Machine Translation Systems for Indian Languages

Latha R. Nair
School of Engineering Cochin University of Science and Technology

David Peter S.
School of Engineering
Cochin University of
Science and Technology

ABSTRACT
Machine processing of Natural (Human) Languages has a long tradition, benefiting from decades of manual and semi-automatic analysis by linguists, sociologists, psychologists and computer scientists among others. This cumulative effort has seen fruit in recent years in the form of publicly available online resources ranging from dictionaries to complete machine translation systems. Machine translation is the name for computerized methods that automate all or part of the process of translating from one language to another. In a large multilingual society like India, there is great demand for translation of documents from one language to another language. Though work in the area of machine translation has been going on for several decades, efficient methods for machine translation continue to be a challenging task. A fully automatic high quality machine translation system is extremely difficult to build. This paper discusses the various approaches which have been applied in translation systems for Indian languages. Some of the important Indian language translation systems implemented with these techniques along with their capabilities and limitations are also discussed.

General Terms
Machine translation

Keywords
Machine Translation; Indian Languages; Transfer based translation; Interlingua.

1. INTRODUCTION
Machine translation is the name for computerized methods that automate all or part of the process of translating from one language to another. In a large multilingual society like India, there is great demand for translation of documents from one language to another language. There are 22 constitutionally approved languages, which are officially used in different states. There are about 1650 dialects spoken by different communities. There are 10 Indic scripts. All of these languages are well developed and rich in content. They have similar scripts and grammars. The alphabetic order is also similar. Some languages use common script, especially Devanagari. Hindi written in the Devanagri script is the official language of the union Government. English is also used for government notifications and communications. India's average literacy level is 65.4 percent (Census 2001). Less than 5 percent of people can either read or write English. As most of the state government works in provincial languages whereas the central government's official documents and reports are in English or Hindi, these documents are to be translated into the respective provincial languages to have an appropriate communication with the people. Moreover, over 95 percent of the population is normally deprived of the benefits of Information Technology due to language barrier [1]. All these make language translation a necessary one.

Work in the area of Machine Translation in India has been going on for several decades. During the early 90s, advanced research in the field of Artificial Intelligence and Computational Linguistics made a promising development of translation technology. This helped in the development of usable Machine Translation Systems in certain well-defined domains. Fully automatic high quality machine translation system (FGH_MT) is extremely difficult to build. In fact there is no system in the world which qualifies to be called FGH_MT. Many organizations like IIT(Kanpur), CDAC(Mumbai), CDAC(Pune), IIIT(Hyderabad) etc. are engaged in development of MT systems under projects sponsored by Department of Electronics (DoE), state governments etc. since 1990[2]. Research on MT systems between Indian and foreign languages and also between Indian languages are going on in these institutions. Translation between structurally similar languages like Hindi and Punjabi is easier than that between language pairs that have wide structural difference like Hindi and English., Translation systems between closely related languages are easier to develop since they have many parts of their grammars and vocabularies in common [3].

The following sections discuss the major projects in machine translation undergoing in India. The first section is about the development of language technology in India. Second section gives an overview of the approaches used in the MT systems developed in India. The other sections describe in detail each of these approaches with some important systems designed using these approaches.

2. DEVELOPMENT OF LANGUAGE TECHNOLOGY IN INDIA
In a large multilingual society like India where there is vast diversity of culture and languages, human communication is a major issue. As trade and businesses widen, people migrate to expand their business activities. In such a scenario, each human being is forced to learn more than one language in order to communicate with others. By providing linguistically cooperative environment, which facilitates smooth communication across different linguistic groups, Information Technology (IT) plays a very important role in reducing language barrier. The development of GIST in the early 1980's, (Graphics and Intelligence - Based Script Technology) was a major breakthrough. Development of GIST was started as a DoE sponsored project at the Indian Institute of Technology, Kanpur (IITK) and later the technology was further matured at the Centre for Development of Advanced Computing (CDAC), Pune. The Department of Electronics initiated a programme on "Electronics Tools for Indian Languages (ETIL)" in 1987. This programme focussed on

development of electronic tools for Indian language processing, their innovative application areas, technology assessment and integration of research findings into deployable systems.

During the VIII Plan of the Government of India, it was decided to give more thrust to language technology for Indian Languages. Accordingly, it was decided to initiate a major programme which will have focus on quality, national relevance and participation of traditional knowledge and consolidate R &D efforts in the area of Information Processing in Indian Languages. Towards this, the Department of Electronics of Government of India, initiated a National level programme during the year 1990-91 on "Technology Development for Indian Languages (TDIL) ".

As a first step, major thrust areas were identified which include: Development of Machine Readable Corpora of Texts of various Languages, Machine Aided Translation among various languages, Human-Machine Interface Systems, Computer Assisted Language Learning/Teaching and Fundamentals of Natural Language Processing. The various programmes developed for promoting use of IT in Indian languages have helped to generate expert manpower in wide areas to offer solutions in indian languages. As part of these programmes, work in the area of some of the foreign languages has also been carried out to provide solutions supporting those languages on personal computers. Research in language computing is on-going in many areas such as:

- machine translation;
- speech processing;
- optical character recognition (OCR);
- standards (character representation, fonts display, etc);
- localization;
- applications like word processors, e-mail clients etc;
- search, information extraction and retrieval;

search engines.

3.MACHINE TRANSLATION IN INDIA

Development of a Machine translation (MT) system requires very close collaboration among linguists, professional translators and computer engineers. In the development process, there are two major goals: (a) accuracy of translation and (b) speed. Accuracy-wise, smart tools for handling transfer grammar and translation standards including equivalent words, expressions, phrases and styles in the target language are to be developed. The grammar should be optimized with a view to obtaining a single correct parse and hence a single translated output. Speed-wise, innovative use of corpus analysis, efficient parsing algorithm, design of efficient Data Structure and run-time frequency-based rearrangement of the grammar which substantially reduces the parsing and generation.

A fully automatic Machine translation system should have different modules such as Morphological analyzer, Part of speech tagger, Chunker, Named entity recognizer, Word sense disambiguator, Syntactic transfer module and Target word generator[4]. The different techniques used for translation differs in the number of modules used and also the way these modules are implemented. Both rule based and statistical approaches have been tried in the implementation of each of these modules. The Machine translation systems developed for Indian languages are listed in Table 1. The various approaches used in the MT systems for Indian languages are discussed in the following sections.

3.1 Direct Machine Translation Systems

As the name suggests, these systems provide direct translation, without using any intermediate representation. This is done on a word by word translation using a bilingual dictionary usually followed by some syntactic arrangement. The steps involved are :

1. Identification of root words by removing suffixes from source language words.
2. dictionary look up to get the target language words/morphemes.
3. word order is changed to match with the target language. For English to Malayalam, this may be reordering of prepositions to postpositions and changing subject-verb-object to subject –object- verb structure.

A sample output from such a system for English to Hindi is given below:

Input(English): Rama played in the garden.

Output (Hindi) after word translation: Rama khela mein baag.

Output after syntactic rearrangement: Rama baag mein khela.

A direct translation system is appropriate for similar languages like Hindi and Punjabi[5,6]. Vishal Goyal and Gurpreet Singh Lehel of Punjab University have developed a web based Hindi to Punjabi MT system with 95% accuracy. Their system has additional modules for training the system for generating the lexicon using already existing corpus, input text font conversion into Unicode format to make the system free from specific font dependency, Hindi text normalization to handle spelling variations for the same word due to variation in dialects, replacement of collocations by keeping a lexicon for collocations, named entity recognition and replacement , word by word translation using bilingual dictionary and transliteration of unknown words. They also perform word sense disambiguation using a dictionary of ambiguous words. It uses a trigram approach with a sample corpus for word sense disambiguation.

A similar direct translation approach has been applied to Anusarak systems which translate between two closely related Indian languages using the principles of paninian grammar[7]. The anusaaraka project started at IIT Kanpur, by Prof. Rajeev Sangal and the research is continued at IIIT Hyderabad**.**

Table 1. Machine Translation systems for Indian Language

Method		Language Pair	Developer	Year
Direct		Anusaarak systems(Telugu,Kannada,Bengali,Marathi to Hindi)	Started in IIT Kanpur, continuing in IIIT, Hyderabad	1995
		Punjabi to Hindi MTS	Punjab University, Patiala	2007
		Web based Hindi to Punjabi MTS	Punjab University, Patiala	2010
Rule Based	Transfer based	Mantra-English to Hindi, Telugu,Gujarathi,Hindi-English,Bengali,Marathi	CDAC,Pune	1995
		Matra-English to Hindi MTS	CDAC,Pune	2004
		Shakti-English to Hindi,Marathi,Telugu MTS (combines rule based and statistical approach)	IIISc Bangalore and IIIT, Hyderabad	2004
		Anubaad-English-Bengali MTS(n-gram approach for pos tagging)	CDAC, Kolkata	2005
		English to Kannada MTS using UCSG	Unversity of Hyderabad	2006
		English – Malayalam MTS	Amrita Institute of Technology	2009
	Interlingua Based	Anglabharti, English-Hindi,Tamil MTS; uses pseudo interlinguaPLIL	IIT,Kanpur	1991
		AnglaHindi(combines example based approach and AnglaBharti approach)	IIT,Kanpur	1991
		English to Hindi MTS Using UNL as interlingua	IIT, Mumbai	2001
		Hindi-English MTS (Additional layer over AnglabhartiII)	IIT,Kanpur	2005
Corpus based	SMT	Tamil-Sinhala MTS	Carnegie-Mellon University, Pittsburgh, USA	2002
		English-Hindi MTS(combines linguistic knowledge)	IIIT, Hyderabad	2010
		English- Malayalam MTS	Cochin University ,Cochin	2010
		Hindi-English MTS	State University of New York.	2010
	EBMT	VAASAANUBAADA - Bengali-Assamese News Texts	Pondichery University	2002
		Anubharti; Hindi-English MTS;Combines syntax	IIT,Kanpur	2004
		Siva- English-Hindi	IISc Bangalore,IIIT Hyderabad	2004
		English-Sanskrit	I.T, Banaras Hindu University	2008

A fully automatic Machine translation system should have different modules such as Morphological analyzer, Part of speech tagger, Chunker, Named entity recognizer, Word sense disambiguator, Syntactic transfer module and Target word generator[4]. The different techniques used for translation differs in the number of modules used and also the way these modules are implemented. Both rule based and statistical approaches have been tried in the implementation of each of these modules. The Machine translation systems developed for Indian languages are listed in Table 1. The various approaches used in the MT systems for Indian languages are discussed in the following sections. Direct Machine Translation Systems

3.2Rule Based Translation

Rule based MT systems parse the source text and produce an intermediate representation, which may be a parse tree or some abstract representation. The target language text is generated from the intermediate representation. These systems rely on specification of rules for morphology, syntax, lexical selection and transfer, semantic analysis and generation and hence are called rule based systems. Depending on the intermediate representation used, these systems are further categorized as Transfer based machine translation and Interlingua based machine translation.

3.2.1Transfer based machine translation

This method needs parsing of input text to get the structure of the input sentence. It has three modules: analysis module, transfer module and generation module[9,10]. The analysis module produces source language structure. The language grammar rules can be used to generate the hierarchical syntax tree for the source language sentence. A set of hierarchical rules for forming the syntax tree for English sentences are given below:

S-> NP VP | VP

NP-> N| NP PP| Det N

VP ->V| VP PP| VP NP

The transfer module transfers the source language structure representation to a target language representation. This modules needs the subtree rearrangement rules by which the source language sentence syntax tree can be transformed into target language sentence syntax tree.

A set of transfer rules for a English to Hindi MT system are:

English structure	Hindi structure
VP->V NP	VP->NP V
PP-> P NP	PP-> NP P
VG->ADV V	VG->V ADV

The generation module generates target language text using target language structure. Syntactic structure transfer for verbs from Tamil to Hindi is discussed in [5]. It involves lexical transfer of verbs, transfer of auxiliary verb for tense, aspect and mood and transfer of gender, number and person information. Syntactic and lexical ambiguities are better resolved in this approach than in direct translation approach. An example for structure transfer from English to Hindi is shown below:

Input(English): Ram sat on a chair.

Output(Hindi): Ram ne ek kursi mein baita.

We have used two transfer rules. One is to switch the preposition and noun in the PP chunk and the other is placing the verb in the end of the sentence. The structural transfer and Hindi generation are shown in Fig.1. Transfer systems are more realistic, flexible and adaptable in meeting the needs of different levels and depths of syntactic and semantic analysis.

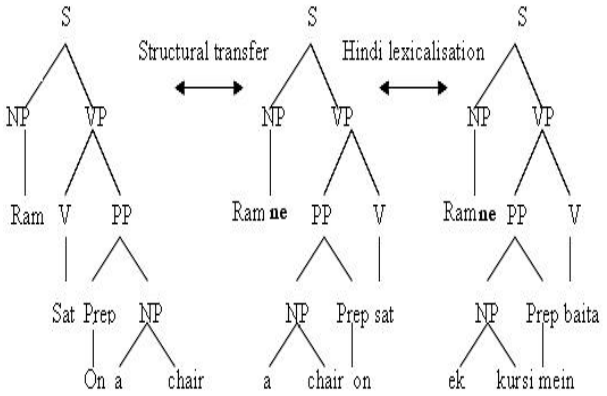

Figure 1. Structural transfer in a English-Hindi MTS

The English to Hindi MT system Mantra, developed by Applied Artifcial Intelligence (AAI) group of CDAC, Bangalore, in 1999 uses transfer based approach. The system translates domain specific documents in the field of personal administration; specifically gazette notifications, office orders, office memorandums and circulars. It is based on lexicalized tree adjoining grammar(LTAG) to represent English and Hindi grammar which are used to parse source English sentences and for structural transfer from English to Hindi[11,12,13]. This system also works well on other language pairs such as English-Bengali, English_Telugu, English_Gujarati and Hindi_English and also between Indian language pairs such as Hindi-Bengali and Hindi-Punjabi. The Mantra approach is general but the lexicon and grammar have been limited to the specific domain of personal Administration. The salient features of Mantra are : i) format retention ii) input can be in .rtf or .htm file or output of a speech recognition program or an optical character recognition package. It uses preprocessing tools like phrase marker, named entity recognizer, spell and grammatical checker. It uses Earley's style bottom up parsing algorithm for parsing . The system provides online addition of grammar rule. The system produces multiple translation results in the case of multiple correct parses. Mantra is one of CDAC's major achievements.

An English to Kannada MT system is developed at Resource centre for Indian Language Technology Solutions (RC_ILTS), University of Hyderabad by Dr. K. Narayan Murthy[14]. This also uses a transfer based approach and it can be applied to the domain of government circulars. The project is funded by Karnataka government. This system uses Universal Clause Structure Grammar (UCSG) formalism[15]. The technique is applied to English_ Telugu translation as well.

Other systems using this approach are : Matra- English to Hindi MTS developed by CDAC, Pune, Sakti- English to Marathi, Hindi and Telugu developed by IISc Bangalore and IIIT, Hyderabad, Anubaad- English to Bengali developed by CDAC, Kolkata, English to Malayalam MTS developed by Amrita Institute of Technology.

3.2.2 Interlingua based machine translation

The interlingua approach is based on Chomsky's claim that regardless of varying surface syntactic structures, languages share a common deep structure[16]. In this approach translation is a two step process: analysis and synthesis. During analysis, the source language text is converted into a language independent meaning representation called interlingua. In synthesis phase the interlingual representation is translated into any target language. Thus it can be used for multilingual translation. The amount of analysis needed in interlingual approach is more than that in a transfer based approach. This requires semantic analysis and the representation can be used for information retrieval.

For the sample sentence "John makes tools", the interlingua generatedis,

(Cat verb

Root make

Tense present

Subject root John

Cat noun-proper

Object root tool

Cat noun

Number plural)

The major difficulty in using this approach is in defining a universal interlingua which preserves the meaning of a sentence.

Anglabharti is an MT system for translation from English to Indian languages which uses pseudo interlingua approach. It was developed in 1991 by Prof. R.M.KSinha and team at IIT Kanpur. The system analyses English sentences and creates an intermediate structure called PLIL(Pseudo Lingua for Indian Languages). It performs most of the disambiguation. The effort required for analysis phase is70% and the generation phase takes 30%. So with an additional effort of 30% a new translator for an Indian language could be built. A context free grammar like structure is used to create the PLIL structure. It also uses statistical analysis of a corpus to identify the movement rules for the PLIL structure. Its beta_version is Angla Hindi for English to Hindi translation and is available at http://anglahindi.iitk.ac.in.[17].

The World Wide Web contents are mostly in English and cannot be accessed without proficiency in this language. The universal networking language has been proposed by the United Nations University for overcoming the language barrier. An English to Hindi MT system which uses Universal Natural Language(UNL) as the interlingua has been developed by Pushpak Bhattacharya and team at IIT Bombay. Their system has an English analyzer which converts the sentence into UNL form which is then given to a Hindi generator which generates the target sentence in Hindi. 95% of the UNL expressions were correctly converted to Hindi. Their system does part of speech disambiguation and some sense disambiguation for postposition markers and wh_pronouns. The system handles language divergence in a better way[16]. Currently work is going on for MT system for English to Marati and Bengali.

3.3 Corpus Based Machine Translation

Corpus based MT systems have gained much interest in recent years. The advantage of these systems are that they are fully automatic and require less human labour than rule based approaches. The disadvantage is that they need sentence aligned parallel text for each language pair and this method can not be employed where these corpora are not available. Corpus based sytems are classified into statistical machine translation(SMT) and Example based Machine Translation(EBMT).

3.3.1 Statistical Machine Translation

In this the input is considered as a distorted version of the target language sentence and the task is to find the most likely source language sentence giving the translation.

The task involves three steps:

1. Estimating the language probability P(t)
2. Estimating the translational model probability p(s/t)
3. Devising an efficient search for the target text that maximizes their product.

We have to find the sentence T for which p(s,t) is maximum.

$$P(s,t) = \arg\max_{t} p(s,t) = \arg\max_{t} p(t)p(s/t)$$

In the above model 's' is the source language sentence and 't' is the target language sentence. The probabilities are to be calculated from the parallel corpus. Smoothing techniques are required for handling data sparsity problem that occurs in any noisy channel model.

A phrase-based Hindi-English translation system was tried by Kamal Kuzhinjedathu and Shravya Shetty at Department of computer science, State University of New York. The translation model was generated using a Hindi-English parallel corpus. Since the parallel corpus is only sentence aligned the freely available online tool called GIZA++ was used to perform word alignment. The alignment produced were then processed to create the phrase based translation model. Two sets of parallel corpuses were used: EMILLE(Enabling minority language engineering) corpus distributed by the European Language Resources Association and the Hindi and English bibles from www.Hindibible.org. Bilingual dictionaries available on the internet were used to to augment the statistical model. The SRI language model toolkit10 was used to obtain a statistical language model. After getting translation model and language model a phrase based decoder, Moses was used to translate the test sentences.

An English to Hindi MT system which combines RBMT and phrase-based SMT approach was developed at IIIT Hyderabad in 2010. Though SMT systems are able to handle local reorderings by themselves, in case of long-distance transformations they benefit considerably from external guidance (the RBMT system in this case). The system works

in two stages. In the first stage, the source analyzer performs extensive linguistic analysis by running Brill's POS tagger and the Stanford dependency parser on the input sentence. It then converts the source into a chunk-based unordered dependency tree. In the next stage, the Transfer Grammar performs local and long-distance reorderings. By chunking the source sentences and converting them into a dependency structure, the RBMT system separates local (intra-chunk) reordering decisions from global (inter-chunk) reorderings. This allows for separate specifications of local and long-distance rules; thus, greatly reducing the number of rules that must be written into the grammar [19].

A SMT system for translating English into the Dravidian language, Malayalam has been tried in Cochin University of Science and Technology, Cochin, in 2010. By using a monolingual Malayalam corpus and a bilingual English/Malayalam corpus in the training phase, the machine automatically generates Malayalam translations of English sentences. The alignment model is improved by incorporating the parts of speech information into the bilingual corpus. Removing the insignificant alignments from the sentence pairs by this approach has ensured better training results. Pre-processing techniques like suffix separation from the Malayalam corpus and stop word elimination from the bilingual corpus are carried out for effective training. The structural difference between the English Malayalam pair is resolved by a decoder using the order conversion rules [20].

3.3.2 Example based Machine translation system (EBMT)

An Example based Machine translation system (EBMT) system maintains a corpus consisting of translation examples between source and target languages. An EBMT system has two modules: Retrieval module and an adaptation module. The retrieval module retrieves a similar sentence and its translation from the corpus for the given source sentence. The adaptation module then adapts the retrieved translation to get the final corrected translation.

Consider the English to Hindi translation for the following sentence.

"Rama sings a song"

The retrieval module retrieves the following sentence and its translation from the corpus from a list of approximately matching sentences. It uses some similarity measures based on word similarity or syntactic and semantic similarity to identify this set of approximately matching sentences. From these the system selects the sentence with closest match with the input sentence.

If the system selects "Rohit sings a song" and its translation "Rohit geet gaata hai" as the closest one, it replaces Rohit with Rama and gaata with gaathi and finally forms the translation.

"Rama geet gaati hai".

Here the adaptation is required to replace the word and suffix replacements. This method may not work in case of translation divergence where structurally similar sentences of the source language get translated into a different structure [21].

Anubharti, is an EBMT approach based MT system developed at IIT Kanpur by Prof. R.M.K Sinha and team. Along with basic EBMT it uses some grammatical analysis to reduce the size of the parallel corpus. This is done primarily by generalizing the constituents and replacing them with abstracted form achieved by identification of syntactic groups from the raw examples [22].

Vaasaanubaada is another system for translating bilingual Bengali-Assamese news texts using EBMT technique. The work involves machine translation of bilingual texts at sentence level. In addition, it also includes preprocessing and post-processing tasks. The work is unique because of the language pair that is chosen for experimentation. The bilingual corpus was constructed and aligned manually by feeding real examples using pseudo code. The longer input sentence is fragmented at punctuations, which resulted in high quality translation. Backtracking is used when an exact match is not found at the sentence/fragment level, leading to further fragmentation of the sentence. Since bilingual Bengali-Assamese languages belong to the Magadha Prakrit group, the grammatical form of sentences is very similar and has no lexical word groups. The system gives quality translation [23]. Siva, an English to Hindi MT system developed jointly by Carneige Mellon University, USA and IIIT Hyderabad is using EBMT approach. In addition to the hard coded linguistic rules it uses a statistical approach for learning new rules [3]. English to Sanskrit EBMT has been tried at Banaras Hindu University.

4 CONCLUSION

Various MT groups have used different formalisms best suited to their applications. Of them transfer based systems are more flexible and it can be extended to language pairs in a multilingual environment. Direct translation is appropriate for structurally similar languages. The interlingua based systems can be used for multilingual translation. The amount of analysis needed in interlingual approach is more than that in a transfer based approach. The universal networking language has been proposed as the interlingua by the United Nations University for overcoming the language barrier. Over the past decade data-driven approaches to machine translation have come to the fore of language processing research. The relative success in terms of robustness of Example Based and Statistical approaches have given rise to a new optimism and an exploration of other data-driven approaches such as Maximum Entropy language modeling. Performance of statistical techniques can be improved through large parallel corpus and usage of linguistic knowledge in the model. Hybrid systems are found to have better performance compared to the ones with the component technology. A number of Machine Translation systems between Indian and non-Indian languages have already been developed. Most of the MT systems are for Hindi and there are only very few systems for south Indian languages. More research has to be done in these areas to overcome the language barrier faced by India. The MT systems so far developed have many shortcomings in terms of rule set, dictionary, translation methodology and it is apparent from the survey that further work is needed in MT as a whole to produce intelligible translations.

5. REFERENCES

[1] Dubey P., Overcoming the Digital Divide through Machine Translation, Translation Journal, Volume 15, 2011, http://translationjournal.net/journal/55mt_india.htm [Dec 12, 2011].

[2] Murthy, B. K., Deshpande W. R., Language technology in India: past, present and future, 1998, http://www.cicc.or.jp/english/hyoujyunka/mlit3/7-12.html [Dec 11,2011]

[3] V Goyal, G S Lehal. "Advances in Machine Translation Systems". Language In India, Vol. 9, No. 11, 2009, pp. 138-150.

[4] Lalithadevi S., Pralayankar P and Kavitha V., Case Marking Pattern from Hindi to Tamil MT, 3rd National Conference on Recent Advances and Future Trends in IT (RAFIT),Punjabi University, Patiala, 2009, Punjab.

[5] Lalitha Devi S., Pralayankar P., et. al., Verb Transfer in a Tamil to Hindi Machine Translation System, International Conference of Asian Language Processing, 2010, Harbin, China, pp.261-264.

[6] G.S.Josan , G.S. Lehal. Evaluation of Hindi to Punjabi Machine Translation System. International Journal of Computer Science Issues, vol4 no1, 2009, pp 243-257.

[7] Goyal V., Lehal G.S., Web Based Hindi to Punjabi Machine Translation System, Journal of Emerging Technologies in Web Intelligence, Vol.2, 2010, pp.148-151.

[8] Bharati A., Chaitanya V., Kulkarni A.P., and Sangal R., Anusaaraka: Machine Translation in Stages, Vivek: A Quarterly in Artificial Intelligence, Vol. 10, No.3, 1997, pp. 22-25.

[9] GoutamKumar S., The EB-Anubad translator: A hybrid scheme, Journal of Zhejiang University SCIENCE 2005, ISSN 1009-3095, pp.1047-1050.

[10] Bandyopadhyay S. , ANUBAAD - The Translator from English to Indian Languages, VIIth State Science and Technology Congress, Calcutta, India, 2000.

[11] Bandyopadhyay S., State and Role of Machine Translation in India. Machine Translation Review, 2000, pp.11: 25.

[12] Bandyopadhyay S., Use of Machine Translation in India-current status. AAMT Journal, 36:, 2004, pp. 25-31.

[13] Bandyopadhyay S., Teaching MT: an Indian perspective, Sixth EAMT Workshop, UMIST, Manchester, England, 2002, pp.13-22.

[14] Dwivedi S.K. et. al., Machine translation systems in Indian perspective, Journal of computer science, pp.1082-1087,2010.

[15] Kumar G.B., Murthy K.N., UCSG Shallow Parser, Proceedings of CICLing-2006, Lecture Notes in Computer Science, Volume 3878 , pp 156-167, Springer-Verlag.

[16] Dave S., Parikh J. and Bhattacharyya P., Interlingua Based English Hindi Machine Translation and Language Divergence, Machine Translation, 2002, Volume 16, pp.251-304.

[17] R.M.K. Sinha et. al., ANGLABHARTI: A Multi-lingual Machine Aided Translation Project on Translation from English to Hindi, 1995 IEEE International Conference on Systems, Man nd Cybernetics, Vancouver, Canada, 1995, pp 1609-1614.

[18] Sinha R.M.K, Jain A., AnglaHindi: an English to Hindi machine-aided translation system, MT Summit IX, New Orleans, USA, 23-27 September 2003, pp.494-497.

[19] Ahsan A., Kolachina P., et. al., Coupling Statistical Machine Translation with Rule-based Transfer and Generation, AMTA- The Ninth Conference of the Association for Machine Translation in the Americas. Denver, Colorado,.2010.

[20] Sebastian M.P., Kurien S.K., Kumar G.S., English to Malayalam translation – a statistical approach, A2CWiC-10, First Amrita ACM –W celebration in women in Computing in India, 2010, article 64, pp.1-5, DOA 10.1145/1858378-1858442.

[21] Gupta D., N. Chatterjee., Identification of Divergence for English to Hindi EBMT ―Proceedings of MT SUMMIT IX, 2003, New Orleans, Louisiana, USA.

[22] Sinha R.M.K, An Engineering Perspective of Machine Translation: AnglaBharti-II and AnuBharti-II Architectures, Proceedings of International Symposium on Machine Translation, NLP and Translation Support System (iSTRANS- 2004), November 17-19, 2004.

[23] Vijayanand K., Choudhury S.I., P., Ratna P., VAASAANUBAADA - Automatic Machine Translation of Bilingual Bengali-Assamese News Texts, *Language Engineering Conference,2002, Hydrabad, India, pp.183-188.*