

SURVEY ON MACHINE TRANSLATION AND ITS APPROACHES

SINDHU D.V¹, Dr.B.M.SAGAR², Mr. RAJASHEKAR MURTHY S³

Department of ISE, R V College of Engineering, Karnataka, India¹

Department of ISE, R V College of Engineering, Karnataka, India²

Department of ISE, R V College of Engineering, Karnataka, India³

Abstract: Machine Translation is the branch of computational linguistics which deals the use of computer software to translate from one natural language to another natural language. The research in the field of Machine Translation has been going on from past few decades but the promising translation work is done from 1990s. Even with advanced research in artificial intelligence and computational linguistics the machines fail to produce the promising translation automatically. So to improve the quality of the translation there is need of machine aid or human aid. This paper focuses on mainly four sections. The first section deals with the introduction the machine translation and the second section deals with the brief history and the third section deals with the machine translation approaches and the last section deals with the existing systems.

Keywords: Computational linguistics, Machine Translation, Machine aid, Human aid

I. INTRODUCTION

India is a linguistically rich country. The languages of India belong to several language families, the majority speakers belong to the Indo- Aryan language family, there are around 73 percent speakers of Indo-Aryan language and second largest speakers in India belong to Dravidian language family that is about 23 percent .There are also other languages in India belong to the Austroasiatic, Tibeto-Burman and also there are few isolates. In order to provide the communication between the different language families there is a need for the translation.

There is a lot of research in the area of machine translation is going on in India for last two decades. Due to the advance in the computational linguistics and artificial intelligence the efficient translation systems are in market during 1990s. Since it has more languages still more work is going on to provide the communication bridge between different language speakers.

The term machine translation can be defined as the computerized system capable of translating from natural

language into another natural language. Machine translation is the branch of computational linguistics which deals with the use of software's for natural language processing. Even with the advance of the processors still machines are failing to provide the perfect translation. So, the machines need human assistance to provide the perfect translation this kind of translation is called as Machine Aided Translation [1].

Fully automatic translation doesn't guarantee the promising translation because the processors are not intelligent to solve the ambiguities. So to provide the promising translation human assistance is required that is for few documents post editing and pre editing are done to satisfy the rules defined by the system. Hence, Machine Aided translation can be defined as the translation from one natural language to another natural language with human assistance [2].

The next section describes the brief history of the machine translation and the following section describes the machine translation approaches and the existing systems.

II. HISTORY

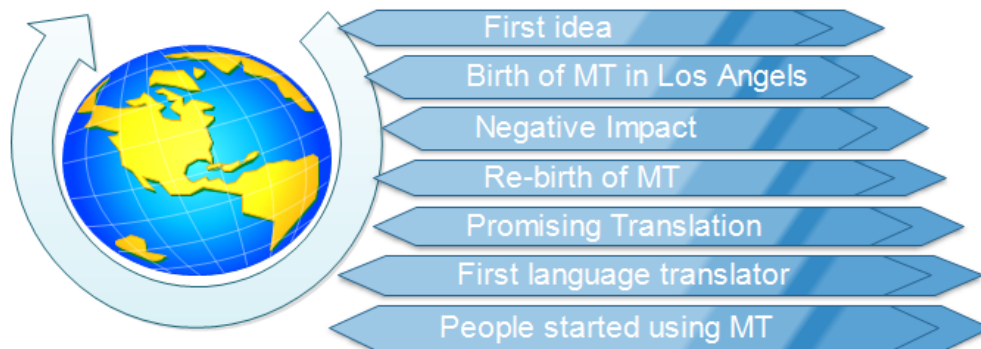


Fig.1 Evolution of MT

During the first decade of translation in the 1940s, there was lot of interest and ideas towards the mechanical translation which uses the mechanical dictionaries of the particular language. But the mechanical translation was failed because of the waste of time and resources in building the mechanical dictionaries and there were many ambiguities which were not solved in mechanical translation. This is the major reason for which the digital computers were turned.

In the year 1949, Warren Weaver proposed the idea of using the digital computers for the translation which gave the birth for the name computer translation now which is termed as a Machine Translation.

During the second decade of translation in the 1950s, many intelligence communities and military in the US started funding for the machine translation projects and also the US government appointed the full time researcher Yehoshua Bar-Hillel, who made the future research outlines very clear. In the year 1952 the first symposium of machine translation was held under the leadership of Yehoshua Bar-Hillel. In the year 1954 the first automatic Russian-English translator was developed by the group of researchers from Georgetown University Leon Dostert in collaboration with IBM. This translator was restricted with vocabulary of 250 words and six grammar rules and in the first demo the translator translated selected sample of 49 Russian sentences into English. In the same year the first journal on the Machine Translation was published, entitled Mechanical Translation.

During the third decade the sponsors of Machine Translation in United States formed the Automatic Language Processing Advisory Committee (ALPAC) to examine the progress of the Machine Translation. In the year 1966 the ALPAC committee released the progress report of Machine Translation stating that Machine Translation was waste of time and money. This negative impact on Machine Translation made funding

organizations to stop their financial support to the researchers; this affected the research on Machine Translation for the few years [2].

During the fourth decade the MT research started outside the US. In the year 1970 the Russian researchers gave rebirth to the Machine Translation by starting the project REVERSO. In the same year researchers of Georgetown University developed the system SYSTRANI which is capable of translating from Russian to English. A research group at Montreal in 1976 developed a system named METEO for translating weather reports for public broadcasting. In the year 1978 the Japanese developed a system based on the rule based approach which works on the few listed rules by the system and this system was able to translate from Korean to Japanese and vice versa.

During the fifth decade the Japanese researchers started developing more systems. In the year 1986 Hitachi developed a translation system Hitachi Computer Aided Translation System (HICAT) which was capable of translating from Japanese to English. They also developed a rule based translator PENSEE, which was also translating from Japanese to English based on the predefined rules in the system. During this decade the researchers were successful in providing the promising translation [2].

During the sixth decade the research on Machine Translation was happening worldwide and they were successful in providing the promising translation systems and the researchers started to build the fully automatic translation systems for the different pair of languages. In the year 1993 the project Consortium for Speech Translation Advanced Research (C-STAR) was started. The system was trilingual it supports English, German and Japanese. This project was defined for the tourism domain. During the seventh decade the people started using the machine translation worldwide. In the year 2005 the Google launched a first website for automatic translation.

III. MACHINE TRANSLATION APPROACHES

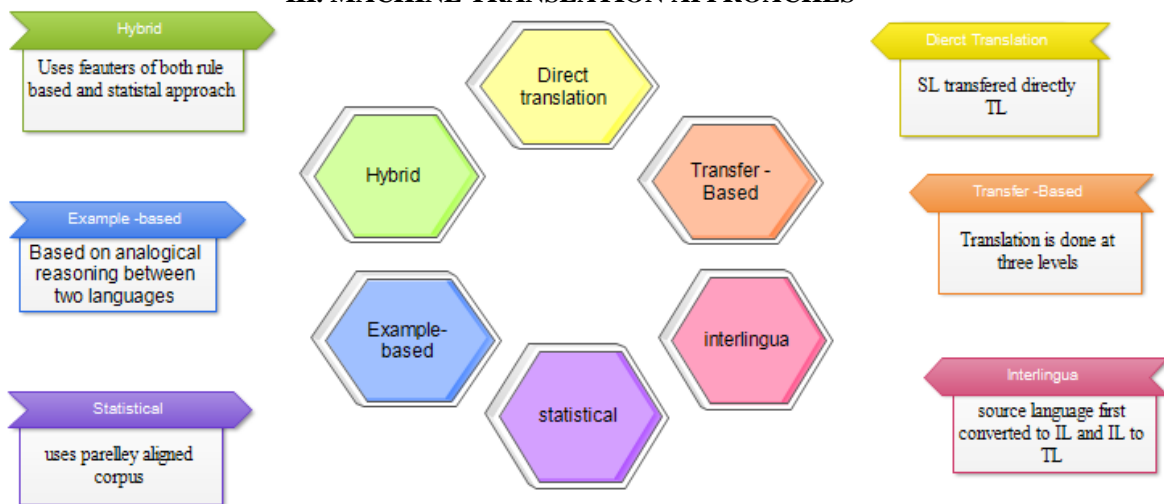


Fig 2. Approaches of MT

III A. Direct translation

In direct translation, the source language is transferred directly into another language. In the direct translation or literal translation word to word translation is done with or

without preserving the sense of the word. For the direct translation it requires both source and target language dictionaries with high quantity and quality and also morphological analyzer. Since direct translation is like

paraphrase the translator looks in to the target language dictionary and replaces the equivalent word without considering any grammatical rules. So the performance of the direct translation depends on the quality of the resources [2][3].

III.B. Transfer –based translation

In the direct translation we can't preserve the meaning of the sentence to overcome this advantage and to produce the promised translation the transfer based translation is proposed. In the transfer based translation, to preserve the meaning of a sentence a set of linguistic rules are listed by analyzing the grammatical structures of both the source language and the target languages. In transfer based translation happens in a three stages: 1) Analysis, 2) Transfer, 3) Generation. In the analysis stage, the source language parser produces the syntactic representation of source language. In the transfer stage, syntactic representation of source language text in to the syntactic representation of the target language. In the generation stage, the morphological analyzer is used to generate the final target language text.

III C. Interlingua –based translation

In the Interlingua translation, first the source language text is converted in to an intermediary language called Interlingua then Interlingua is converted in to the target language. Interlingua can be used for the multiple languages if the structure of the language is same. Interlingua is defined in the way describing the analysis of the source text so that it is possible to convert its morphological, syntactic and semantic characteristics in to meaning full translation[3].

III D. Statistical –Based Approach

In the statistical approach, the source language text is transferred in to the target language text based on the statistical models extracted from the corpus of both the source language and target language. Supervised or unsupervised statistical machine algorithm is used to build statistical tables from the corpora, these statistical tables consists of the statistical information such as

characteristics of the sentences and the structural relation between the two languages. These characteristics are used as statistical models in the statistical approach for translating from one natural language to another language [2].

In statistical translation the document is translated according to the probability distribution function indicated by the $p(e/f)$, this is the probability of translating sentence f in the source language to a sentence e in the target language. There are many ways for handling the probability function.

III E. Example-Based Translation

In the example based translation, a system is defined which contains set of source language sentences and corresponding target language sentences. During the run time, example based translation use bilingual corpus as its database. This database is stored in the translation memory. In translation memory, the user translates text these translations are added to a database, and when the same sentence occurs again during the translation, the previous translation is inserted in to the translated document. The advantage of the example based translation the translation memory saves the user effort of re translating the sentence and this saves the processor time and also the user time. The disadvantage of example based translation is the wastage of memory to store the database.

III F. Hybrid Translation

With the advantages of the statistical based approach and rule based approach, a hybrid approach was developed. The hybrid based translation can be done in different ways. In few cases translations are performed first using rule based approach later the output is produced by correcting the output using the statistical approach and in few cases the rules are used to preprocess the input data as well used to modify the output.

In the hybrid approach of the machine translation the drawbacks of the both the approaches were eliminated to provide the promising translation with high efficiency [3].

IV EXISTING SYSTEMS

The table describes the existing systems with the machine translation approaches used to develop the system.

Table 1: Existing systems of MT

| SL NO | TRANSLATION SYSTEM | SOURCE TARGET LANGUAGE | YEAR OF DEVELOPED | DOMAIN | APPROACH USED |
|-------|-----------------------------|---|-------------------|----------------------------------|-------------------------------------|
| 1 | Anusaaraka | {Telugu, Kannada, Bengali, Punjabi and Marathi to Hindi | 1995 | For translating children stories | Direct machine translation approach |
| 2 | Punjabi to Hindi translator | Punjabi to Hindi | 2007,2008 | General | Direct machine translation |
| 3 | Hindi to Punjabi | Hindi to Punjabi | 2010 | Web pages and email | Direct machine Translation |
| 4 | Hindi to Punjabi | Hindi to Punjabi | 2009,2011 | General | Direct machine translation |

| | | | | | |
|----|---|---|-----------|-------------------------|----------------------------|
| 5 | Mantra | English to Hindi | 1997,1999 | For Rajyasabha | Transfer based approach |
| 6 | English – Hindi translator | English to Hindi | 2002 | Weather forecasting | Transfer based approach |
| 7 | Mat | English to Kannada | 2002 | General | Transfer based approach |
| 8 | Shakti | English to IL | 2003 | General | Transfer based approach |
| 9 | English to Telugu | English to Telugu | 2004 | General | Transfer based approach |
| 10 | Anglabharti | English to IL | 2001 | Health | Interlingua based approach |
| 11 | Unl based English – Hindi MT | English to Hindi ,UNL to Hindi and vice versa | 2001 | General | Interlingua based approach |
| 12 | Anglahindi | English to Hindi | 2003 | General | Interlingua based approach |
| 13 | English to Indian languages MT | English to Indian languages | 2006 | Health care and tourism | Statistical based approach |
| 14 | Anubaad | English to Bengali | 2000,2004 | News | Example based approach |
| 15 | Vaasaanubaa | Bengali to Assamese | 2002 | News | Example based approach |
| 16 | Shiva and Shakti | English to {Hindi, Telugu, Marathi } | 2003 | General | Example based approach |
| 17 | Anglabharti-ii | English to Indian languages | 2004 | General | Example based approach |
| 18 | Hinglish | Hindi to English | 2004 | General | Example based approach |
| 19 | English to {Hindi , Kannada, Tamil} System | English to {Hindi, Kannada ,Tamil } | 2006 | General | Example based approach |
| 20 | Matrex | English to Hindi | 2008 | Conference papers | Example based approach |
| 21 | Anubharati | Hindi to IL | 1995 | General | Hybrid approach |
| 22 | Anubharati-ii | Hindi to IL | 2004 | General | Hybrid approach |
| 23 | Bengali to Hindi MT system | Bengali to Hindi | 2009 | General | Hybrid approach |
| 24 | Lattice based lexical transfer MT framework | Bengali to Hindi | 2009 | General | Hybrid approach |

V CONCLUSION

The survey describes the research done in the machine translation for few decades worldwide. Additionally we tried to describe the machine translation approaches and last section describes the existing machine translation systems and the approaches used to develop the systems and also the purpose for developing the system.

REFERENCES

- [1] GV Garje and G K Kharate “Survey Of Machine Translation Systems In India”, International Journal On Natural Language Computing, Vol.2, No.4,October 2013.
[2] Antony P.J “Machine Translation Approaches And Survey For Indian Languages” , Computational Linguistics And Chinese Language Processing , Vol.18 ,No.1 , March 2013 ,pp.47-78.

- [3] Sanjay Kumar Dwivedi and Pramod Premdas Sukhadeve “Machine Translation System In Indian Perspectives” , Journal of Computer Science vol.6 No.10 ,2010 pp.1111-1116.