I J C T A, 10(9), 2017, pp. 647-657 ISSN: 0974-5572

Machine Translation for Indian Languages: A Review

Niranjan Bhale* and Varsha H. Patil**

ABSTRACT

The evidences of translation are from ancient times. Many a times there is a need to translate from one language to the other. In this era of communication, translation has gained significant importance. By understanding the huge number of languages people use for speaking and writing it remains the challenge to the translators. As far as India is concerned there is tremendous diversity in terms of the spoken languages, scripts and official languages used for communication means. When it comes to the partial or fully automated translation, it poses herculean tasks to the researchers confined to Natural Language Processing, Artificial Intelligence, Machine Translation and Machine Learning. It becomes highly convenient if the documents or the other form of communication happen to be in the language of the user. This paper aims at reviewing the major approaches of Machine Translation, studying existing systems available for translation confined to Indian languages, identifying the domains with pros and cons of the systems, discussing the approaches and finally epitomizing the work for the future directions.

Keywords: Natural Language Understanding, Machine Translation, Statistical Machine Translation, Artificial Intelligence, Dictionary-based Machine Translation

1. INTRODUCTION

Huge textual data are available confined to various languages used for human communication. It is cumbersome to translate the domain specific information to desired target language. So, there is a need to translate the desired domain specific data by using the language of one's choice. The high-end and esoteric information when translated to any of the target language can add to the utility in various applications. There are roughly 6,500 spoken languages in the world today. However, about 2,000 of those languages have fewer than 1,000 speakers. Culturally and linguistically India is the most diverse country in the world. As on today 780 different languages are being spoken in India. Besides that 250 spoken languages have died in the course of last 50 years. Twenty two of the 780 languages are the scheduled Indian languages. Of 780 spoken languages, 122 have been declared by the census as spoken by a population exceeding 10,000 and the rest are spoken by less than 10,000 people.

As per the eighth schedule under the article of Constitution of India the following 22 languages are recognized as official languages:

(i) Assamese, (ii) Bengali, (iii) Gujarati, (iv) Hindi, (v) Kannada, (vi) Kashmiri, (vii) Konkani, (viii) Malayalam, (ix) Manipuri, (x) Marathi, (xi) Nepali, (xii) Oriya, (xiii) Punjabi, (xiv) Sanskrit, (xv) Sindhi, (xvi) Tamil, (vii) Telugu, (viii) Urdu (xix) Bodo, (xx) Santhali, (xxi) Maithili and (xxii) Dogri. Different Indian communities speak approximately 2000 dialects. 13 different scripts are used for the 22 official languages besides English. This diversity creates many challenges as far as the communication among the various entities is concerned. The resultant translation from source language to the target language relies on the knowledge and skill of the interpreter. For official and the other reasons there is always a need to

* Research Scholar: Dept. of Computer Engg. MCERC, Nashik, *Email: nbhale@gmail.com*

** Research Guide: Dept. of Computer Engg. MCERC, Nashik, *Email: varsha.patil@gmail.com*

translate from one language to the other. The recent status of automated translation is confined to the selected domains and the urge for the translation is huge.

With the use of automated translation the correspondence and communication among the different States and Center in India can be better and will add to the overall efficacy of the Government.

This paper is organized into 5 sections. Section 1 is an Introduction; Section 2 includes major different approaches to build a MT system. Section 3 elaborates major MT Systems in India based on approaches of translation along with their characteristic peculiarities, translation quality, domain etc. Section 4 discusses in short the summary of literature review for major MT systems. Section 5 gives the conclusion.

2. MACHINE TRANSLATION

Though the exact definition varies among the scholars, Natural Language can broadly be defined in contrast to Artificial or Constructed languages (such as computer programming languages and international auxiliary languages) and to other communication systems in nature (such as dance of Bird of Paradise). Definitions of "Natural Language" also usually state or imply that a "Natural" language is one that any cognitively normal human infant is able to learn and whose development has been through use rather than by prescription [1]. There are around 6500 natural languages exist in the world today.

Machine Translation (MT) can be defined as a sub-field of Computational Linguistics and Machine Intelligence that investigates the use of software to translate text or speech from one language to another.

2.1. Approaches for Machine Translation

Around the world many Machine Translation Systems were developed based on various approaches. Of the major Natural languages for which such systems were developed and being developed are English, Russian, Spanish, Japanese, Chinese and Hindi.

The major Machine Translation approaches used by the researchers are as shown in the Figure 1.

Every approach has its pros and cons. Statistical Machine Translation is used at large by the researchers.

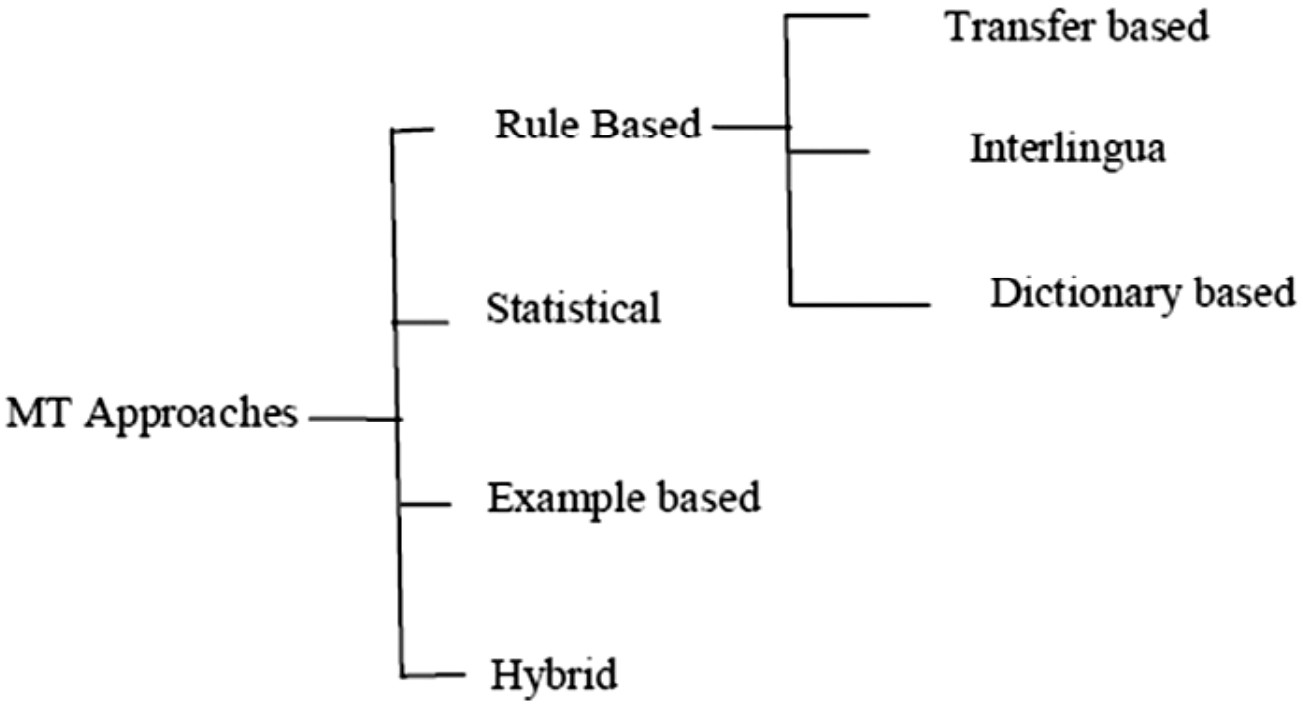

Figure 1: Machine Translation Approaches

3. Machine Translation for Indian Languages

3.1. MT systems developed on Rule based approach

At IIT Kanpur, Anusaaraka project was started by Rajeev Sangal in 1995 which was later continued at IIIT Hyderabad. Ministry of Information Technology, Technology Development for Indian Languages (TDIL), Satyam Computers Pvt. Ltd. and Government of India had funded this project. For the Machine Translation the source languages are Punjabi, Marathi, Telugu, Kannada and Bengali while the target language is Hindi. The system was tested for the translation of the kids' stories from all the source languages to Hindi. By understanding the importance of context the system succeeded in maintaining the information preservation. The system suffered from the major drawback that the generated output of the system follows the grammar of the source language. For Example, a Punjabi to Hindi translation generates the output in Hindi that can be understood by the person who knows Hindi despite the grammatical imperfections. Although word to word translation achieved by the system is about 80%.[30].

Originally the intent of Anusaaraka developers was not the pure Machine Translation, but to provide inter-language access among specified Indian language pairs. Currently, the system is attempting English to Hindi Machine Translation. The system cleverly exploits the close similarity of Indian languages using Paninian Grammar (PG) rules.[30]

In 1997 for English to Hindi Machine Translation system named Mantra was developed by Bharati for information preservation. It has used XTAG based super tagger and light dependency analyzer developed at University of Pennsylvania for analysing of the input English text. Equal load was distributed between human and machine by using a novel way. The system generates multiple outputs corresponding to a given input. Output based on the most detailed analysis of the English input text, used a bilingual dictionary and a full parser. The generated output if the user decides inappropriate then the corrections for precision of the translation can be done manually [15].

In 1999 Authors developed "MANTRA" a Machine assisted Translation tool[41]. This tool converts English text into Hindi for a particular domain of office correspondences and documents. Rigorous testing was carried out while translating official documents like appointment orders, notifications, circulars and notices. Tree Adjoining Grammar (TAG) formalism and Lexicalized Tree Adjoining Grammar (LTAG) were used to represent the English and Hindi grammar. The work was continued providing the future scope for different source and target pair of the languages such as Hindi-English, English-Telugu, English- Bengali, English-Gujarati, Hindi-Bengali and Hindi-Marathi. This approach has provided generality with limited dictionary (lexicon) and grammar rules.

Researchers[4] have developed "ANGLABHARTI", a machine aided translation system in 2001 aiming at translation from English to some of the Indian languages using pseudo-interlingua approach. Using this approach translation from English to more than one Indian language is possible reducing the requirement of separate translation system for English to every considered Indian language. The analysis of English as a source language is done only once and it creates intermediate structure - PLIL (Pseudo Lingua for Indian Languages). The PLIL is then converted to targeted Indian language through a process of text-generation. The effort for PLIL generation is 70 per cent and text generation is 30 per cent. With mere 30 per centadditional effort,s new English to Indian language translation system can be built. The attempt has been made whereby has to do 90 per cent translation task and remaining 10 per cent is left for the post-editing. The domain of this machine translation system was public health.

An English-Hindi translation system was developed by Authors[6] in 2001 using Universal Networking Language (UNL) as the Interlingua structure. UNL is an international project specifically initiated to cover the translation for majority of natural languages. The developers have materialized the translation system for English-Marathi, English-Hindi, Hindi-UNL, UNL-Hindi, and English-Bengali language pairs.

In 2002 Researchers[8] came up with English to Hindi translation system based on transfer based approach using different grammatical rules of source and target languages and a bilingual dictionary for translation. The system was constituted by the modules pre-processor, English tree generator, post- processor for English tree, generator for Hindi tree, Post-processor for Hindi tree and output generator. The domain of the system was weather narration.

Author[20] developed a machine assisted translation (MAT) system for translating English texts into Kannada in 2002, which has used morphological analyzer and generator for Kannada. The English input sentence is parsed by Universal Clause Structure Grammar (UCSG) parser and outputs the number, type and inter-relationships amongst various clauses in the sentence and the word groups. Bilingual dictionary was used to map each word with equivalent target language[23]. At last the target language sentence was generated by placing the clauses and the word groups in appropriate linear order, according to the target language grammar. Additionally, using Post editing tool editing of the translated text could be possible. MAT System 1.0 had shown about 40-60 accuracy. Government circulars was the domain of the translation system.

The translation system named as Shakti was developed by Researchers[23] in 2003. The system translates English to many Indian languages. It combines linguistic statistical approach and rule-based approach. The system was fairly complex with 69 different modules. 9 modules were used for analyzing the source language, 24 modules were used for performing bilingual tasks, and 36 modules were used for generating target Indian language constructs.

In 2003 AnglaHindi a derivative of AnglaBharti MT System was developed by Authors[19] for English to few Indian languages, which was a pseudo interlingual rule-based English to Hindi Machine- Aided Translation System. It had all the modules of AnglaBharti and also abstracted example-base for translating frequently encountered noun phrases and verb phrases. The accuracy of the translation is 90%.

Author[15] has developed a system in 2004 for English - Telugu dictionary containing 42,000 words. A word form synthesizer for Telugu was developed and incorporated in the system. The system was able to handle various complex English sentences.

In 2004 Author developed a Telugu-Tamil translation system using the Telugu Morphological analyzer and Tamil generator for translation. The system made the use of Telugu-Tamil dictionary developed as a part of MAT Lexica. It has also used verb sense disambiguation based on verbs argument structure to handle the verb ambiguity[15][18].

Researchers[42] developed a system , OMTrans in 2004, which translates text from English to Oriya based on grammar and semantics of the source and target languages. Word Sense Disambiguation (WSD) is also handled in this system. OMTrans was designed and developed using principles of object-oriented approach.

In 2004 and 2006 Authors[24] have developed the systems using transfer-based approach with frame-like structured representation. Heuristics were used to resolve the ambiguities in translation. News, technical phrases and annual reports were the domains for the system. It has a text categorization component which determines the type of news story (political, terrorism, economic etc.) before operating on the given story. It has different dictionaries for different domains like politics, terrorism, economy etc. ultimately it chooses an appropriate dictionary depending on the category of news. Considerable human assistance was required in analyzing the input. Another novel component of the system is sentence splitter, which breaks a complex English sentence into simpler sentences. These simple sentences were further analyzed and used to generate Hindi sentences [34][37].

In 2007, 2008 Authors[17] have developed a system which was based on direct word-to-word MT approach for Punjabi to Hindi. This system comprised of modules such as pre-processing, word-to-word

translation using lexicon, morphological analysis, word sense disambiguation, transliteration and post processing. Accuracy of the translation is 90.67%. Word Error Rate is 2.34% and SER is 24.26% [25].

Researcher[20] has developed a system English-Kannada Machine Assisted Translation at Resource Centre for Indian Language Technology Solutions (RC-ILTS), University of Hyderabad in year 2009. The system used a transfer-based approach and it has been applied to the domain of government circulars. The translation system has used Universal Clause Structure Grammar (UCSG) formalism.

In year 2009 Authors[3] developed a Tamil-Hindi MT system which was based on Anusaaraka (started in 1984). MT system architecture was developed by Prof. C N Krishnan. It has used a lexical-level translation and has 80-85% coverage. Both stand-alone and web-based on-line versions have been developed. Tamil morphological analyzer and Tamil-Hindi bilingual dictionary were the bi- products of this system. It has performed exhaustive syntactical analysis. They have also developed a prototype of English-Tamil MAT system. Currently, it has limited vocabulary and small set of Transfer rules [4][33].

TDIL and consortium of 11 Indian institutions have collaboratively initiated a project named as Sampark System: multipart machine translation system for *Indian Language to Indian Language Machine Translation* (ILMT) in 2009[29]. Computational Paninian Grammar (CPG) along with machine learning were used for analyzing languages. It was developed using traditional rules-based and dictionary-based algorithms with statistical machine learning. This consortium has developed language technology for 9 Indian languages resulting in Machine Translation for 18 Indian language pairs.

In 2010 Researchers[4] developed the extended version of web based Hindi-to-Punjabi MT System. The system has several facilities like website translation, email translation etc. They used direct word to word translation approach for translation. The translation has a blend of intelligibility and accuracy of 95.40% and 87.60% respectively. The metrics for the translation yielded WER of 4.58% whereas SER is 28.82% and BLUE score found is 0.7801 [5][11][13][17].

In 2014, Authors[2] have developed ‘Transmuter’, a system for English to Marathi translation for assertive sentences in the domain of Tourism. The characteristic peculiarity of the system is the robust Word Sense Disambiguation model. As compared to Google Translate, Anuvadaksha and Sakava the translation yields better precision while tested with 1000 different sentences. So despite the tedious approach the system was claimed to be precise in translating the sentences from English to Marathi.

3.2. MT systems developed on Statistical based approach

Researchers[23] developed a MT system Shakti in 2003 which translates English text to any Indian language with simple system architecture. It combines linguistic rule based approach with statistical approach. The system contains 69 different modules. Nine modules are used for analyzing the source language (English), 24 modules are used for performing bilingual tasks, and the remaining modules are used for generating target Indian language.

In 2006 E-ILMT (English to Indian Languages MT System) a MT System for English to Indian Languages in Tourism and Healthcare Domains was developed. It was developed by a Consortium of Nine institutions namely C-DAC Mumbai, IISc Bangalore, IIIT Hyderabad, C-DAC Pune, IIT Mumbai, Jadavpur University Kolkata, IIIT Allahabad, Utkal University Bangalore, Amrita University Coimbatore and Banasthali Vidyapeeth, Banasthali[29]. The project is funded by Department of Information Technology, MCIT Government of India. The role of C-DAC Mumbai is to develop statistical models and resources for a statistical MT (SMT) system from English to Hindi/Marathi/Bengali. The engine was initially developed as a baseline system using the state-of-art statistical techniques and the contemporary tools that include the POS tagger (fnTBL), parser (Bikel), decoder (Pharaoh) etc. The primary objective is to initially build an English-Hindi translation system capable of translation of free flow text as found on the web and gradually

adapt it to other Indian language pairs as well. The training corpus (translation model) consisted of 5000 sentences and 800 sentences were split for testing and tuning. The baseline techniques used in this system were inadequate in producing a good quality translation. Therefore, pre-processing stage was included in the system which takes care of syntactic re-ordering on the source language to reduce long distance movements through SMT. It has helped to obtain a better phrase alignment table which resulted in a good improvement in the translation quality using Moses decoder with Giza++ alignment tool. The corpus (translation model) training size for achieving this effort was 12299 sentences with additional 1570 sentences split for testing and tuning Some degradation in the output even after the syntactic processing was observed due to unavailability of sufficient corpus. The syntactically processed corpus was morphologically processed and used for training to counteract the problem of degradation in translation quality. A rule based suffix separation approach was used to separate the root word and the affixes due to the unavailability of sophisticated morphological analyzers.

3.2.1. MT systems developed on Example based approach

ANUBAAD a MT system which translates news headlines from English to Bengali using example based Machine Translation approach was developed by Author[18]. in year 2000, and 2004. An input as an English news headline was given to the system. The input is at first searched in the direct example-base for an exact match. If a match is found, the Bengali headline from the example-base is produced as output. If match is not found, the headline is tagged and the tagged headline is searched in the Generalized Tagged example-base. If a match is found in Generalized Tagged Example-Base, the Bengali headline is to be generated after appropriate synthesis. If a match is not found, the Phrasal example-base will be used to generate the target translation. If the headline still cannot be translated, the heuristic translation strategy is applied where translation of the individual words or terms in their order of appearance in the input headline will be generated. Appropriate dictionarieshave been consulted for translation of the news headlines.

Researchers[7] developed an Automatic Machine Translation system for Bengali-Assamese News Texts, named as VAASAANUBAADA using Example Based Machine Translation (EBMT) approach in 2002. It translates Bengali sentence to Assamese. The system requires preprocessing and post-processing. The bilingual corpus has been constructed and aligned manually by feeding the real examples using pseudo code. Longer sentences are fragmented at punctuations to obtain better translation quality. When the exact match is not found at sentence/fragment level in Example-Base, the backtracking is used and further fragmentation of the sentence is done [22].

In year 2003 Shiva and Shakti are the two Machine Translation systems developed collaboratively by Carneige Mellon University of USA, International Institute of Information Technology, Hyderabad and Indian Institute of Science, Bangalore, India for translating English language to Hindi language. ‘Shiva’ was designed using an Example-based MT and the system Shakti was designed using combination of rule based and statistical approaches of MT. The Shakti system is working for three target languages like Hindi, Marathi and Telgu and and claimed to produce machine translation systems for new languages rapidly. The system is used for translating English sentences into an appropriate target Indian language. The rules used for target language generation are mostly linguistic in nature and the statistical approach tries to infer or use linguistic information. Semantic information is also used by some modules in the system[5][6].

In year 2003 ANGLABHARTI-II was developed by Author[5]. He has suggested a generalized example-base (GEB) approach for hybridization besides a Raw Example-Base (REB). It was found that the modification in the rule-base system was difficult during the development phase and resulted in unpredictable results at times. The example-base approach has grown interactively by augmenting the rule-base. It has the provisions for automated pre-editing and paraphrasing, generalized and conditional multi-word expressions as well as recognition of named-entities. It also contains the modules for an error-analysis and

statistical language-model for automated post-editing. The automatic pre-editing module is used to transform/ paraphrase the input sentence to an intermediate form, which can further be translated easily. Automatic pre-editing may even fragment an input sentence if the fragments can easily be translated and positioned in the final translation. The system also contains a 'failure analysis' module. The failure analysis module consists of heuristics on speculating the reasons for inappropriate translation[6].

Researchers[31] have developed Hinglish - a machine translation system for pure Hindi to pure English forms in year 2004. It has incorporated additional level to the existing English to Hindi translation , AnglaBharti-II and Hindi to English translation , AnuBharti-II systems developed by Sinha. The system has produced satisfactory acceptable results in more than 90% of the cases. The system was not capable of resolving the meaning of few verbs due to a very shallow grammatical analysis used in the process.

Authors[5] developed a MT system for English to {Hindi, Kannada, Tamil} and Kannada to Tamil Language-Pair in 2006 based on a bilingual dictionary. It comprised of sentence dictionary, phrases dictionary, words dictionary and phonetic dictionary. Each of these dictionaries contains parallel corpus of sentences, phrases, words and phonetic mappings of words in their respective files. Example-Base has a set of 75000 most commonly spoken sentences that are originally available in English [6].

The MATREX System was developed by Authors[28] in 2008. The system made a use of marker-based chunking, which is based on the Marker Hypothesis, a psycholinguistic constraint which claims that all languages are marked for surface syntax by a specific closed set of lexemes or morphemes which signify context. Using a set of closed-class (or "marker") words, such as determiners, conjunctions, prepositions, possessive and personal pronouns, aligned source-target sentences are segmented into chunks during a pre-processing step. A chunk is created at each new occurrence of a marker word in such a way that each chunk must contain at least one content (or non-marker) word. In order to align the chunks obtained by the chunking procedures, the system makes use of an "edit-distance style" dynamic programming alignment algorithm.

3.2.2. MT systems developed on Hybrid based approach

Anubharti a system was developed by Sinha[6] in 1995-2004 using a hybridized example-based machine translation approach i.e. a combination of example-based, corpus-based approaches and some elementary grammatical analysis. The example-based approaches follow human-learning process for storing knowledge from past experiences to be used in the future. In this system the traditional EBMT approach has been modified to reduce the requirement of a large example-base. The modification in traditional EBMT was achieved by generalizing the constituents and replacing them with abstracted form from the raw examples. The abstraction was achieved by identifying the syntactic groups. Matching of the input sentence with abstracted examples was carried out, which was based on the syntactic category and semantic tags of the source language structure. The architectures of both AnglaBharti and AnuBharti, have undergone a considerable change from their initial conceptualization. In 2004 these systems were named as AnglaBharti-II and AnuBharti-II respectively. AnglaBharti-II has used a generalized example-base for hybridization besides a raw example-base while AnuBharti-II made use of Hindi as source language for translation to the target language.

MT system ANUBHARTI-II was developed by Author[6]. in 2004 using Generalized Example-Base (GEB) along with Raw Example-Base (REB) approach for hybridizationThe combination of example-based approach and traditional rule-based approach was used in this system. The example based approach mimics human-learning process for storing knowledge from past experiences to be used in future. The Hindi sentence is converted into a standard form to handle the word-order variations. The Hindi sentences converted into standard form are matched with a top level standard form of example-base. If no match is found then a shallow chunker is used to fragment the input sentence into small units and then they are

matched with a hierarchical example-base. The small chunks obtained by shallow chunker are translated and positioned by matching with sentence level example base.

Researchers[26] have developed a hybrid Machine Translation system. It has used an integration of SMT with a lexical transfer based system (RBMT). The experimentation showed that BLEU scores of SMT and lexical transfer based system when evaluated separately were 0.1745 and 0.0424 respectively. The performance of hybrid system claimed to be better [36].

Researchers [6] have developed Bengali to Hindi MT System in 2009, which uses a technique of generation proper lexical translation of Bengali to Hindi Machine Translation. In the baseline system, Bengali word is replaced by the frequently occurring word in Hindi. However, the most frequent translation may fail to generate the output by considering the context of the words [36]. The proposed method finds a better lexical choice amongst the dictionary options with the help of the contextual information of a monolingual corpus of Hindi. The system takes Bengali sentence and converts it to Hindi sentence with the help of lattice-based data structure. The baseline system used for comparison and proposed translation systems are evaluated using the BLEU automatic evaluation tool and human evaluation process. Training Hindi corpus size was 500K.

4. SUMMARY OF THE LITERATURE REVIEW

By reviewing the literature the major observations and characteristic peculiarities are listed below,

- As far as the Machine Translation based on Direct Machine Translation approach the quality of the translation is very coarse. Majority of the Systems based on this approach needs post-editing which is highly undesirable in case of real time applications. Due to the lack of precision the scope of translation is limited to very few domains only.
- The systems developed using Transfer Based approach mainly used the source language as English. Load is shared between the machine and the man. Few systems need pre-processing. Few systems produce multiple translations and choice of precision rest with the user. Though the domain claimed to be general in few cases the loss of precision is observed to meet the generality.
- The use of Interlingua based approach is mainly desirable when there are the choices among the target languages. Significant advantage for the use of this approach is that there is no need to develop separate translation system for each of the language pairs. The efforts are divided into intermediate representation and target language conversion. In few cases post-editing becomes essential. In terms of precision the approach provides around 90% accuracy in most of the cases.
- In the systems built with Statistical Machine Translation approach preprocessing become inevitable in most of the systems. Major advantage is that it does not require language grammar for translation. Rapid translation is the characteristic peculiarity of these systems. Most of the systems developed with this approach try to infer the use of linguistic information. Large number of modules for processing during translation adds to the complexity of such systems. Accuracy observed to be fair in case of this approach.
- Majority of the Translation Systems developed for Indian Languages fall in the category where Example Based Machine Translation approach is used. Few of the systems in this category require pre and/or post-processing. The rules used for the translation are mostly linguistic in nature and number of the rules grows rapidly in case the generality is desired. High accuracy is one of the most desirable features of these systems. Another desirable feature is that it becomes very easy to extend the translation corresponding to new Indian languages. Major disadvantage is the cost associated with the search is comparatively at the higher side.

5. CONCLUSION

This paper attempts to provide the information about the significant development of Machine Translation Systems confined to targeted Indian languages. From the review it is apparent that many of the developed Indian language MT systems are based on Rule-based, Hybrid and StatisticalMachine Translation approaches. Domain specific translation is the characteristic of majority of these systems. It is also evident that the structure and grammatical foundations of the target language has major impact in the performance of the Translators. Though it is highly undesirable most of the MT systems involving Indian languages need significant post-editing for better translation quality.

In the forthcoming years MT systems based on Hybrid approach will produce the desirable effects confined to various important domains of utility.

REFERENCES

[1] Fernandes, Keith (2008). On the Significance of Speech: How Infants Discover Symbols and Structure. Ann Arbor, MI: ProQuest. pp. 7–8. ISBN 978-1243524065.

[2] G V Garje, G.K. Kharate, Harshad Kulkarni : Transmuter: An Approach to Rule-based English to Marathi Machine Translation**,** International Journal of Computer Applications (0975 – 8887) Volume 98 – No.21, July 2014

[3] Sitender & Seema Bawa, (2012) "Survey of Indian Machine Translation Systems", *International Journal Computer Science and Technolgy*, Vol. 3, Issue 1, pp. 286-290, ISSN : 0976-8491

[4] Vishal Goyal & Gurpreet Singh Lehal, (2009) "Advances in Machine Translation Systems", *National Open Access Journal*, Volume 9, ISSN 1930-2940B

[5] Vishal Goyal & Gurpreet Singh Lehal, (2010) "Web Based Hindi to Punjabi Machine Translation System", *International Journal of Emerging Technologies in Web Intelligence*, Vol. 2, no. 2, pp. 148-151, Academy Publishers

[6] Sanjay Kumar Dwivedi & Pramod Premdas Sukhadeve, (2010) "Machine Translation System in Indian Perspectives", *Journal of Computer Science* 6 (10): 1082-1087, ISSN 1549-3636, © 2010

[7] Kommaluri Vijayanand, Sirajul Islam Choudhury & Pranab Ratna "VAASAANUBAADA - Automatic Machine Translation of Bilingual Bengali-Assamese News Texts", *in proceedings of Language Engineering Conference-2002*, Hyderabad, India © IEEE Computer Society.

[8] Lata Gore & Nishigandha Patil, (2002) "English to Hindi - Translation System", *In proceedings of Symposium on Translation Support Systems*. IIT Kanpur. pp. 178-184.

[9] Sitender & Seema Bawa, (2012) "Survey of Indian Machine Translation Systems", *International Journal Computer Science and Technolgy*, Vol. 3, Issue 1, pp. 286-290, ISSN : 0976-8491

[10] Sanjay Kumar Dwivedi & Pramod Premdas Sukhadeve, (2010) "Machine Translation System in Indian Perspectives", *Journal of Computer Science* 6 (10): 1082-1087, ISSN 1549-3636, © 2010

[11] Vishal Goyal & Gurpreet Singh Lehal, (2009) "Advances in Machine Translation Systems", *National Open Access Journal*, Volume 9, ISSN 1930-2940

[12] Latha R. Nair & David Peter S., (2012) "Machine Translation Systems for Indian Languages", *International Journal of Computer Applications* (0975 - 8887) Volume 39- No.1

[13] Vishal Goyal & Gurpreet Singh Lehal, (2010) "Web Based Hindi to Punjabi Machine Translation System", *International Journal of Emerging Technologies in Web Intelligence*, Vol. 2, no. 2, pp.

[14] Shachi Dave, Jignashu Parikh & Pushpak Bhattacharyya, (2002) "Interlingua-based English-Hindi Machine Translation and Language Divergence", *Journal of Machine Translation*, pp. 251-304.

[15] Sudip Naskar & Shivaji Bandyopadhyay, (2005) "Use of Machine Translation in India: Current status"*AAMT Journal*, pp. 25-31.

[16] Sneha Tripathi & Juran Krishna Sarkhel, (2010) "Approaches to Machine Translation", *International journal of Annals of Library and Information Studies*, Vol. 57, pp. 388-393

[17] Gurpreet Singh Josan & Jagroop Kaur, (2011) "Punjabi To Hindi Statistical Machine Transliteration", *International Journal of Information Technology and Knowledge Management* , Volume 4, No. 2, pp. 459-463.

[18] S. Bandyopadhyay, (2004) "ANUBAAD - The Translator from English to Indian Languages", *in proceedings of the VIIth State Science and Technology Congress*. Calcutta. India. pp. 43-51

[19] R.M.K. Sinha & A. Jain, (2002) "AnglaHindi: An English to Hindi Machine-Aided Translation System", *International Conference AMTA(Association of Machine Translation in the Americas)*

[20] Murthy. K, (2002) "MAT: A Machine Assisted Translation System", *In Proceedings of Symposium on Translation Support System(STRANS-2002)*, IIT Kanpur. pp. 134-139.

[21] Lata Gore & Nishigandha Patil, (2002) "English to Hindi - Translation System", *In proceedings of Symposium on Translation Support Systems*. IIT Kanpur. pp. 178-184. International Journal on Natural Language Computing (IJNLC) Vol. 2, No.4, OCtober 2013

[22] Kommaluri Vijayanand, Sirajul Islam Choudhury & Pranab Ratna "VAASAANUBAADA - Automatic Machine Translation of Bilingual Bengali-Assamese News Texts", *in proceedings of Language Engineering Conference-2002*, Hyderabad, India © IEEE Computer Society.

[23] Bharati, R. Moona, P. Reddy, B. Sankar, D.M. Sharma & R. Sangal, (2003) "Machine Translation: The Shakti Approach", *Pre-Conference Tutorial, ICON-2003*.

[24] Ananthakrishnan R, Kavitha M, Jayprasad J Hegde, Chandra Shekhar, Ritesh Shah, Sawani Bade & Sasikumar M., (2006) "MaTra: A Practical Approach to Fully- Automatic Indicative English- Hindi Machine Translation", *In the proceedings of MSPIL-06*.

[25] G. S. Josan & G. S. Lehal, (2008) "A Punjabi to Hindi Machine Translation System", *in proceedings of COLING-2008*: Companion volume: Posters and Demonstrations, Manchester, UK, pp. 157-160.

[26] Sanjay Chatterji, Devshri Roy, Sudeshna Sarkar & Anupam Basu, (2009) "A Hybrid Approach for Bengali to Hindi Machine Translation", *In proceedings of ICON-2009, 7th International Conference on Natural Language Processing*, pp. 83-91.

[27] Vishal Goyal & Gurpreet Singh Lehal, (2011) "Hindi to Punjabi Machine Translation System", *in proceedings of the ACL-HLT 2011 System Demonstrations*, pages 1-6, Portland, Oregon, USA, 21

[28] Ankit Kumar Srivastava, Rejwanul Haque, Sudip Kumar Naskar & Andy Way, (2008) "The MATREX (Machine Translation using Example): The DCU Machine Translation System for ICON 2008", *in Proceedings of ICON-2008: 6th International Conference on Natural Language Processing*, Macmillan Publishers, India, http://ltrc.iiit.ac.in/proceedings/ICON-2008.

[29] Sampark: Machine Translation System among Indian languages (2009) https://www.tdil-dc.in/index.php?option=com_vertical&parenid=74, http://sampark.iiit.ac.in

[30] Akshar Bharti, Chaitanya Vineet, Amba P. Kulkarni & Rajiv Sangal, (1997) "ANUSAARAKA: Machine Translation in stages', *Vivek, a quarterly in Artificial Intelligence*, Vol. 10, No. 3, NCST

[31] R. Mahesh K. Sinha & Anil Thakur, (*th*2005) "Machine Translation of Bi-lingual Hindi-English (Hinglish) Text", *in proceedings of 10 Machine Translation Summit* organized by Asia-Pacific

[32] Parameswari K, Sreenivasulu N.V., Uma Maheshwar Rao G & Christopher M, (2012) "Development of Telugu-Tamil Bidirectional Machine Translation System: A special focus on case divergence", *in proceedings of 11th International Tamil Internet conference*, pp 180-191

[33] Salil Badodekar, (2004) "Translation Resources, Services and Tools for Indian Languages", *a report of Centre for Indian Language Technology, IITB*, http://www.cfilt.iitb.ac.in/Translation

[34] Ananthakrishnan R, Kavitha M, Jayprasad J Hegde, Chandra Shekhar, Ritesh Shah, Sawani Bade & Sasikumar M, (2006) "MaTra: A Practical Approach to Fully-Automatic Indicative English- Hindi Machine Translation", *in proceedings of the first national symposium on Modelling and shallow parsing of Indian languages (MSPIL-06)* organized by IIT Bambay,

[35] CDAC Mumbai, (2008) "MaTra: an English to Hindi Machine Translation System", a report by CDAC Mumbai formerly NCST.

[36] Sanjay Chatterji, Praveen Sonare, Sudeshna Sarkar & Anupam Basu, (2011) "Lattice Based Lexical Transfer in Bengali Hindi Machine Translation Framework", in *Proceedings of ICON- 2011: 9th International Conference on Natural Language Processing*, Macmillan Publishers, India. Also accessible from ltrc.iiit.ac.in/proceedings/ICON-2011.

[37] R. Ananthakrishnan, Jayprasad Hegde, Pushpak Bhattacharyya, Ritesh Shah & M. Sasikumar, (2008) "Simple Syntactic and Morphological Processing Can Help English-Hindi Statistical Machine Translation", *in proceedings of International Joint Conference on NLP (IJCNLP08)*, Hyderabad, India.

[38] Sugata Sanyal & Rajdeep Borgohain, (2013) "Machine Translation Systems in India", *Cornel University Library*, arxiv.org/ftp/arxiv/papers/1304/1304.7728.pdf

[39] Antony P. J., (2013) "Machine Translation Approaches and Survey for Indian Languages", *International journal of Computational Linguistics and Chinese Language Processing* Vol. 18,

[40] Manoj Jain & Om P. Damani, (2009) "English to UNL (Interlingua) Enconversion", *in proceedings of 4th Language and Translation Conference (LTC*-09).

[41] Hemant Darabari, (1999) "Computer Assisted Translation System- An Indian Perspective", *in proceedings of MT Summit VII*, Thialand

[42] S. Mohanty & R.C.Balbantray , (2004) "English to Oriya Translation System (OMTrans)" cs.pitt.edu/chang/cpol/c087.pdf

[43] Smriti Singh, Mrugank Dalal, Vishal Vachhani, Pushpak Bhattacharyya & Om P. Damani, (2007) "Hindi Generation from Interlingua (UNL)", *in proceedings of MT Summit, 2007* language.worldofcomputing.net