44

Language Technology Flash

Jan – Dec 2016

विश्वभारत @ tdil

Ministry of Electronics and Information Technology
Electronics Niketan, 6, CGO Complex, New Delhi – 110003
Telfax : 011-2436 3525 Email : tdilinfo@mit.gov.in Website : http//tdil.mit.gov.in

VishwaBharat@tdil
Jan - Dec 2016

From Editor's Pen...

Dear Readers,

Machine Translation technology is as old as Computer Science discipline. However, it has gained traction in recent years due to globalization. Essentially it has three sub-systems viz. Analysis, Transfer and Generation. The paradigms often used for this are: (i) rule-based machine translation (RBMT), (ii) example-based machine translation (EBMT) and statistical machine translation (SMT). Although translation requires human intelligence and knowledge of source and target language. Yet there are repeated components which can be handled by the machine which can be leveraged to get faster machine translation followed by post-editing which is more popularly known as Machine Assisted Translation (MAT). It reduces the task of translator to validation which makes the quality of machine translation acceptable. Assessing quality of such a translation is a complex subject. An attempt has been made to evolve a standard for this purpose so that technology can be utilized by localization Industry for providing machine assisted translation services.

The readers are welcome to provide feedback towards improvisation of this standard so that it fructifies as a National Standard.

Swaran Lata
slata@meity.gov.in

ISSN No. 0972-6454 has been granted to VishwaBharat@tdil. Google Search engine refers to the contents of the Journal

Website: www.tdil.mit.gov.in **Cover Design Shri Narendera Shrivastava**

Editorial Team

Swaran Lata	-	**slata@meity.gov.in**
Manoj Jain	-	**mjain@meity.gov.in**
Vijay Kumar	-	**vkumar@meity.gov.in**
Somnath Chandra	-	**schandra@meity.gov.in**
Bharat Gupta	-	**bharatg@meity.gov.in**
DC Verma	-	**dc.verma@nic.in**

Technology Development for Indian Languages Programme

www.tdil.mit.gov.in

www.tdil-dc.in

www.ildc.in

VishwaBharat@tdil
Jan - Dec 2016

Contents

Strategies for Development of Machine Translation Systems

Kavi Narayana Murthy
School of Computer and Information Sciences,
University of Hyderabad,
Hyderabad, 500046, India,
knmuh@yahoo.com

Abstract

In this paper we describe the state of the art in Machine Translation. We include a critical review of the challenges faced in MT and the possible reasons behind the failures. We give a number of suggestions for overcoming these limitations and challenges. We support our claims by describing several MT systems we are currently developing. Our approaches hold promise and raise hopes. We believe that the strategies suggested here are applicable to other Indian languages as well. We can develop usable MT system within a very short time and gradually improve the systems as we keep using them.

1. Introduction

Translation is a meaning preserving transformation from one language to another. Machine Translation (MT) or Automatic Translation deals with the design of computational models for translation between human languages. MT systems usually do not attempt to directly capture and preserve meanings, instead, they try to capture and transfer structure, in the hope that structure captures meaning. Thus a natural model for MT is to do analysis of source language (SL) text, and generate the corresponding target language (TL) text, preserving the structure. MT systems normally work sentence by sentence. Morphological analysis and generation to take care of word-internal structure and syntactic analysis and generation to take care of sentence-internal structure are thus natural considerations. A transfer module may be incorporated between the analysis and generation phases to take care of divergences between the two languages.

An alternative view that has become popular in recent times is the Statistical Machine Translation (SMT) view point [5]. Translation involves finding TL words corresponding to the SL words in a given sentence, and re-ordering the words if required to take care of syntactic divergences. SMT models these aspects in a probabilistic fashion. During the Training phase, these probabilities are estimated from a Training Corpus. Then the system can apply the learned models to translate given texts. A large and relatively high quality training corpus is essential for SMT. There is no need for any dictionary, morph analyzer or a parser, nor do we need to explicitly model any divergences.

Of late highly successful MT systems have been built using Deep Learning Neural Networks. These are similar to SMT systems in that they are trained on a large training data set and linguistics is not explicitly used.

In the next few sections, we shall review the state of the art in MT, identify the weak points and suggest strategies to overcome these weaknesses. We shall then mention our own work in MT where we have fruitfully applied some of these suggested strategies.

2. Machine Translation: The State of the Art

Machine Translation is at least 60 years old. Until about 1990, most of research and development effort in MT was in the Rule-Based approach. Included in the Rule-Based Approach are the Direct, Transfer and Interlingua approaches. In hind sight, it would not be wrong to say that MT in general has not been very successful, for any language pair, anywhere in the world. Truely successful systems, which are able to capture and preserve meanings and produce publishable quality translations, are very few and limited in domain and applicability.

End-user expectations are very high and it looks next to impossible to achieve the expected degree of perfection by fully automatic means. In the initial stages, it was felt that the translations produced by the machine can be post-edited by humans to produce high quality translations. However, unless the quality of outputs produced by machines are very good, humans will prefer to translate on their own rather than checking and correcting all the mistakes the machine has made. Psychologically, post-editing is not a very pleasant task [5]. As a result, MT systems rarely came to fully usable level. MT could only be used in those situations where rough translations suffice, and publishable quality translations are not required.

There are many reasons for this failure. Firstly, adequate linguistic resources are not available in many languages. Large, representative, and clean corpora, good dictionaries, good grammar books etc. are not easy to find in all languages of the world. A dictionary meant for human use is very different from the dictionary we need for automatic use by machines and many 'good' dictionaries may still not be good enough for use in MT. Similarly, many works on grammar are not sufficiently detailed and sufficiently precise for implementation on a computing framework. As a result, MT developers end up also doing a good deal of original lexicographic work and grammar discovery. Needless to say, these are hard tasks requiring decades of research work. Also, good linguists are not always available for consultation. More importantly, linguistics is not a finished science, we cannot expect ready-made answers to all the questions we get while developing MT systems. There are many competing theories, theories are constantly undergoing developments and refinements, many problems of critical importance in MT have still not been solved fully and satisfactorily.

Divergences between different languages or language families have not been fully worked out, making the transfer stage a big challenge. MT developers start with great hopes, assuming that languages are rule-governed and the rules can be easily discovered. They encounter harder and harder problems down the line, which can be quite frustrating. A point of saturation is reached and further improvements become very difficult.

Statistical approaches to MT, on the other hand, require large, high quality parallel corpora. Such corpora are available for some language pairs in the world today - English-French, English-German, English-Chinese, for example [4]. A lot of progress has been made in SMT taking such language pairs for study. However, for languages where such large scale parallel corpora are not yet available, these theoretical results are only of academic curiosity, not useful in actually developing usable MT systems. The same observations are also valid for the Deep Learning approaches.

Unless the quality of translations produced by a machine is very good, the outputs cannot be post-edited to produce large scale, high quality parallel corpora. Thus we are left in a chicken-and-egg kind of a situation.

3. MT in India

Interest in MT started in India as early as 1986. [6] MAT [8], MATRA and MANTRA are some of the noteworthy MT systems developed during the 1990s. [3] During the year 1990-91 DIT (Department of IT), Govt. of India initiated the TDIL (Technology Development for Indian Languages) programme. [3] Several major projects have been funded by TDIL since then in MT and related areas. IL-IL-MT, E-IL-MT, Anglabharti-E-IL MT [1] are some of the major projects funded by TDIL in recent times in consortium mode. All these projects generally use rule-based approaches. Shatanuvadak [5] by IIT-Mumbai in 2014 is a notable deviation, it uses SMT in a big way for Indian languages.

Third party evaluations have shown that none of these MT systems have reached a level of performance adequate for deployment and large scale regular use. People are hesitating to come forward to post-edit the translations produced by these MT systems. We do not hear of any big success story.

It appears that everybody is interested in cooking and nobody wants to eat what is cooked. In the next section, we propose a set of ideas to overcome this unpleasant situation.

4. Strategies for Development of MT systems

In this section we analyze the main reasons for not being able to reach the final goal of high quality translations, and our own suggestions for overcoming these challenges. In the next section, we shall describe 'The saara Translator', a set of MT systems being developed by us, that actually try to put these ideas into practice.

Two main tasks an MT system has to do is lexical substitution and re-ordering. SL words need to be substituted with equivalent TL words. We may then need to reorder the words as dictated by the syntactic divergences between the SL and the TL.

Since parallel corpora are not available for many language pairs of interest, we shall restrict our attention to rule-based approaches here. A common belief is that morphology is absolutely essential, especially in Indian languages, which are considered to exhibit rich morphology. That is, a single root word can give rise to a very large number of word forms, through processes such as inflection, derivation, sandhi and compounding. However, developing high performance morph systems has not been easy. Can we bypass morphology?

Generally speaking, it is not an intelligent decision to throw away morphology and try to keep all forms of all words directly in a dictionary. Morphology has its due share of relevance and importance. However, developing proper systems for morphological analysis and generation is a hard task, requiring many years of labour. There are theoretical problems, there is the challenge of finding good linguists, even good books on grammar may not be available, and practical experience shows that approaches to MT which critically depend on morphological analysis, transfer and generation have not worked very well in India. Therefore, while development of complete morphology in a computational framework should remain an important goal, we should keep that on the back burner as a long term strategy and think of developing MT systems that bypass morphology. Our own experiments show that equal or higher performance in translation can be obtained much faster by bypassing morphology. MT systems can be built within a matter of months, taking advantage of ordinary people who are bilinguals, instead of waiting for linguists or try on our own to develop computational systems of morphology. We find that bypassing morphology actually works even for Dravidian languages, which are considered to exhibit exceptional levels of morphological complexity.

Of course we must take good advantage of a system for morphological analysis and generation if we already have one. We can build hybrid MT systems which combine the best of rule-based and statistical approaches. Performance of a hybrid MT system will improve not only with larger and better training corpora that may become available over time, but also as morphology improves over time.

In the simplest statistical MT model, we start with the assumption that all lexical substitutions are equally likely. To give an example, any word in English can map to any word in Hindi. The English word ’table’ may mean ’maa’, ’khaaya’, ’us’, ’mej’, ’idhar’, ’kutta’ or any other word in Hindi, all of these are equally probable. SMT systems then try to adjust these probabilities based on a large training corpus of English-Hindi sentence pairs. In other words, pure SMT systems do not take any advantage of the linguistic knowledge we may have. Pure SMT systems do not use dictionaries, morph, syntax or any other aspect of language and linguistics, even if we have access to such knowledge. That is why they need a very large training corpus. Instead of waiting for a large training corpus of parallel sentences, we can get started off if we make good use of available resources such as bilingual dictionaries. Lexical substitution possibilities are greatly reduced and so we can start seeing good MT performance even before we have any training corpora.

Thus the use of a word-for-word substitution dictionary makes sense both from the rule-based view-point and the SMT view-point. Development of such dictionaries should therefore be given the highest priority.

Traditional SMT systems combine lexical substitution and re-ordering, both of which are learned together from the training corpus. This makes SMT a lot more complex than it needs be. By separating the lexical substitution task from the reordering task, we can greatly simplify the system, both in terms of training corpus requirements and overall simplicity and efficiency. Syntactic divergences can be handled through a transfer grammar if the divergences are already clearly known. Otherwise, while such divergence studies can go in the background, we can develop and use pattern-matching ideas to discover and apply rules of re-ordering. Purely statistical methods will become feasible only after large scale parallel corpora have been developed.

We need to collect large, representative and clean corpora. We need to perform quantitative analysis at each stage, always trying to exploit the exponential nature of distributions we find in the statistics of linguistic material. We need to adapt sound engineering principles. We must base our work on solid theoretical foundations and avoid the tendency to take short cuts. See [7] for more on the theoretical foundations of language engineering.

To build an MT system quickly, we just need to develop a large database of word-for-word substitutions. We do not need any morph, nor do we need a POS tagger, local word grouper or a chunker to start with. We can start getting promising performance in translation very quickly. We can translate at great speed too. We need to carefully observe the outputs and enhance and improve the databases. Once we have sufficient data and sufficient experience, we can then start addressing the remaining research issues one by one.

One of the most widely held beliefs not only in MT but also in the whole field of Natural Language Processing (NLP) is that human languages are highly ambiguous. In fact, disambiguation is considered to be the main focus and priority. Words have multiple senses, they even belong to multiple grammatical categories, there are multiple ways of grouping words to form higher level structures. The degree of ambiguity is claimed to be very high, researchers have talked of hundreds, thousands, Millions, even Trillions of possibilities. The claim is that simply using a dictionary of equivalents will not work.

Upon careful analysis, we find that this is not true. Human languages cannot be so very ambiguous, otherwise, seven and a half Billion people would not be doing their daily business in natural languages. The exponential nature of ambiguities in natural languages is a result of lack of proper understanding of what a word is. We take the written form of language too seriously and we simply go by what we see in a piece of text. Words are taken to be sequences of characters (letters of the alphabet, punctuation marks, special symbols etc.) separated by white spaces. This is not right.

Of critical importance in our approach is the notion of a word. What exactly is a word? We have shown [9] that there is a much better way of defining words, starting from meanings, rather than from spellings. By re-defining words in a proper way, a great deal of ambiguities in languages melt away automatically. Computational complexity is also reduced significantly. In our own work, we find that most words are not ambiguous at all. In the case of ambiguous words, the degree of ambiguity is small. There are also simple ways of disambiguating the real cases of ambiguity.

MT is often projected as a product. This is the big problem. We may never be able to bring MT to a level where end users can directly use it as a ready-to-use product. We have not been able to reach that stage in the last 30 years in India. Instead, MT should be considered as a service. The user submits his translation requirements to a service provider. The service provider runs an MT system and looks at the machine generated output carefully. He may add more entries to the MT databases, he may correct the errors found in the databases if any, he may even edit and clean the input SL texts for the purposes of translation. He may run the MT system several times, each time improving the quality of translations produced as also the MT system itself. Finally, he may proof-read and manually post-edit the machine generated outputs to produce publishable quality translations. We can take good advantage of a synergy between the man and the machine, and semi-automatically produce high quality translations. Our experiments show that this is economically viable and practically feasible too.

Some people argue that high quality is not always essential in translation. Instead of taking this 'sour grape' attitude, we can actually start producing high quality translations by giving up the claim of making it fully automatic. Most successful engineering systems in the world use the best of both the man and the machine. People are ready to do their bit, if only we can guarantee high quality and economic viability. Reasonably good translation performance is possible by using a dictionary of word-to-word mappings, and re-ordering the TL words as required. After reaching this milestone, further improvements can be made through disambiguation rules etc.

Reasonably good translation performance is possible by using a dictionary of word-to-word mappings, and re-ordering the TL words as required. After reaching this milestone, further improvements can be made through disambiguation rules etc.

5. The saara Translator

Based on the saara theory [7] and the ideas and strategies discussed above, three different MT systems are being developed at the School of Computer and Information Sciences, University of Hyderabad, known as MT1, MT2 and MT3 respectively. A brief description of each, along with the current levels of performance, is given below. All the three MT systems are for translating Modern Kannada Prose, written in the so called Standard Dialect, into Modern Telugu Prose.

Quality of translation is measured in terms of Comprehensibility[2], that is, whether the meaning of the sentence can be understood by the reader. Comprehensibility is measured by manual evaluation, on a scale of 0-4, 4 being perfect and 0 indicating total failure. A score of 3 indicates almost perfect translation and a score of 2 indicates that the meaning of the sentence can be fully comprehended, albeit with some difficulty. A sentence is considered to be successfully translated if the score is 2 or more. The performance of the MT system is indicated in terms of the percentage of sentences that are successfully translated. This method of evaluation and scoring has become the defacto standard in India in recent times. We report here the translation performances obtained on a corpus of 4.6 lakh sentences, based on sample studies on several sets of 100 sentences each.

The MT2 system is a dictionary based SMT system. It does not use morphology, syntax or any other linguistic modules, nor does it use a parallel corpus. Its main focus as of now is lexical substitution. This system is very fast - it can translate 1,00,000 sentences per second on an ordinary Desktop PC or a Laptop. The database has about 1.2 lakh entries. Translation performance varies from about 45% to about 55% on first run. Once the outputs are checked and the databases updated as required, translation performance jumps to the range of 85% to 95%. Experiments have shown that the quality of translations so produced is acceptable for the purposes of final proof-reading and post-editing. Also, the time and cost of the entire process is comparable to that of manual translation, actually somewhat more economical as on date. The system improves with time and we believe that it will become a strong competitor to manual translation very soon.

The MT1 system, is an Analysis-Transfer-Generation based MT system. A comprehensive computational grammar of Kannada is used to perform morph analysis of Kannada words. Inflection, derivation and sandhi are all handled. Spelling variations are normalized automatically to a large extent. More than 90% of words are analyzed, with less than 10% error. The morph analyzer produces a fine-grained hierarchical tag for each input word. This tag includes all the necessary lexical, morphological, syntactic and semantic information required for further processing. The MT1 system currently uses a simple tag-for-tag transfer grammar. Telugu word forms are generated using a morph generator. Translation performance varies from about 35% to about 55%.

It may be recalled that this same or better performance is achieved by the MT2 system without using any morphology. This shows that morphology can be bypassed, even for morphologically rich languages, as a practical strategy to start with.

The MT3 system combines the best of the above two MT systems and achieves a performance of 55% to 72% on the first run. As already indicated, we can cross 90% performance using the man-machine synergy.

These MT systems are already much better than the Google Translator. We are in fact ready to take up small translation jobs. Further work is on to improve the databases in the MT2 system and to improve the linguistic modules in the MT1 system. Automatic post-editing modules are being designed to automatically improve the quality of translations, thereby reducing the time and effort required in final proof-reading.

6. Conclusions

In this paper we have described the state of the art in Machine Translation, we have included a critical review of the challenges faced in MT and the possible reasons behind the failures. We have given a number of suggestions for overcoming these limitations and challenges.

We have supported our claims by describing several MT systems we are currently developing. Our approaches, generally labelled 'The saara Translator', hold promise and raise hopes. We believe that the strategies followed in the development of 'The saara Translator' are applicable to other Indian languages as well. We can develop usable MT system within a very short time and gradually improve the systems as we keep using them.

References

1. Paradigm shift of language technology initiatives under tdil programme. Vishwabharat, Apr - jan 2007.
2. Akshar Bharati, Rajni Moona, Smriti Singh, Rajeev Sangal, and Dipti Mishra Sharma. Mteval: An evaluation methodology for machine translation systems. In Proceedings of SIMPLE-Symposium on Indian Morphology, Phonology and Language Engineering, IIT Kharagpur, India, 2004.
3. Hemant Darbari, Anuradha Lele, Aparupa Dasgupta, Ranjan Das, Debasri Dubey, Shraddha Kalele, Shahzad Alam, Priyanka Jain, and Pavan Kurariya. Enabling linguistic idiosyncrasy in anuvadaksh. Vishwabharat, July - Dec 2013.
4. Philipp Koehn. Europarl: A parallel corpus for statistical machine translation. MT Summit, 2005.
5. Philipp Koehn. Statistical Machine Translation. Cambridge Univeristy Press, 2009.

6. R.Mahesh K.Sinha. Man-machine integration in translation processes: an Indian scenario. In Bernadette Sharp, Michael Zock, Michael Carl, and Arnt Lykke Jakobsen, editors, Proceedings of the 8th international NLPSC workshop, Special theme: Human-machine interaction in translation, Copenhagen Studies in Language 41, pages 9–20, Copenhagen Business School, 20-21 August 2011. Samfundslitteratur.
7. Kavi Narayana Murthy. The saara approach to language engineering. Forthcoming.
8. Kavi Narayana Murthy. Mat: A machine assisted translation system. In Proc. of 5th Natural Language Pacific Rim Symposium, Beijing, China, 5-7 Nov 1999.
9. Kavi Narayana Murthy. On defining word. IJDL, XLIV(1):129–161, Jan 2015.

*****

Standard on Machine Translation Acceptance
Version 6.0

Technology Development for Indian Languages,
Ministry of Electronics & Information Technology,
Electronics Niketan,6,CGO Complex, New Delhi - 110 003.

1. Background

Machine Translation is the process of automatically converting the text from one natural language into another natural language. Different organizations have been working for several years to build user acceptable machine translation systems and overcome the language barriers and enable easier communication.

1.1. Worldwide Activity of Machine Translation

A survey of the machine translation systems shows that, huge efforts are being taken to develop various machine translation systems worldwide. Many MT systems across the globe have already been developed for the most commonly used natural languages such as English, Russian, French, Japanese, Chinese, Spanish, Hindi and other Indian languages etc. [2] SYSTRAN, LOGOS, METEO, Weidner and SPANAM are some of the well known results of the same. Some of the companies like CDAC, Google, Microsoft, are also offering Web based Machine Translation Services for limited sentences. In India, domain specific MT systems for English to Indian Language and Indian Languages to Indian Languages have been developed and made available through TDIL Data Centre (www.tdil-dc.in).

1.2. Machine Assisted Translation

The availability of content in multiple languages has become one of the most significant aspects of communicating information between businesses, companies and their customers, organizations, countries and common people. High level of accuracy and speed of producing translated content are important components of translation services. As per studies, a human translator can address at best 5-7 pages per day after which the translation loses its efficiency. As of now no machine translation system can perfectly translate the text into the target language, it can only aid the translation process. Hence, the term machine assisted/ aided translation (MAT) evolved.

MAT acts as a tool whereby translation efficiency may be increased. This is achieved by subsequently post-editing the output translation by the human. The time required to produce correct and readable translation is often seen as a measure of all post editing efforts.

Machine Assisted Translation is a powerful tool that has many purposes and can be used in a number of different ways:

- Quick translation of content to understand the information in another language.
- Aided Tool for human translators for translating bulk documents.
- Instant translation of real time communication of chat, email or customer support communications.

1.3. Different Machine Translation Approaches

Worldwide, there are different technological approaches being used for the development of machine translation systems as briefed below:

- **Rule based MT**

  Rule-based machine translation is based on the classical approach in which linguistic rules, grammar and bilingual dictionaries for source and target languages are used. It mainly covers the semantic, morphological, and syntactic regularities of each language. The objective of RBMT is to convert source language structure to target language structure.

- **Example based MT**

  Example-based machine translation (EBMT) method mainly uses bilingual corpus with parallel texts as its main knowledge base. In this type of translation system, a textual database of large number of translated sentences are stored. It uses case-based reasoning approach in which new translations are made by finding the most similar example in the example-base, and using this as a model for the new translation. [24]

- **Statistical Machine Translation (SMT)**

  Statistical machine translation is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora. [25]

- **Tree Adjoining Grammar based MT**

  The Tree Adjoining Grammar (TAG) consists of a set of elementary trees, divided into initial and auxiliary trees. This works on tree-to-tree translation model and makes use of linguistic syntactic tree for both the source and target language. Parser and Generator modules recognize various grammatical entities in the English sentence, analyze them, represent them in different tree structures and synthesize equivalent Indian Language sentences on the basis of the derivational tree structure and the Transfer Grammar.

- **Analyze-Transfer-Generate based MT**
  The source language text is preprocessed (collected, cleaned, and formatted) and analyzed. After language analysis, transfer of vocabulary and analyzed structure is carried out. And finally the target language text is generated. It is based on linguistics and may also use statistical methods in addition. 7

- **Hybrid & Pseudo Interlingua MT**
  This technology uses a pseudo Interlingua approach. It analyzes English only once and creates an intermediate structure with most of the disambiguation performed. The intermediate structure is then converted to each target language through a process of text generation. It also uses rule based MT and Example based MT approaches.

1.4. Government of India Efforts in Machine Translation System Development

India has multiple languages and scripts, hence need of language translation is immense. Work in this area has been going on for several decades. Many premier academic and R&D organizations are engaged in the development of MT Systems for Indian languages.

Government of India has facilitated development of Machine Translation Systems for translation from English to Indian Languages(IL) and from Indian Languages to Indian Languages using different technological approaches. The language pairs addressed for English to IL are English-Hindi; Assamese; Bodo; Bangla; Gujarati; Malayalam; Marathi; Nepali; Punjabi; Oriya; Tamil; Telugu; and Urdu. Indian Language MT Systems include 9 bidirectional language pairs, i.e. Punjabi – Hindi –Punjabi, Telugu – Tamil – Telugu, Urdu – Hindi – Urdu, Hindi – Telugu – Hindi, Marathi – Hindi – Marathi, Bengali – Hindi – Bengali, Tamil – Hindi – Tamil, Kannada – Hindi – Kannada, Malayalam – Tamil – Malayalam.

2. Importance of MT Evaluation

Machine translation is an important and useful technology, which helps in reducing the time taken for translation of the content, but the quality of machine translation should be acceptable by the end user. Quality of translation may affect the meaning of the entire content as wrong translation may mislead the user by providing incorrect information. Hence, the assessment of translation quality plays the most important role in machine translation development and acceptance of the systems by the end-user. Evaluation of Machine Translation itself is a complex and difficult subject. Researchers have worked on MT evaluation and evolved many subjective (Human) and automatic evaluation techniques and some of the techniques are briefed below.

- **Subjective (Human) Evaluation**

The quality of translation is mainly context dependent and there may be multiple correct translations of a source sentence. To judge whether the output translation quality is in sync with the context of the source sentence, it is 8 necessary to involve human in the evaluation process. Human evaluation allows measuring the quality of MT system by a set of end users. They can recognize the translation errors and grade the output quality adequately. There are different subjective evaluation methods to check the various aspects of translation quality such as comprehensibility, fluency, adequacy and intelligibility. [6]

Following subjective evaluation methods have been evolved for testing the outcomes of the Indian Languages Machine Translation Research and Development projects implemented under TDIL Programme of Ministry of Electronics & IT, Govt of India.

i. 5-Point Scale for Accuracy Evaluation of Indian Languages MT.
ii. Modified 5-Point Scale for Comprehensibility & Fluency for Indian Languages.

- **Limitations of Human Evaluation**

Though human evaluation is very useful & informative, it has several inherent challenges & limitations such as: [11]

i. Expensive & slow: Human evaluation is a time consuming activity and labor-intensive process.

ii. Inter-evaluator agreement: Generally machine translation output is evaluated by multiple evaluators. Two evaluators do not give same score on the same data set, though the same evaluation guidelines/ training are provided. Perception about language, choice of words, may differ from human to human. Therefore, in human evaluation, inter-annotator agreement issue always persists.

iii. Training of evaluators: Providing trainings on the evaluation process to the evaluators is very crucial. Even after detailed training there may be wrong grading by the evaluators as there is a fine difference among grading scales.

- **Automatic evaluation**

Automatic evaluation is a fast, inexpensive and mostly language independent way of evaluating the translation quality. It is useful wherein frequent evaluations are required. 9

The quality of a translation is inherently subjective. Therefore, any metric must assign quality scores such that they correlate with human judgment of quality. Human judgment is the benchmark for assessing automatic metrics, as humans are the end-users of any translation output. Even if a metric correlates well with human judgment in one study on one corpus, this successful correlation may not carry over to another corpus. Automatic evaluation metrics such as BLEU, NIST, METEOR and TER are more suitable for same language family pairs such as Latin based languages, but these are not of much use for distant language pairs such as English to Indian Languages.

- **Limitations of Automatic Evaluation w.r.t. Indian Language MT systems**

Due to the limitations of automatic evaluation, these are not directly applicable for Indian language machine translation evaluation. Automatic measures e.g. BLEU, METEOR, NIST are not diagnostic as these are based on measures of string matching. These metrics do not provide feedback on the ability of MT systems to translate various aspects of the language. [8]

i. Word Order: These most popular metrics do not work well when used for evaluation of translation among distant language pairs like English to Indian languages as English has a different word order from Indian languages.

ii. Multiple correct translations [20]: In case of BLEU, only exact match is considered and synonyms are not considered. All the Indian languages are morphologically rich; there can be multiple correct translations for single input sentence just word to word matching may lead to wrong results and make the evaluation process harder. Multiple correct translations may differ in word choice or the word order choice also.

iii. BLEU is N-gram precision based metric and does not care about the untranslated words in output translation. [23]

iv. No single automatic metrics can perform well for all the Indian language pairs. [17] METEOR uses syntactic parsers, synonym databases, stemming. As of now, these resources are available for few languages.

Details of various methods as listed above may be referred at Annexure-II.

Evaluation of MT system is important and there are various techniques for Indian languages MT evaluation, irrespective of MT development approaches. Both, human evaluation (subjective) and automatic evaluation have some advantages and limitations. In case of Indian languages, these methodologies are not very useful. Hence, a simplified evaluation methodology for defining the criterion of accepting the machine translation output is required.

3. Objective

This Standard (Code of Practice) document defines the criterion for accepting the quality of translated content created by the Machine Translation system. This standard provides "minimum translation quality acceptability score" which indicates that the Machine Translation output is acceptable for post editing and making it useful for the intended purpose.

This standard also defines acceptability metric in the form of various parameters, which measures the quality of the output of Machine Translation system based on post-editing efforts required to make the translated content acceptable by the end user. These parameters are applicable for any language pair irrespective of the source and target language and the machine translation techniques.

A simplified evaluation procedure is defined so that a person who knows both the source and target languages can easily evaluate the translated content.

4. Scope

This document is intended to define the acceptance criteria of Indian language machine translation output. This document may be referred by various stakeholders as mentioned below, to check the acceptability of machine translated content.

- Translation agencies/ Translators to know the quality of the machine translated content with respect to post-editing efforts required to make it completely useful.
- MT evaluation agencies to evaluate the quality of the machine translated content.
- MT Researchers and Developers for analyzing the different types of errors that may occur in the translated content and further improving the quality.

5. Machine Translation Acceptability

Machine translation acceptability is an inspection method that measures the system errors so that a decision could be taken about the usefulness of the machine translated content. The quality of machine translation output is also judged on the time taken for post editing to achieve perfect translation versus the time taken for manual translation. It is expected that use of machine translation should reduce the overall translation time to minimum 50-60% of the time taken over manual translation. Machine translation acceptance may also be defined as a criterion to check whether it is feasible to accept the MT output for post editing or to retranslate the entire content manually.

6. Machine Translation Acceptability Parameters

Machine Translation evaluation is very much subjective and researchers have evolved subjective evaluation techniques by taking comprehensibility as the measure of translation quality. However, comprehensibility score given by different evaluators varies as perception towards the choice of words and acceptable quality differs from human to human. Hence, there is a need of simplified methodology to objectively capture the parameters affecting the translation quality.

In order to judge the acceptability of MT output, various parameters are defined in simplified terminologies so that any native language evaluators can use it. Different grading scales are defined with the suggested weight for each scale to check the acceptance of machine translation output. The suggested weights may be revised based on the large-scale experimentation. Following parameters are defined [16] [21]:

6.1. Acceptability Parameters

i. Sentence Structure

If the sentence structure of the output is not proper or the target language words are correct, but not as per the syntactic rules of target language then this error can be marked as sentence structure error. The following grades are defined to rate this parameter.

Grade	Scale
0	Sentence structure is not at all proper and user prefers to re-translate the same sentence
0.5	Sentence structure is partially correct and user prefers to do the post editing in the translated output
1	Sentence structure is proper

ii. Word Inflection

Word inflection means the root of the word is correct, but the word requires a little modification. This word inflection error may contain error in gender translation, incorrect verb form, incorrect noun form, incorrect adjective form or adverb. Verb related inflection sometimes may alter multiple words.

Grade	Scale
0	Word inflection error present in the translated sentence
0.5	Word inflection error present in less than 50% words of the translated sentence

iii. Spelling Mistakes

The spelling mistake error should be marked if any word in the translated sentence is misspelled. The following grades are defined to rate this parameter.

Grade	Scale
0	The output translation has a spelling mistake/s
0.5	Spelling mistakes present in less than 50% words of the translated sentence
1	The output translation has no spelling mistake/s

iv. Word Sense Disambiguation (WSD)

The process of identifying an appropriate sense of word/s (i.e. meaning) in a sentence, when the word has multiple meanings is defined as WSD. WSD Error should be marked when the meaning of word/s in the sentence context is not appropriate.
Example:
Source Sentence:

The **capital of Bihar** is famous for its rich history and royal architecture.
Translated Sentence with WSD Error:
इसका धनी इतिहास और शाही वास्तुकला के लिये विख्यात बिहार का बड़े अक्षर ।

Correct WSD: बिहार की राजधानी

Grade	Scale
0	Sense of the word/(s) is inappropriate in the sentence context
1	Sense of the word/(s) is appropriate in the sentence context

v. Named Entity (NE) Transliteration

Machine translation system should recognize named entities (proper names; including trade names, brand names, registered trademarks, place names, and personal names) in the source language text and should transliterate these NEs into the target language. Also, there are some borrowed words of source language which are acceptable in the transliterated form in the target language e.g. Computer (कम्प्युटर), Rail (रेल).
The following grades are defined to rate this parameter.

Grade	Scale
0	Named Entity is not identified and transliterated
0.5	Error is present in less than 50% NE words in the sentence
1	Named Entity is identified and transliterated correctly
NA	This parameter is Not Applicable in the given sentence

vi. Punctuations

Sometimes the punctuation marks used in source sentences are not retained in the translated sentences. Wrong placement of punctuation mark(s) may change the meaning of the sentence.

Grade	Scale
0	Punctuation error [punctuation missing or wrongly placed]
0.5	Number of Punctuation errors are less than 50%
1	Punctuation(s) retained
NA	This parameter is Not Applicable in the given sentence

vii. Numerals

"Numerals" category in the translated sentence may contain Numbers, Currency, Date, Time, Percentage, etc.
While translating, if the numerals are not handled properly then it should be marked as "Numeral Error". The following grades are defined to rate this parameter.

Grade	Scale
0	Numerals are not handled properly in the output translation
0.5	Number of numeral errors are less than 50%
1	Numerals are correctly handled and translated
NA	This parameter is Not Applicable in the given sentence

viii. Abbreviations/Acronyms

Acronym is an abbreviation formed from the initial letters of a multiword expression. If some acronym is present in the source sentence, it needs to be transliterated in the target language. If abbreviations/ acronyms are not retained properly, it will be marked as an error. The following grades are defined to rate this parameter.

Grade	Scale
0	Abbreviations/Acronyms not identified and transliterated properly
0.5	More than 50% Abbreviations/Acronyms identified and transliterated correctly
1	All Abbreviations/Acronyms identified and transliterated correctly
NA	This parameter is Not Applicable in the given sentence

ix. Untranslated Words

There might be a case like some of the words are not getting translated into the target language and remain as it is (i.e. in the source language script) in the output sentence. User can mark this error as "Untranslated Words".

Grade	Scale
0	Untranslated word(s) error, i.e. the word is in source language script or word(s) transliterated but are not accepted in the target language
1	All words are translated

x. Irrelevant Word(s) Error

Irrelevant word(s) means there is no corresponding reference word in the source sentence, but these words are present in the output sentence. The following grades are defined to rate this parameter.

Grade	Scale
0	Irrelevant Word(s) error is present
1	Irrelevant Word(s) error is not present

xi. Meaning Conveyed

The main objective of translation is to retain the information (meaning of the context) of the source language in the target language. Therefore, it is important to check how much information from source language is conveyed in the target language. The following grades are defined to rate this parameter.

Grade	Scale
0	Completely irrelevant translation OR Word to word translation, but no meaning conveyed
0.5	Word to word translation, but partial meaning conveyed
1	Most of the meaning conveyed

6.2. Recommended Test Data

- It is observed that sentence length vary significantly across languages and language families. The sentences selected for evaluation should be complete and from the intended domain. Evaluation should be conducted on a data set of 100 sentences covering different grammatical structures and various parameters listed above. It is recommended that the test data set should have at least 5% sentences for each defined parameter so that all the variations are captured in the test data for the detailed MT evaluation.

- It is suggested that minimum 03 evaluators should be engaged to reduce the subjectivity in the evaluation.

6.3. Acceptability Score Calculation

S. No.	Acceptability Parameters	Grade Assigned (Gi)	Weight (Wi)	Assigned Score (Pi-Max / Pi-Min) $P_i = G_i \times W_i$
i.	Sentence Structure	1/0.5/0	20	20/10/0
ii.	Word Inflection	1/0.5/0	10	10/5/0
iii.	Spelling Mistakes	1/0.5/0	5	5/2.5/0
iv.	Word Sense Disambiguation (WSD)	1/0	10	10/0
v.	Transliteration	1/0.5/0	10	10/5/0
vi.	Punctuations	1/0.5/0	5	5/2.5/0
vii.	Numerals	1/0.5/0	5	5/2.5/0
viii.	Abbreviations/Acronyms	1/0.5/0	5	5/2.5/0
ix.	Untranslated Word(s)	1/0	5	5/0
x.	Irrelevant Word(s)	1/0	5	5/0
xi.	Meaning Conveyed	1/0.5/0	20	20/15/00

Where i= 1 to 11

(A) Xi = Sum of scores given by evaluator for the applicable parameters in each sentence.

$$Xi = \frac{(P_{1\text{- Assigned}} + P_{2\text{-Assigned}} + .. + ... + P_{11\text{-Assigned}})}{(P_{1\text{- Max}} + P_{2\text{-Max}} + .. + .. + P_{11\text{-Max}})} \times 100$$

Where
Pi- Assigned = Gi x Wi of applicable parameter
Pi- Max is the maximum score of applicable ith parameter

(B) Formula for calculating Acceptability score for single evaluator (Ei)

$$Ei = [(X_1 + X_2 + X_3 + ... + X_n)/n]$$

Where n = Total number of sentences (100 sentences)
(B) Final Acceptability score (average of evaluator’s score) is calculated below:-

Acceptability Score = [(E1 + E2 + E3+.....+EN)/N]

Where N = Total number of evaluators

6.4. Recommended Acceptability Score

A Machine Translation Output should get minimum Acceptability Score of 50-60 based on above recommended formula. This score indicates that the Machine Translation output is acceptable for post editing and making it useful for the intended purpose.

The Acceptability Score should be calculated on a set of 100 sentences and 3 evaluators. However, more number of sentences and evaluators may be taken for better evaluation.

Annexure I: Guidelines for Grading

The evaluator should have access to the source language text while assigning the grades to the translated content.

- The evaluator should grade the above parameters for each sentence as per the defined grading scales.

- Once a grade is assigned to a word in a sentence for any error, another grade cannot be assigned for the same word in that sentence. This is to avoid wrong grading i.e. if word inflection error is observed in the output sentence, then it should be marked as word inflection error only and not again as a spelling error.

- Sentence lengths vary significantly across languages and language families. The sentences selected for evaluation should be complete and from the intended domain. Evaluation should be conducted on a data set of 100 sentences covering different grammatical structures and various parameters listed above; however, all the above mentioned parameters may not be applicable for each sentence.

- It is suggested that minimum 03 evaluators should be engaged to reduce the subjectivity in the evaluation.

Annexure II: Software Engineering Advisory Parameters

Along with the machine translation output acceptance parameters, this document also defines machine translation software advisory parameters. These parameters are related mainly to the software engineering aspects of MT System. [15]

a. **Translation Speed**
   In case of bulk translations, the speed of the MT Engine is very crucial. The machine translation engine speed shall be calculated in terms of translated-words/second.

b. **Portability**
   The functionality of the MT Software shall be tested for different devices like Mobile, Tablet, Laptop and Operating Systems to ensure inter-operability.

c. Usability

The User Interface of the Machine Translation application should be user friendly and integrated with user manual describing all the functions along with examples.

Annexure III: Details of Human Evaluation Methodologies Adopted Earlier

The quality of translation is mainly context dependent and there may be multiple correct translations of a source sentence. To judge whether the output translation quality is in sync with the context of the source sentence, it is necessary to involve human in the evaluation process. Human evaluation allows measuring the quality of MT system by a set of end users. They can recognize the translation errors and grade the output quality adequately. There are different subjective evaluation methods to check the various aspects of translation quality such as comprehensibility, fluency, adequacy and intelligibility.

Following subjective evaluation methods have been evolved for testing the outcomes of the Indian Languages Machine Translation Research and Development projects implemented under TDIL Programme of Ministry of Electronics & IT, Govt of India.

1. 5-Point Scale: Accuracy (Proposed Scale for Indian Languages)

Various MT Systems have been developed under the TDIL-MeitY funded MT projects. A 5-point evaluation scale was evolved to assess their performance. The approach was formulated with focus on usability and the native speaker's expectations and the translation quality is measured in terms of comprehensibility of output.

Grade 0	No output provided by the engine.
Grade 1	The translated output is not comprehensible.
Grade 2	Comprehensible after accessing the source text.
Grade 3	Comprehensible with difficulty.
Grade 4	Acceptable since the text is comprehensible.

2. Modified 5-Point Scale: Comprehensibility & Fluency

5-Point scale for accuracy mentioned above has Grade 0 for "no output provided by system". This was mainly the result of MT Engine system failure. Experts deliberated on the issue and grade -1 was evolved for No Output by system due to technical reasons such as buffer clearance, etc. Hence, a Modified 5-point scale was evolved as presented below. In modified Scale performance of the MT system is measured on two parameters (1) Comprehensibility & (2) Fluency.

Grade -1	No Output or buffer clearance issue
Grade 0	Nonsense (If the sentence doesn't make any sense at all – it is like someone speaking to you in a language you don't know)
Grade 1	Some parts make sense but is not comprehensible overall (e.g., listening to a language which has lots of borrowed words from your language – you understand those words but nothing more)
Grade 2	Comprehensible but has quite a few errors (e.g., someone who can speak your language but would make lots of errors. However, you can make sense out of what is being said)
Grade 3	Comprehensible, occasional errors (e.g., someone speaking Hindi getting all its genders wrong)
Grade 4	Perfect (e.g., someone who knows the language)

Annexure IV: GLOSSARY

Adequacy	Adequacy refers to the degree to which information present in the source text is communicated in target translation. The objective of the adequacy is to determine the extent to which all of the content of a text is conveyed, regardless the quality of the language in the candidate translation.[15]
BLEU	BLEU stands for Bilingual evaluation under study. It compares the translated machine output with the reference output generated by professional human translator. BLEU is precision based and it uses modified N gram precision. Word precision account for adequacy and n gram precision for n =1, 2, 3 account for fluency
Blind testing	In blind testing, evaluators have no access to the source text, this is to eliminate bias scoring.
Comprehensibility	Comprehensibility is a measure of how easy is a text to understand.
Fidelity	Fidelity is measurement of correctness of the information transferred from source language to the target language. It is a subjective evaluation of the measure in which the information contained in the sentence of the original text reappears without distortion in the translation.[15]

Fluency	Fluency refers to the degree to which the target is well formed according to the rules of the target language. The objective of fluency evaluation is to determine how much like "good fluent" a translation appears to be without taking into account the correctness of information.[15]
Intelligibility	Intelligibility is a measure of how "understandable" the sentence is. Intelligibility is measured without reference to the original source sentence.[15]
METEOR	This metric is based on word to Word alignment between candidate and reference translation. The final Meteor score is between 0 and 1 which is a harmonic mean of unigram precision and unigram recall.
NIST	The NIST metric is based on the BLEU metric. Doddington established NIST in 2002 which is similar to BLEU except, it assigns a weight to each unigram depending upon its uniqueness or how informative the n-gram is
Open testing	Open testing, evaluators will have access to source text while evaluating MT translation output.
TER	TER (Translate Error Rate) was proposed by Snover and Dorr 2006. It represents the number of edits necessarily required to transform the machine output to reference translation.

References:

1. Evaluation of machine translation [Online]. Available: http://en.wikipedia.org/wiki/Evaluation_of_machine_translation.[January 9, 2014]
2. Jonathan Sloculn. “A Survey Of Machine Translation: Its History, Current Status, And Future Prospects.” Computational Linguistics, Volume 11, Number 1, January-March 1985
3. Word error rate [Online]. Available: http://en.wikipedia.org/wiki/Word_error_rate. [January 28,2014]
4. Martin Thoma. Word Error Rate Calculation [Online]. Available: http://martin-thoma.com/word-error-rate-calculation. 2013. [February 4, 2014]
5. Sara Stymne, Machine Translation Evaluation [Online].Available: http://www.ida.liu.se/labs/nlplab/gslt/mt-course/mteval-sarst.pdf. [February 10,2014]
6. Evaluation of machine translation [Online]. Available: http://www.translationdirectory.com/articles/article1814.php. 2008. [February 27,2014]
7. Goyal, Vishal, and Gurpreet Singh Lehal. "Evaluation of Hindi to Punjabi machine translation system." arXiv preprint arXiv:0910.1868 (2009).
8. Bonnie Dorr, Matt Snover, Nitin Madnani, Part 5: Machine Translation Evaluation [Online]. Available: https://www.cs.cmu.edu/~alavie/papers/GALE-book-Ch5.pdf. [February 10,2014]

9. Sinha, R. M. K., and A. Jain. "AnglaHindi: an English to Hindi machine-aided translation system." MT Summit IX, New Orleans, USA (2003): 494-497.
10. Balyan, Renu, et al. "A diagnostic evaluation approach for english to hindi MT using linguistic checkpoints and error rates." Computational Linguistics and Intelligent Text Processing. Springer Berlin Heidelberg, 2013. 285-296.
11. Joshi, Nisheeth, Hemant Darbari, and Iti Mathur. "Human and Automatic Evaluation of English to Hindi Machine Translation Systems." Advances in Computer Science, Engineering & Applications. Springer Berlin Heidelberg, 2012. 423-432.
12. Zhou, Ming, et al. "Diagnostic evaluation of machine translation systems using automatically constructed linguistic check-points." Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 2008.
13. Specia, Lucia, et al. "Predicting machine translation adequacy." Machine Translation Summit. Vol. 13. No. 2011. 2011.
14. Correa, Nelson. "A fine-grained evaluation Framework for machine translation system development." MT Summit IX. 2003.
15. Van Slype, Georges. "Critical study of methods for evaluating the quality of machine translation." Prepared for the Commission of European Communities Directorate General Scientific and Technical Information and Information Management. Report BR 19142 (1979).
16. Jorg Schütz. “Deploying the SAE J2450 Translation Quality Metric in Language Technology Evaluation Projects.” [Translating and the Computer 21. Proceedings... 10-11 November 1999 (London: Aslib)]
17. Kalyani, Aditi, et al. "Assessing the Quality of MT Systems for Hindi to English Translation." arXiv preprint arXiv:1404.3992 2014.
18. Vilar, David, et al. "Error analysis of statistical machine translation output."Proceedings of LREC. 2006.
19. Stymne, Sara, and Lars Ahrenberg. "On the practice of error analysis for machine translation evaluation." LREC. 2012.
20. Ananthakrishnan, R., et al. "Some issues in automatic evaluation of english-hindi mt: more blues for bleu." ICON 2007.
21. Browen Hui, Measuring User Acceptability of Machine Translations to Diagnose System Errors: An Experience Report [Online].Available: https://www.aclweb.org/anthology/W/W02/W02-1609.pdf [July 20,2016]
22. Doherty, Stephen, Sharon O’Brien, and Michael Carl. "Eye tracking as an MT evaluation technique." Machine translation 24.1 (2010): 1-13.
23. Vaishali Gupta, Nisheeth Joshi and Iti Mathur. “Subjective and Objective Evaluation of English to Urdu Machine Translation.” 1310.0578.pdf
24. EBMT Seen as Case-based Reasoning.pdf Harold Somers Centre for Computational Linguistics, UMIST, PO Box 88, Manchester M60 1QD, England Harold.Somers@umist.ac.uk
25. https://en.wikipedia.org/wiki/Statistical_machine_translation

*****

Presentation on Hindi e-Tools and Rajbhasha Prodyaugiki at All India Rajbhasha Hindi Conference and Workshop, Goa

All India Rajbhasha Hindi Conference and Workshop were organized by Parivartan Jankalyan Samiti from 03 Oct 2016 to 05 Oct 2016 at Bogmallo, Goa. Approximately 60-70 Rajbhasha Adhikari from various PSUs and other organizations attended this conference.

On third day session on Hindi e-Tools and Rajbhasha Prodyaugiki from 2:00PM to 03:00PM was scheduled. In this session the advantages of Unicode, Enhanced Inscript, Enhanced Inscript typing tutor, Unicode Typing tool, Bharteeya Libre Office, SakalBharati Font, GIST Data Convertor and other localized software provided in Hindi CD was presented.

It was an interactive session and various queries have been asked and answered like typing sequence of conjuncts, enabling Inscript on computer, switching English and Hindi keyboard etc. The presentation was well received and appreciated by the audience. Hindi Inscript Typing tutor software was specially appreciated by some Rajbhasha Officers. Some of members were curious about speech to text in Hindi and asked about C-DAC's Shrutlekhan and Google STT in Hindi. Members were advised to contact C-DAC Pune AAI group regarding Shrutlekhan software.

Hindi Language CD has been distributed to all participants with registration kit.

*****

Presentation on Hindi e-Tools and Rajbhasha Prodyaugiki at All India Rajbhasha Hindi Conference and Workshop, Thiruvananthapuram

All India Rajbhasha Hindi Conference and Workshop were organized by Parivartan Jankalyan Samiti from 26 May 2016 to 28 May 2016 at Kovlam, Thiruvananthapuram. The conference was inaugurated by Dr. Prasanna Kumar Patasani, Members of Parliament (Lok Sabha). Approximately 100+ Rajbhasha Adhikari from various PSUs and other organizations attended this conference.

On second day session on Hindi e-Tools and Rajbhasha Prodyaugiki from 9:30AM to 11:00 was scheduled. In this session the advantages of Unicode, Enhanced Inscript, Enhanced Inscript typing tutor, Unicode Typing tool, Bharteeya Libre Office, SakalBharati Font, GIST Data Convertor and other localized software provided in Hindi CD was presented.

It was an interactive session and various queries have been asked and answered like typing sequence of conjuncts, switching English and Hindi keyboard etc. The presentation was well received and appreciated by the audience. Hindi Language CD and BOO handbook has been distributed to all participants with registration kit.

Some of members raised the point of limited support of Unicode in printing software like Corel draw, Adobe CS Suit etc and availability of beautiful Unicode based fonts for printing purpose.

*****

Ministry of Electronics and Information Technology
Electronics Niketan, 6, CGO Complex, New Delhi – 110003
Telfax : 011-2436 3525 Email : tdilinfo@mit.gov.in Website : http//tdil.mit.gov.in