

Digital Humanities: Can Machine Translation Replace Human Translation

VenuShree, Iti Mathur, Geetha Yadav, Nisheeth Joshi

“Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.”-Albert Einstein.

Abstract: MT is a sub field of computational linguistics and it uses different software or online translating tools to translate text/documents from one language to another. MT has become quite popular because of its use in business and commerce industries. These industries believe it to be cost efficient, easy and faster. Hence, these industries wish to get rid of human translators and they want their documents, emails etc. to be translated through Google, Bing, Babylon and other similar tools. However, there is a great need of improvements in the quality of translations done with the help of these online tools. At times, they are not at par human translations. This paper provides a detailed description of Machine Translation, its development and its relevance in contemporary scenario. It also focuses on issues like the need for linguists in MT and if MT can replace human translation. We have validated our claim by evaluating the translations of Google and Bing translators on 500 sentences. In all cases the results was not at par human translations.

Keywords: - Linguists, Machine Translation, Technology, Translation tools.

I. INTRODUCTION

Technology is the area which deals with the creation of knowledge and use of technical means. It studies the relationships between society, life and environment. For the development of language and literature web-based technologies and the superb internet connections play a significant role. With the passing of time, language and literature have adopted changes in technical field to keep everything updated. Technical changes are not only bound to the technical field but also to the output and input that is the production and reception of literature and language. Technology has developed gradually from time to time and day by day. If we talk about the past, that is, the history of technology, then in the past, the development of simple things started taking place. The tools were initially made up of wood or shards of rock and then gradually the discovery came out of fire. It provided a way to cook food and create heat and light. These simple technologies helped people to finish their tasks more easily and quickly and the age when these developments were taking place is known as the Stone Age. In the Bronze Age, human beings worked with metal and formed stronger tools.

Revised Manuscript Received on March 20, 2019

Venu Shree department of English and Modern European Languages, Banasthali Vidyapith.

Iti Mathur, Assistant Professor, Banasthali Vidyapith. Department of Electronics and Information Technology (Deity), Govt. of India,

Geetha Yadav, Associate Professor, Banasthali Vidyapith, Department of English and Modern European Languages.

Nisheeth Joshi, Associate Professor, Banasthali Vidyapith. Department of Computer Science,

They also introduced wheel which allowed people to communicate and travel from one place to another. The third one, the Iron Age is known for the development of weapons. People worked with harder metals than tin and copper and learnt to smelt iron and were able to remove iron from ore. All these developments gave the opportunities to people to perform tasks, such as manufacturing and transportation. Thus through innumerable scientific discoveries and inventions through the centuries, humans have reached the 21st century in which the computers have changed the entire scenario. [46]

Computers and internet have become the priority for everyone. Internet has made easy for people to keep themselves updated. Now- a-days, it has become quite popular and writers are fully taking advantage of this technical platform to reach out to their audience and to spread their creativity. There are lots and lots of books published online in their digital form. The response coming from the readers too is irresistible. There are enormous players of technology like Amazon, Google and Apple etc. that have contributed immensely to the production and circulation of literature on digital platform. ‘Kindle’ is the first e-reader platform introduced by Amazon. Now people also follow digital books rather than printed ones. A few decades back there had been a revolutionary development coming up in the fields of computing and communications, and this shows that technology has progressed immensely and will continue to do so. The most important thing that technology has done is doubtlessly electronic commerce over the Internet, and this is a new way of conducting business. It has affected each and every area from large sectors as communications, finance and retail trade and it has expanded to other areas as well such as health services and education. As technology has grown rapidly, all sector related to human activity is now engaging itself in the field of internet and e-commerce. The number of participants is also growing enormously day-by-day. In today’s world, there are innumerable online shopping sites and the competition between them is growing at a very high rate. We can shop 24 hours a day and 7 days a week and these facilities are just a click away from us. Digitized products can reach the customers very easily and can also cross borders within no time at all. Technology also plays an important role when it comes to a company making ads for its products. So, now, in the contemporary world, it is the best for a company to use its creativity and launch the advertisements of its products with the help of latest technologies.

The most remarkable effect of technology has undoubtedly been in the field of communication. It has also helped people to keep themselves connected through video-calling, telephonic communication and also through Short Message Service (SMS) and many other such oral and written modes. It also allows for the work collaborations- the work which involves actors and distributors who rarely meet each other but can work together online. This type of online work mode is active 24*7*365. Communication also has become very easy for the traders as they can communicate through e-mail which decreases the barriers of communication across different countries, states and status levels also. With the help of technology, the problem of distance has reduced and traders can communicate to each other in a very easy manner. With this, it is also possible for the employees to sit in their homes rather than go to centralized workplaces. If someone starts one's business from her/his home, then the opportunities for the local population also increase. In fact, in our day-to-day world, many people are working from their homes and are very successful in their particular field. Technology also helps in creating jobs for the unemployed. As technology has progressed the chances of employment have also increased. There is immense need of employees to work on internet throughout in MNC's or IT sectors like customers' support. So, there are many people who are working at Customer Care Centers at a very low salary. Thus, technology has affected our life in such a way that we can't even think of living without it a single day. Technology has become an integral and indispensable part of our lives in a multilingual, multicultural world that consists of various languages/dialects and it is practically impossible for us to know and understand each and every language. If an individual is not able to read or understand a particular language then it is quite a big disadvantage for the person. So, here, again, technology plays an important role by working on the process of translation. With the help of translation it has become easier for people to read and understand other languages also. Translation, as we know, is a process in which a source text is translated into a target text. It is based on both creativity and technical process. With the help of translation it has become much easier to interact with people who are miles away, almost across the other side of the globe. In the modern world a new type of translation has become quite popular. It is called Machine Translation (MT). MT is a sub-field of computational linguistics. Machine Translation is automated translation or "translation carried out by a computer" as it has been defined in the Oxford English dictionary. It is a type of process which is also referred to as Natural Language Processing (NLP) that uses a set of bilingual data and also other language assets to build language. To translate a text from its source language to its target language, phrase models are used. As we are citizens of globalized world, no doubt, computer and internet have become so popular and essential for each one of us that research and development in MT continue to grow rapidly [6]. MT is a part of computational linguistics and it is the method that works on the software tools to translate a text from its source language to its target language. The concept of language translation has developed widely and it solves the problems of translation divergence. There are around 5,000

languages/dialects spoken all over the world and it is impossible to know and understand each and every language that exists. This shows the need for translation to communicate across languages. Researchers belonging to the field of linguistics are working hard to make it easy for everyone to exchange messages from one language to the other. Government agencies and research institutes are taking interest and working on the process of developing MT tools that could control translation of the texts. This is also useful for worldwide business or for any professional communication. It seems quite interesting and appealing that there are many MT tools that have been developed recently and they are quite easily available without any cost. They are not good for professional business work due to their quality and security. When we need to translate any text or document then we first have to upload the text or document online. So, as soon as the document is uploaded online, there is no assurance of its security. All MT engines use software to translate from one language to the other. This simple and free software do not produce good translations. They produce low-quality translations which are not appropriate and sound unusual to a native speaker. There are different types of MT approaches that work on translation. Dictionary Based MT has been the first generation of automated language translation and it is totally based on electronic dictionaries. This Dictionary Based MT is able to translate phrases but not sentences. The other MT tools that are available and are the most widely used include Rule Based Machine Translation (RBMT), Statistical Machine Translation (SMT), Example Based Machine Translation (EBMT) and Hybrid.

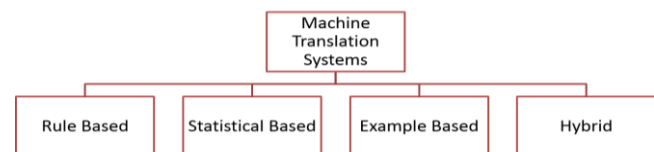


Fig.1: Different Approaches of Machine Translation Systems [1]

II. APPROACHES OF MACHINE TRANSLATION

1. Rule Based Machine Translation (RBMT) – It depends on the number of built-in linguistic rules and millions of bilingual dictionaries for each language pair. In Rule based Machine Translation, the system first parses the text and builds a transitional representation and from here the text in the target language is generated. This process requires major lexicons with the information regarding semantic, syntactic and morphological as well as large set of rules. Then the software uses these complicated rule sets and then it transfers the grammatical structure of the source text into the target text. Rule based MT systems are built on large dictionaries and refined linguistic rules. Human translation or the user who ever works on this software can improve the quality of translation by adding terms into the process of translation by creating user-defined dictionaries, which override the systems default settings [6].

Some work in the Indian context has been done in this area. At lexical level, Ameta et al. [1] developed a rule-based stemmer for Gujarati which they used in a Gujarati-Hindi MT system [2]. This used a transliteration scheme that was developed by Joshi et al. [3][4]. Paul et al. [5][6] developed a lemmatizer for Hindi which used a POS Tagger based on statistical learning [7]. Katyayan and Joshi [8] used this tagger for sarcasm detection. Some more POS Taggers were developed using this approach. Singh et al. [9][10][11] developed a POS Tagger for Marathi while Gupta et al. [12] developed a tagger for Urdu. Gupta et al. [12] further used their tagger in development of a stemmer [13][14] and a lemmatizer [15]. They also used this tagger for development of a multi-word expression system [16][17]. For Punjabi, Bhalla et al. [18] developed a name entity recognition (NER) and translation system. Chopra et al. [19][20] developed NER system for Hindi. Efforts have also been done in the area of parsing and chunking. Asopa et al. [21][22] developed a shallow parser for Hindi. Chopra et al. [23] and Tyagi et al. [24][25] developed techniques performing syntactic transfer for English to Hindi MT. Chopra et al. [26] studied the drawbacks in the MT systems in Indian Languages and developed an MT system [27] which addressed most of the issues raised by them. Singh et al. [28] developed a transfer grammar system for English-Hindi MT using parallel corpus. At semantic level, Kumar et al. [29] and Sharma and Joshi [30] developed mechanisms for word sense disambiguation for Hindi. According to (Bhallah & Jaiswal, 2015 p.424-428), there are three types of RBMT systems which are as follows:

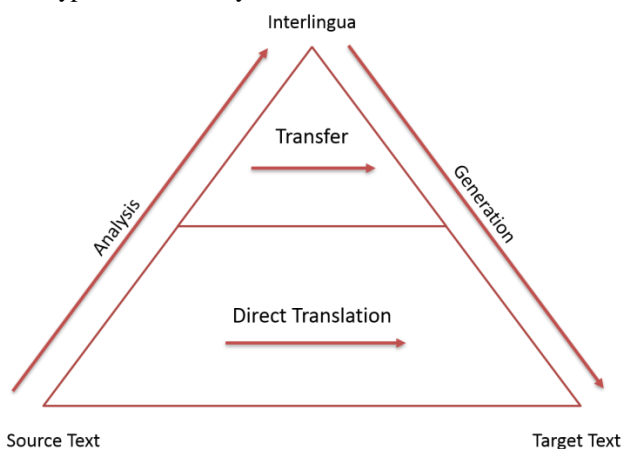


Fig. 2: Sub Approaches to RBMT [31]

a) *Direct Method (Dictionary Based MT)* - This is the part of RBMT system and in this method the source text is translated without passing through an intermediary representation. The words for translation get translated as a dictionary meaning that is word by word, usually without much connection of meaning between them. There is no compulsion of morphological analysis. Using this approach, Joshi et al. [32][33] and Singh et al. [34] developed a translation system for English-Hindi.

b) *Transfer Rules Based Machine Translation Systems*- Morphological and syntactical analysis are the fundamental approaches in Transfer Based Systems. Here source language text is converted into less language specific representation and same level of abstraction is generated

with the help of grammar rules and bilingual dictionaries. In the transfer approach of translation divergence there are transfer rules for transforming a source language (SL) sentence into target language (TL) by performing lexical and structural manipulations.

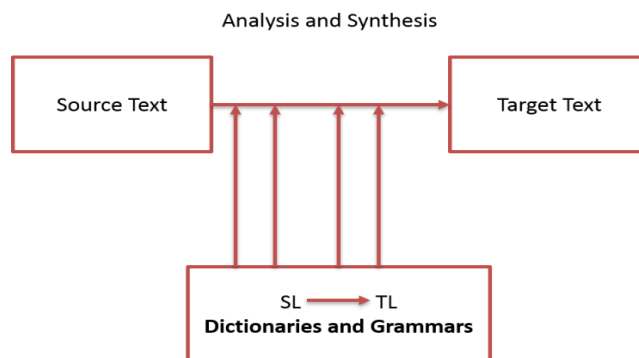


Fig. 3: Dictionary Based Machine Translation [1]

c) *Interlingual RBMT Systems (Interlingua)* - This model is intended to make linguistic homogeneity across the world. In this method, source language is translated into an intermediary representation which does not depend on any language. Target language is derived from this auxiliary form of representation.

2. **Statistical Machine Translation (SMT)**-Statistical models are applied in this method to create translated output with the assistance of bilingual corpora. The concept of Statistical Machine Translation comes from information theory. The important feature of this method is that no customization work is required by linguists because the tool learns translation methods through statistical analysis of bilingual corpora. These systems are not able to produce good translations. Thus, they need to be edited. Yogi et al. [35] and Joshi et al. [36] did a study of post editing on English-Hindi MT.

3. **Example Based Machine Translation (EBMT)**-This method is also called as Memory Based Translation in which a set of sentences from source language are given and it generates corresponding translations in target language with point to point mapping. Here examples are used to convert similar types of sentences and previously translated sentences are repeated, the same translation is likely to be correct again. The main advantage of this model is that it works well with a small set of data and it is possible to generate output more quickly by training the translation program. Example based method is mainly used to translate two totally different language pairs like Japanese and English etc. It is not possible to apply deep linguistic analysis and it is one of the main drawbacks of Example-Based Engine.

4. **Hybrid Machine Translation**- HMT takes the advantages of RBMT and Statistical Machine Translation. It uses RBMT as baseline and refines the rules through statistical models. Rules are used to pre-process data in an attempt to better guide the statistical engine.

Digital Humanities: Can Machine Translation Replace Human Translation

The Hybrid Model differs in various ways from the above models. [37]

All the above concepts describe and explain what Machine Translation is and how it works. So, the next questions that arise are: Can MT replace human translation? Is there any need for linguists in MT? So, when we think about translating a text from one language into another language, we must know that whether the translation is done by a machine and if the source text has been fully transferred to its equivalent meaning in the target language translation. It looks like this is an easy way to translate a text but it is much more difficult and complex because the translation is never a mere word-for-word substitution. So, when it comes to human translation, then it is easy for a human translator to understand and interpret all the elements within the text and understand how each word within the text may influence the context of the text. This obviously will require a translator who is not only an expert in grammar, syntax, the semantics of the languages concerned but is also well versed with the socio-cultural contexts of both the languages, that is the source and the target language. [45] In this area, some work has been done to perform MT evaluation for Indian Languages. Gupta et al. [38][39] evaluated various English-Urdu MT systems while Joshi et al. [40][41][42] performed an evaluation on English-Hindi language pair.

III. METHODOLOGY

There are many tools that are used online for translation. These include Google, Bing, Yahoo, Babelfish, AltaVista etc. Presently Google Translate is totally free for all as an online application. It is used in a wide range of automated translations in any of the available languages. It is very easy to use and it works satisfactorily and it claims to provide translations in over 50 languages. There are limitations and therefore Google cannot provide a sense of security, because it is not possible for Google to find out the accurate linguistic meaning and interpret the cultural nuances involved.

Given below are the sentences translated by Google, Bing and human from Hindi to English which clearly show the difference of translation between online machine translators and human translator.

Example 1

वे बारह थीं - छोटी-छोटी चिड़ियाँ, सुन्दर, रंग-बिरंगी, कोई काली-सफेद, तो कोई पूरी काली और सिर्फ गले में सफेद चाँद की फाँक-सी उड़ली हुई, कोई गेरूए रंग की लेकिन उनके पर काले या सफेद, कोई पीली और लाल।

Google Translation- They were twelve - small birds, beautiful, colourful, some black-white, so there was a full moon in black and white just in the throat, no moon colour, but black or white on them, no yellow and red.

Bing Translation- He was twelve-small-chidriya, elegant, colour-seductive, no black-white, a whole black and just neck safde Clementine-C Uchli, no gerue of colour but black or white, no yellow and red.

Human Translation- They were twelve-small birds, beautiful, colourful, some black-white, some completely black with only a piece of white like the crescent moon

arising on its neck, some grey but with black or white features, some yellow and red.

Example 2

रोहित की माँ के लिए अच्छी व्यवस्था हो गई थी - दिन में दसियों बार देखना कि पिंजड़े में खाना-पीना बराबर रखा है कि नहीं, पिंजड़ा और उसके पास सफाई हुई कि नहीं।

Google Translation- Rohit had got a good arrangement for his mother - seeing ten times a day to see whether the food was kept equal in the cage or not, the cage and the cleaning were done or not.

Bing Translation- Rohit's mother had a good arrangement-to see the Tens bar in the day that cage in the food is not equal to Pinjadera, and that it was not clean.

Human Translation- There was a good arrangement for Rohit's mother-checking ten times a day if the cage had food and water or not, if the cage and its surroundings were cleaned or not.

Example 3

चिड़ियाँ ज्यादा चहचहा रही हैं तो क्यों!

Google Translation- If the birds are talking much more then why!

Bing Translation- Chidriya is more tweet.

Human Translation- Why are the birds chirping so much?

Example 4

चहक तो जैसे बराबर फूटती रहती उनके गले सोसिवाय रात के, जब वे सो जातीं, घर बराबर गूँजता रहता है।

Google Translation- The tart was like a tear of his throat. Except for the night, when you go to sleep, the house keeps echoing.

Bing Translation- The tweet, as equal will, is his throat. Except for the night, when sleeping, the house is equally echoed.

Human Translation- They used to chirp all the time. Except at night, when they slept, the house echoes constantly.

Example 5

चिड़ियाँ बेचने वाले ने कहा था कि ये चिड़ियाँ सिर्फ उचकती-फुदकती हैं, उड़ती नहीं है।

Google Translation- The seller of the birds said that these birds are just fluttering and do not fly.

Bing Translation- Chidriya selling said that these birds are just uchkati-fudkati, not levitating.

Human Translation- The bird seller had said that these birds just hop-leap around, they do not fly.

The above are the sentences which are a little complex which Google Translate has translated to some extent correctly but not exactly. Even Bing Translator has failed badly to obtain the correct translation. One cannot make head or tail of the Bing translations. So these translators are not so efficient in the case of complex sentences. However, there are sentences which Google has translated almost correctly and this involves short sentences of the type subject-verb-object (SVO) pattern.

So, now the below-given examples of short sentences make it clear as to which type of sentences can be translated correctly by Google Translate or Bing. Now let us examine these sentences.

Example 1

गाँव धनी था।

Google Translation- The village was rich.

Bing translation- The village was wealthy.

Human translation- Village was rich.

Example 2

क्या बात है?

Google Translation- What's the matter.

Bing Translation- What's the point?

Human Translation-What is the matter?

Example 3

रास्ते में एक कब्रिस्तान था।

Google Translation- There was a cemetery on the way.

Bing Translation- Was a cemetery on the way.

Human Translation- Human Translation- There was a graveyard on the way.

Example 4

पिंजड़ा रखा था।

Google Translation-The cage was kept.

Bing Translation- Pijda was placed.

Human translation- Cage was kept.

Example 5

रोहित ने देखा।

Google Translation- Rohit saw it

Bing Translation- Saw by Rohit.

Human Translation- Rohit saw.

With the help of the above examples it is clear that Google can translate short sentences in a proper manner to a great extent and Bing too, to some extent but only of the simplest SVO pattern. Bing has failed to translate correctly even the short sentences with simple subject and predicate correctly in most of the cases.

IV. EVALUATION

We evaluated the translation of Google and Bing for 500 sentences which were divided into 5 documents of 100 sentences each. We employed two human annotators for this purpose. First annotator was asked to evaluate the results based on human evaluation metric (HEval) which is developed by Joshi et al. [43]. This metric has 10 parameters which are supplied to the human annotator for judging the quality of MT output. The score is given on a 5-point scale where 5 is the best score and 1 is the poorest. Once all 10 questions have been scored their average is calculated based on equation 1. Here, the sum of all the question scores is divided by the number of questions applicable.

$$HEval - Score = \frac{\sum_{i=1}^{10} question - score_i}{\#questions} \quad (1)$$

The second human annotator was asked to perform the post edits to improve the quality of machine translation and make it at per human translation. For this we used HTER (Human

Translation Edit Rate) metric developed by Specia and Atefeh [44]. This was calculated using Equation 2. Here the total number of edit (shift, insert, delete) operation are divided by the number of reference words in the reference translations. The edit operations are required to make two translations; the translated output (hypothesis) and the human translation (reference translation); exactly same.

$$HTER = 1 - \frac{\# Edits}{\# of Reference Words} \quad (2)$$

	HEval		HTER	
	Google	Bing	Google	Bing
Doc1	0.552926	0.531496	0.423577	0.499932
Doc2	0.507185	0.505481	0.477696	0.455029
Doc3	0.506252	0.506602	0.481573	0.491361
Doc4	0.533271	0.527722	0.52299	0.479944
Doc5	0.515314	0.519334	0.518127	0.481599
Average	0.52299	0.518127	0.484793	0.481573

Table 1: Evaluation Results

As shown in table 1, overall Google's score was better than Bing's, for both humans as well as HTER. On analysing further, we found that Bing's score for slightly higher for Documents 1, 3 and 5, for human evaluation and was had a higher HTER score for documents 1 and 3. The same is shown in figure 4 and 5.

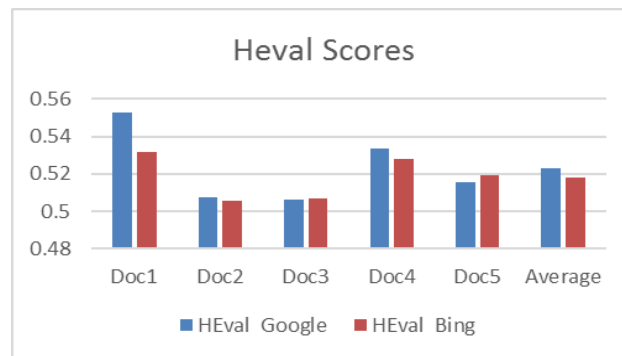


Fig.4: HEval Scores for Google and Bing

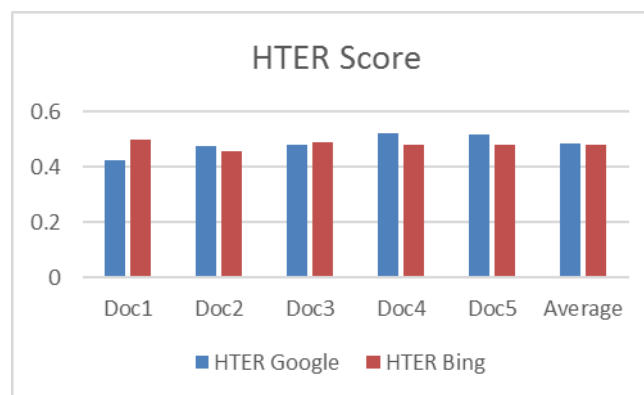


Fig.5: HTER Scores for Google and Bing

V. CONCLUSION

These online tools are free and can be used by anyone having access to the internet with good speed. Moreover, they are also not based on any user agreement between the user and Google or any other online tool which is free. From the examples and explanations given, it is clear that online tools cannot provide translations and these applications are not intended to be used commercially. So, if a translation of a text or document requires a great degree of accuracy, Google Translator or any other free online translators will not be able to replace the human translator. Although Google and Bing give good translations on most of the text, but they at the time produce sub-optimal translations for certain domain texts. This can be verified by our evaluation where we took 500 sentences from children stories, and in all the cases these MT engines were not able to produce good translations. Their evaluation score was under 55%; this means that a human annotator has to put in a significant effort to post edit these translations. This was also verified by our second evaluation where we asked a human annotator to do post editing of the machine translated outputs. The annotator had to put in a lot of effort to make the translation fluent. This score was over 40% across sentences and machine translators.

In future, may be at some point in time, it might be possible for computers to understand every language structure and the software programs that are available may become intelligent enough to grasp and understand languages. The most important point is that language is a dynamic process which keeps on changing and new words keep adding to the dictionary now and then. Above all, language also reflects the psychological condition of human beings which can never be fully captured in its pure sense by a machine. In spite of all their drawbacks, computers and software play an important role in bridging language gaps and helping in inter-language communication. Today's technology is a boon to every one of us because it also opens great opportunities for companies, markets and business people. It also has made everything easily accessible, but when it comes to the field of linguistics and language, then it is lagging behind, and it has to develop a lot.

REFERENCES

1. Ameta, J., Joshi, N., & Mathur, I. (2012). "A lightweight stemmer for Gujarati." arXiv preprint arXiv:1210.5486.
2. Ameta, J., Joshi, N., & Mathur, I. (2013). "Improving the quality of Gujarati-Hindi Machine Translation through part-of-speech tagging and stemmer-assisted transliteration." arXiv preprint arXiv:1307.3310.
3. Joshi, N., & Mathur, I. (2012). "Input Scheme for Hindi Using Phonetic Mapping." arXiv preprint arXiv:1209.1300.
4. Joshi, N., Mathur, I., & Mathur, S. (2010). "Frequency-based predictive input system for Hindi." In Proceedings of the International Conference and Workshop on Emerging Trends in Technology (pp. 690-693). ACM.
5. Paul, S., Tandon, M., Joshi, N., & Mathur, I. (2013). "Design of a rule-based Hindi lemmatizer." In Proceedings of Third International Workshop on Artificial Intelligence, Soft Computing and Applications, Chennai, India (pp. 67-74).
6. Paul, S., Joshi, N., & Mathur, I. (2013). "Development of a Hindi lemmatizer." arXiv preprint arXiv:1305.6211.
7. Joshi, N., Darbari, H., & Mathur, I. (2013). "HMM-based POS tagger for Hindi." In Proceeding of 2013 International Conference on Artificial Intelligence, Soft Computing (AISC-2013).
8. Katyayan, P., & Joshi, N. (2019). "Sarcasm Detection Approaches for English Language." In Smart Techniques for a Smarter Planet (pp. 167-183). Springer, Cham.
9. Singh, J., Joshi, N., & Mathur, I. (2013). "Development of Marathi part of speech tagger using statistical approach." In Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on (pp. 1554-1559). IEEE.
10. Singh, J., Joshi, N., & Mathur, I. (2013). "Part of speech tagging of Marathi text using trigram method." arXiv preprint arXiv:1307.4299.
11. Singh, J., Joshi, N., & Mathur, I. (2014). "Marathi Parts-of-Speech Tagger Using Supervised Learning." In Intelligent Computing, Networking, and Informatics (pp. 251-257). Springer, New Delhi.
12. Gupta, V., Joshi, N., & Mathur, I. (2016). "POS tagger for Urdu using Stochastic approaches." In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (p. 56). ACM.
13. Gupta, V., Joshi, N., & Mathur, I. (2013). "Rule based stemmer in Urdu." In Computer and Communication Technology (ICCT), 2013 4th International Conference on (pp. 129-132). IEEE.
14. Gupta, V., Joshi, N., & Mathur, I. (2015). "Design & development of rule based inflectional and derivational Urdu stemmer 'Usal'." In Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015 International Conference on (pp. 7-12). IEEE.
15. Gupta, V., Joshi, N., & Mathur, I. (2016). "Design and development of a rule-based Urdu lemmatizer." In Proceedings of International Conference on ICT for Sustainable Development (pp. 161-169). Springer, Singapore.
16. Gupta, V., Joshi, N., & Mathur, I. (2017). "Approach for multiword expression recognition & annotation in urdu corpora." In Image Information Processing (ICIIP), 2017 Fourth International Conference on (pp. 1-6). IEEE.
17. Gupta, V., Joshi, N., & Mathur, I. (2019). "Advanced Machine Learning Techniques in Natural Language Processing for Indian Languages." In Smart Techniques for a Smarter Planet (pp. 117-144). Springer, Cham.
18. Bhalla, D., Joshi, N., & Mathur, I. (2013). "Improving the quality of MT output using novel name entity translation scheme." In Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on (pp. 1548-1553). IEEE.
19. Chopra, D., Joshi, N., & Mathur, I. (2016). "Named Entity Recognition in Hindi Using Conditional Random Fields." In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (p. 106). ACM.
20. Chopra, D., Joshi, N., & Mathur, I. (2016). "Named Entity Recognition in Hindi Using Hidden Markov Model." In Computational Intelligence & Communication Technology (CICT), 2016 Second International Conference on (pp. 581-586). IEEE.
21. Asopa, S., Asopa, P., Mathur, I., & Joshi, N. (2016). "Rule based chunker for Hindi." In Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on (pp. 442-445). IEEE.
22. Asopa, S., Asopa, P., Mathur, I., & Joshi, N. (2019). "A Shallow Parsing Model for Hindi Using Conditional Random Field." In International Conference on Innovative Computing and Communications (pp. 295-302). Springer, Singapore.
23. Tyagi, S., Chopra, D., Mathur, I., & Joshi, N. (2015). "Classifier based text simplification for improved machine translation." In Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in (pp. 46-50). IEEE.
24. Tyagi, S., Chopra, D., Mathur, I., & Joshi, N. (2015). "Comparison of classifier-based approach with baseline approach for English-Hindi text simplification." In Computing, Communication & Automation (ICCCA), 2015 International Conference on (pp. 290-293). IEEE.
25. Chopra, D., Joshi, N., & Mathur, I. (2016). "Improving Quality of Machine Translation Using Text Rewriting." In Computational Intelligence & Communication Technology (CICT), 2016 Second International Conference on (pp. 22-27). IEEE.
26. Chopra, D., Joshi, N., & Mathur, I. (2018). "A Review on Machine Translation in Indian Languages." Engineering, Technology & Applied Science Research, 8(5), 3475-3478.
27. Chopra, D., Joshi, N., & Mathur, I. (2018). "Improving Translation Quality By Using Ensemble Approach." Engineering, Technology & Applied Science Research, 8(6), 3512-3514.
28. Singh, S. P., Kumar, A., Darbari, H., Singh, L., Joshi, N., Gupta, P., & Singh, S. (2017, March). Intelligent System for Automatic Transfer Grammar Creation Using Parallel Corpus. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 519-528). Springer, Cham.

29. Kumar, A., Mathur, I., Darbari, H., Purohit, G. N., & Joshi, N. (2016). "Implications of Supervised Learning on Word Sense Disambiguation for Hindi." In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (p. 54). ACM.
30. Sharma, P., & Joshi, N. (2019), "Design and Development of a Knowledge Based Approach for Word Sense Disambiguation by Using WordNet for Hindi", International Journal of Innovative Technology and Exploring Engineering, pp 73-78, Vol 8(3).
31. Ballabh, A., & Jaiswal, UC. "A study of machine translation and their challenges." *International Journal of Advance Research in Science and Engineering* 4, no.1(2015): 424-428. Retrieved from <http://www.ijarse.com>.
32. Joshi, N., Mathur, I., & Mathur, S. (2011). "Translation memory for Indian languages: an aid for human translators." In Proceedings of the International Conference & Workshop on Emerging Trends in Technology (pp. 711-714). ACM.
33. Joshi, N., & Mathur, I. (2012). "Design of English-Hindi Translation Memory for Efficient Translation." arXiv preprint arXiv:1210.5517.
34. Singh, S. P., Kumar, A., Darbari, H., Tailor, N., Rathi, S., & Joshi, N. (2017). "Intelligent English to Hindi Language Model Using Translation Memory." In International Conference on Information and Communication Technology for Intelligent Systems (pp. 487-496). Springer, Cham.
35. Yogi, K. K., Joshi, N., & Jha, C. K. (2015). "Quality Estimation of MT-Engine Output Using Language Models for Post-editing and Their Comparative Study." In Information Systems Design and Intelligent Applications (pp. 507-514). Springer, New Delhi.
36. Joshi, N., Mathur, I., Darbari, H., & Kumar, A. (2016). "Quality Estimation of English-Hindi Machine Translation Systems." In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (p. 53). ACM.
37. Mishr G. (2003), Chuni Hui Rachnayein. New Delhi: Vani Prakashan.
38. Gupta, V., Joshi, N., & Mathur, I. (2013). "Subjective and objective evaluation of English to Urdu Machine translation." In Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on (pp. 1520-1525). IEEE.
39. Gupta, V., Joshi, N., & Mathur, I. (2014). "Evaluation of English-to-Urdu Machine Translation." In Intelligent Computing, Networking, and Informatics (pp. 351-358). Springer, New Delhi.
40. Joshi N., Darbari H., Mathur I. (2012) "Human and Automatic Evaluation of English to Hindi Machine Translation Systems." In: Wyld D., Zizka J., Nagamalai D. (eds) Advances in Computer Science, Engineering & Applications. Advances in Intelligent and Soft Computing, vol 166. Springer, Berlin, Heidelberg
41. Joshi, N., Mathur, I., Darbar, H., Kumar, A., & Jain, P. (2014). "Evaluation of Some English-Hindi MT Systems." In Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on (pp. 1751-1758). IEEE.
42. Joshi, N., Mathur, I., Darbari, H., & Kumar, A. (2015). "Incorporating Machine Learning Techniques in MT Evaluation." In Advances in Intelligent Informatics (pp. 205-214). Springer, Cham.
43. Joshi, N., Mathur, I., Darbari, H., & Kumar, A. (2013). HEval: Yet another human evaluation metric. arXiv preprint arXiv:1311.3961.
44. Specia, L, and Atefeh F. (2010) "Estimating machine translation post-editing effort with HTER." Proceedings of the Second Joint EM+/CNGL Workshop Bringing MT to the User: Research on Integrating MT in the Translation Industry (JEC 10).

Web Sources:

- [1] <http://www.yourdictionary.com/Technologydictionarydefinition/> (Retrieved on 15 July 2018)
- [2] <http://kantanmt.com/documents/machinetranslation> (Retrieved 5 Aug 2018)

Author's Profile



Venu Shree is a Research Scholar in the department of English and Modern European Languages, Banasthali Vidyapith. She did her masters in English literature and M.Phil (ELT) from the department of English and Modern European Languages, Banasthali Vidyapith. Her research interests are Computational Linguistics, Machine Translation, Translation, and Linguistics.



Iti Mathur is an Assistant Professor at Banasthali Vidyapith. Her areas of interest are computational

semantics and ontological engineering. Besides this, she is also involved in the development of MT engines for English to Indian languages. She is one of the experts empanelled with TDIL program, Department of Electronics and Information Technology (Deity), Govt. of India, a premier organisation that oversees Language Technology Funding and Research in India. She has several publications in various journals and conferences and also serves on the program committees and editorial boards of several conferences and journals.



Geetha Yadav is currently working as an Associate Professor at Banasthali Vidyapith in the Department of English and Modern European Languages. She has an experience of 20 years of teaching at UG and PG level, including five years of M.Phil. (ELT) and guiding research scholars for Ph.D. She has presented several papers at various national and international seminars and conferences. She has also delivered invited lectures at various institutions and has been a member of the Board of Studies of some institutions. She has published three papers in various anthologies. Her field of ongoing research is narratology.



Nisheet Joshi is an Associate Professor in Department of Computer Science, Faculty Mathematics and Computing. He has done his PhD in the area of Natural Language Processing. Being involved in teaching and research for over 10 years, he has developed the art of explaining even the most complicated topics in a straight forward and easily understandable fashion. He also has vast experience of handling large scale research projects sponsored by various funding agencies. This has helped him in developing practical insights of complex AI systems. He has delivered invited talks in several conferences, workshops and training programs. He has authored several papers in international journals and conferences on various topics in AI and NLP.