LANGUAGE IN INDIA
Strength for Today and Bright Hope for Tomorrow
Volume 9 : 11 November 2009
ISSN 1930-2940

Advances in Machine Translation Systems

Vishal Goyal, M.Tech.
Gurpreet Singh Lehal, Ph.D.

Advances in Machine Translation Systems

Vishal Goyal, M.Tech.
Gurpreet Singh Lehal, Ph.D.

Abstract

Machine translation system is software designed that essentially takes a text in one language (called the source language) and translates it into another language (called the target language). This paper presents the state of the art in the field of machine translation. First part of this paper discusses the machine translation systems for non-Indian languages and second part discusses the machine translation systems for Indian languages.

Keywords : Machine Translation Systems, Natural Language Processing, MT in India

1. Machine Translation Systems

1.1 Machine Translation System for non-Indian languages

Various machine translation (MT) systems have already been developed for most of the commonly used natural languages. This section briefly discusses some of the existing machine translation systems and the approaches that have been followed.

An English Japanese Machine Translation System (1982) developed by Makoto Nagao et al. The title sentences of scientific and engineering papers are analyzed by simple parsing strategies, and only eighteen fundamental sentential structures are obtained from ten thousand titles. Title sentences of physics and mathematics of some databases in English are translated into Japanese with their keywords, author names, journal names and so on by using fundamental structures. The translation accuracy for the specific areas of physics and mathematics from INSPEC database was about 93%.

RUSLAN (1985), a direct machine translation system between closely related languages Czech and Russian, by Hajic J, for thematic domain, the domain of operating systems of mainframes. The system used transfer based architecture. This project started in 1985 at Charles University, Prague in cooperation with Research Institute of Mathematical Machines in Prague. It was terminated in 1990 due to lack of funds.

The system was rule-based, implemented in Colmerauer’s Q-Systems. The system had a main dictionary of about 8,000 words, accompanied by transducing dictionary covering another 2000 words.

The typical steps followed in the system are Czech morphological analysis, syntactic-semantic analysis with respect to Russian sentence structure and morphological synthesis of Russian. Due to close language pair, a transfer-like translation scheme was adopted with many simplifications. Also many ambiguities are left unresolved due to the close relationship between Czech and Russian. No deep analysis of input sentences was performed.

The evaluations of results of RUSLAN showed that roughly 40% of the input sentences were translated correctly, about 40% of input sentences with minor errors correctable by human post-editor and about 20% of the input required substantial editing or re-translation.

There are two main factors that caused a deterioration of the translation. The first factor was the incompleteness of the main dictionary of the system and the second factor was the module of syntactic analysis of Czech. RUSLAN is a unidirectional system dealing with one pair of language, Czech to Russian.

PONS (1995), an experimental interlingua system for automatic translation of unrestricted text, constructed by Helge Dyvik, Department of Linguistics and Phonetics, University of Bergen. 'PONS' is an acronym in Norwegian for "Partiell Oversettelse mellom Nærstående Språk" (Partial Translation between Closely Related Languages).

PONS exploits the structural similarity between source and target language to make the shortcuts during the translation process. The system makes use of a lexicon and a set of syntactic rules. There is no morphological analysis. The lexicon consists of a list of entries for all word forms and a list of stem entries, or ‘lexemes’. The source text is divided into substrings at certain punctuation marks, and the strings are parsed by a bottom-up, unification-based active chart parser.

The system had been tested for the translation of sentence sets and simple texts between the closely related languages, Norwegian and Swedish, and between the more distantly related English and Norwegian. The developer concluded that in the case of the closely related languages, formally similar constructions will typically share stylistic properties.

CESILKO (2000), a machine translation system for closely related Slavic language pairs, developed by Hajic J, Hric J. K. and Ubon V. It has been fully implemented for Czech to Slovak, the pair of two most closely related Slavic languages.

The main aim of the system is localization of the texts and programs from one source language into a group of mutually related target languages.

In this system, no deep analysis had been performed and word-for-word translation using stochastic disambiguation of Czech word forms has been performed. The input text is passed through different modules namely morphological analyzer, morphological disambiguation, Domain related bilingual glossaries, general bilingual dictionary, and morphological synthesis of Slovak. The dictionary covers over 7, 00,000 items and it is able to recognize more than 15 million word-forms. The system is claimed to achieve about 90% match with the results of human translation, based on relatively large test sample. Work is in progress on translation for Czech-to-Polish language pairs.

Bulgarian-to-Polish Machine Translation system (2000), developed by S. Marinov. The system needs a grammar comparison before the actual translation begins so that the necessary pointers between similar rules are created and system is able to determine where it can take a shortcut. The system has three modes, where mode 1 and 2 enable system to use the source language constructions and without making a deeper semantic analysis to translate to the target language construction. Mode 3 is the escape hatch, when the Polish sentences have to be generated from the semantic representation of the Bulgarian sentence. This system is based on the approach followed by PONS discussed above.

interNOSTRUM (2000), a bidirectional Spanish-Catalan machine translation system, was developed by Marote R.C. et al. The system is available as an internet server and it is being used mainly to obtain draft translation of Spanish documents into Catalan and to browse through Catalan internet servers in Spanish. It is a classical indirect machine translation system using an advanced morphological transfer strategy. Currently, It translates ANSI, RTF, and HTML texts from Castillian Spanish to the central or Barcelona variety of Catalan and vice-versa.

interNOSTRUM has six modules: two analysis modules (morphological analyzer and part-of-speech tagger), two transfer modules (bilingual dictionary module and pattern processing module) and two generation modules (morphological generator and post-generator). The morphological analyzer uses morphological dictionary for source language, which contains lemmas, the inflectional paradigms and their mutual relationships. Bilingual dictionary has been used for translation and morphological analysis.

Antonio M. Corbí-Bellot et. al. (2005) developed the open source shallow-transfer machine translation (MT) engine for the Romance languages of Spain (the main ones being Spanish, Catalan and Galician).

The machine translation architecture uses finite-state transducers for lexical processing, hidden Markov models for part-of-speech tagging, and finite-state based chunking for structural transfer, and is largely based upon that of systems already developed by the Transducens group such as interNOSTRUM (Spanish—Catalan) and Traductor Universia (Spanish—Portuguese). The authors of this system claim that, for related languages such as Spanish, Catalan or Galician, a rudimentary word-for-word MT model may give an adequate translation for 75% of the text, the addition of homograph disambiguation, management of contiguous multi-word units, and local reordering and agreement rules may raise the fraction of adequately translated text above 90%.

Carme Armentano-oller et al (2005) extended the idea of A.M .Corbi-Bellot et. al. and developed an open source machine translation tool box which includes (a) the open-source engine itself, a modular shallow transfer machine translation engine suitable for related languages and largely based upon the systems such as interNOSTRUM and Traductor Universia, (b) extensive documentation (including document type declarations) specifying the XML format of all linguistic (dictionaries, rules) and document format management files, (c) compilers converting these data into the high speed (tens of thousands of words a second) format used by the engine, and (d) pilot linguistic data for Spanish—Catalan and Spanish—Galician and format management specifications for the HTML, RTF and plain text formats.

They use the XML format for linguistic data used by the system. They define five main types of formats for linguistic data i.e. dictionaries, tagger definition file, training corpora, structural transfer rule files and format management files.

Apertium (2005), developed by Carme Armentano-oller et. al is an open-source shallow-transfer machine translation (MT) system for the [European] Portuguese ↔ Spanish language pair. This platform was developed with funding from the Spanish government and the government of Catalonia at the University of Alicante. It is a free software and released under the terms of the GNU General Public License.

Apertium originated as one of the machine translation engines in the project OpenTrad and was originally designed to translate between closely related languages, although it has recently been expanded to treat more divergent language pairs (such as English–Catalan).

Apertium uses finite-state transducers for all lexical processing operations (morphological analysis and generation, lexical transfer), hidden Markov models for part-of-speech tagging, and multi-stage finite-state based chunking for structural transfer. For Portuguese–Spanish language pair, promising results are obtained with the pilot open-source linguistic data released (less than 10000 lemmas and less than 100 shallow transfer rules) which may easily improve (down to error rates around 5%, and even lower for specialized texts), mainly through lexical contributions from the linguistic communities involved.

Tatar (2001) A machine translation system between Turkish and Crimean, developed by Altintas K. et al., used finite state techniques for the translation process. It is in general disambiguated word for word translation. The system takes a Turkish sentence, analyses all the words morphologically, translates the grammatical and context dependent structures, translates the root words and finally morphologically generates the Crimean Tatar text. one-to-one translation of words is done using a bilingual dictionary between Turkish and Crimean Tatar. The system accuracy can be improved by making word sense disambiguation module more robust.

ga2gd (2006), a robust machine translation system, developed by Scannell K.P., between Irish and Scottish Gaelic despite the lack of full parsing technology or pre-existing bilingual lexical resources. It includes the modules Irish standardization, POS Tagging, stemming, chunking, WSD, Syntactic transfer, lexical transfer, and Scottish post processing. The accuracy has been reported to be 92.72%.

SisHiTra(2006), a hybrid machine translation system from Spanish to Catalan, developed by Gonzalez et. al. This project tried to combine knowledge-based and corpus-based techniques to produce a Spanish-to-Catalan machine translation system with no semantic constraints. Spanish and Catalan are languages belonging to the Romance language family and have a lot of characteristics in common. SisHiTra makes use of their similarities to simplify the translation process. A SisHiTra future perspective is the extension to other language pairs (Portuguese, French, Italian, etc.). The system is based on finite state machines. It has following modules: preprocessing modules, generation module, disambiguation module and post-processing module. The word error rate is claimed to be 12.5 for SisHiTra system.

Above discussions about various machine translation system conclude that direct approach is the obvious choice for machine translation system between closely related languages.

2.2 Machine Translation systems for Indian languages

This section will summarize the existing machine translation systems for Indian languages that are as follows:

SYSTRAN System (1968), Russian to English translation system, had been installed for use by United States Air Force (USAF). Large numbers of Russian scientific and technical documents were translated using SYSTRAN under the auspices of the USAF Foreign Technology Division (later the National Air and Space Intelligence Center) at Wright-Patterson Air Force Base, Ohio. The quality of the translations, although only approximate, was usually adequate for understanding content.

The Mark II systems (1974), developed by IBM and Washington University, for translating documents in Russian to English **,** installed at the USAF Foreign Technology Division, the Georgetown University System at the US Atomic Energy Authority and at Eurotom in Italy.

The Meteo System (1975-76), a very high quality machine translation system for weather bulletins that has been in operational use at Envinonnement Canada from 1982 to 2001. The first version of the system (METEO 1) went into operation on a Control Data 7600 supercomputer in March 1977. METEO 1 was formally adopted in 1981, replacing the junior translators in the workflow.

The quality, measured as the percentage of edit operations (inserting or deleting a word counts as 1, replacing as 2) on the MT results, reached 85% in 1985. METEO 2 went into operation in 1983. The software then ran in 48Kb of central memory with a 5Mb hard disk for paging. METEO 2 is believed to have been the first MT application to run on a microcomputer.

In 1996, John Chandioux developed a special version of his system (METEO 96) which was used to translate the weather forecasts (different kinds of bulletins) issued by the US Weather Service during the Atlanta Olympic Games. The latest known version of the system, METEO 5, dates from 1997 and ran on a standard IBM PC network under Windows NT. It translated 10 pages per second, while occupying so little space that it fitted on a 1.44Mb diskette.

ANGLABHARTI (1991), a machine-aided translation system specifically designed for translating English to Indian languages. English is a SVO language while Indian languages are SOV and are relatively of free word-order. Instead of designing translators for English to each Indian language, Anglabharti uses a pseudo-interlingua approach. It analyses English only once and creates an intermediate structure called PLIL (Pseudo Lingua for Indian Languages).

This is the basic translation process translating the English source language to PLIL with most of the disambiguation having been performed. The PLIL structure is then converted to each Indian language through a process of text-generation. The effort in analyzing the English sentences and translating into PLIL is estimated to be about 70% and the text-generation accounts for the rest of the 30%. Thus only with an additional 30% effort, a new English to Indian language translator can be built.

Some of the major design considerations in design of Anglabharti have been aimed at:

- providing a practical aid for translation wherein an attempt is made to get 90% of the task done by the machine and 10% left to the human post-editing;

- a system which could grow incrementally to handle more complex situations;

- an uniform mechanism by which translation from English to majority of Indian languages with attachment of appropriate text generator modules; and

- a human engineered man-machine interface to facilitate both its usage and augmentation.

Anglabharti is a pattern directed rule based system with context free grammar like structure for English (source language) which generates a `pseudo-target' (PLIL) applicable to a group of Indian languages (target languages). A set of rules obtained through corpus analysis is used to identify plausible constituents with respect to which movement rules for the PLIL is constructed. The idea of using PLIL is primarily to exploit structural similarity to obtain advantages similar to that of using interlingua approach. It also uses some example-base to identify noun and verb phrasals and resolve their ambiguities.

Anusaaraka (1995) project which started at IIT Kanpur, and is now being continued at IIIT Hyderabad, was started with the explicit aim of translation from one Indian language to another. It produces output which a reader can understand but is not exactly grammatical.

For example, a Bengali to Hindi Anusaaraka can take a Bengali text and produce output in Hindi which can be understood by the user but will not be grammatically perfect. Likewise, a person visiting a site in a language he does not know can run Anusaaraka and read the text. Anusaaraka's have been built from Telugu, Kannada, Bengali, and Marathi to Hindi.

The Mantra (MAchiNe assisted TRAnslation tool) (1999) translates English text into Hindi in a specified domain of personal administration, specifically gazette notifications, office orders, office memorandums and circulars. Initially, the Mantra system was started with the translation of administrative document such as appointment letters, notification, and circular issued in Central government from English to Hindi. The system is ready for use in its domains.

English – Hindi translation system (2002) with special reference to weather narration domain has been designed and developed by Lata Gore et. al.

VAASAANUBAADA (2002), an Automatic machine Translation of Bilingual Bengali-Assamese News Texts using Example-Based Machine Translation technique, developed by Kommaluri Vijayanand et. al.

ANGLABHARTI-II (2004) addressed many of the shortcomings of the earlier architecture. It uses a generalized example-base (GEB) for hybridization besides a raw example-base (REB). During the development phase, when it was found that the modification in the rule-base was difficult and might result in unpredictable results, the example-base is grown interactively by augmenting it.

At the time of actual usage, the system first attempts a match in REB and GEB before invoking the rule-base. In AnglaBharti-II, provision were made for automated pre-editing & paraphrasing, generalized & conditional multi-word expressions, recognition of named-entities and incorporated an error-analysis module and statistical language-model for automated post-editing.

The purpose of automatic pre-editing module is to transform/paraphrase the input sentence to a form which is more easily translatable. Automated pre-editing may even fragment an input sentence if the fragments are easily translatable and positioned in the final translation Such fragmentation may be triggered by in case of a failure of translation by the 'failure analysis' module. The failure analysis consists of heuristics on speculating what might have gone wrong. The entire system is pipelined with various sub-modules. All these have contributed significantly to greater accuracy and robustness to the system.

The Matra system (2004), developed by the Natural Language group of the Knowledge Based Computer Systems (KBCS) division at the National Centre for Software Technology (NCST), Mumbai (currently CDAC, Mumbai) and supported under the TDIL Project is a tool for human aided machine translation from English to Hindi for news stories.

It has a text categorization component at the front, which determines the type of news story (political, terrorism, economic, etc.) before operating on the given story. Depending on the type of news, it uses an appropriate dictionary.

It requires considerable human assistance in analyzing the input. Another novel component of the system is that given a complex English sentence, it breaks it up into simpler sentences, which are then analyzed and used to generate Hindi. They are using the translation system in a project on Cross Lingual Information Retrieval (CLIR) that enables a person to query the web for documents related to health issues in Hindi.

ANUBHARTI (2004) approach for machine-aided-translation is a hybridized example-based machine translation approach that is a combination of example-based, corpus-based approaches and some elementary grammatical analysis. The example-based approaches emulate human-learning process for storing knowledge from past experiences to use it in future. In Anubharti, the traditional EBMT approach has been modified to reduce the requirement of a large example-base. This is done primarily 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.

Shiva and Shakti (2004), Two machine translation systems from English to Hindi, developed jointly by Carnegie Mellon University USA, Indian Institute of Science, Bangalore, India, and International Institute of Information Technology, Hyderabad. Shakti machine translation system has been designed to produce machine translation systems for new languages rapidly. Shakti system combines rule-based approach with statistical approach whereas Shiva is example based machine translation system. The rules are mostly linguistic in nature, and the statistical approach tries to infer or use linguistic information. Some modules also use semantic information. Currently Shakti is working for three target languages (Hindi, Marathi and Telugu).

English-Telugu Machine Translation System is developed jointly at CALTS with IIIT, Hyderabad, Telugu University, Hyderabad and Osmania University, Hyderabad. This system uses

English-Telugu lexicon consisting of 42,000 words. A word form synthesizer for Telugu is developed and incorporated in the system.

Telugu-Tamil Machine Translation System is also being developed at CALTS. This system uses the Telugu Morphological analyzer and Tamil generator developed at CALTS. The backbone of the system is Telugu-Tamil dictionary.

English-Kannada Machine Aided Translation system is developed at Resource Centre for Indian Language Technology Solutions, University of Hyderabad by Dr. K. Narayana Murthy. Their approach is based on using the Universal Clause Structure Grammar (UCSG) formalism. This is essentially a transfer-based approach, and has been applied to the domain of government circulars, and funded by the Karnataka government.

ANUBAAD (2004), a hybrid MT system for translating English news headlines to Bengali, developed by Sivaji Bandyopadhyay at Jadavpur University Kolkata and. The current version of the system works at the sentence level.

Hinglish (2004), a machine translation system for pure (standard) Hindi to pure English forms developed by R. Mahesh K. Sinha and Anil Thakur. It had been implemented by incorporating additional layer to the existing English to Hindi translation (AnglaBharti-II) and Hindi to English translation (AnuBharti-II) systems developed by Sinha. The system claimed to be produced satisfactory acceptable results in more than 90% of the cases. Only in case of polysemous verbs, due to a very shallow grammatical analysis used in the process, the system is unable to resolve their meaning.

Tamil-Hindi, Machine-Aided Translation system developed by Prof. C.N. Krishnan at AU-KBC Research Centre, MIT Campus, Anna University Chennai. This system is based on Anusaaraka Machine Translation System. It uses a lexical level translation and has 80-85% coverage. Stand-alone, API, and Web-based on-line versions are developed. Tamil morphological analyser and Tamil-Hindi bilingual dictionary (~ 36k) are the byproducts of this system. They also developed a prototype of English - Tamil MAT system. It includes exhaustive syntactical analysis. At present it has limited Vocabulary (100-150) and small set of Transfer rules.

English-Hindi example based machine translation system, developed by IBM India Research Lab at New Delhi. Now, they have recently initiated work on statistical machine translation between English and Indian languages, building on IBM's existing work on statistical machine translation.

English to {Hindi, Kannada, Tamil} and Kannada to Tamil language-pair example based machine translation (2006) developed by Prashanth Balajapally. It is based on a bilingual dictionary comprising of sentence-dictionary, phrases-dictionary, words-dictionary and phonetic-dictionary is used for the machine translation. Each of the above dictionaries contains parallel corpora of sentence, phrases and words, and phonetic mappings of words in their respective files. Example Based Machine Translation (EBMT) has a set of 75000 most commonly spoken sentences that are originally available in English. These sentences have been manually translated into three of the target Indian languages, namely Hindi, Kannada and Tamil.

Punjabi to Hindi Machine translation System (2007) developed by Gurpreet Singh Josan et. al. at Punjabi University Patiala.This system is based on direct word-to-word translation approach. This system consists of modules like pre-processing, word-to-word translation using Punjabi-Hindi lexicon, morphological analysis, word sense disambiguation, transliteration and post processing. The system has reported 92.8% accuracy.

Sampark: Machine Translation System among Indian languages (2009), developed by the Consortium of institutions. Consortium of institutions include IIIT Hyderabad, University of Hyderabad, CDAC(Noida,Pune), Anna University, KBC, Chennai, IIT Kharagpur, IIT Kanpur, IISc Bangalore, IIIT Alahabad, Tamil University, Jadavpur University. Currently experimental systems have been released namely {Punjabi,Urdu, Tamil, Marathi} to Hindi and Tamil-Hindi Machine Translation systems.

Hindi to Punjabi Machine translation System (2009) developed by Vishal Goyal et. al. at Punjabi University Patiala. This system is based on direct word-to-word translation approach. This system consists of modules like pre-processing, word-to-word translation using Hindi-Punjabi lexicon, morphological analysis, word sense disambiguation, transliteration and post processing. The system has reported 95% accuracy.

Conclusion

As discussed in the above section, systems utilizing simpler approach like direct MT for translating between similar languages have built the general opinion that it is easier to create an MT system for a pair of related languages. It is concluded from the above references that lot of research is going in the area of NLP and number of machine translation systems has been developed and regular efforts are being done for its improvements. For closely related languages, direct approach is most suitable approach.

References

1. González J., A.L.Lagarda, J.R.Navarro, L.Eliodoro, A.Giménez, F.Casacuberta, J.M.de Val, & F.Fabregat, “SisHiTra: a Spanish-to-Catalan hybrid machine translation system:, LREC-2006: Fifth International Conference on Language Resources and Evaluation. 5th SALTMIL Workshop on Minority Languages: “Strategies for developing machine translation for minority languages”, Genoa, Italy, 23 May 2006; pp.69-73.
2. Makoto Nagao , Jun-ichi Tsujii , Koji Yada , Toshihiro Kakimoto, An English Japanese machine translation system of the titles of scientific and engineering papers, Proceedings of the 9th conference on Computational linguistics, p.245-252, July 05-10, 1982, Prague, Czechoslovakia
3. Hajic J., "Ruslan-An MT System between closely related languages", In Proceedings of the 3rd Conference of The European Chapter of the Association for Computational Linguistics, Copenhagen, Denmark, 1987, pp.113-117.
4. Dyvik, Helge 1995: Exploiting Structural Similarities in Machine Translation. Computers and the Humanities 28:225 - 234.

5. HAJIC J, HRIC J, KUBON V., "CESILKO– an MT system for closely related languages", In ACL2000, Tutorial Abstracts and Demonstration Notes, pp. 7-8. ACL, Washington.
6. S. Marinov, “Structural Similarities in MT A Bulgarian-Polish Case”, from website http://www.gslt.hum.gu.se/~svet/courses/mt/termp.pdf.
7. Marote R. C, Guillen E., Alenda A.G., Savall M.I.G., Bellver A.I., Buendia S.M., Rozas S.O., Pina H.P., Anton P.M.P., Forcada M.L., “The Spanish-Catalan machine translation system interNOSTRM”, In proceedings of MT Summit VIII, 18-22 Sept. 2001, Santiago de Compostela, Galicia, Spain.
8. A. M. Corbí-Bellot, Mikel L. Forcada, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Gema Ramírez-Sánchez, Felipe Sánchez-Martínez, Iñaki Alegria, Aingeru Mayor, Kepa Sarasola. (2005) “An open-source shallow-transfer machine translation engine for the Romance languages of Spain.”, In Proceedings of the Tenth Conference of the European Association for Machine Translation, May 30-31, Budapest, Hungary, pp 79-86.
9. Carme A., Rafael C., Antonio M., Mikel L., Mireia G., Sergio O., Juan A., Gema R., Felipe S., Miriam A., “Open-source Portuguese-Spanish machine translation.”, In Lecture Notes in Computer Science 3960 (Computational Processing of the Portuguese Language, Proceedings of the 7th International Workshop on Computational Processing of Written and Spoken Portuguese, PROPOR 2006), May 13-17, 2006, ME - RJ/Itatiaia, Rio de Janeiro, Brazil, p. 50-59.
10. K. Altintas; Turkish To Crimean Tatar Machine Translation System; 2001; Master Thesis; Department Of Computer Engineering And The Institute Of Engineering And Science, Bilkent University, Turkey.
11. Scannell K.P.,"Machine Translation for Closely Related language Pair", Proceedings of the Workshop on Strategies for developing machine translation for minority languages at LREC 2006, Genoa, Italy, May 2006, pp103-107.
12. R. M. K. Sinha, Jain R., Jain A., “Translation from English to Indian languages:ANGLABHARTI Approach”, In proceedings of Symposium on Translation Support System STRANS 2001, Feb 15-17, IIT Kanpur, India.
13. R.M.K. Sinha & A. Jain, “AnglaHindi: an English to Hindi machine-aided translation system”, MT Summit IX, New Orleans, USA, 23-27 September 2003; pp.494-497.
14. R. Mahesh K. Sinha & Anil Thakur, “Machine translation of bi-lingual Hindi-English (Hinglish) text”, MT Summit X, Phuket, Thailand, September 13-15, 2005, Conference Proceedings: the tenth Machine Translation Summit; pp.149-156.
15. R.M.K. Sinha, “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, Tata Mc Graw Hill, New Delhi.
16. Sivaji Bandyopadhyay. 2004. Use of Machine Translation in India. AAMT Journal, 36: 25-31.
17. R.M.K. Sinha and A. Jain. 2003. AnglaHindi: An English to Hindi Machine-Aided Translation System. In “Proceedings of MT SUMMIT IX”, New Orleans, Louisiana, USA.
18. D. Gupta and N. Chatterjee. 2003. Identification of Divergence for English to Hindi EBMT. In “Proceedings of MT SUMMIT IX”, New Orleans, Louisiana, USA.
19. Sivaji Bandyopadhyay. 2002. Teaching MT – An Indian Perspective. In “Proceedings of the 6th EAMT Workshop on Teaching Machine Translation”, Manchester, UK, 13-22.

20. Shachi Dave, Jignashu Parikh and Pushpak Bhattacharyya. 2001. Interlingua-based English-Hindi Machine Translation and Language Divergence. Journal of Machine Translation, 16 (4): 251-304.
21. Durgesh Rao. 2001. Machine Translation in India: A Brief Survey. In "Proceedings of SCALLA 2001 Conference", Banglaore, India.
22. Sivaji Bandyopadhyay. 2000. State and Role of Machine Translation in India. Machine Translation Review, 11: 25-27.
23. Bharati, Akshar, Vineet Chaitanya, Amba P Kulkarni, and Rajeev Sangal (1997), "Anusaaraka: Machine Translation in Stages", Vivek: A Quarterly in Artificial Intelligence, Vol. 10, No.3, pp. 22-25.
24. Bharati, Akshar, Chaitanya, Vineet, Kulkarni, Amba P., Sangal, Rajeev Anusaaraka: Machine Translation in stages. Vivek, A Quarterly in Artificial Intelligence, Vol. 10, No. 3 (July 1997), NCST, India, pp. 22-25.
25. Dave, Shachi, Parikh, Jignashu and Bhattacharyya, Pushpak Interlingua Based English Hindi Machine Translation and Language Divergence, Journal of Machine Translation, Volume 17, September, 2002.
26. Hutchins, W. John, Somers, Harold L. An Introduction to Machine Translation. Academic Press, London, 1992.
27. Murthy, B. K., Deshpande, W. R. 1998. Language technology in India: past, present and future. http://www.cicc.or.jp/english/hyoujyunka/mlit3/7-12.html
28. A. Bharati, R. Moona, P. Reddy, B. Sankar, D.M. Sharma, R. Sangal, Machine Translation: The Shakti Approach, Pre-Conference Tutorial at ICON-2003.
29. Kommaluri Vijayanand, Sirajul Islam Choudhury, Pranab Ratna, "VAASAANUBAADA - Automatic Machine Translation of Bilingual Bengali-Assamese News Texts", Language Engineering Conference, Hydrabad, India 2002.
30. Bandyopadhyay S. (2000), "ANUBAAD - The Translator from English to Indian Languages", In Proceedings of the VIIth State Science and Technology Congress, Calcutta, India.
31. Bandyopadhyay S. (2002), "Teaching MT: an Indian perspective", Sixth EAMT Workshop "Teaching machine translation", November 14-15, UMIST, Manchester, England. pp.13-22.
32. G. S. Josan and G. S. Lehal (2008), "A Punjabi to Hindi machine Translation System", Coling 2008: Companion volume: Posters and Demonstrations, Manchester, UK, pp. 157-160.
33. G Bharadwaja Kumar and Kavi Narayana Murthy, "UCSG Shallow Parser", Lecture Notes in Computer Science, Volume 3878 / 2006, pp 156-167 Springer-Verlag Proceedings of the Sixth International Conference - CICLing-2006 Conference on Intelligent Text Processing and Computational Linguistics, February 19-25, 2006, Mexico City, Mexico
34. R. M. K. Sinha, Jain R., Jain A., "Translation from English to Indian languages:ANGLABHARTI Approach", *In proceedings of Symposium on Translation Support System STRANS 2001, Feb 15-17, IIT Kanpur, India.*
35. Bandyopadhyay S. (2000), "ANUBAAD - The Translator from English to Indian Languages", *In Proceedings of the VIIth State Science and Technology Congress*, Calcutta, India.
36. Bandyopadhyay S. (2002), "Teaching MT: an Indian perspective", *Sixth EAMT Workshop "Teaching machine translation"*, November 14-15, UMIST, Manchester, England. pp.13-22.

37. R. Mahesh K. Sinha & Anil Thakur, "Machine translation of bi-lingual Hindi-English (Hinglish) text", *MT Summit X, Phuket, Thailand, September 13-15, 2005, Conference Proceedings: the tenth Machine Translation Summit*; pp.149-156.
38. Sobha. L, Arulmozhi. P. (2006). "Translingual Information Accessor Using Information Extraction - English to Tamil". presented in 34th All India conference for Dravidian Linguistics held at International School of Dravidian Linguistics, Trivandrum.
39. Sobha L, Pralayankar P,and Kavitha V. (2009)."Case Marking Pattern from Hindi to Tamil MT", In 3rd National Conference on Recent Advances and Future Trends in IT (RAFIT),Punjabi University, Patiala, Punjab.
40. Manish Shrivastava, Nitin Agrawal, Bibhuti Mohapatra Smriti Singh, Pushpak Bhattacharya, "Morphology Based Natural Language Processing tools for Indian Languages", *The 4th Annual Inter Research Institute Student Seminar in Computer Science*, Indian Institute of Technology, Kanpur April 1-2, 2005
41. Ananthakrishnan R., Pushpak B., Sasikumar M. and Ritesh M.S. (2007), "Some issues in automatic evaluation of English-Hindi MT: more blues for BLUE", *In Proceedings of 5th International conference on natural language processing (ICON-2007),* January 4-6, IIIT Hyderabad, pp 53-60
42. Kommaluri Vijayanand, Sirajul Islam Choudhury, Pranab Ratna, "VAASAANUBAADA - Automatic Machine Translation of Bilingual Bengali-Assamese News Texts", *Language Engineering Conference, Hydrabad, India 2002.*
43. Lata Gore and Nishigandha Patil, "Paper On English To Hindi - Translation System", *Proceedings of Symposium on Translation Support Systems STRANS-2002, IIT Kanpur, March. 15-17,2002*
44. Prashanth Balajapally, Phanindra Pydimarri, Madhavi Ganapathiraju, N. Balakrishnan, Raj Reddy, "Multilingual Book Reader: Transliteration, Word-to-Word Translation and Full-text Translation", *VALA 2006: 13th Biennial Conference and Exhibition Conference of Victorian Association for Library Automation*, Melbourne, Australia, February 8-10, 2006.

Vishal Goyal, M.Tech.
Department of Computer Science
Punjabi University
Patiala-147002
Punjab, India
vishal.pup@gmail.com

Gurpreet Singh Lehal, Ph.D.
Advanced Centre for Technical Development of Punjabi Language, Literature & Culture
Punjabi University
Patiala 147002
Punjab, India
gslehal@gmail.com