A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)

Abstract

The paper demonstrates the qualitative evaluation of the English to Urdu Machine Translation Systems, namely PBSMT and NMT hosted on Google's Translate. This system is popularly known as Rosetta, formerly governed by Phrase-based approach and is presently governed by the neural module of source and target languages. In this study, a model corpus set of 100 English sentences has been applied out of 1k cross-domain data considering various types of verbs as input text to evaluate the output of the online systems in Urdu.

In order to evaluate the output text in a qualitative manner, the Inter-translator Agreement (IA) of three human translators has been considered with their scores on a five-point scale. The scores are calculated by the Fleiss' Kappa statistical measure with regard to comprehensibility and grammaticality on the basis of which error analysis and suggestions have been provided for improvement. The Kappa scores of PBSMT for comprehensibility and grammaticality are 0.24 and 0.22 respectively which is indicative of the fact that on both counts the scores are not up to the mark. Furthermore, the system has also been quantitatively evaluated on the basis of word error rate (21.11%) and sentence error rate (72.39%). On the contrary, NMT module has Kappa scores of 0.61 and 1 on comprehensibility and grammaticality respectively. So far as WER and SER are concerned, NMT has 32.58% and 28% respectively. In addition, all the erroneous entities have been analyzed through computational typology. The strategy for evaluation is to evaluate the Urdu output text based on the five-point scale with scores that range from 0-4 where 0 refers to incomprehensible or ungrammatical, 1 = little meaning or disfluent, 2 = neutrality, 3 = comprehensible or grammatical and 4 suggests flawless in both cases.

Keywords: PBSMT, NMT, Google's Translate, MT, Urdu, Indo-Aryan, NLP, Fleiss' Kappa

Overview

As discussed in Castilho et al. (2018), since the advent of the Machine Translation (MT) or automated translation, new methods, approaches and techniques have really created high expectations among researchers. On one hand, the qualitative approaches have paved the way for a graded or incremental improvements in contrary to the significant improvements exhibited by the statistical approaches. Among the statistical techniques, Neural Machine Translation (NMT) has recently emerged as an innovative and robust technique as it has generated a lot of attention because of its high qualitative outputs in comparison to its counterparts.

===
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   154
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

This study represents the qualitative evaluation of the English-Urdu Machine Translation systems namely Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) hosted on Google's Translate. The system is also otherwise known as Rosetta earlier governed by the phrase-based model which is presently replaced by the neural model. The rationale for the consideration of the qualitative evaluation of the systems is that quantitative evaluation doesn't prove to be adequate and sufficient in bringing out the reasons behind the error-prone outputs. Although every NLP MT platform is vying for adapting neural framework presently, it is not a panacea for all the issues in the domain. We have considered a representative corpus of 100 sentences out of 1k corpus ranging across various categories of verbs as input in English for the evaluation of Urdu output sentences.

Machine Translation

It is an automated translation process of text from source language (SL) to target language (TL). It is one of the sub-fields of Natural Language Processing (NLP) the sole objective of which is to enquire the application of the software for translating speech or text from one language to another one.

MT Systems - A Review of Literature

AnglaBharti is an English to Indian languages computer aided translation (CAT) system launched by Sinha et al. (1995) at Indian Institute of Technology, Kanpur in 1991. AnglaBharti-II was developed by Sinha in 2004 addressing shortcomings of the latter model and incorporating Generalized Example Base and Raw Example Base.

Anubharti is a template-based Machine Translation of Hindi-English which applies Hybrid Example-based model which is an amalgamation of the strategies used in both the approaches of rule-based and example based for translation.

Anusaaraka (1995) project was started by Prof. Rajeev Sangal and is presently the director at IIT BHU. It is a software which translates texts from English to Hindi languages. Anusaaraka is modelled upon Panini's Ashtadhyayi which is based upon grammar rules and aims at mixing ancient Indian shastras and modern technologies.

MaTra is a hybrid system trained on cross-domain corpus text and represents a pragmatic approach to language engineering. It is primarily utilized in the project on Cross Lingual Information Retrieval (CLIR) (Rao, 2001).

AnglaHindi (Sinha, 2003) is an example-based English-Hindi version of AnglaBharti which can handle all types of sentences up to maximum 20 words each. It further integrates a rule and example-based approaches during the process of post-editing.

Mantra, an English-Hindi MT trained specifically in the personal administration domain data, is developed by CDAC, Pune. The system applies Lexicalized Tree Adjoining Grammar (LTAG) which maps a lexical tree in SL to its counterpart lexical tree in TL.

===
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   155
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

Shiva and Shakti are the Machine Translation systems jointly developed by Carnegie Mellon University, USA, IIIT, Hyderabad and IISc, Bangalore, India that translate texts from English to Hindi.

Shakti MT system (Bharati et al., 2003) has been modelled and developed in such a way that they can produce MT systems for newly incorporated languages frequently and is hybrid in nature whereas Shiva is Example-Based in nature.

Sampark MT system has been developed by 11 consortia member institutions in the project named ILMT project funded by the TDIL program of the DeiTY, Govt. of India. It has created NLP resources for 9 Indian language pairs resulting in Machine Translation for 18 languages.

Anuvaadaksh system has been developed in the EILMT project, funded by the TDIL program of the DeiTY, Govt. of India which translates English text into 8 Indian languages. It has been conceptually designed and prepared by 13 institution consortium members and has integrated four MT technologies: Tree Adjoining Grammar, Statistical-base, EBMT and Analyze and Generate Rules.

The first version of Microsoft's MT system such as Bing Translator has been designed, developed and managed by Microsoft Research between the years 1999 and 2000. It was exploited to translate the whole gamut of Microsoft Knowledge Base into Spanish, French, German and Japanese.

UCSG MAT is developed by the University of Hyderabad which is a machine-aided translation platform utilized in order for translating English texts as input into Kannada as output and also needs post-editing. Its primary purpose is to parse an English input sentence applying the UCSG parsing technology which was developed by Dr. K. Narayana Murthy and thereafter translates it into Kannada language applying the bilingual dictionary of English-Kannada, Morphological Generator of Kannada and the linguistic rules for translation.

Universal Networking Language (UNL) is an international project of United Nations University in which IIT, Mumbai participates. It is an inter-lingua for semantic representation. Currently, this project is working in languages such as English, Hindi and Marathi where any of these languages is taken as SL and converted into UNL and then again de-converted from UNL to TL.

Tamil Anusaaraka has been developed by K. B. Chandrasekhar Research Centre, Anna University, Chennai. Its primary aim was to develop a Human Aided Machine Translation System for the language pairs English-Tamil. It has three major components viz. morphological analyzer of English, mapping system unit and the Tamil language generator.

MAT by Jadavpur University: Rule-based English-Hindi MAT is in Jadavpur University, Kolkata. It uses transfer-based approach and its purpose is to work for new sentences.

Anuvaadak 5.0 system was developed for a general purpose of automatic translation from English-Hindi by Super Infosoft private limited, Delhi under the leadership and supervision of Mrs. A. R. Choudhury. For each specific domains like official, formal, agriculture, linguistics, technical and administrative, it contains inbuilt dictionaries.

==
Language in India www.languageinindia.com ISSN 1930-2940 18:10 October 2018
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   156
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

Anubaad Hybrid Machine Translation System was developed by Bandyopadhyay at Jadavpur University, Kolkata in 2004 for translating English news headlines into Bengali. The current version of the system works at the sentence level. Statistical MT was developed by IBM India Research Lab at New Delhi and its sole purpose was to translate texts between English and Indian languages.

Oriya Machine Translation System (OMTrans) is developed by Utkal University, Vanivihar. The SL is English while TL is Oriya in this system. It serves the purpose of sense disambiguation using the N-gram model.

Tamil-Hindi machine-aided translation system was developed by Prof. C. N. Krishnan at Anna University at KB Chandrashekhar (AU-KBC) Research Centre, Chennai. It is based on Anusaaraka Machine Translation System and applies a lexical level translation.

Dr. K. Narayana Murhty has developed English-Kannada MAT system which is situated at Resource Centre for Indian Language Technology Solutions (RC-ILTS), University of Hyderabad. It is essentially based on a transfer-based approach which is applied to the documents related to government circulars.

Hinglish machine translation system has been developed by Sinha and Thakur (2005) in 2004 which is based on pure Hindi to English forms. It has been executed after having incorporated an additional layer to the existing English to Hindi (AnglaBharti-II) and Hindi to English machine translation (AnuBharti-II) systems which was also developed by Sinha himself. English to (Hindi, Kannada, Tamil) and Kannada to Tamil language-pair example-based machine translation systems were developed by Balajapally et al. (2006). It encapsulates a bilingual dictionary which comprises of phonetic-dictionary, words-dictionary, phrases-dictionary and sentence dictionary.

Punjabi to Hindi machine translation system was developed by Josan and Lehal at Punjabi University Patiala in 2007. It is based on a direct word-to-word translation mapping approach. Hindi to Punjabi machine translation system was conceptualized and developed by Goyal and Lehal (2010) at Punjabi University Patiala in 2009. It is also based on direct word-to-word mapping from SL to TL.

Apni Urdu is an English-Urdu MT Platform which is an incorporation of the English-Urdu machine translated texts. As inputs, it applies some English texts which are readily available online and for outputs; it supports Urdu Unicode fonts. This platform is beneficial for simple constructions.

The Apertium Machine Translation Platform was developed by Forcada et al. (2009) which provides a readymade framework for developing new platforms for any language pairs. Lavie et al. (2004) describe an MT system known as a trainable transfer-based Hindi-English MT platform which is designed to further the development of MT for less resourced languages.

Behera et al. (2016a) have discussed about the divergence patterns between English to Bhojpuri language pairs where they have discussed about various syntactic and semantic divergences between English and Bhojpuri language pairs.

===
Language in India www.languageinindia.com **ISSN 1930-2940** **18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   157
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

Behera et al. (2016b) have proposed to transform the IMAGACT4ALL ontology into an MT platform where animated videos along with their sentential written expressions could be translated into other Indian languages either as visual representations or as written forms.

Google Translate

It was released by Google Inc. on 12 Jan, 2010. Its revised version came in 2011. This platform is providing services for many languages around the globe. Among Indian languages, it supports Hindi, Urdu, Bengali, Gujarati, Punjabi, Telugu, Tamil, Kannada, Marathi, Malayalam and Nepali (Muzaffar and Behera, 2014).

Phrase-based Statistical Machine Translation

It is modeled upon the phrase-based language translation model which translates the SL phrase into the TL phrase. The PBSMT used is Moses (Koehn, 2009), MGIZA (Gao and Vogel, 2008) is applied for training word alignments, and KenLM (Heafield, 2011) is applied to language model training and scoring. It is a linear combination of different features such as phrase and word penalty, 5-gram language model and phrase translation probabilities with some modification techniques of Kneser-Ney smoothing (Kneser and Ney, 1995; Chen and Goodman, 1998) and the below mentioned advanced features such as a 5-gram operation sequence model (Durrani et al., 2013); a hierarchical lexicalized reordering model (Galley and Manning, 2008); sparse features which indicate phrase pair frequency, length of phrase, and sparse lexical features. For English-Russian pair of languages, it employs a transliteration mapping model for unknown sort of words (Durrani et al., 2014). Feature weights optimization is applied in order to increase the level of BLEU score with the batch MIRA (Cherry and Foster, 2012) on a within-domain tuning set that has been extracted (and held out) from the in-domain training data. Muzaffar and Behera (2014) have provided a detailed description on the errors pertaining to verb markers in Urdu while dealing with the translation platform of Google and Bing. Muzaffar et al. (2016a) have proposed a parser based on Pāniniān framework for successfully analyzing errors related to case markers in English-Urdu Machine Translation in general. Muzaffar et al. (2016b) have provided a detailed description on the divergent patterns between English and Urdu after observing the outputs collected from Google and Bing MT platforms. Muzaffar and Behera (2016c) have dealt with the concepts of equivalence, gain and loss in Machine Translation while experimenting on Google and Bing. Gupta and others (2013) have conducted both subjective and objective evaluations of English to Urdu Machine Translation.

Neural Machine Translation

It focuses on semantics of SL and TL and thus semantically makes efficient translation than PBSMT. It involves building a single neural network that maps SL and TL aligned bilingual texts and is designed and trained so as to "maximize the probability of a correct translation" (Bahdanau et al., 2014), when given input text to translate without external linguistic information. This interest is shared by many in the language service industry, where there is a need for improved MT quality and better quality estimation to "help reduce the frustrating aspects of post-editing" (Etchegoyhen et al., 2014). NMT results in the latest shared tasks have quickly matched or surpassed those of PBSMT systems, even after so many years of development of PBSMT systems (Sennrich et al., 2017; Bojar et al., 2016). As according to the reported recent studies on NMT, it can be vehemently affirmed that one can observe a significant increase in quality if one considers the comparison of NMT with PBSMT applying either automatic (Bahdanau et al., 2014; Jean et al., 2015), or human evaluations (Bentivogli et al., 2016; Wu et al., 2016). Although the initial NMT experiments have exhibited a significant increase in results, human evaluations on NMT output have not been conducted on a large-scale basis.

===
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics) 158
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

Urdu Language
Urdu is an Indo-Aryan language (Muzaffar et al, 2015) and is a member of the New-Indo-Aryan which is a subgroup of Indo-European family of languages. It is spoken in most of the areas of Indian sub-continent (Muzaffar & Behera, 2014). According to the census 2011, there are approximately fifty million speakers of Urdu in India.

Methodology
This section has been divided into corpus collection method and data analysis method.

Method of Corpus Collection
We have collected 1000 corpus of English sentences selectively considering different types of verbs. Out of them, 100 representative sentences have been taken as input to evaluate the output data in Urdu.

Method of Data Analysis
Data has been analysed by considering the Inter-translator Agreement (IA) of three human translators with their scores on a five-point scale which range from 0-4 where 0 stands for incomprehensible or ungrammatical, 1 means little meaning or disfluent, 2 refers to neutrality, 3 stands for comprehensible or grammatical and 4 suggests flawless. The scores are calculated by the Fleiss' Kappa statistical measure with regard to comprehensibility and grammaticality on the basis of which error analysis and suggestions have been provided for improvement.

Evaluation
Evaluation is considered to be one of the stepping stones for measuring the efficiency of an NLP application (Mitkov, 2003). It can be of two broad categories: human and automatic or statistical based on two approaches to research i.e. qualitative and quantitative.

Qualitative vs Quantitative Evaluation
In qualitative evaluation, judgments of different translators have been considered for measuring the output texts of the MT systems. Contrarily, the quantitative evaluation conducts a statistical measurement of the performance. Therefore, it is indispensable that we need to perform a qualitative evaluation of the systems so as to figure out their performance and various bottlenecks constricting the efficiency (Behera et al., 2016).

The Role of Qualitative Evaluation
Qualitative evaluation leaves us some room, for it evaluates any system with regard to reliability and adequacy or comprehensibility and acceptability, and so on. It gives us much background to the underlying issues and challenges as to why a given system under performs. Thus, we have undertaken a detailed qualitative evaluation of Google Translate where we have taken English as the SL and Urdu as the TL. Evaluation has been conducted at three levels: percentage agreement, Fleiss' Kappa Agreement, WER and SER.

Fleiss' Kappa
"Kappa" is a statistical measure applied to test the inter-rater reliability judgements between two or more raters qualitatively. There are two types of Kappa: e.g. Cohen and Fleiss. The former is

==
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   159
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

applied to the evaluation of the output for the agreement between two persons. On the other hand, the latter is applied to the agreement on the evaluation among multiple raters.

In 1971, Fleiss extended Cohen's Kappa for measuring IA reliability of more than two raters. Fleiss' Kappa is defined as the output of the agreement above chance divided by the degree of agreement actually achieved. It takes values between 0-1 where 1 signifies complete agreement. The scores of three native translators according to Fleiss 'Kappa have been taken into account.

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

Categories	Qualitative/Human					
	Comprehensibility			Grammaticality		
	p_a	p_e	K	p_a	p_e	K
Urdu (PBMT)	0.450	0.270	0.246	0.450	0.286	0.229
Urdu (NMT)	0.7	0.22	0.61	1	0.37	1

Table 1. Qualitative Evaluation of Google's MT Platform on Kappa Statistics

On one hand, the Kappa score of PBSMT for comprehensibility is 0.24 and grammaticality is 0.22 which are not up to the mark. On the other hand, NMT has Kappa scores of 0.61 for comprehensibility and 1 for grammaticality. The Kappa scores of PBSMT and NMT for comprehensibility are 0.24 and 0.61 respectively. A comparatively higher score of NMT depicts that it performs better than PBSMT in terms of comprehensibility. So far as grammaticality is concerned, PBSMT has 0.22% and NMT has 1%. The score of NMT suggests to the fact that it has almost the perfect agreement among raters and thereby there are no erroneous patterns observed.

Quantitative/Statistical (PBMT)		Quantitative/Statistical (NMT)	
WER[1]	**SER**[2]	**WER**	**SER**
21.11%	72.39%	32.58%	28.00%

Table 2. Quantitative Evaluation of Google's MT Platform on WER & SER

In this section, the higher number of scores is proportionate to the higher amount of erroneous linguistic patterns at the corresponding level. PBSMT and NMT have further been quantitatively evaluated on the basis of word error rates: 21.11% and 32.58% respectively. WER being higher for NMT implies that at the word level, the PBSMT outperforms its counterpart. So far as sentence error rates are concerned, PBSMT has 72.39% whereas NMT has 28.00%. The SER score being higher for the PBSMT is indicative of the fact that NMT outperforms its counterpart at the sentence level.

	PBSMT	NMT
Computational Errors	2.5%	1.8%

Table 3. Distribution of Computational Errors

---

[1] Word Error Rate
[2] Sentence Error Rate

==
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   160
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

The scores of Computational Errors on PBSMT and NMT are 2.5% and 1.8% respectively. A comparatively lower score of NMT refers to the fact that it is less-prone to errors and thus performs better than the PBSMT.

Analysis of Computational Errors
Tokenization
It is one of the computational processes of text segmentation which classifies different characters from the other preceding and following ones. In the following example, the apostrophe is not usually translated in most of the instances. The following representative example exhibits that although the genitive form is translated properly, the plural oblique form (laDakI-yoM) has not been translated from the English counterpart.
English: Girls' college (apostrophe + plural)
Itrans: laDakI ke mezabAna

Named Entities ('institute' missing)
Named entities are the proper nouns of any language. In the below-instantiated example, the part 'institute' is missing when translated from English to Urdu. Firstly, the Urdu output suggests that it is merely a transliteration output and not a translation of the original SL input text. Furthermore, the word 'institute' is not translated at all.
English: All India Institute of Medical Sciences
Itrans: Al InDiA medikAl sAIns

Morphological
These errors pertain to both the noun and verb morphology of Urdu getting wrongly translated. In the below example, 'left over' English adjectival phrase is getting erroneously translated into Urdu as a verbal phrase.
English: left over pieces of food
Itrans: khAne ke TukaDe TukaDoM para ChoDa diyA

Chunking
Chunking is the computational process of grouping the local words, viz. nominal categories and verbal categories to be grouped under their respective single broad category. Here, in the below-stated example, 'call off' as a verbal phrase has been inverted and hence causing problems during translations.
English: I am **calling** the meeting **off**.
Itrans: maiM mulAqAta kara rahA huM.

Parsing (covert you)
It is the computational process of having syntactic relations between and among different parts of the sentences. In the following example, the covert you in both the sentences has not been translated appropriately.
English: Silence please!!!
Itrans: barAye karam qhAmoshI

Multi-words (idioms)

===
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   161
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

It is one of the computational categories, the meaning of which is different from their constituent parts. There are many sub-categories such as idioms and phrases, reduplications, echo-word formations. In the following instance, the idiomatic expression has been literally translated.
English: Pull up the socks
Itrans: jarAbeM khichoM

Conclusion

One of the limitations of the current research is that we have taken into consideration only the computational errors for the purpose of analysis. Another most important limitation is that we have applied a limited amount of data of general domain and fed the Google MT system for this study. Depending upon the specificity and nature of the domain of the data, there will definitely be positive or negative impact on the quality of the output.

In this paper, we have presented a qualitative evaluation of Google's PBSMT and NMT between English and Urdu. We have applied the Fleiss' Kappa method to measure agreement among multiple raters-cum-translators. We have further demonstrated analysis of errors computationally. We have compared Google's PBSMT and NMT platforms and have observed that NMT performs well for this pair of languages. We would further like to replicate and extend this study to other Indian languages. We would like to reiterate a point that certainly NMT outperforms the PBSMT on the yardstick of the qualitative nature of the TL output. But it is not a panacea for all the issues and challenges pertaining to MT. For machines to perform at par with the humans or outperform them, let's say, Machine Learning and NLP have to go a long way.

Acknowledgements

===

References

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Behera, P., Singh, R. & Jha, G. N. (2016). Evaluation of Anuvadaksh (EILMT) English-Odia Machine-assisted Translation Tool. In *Proceedings of WILDRE-3 (LREC-2016),* Portoroz, Slovenia, pp. 110-117.

Behera, P., Maurya, N., Pandey, V. & Banerjee, E. (2016a). Dealing with Linguistic Divergences in English-Bhojpuri Machine Translation. In Proceedings of WSSANLP-6, COLING-2016, Osaka, Japan.

_____, Muzaffar, S., & Jha, G. (2016b). The IMAGACT4ALL ontology of animated images: implications for theoretical and machine translation of action verbs from English-Indian languages. In *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016)* (pp. 64-73).

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: a case study. *arXiv preprint arXiv:1608.04631*.

Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huck, M., ... & Negri, M. (2016). Findings of the 2016 Conference on Machine Translation. In *ACL 2016 First Conference on Machine Translation (WMT16)* (pp. 131-198). The Association for Computational Linguistics.

==
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   162
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Way, A., & Georgakopoulou, P. (2018). Evaluating MT for massive open online courses. *Machine Translation*, 1-24.

Durrani, N., Haddow, B., Heafield, K., & Koehn, P. (2013). Edinburgh's machine translation systems for European language pairs. In *Proceedings of the Eighth Workshop on Statistical Machine Translation* (pp. 114-121).

Durrani, N., Koehn, P., Schmid, H., & Fraser, A. (2014). Investigating the usefulness of generalized word representations in SMT. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 421-432).

Dwivedi, S. K., & Sukhadeve, P. P. (2010). Machine translation system in Indian perspectives. *Journal of computer science*, *6*(10), 1111.

Etchegoyhen, T., Bywood, L., Fishel, M., Georgakopoulou, P., Jiang, J., Van Loenhout, G., ... & Volk, M. (2014). Machine Translation for Subtitling: A Large-Scale Evaluation. In *LREC* (pp. 46-53).

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5): 378.

Forcada, M. L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., ... & Tyers, F. M. (2011). Apertium: a free/open-source platform for rule-based machine translation. *Machine translation*, *25*(2), 127-144.

Garje, G. V., & Kharate, G. K. (2013). Survey of Machine Translation Systems in India. *International Journal on Natural Language Computing (IJNLC)*, *2*(4), 47-67.

Gao, Q. and Vogel, S. (2008). Parallel implementations of word alignment tool. In Software Engineering, Testing, and Quality Assurance for Natural Language Processing. Association for Computational Linguistics, 49–57.

Goyal, V., & Lehal, G. S. (2010). Web based Hindi to Punjabi machine translation system. Journal of Emerging Technologies in Web Intelligence, *2*(2), 148-151

Gupta, V., Joshi, N. & Mathur, I. (2013, August). Subjective and objective evaluation of English to Urdu Machine translation. In *Advances in Computing, Communications and Informatics (ICACCI),* pp. 1520-1525. IEEE.

Heafield, K. (2011, July). KenLM: Faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation* (pp. 187-197). Association for Computational Linguistics.

Jean, S., Firat, O., Cho, K., Memisevic, R., & Bengio, Y. (2015). Montreal neural machine translation systems for WMT'15. In *Proceedings of the Tenth Workshop on Statistical Machine Translation* (pp. 134-140).

Koehn, K. (2009). Statistical Machine Translation. Cambridge University Press.

Mitkov, R. (2003). The Oxford Handbook of Computational Linguistics. Oxford University Press, New York.

Muzaffar, S. & Behera, P. (2014). An Error Analysis of the Urdu Verb Markers: A Comparative Study on Google and Bing Machine Translation Platforms. *Aligarh Journal of Linguistics*, 4 (1-2), pp. 199-208.

_____, Behera, P., Jha, G. N., Hellan, L. & Beermann, D. (2015). The TypeCraft Natural Language Database: Annotating and Incorporating Urdu. *Indian Journal of Science and Technology*, *8*(27), October, Chennai, India.

_____, Behera, P. & Jha, G. N. (2016a). A Pāniniān Framework for Analyzing Case Marker Errors in English-Urdu Machine Translation. *Procedia Computer Science (Elsevier)*, *96*, pp. 502-510.

===
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   163
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems

_____, Behera, P. & Jha, G. N. (2016b). Classification and Resolution of Linguistic Divergences in English-Urdu Machine Translation. In *Proceedings of WILDRE-3 (LREC-2016), (ISBN: 978-2-9517408-8-4),* Portoroz, Slovenia, pp 43-49.

_____. & Behera, P. (2016c). The Concepts of Equivalence, Gain and Loss (Divergence) in English-Urdu Web-Based Machine Translation Platforms. In *Proceedings of International Conference of South Asian Languages (ICOSAL-12)*, CELMTS, University of Hyderabad.

Sennrich, R., Firat, O., Cho, K., Birch, A., Haddow, B., Hitschler, J., ... & Nădejde, M. (2017). Nematus: a toolkit for neural machine translation. *arXiv preprint arXiv:1703.04357*.

Sinha, R. M. K., & Jain, A. (2003). AnglaHindi: an English to Hindi machine-aided translation system. *MT Summit IX, New Orleans, USA*, 494-497.

Sinha, R. M. K. (2004, November). An engineering perspective of machine translation: anglabharti-II and anubharti-II architectures. In *Proceedings of International Symposium on Machine Translation, NLP and Translation Support System (iSTRANS-2004)* (pp. 10-17).

Sinha, R. M. K., & Thakur, A. (2005). Machine translation of bi-lingual hindi-english (hinglish) text. *10th Machine Translation summit (MT Summit X), Phuket, Thailand*, 149-156.

Sinha, R. M. K., Sivaraman, K., Agrawal, A., Jain, R., Srivastava, R., & Jain, A. (1995, October). ANGLABHARTI: a multilingual machine aided translation project on translation from English to Indian languages. In *Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century, IEEE International Conference on* (Vol. 2, pp. 1609-1614). IEEE.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Klingner, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

===

Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Department of Linguistics, Faculty of Arts
Aligarh Muslim University
Aligarh, Uttar Pradesh-202002, India
Sharmin.muzaffar@gmail.com

Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)
Centre for Linguistics
School of Language, Literature and Culture Studies
Jawaharlal Nehru University
New Delhi-110067, India

==
Language in India www.languageinindia.com **ISSN 1930-2940 18:10 October 2018**
Sharmin Muzaffar, M. A., Ph.D. (Linguistics)
Pitambar Behera, M.A., B.Ed., M.Phil., Ph.D. (Linguistics)   164
A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems