SURVEY OF MACHINE TRANSLATION SYSTEMS IN INDIA

G V Garje[1] and G K Kharate[2]

[1]Department of Computer Engineering and Information Technology PVG's College of Engineering and Technology, Pune, India

[2]Principal, Matoshri College of Engineering and Research Centre, Nashik, India

ABSTRACT

The work in the area of machine translation has been going on for last few decades but the promising translation work began in the early 1990s due to advanced research in Artificial Intelligence and Computational Linguistics. India is a multilingual and multicultural country with over 1.25 billion population and 22 constitutionally recognized languages which are written in 12 different scripts. This necessitates the automated machine translation system for English to Indian languages and among Indian languages so as to exchange the information amongst people in their local language. Many usable machine translation systems have been developed and are under development in India and around the world. The paper focuses on different approaches used in the development of Machine Translation Systems and also briefly described some of the Machine Translation Systems along with their features, domains and limitations.

KEYWORDS

Machine Translation, Example-based MT, Transfer-based MT, Interlingua-based MT

1. INTRODUCTION

India is a multilingual country where the spoken language changes after every 50 miles. There are 22 official languages; and approximately 2000 dialects are spoken by different communities in India. English and Hindi are used for official work in most states of India. The state governments in India predominantly carry out their official work in their respective regional language whereas the official work of Union government is carried out in English and/or Hindi. All the official documents and reports of Union government are published in English or Hindi or in both English and Hindi. Many newspapers are also published in regional languages. Translating these documents manually is very time consuming and costly. Hence there is need to develop good machine translation (henceforth referred as MT) systems to address all these issues, in order to establish a better communication between states and Union governments and exchange of information amongst the people of different states with different regional languages. Indian languages are divided into five language families; viz. Indo-Aryan (76.87% speakers), Dravidian (20.82% speakers), Austro-Asiatic (1.11% speakers), Tibeto-Burman (1% speakers) and Andmanese (less than 0.001% speakers). Many Indian languages being *low resource* languages become a major hurdle in the development of MT systems for Indian languages. [1][37].

Many researchers, institutions and research organizations in India have started working on MT systems for English to Indian languages and among Indian languages have succeeded in obtaining

very satisfactory results. The Government of India has decided to give more thrust to Language Technology for Indian languages during VIIIth plan and to initiate a programme that would emphasize on quality, national relevance and participation of traditional knowledge and R & D efforts in the area of information processing in Indian languages. The Department of Electronics of Government of India launched a National level programme during the year 1990-91 on *Technology Development for Indian Languages* (TDIL) [16]. Other institutions like IIT Kanpur, IIT Bombay, IIIT Hyderabad, University of Hyderabad, NCST Mumbai, CDAC Pune, CDAC Noida, Department of Computer Science and Engineering Jadavpur University, Kolkata, JNU New Delhi etc are playing a major role in developing the MT systems in India. Many MT systems have been developed and are being developed. The MT systems have been developed using different machine translation approaches. This paper provides brief information about development year, source & target language, translation approach, domain, salient features, and translation accuracy of major Machine Translation systems in India [2]. There is an immense need to translate these documents in respective state's local language for proper communication with common people of the state. More than 95% of the Indian population is deprived of the benefits of Information Technology due to language barrier [16].

This paper is organized into 6 sections. Section 2 gives a brief history of MT system; section 3 gives an idea of the different approaches to build a MT system. Section 4 discusses major MT systems in India based on translation approaches along with their features, translation quality, domain etc. Section 5 describes the summary of literature review in brief for major MT systems. Section 6 gives the comparison of major translation approaches.

2. BRIEF HISTORY OF MT

Table 1 shows the time line chart of Machine translation at International level [3][23][24][25][27].

Table 1. Time Line Chart of MT

Period		Year	Activity
1948 to 1960	The beginning	1949	Warren Weaver proposed the first idea on the use of computers in translation by adopting the term computer translation.
		1952	The first symposium of MT was held at MIT under leadership of Yehoshua Bar-Hillel.
		1954	The first basic automatic Russian-English translator was developed by a group of researchers from Georgetown University in collaboration with IBM that translated more than sixty Russian sentences. Victor Yngve published the first journal on MT, entitled Mechanical translation devoted to the translation of languages by the aid of machines.
1960 to 1966	Parsing and disillusionment	1961	The computational linguistics was born due to weekly lectures organized by David G. Hays at the Rand Corporation in Los Angeles. First International Conference on MT of Languages and Applied Language Analysis of Teddington was held with the participation of linguists and computer scientists. The scientists involved in the translation work were Paul Garvin, Sydney M. Lamb, Kenneth E. Harper, Charles

1960 to 1966	Parsing and disillusionment		Hockett, Martin Kay and Bernard Vauquois.
		1964	Creation of committee ALPAC (Automatic Language Processing Advisory Committee) with American government to study the perspectives and the chances of machine translation.
		1966	ALPAC published its famous rapport in which it concluded that its work on machine translation was just waste of time and money. Conclusion of this rapport made a negative impact on the MT research for number of years.
1966 to 1980	New birth and hope	1970	Start of the project REVERSO by a group of Russian researchers. Development of System SYSTRAN1 (Russian-English) by Peter Toma, who was member of a group search for Georgetown at that time.
		1976	Developed a MT system WEATHER in the project TAUM (machine translation in the university of Montreal) under the direction of Alai Colmerauer for the machine translation weather-forecasts for the general public. This system was created by group of researchers.
		1978	Developed a MT system ATLAS2 by the Japanese firm FUJITSU. This translator was based on rules. It was able to translate from Korean to Japanese and vice versa.
1980 to 1990	Japanese invaders	1981	The Japanese firm SHARP developed Automatic translator DUET (English - Japanese), which was based on rules and transfer approach.
		1983	NEC developed a system based on algorithm called PIVOT named as Honyaku Adaptor II, used for Interlingua approach.
		1986	OKI3 Developed a Japanese-English system PENSEE, which was rule based translator. Hitachi developed a translation Japanese-English system HICATS (Hitachi Computer Aided Translation System).
1990 to 2000	Web & new vague of translators	1993	The project C-STAR (Consortium for Speech Translation Advanced Research) was initiated. It was trilingual (English, German & Japanese) MT system for the parole in the field of tourism (dialogue client travel agent)
		1998	Marketing of machine translator REVERSO was done by Softissimo.
2000 to 2010		2000	Japanese Laboratory ATR developed a (Japanese-English & Chinese - English) system ALPH. This system used Example based approach of MT.
		2005	The first web site for automatic translation by Google was launched
		2007	A hybrid MT METIS-II was developed that used the SMT, EBMT, and RBMT machine translation approaches.
		2008	23% of Internet users used the MT and 40 % considering doing so.
2000 to 2010		2009	30% the professionals have used the MT and 18% perform a proofreading.
		2010	28% of Internet users used the MT and 50% planned to do so.

3. MACHINE TRANSLATION APPROACHES

Many MT systems across the globe have already been developed for the most commonly used natural languages such as English, Russian, Japanese, Chinese, Spanish, Hindi and other Indian languages etc. Figure 1 depicts the existing machine translation systems and various approaches used in developing these systems.

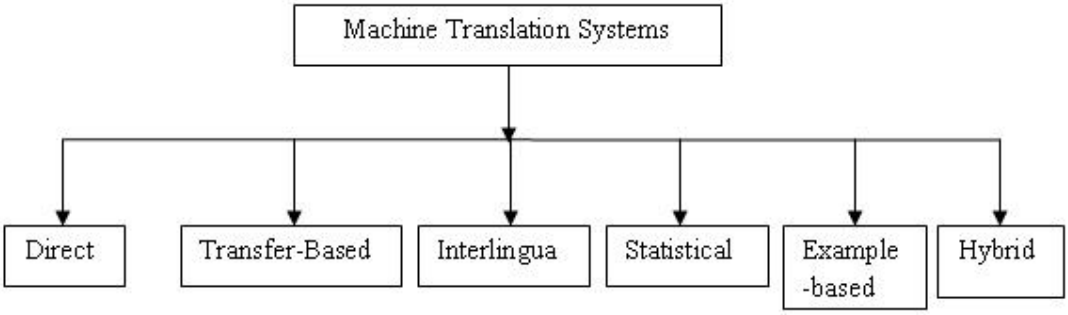

Figure 1. Machine Translation Systems

4. INDIAN MACHINE TRANSLATION SYSTEMS

4.1. Direct Machine Translation Systems

4.1.1. Anusaaraka systems among Indian Languages (1995)

Anusaaraka project started at IIT Kanpur by Rajeev Sangal is now being continued at IIIT Hyderabad. The purpose of the project was the MT of one Indian language to another Indian language. The project is being funded by Technology Development for Indian Languages (TDIL), Ministry of Information Technology, Government of India and Satyam Computers Private Limited. The source languages are {Telugu, Kannada, Bengali, Punjabi and Marathi} and the target language is Hindi. It is not domain specific but the system has been tested mainly for translating children's' stories. The system was mainly developed for the purpose of perfect "information preservation". The output of the system followed the grammar of the source language only.

For Example, a Bengali to Hindi translation can take a Bengali text and produce output in Hindi that can be understood by the user but may not be grammatically perfect. For 80% of the Kannada words in the Anusaaraka dictionary [26], 1997) of 30,000 root words, there is a single equivalent Hindi word which covers the sense of the original Kannada word.

The focus of Anusaaraka was not mainly on MT, but it was on language access between Indian languages. It is currently attempting an English-Hindi machine translation. It uses a Paninian Grammar (PG) and exploits the close similarity of Indian languages [26][27].

4.1.2. Punjabi to Hindi MT System (2007, 2008)

G S Josan and G S Lehal have developed a system which is based on direct word-to-word MT approach. This system comprised of modules such as pre-processing, word-to-word translation using Punjabi-Hindi lexicon, morphological analysis, word sense disambiguation, transliteration

and post processing. Accuracy of the translation produced by this system is 90.67%. Word Error Rate is 2.34% and SER is 24.26% [19][38].

4.1.3. Web based Hindi-to-Punjabi MT System (2010)

Goyal V and Lehal G S developed the extended version of Hindi-to-Punjabi MT System to Web. The system has several facilities like website translation, email translation, etc. [6].

4.1.4 Hindi-to-Punjabi MT System (2009, 2011)

Goyal V and Lehal G S developed a system that uses direct word to word translation approach at Punjabi University, Patiala. The translation accuracy of the system is 95.40% on the basis of intelligibility test and 87.60% on the basis of accuracy test. In the quantitative tests the Word Error Rate is 4.58% whereas Sentence Error Rate is 28.82% and BLUE score found is 0.7801[1][10][21][32].

4.2. Transfer-Based MT Systems

4.2.1. Mantra MT (1997)

Mantra is English to Hindi MT system developed by Bharati for information preservation. The text available in one Indian language is made accessible in another Indian language with the help of this system. It uses XTAG based super tagger and light dependency analyzer developed at University of Pennsylvania for performing the analysis of the input English text. It distributes the load on man and machine in novel ways. The system produces several outputs corresponding to a given input. Output based on the most detailed analysis of the English input text, uses a full parser and a bilingual dictionary. The parsing system is based on XTAG (Bandyopadhyay 2002, consisting of super tagger and parser) with minor modification for the task at hand. A user may read the output produced after the full analysis, but when she finds that the system has "obviously" gone wrong or failed to produce the output, she can always switch to a simpler output [8].

4.2.2. MANTRA MT(1999)

Hemant Darbari and Mahendra Kumar Pandey developed a MAchiNe assisted TRAnslation tool (MANTRA). It translates English text into Hindi in a specific domain of personal administration that includes gazette notifications, office orders, office memorandums and circulars. The system was tested for the translation of administrative documents such as appointment letters, notification and circular issued in central government from English to Hindi in initial stage. It uses the Tree Adjoining Grammar (TAG) formalism developed by University of Pennsylvania and Lexicalized Tree Adjoining Grammar (LTAG) (Bandyopadhyay, 2004) to represent the English and Hindi grammar. It is based on synchronous Tree Adjoining Grammar and uses tree transfer for translating from English to Hindi. It has become a part of "The 1999 Innovation Collection" on information technology at Smithsonian institution's National museum of American history, Washington DC, USA. The project was funded by the Rajya Sabha Secretariat. The grammar of the system was specifically designed to accept, analyze and generate sentential constructions in "Official" domain and the lexicon was restricted to deal with meanings of English words that are used in its subject domain.

The system is developed for the Rajya Sabha Secretariat, the Upper House of Parliament of India and used to translate the proceedings of parliament such as study to be laid on the Table, Bulletin Part-I and Part-II. This system can also be used for other language pairs such as English- Bengali,

English-Telugu, English-Gujarati and Hindi-English and also among Indian languages such as Hindi-Bengali and Hindi-Marathi. The Mantra approach is general, but the dictionary (lexicon) and grammar rules developed were limited to the sub-language of the domain [8][28].

4.2.3. An English–Hindi Translation System (2002)

Gore L and Patil N developed a system based on transfer based translation approach, which uses different grammatical rules of source and target languages and a bilingual dictionary for translation. The translation module consists of pre-processing, English tree generator, post-processing of English tree, generation of Hindi tree, Post-processing of Hindi tree and generating output. The domain of the system was weather narration [14].

4.2.4. MAT (2002)

Murthy K developed a machine assisted translation system for translating English texts into Kannada, which 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. For each word, suitable target language equivalence is obtained from the bilingual dictionary. Finally, the target language sentence is generated by placing the clauses and the word groups in appropriate linear order, according to the target language grammar. Post editing tool is provided for editing the translated text. MAT System 1.0 had shown about 40-60% of fully automatic accuracy. The domain of the translation system was government circulars [13].

4.2.5. Shakti (2003)

Bharati, R Moona, P Reddy, B Sankar, D M Sharma and R Sangal have developed a system which translates English to any Indian languages with simple system architecture. It combines linguistic rule-based approach with statistical approach. The system consists of large number of modules (69 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 [16].

4.2.6. English-Telugu MT System (2004)

Bandyopadhyay S developed a system that uses English - Telugu dictionary containing 42,000 words. A word form synthesizer for Telugu is developed and incorporated in the system. The system handles various complex English sentences [30].

4.2.7. Telugu-Tamil MT System (2004)

Bandyopadhyay S developed a system that uses the Telugu Morphological analyzer and Tamil generator for translation. The system makes use of Telugu-Tamil dictionary developed as a part of MAT Lexica. It also uses verb sense disambiguation based on verbs argument structure to take care of ambiguity in the meaning of the verb [30].

4.2.8. OMTrans(2004)

Mohanty S, Balabantaray R C developed a system that translates text from English to Oriya based on grammar and semantics of the source and target language. Word Sense Disambiguation

(WSD) is also handled in this system. OMTrans is designed and developed using principles of object-oriented approach [17][39].

4.2.9. The MaTra System (2004, 2006)

Ananthakrishnan R, Kavitha M, Hegde J J, Chandra Shekhar, Ritesh Shah, Sawani Bade, and Sasikumar M have developed a system that uses transfer-based approach using a frame-like structured representation. It also uses heuristics to resolve ambiguities. The domain of the system is news, annual reports and technical phrases. 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 political, terrorism etc. and chooses an appropriate dictionary depending on the type of news. It requires considerable human assistance in analyzing the input. Another novel component is sentence splitter which breaks a complex English sentence into simpler sentences. These simple sentences are further analyzed and used to generate Hindi sentences [18][32][33].

4.2.10. English-Kannada machine-aided translation system (2009)

K Narayana Murthy has developed a system English-Kannada Machine Assisted Translation at Resource Centre for Indian Language Technology Solutions (RC-ILTS), University of Hyderabad. The system uses a transfer-based approach and it has been applied to the domain of government circulars. English-Kannada machine translation system uses Universal Clause Structure Grammar (UCSG) formalism. The system is funded by the Karnataka government [4][5].

4.2.11. Tamil-Hindi Machine-Aided Translation system (2009)

Sobha L, Pralayankar P and Kavitha V developed a system which is based on Anusaaraka (started in 1984). MT system architecture is developed by Prof. C N Krishnan. It uses 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 (36K) are the bi-products of this system. It performs exhaustive syntactical analysis. They have also developed a prototype of English-Tamil MAT system. Currently, it has limited vocabulary (100-150 sentences) and small set of Transfer rules [1][4][31].

4.2.12. Sampark System: Automated Translation among Indian Languages (2009)

A consortium of 11 institutions in India have developed a multipart machine translation system for *Indian Language to India Language Machine Translation* (ILMT) funded by TDIL program of Department of Electronics and Information Technology (DeitY), Govt. of India. It uses Computational Paninian Grammar (CPG) for analyzing language and combines it with machine learning. It is developed using both 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 [25].

4.3. Interlingua Machine Translation Systems

4.3.1. ANGLABHARTI (2001)

R M K Sinha, Jain R, Jain A developed a machine aided translation system designed for translating English to Indian languages. It is developed using pseudo-interlingua approach. The

interlingua approach made it possible to use the same system for translating English to more than one Indian language and has eliminated the need of developing separate translation system for English to each 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 each Indian language through a process of text-generation. The effort for PLIL generation is 70% and text generation is 30%. Only with an additional 30% effort, new English to Indian language translation system can be built. The attempt has been made whereby has to do 90% translation task and remaining 10% is left for the human post-editing. The domain of this machine translation system has been public health [4].

4.3.2. UNL-based English-Hindi MT System (2001)

Dave S, Parikh J and Bhattacharyya P developed a translation system using Universal Networking Language (UNL) as the Interlingua structure. The UNL is an international project aimed to create an Interlingua for all major human languages. IIT Mumbai is the Indian participant in UNL project. English-Hindi, Hindi-UNL, UNL-Hindi, English-Marathi and English-Bengali were also developed using UNL formalism [2][7][40][41].

4.3.3. AnglaHindi (2003)

AnglaHindi is a derivative of AnglaBharti MT System developed by R M K Sinha and Jain A for English to Indian languages, which is a pseudo interlingual rule-based English to Hindi Machine-Aided Translation System. It uses all the modules of AnglaBharti and also uses abstracted example-base for translating frequently encountered noun phrases and verb phrases. The accuracy of the translation is 90% [12].

4.4. Hybrid Machine Translation Systems

4.4.1. Anubharti Technology (1995, 2004)

Anubharti (Sinha, 2004) is developed 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 and to be used it in future. In Anubharti, the traditional EBMT (Gupta and Chatterjee, 2003) approach has been modified to reduce the requirement of a large example-base. The modification in traditional EBMT is achieved by generalizing the constituents and replacing them with abstracted form from the raw examples. The abstraction is achieved by identifying the syntactic groups. Matching of the input sentence with abstracted examples is done 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 uses a generalized example-base for hybridization besides a raw example-base and the AnuBharti-II makes use of Hindi as source language for translation to any other language. The generalization of the example-base is dependent upon the target language [2][37].

4.4.2. ANUBHARTI-II (2004),

R M K Sinha developed a MT system using Generalized Example-Base (GEB) along with Raw Example-Base (REB) MT approach for hybridization. The combination of example-based approach and traditional rule-based approach is used in this system. The example based approach emulates human-learning process for storing knowledge from past experiences and to be used in

future. The source language is Hindi. The inputted 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 [2][37].

4.4.3. Bengali to Hindi MT System (2009)

Chatterji S, Roy D, Sarkar S and Basu A developed a hybrid Machine Translation system. It uses an integration of SMT with a lexical transfer based system (RBMT) i.e. multi-engine Machine Translation approach. The experimentation shows that BLEU scores of SMT and lexical transfer based system when evaluated separately are 0.1745 and .0424 respectively. The performance of hybrid system is better and its BLEU score is 0.2275 [20].

4.4.4. Lattice Based Lexical Transfer in Bengali Hindi MT Framework (2011),

Sanjay Chatterji, Praveen Sonare, Sudeshna Sarkar, and Anupam Basu described a method for making proper lexical translation in Bengali to Hindi Machine Translation Framework and used a transfer based MT approach. In the baseline system, Bengali word is replaced by the most frequent word in Hindi. However, the most frequent translation may not be correct if we consider the context of the word(s). 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. It is observed that the proposed system performs better. Training corpus size is 500K Hindi corpus and is tested on Hindi to Bengali [34].

4.5. Example Based Machine Translation (EBMT) Systems

4.5.1. ANUBAAD (2000, 2004)

Bandyopadhyay S developed a MT system which translates news headlines from English to Bengali using example based Machine Translation approach. An English news headline given to the system as an input is initially 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 dictionaries have been consulted for translation of the news headlines [11].

4.5.2. VAASAANUBAADA (2002)

Vijayanand K, Choudhury S I and Ratna P developed an Automatic Machine Translation system for Bengali-Assamese News Texts using Example Based Machine Translation (EBMT) approach. It involves Bengali-Assamese sentence level Machine Translation for Bengali text. It includes preprocessing and post-processing tasks. 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 quality translation. 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[15].

4.5.3. Shiva and Shakti MT System (2003)

MT system 'Shiva' is designed using an Example-based and the system Shakti is designed using combination of rule based and statistical approaches. The Shakti system is working for three target languages like Hindi, Marathi and Telgu and can produce machine translation systems for new languages rapidly. Shiva and Shakti are the two Machine Translation systems from English to Hindi developed jointly by Carneige Mellon University of USA, International Institute of Information Technology, Hyderabad and Indian Institute of Science, Bangalore, India. 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. Currently the system is working for three languages (Hindi, Marathi and Telugu) [2][4].

4.5.4. ANGLABHARTI-II (2004)

R M K Sinha suggested a generalized example-base (GEB) approach for hybridization besides a Raw Example-Base (REB). It is found that the modification in the rule-base system is difficult during development phase and may result in unpredictable results; the example-base approach is grown interactively by augmenting rule-base base. The system first attempts a match in REB and GEB before invoking the rule-base at the time of actual use. It also provides 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 a form which can be easily translated. Automatic pre-editing may even fragment an input sentence if the fragments are easily translatable 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 wrong translation. The system includes various sub-modules [2][4].

4.5.5. Hinglish MT System (2004)

Sinha and Thakur developed Hinglish - a machine translation system for pure Hindi to pure English forms. It incorporates 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 is not capable of resolving the meaning of polysemous verbs due to a very shallow grammatical analysis used in the process [29].

4.5.6. English to {Hindi, Kannada, Tamil} and Kannada to Tamil Language-Pair Example Based MT (2006)

Balajapally P, P Pydimarri, M Ganapathiraju, N Balakrishnan and R Reedy developed a MT system based on a bilingual dictionary comprising of sentence dictionary, phrases dictionary, words dictionary and phonetic dictionary. Each of the 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. All

the sentences in Example-Base have been manually translated into three target Indian languages, namely Hindi, Kannada and Tamil [2][4].

4.5.7. The MATREX (MT using Example) System (2008)

Ankit Kumar Srivastava, Rejwanul Haque, Sudip Kumar Naskar and Andy Way developed a DCU Machine Translation System for ICON 2008. The MATREX system makes use of marker-based chunking, which is based on the Marker Hypothesis (Green, 1979), a psycholinguistic constraint which posits 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 (Gough and Way, 2004) 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 [22][36].

4.6. Statistical Machine Translation Systems

4.6.1. Shakti (2003)

Bharati, R Moona, P Reddy, B Sankar, D M Sharma and R Sangal developed a MT system 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 [2][4].

4.6.2. English to Indian Languages MT System (E-ILMT) (2006)

The EILMT is a MT System for English to Indian Languages in Tourism and Healthcare Domains. It is 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. 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. The system is extended and tested for English-Marathi and English-Bengali pairs with the statistics shown in Table 2 [35].

Table 2. Performance of E-ILMT

Language Pair	Training size	Testing + Tuning size
English-Marathi	13598	1500 (750+750)
English-Bengali	13015	1550

5. SUMMARY OF LITERATURE REVIEW FOR MAJOR INDIAN MT SYSTEMS

5.1. Direct Machine Translation Approach

The following table describes the summary of MT systems developed using Direct MT approach and their salient features and/or limitations

Table 3. Summary of Direct MT Approach

SN	MT System name	Languages	Domain/main application	Salient Features/Limitations
01	Anusaaraka systems among Indian languages (1995)	{Telugu, Kannada, Bengali, Punjabi and Marathi} to Hindi	Domain free but the system has been applied mainly for translating children's stories	-Translation quality is very coarse -The focus is not mainly on MT but on language access between Indian languages. -It is currently attempting an English-Hindi MT
02	Punjabi to Hindi Machine Translation System (2007, 2008)	Punjabi to Hindi	General	-Translation quality is very coarse -Requires post processing -Accuracy 90.7%
03	Web based Hindi-to-Punjabi Machine Translation System(2010)	Hindi to Punjabi	Web pages, email	-Translation quality is very coarse , extension of Punjabi to Hindi MT -Requires post processing -Web-based
04	Hindi-to-Punjabi Machine Translation System (2009, 2011)	Hindi to Punjabi	General	-Translation quality is very coarse -Requires post processing -Accuracy 95.4% -BLEU score 0.7804 -Word Error Rate is 4.58% and Sentence Error Rate is 28.82%

5.2. Transfer-Based Machine Translation Approach

The following table describes the summary of MT systems developed using Transfer-Based MT approach and their salient features and/or limitations

Table 4. Summary of Transfer-Based MT Approach

SN	MT System name	Languages	Domain/main application	Salient Features/Limitations
01	MANTRA System(1997) MANTRA-Rajayasabha (1999)	English and Hindi English and Hindi	Office administration documents Proceeding of Rajyasabaha	- Uses a TAG and LTAG - Distributes a load on man and machine in a novel way - Uses a TAG, Synchronous TAG and Tree Transfer - The system produces several outputs corresponding to a given input - Output based on the most detailed analysis of the English input text uses a full parser and bilingual dictionary -The Mantra approach is general but the lexicon/grammar has been limited to sub-language of the domain
02	An English–Hindi Translation System(2002)	English And Hindi	Weather narration	-Rule-based -The translation module consists of Pre-processing, English tree generator, post-processing of English tree, generation of Hindi tree, Post-processing of Hindi tree and generating output
03	MAT System (2002)	English and Kannada	General	-UCSG(Universal Clause Structure Grammar), morphological analyzer & post-editing -MAT System 1.0 has shown about 40-60% fully automatic accuracy -The system is tested for government circulars
04	Shakti (2003)	English and IL	General	-Linguistic rule-base with Statistical processing -The system consists of 69 different modules, of which 9 modules are used for analyzing the source language, 24 modules are used for performing bilingual tasks, and the remaining modules are used for generating target Indian language.
05	English-Telugu Machine Translation System (2004)	English and Telugu	General	-Rule-based -Lexicon consists of 42,000 words -Handles various complex English sentences

5.3. Interlingua-Based Machine Translation Approach

The following table describes the summary of MT systems developed using Interlingua-Based MT approach and their salient features and/or limitations

Table 5. Summary of Interlingua-based MT Approach

SN	MT System name	Languages	Domain/main application	Salient Features/Limitations
01	ANGLABHARTI (2001)	English to IL	Public health	-Uses pseudo-interlingua -Possible to use the same system for translating English to more than one Indian language due to interlingua approach -No need to develop separate translation system for English to each Indian language English (source) language creates intermediate structure – PLIL (Pseudo Lingua for Indian Languages) -The effort for PLIL generation is 70% and text generation is 30% -Only with an additional 30% effort, new English to Indian language translation system can be built 90% translation task is done by machine and 10% left to the human post-editing.
02	UNL-based English-Hindi MT System (2001)	English to Hindi, UNL to Hindi, Hindi to UNL	General	-Universal Networking Language (UNL) as Interlingua -Developed for English-Hindi, Hindi-UNL, UNL-Hindi, English-Marathi and English-Bengali -Easy to add new language for translation -UNL is an international project with an aim to create inerlingua for all major human languages
03	AnglaHindi (2003)	English to Hindi	General	-Pseudo interlingual rule-based -Uses all the modules of AnglaBharti -Makes use of an *abstracted example-base* for translating frequently encountered noun phrases and verb phrases -The translation accuracy is 90%.

5.4. Statistical Machine Translation Approach

The following table describes the summary of MT systems developed using Statistical MT approach and their salient features and/or limitations

Table 6. Summary of Statistical MT Approach

SN	MT System name	Languages	Domain/main application	Salient Features/Limitations
01	English to Indian Languages Machine Translation System (E-ILMT 2006)	English to Indian Languages	Tourism and Healthcare	-Rule-based and Statistical -The engine was developed using the statistical techniques and tools which includes the POS tagger (fnTBL), parser (Bikel), decoder (Pharaoh) etc -The objective was to 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 -The training corpus consisted of 5000 sentences and 800 sentences were split for testing and tuning -Pre-processing phase was included to take care of syntactic re-ordering on the source language to reduce long distance movements through SMT -The syntactically processed corpus was morphologically processed and used for training to tackle the problem of degradation in translation quality. -A rule based suffix separation approach was used to separate the root word and the affixes

5.5. Example-Based Machine Translation Approach

The following table describes the summary of MT systems developed using Example-Based MT approach and their salient features and/or limitations

Table 7. Summary of Example-based MT Approach

SN	MT System name	Languages	Domain/main application	Salient Features/Limitations
01	ANUBAAD (2000, 2004)	English to Bengali	News Headlines	-Example-base, Generalized Tagged example- base and Phrasal example-base are separately maintained - Bengali headline is generated after appropriate synthesis if the headline is found in Generalized Tagged example-base

				If the headline cannot be translated using Example-base, Generalized Tagged example-base or Phrasal example-base then the heuristic translation strategy is used
02	VAASAANUBAADA (2002)	Bengali to Assamese	News Text	-It includes pre-processing and post-processing tasks -The bilingual corpus is constructed and aligned manually -Longer sentences are fragmented at punctuations to obtain better quality translation
03	Shiva and Shakti machine translation system (2003)	English to {Hindi, Telugu, Marathi}	General	-Shiva is designed using an Example-based and Shakti is designed using combination of rule based and statistical approaches -Easy to extend this system for new target language -The rules used for target language generation are mostly linguistics in nature -Semantic information is also used by some modules in the system. -Currently system is working for three languages
04	ANGLABHARTI-II (2004)	English to Indian languages	General	-Uses a generalized example-base (GEB) approach along with a Raw Example-Base *(REB)* -Provides provisions for automated pre-editing & paraphrasing, generalized and conditional multi-word expressions as well as recognition of named-entities -Contains the modules for an error-analysis, statistical language-model for automated post-editing and failure analysis module
05	Hinglish machine translation system (2004)	Hindi to English	General	-Based on AnubBarti-II and AnglaBharti-II -Produced satisfactory acceptable results in more than 90% of the cases -Performs very shallow grammatical analysis -The system is not capable of resolving the meaning of polysemous verbs
06	English to {Hindi, Kannada, Tamil} and Kannada to Tamil Language-Pair Example Based Machine Translation (2006)	English to {Hindi, Kannada, Tamil} and Kannada to Tamil	General	-Maintains a bilingual sentence dictionary, phrases dictionary, words dictionary and phonetic dictionary -Example-Base of 75000 most commonly spoken sentences that are originally available in English -Sentences in the Example-Base have been manually translated into three Indian languages namely

				Hindi, Kannada and Tamil.
07	The MATREX (Machine Translation using Example) System (2008) The DCU Machine Translation System for ICON 2008	English to Hindi	Conference papers	-Makes use of marker-based chunking, which is based on the Marker Hypothesis, a psycholinguistic constraint which signifies context -A set of "marker" words, such as determiners, conjunctions, prepositions, possessive and personal pronouns is used to split sentences into chunks during pre-processing -In order to align the chunks the system makes use of an "edit-distance style" dynamic programming alignment algorithm

5.6. Hybrid Machine Translation Approach

The following table describes the summary of MT systems developed using Hybrid MT approach and their salient features and/or limitations

Table 8. Summary of Hybrid MT Approach

SN	MT System name	Languages	Domain/main application	Salient Features/Limitations
01	Anubharti Technology (1995)	Hindi to IL	General	- Hybrid Example-based - Combination of example-based, corpus-based approaches and some elementary grammatical analysis - Reduced the requirement of a large example-base -The generalization of the example-base is dependent on the target language.
02	ANUBHARTI-II (2004)	Hindi to IL	General	-Generalized Example-Base(GEB) along with Raw Example-Base(REB) -Emulates human-learning process for storing knowledge from past experiences to use it in future -S*hallow chunker* is used to fragment the input sentence into small units and then they are matched with a hierarchical example-base
03	Bengali to Hindi Machine Translation System (2009)	Bengali to Hindi	General	-Multi-engine Machine Translation approach -Uses an integration of SMT with a lexical transfer based system (RBMT) -The BLEU scores of SMT and lexical transfer based system when evaluated separately are 0.1745 and .0424 respectively -The BLEU score of hybrid system is

				better and it is 0.2275
04	Lattice Based Lexical Transfer in Bengali Hindi Machine Translation Framework	Bengali to Hindi	General	-Lattice based integrated with transfer based -The lattice based lexical translation system has been integrated with transfer based -Uses a lattice-based data structure i.e. a weighted directed acyclic graph with one start node and one end node. -BLEU score of proposed system is better than baseline system -It is tested for 500k Hindi sentences

6. COMPARISON OF MT APPROACHES

The following table shows the comparison of major MT approaches

Table 9. Comparison of MT Approaches

MT approach	Advantages	Disadvantages
Rule-based	1. Easy to build an initial system 2. Based on linguistic theories 3. Effective for core phenomena 4. Better choice for domain specific translation 5. The quality of translation is good for domain specific systems	1. Rules are formulated by experts 2. Difficult to maintain and extend 3. Ineffective for managerial phenomena 4. The number of rules will grow drastically in case of general translation systems
Knowledge-based	1. Based on taxonomy of knowledge 2. Contains an inference engine 3. Interlingual representation	1. Hard to build knowledge hierarchy 2. Hard to define 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 more 3. Knowledge acquisition problem still persists
Statistical	1. Does not consider language grammar for translation 2. Extracts knowledge from corpus 3. Reduces the human errors 4. Model is mathematically grounded	1. No linguistic background 2. Search cost is expensive 3. Hard to capture long distance phenomena 4. Require huge amount of parallel corpora 5. The translation quality will be very coarse due to lack of sufficient corpora

7. CONCLUSIONS

This paper described MT techniques in a longitudinal and latitudinal way with an emphasis on the MT development for Indian languages as well as non-Indian languages. From the study, we found that almost all existing Indian language MT systems are based on rule-based, hybrid and statistical approaches. We identified the following reasons to justify as to why most of the

developed MT systems for Indian languages have followed the rule-based, hybrid and statistical approach.

- Most of the Indian and non-Indian MT systems developed so far are developed for specific domains such as tourism, health care, children stories, medical, news headlines, technical documents, government circulars and notifications etc. They have used rule based approach as it provides better performance and accuracy if the set of rules is under control.
- The Indian languages are morphologically rich in features and agglutinative in nature, hence rule-based approaches may fail in situations where full-fledged general purpose MT systems are to be developed because the number of rules would be very high.
- Support of linguistic experts is essential for developing rule-based MT systems hence many researchers are now working on statistical and hybrid approaches.
- Statistical and Example-based MT systems require huge bilingual parallel corpora but even monolingual corpus in not available for many Indian languages including Marathi.
- Not much of the work is done for developing English to Marathi translation system. The only MT system available for English to Marathi is the β-version of Anuvadak developed and made available by TDIL (Technology Development for Indian Languages), New Delhi.
- Most of the Indian MT systems requires post editing for producing better results

REFERENCES

[1] 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 (Online) | ISSN : 2229-4333 (Print)

[2] 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 Science

[3] John Hutchins, (2005) "Current commercial machine translation systems and computer-based translation tools: system types and their uses", *International Journal of Translation* vol.17, no.1-2, pp.5-38.

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

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

[6] 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 PUBLISHER

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

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

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

[10] 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.

[11] 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

[12] 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)*

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

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

[15] 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.

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

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

[18] 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.*

[19] 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.

[20] 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.

[21] 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 June 2011.

[22] 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.

[23] hutchinsweb.me.uk/Nutshell-2005.pdf

[24] John Hutchins "Historical survey of machine translation in Eastern and Central Europe", *Based on an unpublished presentation at the conference on Crosslingual Language Technology in service of an integrated multilingual Europe*, 4-5 May 2012, Hamburg, Germany. (www.hutchinsweb.me.uk/Hamburg-2012.pdf)

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

[26] 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 Mumbai, pp. 22-25

[27] Akshar Bharti, Chaitanya Vineet, Amba P. Kulkarni & Rajiv Sangal, (2001) "ANUSAARAKA: overcoming the language barrier in India", *published in Anuvad: approaches to Translation*

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

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

[30] 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

[31] 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-survey/survey.pdf

[32] 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, 202.141.152.9/clir/papers/matra_mspil06.pdf

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

[34] 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.

[35] 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.

[36] Yanjun Ma, John Tinsley, Hany Hassan, Jinhua Du & Andy Way, (2008) "Exploiting Alignment Techniques in MATREX: the DCU Machine Translation System for IWSLT 2008', *in proceedings of IWSLT 2008*, Hawaii, USA

[37] projects.uptuwatch.com/cs-it/anubharti-an-hybrid-example-based-approach-for-machine-aided-trapnslation/

[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, No. 1, pp. 47-78.

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

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

[42] language.worldofcomputing.net

[43] sampark.iiit.ac.in

[44] www.cdacmumbai.in/xlit

[45] www.cdacmumbai.in/rupantar

[46] translationjournal.net/journal/29computers.htm

[47] www.cfilt.iitb.ac.in/resources/surveys/MT-Literature%20Survey-2012-Somya.pdf

[48] www.cdacmumbai.in/e-ilmt

[49] www.iiit.net/ltrc/Anusaaraka/anu_home.html

[50] cdac.in/html/aai/mantra.asp

[51] translate.google.com/about/intl/en_ALL/

Authors

G V Garje (gvg_comp@pvgcoet.ac.in) has completed ME in Computer Science and Engineering from NITTR, Chandigarh, India in 1998. Currently he is working as Associate Professor and Head of Computer Engineering and Information Technology Department at Pune Vidyarthi Griha's College of Engineering and Technology, Pune. Presently, he is a Chairman, Board of Studies in Information Technology, University of Pune and formerly, chairman, Board of Studies in Computer Engineering, University of Pune. Presently he is pursuing his Ph.D. from University of Pune, Maharashtra, India, in the area of Machine Translation. His area of research are NLP, Machine Translation specifically English-Marathi Language Pair. He has developed a tool for translating simple English interrogative sentences to Marathi sentences funded by University of Pune. His areas of interest are Data Structures, Operating Systems and Software Architecture.

G K Kharate (gkkharate@rediffmail.com) has completed his Ph.D. in Electronics and Telecommunication Engineering from University of Pune and ME Electronics and Communication from Walchand College of Engineering, Sangali, Maharashtra. Currently he is a Principal at Matoshri College of Engineering and Research Centre, Nashik, Maharshtra. He is a Dean, Faculty of Engineering and Member of Man agement Council, University of Pune. He is former Chairman, Board of Studies in Electronics Engineering, University of Pune. His areas of research are Image Processing, Pattern Recognition, and Artificial Intelligence. His areas of interest are Digital Electronics, Computer Networks, Image Processing and Natural Language Processing.