexical*

1. Introduction [2][4][6]

Machine Translation is very important application of Natural Language Processing; it removes the barrier so that humans can transform information, share ideas, know one another cultures, technological discussions etc. Machine translation helps to unite the world socially, culturally and technologically. There is big necessity for inter-language translation for transfer and sharing of information and ideas. Using machine translation one natural language can be translated to other. Natural Language Processing (NLP) involves making computers to perform useful tasks using languages used by humans.NLP have to face a lot of ambiguity during its processing and Sanskrit language overcomes all of these hurdles, because of "formally defined grammar", to become the best suited natural language for machine translation.

2. Natural Language Processing [6]

Natural language processing is the sub field of artificial intelligence dedicated to make computers understand statements or word written or spoken in human language. Natural language understanding at the input side and the natural language generation at the output side are the two major parts of natural language processing

2.1 Application of Natural Language Processing

Natural language processing provides a better human-computer interface that could artificial intelligence systems to pervade more efficiently into the present day applications like:

- Translate one human language to another using translation program.
- A program for grammatical errors checking in a given text.

ISSN: 2233-7857 IJFGCN

- A system for blind people with speech input.
- The chair of Stephan hawking which converts text into speech.

3. Difficulties in Language Translation [7]

Due to different types of ambiguities Machine translation is a difficult task. For translation, ambiguity needs to be resolved. These difficulties are inbuilt in English but are not fundamental to all natural languages. Scientific Sanskrit is particularly specific i.e. clear and accurate. There are different types of ambiguities which depending upon study meaning of word, problem solving, explanation or understanding and number of meaning of one word etc:

- Language translation in structural form- In this, words in a sentence is interpreted after the sentence combined into groups of words, which are without definite verb.
- Difficulty or ambiguity in problem solving in specific way- This is related to context of sentence.
- Lexical difficulty or ambiguity- when a single word has many different meanings and in this all meanings are potentially valid.
- Difficulty in meaning of words or semantic ambiguity- This is related to sentence

4. Silent Features of Sanskrit Language [4][7]

India is multilingual country with as many as 22 scheduled languages of which Sanskrit is one among them and it's official language of state of Uttarakhand, India. It is considered as the oldest Indo-European language. It is holy and philosophical language in Hinduism, Buddhism, and Jainism. The Sanskrit is mother of most Indian languages. Vedas, Extensive epic, Upanishads, philosophical, mathematical, scientific, dramatic, poetic texts include in Sanskrit work. Grammar of Sanskrit is well organized and ambiguity less compared to other natural languages. Sanskrit grammar is given by Panini as “Astadhyayi”. Feature of generating new words is most The most distinctive feature of Sanskrit language is feature of generating new words . 14 sets are given by Panini in Sanskrit language are called “Maheshwara Sutras‟, which explain Sanskrit in mathematical representation or form. Fibonacci series correlated with mathematical expression of language which explains every natural problem. Fibonacci series is so simple and nature follows it because it generates the patterns. And this theory explained in Sanskrit as regenerative for computation. Context developing is based on language’s grammar. Sanskrit language is a set of 14 rules given by Panini. These rules are used to form all sentences in Sanskrit. All these are possible only through object oriented approaches which is available in Sanskrit grammar.

Following are some salient features of Sanskrit language-

- Sanskrits language is promoted as the language of processing for its relatively *unambiguous naturean* and *well laid-out grammatical structure.*
- Sanskrit has a more *strictly defined syntax*, so it is technically more computable.
- Sanskrit is the most *Scientific and Structured* language. There are many hidden algorithms in Sanskrit as a part of its vast scientific treatises, to analyze "Meanings" or "Word sense" from many perspectives.
- The *word representation* in Sanskrit is done *by its property*, not according to the objects. Any object or a thing is named by the property it possesses.
- All Sanskrit *words* are made of *characters*, either *vowels or consonants*. Vowels exist independently, while consonants depend on vowels. The process of *Sandhi* is defined.

ISSN: 2233-7857 IJFGCN

- Sanskrit *words are composed of two parts*, a fixed base part and a variable affix part, both forming an integral unit. The meaning of the word base, depending on a set of given relationships is modified by the variable part.
- Sanskrit is a very *predictable language*. It is easy to formulate sentences and obtain meanings from words. It is easy to make words plural. So that a computer can inherently formulate sentences very easily.
- Words are of either *nominal type or verbal type* i.e. denoting either entities or actions.
- Sanskrit is clearly *differentiate between dual and plural case* and thus we can get an error free NLP.
- *Vibhaktis* (cases) provides an efficient way of *segmenting* the sentences into *logical constructs* for natural language processing (NLP). The splitting of the sentences in Sanskrit is very similar to the semantic net models used for artificial intelligence systems.
- Sentence formation in Sanskrit is done with the help of two well known tools *Vibhakti and Karaka*. Vibhakti assists for making sentence in Sanskrit, there are seven kinds of vibhakti which also provide information on respective karaka. Karaka approach guides for generating grammatical relationship of nouns and pronouns for other words in a sentence.
- Sanskrit has *inflection based syntax* which makes the overall meaning of a sentence independent on the position of its constituent words. An inflection of a word means a different form of that word and is used for enhancing the meaning of the original word.

5. Sanskrit And Machine Translation [4] [5]

Sanskrit has formal defined grammar. For any language to become computationally doable, it should have following features

- Less or Unambiguous Grammar
- Guard against Mispronunciation/ Misspelling Resulting in Misconception
- Total precision
- Co-relation between written and Spoken form of words
- Potential Grammatical Tools

Sanskrit language can be treated as best suited natural language for machine translation as it holds most of these features. The linguistic aspects of Sanskrit language that need to be considered while dealing with complexity in machine translation are as follows-

- *Phonetics and Phonology*—knowledge about linguistic sounds - In Sanskrit it is known as Panini Shiksha shastra which connects to the Grammar and the rules of the grammar also abide by the rules of the Phonetics.
- *Morphology*—knowledge of the meaningful components of words from stems and their generation and utilization - In Sanskrit this is called as 'pada vyutpatti'. In addition the methods for generating words are also explained step-by-step in Panini's Ashtadyayi like a mathematical equation.
- *Lexical*—knowledge of meanings and equivalent words. Every Sanskrit lexical item has a one-one correspondence. So a particular word used in different places is the same from a semantics point of view.

ISSN: 2233-7857 IJFGCN

- *Syntax*—knowledge of the structural relationships between words - declensions of nominal forms /stems - In Sanskrit Vibhakti play this role – it has very tight rule thus there is no ambiguity.
- *Semantics*—the meanings of words in a sentence related knowledge. Sanskrit has many ways of sentence meanings and their analysis on a scientific basis.
- *Pragmatics*— knowledge of the relationship of meaning with respect to the context - this is the most complex as meanings change, based on context and many other factors. For pragmastics a wonderful Vyakarana treatise available in Sanskrit called as "Vakyapadiyam" by Maharishi Bhartrhari.

5.1Comparative View of English and Sanskrit [8]

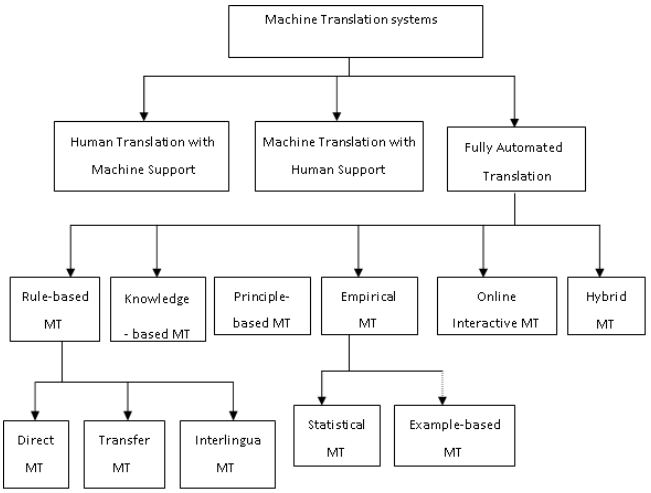
English and Sanskrit comparitative view on different basis as below :

Basis	**English**	**Sanskrit**
Alphabet	26 character	42 character
Number of vowel	Five vowels	Nine vowels
Number of consonant	Twenty one consonant	Thirty three consonant
Number	Two: singular and plural	Three: singular, dual and plural
Sentence Order	SVO (Subject-Verb-Object)	Free word order
Tenses	Three: present, past and future	Six: present, aorist, imperfect, perfect. 1st future and 2nd future
Verb Mood	Five: indicative, imperative, interrogative, conditional and subjunctive	Four: imperative, potential, benedictive and conditional

6. Machine Translation Approaches [2][3]

ISSN: 2233-7857 IJFGCN

Machine Translation is classified into seven broad categories: rule-based, statistical-based, hybrid-based, example-based, knowledge-based, principle-based, and online interactive based methods. First three Machine Translation approaches are the most widely used and earliest methods.



7. Sankrit Machine Translation Systems [1][2]

Comparison between some Machine Translation System either developed or under development.

Machine Translation System	**Approach**	**Source- Target Language Pair**	**Features**
DESIKA	Rule based	Sanskrit to Sanskrit	Desika is a Paninin grammar based system which includes Vedic pro-cessing and shabda-bodha as well
ANGLABHARTI	pseudo-interlingua approach.	English to any Indian Language	English is analyzed only once and creates an intermediate structure called PLIL. Further PLIL is converted to each Indian language through a process of

ISSN: 2233-7857 IJFGCN

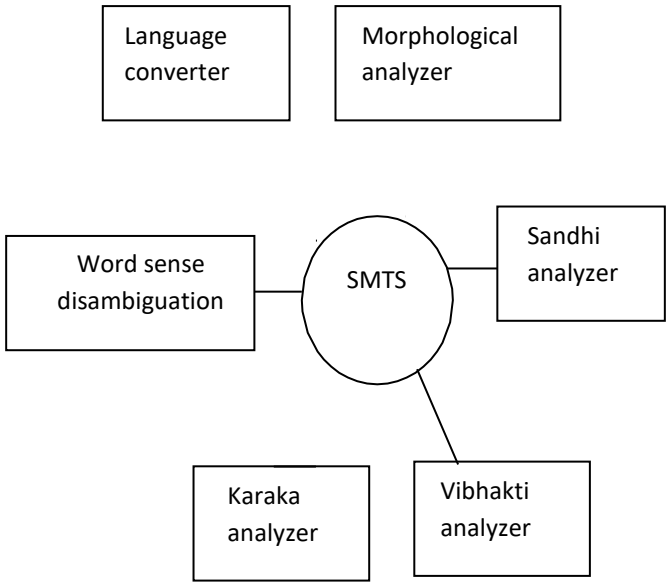
			text-generation..
GOOGLE TRANS-LATOR	statistical machine translation approach.	English to Hindi/Urdu/Sanskrit	Translation is provided only for Hindi, Urdu and Sanskrit.
ETSTS	Rule and Example based	English to Sanskrit	Using Bilingual dictionary and Modular design for converting target sentence to speech output
ESSS	Rule based	English to Sanskrit	English Speech to Sanskrit speech is converted via English and Sanskrit words
E-tranS	Rule based	English to Sanskrit	Synchronous Context Free Grammar (SCFG) is formed and used for language representation of syntax, Lexicon used for Morphological analysis
Sanskrit to English Translator by Subramania m A.	Rule based	Sanskrit to English	Focus on *Sandhi Vichheda* ,Morphological Analysis.

ISSN: 2233-7857 IJFGCN

English to Sanskrit MT by Mishra and Mishra	Example based	English to Sanskrit	POS tagger Module, Uses ANN for verb selection, GNP Module.
English to Sanskrit MT by Warhade S, et al	Statistical based	English to Sanskrit	Phrase based
English to Sanskrit MT by Mane D.T. , et al	Rule based	English to Sanskrit	Use of bilingual dictionary and grammar rules file.

8. Components of Sanskrit Translational System [4]

Developing a Sanskrit Machine Translation System (SMTS) is much more fascinating and challenging task. MT is difficult because words can have several meanings. It is possible only by replacing the words in text by their equivalent words. Then modifying and arranging these words according to grammar. The components of proposed Sanskrit Machine Translation system (SMTS) include the modules as shown in figure



7. Conclusion

Sanskrit language has specific, unambiguous nature and vast literature and vocabulary, which prompt to be used as source language in machine translation. Machine Translation is a difficult task because words

ISSN: 2233-7857 IJFGCN

can have several meanings. Sanskrit is used as source or target language for development of Machine Translation systems. Still some systems are particular to specific domain, restricted to short sentences and phrases. Sanskrit language becomes challenging in Machine Translation application using Corpus based Machine Translation techniques due to its rich morphological nature. Systems using different translation techniques, suitable for particular domain are available for converting English to Sanskrit language. Additionally, we tried to describe briefly the different existing approaches that have been used to develop Machine Translation systems and proposed a general structure of SMTS which tried to utilize the salient grammatical features of Sanskrit language.

References:

[1] Mane, Deepak, and Aniket Hirve. "Study of various approaches in Machine Translation for Sanskrit Language." *International Journal of Advancements in Research & Technology* 2.4 (2013).

[2] Raulji, Jaideepsinh K., and Jatinderkumar R. Saini. "Sanskrit machine translation systems: A comparative analysis." *International Journal of Computer Applications* 136.1 (2016): 1-4.

[3] Antony, P. J. "Machine translation approaches and survey for Indian languages." *International Journal of Computational Linguistics & Chinese Language Processing, Volume 18, Number 1, March 2013*. 2013.

[4]Glida, Dixit, and Narote "General Structure of Machine Translation System" *Journal of Emerging Technologies and Innovative Research (JETIR)* (May 2019)

[5] Mishra, Vipin. "Sanskrit as a Programming Language: Possibilities & Difficulties."

[6] Bathulapalli, Chandana, Drumil Desai, and Manasi Kanhere. "Use of Sanskrit for natural language processing." (2016): 78-81.

[7] Inderjeet "An Approch to Sanskrit as Computational and Natral Language Processing" (sept 2015)

[8] Mishra, Vimal, and R. B. Mishra. "Study of example based English to Sanskrit machine translation." *Polibits* 37 (2008): 43-54.

ISSN: 2233-7857 IJFGCN