A Review Study of Machine Translation Systems for Indian Languages and Their Issues

Bhanudas Panchbhai and Varsha Pathak

April 6, 2022

A Review Study of Machine translation systems for Indian languages and their issues

Bhanudas Suresh Panchbhai[1]

[1]Department of computer science, R.C. Patel Arts Commerce and Science College, Shirpur, India

Bharat.panchbhai@gmail.com

Dr.Varsha Makarand Pathak[2]

[2]Department of computer application, KCE'S Institute of Management and Research, Jalgaon, India

pathak.vmpathak.varsha@gmail.com

Abstract

The main objective of this study is to see how far machine translation in Indian languages has advanced. The three types of machine translation methodologies examined in the survey are rule, empirical, and hybrid-based approaches. It has a set of benefits and drawbacks all of its own. The primary purpose of machine translation is to bridge the linguistic divide between two people, companies, or countries by converting one natural language into a new one using automated computing tools. Machine translation (MT) is difficult because it entails a variety of difficult subtasks, such as ambiguous language, linguistic intricacies, and differences between source and target languages. The aim of the research is to give a general overview of machine translation in Indian languages. The current condition of machine translation on a nationwide scale was investigated in this study.

Keywords

Machine Translation, National Language Machine Translation, International Machine Translation, Indian language

1. INTRODUCTION

Language is a powerful and systematic way of communication that allows people to convey their thoughts, feelings, wishes, and other emotions. According to [1] the world has around 6,800 living languages. As a result, We must create tools for translating information from one language to another in order to access information given in another. The process of changing the meaning of a text from one language to another for a new relationship is known as "translation."

Machine conversion research began in the 17th century with the purpose of using dictionaries to bridge linguistic divides across different parts of the world. Studies in India began a little later, nearly two decades ago. Due to their morphological variation and dozens of dialects, machine translation for Indian languages is a little more complex than for other languages due to machine translation. In the corporate world, the language issue is a stumbling block to the expansion of international business deals. It's difficult to get into worldwide trade since global audiences try to read and speak in their native tongues. To begin with, human translators were used, but they couldn't keep up with

the demand for translated content. Due to the limits of human translation, machine translation was developed. Language computing technologies may now be used to improve products and services in high-growth markets [2], such as health care, mobile communication, online retail, call centers, the media, and publishing, thanks to MT research.

2. Machine Translation Techniques

There are several approaches to machine translation [6]. The points that will be discussed using **Fig. 1** are as follows.
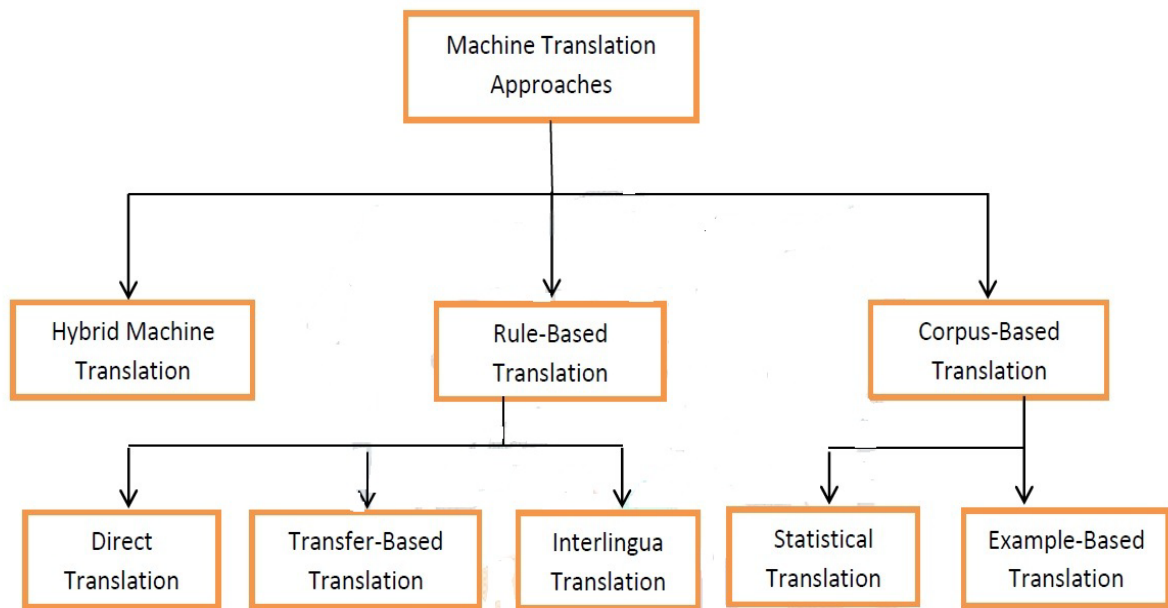

Fig. 1. **Machine Translation approaches**

2.1 Machine Translation using Rules

The earliest commercial machine translation systems were Principles-Based Machine Translation (RBMT) systems, which are based on linguistic rules that allow words to be placed in multiple contexts and have different meanings [6]. Language proficiency is required for this strategy. An expert system, for example, could assist a clinician in making an accurate diagnosis based on a group of symptoms. Rule-based machine translation can be divided into three types: direct machine translation, transfer-based machine translation, and Interlingua machine translation.

2.1.1 Using a direct method

One of the most fundamental ways of machine translation is direct machine translation. Direct machine translation [5] employs a bilingual dictionary to do a literal word-for-word translation of the source material, followed by some syntactical reorganization. Borrowing and calque are two direct translation procedures.
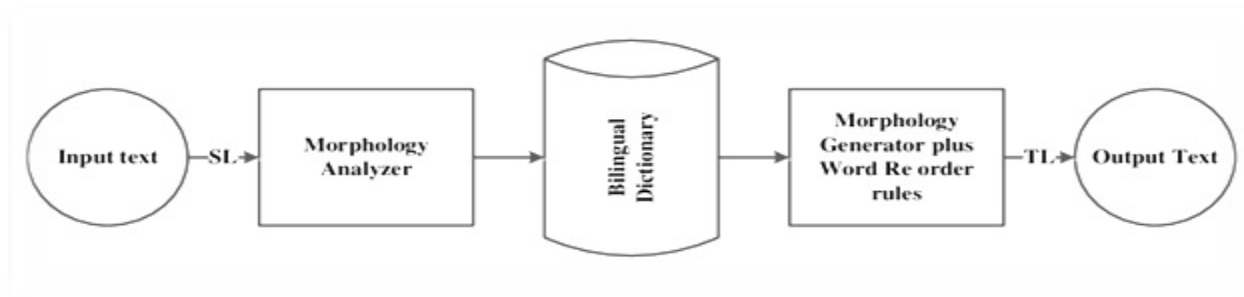

Fig. 2. **Direct MT system**

2.1.2 Transfer Based MT

Using a database of translation rules, this method converts source text to the other. When a sentence is explicitly translated using a dictionary, it follows one of the rules or is an illustration of one of the rules. This creates as or to Interlingua by performing morphological and syntactic analysis on the source language's basic forms, then translating them to the destination language's fundamental forms and enriching the translation [7].
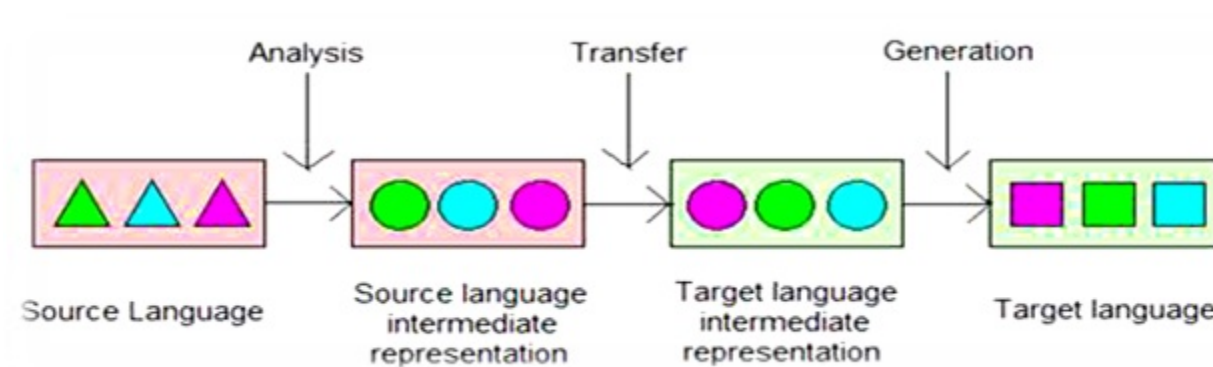

Fig. 3. Transfer Based System

2.1.3 Interlingua Based MT

Another traditional method of computer translation is Interlingua machine translation. This is a transfer technique that is a more efficient alternative to straight translation. The source language is translated using Interlingua, a language-independent intermediate abstract language [6].

Fig. 4. Interlingua language systems

2.2 Corpus Method

Sentence-aligned parallel text is required for each language pair in corpus designed method systems. Additionally, Additional classifications are carried out for corpus built statistical research and example-based machine translation systems [5].

2.2.1 MT based on statistics

In 1949, Warren Weaver proposed the concept of statistical machine translation. This methodology employs statistical methodologies to create translated forms from bilingual corpora. Factual translation models are used in statistical machine translation, and the parameters are derived using corpora in monolingual and bilingual languages. The development of statistical translation models is straightforward, but it is heavily reliant on existing multilingual corpora for innovation [4].

2.2.2 Example Designed Method

Example-built systems produce translations for a given input based on past translation examples. When the system receives a sentence as input, it looks up a similar source sentence in the example database and translates it. t is extremely difficult to create a parallel corpus with minimum resources, and the results are unpredictable depending on the quality of the resources available [5].
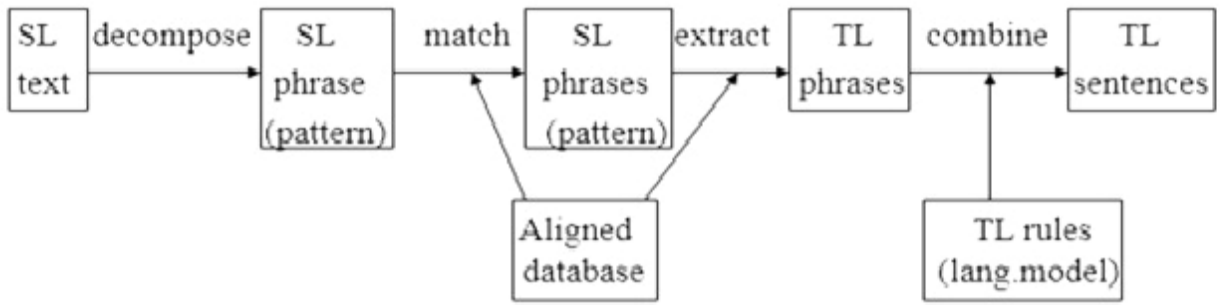
Fig. 5 .Example Based MT System

2.3 Hybrid MT

The hybrid MT method integrates statistical and rule-based techniques to create a revolutionary hybrid approach that tackles the shortcomings of both systems while outperforming current systems. Many businesses now adopt a hybrid approach that combines data with user-defined criteria. Rule-based translations are employed in a few circumstances, with subsequent output dependent on statistical data or the other way around [5]. Hybrid-based MT is more versatile than others in terms of performance enhancement.
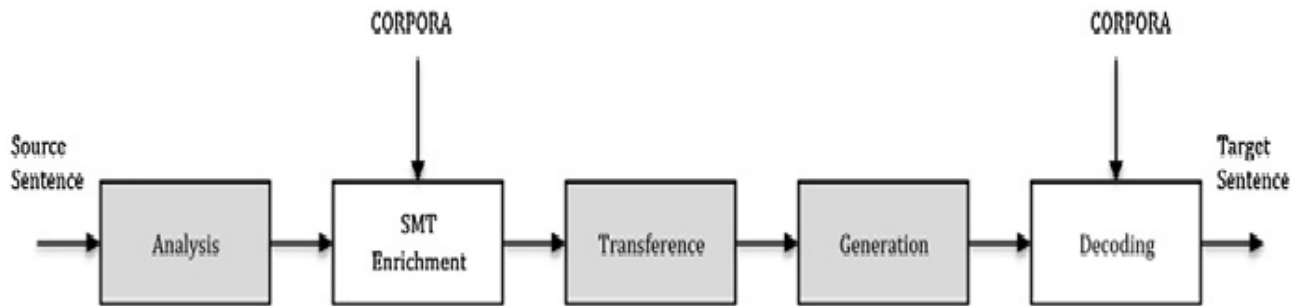
Fig. 6. Hybrid MT System

3. Approaches to Machine Translation's Strengths and Weaknesses

As previously stated, several machine translation techniques exist, each with its own set of advantages and disadvantages. In different languages, MT systems are used in a variety of ways, depending on the needs and structure of the language. The benefits and drawbacks of each technique are summarized in **Table 1.**

Table 1: Differences in MT Approaches

Sr. No.	Machine Translation Approach	Advantages	Disadvantage
1.	**Direct method**	1. The translation is simple to grasp and requires little effort on the part of the reader.	1. It does not merely focus on the lexical structure of the word and its ties to other words because it requires analysis. 2. It is designed particularly for the development of certain language pairs and is not suitable for other language pairs. 3. Developing in several languages is more expensive.
2.	**Interlingua system**	1. It works well with a domain-based approach. 2. It is especially useful in multilingual environments.	1. There is decreased time efficiency when compared to straight machine translation. 2. It's tough to turn those results into a meaningful statement using the intermediary representation. 3. Due to cultural and organizational variations, representing a multitude of languages is difficult.
3.	**Transfer based**	1. In a range of languages, the system can deal with ambiguity. 2. It lays the groundwork for the modular framework.	1. The text's significance may be lost by the time it reaches the end.
4.	**Dictionary based**	1. It aids in the human translation process by improving the sentence's syntax and grammar.	1. less applicable in sentence translation
5.	**Example based**	1. It is ideally suited to languages with similar structural similarities.	1. It necessitates a large bilingual corpus of the language pair for which the translation is required.
6.	**Statistical based**	1. It was not built for a specific language pair, hence it is language agnostic. 2. It can be used with any language pair. 3. It is less costly than a rule-based approach. 4. As they were trained by real-time messages, translations used to be natural.	1. It necessitates a lot of hardware configuration. 2. A domain-based translation necessitates a corpus of at least a million words.

4. Indian Language Machine Translation Systems

Since 1991, machine translation has been a popular research area in India. The first research was done at IIT Kanpur, and it has subsequently been repeated at several other institutes. We'll look at some big national (Indian) language MT efforts in this section. The following are the major parameters that we will discuss: The name of the application and the year it was published for the MT System. The status levels of MT languages in this paper are state, national, and international.

Table 2. Shows the target and source translations of a range of national languages.

Table 2. Source or target language translation system for "Marathi", "Hindi","Urdu"							
Sr. No	Author	System of Machine Translation	Year	Translational languages	Application /Domain	Approach Used	Observations
1.	Manoj Chinnakotla, Om P. Damani	Evaluation of cross-language information retrieval from Hindi and Marathi to English [8]	2007	Marathi - English and Hindi – English	A system that allows you to search for information in several	Dictionaries that are bilingual	The CLIR Hindi-to-English and Marathi-to-English systems, respectively, A simple rule-based technique is employed to translate terms that aren't found in dictionaries.
2.	Abbas Malik ,Laurent Besacier Christian Boitet, Pushpak Bhattacharyya	A Hindi-to-Urdu Transliteration System [27].	2009	Hindi–Urdu	The process of transliteration from Hindi to Urdu is quite precise.	Rule Based	Language barriers between Hindi and Urdu have been overcome by developing special rules and utilising resources such as the Hindi spell checker. Multiple/zero character mappings, pronunciation and orthography variations, proper noun transliteration, Urdu word boundary, and other issues have been addressed.
3.	Nadir Durrani, HassanSajjad	Machine Translation from Hindi to Urdu Using Transliteration [9]	2010	Hindi – Urdu	General	Based on Statistics	This system proposes two innovative probabilistic models for the problem, based on conditional and joint probability formulations.
4.	Bushra Baig ,M. M. Kumar , Sujoy Das,	Transliteration system from Hindi to Urdu based on rules [10].	2012	Hindi – Urdu	General	Rule Based	Some issues, such as unclear characters and nukta-related errors, have been overcome by developing unique rules and utilizing a database.
5.	S. B. Kulkarni, D.Deshmukh,K. V. Kale	Machine Translation of English to Marathi: [11]	2013	English-Marathi	General	Translation Divergence Classification	We looked into the topic of translation divergence classification for MT between English and Marathi. Established that the discrepancy in translation between.
6.	AbhayAdapanawar,AnitaGarje,Paurnima Thakare,Prajakt Etc.al.	Assertive Sentence Translation From English To Marathi Based On Rules[12]	2013	English-Marathi	General	Rule Based	The developer is working on rule-based English to Marathi aggressive sentence translation. A bilingual dictionary is used in this method.
7.	SudhaMorwal ,NusratJahan	An Experimental Result on Hindi, Urdu, and Marathi Languages [13].	2013	An Experiment on the Languages of Hindi, Urdu, and Marathi	General	Machine learning-based approach, linguistic	The main goal is to use a Hidden Markov Model (HMM) to conduct Named Entity Recognition in Natural Languages and to propose strategies to improve accuracy and performance metrics (precision,

						approach	recall, and F-measure).
8.	Piyush Dungarwal, Rajen Chatterjee, Abhijit Mishra,	System for translating Hindi into English [14].	2014	Hindi to English	General	Statistical Based	The usage of parameters such as number, case, and Tree Adjoining Grammar Information (TAGI) aids in improving English-Hindi translation.
9.	Ruchika b.Sinhal	An English-to-Hindi Sentence Translation System [15]	2014	English to Hindi	General	Example Based	This study focuses on a simple technique to extract the translation by comparing sentences.
10.	S. B. Kulkarni , P. D. Deshmukh , M. M. Kazi, K. V. Kale	Linguistic Divergence in English-to-Marathi Translation[16] Patterns of Linguistic Divergence in English-to-Marathi Translation	2014	English to Marathi	General	Translation of lexico-syntactic divergence	Be aware of the various types of divergence problems that can occur when translating from English to Marathi. This topic has been examined from several perspectives, and a variety of solutions have been proposed to address it.
11.	G V Garje, G K Kharate,Hars hadKulkarni	An English-to-Marathi Machine Translation Method [17]	2014	English to Marathi	Transmuter	Rule Based	The core mechanism for identifying the correct word order in the target language was designed based on certain traversals of the parse tree. One of the system's distinctive characteristics is a word sense disambiguation model.
12.	Amarpreet Kaur, Jyoti Rani	A Statistical Machine Translation System for Punjabi to Hindi [18].	2015	Punjabi– Hindi	General	Machine Translation Using Statistics	The framework for resolving the challenges is provided by algorithms such as the Unigram algorithm and the N-gram string matching technique. Subjective assessments, such as the intelligibility and accuracy tests, were used to assess the system's accuracy. This system can also be used in reverse.
13.	Sanjay Dwivedi and Pramod Sukhadeve	English to Hindi [19]	2015	English–Hindi	Homoeopathy	Rule Based	The grammatical rules for an English-to-Hindi machine translation system will be outlined in this study, which will be used to translate homoeopathic literature, medical reports, prescriptions, and other documents.
	Akanksha Gehlot,	Machine Translation				Machine	This system parses an input text to determine its structure. The text in

14.	Vaishali Sharma, Shashipal Singh	System for Hindi to English Transfer [20].	2015	English–Hindi	General	Translation Based on Transfer	Target language is generated using reordering principles.
15.	Jayashree Nair, Amrutha Krishnan K, Deetha R	An Efficient English-to-Hindi Machine Translation System Using a Hybrid Mechanism [21]	2015	English to Hindi	General	Hybrid	Machine translation from English to Hindi Declension rules are used to develop systems. They offered a useful methodology, and recommended a novel approach to MT system design that has yet to be explored in any of the existing MT systems.
16.	Goraksh V. Garje,	Sentences of Assertion and Interrogation [22]	2016	Marathi–English	Assertive and interrogative statements must be translated.	Rule Based	The suggested system's main purpose is to create software that will convert Marathi simple assertive and interrogative sentences into English sentences. To produce higher-quality translations, the system will employ shallow parsing, bilingual lexicon, and rearrangement algorithms.
17.	PramodSalunkhe,Aniket.D. Kadam	From English to Marathi: A Machine Translation Evaluation [23]	2016	English-Marathi	Agriculture-related web pages and text documents have been translated.	Hybrid Translation	Different ways of machine translation have been discussed by the developer. And there's a difference in divergence. The author has developed UI tags for web page translation, which suggests a hybrid technique that uses the RBI portal to construct a multilingual dictionary and a C parser.
18.	NitinBansal, Dr. Ajit Kumar	Punjabi and Urdu Language Machine Translation Survey[24]	2017	Urdu-English, Punjabi, Hindi	Survey	Different approach	Examine the various types of machine translation systems for Punjabi and Urdu languages, as well as the tools for converting source language text into target language text for regional and international languages and the various methods for calculating the accuracy of the translated output of systems designed for Punjabi and Urdu languages.
19.	Sudha Morwal , Nusrat Jahan	Validation and Reliability of the 120-Item Big Five IPIP Personality Scale in Urdu [25].	2017	Validation and Reliability in Urdu	Item Pool of 120 International Personalities (IPIP)	Model of Darwish translation	Researchers generated a 120-item Urdu version of the International Personality Item Pool using the Darwish translation method (IPIP). The translation was reviewed by a panel of engineering professionals as well as Urdu and English language experts. In addition, the internal consistency, reliability, and construct validity of the Urdu version were investigated empirically.

| 20. | I.Ahmed Khan, Babar Nazir, Syed | Sindhi to Hindi Sentence Translation System [26]. | 2018 | Sindhi – Hindi | General | Example Based | This research focuses on a simple way of comparing sentences to determine the translation. The system made use of the training algorithm. |

5. Conclusion

Although machine translation has been around since the dawn of computers, it has only lately been a hot topic in natural language processing research. Machine translation is a difficult problem for Indian languages because of their complex language patterns. According to the survey for Indian languages, many Indian systems are based on statistical and hybrid methodologies. The current research examines the numerous language translation systems that have been developed in India and around the world, each with its own unique methodology. The study's source and target languages are Hindi, Marathi, and Urdu. The direct approach of translation has been discovered to be most suited for closely related languages, i.e., those with similar structures. For languages with a wide range of dialects, indirect and statistical approaches are beneficial.

6. References

[1]. Masaru Tomita & Jaime G. Carbonell. (April-1986). Carnegie Mellon University "Knowledge-Based Machine Translation, The CMU Approach", Institute of Software Research, School Computer Science. PA 15213.

[2]. Abiola O.B, Adetunmbi A.O and Oguntimilehin A. (June-2015). Department of Computer Science AfeBabalola University, Ado Ekiti, Nigeria and Department of Computer Science, Federal University of Technology Akure, Nigeria: "A Review of the Various Approaches for text to Text Machine Translation", International Journal of Computer Applications: Volume 120-No.18.

[3]. Shachi Mall, Umesh Chandra Jaiswal. (July-2018). "Survey: Machine Translation for Indian Language". Volume 13, Number 1 pp. 202-209.

[4]. Neeha Ashraf, Manzoor Ahmad. (Sept.2015) "Machine Translation Techniques and their Comparative Study", International Journal of Computer Applications (0975 – 8887) Volume 125 – No.7.

[5]. Nakul Sharma. (Feb-2020) "English To Hindi Statistical Machine Translation System" Thesis Submitted In Partial Fulfillment Of The Requirements For The Award of Degree.pp.1-7.

[6]. Sandeep Saini, Vineet Sahula (2015) "Survey of Machine Translation Techniques and Systems for Indian Languages" IEEE International Conference on Computational Intelligence & Communication Technology.978-1-4799-6023-1/15.

[7]. Mirjam Sepesy and Gregor Donjay (Sept.-2019) "Machine Translation & the evaluation of its Quality".

[8]. Manoj Chinnakotla, Om P. Damani. (2007) "Hindi to English and Marathi to English Cross Language Information Retrieval Evaluation", Conference paper, Research Gate. Advances in Multilingual and Multimodal Information Retrieval pp 111–118.

[9]. Nadir Durrani, Hassan Sajjad. (Nov.-2010) "Hindi-to-Urdu Machine Translation through Transliteration", Research Gate. Conference: ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics.

[10]. BushraBaig ,M.Kumar , Sujoy Das.(Aug-2012). "Rule Based Hindi to Urdu Transliteration System", Journal of Emerging Trends in Computing and Information Sciences, Vol.3, No. 8.

[11]. S. B. Kulkarni,. D. Deshmukh, K.V. Kale, (Feb.2014) "Syntactic and Structural Divergence in English-to- Marathi Machine Translation" International Symposium on Computational and Business Intelligence. International Journal of Computer Applications (0975 – 8887) Volume 87 – No.4.

[12]. Abhay Adapanawar,Anita Garje,Paurnima Thakare,Prajakta Gundawar,Priyanka Kulkarni (May-2013) "Rule Based English To Marathi Translation Of Assertive Sentence", International Journal of Scientific & Engineering Research, Volume 4,Issue 5.

[13]. Sudha Morwal ,Nusrat Jahan (2013) "Named Entity Recognition Using Hidden Markov Model (HMM): An Experimental Result on Hindi, Urdu and Marathi Languages" International Journal of Advanced Research in Computer Science and Software Engineering IJARCSSE All Rights Reserved.

[14]. Piyush Dungarwal, Rajen Chatterjee, Abhijit Mishra, Anoop Kunchukuttan,RiteshShah, Pushpak Bhattacharyya, (2014) "The IIT Bombay Hindi , English Translation System WMT2014".

[15]. Ruchika Sinhal. (July-2017). "A Pure EBMT Approach for English to Hindi Sentence Translation System"I.J.Modern Education and Computer Science, 7, 1-8 Published Online.

[16]. S. B. Kulkarni , P. D. Deshmukh , M. M. Kazi, K. V. Kale .(Feb-2014)." Linguistic Divergence Patterns in English to Marathi Translation" International Journal of Computer Applications(0975–8887) Volume 87– No.4.

[17]. G V Garje, G K Kharate,Harshad Kulkarni.(July 2014). "Transmuter: An Approach to Rule-based English to Marathi Machine Translation",International Journal of Computer Applications (0975 – 8887) Volume 98 –No.21.

[18]. Amarpreet Kaur, Jyoti Rani. (Dec.-2015). "A Web Based Punjabi to Hindi Statistical Machine Translation System" Proceedings of 2015 RAECSUIET Punjab University Chandigarh.

[19]. Sanjay Dwivedi and Pramod Sukhadeve, (2015)"Translation Rules for English to Hindi Machine Translation System: Homoeopathy Domain". The International Arab Journal of Information Technology,Vol. 12,No. 6A.

[20]. AkankshaGehlot, Vaishali Sharma, Shashipal Singh, Ajai Kumar. (June-2015). "Hindi to English Transfer BasedMachine Translation System" International Journal of Advanced Computer Research.

[21]. Jayashree Nair, Amrutha Krishnan K, Deetha R (Sept-2016) "An Efficient English to Hindi Machine Translation System Using Hybrid Mechanism", Conference on Advances in Computing, Communications and Informatics(ICACCI),21-24.

[22]. Goraksh V. Garje. (March-2016). "Marathi to English Sentence Translator for Simple Assertive and Interrogative Sentences", Research Gate, International Journal of Computer Applications.

[23]. PramodSalunkhe,Aniket.D.Kadam,Prof.ShashankJoshi,Prof.Shuhaspatil,Dr.DevendrasinghThakore,Shri kantJadhav.(2016) "Hybrid Machine Translation For English to Marathi: A Research Evaluation In Machine Translation" International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT).

[24]. Nitin Bansal, Dr. Ajit Kumar. (2017) "Machine Translation Survey for Punjabi and Urdu Languages".

[25]. Sudha Morwal ,Nusrat Jahan. (Oct-2017). "Named Entity Recognition Using Hidden Markov Model (HMM): An Experimental Result on Hindi, Urdu and Marathi Languages" International Journal of Advanced Research.

[26]. Iftikhar Ahmed Khan, Ahmad Khan,Babar Nazir,Syed SajidHussain,FiazGul Khan,Imran AliKhan. (2017). "Urdu Translation: the Validation and Reliability of the 120-Item Big Five IPIP Personality Scale".

AUTHORS PROFILE

Research Scholar
Mr.Bhanudas Suresh Panchbhai, (Pursuing PhD, M.Phil, M.C.S.), Head Department of Computer Science, R.C.Patel Arts, Commerce and Science College, Tal. Shirpur, Dist.Dhule , Maharashtra, India.

Research Guide
Dr.Varsha Makarand Pathak, (PhD, SET, MCA) ,Department of computer application, KCE'S Institute of Management and Research ,Jalgaon , Maharshtra.,India