

A Survey on “Machine translation Approaches for Indian Languages”

Pooja P. Walke^a, Dr FarhaHaneef^b

^aResearch scholar, CSE, Oriental University, Indore, India. Assistant Professor, St Vincent Pallotti College of Engineering & Technology, Nagpur, India.

^bAssociate Professor, Department of CSE, Oriental Institute of Technology & Research, Bhopal, India.

^apwalke@stvincentngp.edu.in, ^bfarhahaneef2014@gmail.com

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: Translation has always helped India to knit Indians together with respect to its rich culture and literature. Ideas and concepts like ‘Indian ancient literature’, ‘Indian rich culture’, ‘Indian philosophy’ and ‘Indian knowledgeable systems’ would have been impossible in the absence of translations with their natural integrationist mission. Machine Translation assist to translate Information presented in one language to other language. Information can be present in form of text, speech and image translating this information helps for sharing of information and ultimately information gain. Translation process is an extremely complex & challenging process. It requires an in-depth knowledge about grammar of both the languages i.e. Source language and Target language to frame the rules for target language generation. Marathi is a regional Indian language and consists of a lot of literature that could be useful if projected in the universal English language. As manual translation is a tedious task, we propose a literature survey about machine translation systems that translates Indian Languages into English Language using various Machine translation approaches like RBMT, SMT, NMT, Hybrid translation

Keywords: Machine Translation System(MTS), Marathi Language Translation, Rule-Based Machine Translation(RBMT), Statistical Machine Translation(SMT), Neural Network Machine Translation(NMT), Hybrid Machine Translation.

1. Introduction

Language is one of the mode to communication. Language helps us to share our thoughts/vision with other people. Human beings are a very much attach with the community. We love to share our information with other at the same time we love to know about other people. India is a country where people uses multiple languages as per their culture or region. Being a multilingual country we all know that India is having a very rich culture and literature. But sometimes it is observed that the language which is to be used for communication becomes a barrier for sharing of thoughts or

ideas. This arise the need of language translation so that people with different language can easily communicate their thoughts and can understand other person’s thought. Language translation become very important and needful when it comes for literature also. One can easily access literature of other language if translation is applied. It is observed that much of work is done for English, Hindi, Tamil, Bengali languages but till now very few work is carried out for Marathi language. So i propose a Machine Translation System which will accept the source language as Marathi and will generate the appropriate English language translation. Their are various language translation approaches like RBMT, SMT, NMT, Hybrid translation. One can use either of the approach for doing the translation.

2. Literature Survey

PramodSalunkhe et al.[1] The author shared the his views on how the language is becoming barrier in spreading the literature available in various languages. Language is for communication but if its becoming the hurdle than we need to translate the source language to target language in order continue the spreading of information. Now a days English became a global language and that’s why majority of official documents are available in English language only. But it becomes difficult for the user who do not understand English and than his/her work gets hamper. Hence translation of English to any other Indian or foreign language will help people to understand the official documents or available literature. It is observed that lots of translation work for English to Hindi, Tamil Bangla and other foreign languages are done but English to Marathi language translation is still missing. So there is huge scope in this language pair translation. This will open the access to official documents and literature for a Marathi person. The author also shares the comparative study of various MTS approaches like SMT,

Rule-based and Hybrid approach. The author suggested the user interface tags for web pages translation from English to Marathi and proposes the Rule-based system and Hybrid system.

Sandeep Saini et al.[2] The author provides the survey regarding various MTS as the scope for language translation is huge in the country like India where multiple languages are spoken by people. The author presents the work done for Hindi language and also claims that very few work has been done for South Indian languages. So one can use transfer based MTS as they provides more flexibility and supports multilingual environment. It also covers the inefficiency of MTS with respect to rules, translation methodology, dictionary. It shows the facts where we can improve the MTS for future work.

Stephen Doherty et al.[3] The author gives the study of effects of technology on language translation and society. The author also discussed regarding the technological growth of TM and MT. He also shed the light on the usage of some renewed MTS like Google and Microsoft's Bing. The author focuses on the two approaches for contemporary language translation: computer-assisted translation(CAT) tools and machine translation system(MTS). The CAT uses the translation memory as a core of translation where one can have the source input as well as targeted output side-by-side. MTS provides various approaches like SMT, Rule based, EBMT etc. One can select the required approach as per their application need.

Pankaj Kumar et al.[4] The author aims the translation of English language input into appropriate Hindi language output in a stepwise manner with the help of Machine Translation approach. MT being a part of NLP allows the researcher follow their selected MT approach. In The author the author had used a stepwise approach in which the English input text will be stored in a file, then extraction of words and punctuation will be done by storing them in an array. Once this is done then the grouping and sequencing of words will be done as per the grammatical structure of the target language ie Hindi and than this will be converted into target language. The researcher also address the issues like ambiguity, idioms, word sense etc and is also managing the dictionary for the words whose meaning is not confirm.

Na Li et al. [5] The researcher focuses on the automatic evaluation metrics by using Machine Translation approach for natural language translation from source to target language. It is capable to defeat the errors in translation and can enhance its development. The paper represents following

3 automatic evaluation methods 1- Vocabulary-based Automatic Evaluation Method for MT, 2- Syntax-based

Automatic Evaluation Method for MT, 3- Semantic-based Automatic Evaluation Method for MT and provides the comparative study with respect to their advantages and disadvantages. The author analyzes the MT automatic evaluation metric by using extended reference.this method works on different levels of lexical form to deep syntactic and grammatical information and has received good accuracy in the result.

Mangala Madankar et al. [6] The author focuses on the Information Retrieval areas like cross-lingual (CL) , Multi-lingual (ML)IR for retrieval of documents from the local database. It also shed a light on how Machine translation approaches and its techniques can be useful for Information Retrieval in CLIR & MLIR. Machine translation is one of the active research area in the domain of AI and IR. In general its difficult to translate the natural languages from source to target language. This hard problem can be solved by using various MT approaches and techniques so the author presents some of the latest research work in the area of crosslingual and multilingual information retrieval.

Nagmani Wanjari et al. [7] The author focuses on the identification of end-of-sentence correctly for Marathi language. It is observed that lots of work has be done on foreign languages but very few work is done for Indian language. The biggest complexity in doing this task is that in English language a new sentence starts with a capital letter which is not available in Marathi language. So by using Rule based approach multiple rules have been design to identify the end-of-sentence correctly. Different tools like Tokenizer, CoreNLP, Lingpipe, MxTerminator etc are used for English language along with the approaches like Rule based approach, Stazsystem, Punkt system, Maximum Entropy approach etc. Same kind of work has been done for several foreign and Indian languages like French, German, Spanish Malayalam, Kannada, Bengali etc and received distinct accuracy. Unfortunately no such work is yet done for Marathi language and hence author proposed a rule based system approach for Marathi language. As rule-based approach allows us to design the rule as per the grammatical syntax for the formation of sentence. These rules will help in identification of correct end-of-sentence.

Omkar Dhariya et al. [8] The author came up with an new Hybrid MTS approach by combining the EBMT, SMT and RBMT. This approach works majorly on Rule based MT after deriving the the set of partial translation

received by the subsystems EBMT and SMT.It also shares the comparison of proposed system with Google, Bing and Babylonian translators.