Journal of Theoretical and Applied Information Technology
15th August 2019. Vol.97. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANALYSIS WITH VARYING QUESTION SIZE IN QUESTION PAPER TRANSLATION

SHWETA VIKRAM[1] and SANJAY K. DWIVEDI[2]

[1]Research Scholar, Department of Computer Science

[2]Professor, Head of the Department, Department of Computer Science

[1]shwetavikram.2009@rediffmail.com, [2]skd.200@yahoo.com

ABSTRACT

Word Sense Disambiguation in question paper translation is a challenging task. Some words in the question sentence can make the entire sentence ambiguous. WSD is a process to remove the ambiguity in a natural sentence to provide the correct sense of the word according to the sentence/context. Works have been done in question answering system to deal with ambiguity, however there has not been much work in resolving ambiguity related issues specially when it comes to translate questions rather than simple text. This paper specially highlights issues in the translation of Wh-questions from English to Hindi. We used five translators to show the impact of translation of Wh-questions using these translations. The experimental analysis of some English questions classified in three categories based on the number of words in each question. After analyzing these translations through MT tools for the three categories of questions, we found that the performance of translations of small questions is much better than that of other category questions having size medium to large. Further the average BLEU score (for all categories) has been found 0.483 for Babelfish which is best whereas Babylon performed poorly with 0.429.
Keywords: *Machine Translation, Word Sense Disambiguation, Questions, English and Hindi.*

1. INTRODUCTION

Translation of questions appearing in various competitive examinations from English to Hindi and other Indian languages are mostly being carried manually. It involves the timely availability of human experts in order to correctly translate questions to and from various Indian languages. The translations of question papers using an MT tool may highly help in such circumstances to cut time and energy. Though there are many good Indian languages MT tools available (both offline and online) such as Anusaaraka [33], Babelfish [32], Babylon [31], Bing [30], Google [34], they still perform fairly while translating many natural language sentences and the issues such as ambiguity, ordering, Tense-Aspect and Modality (TAM), gender, synonym aspect [17] often causes translation to become vague. Among these, ambiguity during translation is the most critical aspect. There have been many studies and successful implementations of WSD algorithms to minimize the issue, however, the high-level accuracy of translation still remains a challenging task in MT research. This issue becomes more serious when an MT tool is used to translate questions of various examinations because a slight change in the meaning after translation might change the expected answer to a question and that may result in wrong evaluation.

Apart from MT tools, researches have also been carried in question answering system which usually analyzes the patterns of the question for giving exact answers to the users. These systems used different patterns such as surface text patterns, regular expression and symmetric information [1, 14]. Question answering systems can also be multilingual wherein the question of one language may result in the answer in a different language as intended by researchers [5, 7]. However, these systems also face the issues as raised above and may suffer with poor or inter relevant answers. So, in this paper we have analyzed the issues of ambiguity in translation wh-question from English to Hindi. Questions have been categorized in three

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sets based on their size and each set have been analyzed separately [35-36].

2. RELATED WORK

Word Sense Disambiguation (WSD) has been the field of linguistic research and a large number of researches have been carried towards WSD for different natural languages in the context of Indian languages, there have been many works reported in the literature, for example in Indian a number of authors context have worked in Hindi WSD [17-21]. The author focused on three Indian languages such as Hindi, Marathi and Malayalam to handle morphological inflections problem from English to These Three Indian language translation using factored translation model. The author also observed this morphology injection improves the quality of translation in terms of adequacy and fluency [28-29].

Translation of question from one language to another may also result in ambiguity that may impact the interpretation of questions. In our earlier work [3], we tried to establish this fact by taking the set of questions in English and translating them into equivalent Hindi version using popular translators and also discussed various word sense ambiguities that may occur in different types of questions sentences. In our previous work [15], we discussed the impact of ambiguity on question paper translation by taking a set of examination questions in English using five translators. The question can be asked in different ways and can produce ambiguities differently. To study the impact of ambiguity work to different types of questions, classification of the question may be a great help. In a work [2], authors have provided question classification based on taxonomy, focus word, and question corpus for the purpose of question answering system. Wh-question (who, how) are one most common questions occurring in the examination. These questions are also the case for this research.

This paper also gives some rules for wh- question for extracting focus word and question class. In this author define wh-word (who, why, what, when, which, where and how) and also give some rules for Question answering system. In a work [26], the authors present a trained question answered pairs system with different type of questions. The new model of Q-A system makes the system trainable and gives the good result. This Q-A system uses POS tagger, Parser, lexical network and some supervised learning algorithms. Through a simple experiment, it was found that ambiguity affects the translation accuracy of the question.

Hao T. et al. [4] discuss the semantic pattern of the question for user interactive question answering system. The authors define five components of the question in the semantic pattern such as question target, question type, concept, event, and constraint. This paper also defines how semantic patterns help for answer extraction but it defines for English Question-Answering system. Authors Dave S. et.al [6] discusses the complexities that arise due to Hindi language structure and solve these complexities with the help of knowledge extraction with a case study. This paper also discusses simple, Interrogative, complex and compound sentences. Mishra A. and Jain S.K [22] discuss many different types of question in question answering system. The author classifies the question as the application domain, general domain, and restricted domain. Paper also defines word wh-question as factoid type, list-type, hypothetical-type, and causal and confirmation question. Many other works have also been done in question answering system such as Bouziane et. Al [23], Pechsiri, C.and R. Piriyakul [24], Zayaraz, Godandapani [25], Ramakrishnan et. [26] however, issues related to analyzing the questions translations has not been explored much in literature.

3. EXPERIMENTAL ANALYSIS

We took 110 Wh-questions in English from various authentic sources (such as NCERT-National Council of Educational Research and Training) and divided them into three different categories according to the length of words in each question. The first category has questions having length up to 6 words. The second category of questions has length between 7 to 12 words and all other remaining questions are placed in category third. Of the total questions, the first category has **21.81%** question, the second category has **50 %** and **28.18%** questions belong to the third category. This division of question sentences has been done according to source language (English).

3.1 Translation Tools

Five different types of machine translation tools are used for translating wh-questions from English to Hindi. The analysis of questions translation of different sizes will also help to understand as to how there popular tools behave when given

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

questions of varying sizes. BLEU score of MT tool output shows how these output translation matched from reference translation [12-13, 16, 27]. Following MT tools have been considered for our analysis.

1. *Anusaaraka:* Anusaaraka is free online machine translation tool for an English – Hindi language. It is based on the rule-based translation system. It gives layered output and source data should be in text form for input.
2. Babelfish: It is a free online Machine Translation tool to translate phrases in entire web pages, blogs, documents and sentences into 15 different languages.
3. Babylon: Babylon is unique tool and was developed using Optical Character Recognition (OCR) and it supports both texts, as well as the user, define the term to translate.
4. *Bing*: MT Bing is a free online translation tool which is developed by Microsoft. In this word limit that is maximum 5000 words at a time.
5. *Google*: It is an automatic machine translation service [1]. It is a multilingual machine translation facility, to translate text. It supports more than 100 languages at various levels

The reason behind choosing these translators is that the actual impact of ambiguity in wh-question could be better understood by using a number of translators, for example, if most of the translators translate questions accurately, there is no ambiguity in the questions despite the question might be ambiguous. Similarly, if one translator is able to correctly translate the questions and others fail to do so, it means the ambiguity issue affects the translation.

3.2 Performance Measurement

The widely used criteria of computing the BLEU score has been used it stand for Bilingual Evaluation Understudy (BLEU), shows the result of how the MT translated sentence varies from reference translation [11, 13]. BLEU is a matrix which is based on N-Gram precision. BLEU is designed to approximate human judgment at a corpus level and performs badly if used to evaluate the quality of individual sentences.

BLEU score does not focus on the ordering of word that means word matching is position independent, it is only focused on the correct meaning of the particular word. In this experiment, we used 1 gram precision. The computation of BLEU is done using the following formula.

$$\mathbf{BLEU=min}\left(1,\frac{\mathbf{Output-Length}}{\mathbf{Reference-Length}}\right)\left(\prod_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{precision_i}\right)^{\mathbf{1/n}}$$

For experimental analysis, BLEU score has been computed for each translation carried by different translators. Further, the more detailed analysis score has been divided into three parts that is, a score of "1" means translations are as per reference, the score between 0.5 and 1 indicating average translations accuracy, and score below 0.5 indicating poor translation accuracy.

3.3 Performance Evaluation

It can be seen from table 1 wherein questions have been divided into three categories I, II and III, the performance of translations of small questions (category I) is much better than that of category II and III in fact, all translators used have produced an absolute BLEU score of 1 to some questions. In category II and III, none of the questions have achieved absolute BLEU score of 1 (except for the Google translation which has one question for category II). It indicates that getting an absolute translation is difficult as we move from small to large size questions.

A large percentage of the questions have been translated into the category I which have BLEU score between 0.5 to 1, that shows many translations in this category are understandable, though not so accurately. Only a few questions have achieved poor translation accuracy (i.e. less than 0.5). In category II a large number of questions have been translated with BLEU score less than 0.5 this is again a clear indication of deteriorating performance of translation when the size of questions gets increased.

Journal of Theoretical and Applied Information Technology
15th August 2019. Vol.97. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table1: Question categorization according to BLEU score

Questions Category	I Category			II Category			III Category		
Number (Percentage) of questions	24 (21.81%)			55 (50%)			31 (28.18%)		
BLEU score	1	0.5 < 1	0.5 ≥0	1	0.5 < 1	0.5 ≥0	1	0.5 < 1	0.5 ≥0
Anusaaraka	4	17	3	0	20	35	0	6	25
Babelfish	4	14	6	0	30	25	0	9	22
Babylon	4	13	7	0	24	31	0	8	23
Bing	4	11	9	0	31	24	0	7	24
Google	5	10	9	1	19	35	0	6	25

Category III set of questions have been translated with least accuracy (i.e. lowest BLEU score). In fact, a majority of questions of this category have shown the BLEU score less than 0.5. The table also indicates that almost all translators we considered for the experiment have shown more or less similar performances for the three categories of wh-questions. Their performances gradually deteriorate as we move from category I to II that means the ambiguity and other related issues in larger wh questions dominates an affect the accuracy of the translation.

If we compare the translation accuracy of questions it is evident from the table that all tools have performed much better in translating category I question. Translation accuracy deteriorates constantly as we move towards category II and III. In fact, for the III category questions which are largest in size, all translators produced a very poor translation. As an example, we took one question for each category to show this trend.
Category I question,
Source Sentence: What is a mineral?
Reference Sentence: खनिज क्या हैं?

MT (Anusaaraka): खनिज क्या है?
MT (Babelfish): क्या एक खनिज है?
MT (Babylon): क्या है? एक खनिज
MT (Bing): क्या एक खनिज है?
MT (Google): एक खनिज क्या है?

In this example, all MT translated versions have the correct translation for category I question because word "mineral" is ambiguous. The correct meaning of the word "mineral" is "खनिज (KHANIJ)" and all MT has "खनिज (KHANIJ)" meaning of word "mineral" which is matched in context.
A majority of the questions of this category have translated correctly by most of the tools. However, only in a few cases for category I question the ambiguity issue has not been properly resolved by translators, for example

Source Sentence: What are body waves?
Reference Sentence: भूगर्भीय तरंगें क्या हैं?

MT (Anusaaraka): शरीरिक लहरें क्या हैं?
MT (Babelfish): क्या शरीर लहरों कर रहे हैं?
MT (Babylon): क्या हैं? लहरें शरीर
MT (Bing): क्या शरीर लहरों कर रहे हैं?
MT (Google): शरीर लहरों क्या हैं?

In this example, all MT translated versions have the incorrect translation for the ambiguous word "body". The correct meaning of the word "body" is "भूगर्भीय (BHUGARBHIY)" however, all MT tools have translated it as "शरीर (SHARIR)" meaning which does not match the context.

For category II, we found that majority of questions belong to the poorest range of score that is up to 0.5 only and considerable number of questions also lie between average score. Summering this, it can be said that question belonging to this category have average to poor accuracy.
As an example for this category,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Source Sentence: Where do they meet to form the Ganga?
Reference Sentence: ये कहाँ पर एक-दूसरे से मिलकर गंगा नदी का निर्माण करती हैं?

MT (Anusaaraka): वे गङ्गा बनाने के लिए कहाँ मिलते हैं?
MT (Babelfish): जहां वे गंगा फार्म को पूरा करते हैं?
MT (Babylon): वे कहां मिलेंगे रूप से गंगा?
MT (Bing): जहां वे गंगा फार्म को पूरा करते हैं?
MT (Google): वे कहाँ गंगा के लिए फार्म मिलना है?

None of the tools above have translated the question correctly. Likewise, other questions also suffer the same issue after translation, sometimes due to structure while it is the ambiguity for the other cases.

For category III, we found that majority of questions belong to the poorest range of score that is up to 0.5 however, only a few question lie between average score. Summering this, it can be said that majority of question belonging to this category have poor accuracy.
As an example for this category,

Source Sentence: What are the effects of propagation of earthquake waves on the rock mass through which they travel?
Reference Sentence: भूकंपीय गतिविधियों के अतिरिक्त भूगर्भ की जानकारी संबंधी अप्रत्यक्ष साक्ष्यों का संक्षेप में वर्णन करें।
MT (Anusaaraka): चट्टान परिमाण पर भूकम्प लहरों के प्रसारण के परिणाम हैं कौन सा में से वे क्या यात्रा करते हैं?
MT (Babelfish): जिसके माध्यम से वे यात्रा रॉक मास पर प्रचार की भूकंप तरंगों के प्रभाव क्या हैं?
MT (Babylon): क्या प्रभाव का प्रचार माध्यम से वे यात्रा पर आए भूकंप लहरों रॉक मास है?
MT (Bing): जिसके माध्यम से वे यात्रा रॉक मास पर प्रचार की भूकंप तरंगों के प्रभाव क्या हैं?
MT (Google): रॉक मास जिसके माध्यम से वे यात्रा पर आए भूकंप तरंगों के प्रसार के प्रभाव क्या हैं?

None of the tools above have translated the question correctly. Likewise, other questions also suffer the same issue after translation, sometimes due to structure while it is the ambiguity for the other cases.

Long questions have many ambiguous words so the MT translated versions have the ambiguous sense of the question. As we know that the size of the question increased as well as the accuracy of translation decreased.

Table 2, indicates the average value for subcategories wise and also contain the total average value for category wise. MT Bing has maximum average BLEU score in subcategory II under category I and MT Anusaaraka have minimum average BLEU score.

In the second subcategory, MT Bing have maximum and MT Anusaaraka has minimum average BLEU score. In category I, MT Babelfish and MT Babylon have maximum and minimum BLEU score.

For category II only MT Google has 1 average BLEU score in the first subcategory. In subcategory II Google have maximum and MT Babylon have the minimum average score, subcategory III MT Google have maximum and MT Babelfish have minimum BLEU score. For category II, Bing has a maximum average and MT Anusaaraka has minimum average BLEU score.

The subcategory of category III has 0 averages BLEU score for all MT tools. In this second subcategory, MT Babylon and MT Anusaaraka have maximum and minimum average BLEU score. For subcategory II, MT Babelfish and MT Anusaaraka have maximum and minimum average BLEU score. For category III MT Bing and MT Babelfish have maximum and minimum average BLEU score. For small size questions performance of MT Babelfish is best in our experiment, for the medium in size questions which come under category II MT Bing have better performance and at last at last category III for long wh questions performance of MT Bing again better than all taken MT tools.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Average BLEU Score For Different Question Categorization

Questions Category	I Category				II Category				III Category			
BLEU score	1	0.5 ≤1	0.5 >0	Total avg	1	0.5 ≤1	0.5 > 0	Total avg	1	0.5 ≤1	0.5 >0	Total Avg
Anusaaraka	1	0.628	0.193	0.635	0	0.629	0.299	0.419	0	0.533	0.248	0.303
Babelfish	1	0.665	0.329	0.637	0	0.623	0.275	0.467	0	0.569	0.309	0.234
Babylon	1	0.655	0.200	0.58	0	0.619	0.276	0.437	0	0.616	0.250	0.345
Bing	1	0.707	0.338	0.618	0	0.625	0.295	0.48	0	0.606	0.296	0.366
Google	1	0.672	0.289	0.597	1	0.644	0.325	0.429	0	0.549	0.307	0.354

In table 2 the average BLEU score of different category questions is shown. Few questions have the average score of 1 as all these questions have the individual score of 1. In this category, most of the questions fall in the score range of 0.5 to 1 and the average score for different translations are above 0.62. All through the highest number of the question had scored between 0.5 to 1 by Anusaaraka (table 1) but the average score for this subcategory shows that *Bing* is strictly better among others. The overall average of the scores of three subcategories of category I suggest that all the questions have a satisfactory score of around 0.6 and on comparing results table 1 and 2 we understand that in order to find the accuracy of translation through Blue matrix the average scores as computed in table 2 is also important.

In category II we see a sharp drop in the average score by different MT compared to category I questions.

The total average of score further drops for category III questions for all translators. It is also significant to known that the performance of one MT tool for all category of questions is not same for example the Babelfish has been good for category I and II questions whereas its perform fairly for category III questions.

Table 3, indicates the average value for all five different MT tools. For all wh questions translation, Babelfish (0.483) is best and Anusaaraka (0.433) is poor performance in our performance evaluation.

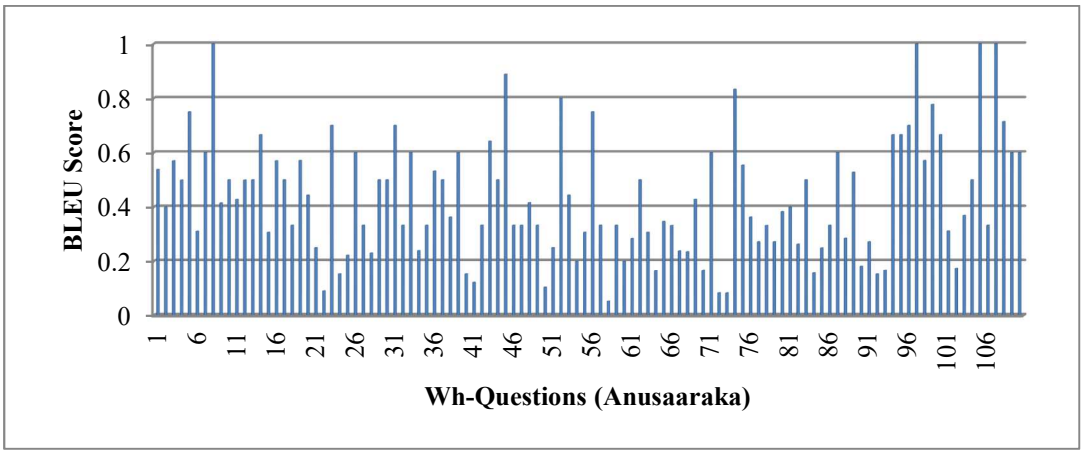
Table 3: Average BLEU Score For Machine Translation.

MT	Anusaaraka	Babelfish	Babylon	Bing	Google
Avg BLEU score	0.433	0.483	0.429	0.478	0.458

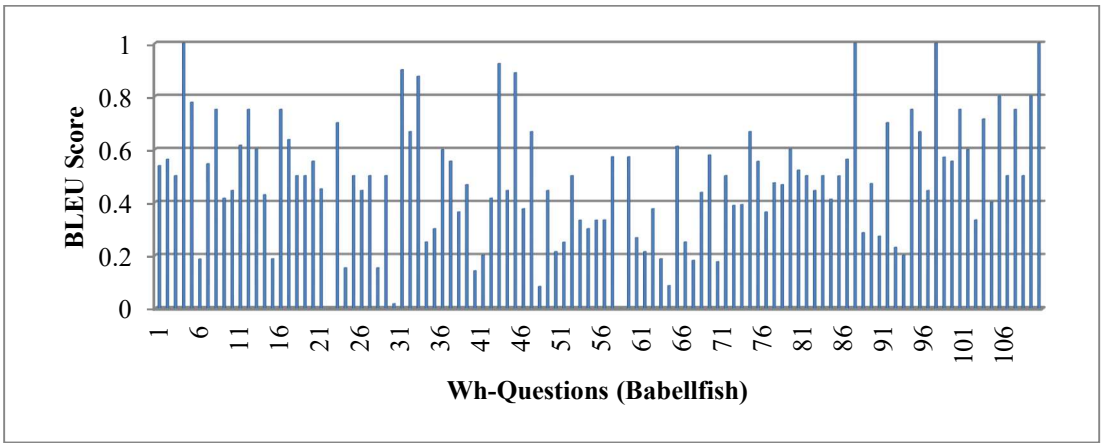
4. RESULTS

Here we will discuss the result of our experimental analysis according to Bleu score for taken all MT tools. Wh question categorized into three categories according to question length and this category again subcategories with respect to Bleu score. In figure 2 all graphs have four questions which BLEU score is "1" score except MT Google and MT Babelfish have the best result with 0.483 average score, Babylon has the poor result with 0.429 average BLEU score and all remaining MT are lies between Babelfish and Babylon.

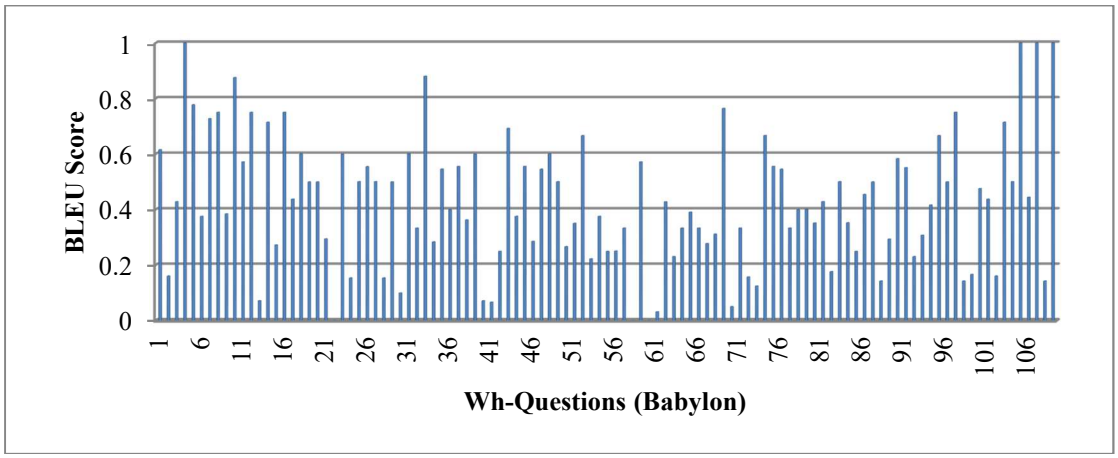
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



Graph: 1(A)

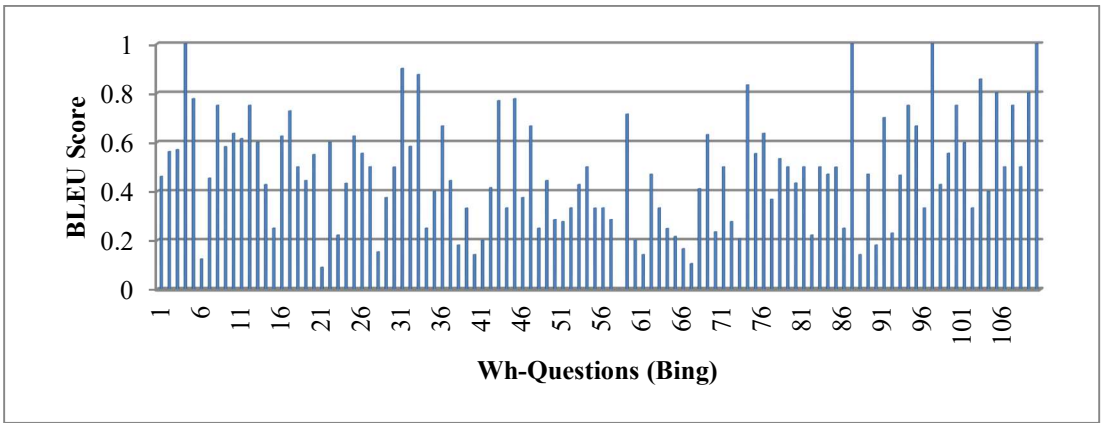


Graph: 1(B)

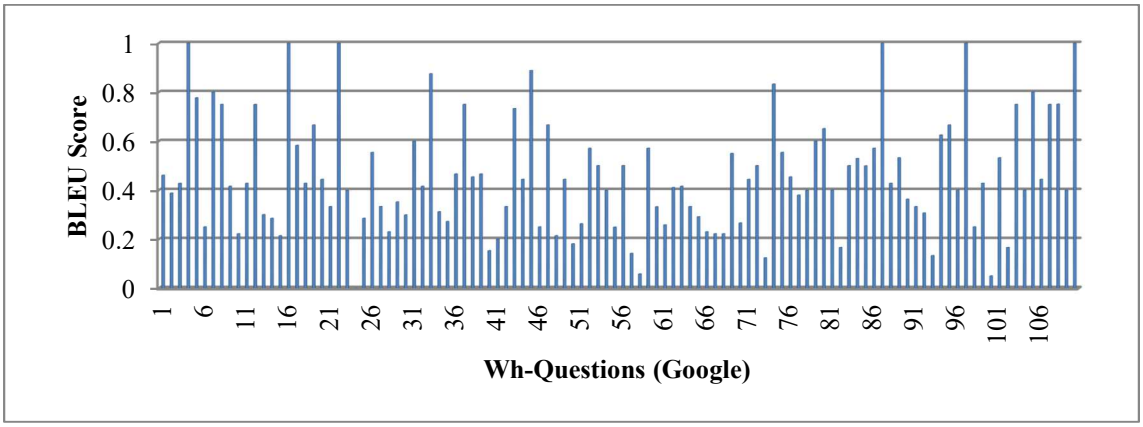


Graph: 1 (C)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



Graph: 1 (D)



Graph: 1 (E)

Graph: 5.1 (A - E): Shows The BLEU Score For All Five Different MT Tools With 1-Gram Precision.

5. DISCUSSION

The comprehensive experiment carried in this paper reveals that only one MT tool Anusaaraka which does not have a single wh question with BLEU score as 0, otherwise all remaining four MT have one or more than one questions having 0 BLEU score.

In fact MT Anusaaraka is the only one translation tool in our experiment which does not have 0 BLEU score in all three categories. All MT have 1 BLEU score for the small question which comes under the category I, only MT Google have 1 BLEU score for one question in category II (medium size questions).

Table 1 shows question categorization according to the size of the question and BLEU score also has been subcategorized in three ranges. Table 2 shows average BLEU score for all popular five MT tool according to sub categorization. Table 3 shows the overall average BLEU score for all five MT tools.

The size of question (under Wh Type) has the major impact on its translation accuracy. Among the large questions (category III), all translations generally failed to give high score. This shows that, despite the type of questions, the size also has a big impact in the translation accuracy

For all MT tools, sometimes the inferred meaning is somewhat different. MT has given word to word (literal) translation of the same question. Clearly, these translations also are often not appropriate. MT Anusaaraka is given slightly better the translation of the small questions. Through its translation is not exactly the same as reference translation, one can easily understand the meaning as to what is actually being asked in the question.

However, for category II and III questions it performs poorly. Babelfish has best overall score and Bing gives the lowest score among all the MT

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

tools. MT (Google) lies in between all MT. The order of all MT tool result shown as ascending order is Babelfish, Bing, Google, Anusaaraka and Babylon.

The results indicate that the average accuracy of translation for all categories of question is less than 50%.
The poorest among these is Babylon. While analyzing these questions through various translators we considered all those translation as correct wherein the overall meaning and its interpretation are similar to the reference translation.

6. CONCLUSION

In this paper mainly we focus on the analysis according to the size of questions and find some statistical result. In these five MT, Babelfish stands better in the average. Smaller questions have performed better in terms of accuracy whereas long questions have shown poor accuracy. The average accuracy of all translators for all categories combined together is found to be below 50% which indicates that the tools cannot be relied upon. That means MT Babelfish gives 48.2 % accurate result vis-à-vis same as reference translation. So improvement is the need for wh-question translation specially when the size of questions increases.

REFERENCES:

[1] Li, Haiying, Arthur C. Graesser, and Zhiqiang Cai. "Comparison of Google translation with human translation." *The Twenty-Seventh International Flairs Conference,* 2014.

[2] Dwivedi, S. K., & Singh, V., "Integrated question classification based on rules and pattern matching", In *Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies, ACM, 2014,* pp. 39.

[3] Dwivedi, S. K., & Vikram, S., "Word Sense Ambiguity in Question Sentence Translation: A Review", In *International Conference on Information and Communication Technology for Intelligent Systems, 2017,* pp. 64-71.

[4] Hao, T., & Wenyin, L., "Automatic question translation based on semantic pattern", In *Semantics, Knowledge and Grid, SKG'08. Fourth International Conference on* IEEE, 2008, pp. 372-375.

[5] Hao, T., Hu, D., Wenyin, L., & Zeng, Q., "Semantic patterns for user-interactive question answering", *Concurrency and Computation: Practice and Experience,* 2008, pp. 783-799.

[6] Dave, S., & Bhattacharyya, P., "Knowledge extraction from Hindi text", *IETE Technical Review*, 2001, pp. 323-331.

[7] Gupta, Vishal. "A Survey of Word-sense Disambiguation Effective Techniques and Methods for Indian Languages", *Journal of Emerging Technologies in Web Intelligence* 5.4 2013.

[8] Madankar, M., Chandak, M. B., & Chavhan, N., "Information retrieval system and machine translation: a review", *Procedia Computer Science*, 2016, pp. 845-850.

[9] Chatterjee, N., Goyal, S., & Naithani, A., "Resolving pattern ambiguity for english to hindi machine translation using WordNet", In *Workshop on Modern Approaches in Translation Technologies*, 2005.

[10] Kunchukuttan, A., Mishra, A., Chatterjee, R., Shah, R., & Bhattacharyya, "Sata-anuvadak: Tackling multiway translation of indian languages". *pan*, 2014, pp. 4-135.

[11] Chaudhury, S., Rao, A., & Sharma, D. M., "Anusaaraka: An expert system based machine translation system", In *Natural Language Processing and Knowledge Engineering (NLP-KE), International Conference on* IEEE, 2010, pp. 1-6.

[12] Gautam, S., & Bhattacharyya, P., "Layered: Metric for machine translation evaluation" In *Proceedings of the Ninth Workshop on Statistical Machine Translation, 2014,* pp. 387-393.

[13] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J., "BLEU: a method for automatic evaluation of machine translation" In *Proceedings of the 40th annual meeting on association for computational linguistics* Association for Computational Linguistics, 2002, pp. 311-318.

[14] Dreyer, M., & Marcu, D., "Hyter: Meaning-equivalent semantics for translation evaluation", In *Proceedings of the Conference of the North American Paper of the Association for Computational Linguistics: Human Language Technologies* Association for Computational Linguistics, 2012, pp. 162-171.

[15] Vikram, S., & Dwivedi, S. K., "Ambiguity in Question Paper Translation", *International Journal of Modern Education and Computer Science*, 2018, pp. 13.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[16] Koehn, P., "*Statistical machine translation*", Cambridge University Press, 2009.

[17] Navigli, R., "Word sense disambiguation: A survey". *ACM computing surveys (CSUR)*, *2009*, pp. 10.

[18] Carpuat, M., & Wu, D., "Improving statistical machine translation using word sense disambiguation" In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL),* 2007.

[19] Carpuat, M., & Wu, D., "Word sense disambiguation vs. statistical machine translation", In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* Association for Computational Linguistics, 2005, pp. 387-394.

[20] Ramakrishnan, G., Prithviraj, B. P., Deepa, A., Bhattacharyya, P., & Chakrabarti, S., "Soft word sense disambiguation", In *Proceedings of GWC* , 2004.

[21] Singh, S., Siddiqui, T. J., & Sharma, S. K., "Naïve Bayes classifier for Hindi word sense disambiguation", In *Proceedings of the 7th ACM India Computing Conference* ACM, 2014, pp. 1.

[22] Mishra, A., & Jain, S. K., "A survey on question answering systems with classification", *Journal of King Saud University-Computer and Information Sciences*, *2016, pp* 345-361.

[23] Bouziane, A., Bouchiha, D., Doumi, N., & Malki, M., "Question answering systems: survey and trends", *Procedia Computer Science*, 2015, pp. 366-375.

[24] Pechsiri, C., & Piriyakul, R., "Developing a Why–How Question Answering system on community web boards with a causality graph including procedural knowledge", *Information Processing in Agriculture*, *2016,* pp. 36-53.

[25] Zayaraz, G., "Concept relation extraction using Naïve Bayes classifier for ontology-based question answering systems", *Journal of King Saud University-Computer and Information Sciences*, *2015,* pp. 13-24.

[26] Ramakrishnan, G., Chakrabarti, S., Paranjpe, D., & Bhattacharya, P., "Is question answering an acquired skill?", In *Proceedings of the 13th international conference on World Wide Web ACM, 2004,* pp. 111-120.

[27] Ananthakrishnan, R., Bhattacharyya, P., Sasikumar, M., & Shah, R. M., "Some issues in automatic evaluation of english-hindi mt: more blues for bleu", *ICON,* 2007.

[28] Bhattacharyya, P., "Role of Morphology Injection in SMT: A Case Study from Indian Language Perspective", *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, *2017, pp.* 1.

[29] Gupta, D., Kumari, S., Ekbal, A., & Bhattacharyya, P., "MMQA: A Multi-domain Multi-lingual Question-Answering Framework for English and Hindi", In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC),* 2018.

[30] MT Bing: https://www.bing.com/translator/help/#About MicrosoftTranslator

[31] MT Babylon: http://translation.babylon-software.com/english/to-french/

[32] MT Babelfish: https://www.babelfish.com/success/

[33] MT Anusaaraka: https://anusaaraka.iiit.ac.in/drupal/node/2

[34] MT Google: https://translate.google.com/

[35] NCERT: http://epathshala.nic.in/e-pathshala-4/flipbook/

[36] NCERT: http://ncert.nic.in/NCERTS/textbook/textbook.htm