Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

Survey on Machine Translation systems for Ancient Indian Languages

Sreedeepa H. S
Assistant Professor
Computer Science & Engineering
College Of Engineering Kallooppara
Thriuvalla

Divya Madhu
Assistant Professor
Computer Science & Engineering
Vidya Academy of Science & Technology
Technical Campus Kilimanoor

***Abstract*— Sanskrit is a less ambiguous, language suitable for natural language processing. Most of the ancient In-dian books were written Sanskrit. This paper is a survey done on different Sanskrit involved machine translation systems.**

***Keywords*— Machine translation; Interlingua; source language; direct translation; destination language.**

I. INTRODUCTION

Sanskrit is a less ambiguous language. As its less ambiguous in nature it is more suitable for natural language processing. [1] Sanskrit is a free word order language. Sanskrit, considered as the mother of most of all languages, possesses a rich grammar which was developed by Panini around 3000 years ago and it includes 3,959 rules. NASA, the most advanced research center in the world has discovered that Sanskrit is the less ambiguous spoken language on the planet. There is saying that Sanskrit is the best suitable language for computers. Due to the unambiguous nature of the language Sanskrit is the simplest language that is most suited for Artificial Intelligence and Natural Language Processors.

Machine translation (MT) is the process of converting one natural language to another using application software. Mainly there are three types of rule based machine translation techniques- direct approach, transfer based approach and interlingua based approach. Most of the translators developed were concern about word translation, bilingual dictionaries based on direct translation.

II. MACHINE TRANSLATION SYSTEMS

Machine translation (MT) is the process of converting sentences in one natural language called source language to another called destination language. One of the the major classification of machine translation approach include Rule based machine translation, Statistical, Example-based, Hybrid machine translation and Neural machine translation. In rule based approach large set of rules are manually developed and apply these rules to map structures from source to target language TABLE I.[2] summarizes the advantages/disadvantages of major machine translation approaches.

I. MACHINE TRANSLATION APPROACHES

Approaches	Advantages	Disadvantages
Rule based	1. Easy to build an initial system 2. Based on linguistic theories 3. Effective for core phenomena	1. Rules are formulated by experts 2. Difficult to maintain and extend 3. Ineffective for marginal phenomena
Knowledge based	1.Based on taxonomy of knowledge. 2. Contains an inference engine. 3.Interlingual representation	1. Hard to build a knowledge hierarchy. 2. Hard to define the granularity of knowledge 3. Hard to represent knowledge
Example based	1.Extracts knowledge from corpus. 2.Based on translation patterns in corpus. 3. Reduces the human cost	1. Similarity measure is sensitive to system. 2. Search cost is expensive. 3.Knowledge acquisition is still problematic.
Statistics based	1.Numerical knowledge 2. Extracts knowledge from corpus. 3. Reduces the human cost 4. The model is mathematically grounded.	1. No linguistic background. 2. Search cost is expensive. 3. Hard to capture long distance phenomena.

Direct translation, transfer based and interlingua based approaches are the major rule based machine translation techniques.

A. *Direct Translation*

Direct translation is the simplest form translation in which words in the source sentence are directly converted into a destination language .In this translation is done with the help of a bilingual dictionary. Word by word translation is performed here. Anusaaraka is an example of direct ta-ranslation based well known machine translation system.

B. *Transfer based Translation*

A database of translation rules is used to translate a text in source language to target language. In this approach whenever a sentence is matched to any one of the rules present in the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

database its directly translated using a dictionary. The dictionary is such as source language(SL) dictionary, target language(TL) dictionary, and a bilingual dictionary. There are mainly two steps in this approach, syntactic transfer and semantic transfer.

In syntactic transfer the SL sentence is analysed to generate asyntactic structure called parse tree and this parse tree of SL is then transfers to TL parse tree. At semantic transfer analyse a SL input to a language specific semantic representation and transfer this to TL semantic representation. Case frames and logical forms are the two constructs used for semantic representation. Finally, these representations are used generate syntactic structure and then surface sentence in the TL.

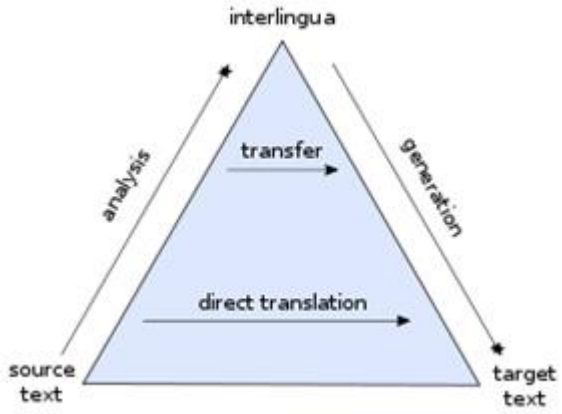
C. Interlingua based Translation

In Interlingua based approach a language independent frame work is developed for translation of source language to destination language. [2] The interlingua approach has a number of advantages. It requires fewer components for the translation of the source language to each target language, and to add a new language. It allows both the analyzers and generators to be written by monolingual system developers. Also, it can handle languages that are different from each other.

DeryleW. Lonsdale, Alexander M. Franz, and John R. R. Leavitt presented the design and development of an interlingua for a large-scale MT project, 1SL-nTL. They also discussed how the resulting Knowledge-based, Accurate Natural-Language Translation (KANT) interlingua handles complexity,and development of different stages efficiently. It is developed in a balanced fashion with maximal coverage. They use, a recursive list-based structural representation of source sentences in this approach. An interlingua frame consists of a head concept, feature-value pairs, and semantic slots. It may contain nested interlingua frames. The source language expressions and semantic units from the domain were considered for the concept generation. The overall format is modeled using frame-based structures. The f-structure reflects deep semantic relationships between major constituents. [8]

The Interlingua approach is based on the concept that MT must go beyond purely linguistic information, syntax and semantics, and should understand the content of texts. Interlingua based translation is divided into two monolingual components: analyzing the source language text into an abstract universal language-independent representation of meaning, the interlingua, and generating this meaning using the lexical units and the syntactic constructions of the target language. [9]

Fig.1 represents the vaquous triangle for machine translation approaches. It depicts the three types of rule based translation system. The main phases present in the translation are analysis transfer and generation phase.



1. .Vaquous triangle for machine translation

III RELATED WORKS

A detailed study on machine translation system on Sanskrit, Interlingua based machine translation system and Paninian framework for translation were done in developing the proposed system. Akshar Bharathi et.al. provided details of the Paninian framework [1], Parsing Free Word Order Languages in the Paninian Framework [2], and Karaka analysis [3]. He also explains the use of lexical functional grammar (LFG) in unification for specifying mapping to grammatical relations[4]. The parsing of Sanskrit sentences using LFG is explained by Mrs. Namrata Tapaswi et.al. [5]. Paul Kiparsky gives detailed description of different levels of Paninian framework with examples and rules of Ashtadhyayi and rule formation on different levels of Paninian framework. [6] Sudhir Kumar Mishra et.al. [7] gives a detailed study on the Karaka analysis system based on rules of Ashtadhyayi with examples.

Sameh AlAnsary et.al. briefly reviews three of the most renowned interlingua-based machine translation projects, Distributed Language Translation (DLT), UNIversal TRANslator(UNITRAN) and KANT system. DLT, a research project developed in Utrecht, The Netherlands, is an interactive system developed to operate over computer networks. Translation is distributed between two independent terminals; one for the analysis and another for generation.UNITRAN is a translation system developed at Massachusetts Institute of Technology. The system operates bidirectionally between Spanish and English. KANT system has been developed at Carnegie- Melon University (CMU) in Pennsylvania, USA in 1989". KANT is the only interlingua-based MT system to be operational commercially. It has been used in translating English technical documents into French, Spanish and German.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

Translation system developed JNU uses word sense disambiguation module and Anaphora Resolution module Here they used Sanskrit as SL and Hindi as TL. Sanskrit to English machine translation developed by Subramanian focus on sandhi vicheda,,and morphological analysis.

IV SANSKRIT INVOLVED MACHINE TRANSLATION SYSTEMS

Some of the Sanskrit involved machine translation systems were shown in the TABLE II.[3]. Most of the systems were developed on the rule based approach.

TABLE II MACHINE TRANSLATION SYSTEMS DEVELOPED FOR SANSKRIT

Machine Translation System	Approach	Source-target Language Pair	Features
ETSTS	Rule and example based	English to Sanskrit	Converts target sentence to speech output, Use of Bilingual dictionary
Sanskrit to English Translator by Subramaniam	Rule based	Sanskrit to English	Focus on Sandhi Vichheda , Morphological Analysis.
English to Sanskrit machine translation by Mishra and Mishra	Rule based	English to Sanskrit	POS tagger Module, Uses ANN for verb selection, GNP Module.
English to Sanskrit machine translation by Mane D.T.etal	Rule based	English to Sanskrit	Use of bilingual dictionary and grammar rules file.
Sanskrit to Hindi MT by JNU.	Rule based	Sanskrit to Hindi	WSD module, Anaphora Resolution module.
Interlingua based Sanskrit to English machine translation	Knowledge bsed	Sanskrit- English	Based on Paninian Grammer

V. CONCLUSION

Linguistic studies on Sanskrit are less compared to other Indian natural languages Rule based translation scheme is used in most of the Sanskrit involved translation systems Most of the systems were developed either in direct or transfer based approaches and for simple sentences. Very rare translation systems uses Sanskrit as source language. There is an interesting and more efficient machine translation system developed based on interlingua approach . As Sanskrit considered as mother of many Indian languages a translation system based on interlingua approach seems to be more efficient and useful.

REFERENCES

[1] R. Briggs, "Knowledge representation in Sanskrit and artificial intelligence," AI magazine, vol. 6, 1985, p. 32. Springer, 2009, pp. 200-218. Annual Conf. Magnetics Japan, p. 301, 1982.

[2] H. S. Sreedeepa and S. M. Idicula, "Interlingua based Sanskrit-English machine translation," *2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT)*, Kollam, 2017, pp. 1-5. doi: 10.1109/ICCPCT.2017.8074251

[3] Jaideepsinh K. Raulji, "Sanskrit Machine Translation Systems: A Comparative Analysis", International Journal of Computer Applications,2016.

[4] Sameh AlAnsary Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University ElShatby, Alexandria, Egypt., "Interlingua-based Machine Translation Systems: UNL versus Other Interlinguas",2014.

[5] Akshar Bharati Rajeev Sangal Department of Computer Science and Engineering Indian Institute of Technology Kanpur Kanpur 208016 India Internet: sangal@iitk.ernet.in, "Parsing Free Word Order Lqanguages in the Paninian Framework," ACL '93 Proceedings of the 31st annual meeting on Association for Computational Linguistics Pages 105-111, June 22 - 26, 1993

[6] Akshar Bharati, Medhavi Bhatia, Vineet Chaitanya, Rajeev Sangal Department of Computer Science and Engineering Indian Institute of Technology Kanpur sangal@iitk.ernet." Paninian Grammar Framework Applied to English"in February 1996

[7] Akshar Bharati, Vineet Chaitanya, Rajeev Sanga" Paninian framework and its application to Anusaraka",Springer, February 1994, Volume 19, Issue 1, pp 113–127

[8] Akshar Bharati, NLP:A Paninian Perspective, PHI Learning, 1996

[9] Namrata Tapaswi, Suresh Jain and Vaishali Chourey, "Parsing Sanskrit Sentences Using Lexical Functional Grammar" Systems and Informatics (ICSAI), International Conference on 19-20 May 2012 pp.2636 – 2640..

[10] Paul Kiparsky, Stanford University, "On the Architecture of Paninian Grammar," UCLA.2002.

[11] Sudhir kumar Mishra, JNU, "Sanskrit Karaka Analyzer for MT,"2007.

[12] Deryle W. Lonsdale, A. M. Franz and J. R. R. Leavitt. "Large Scale Machine Translation: An Interlingua Approach". in Proceedings of the 7 th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Austin, Texas, The United States. 1994

[13] Sameh AlAnsary Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University ElShatby, Alexandria, Egypt., "Interlingua-based Machine Translation Systems: UNL versus Other Interlinguas",2014

[14] P. Goyal, V. Arora and L. Behera, "Analysis of Sanskrit text: Parsing and semantic relations," in Sanskrit Computational Linguistics, Sanskrit Computational Linguistics, 200-218, 2009

[15] Ved Kumar Gupta, Prof. Namrata Tapaswi, Dr. Suresh Jai,Knowledge representation of Grammatical constructs of Sanskrit language using Rule based Sanskrit to English MT, 2016 http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6524744

[16] Dinesh kumar, Gurpreet Sing, POS Tagger for Morphology rich Indian languages, International Journal of Computer Applications (0975 – 8887) Volume 6– No.5, September 2010