

# A SURVEY OF MACHINE TRANSLATION APPROACHES

Harjinder Kaur, Dr. Vijay Laxmi

*Abstract – Machine translation (MT) plays an important role in benefiting linguists, sociologists, computer scientists, etc. by processing natural language to translate it into some other natural language. The demand of translation has become more in recent years due to increase in the exchange of information between various regions using different regional languages. Due to this reason machine translation has become an important research sunbfield under the Artificial Intelligence(AI). Many approaches have been used in the recent times to develop an MT system. Each of these approaches has its own advantages and disadvantages. The performance of an MT system depends on the approach used to design the system. In this paper we are presenting a brief overview of the MT and various techniques of designing an MT system. Also we are discussing the challenges faced while translating one language into another..*

*Index Terms- Machine Translation, Rule Based Approach Direct Approach, Tranfer Based Approach, Interlingual approach, Hybrid Approach, EBMT, SMT.*

*Harjinder Kaur, M.Tech. CSE, Guru Kashi University Talwandi Sabo, Bathinda, Punjab, India.*

*Dr. Vijay Laxmi, Associate Professor, CSE Department, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India*

## I. INTRODUCTION

Machine translation can be defined as the study of designing the systems that can translate one human language into another. These systems take input in one natural language and convert it into another human language. The language that is given as an input is called Source Language and the language in which we get the output is called Target language.

### Need of Translation

India is a multilingual country. Major population is not familiar with English. As most of the information available on web or electronic information is in English, people who are unaware of English cannot make use of this electronic information without anyone's help. In order to make it possible for evryone to use web, automatic language translators are essential. Also in state like Punjab where most of the official work is done in punjabi and not english, there has to be some way to convert the english documents from outside the state into punjabi.

## II. LITERATURE REVIEW

**R.M.K. Sinha and A. Jain, AnglaHindi: An English to Hindi Machine-Aided Translation System.** presented a machine translation system called AnglaHindi which is an English to Hindi version of the ANGLABHARTI translation methodology with a mixture of some example-based translation methodology. AnglaHindi system has been webenabled and is available at URL: <http://anglahindi.iitk.ac.in> for free translation. The system generates approximately 90% acceptable translation in case of simple, compound and complex sentences upto a length of 20 words. [3]

**Latha R. Nair and David Peter S., Machine Translation systems for Indian languages,** discusses various approaches to machine translation and various machine translation systems that have been developed for Indian languages. An English to Hindi MT system called Mantra was developed by CDAC, Bangalore in 1999. It translates domain specific documents in the field of personal administration.[1]. This paper also discusses various other translation systems in India that are converting English to Hindi using different approaches of MT. It is concluded that translation based systems are more flexible. Direct translation is appropriate for structurally similar language [1]. Interlingua approach is used in case of multilingual translations with the help of UNL.

**Cheragui A.M, Theoretical Overview of Machine Translation** discusses the basics and history of machine translation from 1948 to 2010. This paper

also presents various approaches of translation based on linguistic and computational architectures. Various types of machine translations their evaluation strategies are discussed.

## III. APPROACHES TO MACHINE TRANSLATION

A machine translation system first analyses the source language input and creates an internal representation. Then this representation is manipulated and transferred to a form suitable for the target language. Then at last output is generated in target language. Based upon the degree of dependence of internal representation on the source and target languages approaches, MT can be classified as follows:

- **Direct Approach:** This approach directly translates the source language to the target language. Due to this direct mapping such systems are highly dependent on both the source and target languages. Direct translation systems are basically bilingual and uni-directional. Only a little syntactic and semantic analysis is required in this approach.
- **Rule Based Approach:-** Rule based MT systems parse the source text and produce an intermediate representation. Depending on the intermediate representation used, this approach is further classified into following approaches[1]:

- **Transfer Based Approach:** It first converts the source language into an internal representation that depends on the source but not the target language. This representation is then converted to another internal representation which is dependent on the target language only. Finally target language is generated.
- **Inter-lingua Approach:** It converts the input into a single internal representation that is independent of both the source and the target language. This internal representation is then converted into final output.
- **Corpus-based Approach:** this approach uses a large amount of raw data in the form of parallel corpora. This raw data contains text and their translations. These corpora are used for acquiring translation knowledge. Corpus based approach is further classified into following two sub types.
  - **Statistical Machine Translation (SMT):** In this approach translations are generated on the basis of statistical models. SMT depends on a language model, a translation model and a decoding algorithm.
  - **Example-Based Machine Translation (EBMT):** In this the examples of existing translation are reused for new translation. First of all the examples matching with the input are found out. Then alignment is done in order to find out the parts of translation that can be reused. Finally recombination is done to make sure that the reusable parts identified during alignment are put together in a legitimate way.
- **Hybrid Approach:** This approach uses a combination of transfer approach and one of the corpus based approaches in order to overcome their limitations.

#### IV. CHALLENGES IN MACHINE TRANSLATION

- Two languages may have completely different structures.eg. English has SVO (Subject- Verb- Object) structure and Tamil has SOV (Subject- Object- Verb) structure. This difference may make the translation process tedious.
- The way in which the sentences are put together in different languages may differ.
- All words in one language may not have equivalents in other language. In some cases a single word in one language is expressed as a group of words in some other language. Such words are difficult to translate.
- There are certain words in some languages where transliteration is required before translation.eg. word in a language is used as noun.

- Ambiguity can have adverse effect on translation process. Ambiguity means a word or a Whole sentence in a language has entirely different meaning in some other language.
- The natural language is open and keeps on changing from time to time. So complete automatic simulation of natural language is almost impossible.

#### V. CONCLUSION

Machine translation has been an active research sub-field of AI from years. But the challenges faced during translation need to be solved for which more detailed study of various natural languages is required. So still a lot of work is required to develop a completely automatic translation system.

#### VI. REFERENCES

1. Latha R. Nair and David Peter S., Machine Translation systems for Indian Languages, IJCA, Feb 2012.
2. Goyal V, Lehal G.S., “Advances in Machine translation System Languages in India”, 2009.
3. R.M.K. Sinha and Ajay Jain, AnglaHindi: An English to Hindi Machine Translation System, MT Summit IX, New Orleans, USA, Sept.23-27, 2003.
4. Rao D.D., “Machine Translation - A Gentle Introduction”, Resonance, July 1998.
5. Hutchins J.W., “Introduction to Machine Translation”, Academic Press, 1992.

6. Ambati V., Rohini U., “A Hybrid Approach to Example based Machine Translation for Indian Languages”, Proceedings ICON 2007.

7. Cheragui M.A., “Theoretical Overview of Machine Translation”, Proceedings ICWIT 2012.

#### VII ABOUT THE AUTHORS



**Harjinder kaur** received her B.E. Computer Engineering Degree from AISSMS WCOE, University of Pune in 2007. Currently she is pursuing M. Tech CSE from Guru Kashi University, Talwandi Sabo , Punjab. She is doing her thesis work on Machine Translation of English to Punjabi. Her areas of interest include Machine translation, OCR.



**Dr Vijay Laxmi** received her B-Tech degree in computer science & Engineering from SLIET Longowal in 2003, and Ph.D degree in Computer Science & Engineering in year 2012. Her research area is Grid Computing & OCR. She has published 25 research papers in various National/International Conferences and Journals. At present, she is engaged in Guru Kashi University, Talwandi Sabo, and Punjab as Dy. Dear Research and an Associate Professor in Computer Science Engineering & Information Technology department.