"Machine Translation for Indian Languages a Review"

Aqsa Shaikh
M-phil Research Student
Dr. B. A. M. U. Aurangabad
Aurangabad, India

Guide: S. B. Kulkarni
Assistant Professor
Dept. of CS & IT,
Dr. B. A. M. U.
Aurangabad, India

Abstract:

Machine Translation Refers to Translation of one natural language to other by using automated computing facilities the main aim is to fill the language gap between two people, communities or countries. Machine Translation (MT) is exigent because it involves several thorny subtasks such as intrinsic language ambiguities, linguistic complexities and diversities between source and target language. This paper presents a review regarding the machine translation of Indian languages. This paper focused on the current scenario of machine translation nationally and internationally. This Literature Survey on machine translation considers three languages such as Hindi, Marathi, and Urdu.

Keywords:

Machine Translation, National Language Machine Translation, International Language Machine Translation

1.   Introduction:

In this Section First described what is Machine Translation (MT) and Its Multiple approaches also discussed national and internationally work done in machine translation.
 Machine Translation is the name for computerized methods that automate all or part of the process of translating from one language to another. In a large multilingual society like India, there is great demand for translation of documents from one language to another language. There are 22 constitutionally approved languages, which are officially used in different states. There are about 1650 dialects spoken by different communities. There are 10 Indic scripts. All of these languages are well developed and rich in content. They have similar scripts and grammars [22]. The alphabetic order is also similar. Multiple Languages use common scripts. Like devnagari.
 Hindi written in the Devanagri script is the official language of the union Government. English is also used for government notifications and communications. India's average literacy level is 65.4 percent (Census 2001).

 Research on MT systems between National and international based and also between Indian languages are going on in these institutions. Translation between structurally similar languages like Hindi and Punjabi is easier than that between language pairs that have wide structural difference like Hindi and English., Translation systems between closely related languages are easier to develop since they have many parts of their grammars and vocabularies in common [23].

2.   Machine Translation:

        The Aim of Machine translation is to translate one language to another language or source language to target language. Many people can use this Translator for Translation. Machine translation is from the broad area of Artificial Intelligence Natural language processing is based on different corpora

(vocabulary), these corpora are used for the processing of NLP to generate and develop a standard model which can be used for many purposes such as speech recognition technique, etc. [24].

2.1   Approaches to MT

There are multiple approaches to Machine Translation. These are discussed as follows.
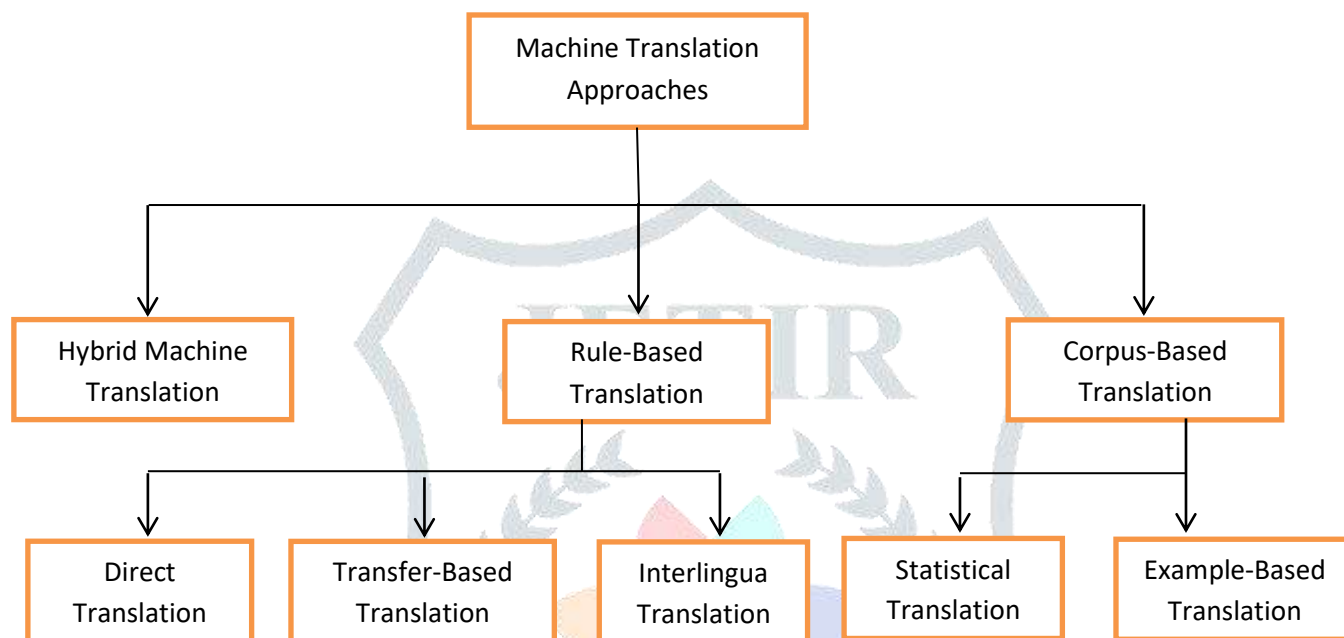

Figure2.1: Machine Translation approaches [27]

2.1.1   Rule-based MT

A Rule-based M T system parses the source text and produces an intermediate representation, which may be a parse tree or some abstract representation [26].

2.1.1.1 Direct-based MT

Direct Machine Translation is the one of the simplest machine translation approach. In Direct Machine Translation, a direct word by word translation of the input source is carried out with the help of a bilingual dictionary and after which some syntactical rearrangement are made. [27]

2.1.1.2   Transfer Based MT

In this translation system, a database of translation rules is used to translate text from source to target language. Whenever a sentence matches one of the rules, or examples, it is translated directly using a dictionary. It goes from the source language to a morphological and syntactic analysis to produce asor to Interlingua on the base forms of the source language, from this it translates it to the base forms of the target language and from there a better translation is made to create the final step in the translation.
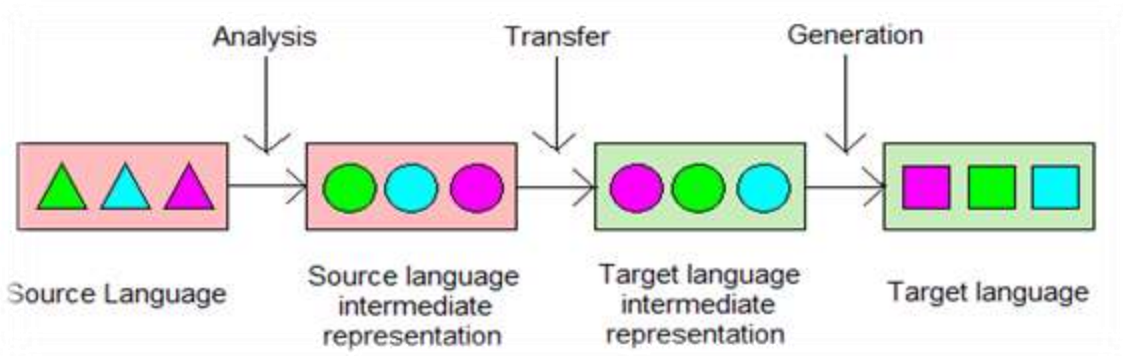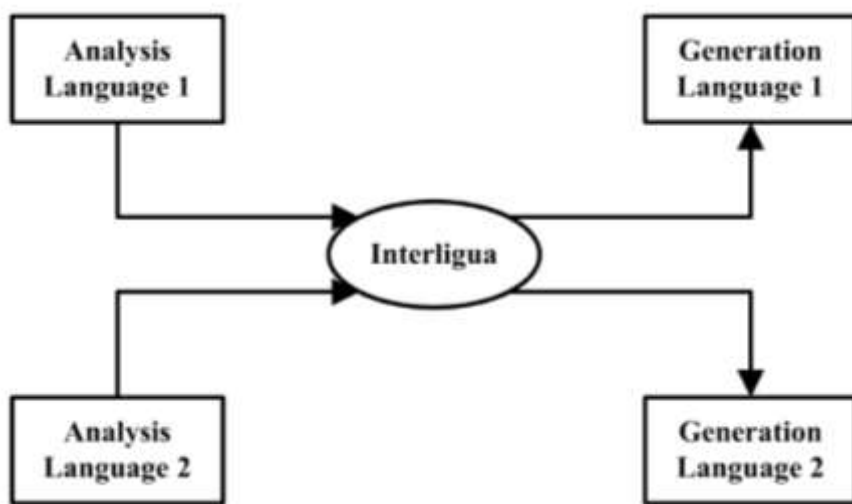
Fig2.2. Description of Transfer-Based Machine Translation

2.1.1.3   Interlingua Based MT

Interlingua machine translation is another classical approach to machine translation. This is an alternative to less efficient direct translation approach and includes transfer approach. In this approach, the source language is transformed into an Interlingua, which is an intermediate abstract language-independent representation. Then target language is generated from this Interlingua.

This approach is more efficient than direct translation as it is not merely a dictionary mapping of two languages. In this approach linguistic rules which are specific to the language pair transform the source language representation into an abstract target language representation and from this the target sentence is generated.   [27] Figure 3 shows how   different   languages can   be   translated through this system.

Fig2.3. Interlingua language system

2.1.3.   Corpus-based MT

Corpus based MT systems require sentence-aligned parallel text for each language pair. The corpus based approach is further classified into statistical and example-based machine translation approaches [26].

2.1.3.1   Statistical Based MT

In 1949, Warren Weaver presented the thought of statistical machine translation. In this methodology, statistical methods are employed to create translated form utilizing bilingual corpora. Statistical machine translation uses factual translation models whose parameters stem from the examination of monolingual and bilingual corpora. Building statistical translation models is a fast process; however the innovation depends intensely on existing multilingual corpora. At least 2 million words for a particular space and considerably more for general dialect are needed. Hypothetically it is conceivable to achieve the quality edge however most organizations don't have such a lot of existing multilingual corpora to construct the important translation models. Also, statistical machine translation is CPU concentrated and requires a broad equipment arrangement to run translation models for normal execution levels [25].

2.1.3.2   Example Based MT

Example based systems use previous translation examples to generate translations for an input provided. When an input sentence is presented to the system, it retrieves a similar source sentence from the example-base and its translation. The system then adapts the example translation to generate the translation of the input sentence.
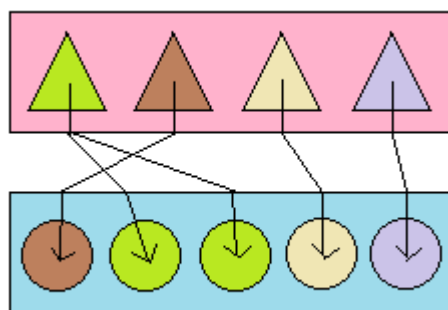

Fig: 2.4. Translation Template of a phrase in two different languages

2.1.4   Knowledge-based MT

Early MT systems are characterized by the syntax. Semantic features are attached to the syntactic structures and semantic processing occurs only after syntactic processing. Semantic-based approaches to language analysis have been introduced by AI researchers. The approached require large knowledge-base that includes both ontological and lexical knowledge [26].

LITERATURE SURVEY

3.   National Language Machine Translation

Basically Machine Translation is an active topic of research in India from 1991 onwards. The first work was started at IIT Kanpur and nowadays it has spread too many Universities. In this section now we look at some major National (Indian) Language MT Project. The Main Parameter we will cover here are: Language Pair(s), Approaches used for handling problems, Year of publication and domain name of MT system. Here I have discussed in table1, multiple national Languages Translation as Target Language or Source Language.

3.1 TRANSLATION SYSTEM FOR "HINDI", "MARATHI", "URDU" AS SOURCE OR TARGET LANGUAGE

Sr. No	Machine Translation System	Year	Languages for Translation	Domain/ Application	Approach Used	Observations
1	A Web Based Punjabi to Hindi Statistical Machine Translation System[1]	2015	Punjabi – Hindi	General	Statistical Based Machine Translation	Unigram algorithm, N-Gram string matching Algorithm etc. is formed the basis for solving the issues. The accuracy of the system has been evaluated using subjective tests i.e. intelligibility test and accuracy test. This system also works in reverse mode.
2.	The IIT Bombay Hindi to English Translation System at WMT[2]	2014	Hindi - English	General	Statistical Based	the use of number, case and Tree Adjoining Grammar Information as factors helps to improve English-Hindi translation, primarily by Generating morphological inflections correctly.
3.	A Pure EBMT Approach for English to Hindi Sentence Translation System[3]	2014	English – Hindi	General	Example Based	This research focuses on simple way of comparing Sentence to extract the translation.
4.	Translation Rules for English to Hindi Machine Translation System[4]	2015	English –Hindi	Homoeopathy	Rule Based	This paper have described the grammar rules intended for the English to Hindi machine translation system to translate the homoeopathic literatures,

					medical reports, prescription etc.
5.	Hindi to English Transfer Based Machine Translation System[18]	2015	English – Hindi	General	Transfer Based Machine Translation	This system takes an Input text checks its structure through parsing. Reordering rules are used to generate the text in Target language.
6.	An Efficient English to Hindi Machine Translation System Using Hybrid Mechanism[20]	2015	English – Hindi	General	Hybrid Machine Translation	English to Hindi machine translation System design based on declension rules. presented an effective methodology, proposed a new approach to MT system design which has not been considered in any of the existing MT systems so far
7.	EBMT Sindhi to Hindi Sentence Translation System[5]	2018	Sindhi – Hindi	General	Example Based	This research focuses on simple way of comparing sentence to Extract the translation. System have used training algorithm.
8.	Syntactic and Structural Divergence in English-to-Marathi Machine Translation[33]	2013	English - Marathi	General		we have examined the issue of Classification of translation divergence for MT between English and Marathi. shown that the translation divergence between

						Marathi and English machine translation is more varied and complex than the works in the existing literature can Accommodate and account for.
9.	Linguistic Divergence Patterns in English to Marathi Translation[31]	2014	English - Marathi	General	-	The primary objective of this paper is to understand the types of divergence problems that operate behind English to Marathi translation. This topic has been studied from different perspective and a number of approaches have been proposed to handle them.
10.	Hindi to English and Marathi to English Cross Language Information Retrieval Evaluation[6]	2007	Hindi – English And Marathi – English	Cross-Lingual Information Retrieval System	bi-lingual dictionaries	This paper present hindi to English and Marathi to English CLIR systems developed as of their participation in the CLEF 2007 Ad-HOC bilingual task. Translation of words which are not found in the dictionary is done using a simple rule based approach.
11.	Rule Based English To Marathi Translation Of Assertive	2013	English - Marathi	General	Rule Based	The developer dealing with the rule based English to Marathi

	Sentence [7]					translation of assertive sentence. In this system there is a use of bilingual dictionary.
12.	An Approach to Rule-based English to Marathi Machine Translation[8]	2014	English - Marathi	Transmuter	Rule Based	The basic algorithm for obtaining the correct word order in the target language was developed based on specific traversals of the parse tree. One of the special features of the system is a Word Sense Disambiguation model.
13.	Marathi to English Sentence Translator for Simple Assertive and Interrogative Sentences[9]	2016	Marathi – English	Translate Assertive and Interrogative sentences	Rule Based	The major goal of proposed system is to develop software system which would translate Marathi Simple Assertive and Interrogative Sentences to corresponding English sentences. The system will make use of Shallow parser, Bilingual Lexicon and Rearrangement algorithms to generate better quality translations.
14.	Hybrid Machine Translation For English to Marathi: A Research Evaluation In Machine	2016	English - Marathi	Hybrid Translator	translated Web pages, text Documents on Agriculture	The developer has discussed different approaches to machine translation. And different

	Translation[17]					divergence. author has proposed UI tags for web pages translation which proposes hybrid process that builds bilingual dictionary on RBI portal and parser is built in C
15.	Hindi-to-Urdu Machine Translation Through Transliteration [10]	2010	Hindi – Urdu	General	Statistical Based	This system propose two probabilistic models, based on conditional and joint probability formulations, that are novel solutions to the problem. used Kevin Gimpel's tester (http://www.ark.cs.cmu.edu/MT/) which uses bootstrap Resampling (Koehn, 2004b), with 1000 samples.
16.	Rule Based Hindi to Urdu Transliteration System[11]	2012	Hindi – Urdu	General	Rule Based	Some challenges have been handled such as ambiguous character, nukta related errors etc. by formulating special rules and using Database.
17.	A Hindi to Urdu Transliteration System[15]	-	Hindi – Urdu	high accuracy Hindi to Urdu transliteration system	Rule Based	The various challenges such as multiple/zero character mappings, variations in pronunciations and orthography, transliteration of

					proper nouns, Urdu word boundary etc. have been handled by generating special rules and using various lexical Resources such as Hindi spell checker, Urdu and Hindi word frequency lists, Urdu word bigram list, Hindi-Urdu lookup table etc.
18.	Machine Translation Survey for Punjabi and Urdu Languages[16]	2017	Urdu- English, Punjabi, hindi	survey	Different approaches	study different types of machine translation systems available for Punjabi and Urdu languages, about the tools available for converting source language text into target language text for regional as well international languages, discussed various methods for calculating accuracy of the translated output of the systems designed for the Punjabi and Urdu languages
19.	Named Entity Recognition Using Hidden Markov Model (HMM): An Experimental Result on Hindi, Urdu and Marathi Languages [19].	2013	An Experimental Result on Hindi, Urdu and Marathi Language	General	Linguistic Approach, Machine learning based Approach.	Main objective is to perform Named Entity Recognition in Natural languages using Hidden Markov Model (HMM) and provide ways to increase accuracy and the

						Performance Metrics (Precision, Recall, F-Measure).
20.	Urdu Translation: the Validation and Reliability of the 120-Item Big Five IPIP Personality Scale[34]	2017	Urdu Validation and Reliability	120-item International Personality Item Pool (IPIP)	Darwish translationmodel	In this study, developed the 120-itemInternational Personality Item Pool (IPIP) Urdu version using the Darwish translation model. The translation was verified by a panel of engineering experts and Urdu and English language experts. Moreover, an empirical investigation was conducted to determine the internal consistency, reliability and construct validity of the Urdu version.

4.   International Urdu Language Machine Translation

4.1   TRANSLATION SYSTEM FOR "URDU" LANGUAGE AS SOURCE OR TARGET LANGUAGE

Sr. No	Machine Translation System	Year	Languages for Translation	Domain/ Application	Approach Used	Observations
1.	Urdu to English Machine Translation using Bilingual Evaluation Understudy[12] (Kohat,	2013	Urdu – English	The Bilingual Evaluation Understudy (BLEU)	Rule Based, Statistical Based, Example Based	Analyzed and evaluated the main MT techniques using qualitative as well as quantitative approaches.

	Pakistan)				
2.	Model for English-Urdu Statistical Machine Translation[13] (Lahore, Pakistan)	2013	English – Urdu	General	Statistical Based	Discuss the issues of corpus alignment and share the results of baseline system prepared using Moses Decoder and other supporting tools.
3.	Hindi to Urdu Conversion: Beyond Simple Transliteration [14]	-	Hindi – Urdu	General		This paper detailed analysis of existing work on Hindi to Urdu transliteration systems and finds the enhancements they required. It lists the issues that are beyond the scope of character by character mapping.
4.	Lexical-Semantic Divergence in Urdu-to-English Example Based Machine Translation[28]	2010	Urdu - English	General	Example Based Machine Translation	The focus in this research is on lexical semanticdivergence and six different types are identified and generalizations are made on the basis of examples, for Urdu to English translation. Strategies are also presented for the identification of these types.
5.	Conversion between Hindi and Urdu[29] (Dammam, Saudi Arabia)	2015	Hindi - Urdu	General	Interlingua Based Machine Translation	This paper discusses the similarities and dissimilarities between Hindi and Urdu languages, delineates the issues in simple

					transliteration of Hindi to Urdu or vice versa, presents the effect of differences in spellings, pronunciation and writing style on conversion, lists the issues and solution to these issues which make conversion between Hindi to Urdu or Urdu to Hindi more than just simple transliteration
6.	Sequence to Sequence Networks for Roman-Urdu to Urdu Transliteration [30] (Islamabad, Pakistan)	2017	Roman-Urdu	General	Statistical Based Machine Translation	We create the first ever parallel corpora of Roman-Urdu to Urdu, create the first ever distributed representation of Roman-Urdu and present the first neural machine translation model that transliterates text from Roman-Urdu to Urdu language.
7.	Knowledge Based Machine Translation Semantically Enriched English-to-Urdu Machine Translation Using Data Mining Techniques[32] (Islamabad Pakistan)	2010	English - Urdu	ApniUrdu	Transfer Based Machine Translation	Proposed and designed a new Knowledge Based Machine Translation System to overcome the above mentioned problems by using data mining and text mining techniques.

5.   Conclusion

The present paper discusses the various language translation systems developed in the India as well as internationally which follows different approaches.3 main languages is considered in this paper such as Hindi, Marathi, Urdu as Source language or Target language.

It is concluded that direct approach for Machine Translation is most suitable for closely related languages i.e. the languages with similar structure. The indirect and statistical approach is suitable for languages with different structures

6.   References:

1]   AmarpreetKaur, Jyoti Rani "A Web Based Punjabi to Hindi Statistical Machine Translation System" Proceedings of 2015 RAECS UIET Panjab University Chandigarh 21-22nd December 2015, ©2015 IEEE

2]   PiyushDungarwal, RajenChatterjee, Abhijit Mishra, AnoopKunchukuttan,
Ritesh Shah, Pushpak Bhattacharyya, "The IIT Bombay Hindi,English Translation System at WMT 2014".

3]   RuchikaSinhal, "A Pure EBMT Approach for English to Hindi Sentence Translation System"I.J. Modern Education and Computer Science, 2014, 7, 1-8 Published Online July 2014 in MECS (http://www.mecs-press.org/).

4]   Sanjay Dwivedi and PramodSukhadeve, "Translation Rules for English to Hindi Machine Translation System: Homoeopathy Domain". The International Arab Journal of Information Technology, Vol. 12, No. 6A, 2015.

5]   Nisha S. Tathe, Jayasha S. Kriplani, "EBMT Sindhi to Hindi Sentence Translation System". International Journal of Advance Research, Ideas and Innovations in Technology, Volume 4, Issue 2, 2018.

6]   ManojChinnakotla, Om P. Damani, "Hindi to English and Marathi to English Cross Language Information Retrieval Evaluation", Conference paper, Research Gate, 2007.

7]   AbhayAdapanawar, Anita Garje, PaurnimaThakare, PrajaktaGundawar, PriyankaKulkarni " Rule Based English To Marathi Translation Of Assertive Sentence", International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013.

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

9]   Goraksh V. Garje, "Marathi to English Sentence Translator for Simple Assertive and Interrogative Sentences", ResearchGate, International Journal of Computer Applications March 2016.

10]   Nadir Durrani, Hassan Sajjad, "Hindi-to-Urdu Machine Translation Through Transliteration", ResearchGate, https://www.researchgate.net/publication/220873998, November 2010.
11]   BushraBaig ,M.Kumar , 3Sujoy Das, " Rule Based Hindi to Urdu Transliteration System", Journal of Emerging Trends in Computing and Information Sciences, Vol. 3, No. 8 Aug, 2012

12]   Asad Abdul Malik ,AsadHabib, " Urdu to English Machine Translation using Bilingual Evaluation Understudy", International Journal of Computer Applications (0975 – 8887) Volume 82 – No 7, November 2013.

13]   Aasim Ali, ArshadHussain and Muhammad Kamran Malik, "Model for English-Urdu Statistical Machine Translation", World Applied Sciences Journal 24 (10): 1362-1367, 2013.

14]   BushraJawaid, Tafseer Ahmed, "Hindi to Urdu Conversion: Beyond Simple Transliteration".

15]   urpreet Singh Lehal, Tejinder Singh Saini, "A Hindi to Urdu Transliteration System", 8th International Conference on Natural Language Processing Macmillan Publishers, India. Also accessible from http://ltrc.iiit.ac.in/proceedings/ICON-2010.

16]   NitinBansal, Dr. Ajit Kumar, "Machine Translation Survey for Punjabi and   Urdu Languages" © **2017 IEEE.**

17]   PramodSalunkhe,   Aniket .D. Kadam,   Prof. Shashank   Joshi,   Prof.Shuhaspatil, Dr.DevendrasinghThakore, ShrikantJadhav, "Hybrid Machine Translation For English to Marathi: A Research Evaluation In Machine Translation" International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) **©2016 IEEE.**

18]   AkankshaGehlot, Vaishali Sharma, Shashipal Singh, Ajai Kumar, "Hindi to English Transfer Based Machine Translation System" International Journal of Advanced Computer ResearchISSN, **ResearchGate** 19 June 2015

19]   SudhaMorwal ,NusratJahan, "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, © 2013, IJARCSSE All Rights Reserved.

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

21]   AmrutaGodase,   andSharvariGovilkar, "Machine Translation Development For Indian Languages And Its Approaches" International Journal on Natural Language Computing (IJNLC) Vol. 4, No.2,April 2015.

22]   Latha R. Nair, David Peter S. "Machine Translation Systems for Indian Languages", International Journal of computer Application, volume 39, 2012.

23]   V Goyal, G S Lehal. "Advances in Machine Translation Systems". Language In India, Vol. 9, No. 11, 2009, pp. 138-150.

24]   Shachi Mall, Umesh Chandra Jaiswal, "Survey: Machine Translation for Indian Language", International Journal of Applied Engineering Reasearch, Volume 13, 2018

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

26]   Nakul Sharma "English To Hindi Statistical Machine TranslationSystem" Thesis Submitted In Partial Fulfillment OfThe RequirementsFor The Award Of Degree.

27]   SandeepSaini, VineetSahula "Survey of Machine Translation Techniques and Systems for Indian Languages" IEEE International Conference on Computational Intelligence & Communication Technology, **© 2015 IEEE**

28]   AbdusSaboor, Mohammad Abid Khan "Lexical-Semantic Divergence in Urdu-to English Example Based Machine Translation" 2010 6th International Conference on Emerging Technologies (ICET) **©2010 IEEE**

29]   Shahnawaz, "Conversion between Hindi and Urdu" International Conference on Computing, Communication and Automation (ICCCA2015) **©2015 IEEE**

30]   MehreenAlam, SibtulHussain, "Sequence to Sequence Networks for Roman-Urdu to Urdu Transliteration" 20th International Multitopic Conference (INMIC' 17)**©2017 IEEE**

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

32]   GhulamRasoolTahir, SohailAsghar, NayyerMasood, " Knowledge Based Machine Translation Semantically Enriched English-to-Urdu Machine Translation Using Data Mining Techniques" ResearchGate**©2010 IEEE**

33]   S. B. Kulkarni,. D. Deshmukh, K.V. Kale, **"**Syntactic and Structural Divergence in English-to- Marathi Machine Translation" International Symposium on Computational and Business Intelligence **© 2013 IEEE**

34]   Iftikhar Ahmed Khan, Ahmad Khan, Babar Nazir, Syed SajidHussain, FiazGul Khan, Imran Ali Khan, "Urdu Translation: the Validation and Reliability of the 120-ItemBig Five IPIP Personality Scale",© **Springer**Science+Business Media, LLC 2017.