Performance Enhancement of Machine Translation Evaluation Systems for English – Hindi Language Pair

Pooja Malik[1, 2]
[1] Gautam Buddha University, Greater Noida, India-201308
[2] Shiv Nadar University, Greater Noida, India-201314
Email: singh.pooja902@gmail.com

Anurag Singh Baghel
Gautam Buddha University, Greater Noida, India-201308
Email: anuragsbaghel@gmail.com

Abstract—Machine Translation (MT) is a programmed conversion in which computer software is utilized to convert manuscripts from one Natural Language (like English) to a different Language (such as Hindi). To process any such conversion, through human or through automatic means, the conversion must be established such that it reinstate the complete sense of a manuscript from its base (source) linguistic into the target language. In this paper, the study of prevailing evaluation systems along with assessing their performance is achieved through the similarity metrics. Moreover, the authors have also presented an improved technique of translation employing features of Natural Language Processing and consequently, to acquire an enhanced and more accurate assessing Machine Translation system, a corpus is selected and the outcomes are compared with the prevailing methods. Besides this, two well-known systems such as Google and Bing decoders are selected to inquire and to assess the study of metrics called similarity metrics through Assessment of Text Essential Characteristics score. This is found to provide more accuracy than prevailing methods. Furthermore, evaluations are tested under various metrics systems like Jaccard similarity metrics, cosine similarity metrics, and sine metrics to deliver enhanced accuracy than prevailing methods.

Index Terms—Machine Translation, Machine Translation Evaluation, Similarity metrics, ATEC Score, Google and Bing Translators.

I. Introduction

MT is a significant computational language field. Computational linguistics fits in the branch of science that pact the linguistic characteristics with the aid of computer science technology [1, 2]. Here, all processing on natural language is completed by machinery (computers). Computation is carried by making an allowance for every known together with possible and essential principles of composition, semantics, and morphology of the elements of speech [3]. The machine must be aware of every of these probable language aspects, but earlier work does not handle the additional needs in MT. Decoding is not a fair substitution in word-for-word [4]. A decipherer needs to comprehend then analyze every element in the manuscript then know in what manner every term might impact another [5]. This needs a complete acquaintance of syntax, arrangement (structure of the sentence), meanings, etc., in the original language along with target languages besides the consciousness of every native region [6].

Machine and Human translation respectively have their portion of encounters [7]. For instance, no two dissimilar translators can generate identical conversions of the analogous script in the pair of the same language, and then this can deliberate numerous rounds of revisions to get customer gratification [8]. Conversely, the better challenge includes in how MT can do publishable quality translations.

Current online along with desktop machine translation schemes neglects several characteristics of languages throughout the conversion [9]. On account of this issue, various uncertainties occur namely polysemy, syntactic ambiguity, structural ambiguity, lexical ambiguity, anaphoric/reference ambiguity, and discourse. Due to these ambiguities, current machine translator is unable to produce precise translation [10].

II. Our Contribution

1. To inquire and assess the performance of Machine Translators via Similarity metrics.
2. Evaluations are tested under various metrics systems like Jaccard similarity metrics, cosine similarity

metrics, and sine metrics to offer more accuracy over prevailing methods.
3. To obtain the accuracy and speed besides computational time.
4. Exact translation within the stipulated time and improved accuracy of the metric tool when associated with an existing method and also human assessment system.

III. RELATED WORKS

Chen et al., (2018) suggested a dissimilar neural methodology to origin dependence-based context illustration for forecasting the conversion. The recommended construction was accomplished for encoding source long-distance dependencies besides capturing functional similarities to enhance predicts translations. To validate this technique, the suggested mode was combined into phrase-based along with hierarchical phrase-based conversion representations, correspondingly. Experimentations on large-scale Chinese-to-English along with English-to-German translation tasks displayed that, the recommended method attained substantial enhancement through baseline systems and then did better over numerous existing context-enhanced approaches [11].

Berger et al., (2017) suggested an innovative probabilistic method for information recovery with the help of ideas in addition to techniques of arithmetic machine translation. The vital constituent of this technique is a numerical model of a manner in which the user might concentrate or "translate" a specified phrase to a question. To estimate the significance of the document to user's query, the researchers assessed the prospect that, the query might be developed as the conversion of the document, along with factor in user's general preferences using a prior distribution over documents [12].

Mallinson et al., (2017) revised bilingual pivoting in neural MT background and then offered a rephrasing model that depended virtuously on neural networks. This model paraphrases significantly in an unremitting space, assessed the grade of the semantic relationship among text fragments of arbitrary length, or produces candidate summaries for every source input. Investigational outcomes across tasks along with datasets presented that neural paraphrases were better related with those acquired with conventional phrase-based pivoting techniques [13].

Koehn et al., (2017) deliberated six trials for decoding of neural mechanism; specifically beam search, uncommon words, extended verses, and the quantity of data for training, alignment of the word, and domain mismatch. The researchers revealed both deficiencies and enhancements through the eminence of phrase-based numerical MT. It was presented that, irrespective of its modern achievements, neural MT still has to beat numerous encounters, most particularly performance out-of-domain then under low supply environments [14].

Zhang et al., (2017) suggested RNN-embed, a character-level sequence-to-sequence technique for learning. This technique stated quantized characters in the scheme of translation are then utilizing a predetermined vocabulary containing a restricted amount of words. In the course of English-to-Chinese subtitle translation, a Recurrent Rural Network (RNN) was inserted in the encoder-decoder system for producing character-level sequence representation. The Gated Recurrent Unit (GRU) in the linguistic system of the encoder was also enhanced. The recommended model was studied in large-scale subtitle dataset compiled by the researchers [15].

Wijaya et al., (2017) offered an enhanced structure for uniting diverse, sparse along with potentially noisy multimodal signals for translations. The concern of learning translations as matrix achievement task was established and hence utilized an efficient and extendable matrix factorization method with Bayesian Personalized Ranking (BPR) for studying translations. The usage of this method was presented in large-scale experimentations. Beginning from minimally trained monolingual word embedding, this technique reliably and very suggestively outperformed contemporary methods by relating these features with extra topographies in a controlled method with BPR [16].

Kong et al., (2017) suggested a rearrangement table-filtering model by making an allowance for the deep neural network in enhancing the concerns of reordering in Statistical MT. The suggested model was assessed in the arena of Uyghur – Chinese as well as English – Chinese machine conversion. The investigation outcomes presented that the excellence of MT in Uyghur-Chinese along with English-Chinese acquired clear enhancements when considered the novel filtered reordering table in decoding procedure and reorganization capability became enhanced [17].

Cox et al., (2018) offered SnipSnap which was an innovation of this technology to protect memory acquisition. SnipSnap has hardware TCB and permitted forensic predictors to gather reliable memory snapshots from the objective machine while offering performance isolation for applications implementing over the target. Investigational assessment of numerous data-intensive workloads exhibited the advantage of this method [18].

IV. METHODOLOGY

This section presents the procedure followed to assess an improved technique of conversion and is compared with two familiar Free Online Machine Translation (FOMT), i.e. Google and Bing translators. In this proposed technique, Natural Language Processing (NLP) was employed for deciphering texts from its base linguistic to the required linguistic. The framework of this translation process is explained in Fig. 1.
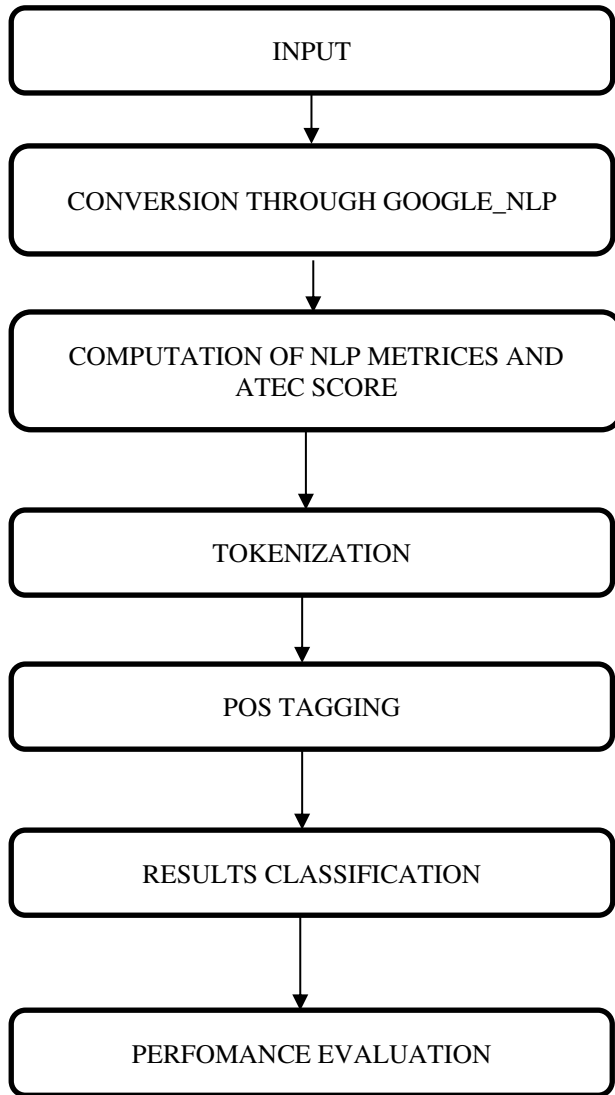
```
        ┌─────────────────────────┐
        │          INPUT          │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐
        │ CONVERSION THROUGH       │
        │ GOOGLE_NLP               │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐
        │ COMPUTATION OF NLP       │
        │ METRICES AND ATEC SCORE  │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐
        │      TOKENIZATION        │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐
        │       POS TAGGING        │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐
        │  RESULTS CLASSIFICATION  │
        └─────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────┐
        │  PERFOMANCE EVALUATION   │
        └─────────────────────────┘
```

Fig.1. The Framework of the Translation Process

Based on this framework, the procedure starts with the selection of input strings or verses that needs to be translated. Here, the conversion is established from English to Hindi, and we have gathered a corpus of judgements of the House of Lords (the HOLJ corpus). Each document contains a header providing structured information (e.g., respondent, appellant, date of hearing), followed by a sequence of (usually five) Law Lords' judgments consisting of free-running text, at least one of which is a substantial speech. Fig. 2 Displays the home screen of MT program by which the translation can be achieved.
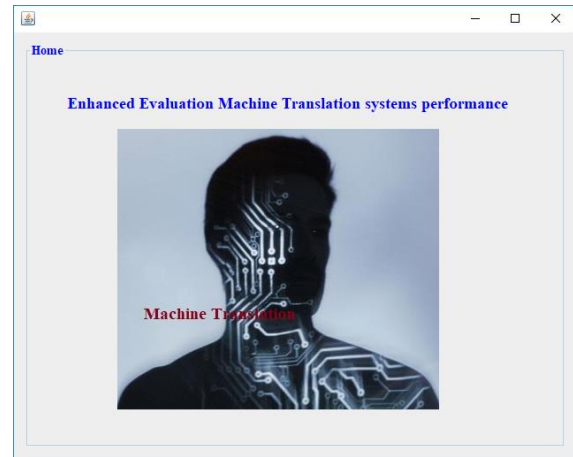

Fig.2. The Home Screen of the Machine Translation System

The subsequent segment is decoding of input verses from English to Hindi by applying the technique of Google_NLP. Also, the corresponding verses are translated using FOMT tools. In comparisons with other translation tools like Bing and Google, this type of conversion appears to be a better option as it involves conversion through NLP technique.

At this point, the chosen sections are compared with reference translation verses for MT. In translation window, the designated verses are displayed on the topmost corner followed by space allocated for translated verses. The extremity of the page is supported with for reference translation sentences region for MT. The translation window in NLP program is shown in Fig. 3.
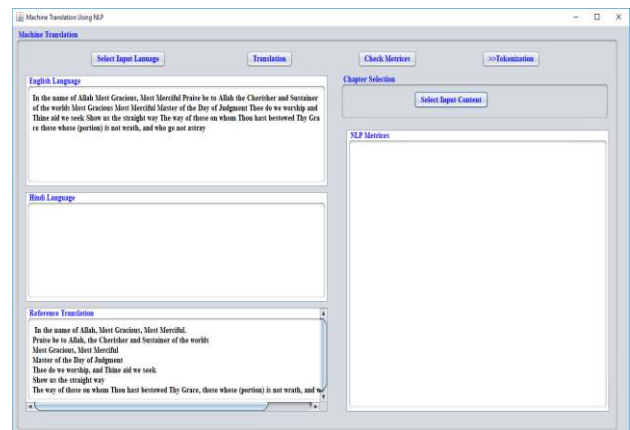

Fig.3. Translation window in Machine Translation using NLP

The translated verses are displayed on translated area text box allocated for it. Fig. 4. Shows the conversion window with translated verses.
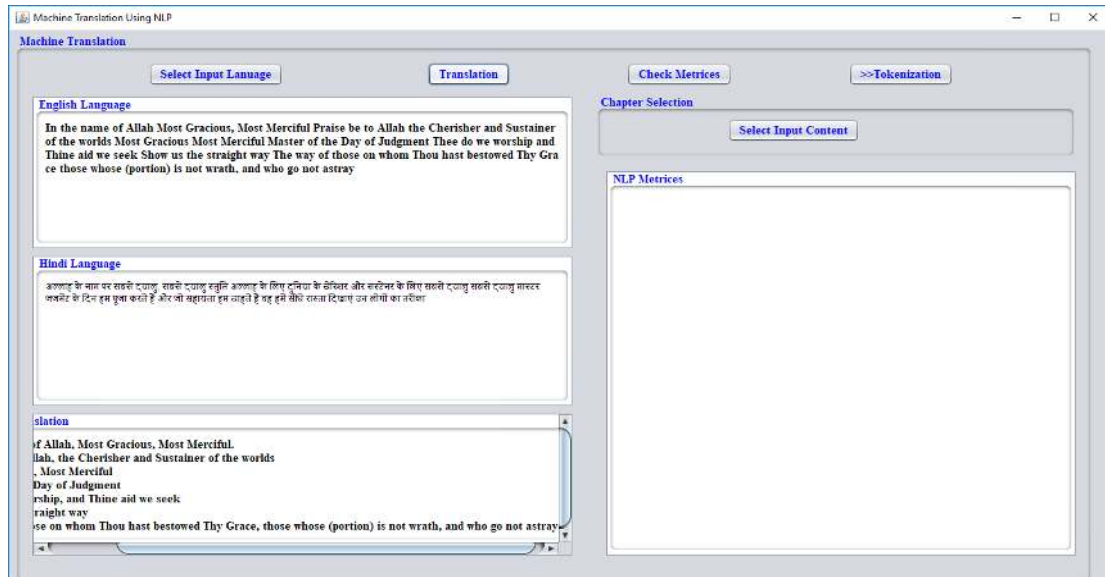
Fig.4. Translation window with translated Hindi verses

This is followed by the computation of NLP metrices and Assessment of Text Essential Characteristics (ATEC) score of translated verses. Here, the calculation of NLP metrices includes the assessment of Input instances, number of correctly classified instances, the correctly classified rate, number of incorrectly classified instances, the incorrectly classified rate and total number of instances.

The computation of score regarding ATEC depends on unigram F-measure to measure the selection of words, i.e. the best matching amid the words for translation and reference translation. Hence, this computation comprises the valuation of Precision (P), Recall (R), F-measure, Accuracy (A), Penalty and ATEC score. Also, the calculation of Precision, Recall, and F-measure requires the tabular assessment of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The valuations can be established by applying the below formulae,

$$\text{Precision (P)} = \frac{True \text{ Positive (TP)}}{True \text{ Positive (TP)} + \text{False Positive (FP)}} \quad (1)$$

$$\text{Recall (R)} = \frac{True \text{ Positive (TP)}}{True \text{ Positive (TP)} + \text{False Negative (FN)}} \quad (2)$$

$$\text{F-Measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (3)$$

$$\text{Accuracy (A)} = \frac{TP + \text{ TN}}{TP + \text{TN} + \text{FP} + \text{FN}} \quad (4)$$

These evaluations are made to estimate the penalty and ATEC score of the translator. The penalty rate depends on the variations in the location of words amid the verse selected for translation and one or more reference translations. The formulation for penalty rate calculation is expressed as,

$$Penalty = \begin{cases} 1 - (\text{PosDiff} \times 4) & \text{if PosDiff} \leq 0.25 \\ 0 & \text{if PosDiff} > 0.25 \end{cases} \quad (5)$$

Here, $PosDiff$ is the degree of alteration in place of words.

With the aid of all values, the ATEC score of the translator is evaluated by employing the formula,

$$ATEC = (F - Measure) \times (Penalty) \quad (6)$$

These assessed values are used to estimate NLP metrices and ATEC score of translator. Subsequently, the procedure of Tokenization in NLP translation flows in the next phase through conversion process. The idea of migrating an arrangement of sections into fragments like into words, symbols, phrases and supplementary elements is known as tokenization. These integrated elements are called as tokens. At this process, the exceptional fonts such as punctuations are rejected and the remaining tokens are feeded as input for text mining process.

After the process mentioned above, the NLP probability of every token is determined in the same window. This is achieved by pressing the pushbutton stating NLP probability. The next course in the conversion procedure deals with POS tagging. POS is nothing but the Parts of Speech of the sentences those are being translated. The POS might be broadly classified into two categories as

1. Closed class types and
2. Open class types

The former comprises relatively fixed membership like prepositions, as new prepositions are rarely coined. On the contrary, the nouns and verbs belong to the latter as new nouns and verbs are continually being created. Four chief open classes occurring in almost all languages are

nouns, verbs, adjectives, and adverbs. This technique of tagging also provides the precision measurement of every token.

This is followed by the procedures of results classifications and assessment of the performance of translator. From the POS tagged tokens and its accuracy value, the outcomes are analyzed and are categorized in the distinct window. In the performance estimation, the performances of conversion through Google, Bing, and Google_NLP are compared. Here, the level of accuracy and score based on ATEC of developed Google_NLP is validated with those of two FOMT translators.

V. RESULTS AND DISCUSSION

This section presents the outcomes acquired through translation of English verses into Hindi verses using NLP technique by Machine translation. This proposed technique is intended in interpreting the stanzas to our convenient language with the help of Natural Language Processing (NLP) technique. The computation of NLP metrices and ATEC score of the translator are shown in Fig. 5. This score is essential for comparing the rate of translation meticulousness among various translation tools.
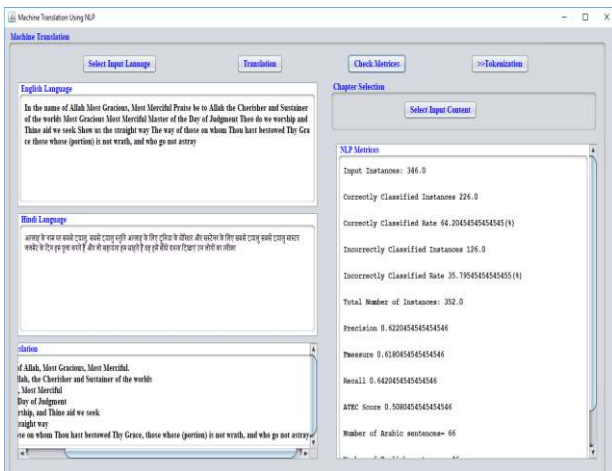
Fig.5. Computation of NLP metrices and ATEC score

After conversion to Hindi, we split the words for NLP process which is called Tokenization process. Fig. 6 shows the outcomes achieved through this process of tokenization.
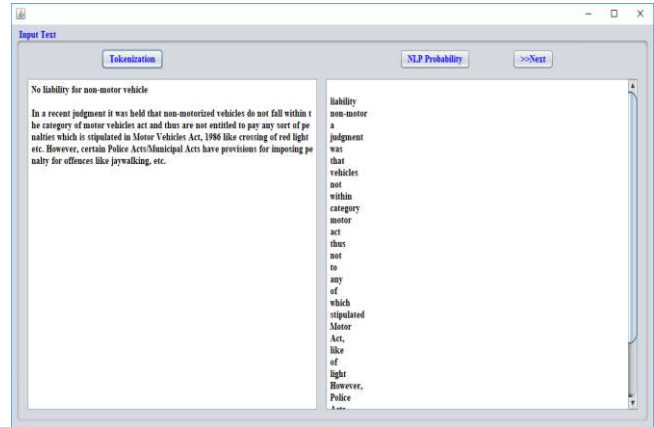
Fig.6. Tokenization

After tokenization, we have calculated probability value for all kinds of tokens. From the obtained results, it may be observed that the detailed exactness of this type of conversion is delivered in a discrete window. This depicts the precision, recall, F-measure and ATEC score of NLP translation technique. This ATEC score regulates the efficiency of translation technique as it determines the precision and penalty of translation technique.
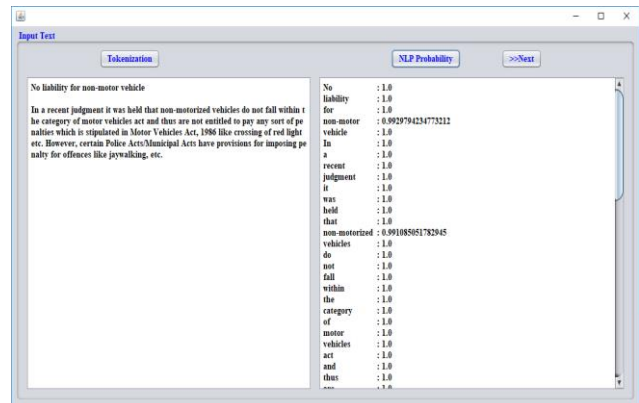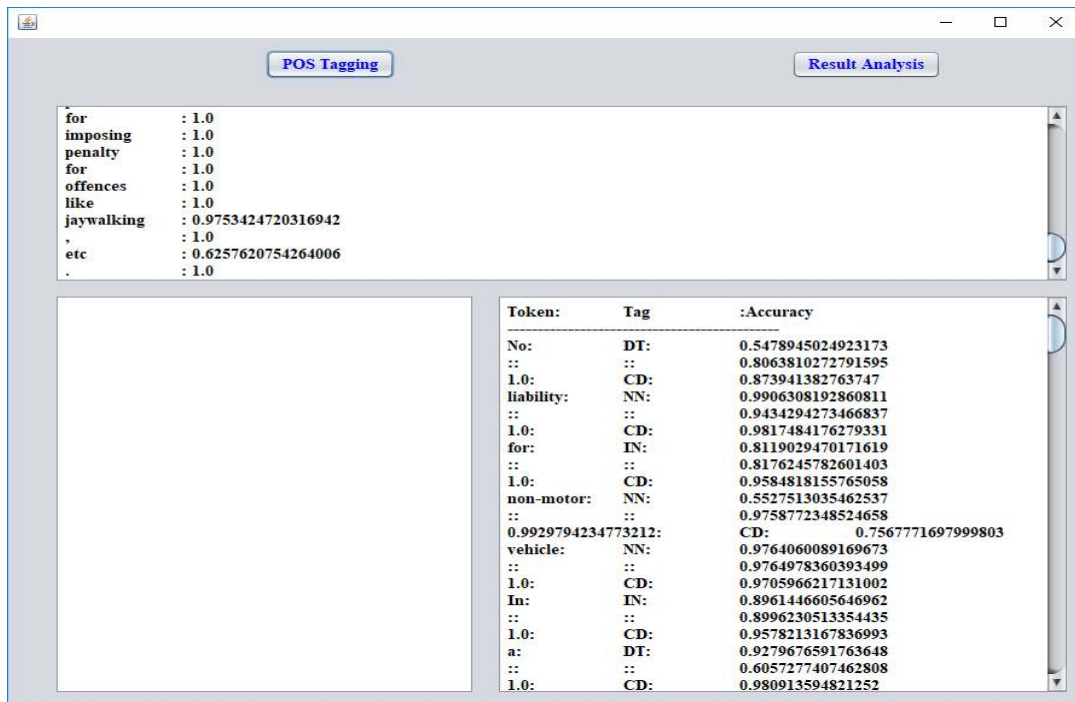
Fig.7. NLP probability

Fig.8. POS Tagging

Fig. 8. shows all the tokens present in a sentence along with their POS tag and accuracy. By accuracy, we mean that how accurately a particular token is being tagged.
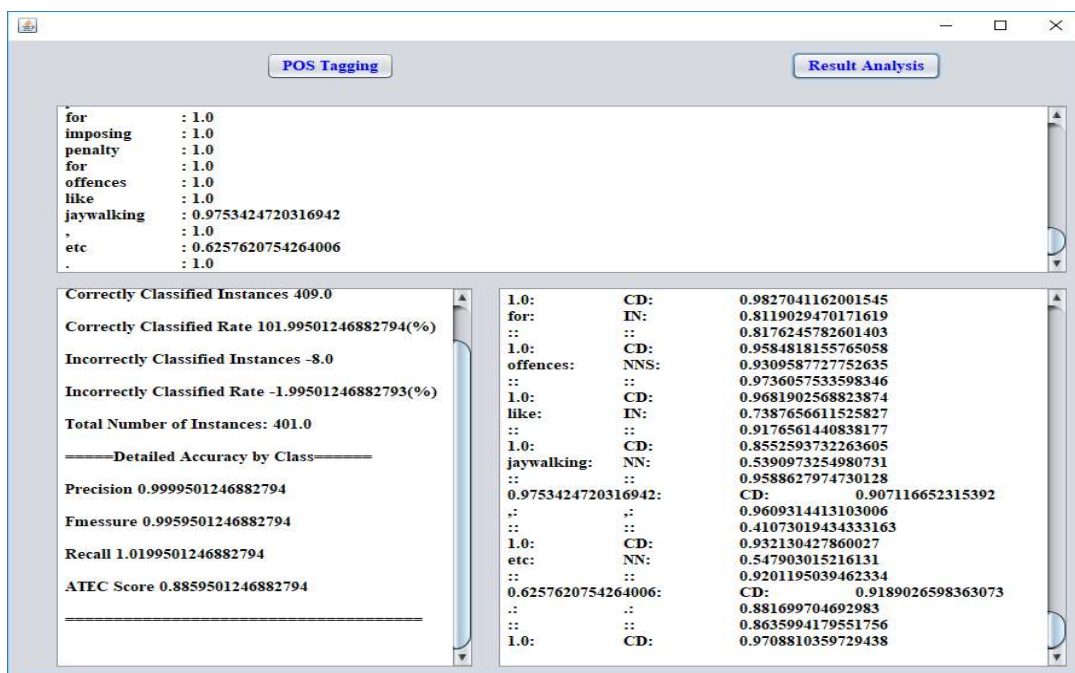

Fig.9. Classification Results

Finally, we have compared ATEC Scores of translated text through Google_NLP Translator and through Google and Bing Translators and the results are shown in Fig. 9.
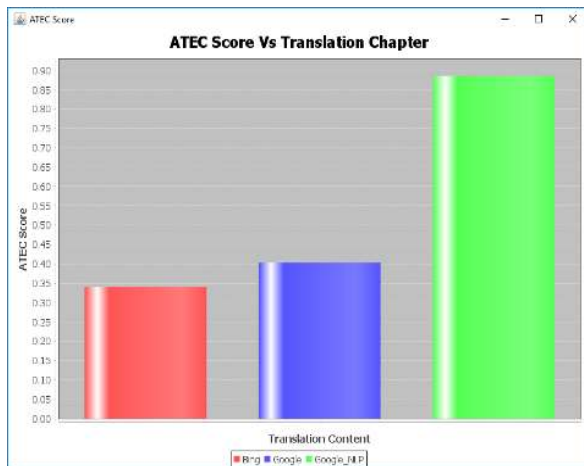
Fig.10. ATEC score comparison

On comparison of ACET scores of two FOMT tools and NLP based translator, the ACET score of NLP based translator came out as 0.508 whereas the score of Google and that of Bing was 0.402 and 0.342 respectively.

Fig. 10 shows the comparison of all the three translators, i.e., Google, Bing, and Google_NLP. Also, the length of the translated string plays a chief role in the determination of ATEC score of the translator.

The ATEC score depending on the length of translated verses is displayed in Fig. 11.
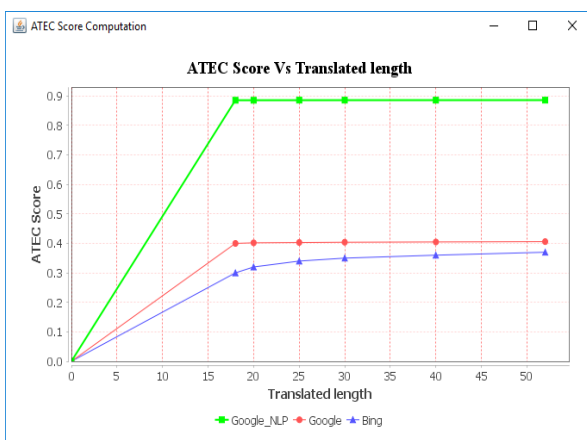

Fig.11. ATEC score Vs. Translated length

VI. COMPARISON WITH EXISTING TECHNIQUES

From Fig 11, it was observed that, the ATEC score of Google_NLP is found to be higher than the other two already existing translators. Hence, it is evident that the usage of NLP tool in rendition provides more accurate translation than other two translation techniques.

VII. CONCLUSION

The necessity of translating English to any natural linguistics of India is greatly increasing. Among those natural languages, Hindi is the most spoken Indian language as almost in every Indian state, people can converse in Hindi. In this study, an enhanced translator tool is proposed which utilizes the NLP technique for translation. With the help of this tool, the verses from the HOLJ corpus were translated from English to Hindi. The results obtained from this Machine translator was evaluated with two renowned FOMT systems and are validated. The outcomes depicted that, the accuracy and ATEC score of this Google_NLP translator system is much better as compared to other two FOMT systems. The drawbacks of these kinds of Machine translators are that they cannot achieve 100 percent accuracy in translation like that of the translation achieved through human efforts.

REFERENCES

[1] Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[2] Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[3] Cabrerizo, F. J., Morente-Molinera, J. A., Pedrycz, W., Taghavi, A., & Herrera-Viedma, E. (2018). Granulating linguistic information in decision making under consensus and consistency. *Expert Systems with Applications*, *99*, 83-92.

[4] Berthiaume, R., Daigle, D., & Desrochers, A. (Eds.). (2018). *Morphological Processing and Literacy Development: Current Issues and Research*. Routledge.

[5] Romero-Fresco, P., & Pöchhacker, F. (2018). Quality assessment in interlingual live subtitling: The NTR Model. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, *16*.

[6] Sárosi-Márdirosz, K. (2014). Problems related to the translation of political texts. *Acta Universitatis Sapientiae, Philologica*, *6*(2), 159-180.

[7] Woll, N. (2018). Investigating dimensions of metalinguistic awareness: what think-aloud protocols revealed about the cognitive processes involved in positive transfer from L2 to L3. *Language Awareness*, 1-19.

[8] Alkhatib, M., & Shaalan, K. (2018). The Key Challenges for Arabic Machine Translation. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 139-156). Springer, Cham.

[9] Lin, X. V., Wang, C., Zettlemoyer, L., & Ernst, M. D. (2018). NL2Bash: A Corpus and Semantic Parser for Natural Language Interface to the Linux Operating System. *arXiv preprint arXiv:1802.08979*.

[10] Ciobanu, D. (2018). Collaborative Student Translation Projects. *Multilingual Writing and Pedagogical Cooperation in Virtual Learning Environments*, 222.

[11] Mollo, G., Jefferies, E., Cornelissen, P., & Gennari, S. P. (2018). Context-dependent lexical ambiguity resolution: MEG evidence for the time-course of activity in left inferior frontal gyrus and posterior middle temporal gyrus. *Brain and language*, *177*, 23-36.

[12] Chen, K., Zhao, T., Yang, M., Liu, L., Tamura, A., Wang, R. & Sumita, E. (2018). A Neural Approach to Source Dependence Based Context Model for Statistical Machine Translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *26*(2), 266-280.

[13] Berger, A., & Lafferty, J. (2017, August). Information retrieval as statistical translation. In *ACM SIGIR Forum* (Vol. 51, No. 2, pp. 219-226). ACM.

[14] Mallinson, J., Sennrich, R., & Lapata, M. (2017). Paraphrasing revisited with neural machine translation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (Vol. 1, pp. 881-893).

[15] Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. *arXiv preprint arXiv:1706.03872*.

[16] Zhang, H., Li, J., Ji, Y., & Yue, H. (2017). Understanding subtitles by character-level sequence-to-sequence learning. *IEEE Transactions on Industrial Informatics*, *13*(2), 616-624.

[17] Wijaya, D. T., Callahan, B., Hewitt, J., Gao, J., Ling, X., Apidianaki, M., & Callison-Burch, C. (2017). Learning Translations via Matrix Completion. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 1452-1463).

[18] Kong, J., Yang, Y., Wang, L., Zhou, X., Jiang, T., & Li, X. (2017). Filtering Reordering Table Using a Novel Recursive Autoencoder Model for Statistical Machine Translation. *Mathematical Problems in Engineering*, *2017*.

[19] Cox, G., Yan, Z., Bhattacharjee, A., & Ganapathy, V. (2018). Secure, Consistent, and High-Performance Memory Snapshotting.

Authors' Profiles

Pooja Malik received her M. Tech (Computer Science and Engineering) degree from Banasthali University, Rajasthan, India and B. Tech (Computer Science and Engineering) degree from Uttar Pradesh Technical University, Lucknow, India. Currently, she is a research scholar at the School of Information and Communication Technology, Gautam Buddha University, Greater Noida, India. Her research interests include Natural Language Processing, Machine Translation and Machine Translation Evaluation.

Anurag Singh Baghel has completed his M.Tech (Electronics) in 2000 and D.Phil in 2010 both from University of Allahabad, Allahabad, India. He served as Lecturer (Electronics) from 2004 to 2011 in Banasthali University, Tonk, India and since then he is working as Assistant Professor (Computer Science) in Gautam Buddha University, Greater Noida, India. His areas of interest are – Metaheuristics and applications, Software Engineering, and Big Data. He has published more than 40 research publications in various journals and international conferences. He has supervised more than 40 M.Tech Dissertations. Presently, seven scholars are pursuing Ph.D. under his supervision.