

# Survey of Approaches Used in Machine Translation System

Mamta<sup>1</sup>, Tanuj Wala<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Career Point University Hamirpur, India

<sup>2</sup>Assistant Professor, Department of Computer Science, Career Point University Hamirpur, India

**Abstract**— Machine translation is one of the most important applications of Natural Language Processing (NLP). Many techniques and methods for different language have been proposed and developed using different approaches of machine translation. Machine translation translate the one language (source text) into the another language (target text) without human intervention. In this paper, we will provide the survey on various approaches of machine translation and detail of approach which is mostly use.

**Index Terms**— Machine translation, example based machine translation, statistical based machine translation.

## I. INTRODUCTION

Machine translation is one of the most important applications of computational linguistics that uses the computer software or web to translate text from one language to another.

As the world becomes more globalized, the problem of understand other language information to be translated by hand can only become more acute. Human translators have expensive resources and limited resources. Machine translation can increase the efficiency of human translators. People from different countries have different languages and different culture, so there is a big need of inter language translation to transfer their information, share ideas and communicate with one another. Email, mobile texting, instant messaging, online social media and video conferencing are an integral part of today's information society. Machine translation offers the direct and immediate response that would be difficult to achieve with human translators.

Machine translation is one of the most important applications of Natural Language Processing. Machine translation that helps people from different places to understand an unknown language without the aid of a human

*Mamta, Department of Computer Science, Career Point University Hamirpur, India*

translator. Machine translation is often perceived as low quality based an outdated perception created by older translation technologies or freely available generic translation tools from Google or Bing that have not been customized for a specific purpose.

In this paper, we will discuss the approaches of machine translation. The rest of the paper organized as follows: Section II, covers the related work behind these approaches, Section III, provides the overview of various machine translation approaches, In Section IV, we will cover the detail of machine translation approach, and in last Section V, provides the conclusion of complete study.

## II. RELATED WORK

In recent years, machine translation became the most encouraging field for researchers in India. Lots of research projects are going on machine translation in the world. Indian government is also sponsoring the various projects for removing the language barriers and developing the multilingual translators.

Sitender et. al. (2012) [1], the author's discussed the introduction of various Indian machine translation systems either developed or under the development on different Indian languages. They also give brief description on the Indian Machine Translation Systems on the basis of machine translation approaches and domain etc. Antony P. J. (2013) [2] gives a brief description of the various approaches and major machine translation developments in India and shows comparison of different MT systems in India on the basis of approaches. W. John Hutchins (2010) [3], has present the history of efforts to develop computer programs (software) for the translation of natural languages. Marta R. Costa-Juss'A et. al. (2012) [4], has analyse the main differences between rule-based and statistical machine translation paradigms in the specific case of Catalan-Spanish pair. Vishal Goyal et. al. (2009) [5], has discussed the machine translation systems for non-Indian languages and second part discusses the machine translation systems for Indian languages Durgesh D Rao (1998) [7], has presents that Machine translation is the study of designing systems

that translate from one human language into another. They introduce the main concepts, issues and techniques involved in machine translation, and looks at some applications.

### III. OVERVIEW OF MACHINE TRANSLATION APPROACHES

Different machine translation approaches has been proposed. In this section, we will provide an overview of few approaches for machine translation. There are different machine translation approaches that are explained in this section. Figure.1 shows the different approaches in the machine translation system.

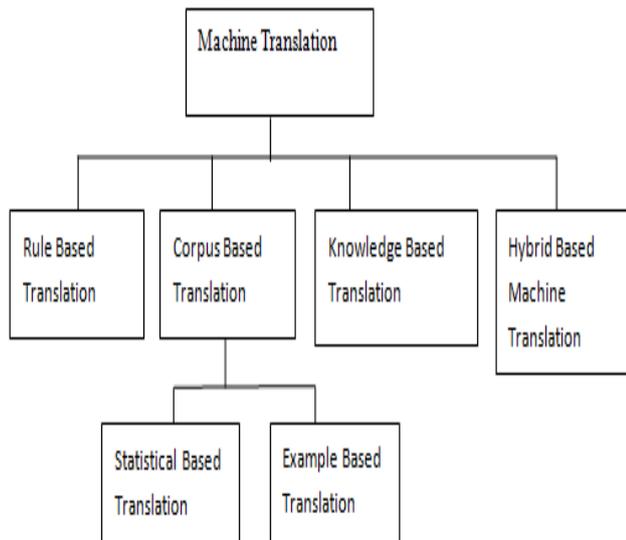


Fig. 1 Machine translation approaches

#### 1 Rule Based Machine Translation (RBMT)

RBMT systems rely on linguistic rules and dictionaries. The linguistic modules in machine translation system are responsible for the *analysis* of the source text, *transfer* between two languages and the *generation* or synthesis of the target text [2].

Rule based system requires: syntax analysis, semantic analysis, syntax generation and semantic generation. The analysis produces a complete parsing of a source language sentence. In the analysis and generation stages, most systems have clearly separated components with different levels of linguistic description: morphology, syntax and semantics etc. Analysis is divided into morphologic analysis, POS tagging, parsing, chunking, dependency analysis. Transfer phase consists of local reordering and long distance reordering. In the final, generation phase have lexical transfer, mapping and agreement.

#### 2 Corpus Based MT

The corpus based MT is divided into two parts: Statistical Based approach and Example Based approach. These approaches are explained as follow:

##### 2.1 Statistical Based Approach

Statistical machine translation (SMT) is a machine translation where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora. The SMT is a corpus based approach, where a massive parallel corpus is required for

training the SMT systems. The SMT systems are built based on two probabilistic models: language model and translation model. The advantage of SMT system is that linguistic knowledge is not required for building them. The difficulty in SMT system is creating massive parallel corpus. SMT systems work well for machine translation of English to European languages

because the word order is almost preserved in such translations. For machine translation of English to Indian languages, the parallel corpora have to be pre-processed (changing word-order) and trained in SMT.

##### 2.2 Example Based MT (EBMT) approach

The example based machine translation (EBMT) is the corpus based approach without any statistical models. The example based systems are trained with the parallel corpus of example sentences, similar to SMT systems. The example based systems generally don't learn from the corpus. They store the parallel corpus and uses matching algorithms to search and retrieve the sentences.

#### 3 Knowledge Based Approach

Knowledge based approach based on taxonomy of knowledge and contains an inference engine. Machine translations are characterized by the syntax and there is little (semantics) analysis.

#### 4 Hybrid Based Approach

Hybrid Machine Translation (HMT) was built due to the limitations of the two approaches and their possibility to be integrated. Statistical Machine Translation and Rule-Based Translation are two MT approaches which work oppositely. SMT did not need to learn about the language at all, while RBMTs basis is gathering language rules. Due to this difference, SMT and RBMT give a different performance. Although there are several forms of hybrid machine translation such as Multi-Engine, statistical rule generation and multi-pass.

### IV. EVALUATE BEST APPROACH IN MACHINE TRANSLATION

These days, Statistical Machine Translation (SMT) is considered to be good approach for the machine translation, as implementation of this approach with tools SRILM, GIZA++ and Moses decoder.

Statistical Machine Translation works on statistical models. These models are language model, translation model and decoder. *Language Model* gives the probability of the sentence which depends on the probability of individual words *Translation Model* computes the probability of source sentence for target sentence *Decoder* decodes the language and gives maximum probability.

**Accuracy:** Accuracy of the system is good enough to understand the sentences after translation.

**Robustness:** The system provides both the robustness in case of failure and no loss of information while translating the text. The output of the system follows the grammar of the source language.

**Statistical Machine Translator:** *Google Translator* is one of the statistical machine translator system which translate Any language pair, for which there are enough training data. It is server-based solutions and having Parallel corpora (at least, 5,000,000 words / 500,000 translation units). Google translator developed by **Franz Josef Ochs** in 2007. Google announced that *Google Translate* translates roughly enough text to fill 1 million books in one day (2012). Statistical model parameters are estimated from bi-lingual and mono-lingual corpora. There are various statistical machine translators that use to translate the language pairs.

## V. CONCLUSION

It is concluded from the above discussion that lot of research is going in the area of NLP and each approach has capability to translate the language pairs with some limitation. Statistical machine translation system translates any language pair if we train the data in the system. It provides maximum accuracy and there is no loss of information when we translate the text.

## VI. REFERENCES

- [1] Sitender, Seema Bawa, "Survey of Indian Machine Translation Systems", *International Journal of Computer Science and Technology*, March (2012).
- [2] Antony P. J. "Machine Translation Approaches and Survey for Indian Languages", *Computational Linguistics and Chinese Language Processing*, March (2013).
- [3] W. John Hutchins, "Machine translation: a concise history", *Journal of Translation Studies*, (2010).
- [4] Marta R. Costa-Jussà, Mireia Farrús, José B. Mariño, José A.R. Fonollosa, "Study and Comparison of Rule-Based and Statistical Catalan-Spanish Machine Translation Systems", *The Journal Computing and Informatics*, April (2012).
- [5] Vishal Goyal and Gurpreet Singh Lehal, Ph.D, "Advances in Machine Translation Systems", 11 November (2009).
- [6] Jörg Tiedemann, "Machine Translation Rule-based MT & MT evaluation", *Department of Linguistics and Philology Uppsala University*, September (2009).
- [7] D.D. Rao, "Machine Translation A gentle Introduction", *RESONANCE*, July (1998).
- [8] S.K. Dwivedi and P.P. Sukadeve, "Machine Translation System Indian Perspectives", *Proceeding of Journal of computer science*, May (2010).