

# Assessing Query Translation Quality Using Back Translation in Hindi-English CLIR

**Ganesh Chandra**

Department of Computer Science, BBAU (A Central University), Lucknow, U.P, India  
E-mail: ganesh.iiscgate@gmail.com

**Sanjay K. Dwivedi**

Department of Computer Science, BBAU (A Central University), Lucknow, U.P, India  
E-mail: skd200@yahoo.com

**Abstract**—Cross-Language Information Retrieval (CLIR) is a most demanding research area of Information Retrieval (IR) which deals with retrieval of documents different from query language. In CLIR, translation is an important activity for retrieving relevant results. Its goal is to translate query or document from one language into another language. The correct translation of the query is an essential task of CLIR because incorrect translation may affect the relevancy of retrieved results.

The purpose of this paper is to compute the accuracy of query translation using the back translation for a Hindi-English CLIR system. For experimental analysis, we used FIRE- 2011 dataset to select Hindi queries. Our analysis shows that back translation can be effective in improving the accuracy of query translation of the three translators used for analysis (i.e. Google, Microsoft and Babylon). Google is found best for the purpose.

**Index Terms**—Back-Translation, BLUE, METEOR, TER & query translation, transliteration.

## I. INTRODUCTION

Information retrieval (IR) has become the primary way for users to understand the world by exchanging the different types of information. The purpose of IR is to search relevant documents from a large collection of documents against a user's query [1].

IR can be classified into three types: monolingual information retrieval (MIR), cross-lingual information retrieval (CLIR) and multi-lingual information retrieval (MLIR). In MIR, query and document are of same language whereas in CLIR, query and document are of different languages. In MLIR, a user searches documents from a multilingual collection of documents against a query of single language [2, 53].

With the enormous increase of information in different languages on Internet, search engine allows users to retrieve documents different from his/her language [52]. Such type of information retrieval is known as Cross - Lingual Information Retrieval (CLIR) [3, 43, 44]. The development of network technology and information globalization increases the demand of CLIR contents

because it removes language barrier, reduces communication cost and promote information exchange and usage [4, 5, 51].

Various forums such as TREC, CLEF & NTCIR organizes a large number of conferences, tracks and workshops on CLIR [6]. Each of these forums represents the following list of languages:

- FIRE (Forum for Information Retrieval Evaluation): Hindi, English, Bengali, Marathi, Tamil, Telugu, Gujarati, Odia, Punjabi & Assamese.
- TREC (Text Retrieval Conference): Spanish, Chinese, German, French, Italian & Arabic.
- CLEF (Cross Language Evaluation Forum): French, German, Italian, Spanish, Dutch, Finnish, Russian.
- NTCIR (NII Testbeds and Community for Information access Research): Japanese, Chinese and Korean.

These forums provide an evaluation infrastructure and suitable facilities for testing various techniques of CLIR. A huge amount of information on the Web is available in English. India is a multilingual country where most of the people used the Hindi language for communication and searching of documents. The number of Web users is increasing continuously day by day that creates a strong platform for bilingual research [54].

CLIR depends on machine translation for removing the language barrier between source language and target language. Query translation is an important activity of CLIR that can be defined as the process of obtaining the correct equivalent translation(s) of each word of query into another language(s) by various resources. The accuracy of the translated query depends on translating mechanism. Some of the most effective resources used for query translation are bi-lingual dictionaries, parallel corpora and comparable corpora [7].

Evaluation of machine translation (either a query or document) is a challenging task [55, 56, 57]. Various human judgments are used to evaluate the translation quality like fluency and adequacy [8, 58].

The accuracy of machine translation (MT) is usually

evaluated by comparing the translated output with reference output or by human judgment. Some important strategies used for evaluation of translation accuracy are BLUE, METEOR, TER, GTM, NIST, PORT, LEPOR, AMBER, ROUGE, WER and ROSE etc.

BLUE (Bi Lingual Evaluation Understudy) is one of the most important techniques which is based on n-gram match precision. Its concept was introduced by Papineni, Roukos, Ward, and Zhu [9].

In METEOR [10, 45], evaluation of translation is based on unigram matching between machine-produced translation and human-produced reference translation. It resolves the problems of BLUE.

The concept of TER (Translation Edit Rate) was introduced by Snover and Dorr in 2006 [11]. It works on counting transformations rather than n-gram matches. This method represents the number of edits needed to change a candidate translation to the reference translation, normalized by the length of the reference translation. Possible edits include insertion, deletion, substitution of a single word and word sequence.

GTM (General Text Matcher) measures the similarity of different texts. It computes precision, recall and f-measure for accuracy measurement of text translations [12].

The name NIST came into existence from National Institute of Standards and Technology which is based on n-gram technique as similar to BLUE. In this, for computing the brevity penalty shortest length of references is used, whereas BLUE uses average length of references. Another big difference between BLUE and NIST is informativeness. BLUE treats n-gram equally whereas NIST does not treat equally all n-gram. It assigns more weights to that n-gram which more is informative and assigns less weight to those that are less informative [13].

PORT (Precision-Order-Recall Tuning) is an evaluation metric that performs an automatic evaluation of machine translation [14]. This metric has five components such as precision, recall, strict brevity penalty, ordering metric and redundancy penalty. It does not require any external resources for tuning of machine translation. It performs better evaluation than BLUE when translation is hard or at the system level and segment level [59].

LEPOR, an evaluation metric combines many factors such as precision, recall, sentence-length penalty and n-gram based word order penalty. This metric develops the higher system level correlation with human judgments in comparison to other metrics such as BLUE, METEOR, and TER. The hLEPOR metric is the higher version of LEPOR that utilizes the harmonic mean [15].

AMBER (A Modified Blue, Enhanced Ranking), one of the automatic translation evaluation metric which is based on BLUE but includes some additional features such as recall, extra penalties and some text processing variants [16]. It describes four different strategies: N-gram matching, Fixed-gap n-gram, Flexible -gap n-gram and Skips n-gram [66].

ROUGE (Recall-Oriented Understudy for Gisting

Evaluation) is a set of metrics which came into existence in 2003 [60]. It uses a unigram co-occurrence method between summary pairs [17]. This metrics set contain following evaluation metrics: ROUGE-N (based on n-gram co-occurrence statistics), ROUGE-L (based on Longest Common Subsequence (LCS)), ROUGE-W (based on weighted LCS statistics), ROUGE-S (based on Skip-bigram co-occurrence statistics) and ROUGE-SU (based on a Skip-bigram plus unigram-based co-occurrence statistics).

The concept of WER (Word Error Rate) was introduced by Niessen et al. in 2000 for automatic and quick MT evaluation [18]. It is based on Levenshtein distance which was given by Vladimir Levenshtein in 1965 [65]. This distance can be defined as the minimum number of operations (i.e. insertion, deletion or substitution) between two strings that are required to transform one string into another.

ROSE is sentence level automatic evaluation metric which contains only simple features for quick computation. It can be defined as a linear model where Support Vector Machine (SVM) is used to train its weight. It is based on two training approaches: linear regression and ranking [19].

The rest of the paper is organized as follows. In Section 2, we describe the related work. Section 3 & 4, presents query translation and back-translation respectively. Section 5 describes experimental results and analysis. Section 6 discusses this work and last but not least Section 7 presents the conclusion.

## II. RELATED WORK

In CLIR, different translation approaches have been used for query translation. There are three types of resources have been widely used in CLIR for query translation: dictionary based approach, corpora based approach (parallel & comparable) and machine translation based approach.

In 1996, Hull and Grefenstette [20] used a bilingual dictionary to derive all possible translation of query for retrieving the relevant result. This is the simplest method but decreases the time efficiency of retrieved documents. To resolve this problem, Hull [21] in 1997 used "OR" operator for translating query and also used weighted Boolean method for a assigning degree to each translation.

In 1997, Ballesteros and Croft used [22] "local context analysis" method to enhanced the dictionary-based query translation. In 1997, Carbonell et al. [24] uses corpus - based approach for query translation in CLIR, where bilingual corpora used for extracting translations of query term. Their experimental result shows that corpus-based query translation performed much better than other.

In 1998, Dorr and Oard [23], evaluate the effectiveness of semantic structure for query translation and found that the technique of semantic structure was less effective than dictionary and MT-based query translation

In 1999, Xu et al. [25] performs the comparison of three techniques: machine translation, structural query translation and their own technique. In this research work

they used Linguistic Data Consortium (LDC) lexicon of English and Chinese languages. Their experimental result shows that the success rate can increase by using a bilingual lexicon and parallel text.

Gao et al. [26] perform the experimental analysis of three techniques: decaying co-occurrence, noun phrase and dependency translation for Chinese –English CLIR. In this work, they used TREC collection of Chinese dataset. The outcome of this work indicates that decaying co-occurrence method performs 5% better than the other model.

In 2004, Braschler [27] used three types of approaches for query translation: output of an MT system, novel translation approach (based on thesaurus) and dictionary-based translation. Unfortunately, this combination does not provide much better results due to lower coverage of thesaurus-based and dictionary-based translation methods. In 2009, Gao et al. [28], used machine learning methods for query translation in CLIR.

In 2011, Herbert [29] use a similar approach as used by Braschler for translating certain phrases and entities using Wikipedia on Google MT system, found improvement in retrieved result of English-German CLIR. In 2012, Ture [30] used an internal representation of MT system for query translation and found significant improvement in retrieved results.

In 1970, R.W. Brislin [31] used back translation and found that it is a highly useful method for translating international questionnaires and surveys, as well as diagnostic and research instruments.

In 2002, Dasqing He et al [32], worked on query translation of English/German CLIR by using two methods: (i) back translation (ii) Keyword in Context (KWIC). Their analysis suggests that the combined result of these two methods can provide effective results.

In 2006, Grunwald [33] also used the back translation for the purpose of quality control. In 2008, U.Ozolins [34] worked on back translation and found that back translation is a quality control approach that can help to achieve the good transfer of meaning across languages in international health studies.

In 2009, Rapp [35] used OrthoBLEU method for solving the problem of evaluation methods such as BLUE which require reference translation. Their result shows that OrthoBLEU can improve the evaluation accuracy of the back translation.

In 2015, M. Miyabe et al. [36] worked to verify the validity of back translation. Results show that back-translation is a useful method only when high level translation accuracy is not needed.

### III. QUERY TRANSLATION

Translation is the process of transferring information into an equivalent structure of one language into another language [47]. It is an important factor that can reduce the performance of CLIR as compared to MIR (Monolingual Information Retrieval).

In CLIR three types of translation are possible: query translation, document translation and dual translation

(both query & document translation) [38].

Query translation is the process of translating each term present in user query of one language into another language. The effectiveness of query translation depends on the method of translation that can express user's need.

Query translation can be achieved by a dictionary, corpus and machine translation [37]. In dictionary translation, query terms are processed linguistically and only keywords are translating using machine-readable dictionaries. Dictionary based approach also has some drawbacks and benefits. Uses of dictionaries are very simple and these are also available for many language pairs. Unfortunately, these also have some shortcomings: limited coverage. For example, usually, dictionaries do not contain a proper noun.

In corpus based translation, query terms are translated on the basis of multilingual terms extracted from parallel or comparable documents collection. In parallel corpus, collections of text are translated into one or more languages. In comparable corpus, collections of text are not translated text but cover the same topic area like news on BBC and CNN. Translations that can be obtained through parallel corpora are more accurate than comparable corpora. Comparable corpora are noisier because these are not an exact translation of documents.

In machine translation, query terms are automatically translated from one language into another language by using a context.

In CLIR, the relevancy of retrieved documents typically depends on the size of queries. Query translation approach performs better than document translation because of less implementation cost & computational time. Query translation also requires less space as compared to document translation. The small size of queries makes query translation simple and economically efficient for researchers.

### IV. BACK TRANSLATION

Transliteration and translation are the two ways used to convert words from one language into another language. It plays an important role in CLIR and can be defined as phonetics translation of words between two languages with different writing system [61]. It is highly useful in the development of speech processing, multilingual resources, and text [38, 62].

In CLIR transliteration can be performed by two methods: pivot method and direct method. In pivot method, before converting the words of a source language into the target language, source language words are firstly converted into pronunciation symbol and then converted into target language words. Pronunciation symbol is the International Phonetic Alphabet for notation of all languages [40, 63]. The direct method is corpus-based where an intermediate state is not required. Transliteration solves the OOV (out-of-vocabulary) problem which occurs in the translation of queries/documents. For example, in Hind-English CLIR, if translation system fails to translate Hindi words into the English language than transliteration can be used to

translate such words.

Translation helps individual to communicate in nonnative languages. But it is still very difficult to remove the language barrier. So, there is the great importance of correct translation in today's cross-lingual or multilingual environment. It is the major contributing factor for the development of the cross cultural environment in the world. It also helps in the development of science and technology.

In CLIR, language barrier or inaccurate translation prevents a user from retrieving effective results [48]. In order to retrieve relevant results across languages, machine translation plays an important role [49]. Accurate translation of user queries is required for retrieving documents in CLIR.

Back-translation [34, 46, 50] can be defined as the process of translating, translated query back to original query. Back-translated queries are obtained by two step procedure: (1) translation of original query to target language query and (2) translation of target language query back to original language query.

For example as shown in figure1, Hindi query i.e. "दुर्लभ खगोलीय घटनाएँ, (Durlabh Khagoliye Ghatnayn)" is translated into the English language i.e. "Rare Astronomical Events" than again English query is translated back into Hindi language i.e. "दुर्लभ खगोलीय घटनाओं, (Durlabh Khagoliye Ghatnao)". Morphological factor occurs with the word (घटनाएँ, घटनाओं) in a query that may affect the relevancy of retrieved documents.

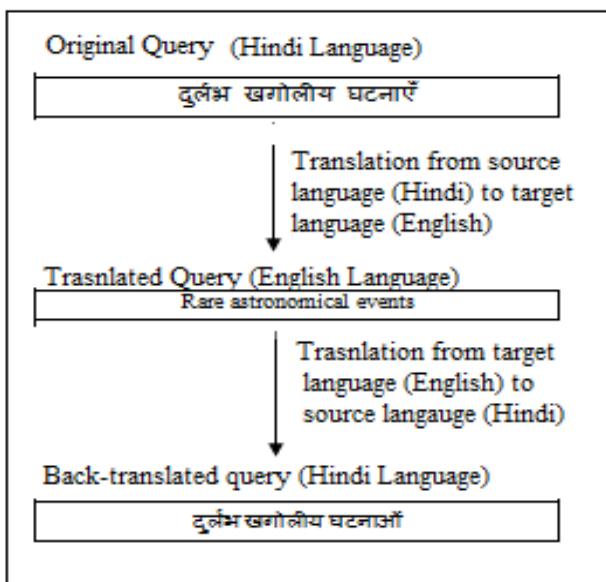


Fig.1. Procedure of back-translation for Hindi-English CLIR

Back-translation can also be called as round-trip translation because it performs the two journeys: the outward journey and forward journey. If back-translation result found bad, it becomes very difficult to tell where the translation (i.e. outward or return translation) went wrong.

Many professional used back-translation for evaluating

the quality and accuracy of the translation. This process does not require the prior knowledge of target language. It is an excellent way of avoiding errors in making a decision.

Back-translation is very useful in a global market because it creates the bridge between cultures and distances.

Many areas such as medical, academic, business etc used back-translation as an effective way of transferring information. For example, WHO (World Health Organisation) controls many medical organizations that used back-translation as a quality control process in various health studies at international level [32]. The process of back-translation involves a technique called decentering. Decentering technique means the process of modifying the translation of original and target language version [64].

Back-translation and translation are two different techniques that differ from each other. Table1 describes the comparative analysis between back-translation and translation.

Table 1. Comparison of Translation and Back Translation

Properties	Translation	Back Translation
Accurate Evaluation	Not Easy (reference translation is required)	Easy (reference translation is not required)
Time complexity	Less (due to single translation)	More (due to double translation)
Precision	Cannot be calculated for all queries (reference translation is not possible for all queries)	Can be calculated for all queries (original query can be treated as reference translation)
Pre-knowledge	Knowledge of translated language is required	Not required
User's	Experts	Common man

## V. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, an experiment is performed on 50 Hindi queries of FIRE (Forum for Information Retrieval Evaluation) dataset for Hindi-English CLIR. In order to evaluate the translation accuracy following steps are performed:

- Step1: Run original query of Hindi language.
- Step2: Translate Hindi query to the English language.
- Step3: Perform back-translation for translated query.
- Step4: Apply 1-gram (word-to-word match) method for evaluation of translation and back-translation.

The concept of Weighted N-gram Model was introduced by Babych and Hartely in 2004 [41]. An n-gram is an excellent technique for efficient evaluation of machine translation. It is widely used in various fields such as probability, communication theory, data compression and computational linguistics.

We performed the translation and back translation by using ImTranslator which provides the most convenient access to the online translation services offers by Google

and other translators for more than 50 languages [42].

Table 2, 3 & 4 describes the Precision, Recall & F-measure values of translation (Hindi-English) and back translation (Hindi-English-Hindi) respectively that are performed using Google, Microsoft, and Babylon translators and are calculated as follows:

$$\text{Precision} = \frac{\text{correct}}{\text{output-length}} \quad (1)$$

$$\text{Recall} = \frac{\text{correct}}{\text{reference-length}} \quad (2)$$

$$\text{F-measure} = \frac{\text{precision} \times \text{recall}}{(\text{precision} + \text{recall})/2} \quad (3)$$

Table 2. Precision value

Translator	Translation (Hindi-English)	Back Translation (Hindi-English-Hindi)
Google	75.44%	69.64%
Microsoft	71.16%	62.42%
Babylon	75.97%	70.08%

Table 3. Recall value

Translator	Translation (Hindi-English)	Back Translation (Hindi-English-Hindi)
Google	73.44%	71.42%
Microsoft	66.32%	64.15%
Babylon	71.2%	69.62%

Table 4. F-Measure

Translator	Translation (Hindi-English)	Back Translation (Hindi-English-Hindi)
Google	75.77%	73.12%
Microsoft	70.73%	62.44%
Babylon	72.45%	69.49%

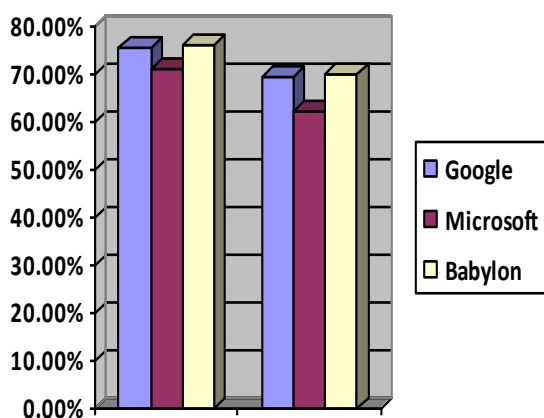


Fig.2. Precision of translation and back translation

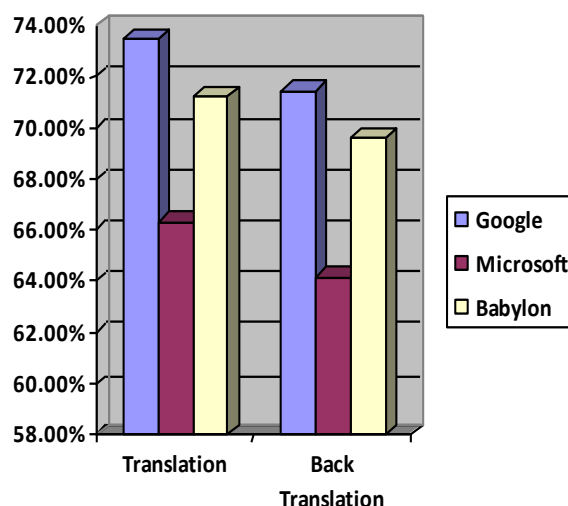


Fig.3. Recall value of translation and back translation

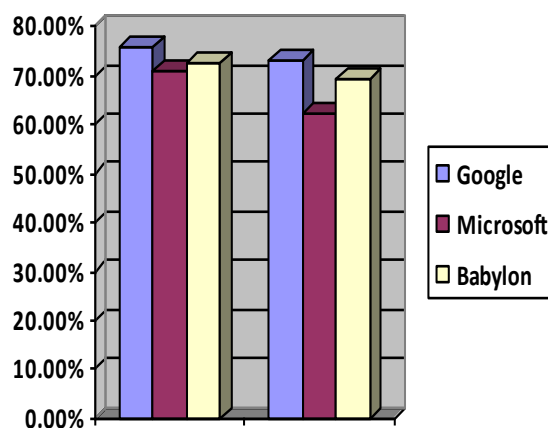


Fig.4. F-measure value of translation and back translation

## VI. DISCUSSION

In this section, we present an experimental analysis that validates the utility of translation & back translation in CLIR. We performed the two types of experimental analysis: (i) comparison of translation and back translation (ii) comparison of query translation by three translators i.e. Google, Microsoft & Babylon.

Values of Precision, Recall, and F-Measure for both automated translation and back translation are computed manually using 1-gram method but without considering the order of translated words in query.

In a case of Hindi-to-English translation, an accuracy of English translated queries is computed by comparing the queries with the expert's English queries (FIRE experts' query).

In case of English (translated query)-to-Hindi query translation, accuracy of Hindi translated queries are computed by comparing the translated queries (i.e. translated Hindi queries) with original Hindi queries.

The Precision (table2), Recall (table3) & F-measure (table4) are computed for evaluation of translation and back-translation using three translators (i.e. Google, Microsoft & Babylon). The purpose of these three scores is to compute the translation quality.

The F-measure score (which considered both Precision and Recall results) of query translation are 75.77%, 70.73% & 72.45% for Google, Microsoft & Babylon translators respectively. The F-measure score of query back-translation are 73.12%, 62.44% & 69.49% for Google, Microsoft & Babylon respectively.

Table 5. Comparison of query translation & back translation

S.No.	Google	Microsoft	Babylon
Correct Translation & Back-Translation (CTBT)	84%	78%	82%
Incorrect Translation & Back-Translation (ITBT)	4%	10%	4%
Correct Translation & Incorrect Back-Translation (CTIBT)	10%	10%	10%
Incorrect Translation & Correct Back – Translation (ITCBT)	2%	0.0%	2%
No Translation & Back-Translation (NTBT)	0.0%	2%	2%

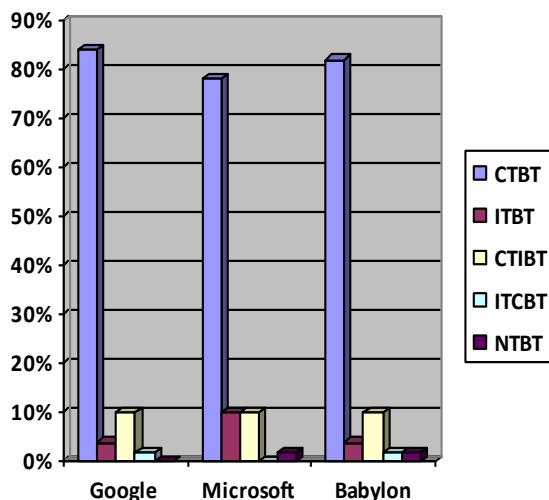


Fig.5. Comparison of translators

Here, incorrect translation of query means that all words of a query are wrongly translated or provides incorrect sense. Correct or incorrect query translations are judged using linguistics, reference queries and original queries (in the case of back translation).

Results of table 5 are divided into 5 types: (i) Correct

Translation & Back-Translation (CTBT), (ii) Incorrect Translation & Back-Translation (ITBT), (iii) Correct Translation & Incorrect Back-Translation (CTIBT), (iv) Incorrect Translation & Correct Back – Translation (ITCBT) and (v) No Translation & Back-Translation (NTBT). Case3 states that the back-translation gives 10% wrong translation (i.e. Google, Microsoft, Babylon) when queries are correctly translated. Hence, back translation is not a 100% reliable method for accessing query translation quality.

Out of three translators, results of Google translator in Hindi-English translation and English-Hindi translation (back translation) are higher than the remaining two translators (Microsoft and Babylon). So Google translator is an effective translator which can be used for the relevancy improvement of CLIR.

## VII. CONCLUSION

Query translation is the major issues which are responsible for poor performance of retrieved results in CLIR. We work out here for evaluation of automatic query translation in CLIR. In this paper, the concept of back translation is used to check the effectiveness of translation in CLIR. We also performed the comparative analysis of three translators for CLIR on 50 queries on a FIRE-2011 dataset.

Back translation is the simplest technique to check the correctness of any translation for a common man. Back translation is also beneficial, when reference translations are not available. An experimental result also shows that the Google translator is more effective in CLIR in comparison to other translators: Microsoft and Babylon.

## REFERENCES

- [1] B. Zhou, Y. Yao, "Evaluating information retrieval system performance based on user preference," journal of intelligent information systems, Springer link, vol. 34, issue 3, pp. 227-248, June. 2010.
- [2] Liang, Ye, et al. "Multilingual Information Retrieval and Smart News Feed Based on Big Data." 12<sup>th</sup> Web Information System and Application Conference (WISA). IEEE, 2015.
- [3] Grefenstette, Gregory, ed. Cross-language information retrieval. Vol. 2. Springer Science & Business Media, 2012.
- [4] Campos, Ricardo, et al. "Survey of temporal information retrieval and related applications." ACM Computing Surveys (CSUR) Vol. 47.2, 2015.
- [5] Kumar, Aarti and Das Sujoy. "Topology for Linguistic Pattern in English-Hindi Journalistic Text Reuse." International Journal of Information Technology and Computer Science (IJITCS) Vol 8, pp 75-86, 2016.
- [6] S.K. Dwivedi and G. Chandra. "A Survey on Cross-Language Information Retrieval," International Journal on Cybernetics & Informatics (IJCI) Vol.5, No.1, Feb 2016.
- [7] Phye, Soe Lai. "Development of Myanmar-English Bilingual WordNet like Lexicon." International Journal of Information Technology and Computer Science (IJITCS) 6.10: 28, 2014.
- [8] Sharma, Manisha, and G. N. Purohit. "Evaluation of machine translation." Proceedings of the International

- Conference & Workshop on Emerging Trends in Technology. ACM, 2011.
- [9] K. Papineni, S. Roukos, T. Ward, T. and W.J. Zhu, 2002. "BLEU: a method for automatic evaluation of machine translation," In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02). Stroudsburg, PA, USA, pp. 311-318, 2002.
- [10] Denkowski, Michael, and Alon Lavie. "Meteor 1.3: Automatic metric for reliable optimization and evaluation of machine translation systems." Proceedings of the Sixth Workshop on Statistical Machine Translation. Association for Computational Linguistics, 2011.
- [11] M. Snover, B. Dorr, R. Schwartz, L. Micciulla & J Makhoul, "A study of translation edit rate with targeted human annotation," Proceedings of association for machine translation in the Americas. Vol. 200. No. 6. 2006.
- [12] O'Brien, Sharon. "Towards predicting post-editing productivity." Machine translation Vol. 25 No.3, pp 197-215, September 2011.
- [13] G. Doddington, "Automatic Evaluation of Machine Translation Quality using N-gram Co-occurrence Statistics", Proceedings of 2<sup>nd</sup> Human Language Technologies Conference (HLT-02). San Diego, CA, pp. 128-132. 2002.
- [14] Chen, Boxing, Roland Kuhn, and Samuel Larkin. "Port: a precision-order-recall mt evaluation metric for tuning." Proceedings of the 50<sup>th</sup> Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 2012.
- [15] L. F. Aaron and S. Lidia, "LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors", Proceedings of COLING 2012: Posters, Mumbai, pp. 441-450, December 2012.
- [16] Chen, Boxing, Roland Kuhn, and George Foster. "Improving AMBER, an MT evaluation metric." Proceedings of the Seventh Workshop on Statistical Machine Translation. Association for Computational Linguistics, 2012.
- [17] Bojar, Ondrej, et al. "Findings of the 2014 workshop on statistical machine translation." Proceedings of the Ninth Workshop on Statistical Machine Translation. Association for Computational Linguistics Baltimore, MD, USA, 2014.
- [18] S. Niessen, F. J. Och, G. Leusch and H. Ney, "An evaluation tool for machine translation: Fast evaluation for MT research." In Proceedings of the 2<sup>nd</sup> International Conference on Language Resources and Evaluation, 2000.
- [19] X. Song and T. Cohn. "Regression and ranking based optimization for sentence level machine translation evaluation." Proceedings of the Sixth Workshop on Statistical Machine Translation. Association for Computational Linguistics, 2011.
- [20] D.A. Hull and G. Grefenstette. "Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval". In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 49-57, 1996.
- [21] D.A. Hull, "Using Structured Queries for Disambiguation in Cross-Language Information Retrieval," In Electronic Working Notes of the AAAI Spring Symposium on Cross-Language Text and Speech Retrieval, 1997.
- [22] L. Ballesteros and W. B. Croft, "Phrasal Translation and Query Expansion Techniques for Cross-Language Information Retrieval," In Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 84-91, 1997.
- [23] B.J. Dorr and D. W. Oard, "Evaluating Resources for Query Translation in Cross-Language Information Retrieval," In Proceedings of the 1st International Conference on Language Resources and Evaluation, pp. 759-764, 1998.
- [24] J.G. Carbonell, Y. Yang, R. E. Frederking, R. D. Brown, Y. Geng and D. Lee. "Translingual Information Retrieval: A Comparative Evaluation". In Proceedings of the 15th International Joint Conference on Artificial Intelligence, pp. 708-714, 1997.
- [25] J. Xu, R. Weischedel and C. Nguyen, "Evaluating a probabilistic model for cross-lingual information retrieval." Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2001.
- [26] J. Gao, J. Y. Nie and M. Zhou, "Statistical query translation model for cross language information retrieval", ACM Transactions on Asian Language Information Processing (TALIP), Volume 5, Issue 4, Pages: 323 - 359, 2006, ISSN: 1530-0226, December 2006.
- [27] M. Braschler, "Combination approaches for multilingual text retrieval. Information Retrieval," Vol.7 (1-2):183-204, January, 2004.
- [28] W. Gao, C. Niu, M. Zhou and K.F. Wong, "Joint Ranking for Multilingual Web Search," In Proceedings of the 31<sup>st</sup> European Conference on Information Retrieval (ECIR), Toulouse, France, pp.114-125, 2009.
- [29] B. Herbert, G. Szarvas, and I. Gurevych, "Combining query translation techniques to improve cross-language information retrieval," In Proceedings of the 33<sup>rd</sup> European Conference on Advances in Information Retrieval, ECIR'11, Berlin, Heidelberg. Springer-Verlag, pages 712-715, 2011.
- [30] F. Ture, J.J. Lin, and D. W. Oard, "Combining Statistical Translation Techniques for Cross-Language Information Retrieval," In Proceedings of the 24<sup>th</sup> International Conference on Computational Linguistics, COLING '12, pp. 2685-2702, 2012.
- [31] R.W. Brislin, "Back-Translation for Cross-Cultural Research," Journal of Cross-Cultural Psychology 1, 1970.
- [32] D. He, J Wang, D.W Oard and M Nossal, "Comparing user-assisted and automatic query translation," Workshop of the Cross-Language Evaluation Forum for European Languages. Springer Berlin Heidelberg, 2002.
- [33] D. Grunwald, D and N.M. Goldfarb, "Back Translation for Quality Control of Informed Consent," Journal of Clinical Research Best Practices, 2 (2). Available at: [www.translationdirectory.com/article1043.htm](http://www.translationdirectory.com/article1043.htm).
- [34] U. Ozolins, "Issues of back translation methodology in medical translations," Proceedings, FIT [International Federation of Translators] XVIII Congress, Shanghai, 2008.
- [35] R. Rapp. "The back-translation score: automatic MT evaluation at the sentence level without reference translations." In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, ACL Short' 09, pages 133-136, 2009.
- [36] M. Miyabe and T. Yoshino, "Evaluation of the Validity of Back-Translation as a Method of Assessing the Accuracy of Machine Translation," International Conference on Culture and Computing (Culture Computing), Kyoto, pp. 145-150, 2015.
- [37] D. Zhou. M. Truran, T. Brailsford, V. Wade and H. Ashman, "Translation techniques in cross language information retrieval," ACM Comput. Surv.45, 1, Article 1, 44 pages, November, 2012.

- [38] Raju, BN V. Narasimha, MSVS Bhadri Raju, and K. V. V. Satyanarayana. "Translation approaches in Cross Language Information Retrieval." *Computer and Communications Technologies (ICCCT)*, 2014 International Conference on. IEEE, 2014.
- [39] Kumaran, A., Mitesh M. Khapra, and Pushpak Bhattacharyya. "Compositional machine transliteration." *ACM Transactions on Asian Language Information Processing (TALIP)* 9.4: 13, 2010
- [40] Saravanan, K., Raghavendra Udupa, and A. Kumaran. "Improving Cross-Language Information Retrieval by Transliteration Mining and Generation." *Multilingual Information Access in South Asian Languages*. Springer Berlin Heidelberg, 310-333, 2013.
- [41] B. Babych and A. Hartley, "Extending BLUE MT evaluation method with frequency weighting." *Proceedings of the 42<sup>nd</sup> annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2004.
- [42] ImTranslator Available at: <http://imtranslator.net/translation/hindi/to-english/translati on/>.
- [43] G. Salton, "Automatic processing of foreign language documents," *Journal of the American Society for Information Science* 21.3, 187-194, 1970.
- [44] C. Peters, "Cross-Language Information Retrieval and Evaluation", *Lecture Notes in Computer Science* 2069, Springer-Verlag, Germany, 2001.
- [45] S. Banerjee and A. Lavie, "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments" *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, Vol.29, 2005.
- [46] R. W. Brislin, "Back-translation for cross-cultural research," *Journal of Cross-Cultural Psychology* 1.3, 185-216, 1970.
- [47] E.A. Nida, "Towards a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating," Leiden, the Netherlands: Brill, 1964.
- [48] M.Aiken, "Multilingual communication in electronic meetings," *ACM SIGGROUP Bulletin* 23.1, 18-19, 2002.
- [49] S. Climent, J. More and A Oliver, "Bilingual newsgroups in Catalonia: A challenge for machine translation," *Journal of Computer Mediated Communication* 9(1), 2003.
- [50] M. Miyabe, T. Yoshino, and T. Shigenobu, "Effects of repair support agent for accurate multilingual communication," *Pacific Rim International Conference on Artificial Intelligence*. Springer Berlin Heidelberg, 2008.
- [51] Qu Peng, Li Lu and Zhang lili. "A Review of Advanced Topics in Information Retrieval," *Library and Information Service*, Vol.52.No 3.China.2008, pp.19-23.
- [52] K. Kishida, "Technical issues of cross-language information retrieval: a review," *Information Processing and Management international journal*, science direct, vol. 41, issue 3, pp. 433-455, May.2005.
- [53] W. C. Lin and H. H. Chen, "Merging mechanisms in multilingual information retrieval." *Workshop of the Cross-Language Evaluation Forum for European Languages*. Springer Berlin Heidelberg, 2002.
- [54] R.M.K Sinha, K. Sivaraman, and A. Agrawal, R. Jain, R Srivastava and A Jain, "ANGLABHARTI: a multilingual machine aided translation project on translation from English to Indian languages". Sinha, R. M. K., et al. "ANGLABHARTI: a multilingual machine aided translation project on translation from English to Indian languages." *Systems, Man and Cybernetics*, 1995. *Intelligent Systems for the 21st Century*, IEEE International Conference on. Vol. 2. IEEE, 1995.
- [55] Y. J. Zhang and T. Zhang. "Research on English-Chinese Cross-Language Information Retrieval." 2007 International Conference on Machine Learning and Cybernetics. Vol. 6. IEEE, 2007.
- [56] D.W. Oard and P. Hackett, "Document Translation for the Cross-Language Text Retrieval at the University of Maryland," in *Proceedings of TREC Conference*, 1997.
- [57] K. K. N. Kando, N. "Hybrid approach of query and document translation with pivot language for cross-language information retrieval," in *Proceedings of CLEF Conference*, 2005.
- [58] I. D. Melamed, R. Green and J. P. Turian, "Precision and recall of machine translation." *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers-Volume 2*. Association for Computational Linguistics, 2003.
- [59] A.L.F. Han, D.F. Wong, L.S. Chao, L. He, Y. Lu, J. Xing and X. Zeng, "Language-independent Model for Machine Translation Evaluation with Reinforced Factors", *Proceedings of the XIV International Conference of Machine Translation Summit*, Nice, France, pp. 215-222, International Association for Machine Translation Press September 2-6, 2013.
- [60] C.Y. Lin, "ROUGE: Recall-oriented understudy for gisting evaluation." 1-12, 2003. <http://berouge.com/>.
- [61] P. Virga. and S. Khudanpur. "Transliteration of proper names in cross-language applications." *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003.
- [62] K. Knight and J. Graehl, "Machine Transliteration". *Computational Linguistics*, 24(4), pp 599-612, 1998.
- [63] A. Fujii and T. Ishikawa. . "Japanese/English cross-language information retrieval: Exploration of query translation and transliteration." *Computers and the Humanities* 35.4, pp 389-420, 2001.
- [64] R. W. Brislin, W. J. Lonner, R.M. Thorndike, "Cross-Cultural Research Methods." New York: John Wiley & Sons, 1973.
- [65] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals." *Soviet physics doklady*. Vol. 10. pp. 707-710, 1966.
- [66] B. Chen and R. Kuhn. "Amber: A modified BLEU, Enhanced Ranking Metric." *Proceedings of the 6th Workshop on Statistical Machine Translation*. 2011.

### Authors' Profiles



**Ganesh Chandra** received his M. Tech. degree from KNIT Sultanpur, U.P. India and is pursuing his Ph.D. degree in Department of Computer Science at Babasaheb Bhimrao Ambedkar Central University, Lucknow, India. His major research interests include Information

Retrieval, WSD and Machine Translation. He is approachable at [ganesh.iiscgate@gmail.com](mailto:ganesh.iiscgate@gmail.com).





**Prof. Sanjay K. Dwivedi** is working as Professor & Head, Department of Computer Science at Babasaheb Bhimrao Ambedkar Central University, Lucknow, India. His research interest includes Artificial Intelligence, Information Retrieval, Web Mining, NLP and WSD.

He has published number of research papers in reputed journals and conferences. He is approachable at [skd200@yahoo.com](mailto:skd200@yahoo.com).

**How to cite this paper:** Ganesh Chandra, Sanjay K. Dwivedi, "Assessing Query Translation Quality Using Back Translation in Hindi-English CLIR", *International Journal of Intelligent Systems and Applications (IJISA)*, Vol.9, No.3, pp.51-59, 2017. DOI: 10.5815/ijisa.2017.03.07