

Analysis on Translation Quality of English to Hindi Online Translation Systems- A Review

Ekta Gupta

M.Tech, Scholar

Samrat Ashok Technology Institute Vidisha (M.P.)
India

Shailendra Shrivastava, PhD

Head of Department

Samrat Ashok Technology Institute Vidisha (M.P.)
India

ABSTRACT

In developing countries like Asian country and India where English is mainly half-hour recognize there automatic computational linguistics systems in education, analysis and industrial activities of very necessary role. Asian country has state a large assembly in Hindi is the language you speak and in a number of areas it works in all kinds of study and official. Nowadays several on-line translator technologies use completely different computational linguistics approach. Like every translation approaches completely different characteristics, the outcome of the explanation would differ. Bing Translator and Google Translate free on-line machine translators is exploitation statistical computational linguistics each translators are of the most accepted. It keeps increasing the language selection and increasing its usability. As a result of the most important reverse characteristics of every machine translator services and their needed role for the development of machine translators mainly in web platform, it's determined to own a study concerning their comparison. Bing translator and Google translate within the whole world have to be obliged to use automatic on-line translation system is employed broadly in India is additionally as a outcome of it's a free and reliable. The aspire of this study is to make understanding concerning the different performance of the 2 on-line translation services due to the same actions they need. The experimentation designed is meant show however the 2 on-line translation services possess advantages and disadvantages which may have a consequence on their performance. Secondary aim of this study is to look for out the typical problems that will occur in translation between English and Hindi and to look for out from the 2 on-line translation services, which is more proper. To get a completely automatic top quality machine translation system is a troublesome task. In this document, we have a tendency to explain the web translation systems like Microsoft Bing Translator and Google Translate for English to Hindi translation and its conversion quality. Researches focused on survey of on-line translation resolutions for English to Hindi translation and examine its translation quality.

The evaluation parameters that are considered here are Accuracy, translation speed, output fluency, Adequacy,

Keywords

Translation Quality Analysis, Translation Quality of Online Translation System for English to Hindi Translation.

1. INTRODUCTION

Machine translation is the name for processed strategies that automatize all or a part of the method of translating from one language to a different. During a massive multilingual society like India, there's great demand for translation of documents from one language to a different language. Hindi written within the Devanagri script is the official language of the union Government. English is also use for government

notifications and communications. India's average skill level is seventy percent. But five % of individuals will either scan or write English. As most of the higher study material, research journals and different commonplace communication tools are in English; these materials are to be translated into Hindi or particular local languages to Possess an applicable higher study and communication with the individuals. Moreover, over 95 percent of the population is generally deprived of the advantages of Information Technology because of barrier. All these make language translation a necessary one. Work in the area of computational linguistics in world has been going on for many decades. Throughout the early 50s, advanced research within the field of artificial intelligence and Computational Linguistics created a promising development of translation technology. This helped within the development of usable computational linguistics Systems in certain well-defined domains. Totally prime quality automatic computational linguistics system to induce could be a difficult task several organizations like Google, Microsoft, IBM, and lots of different etc. are engaged in development of MT systems.

In this A Survey of Translation Quality of English to Hindi Online Translation Systems paper [1], presents comparison between two online machine translation systems on the basis of their structure and architecture which leads to difference in semantics. In this Mining Hindi-English Transliteration pair from Online Hindi Lyrics paper [2], presents a procedure to mined Hindi-English transliteration pairs from online Hindi song lyrics. The technique is based on the clarification that lyrics are transliterated word-by-word, maintaining the precise word order. In this Translation Rules for English to Hindi Machine Translation System: Homoeopathy Domain paper [3], presents the grammar rules proposed for our English to Hindi machine translation system to interpret the homoeopathic literatures, prescription, medical reports etc. In this Evaluation of Hindi-English MT Systems paper [4], presents checked Health and General cooking data and evaluated the English output text. Human assessment policy has been used for the purpose of evaluation, on the basis of which problem areas in both the MT systems were recognized and measure up to reach a conclusive analysis in terms of the output's fluency and comprehensibility. In this Automatic Translation of English Nominal Compound in Hindi paper [5] presents an automatic translation arrangement for converting English bigram nominal compound into Hindi. In this English from Hindi viewpoint: A Paninian Perspective paper [6] proposed whole purpose of this exercise is to look at the structural differences between English and Hindi from an information theoretic point of view. In this An English to Hindi Machine-Aided Translation System paper [7] proposed a system overview of an English to Hindi Machine-Aided Translation System named AnglaHindi. Its beta-version has been made available on the internet for free translation. In this Hindi to English Machine Translation: Using Effective

Selection in Multi-Model SMT paper [8] proposed an approach to estimate the quality of machine translation and dynamically select the better translation at run-time. Combining the text analysis and linguistic features. In this Developing Lexicon Databases for English to Sinhala Machine Translation paper [9] proposed the design and implementation of the lexical database for English to Sinhala Computer Aided Translation. In this Machine Translation Systems for Indian Languages paper [10] the various approaches which have been applied in translation systems for Indian languages. Some of the important Indian language translation systems implemented with these techniques along with their capabilities and limitations are also discussed.

2. LITERATURE SURVEY

A Survey of Translation Quality of English to Hindi Online Translation Systems (Google and Bing) [1], “India has declared a large assembly in Hindi is the language they speak and in several areas it works in all type of official and study. Many online translator technologies today use different machine translation approach like each translation approaches diverse characteristics; the outcome of the translation would be different. Bing Translator and Google Translate free online machine translators is using statistical machine translation. Both translators are of the most accepted. It keeps growing the language option and increasing its usability. The procedure used by us is as follows

1. Data collection.
2. Translation data into English to Hindi.
3. Manual analysis of collected data.
4. Compare evaluation results.

Due to the major reverse characteristics of both the translator services and their important role for the growth of machine translators particularly in internet platform, it is determined to have a study about their similarity. The idea of this study is to create understanding about the special performance of the two online translation services due to the same procedures they include. The testing designed is intended to show how the two online translation services have its own benefit and disadvantage which can influence their performance. From this investigation it can be learned that the characters of two online machine translator application, Bing machine translators and Google machine translator. They have difference in basic features, where it leads to differences in its structure architecture. Bing and Google machine translator exercise statistical information of earlier translation to learn about the language which is to be translated. Statistical machine translators are separated into three kinds, word-based, syntax based and phrase-based. They chose Bing Translate and Google Translator represent English to Hindi translation as it is the most known online machine translators.

Mining Hindi-English Transliteration pair from Online Hindi Lyrics, “Approximate string-matching algorithm” [2] explains a procedure to mined Hindi-English transliteration pairs from online Hindi song lyrics. The technique is based on the clarification that lyrics are transliterated word-by-word, maintaining the precise word order. The mining task is nevertheless difficult because the Hindi lyrics and its transliterations are usually available from different, often unrelated, websites. Thus, it is a non-trivial assignment to match the Hindi lyrics to their transliterated counterparts. Moreover, there are different types of noise in lyrics data that needs to be appropriately handled before songs can be aligned at word level. The supply data of 30823 unique Hindi-English

transliteration pair with a precision of more than 92% is accessible publicly. Even though the present work information mining of Hindi-English word pairs, the same procedure can be easily adapted for supplementary languages for which song lyrics are accessible online in local and Roman scripts. In this paper explain a technique for extracting transliteration correspondent using online song lyrics. The mined data is of high quality and comparatively noise-free. This data can be use for training English-Hindi backward transliteration engine for building IME for Hindi and for other general applications of transliteration engines in IR and MT. The mined pairs will as well be useful for linguistic learning on spelling difference and typing patterns in transliterated texts. This data is accessible as supplementary material along with this document and will be circulated freely for research. Although the current work only explains mining of English-Hindi transliteration pairs, the process described here is generic and appropriate for mining similar data for every language that has song lyrics or other similar contents presented online in both local and Roman scripts.

Translation Rules for English to Hindi Machine Translation System: Homoeopathy Domain [3], “Stemming Rules and PoS Tagging Rules” regulation based machine translation system hold a set of grammar rules which are compulsory for the mapping of syntactic demonstration of a source language, on the target language. The scheme necessitates good linguistic awareness to write rules and need of acquaintance source such as corpus and bilingual dictionary. In this paper described the grammar rules proposed for our English to Hindi machine translation system to interpret the homoeopathic literatures, prescription, medical reports etc. The rules which have been printed track the transfer based approach for reordering of rules between the two languages. The document first talk about our developed stemmer and its set of rules, in advance we discuss the Part of Speech tagging rules for classify each word of the sentence grammatically and our developed homoeopathy corpus in Hindi and English of size 20085 and 20072 words correspondingly and at the end in this paper discuss the translation / agreement rules for translating various homoeopathic sentences. The present effort, attempts have been complete to deal with the translation rules of homoeopathic language with respect to their sentence structure. By going all the way through the literature related to MT, we experience that MT systems frequently fail to convey good accuracy for detailed domains. This is factual even for a lot of open domain systems including Google translator. The cause is the lack of relevant corpus and domain definite mapping rules. So that, Homoeopathy corpuses for English and Hindi of sizes 20085 and 20072 words respectively have been developed. We utilize 500 sentences of training data for calculating the performance of gathering rules on the basis of sentence categories like simple sentence, compound and Interrogative sentences. The advantage of this system is good accuracy.

Evaluation of Hindi-English MT Systems [4], “evaluating the output of Hindi-English language pair through two MT systems Bing and Google.” Evaluation of any Machine Translation (MT) system is a significant step towards improving its precision can be done by two ways (a) Automatic Evaluation, (b) manual or Human evaluation . In this paper, we are trying to evaluate Hindi-English module through two most commonly used MT systems - Google and Bing (Microsoft). These MT systems are Statistics-Based MT systems and are capable of providing translation in several languages across the globe other than Hindi-English. For the reason of assessment, we checked Health and General

cooking data and evaluated the English output text. Human assessment policy has been used for the purpose of evaluation, on the basis of which problem areas in both the MT systems were recognized and measure up to reach a conclusive analysis in terms of the output's fluency and comprehensibility. The relative study helps in understanding not only which system is better but also what works best for automatic translation and under what conditions. The discrepancies found are discussed with some suggestions towards their solution. In this document, two major points have been presented: The first is the errors in MT output and second is the evaluation of Bing MT & Google MT systems. When we observed and evaluated these systems, we found many errors. We have previously talked about above the discrepancies in Tokenization, Morph Error, Structural Errors and Parser issues etc. And when, we estimated MT systems, the fluency was set up to be very low but nonetheless it was almost comprehensible. On judgment, Google was found to be superior to Bing MT in comprehensibility. When we compute the ratio of results attained in Table 4 then Bing is found to be improved than Google. If the problems mentioned for Bing MT system are resolved then it can outperform Google in future. The advantage of the increase fluency as well as in comprehensibility of the Bing and Google MT systems.

Automatic Translation of English Nominal Compound in Hindi [5], "context based translation system" English nominal compounds is able to be variously translated into Hindi. This paper presents an automatic translation arrangement for converting English bigram nominal compound into Hindi.

The process comprises of the following steps:

1. Translation template generation,
2. Removal of nominal compound from English corpus (3) discovery the appropriate sense of the components of the compound by using WSD tool
3. Lexical substitution of component nouns using Bi-Lingual Dictionary
4. Corpus Search by using translation templates and Ranking of possible candidates.

We have made known that the correct sense choice of the component nouns of a known nominal compound throughout the analysis stage considerably enhanced the performance of the system and makes the present effort distinct from all the earlier works made for automatic bilingual translation of Nominal compounds. This document explains the structural design of a template based translation system for converting English nominal compound into Hindi. We have examined that English nominal compounds can variously be translated into Hindi. Although no clue is available to determine which type of Hindi construct a given English nominal compound would be converted into. We have, therefore, adopted a corpus search approach that performs the search of candidate templates in a Hindi indexed corpus. at the same time as generating templates, we found out that adjectival templates are hard to generate for the reason that adjective formation from noun is a complex derivational process in Hindi. It does not simply involve adding an adjectival suffix on the noun however also many time needed a change in the vowel of the stem. In the present work, we have achieved poorly for adjective noun translation templates.

English from Hindi viewpoint: A Paninian Perspective [6], "the Paaninian way of analysis to discover the structural differences" proposed whole purpose of this exercise is to

look at the structural differences between English and Hindi from an information theoretic point of view. The major reason behind the structural differences between English and Hindi is the absence of accusative marker analyze marker in English. To compensate for this absence, English resorts to the word order. This further gives rise to more structural differences between the two languages. The advantage of this system is reduces the language barrier.

English to Hindi Machine-Aided Translation System [7], "Anglabharti is a pseudo-interlingual rule-based translation methodology" is an overview of English to Hindi Machine-Aided Translation System named AnglaHindi. Its beta-version has been made available on the internet for free translation at AnglaHindi is English to Hindi version of the ANGLABHARTI translation methodology developed by the author for translation from English to all Indian languages. Anglabharti is a pseudo-interlingual rule-based translation methodology. AnglaHindi, besides using the rule-bases, uses example-base and statistics to obtain more acceptable and accurate translation for frequently encountered noun and verb phrasals. This way a limited hybridization of rulebased and example-based approaches has been incorporated.

Hindi to English Machine Translation: Using Effective Selection in Multi-Model SMT [8] "aforementioned prediction methodology" Combining the text analysis and linguistic features, results in a system which shows improvement over the baseline system and shows high agreement with human judgment. Sequentially running both the phrase and hierarchical system may result in increase in time of computation as parse tree and other feature computation add to decoding time? To overcome this issue we have employed distributed computing so as to compute all features in parallel for phrase and hierarchical systems. The advantage of this model is better translation at run time. Computation time is increased in parse tree and other feature computation add to decoding time is disadvantage of that model.

Developing Lexicon Databases for English to Sinhala Machine Translation [9] "development of lexicon database sub-system for English to Sinhala Computer-Assisted Translation system" the design and implementation of the lexical database for English to Sinhala Computer Aided Translation. The six dictionaries, namely, English word dictionary, English concepts dictionary, English-Sinhala bilingual dictionary, Sinhala word dictionary, Sinhala rule dictionary and Sinhala concept dictionary, have been designed and developed. At present all these dictionaries contain a limited number of words with their lexicon information. The limited number of words is the disadvantage of that system.

Machine Translation Systems for Indian Languages [10] "Rule Based Translation" Various MT groups have used different formalisms best suited to their applications. Of them transfer based systems are more flexible and it can be extended to language pairs in a multilingual environment. Direct translation is appropriate for structurally similar languages. The interlingua based systems can be used for multilingual translation. The amount of analysis needed in interlingual approach is more than that in a transfer based approach. The universal networking language has been proposed as the interlingua by the United Nations University for overcoming the language barrier. Over the past decade data-driven approaches to machine translation have come to the fore of language processing research. The advantage of this system is more flexible. The disadvantage of this system is language barrier.

3. METHOD

3.1 Translation Quality Analysis

Translation quality assurance is that the primary concern of any customer of translation services. this can be absolutely comprehensible considering that, unless you speak the language you're translating into, you have got no method of assessing the standard of the ultimate translation. Qualities are some things that much each Translation Service supplier claims to supply. though it's typically argued that translation quality is subjective, it's however potential to determine objective quality criteria for the interpretation itself, the work method and also the overall service. the essential criteria that a Translation Service supplier providing a high quality service ought to fulfill.

Translation is correct transfer of data from the supply text to the target text. acceptable selection of language, vocabulary, idiom and register within the target language. Acceptable use of synchronic linguistics, spelling, punctuation and syntax, further as correct transfer of dates, names, figures, etc. within the target language. Applicable vogue for the aim of the text. Translation is each a psychological feature procedure that happens during a human being's, the translator's, head, and a social, communication and society observe. Any valid theory of translation should embrace these 2 aspects. To do this, a multi disciplinary approach to translation theory integration these aspects during a plausible manner are required. Further, a theory of translation isn't potential while not a mirrored image on the role of 1 of its core concepts: equivalence in translation. At equivalence leads directly into a discussion of however one would approach assessing the standard of a translation. Translation quality assessment will therefore be same to be at the center of any theory of translation. Translation may be outlined because the results of a linguistic-textual operation during which a text in one language is re-contextualized in another language. As a linguistic matter operation, translation is, however, subject to, and considerably influenced by, a spread of extra-linguistic factors and conditions.

3.2 Machine Translation

Machine Translation (MT) primarily deals with the transformation from one language to a different. Language Interface provides the user freedom to move with the pc during a language like English, Malayalam, Telugu, and Hindi or the other language used for day to day communication. one amongst the necessary goals of linguistics could be a absolutely automatic AI between such natural languages. this can be necessary as a result of communication between individuals from completely different linguistic backgrounds still poses as a serious drawback. The most objective of MT is to interrupt the barrier during a multilingual nation like Republic of India. Analysis of AI (MT) like MT itself has proved to be a really tough task since the instigation. the issue arises primarily from the rationale that almost all sentences may be translated in several acceptable ways in which.

Many MT systems across the world have already been developed for the foremost normally used natural languages like English, Russian, Japanese, Chinese, Spanish, Hindi and different Indian languages etc. Figure 1 depicts the present AI systems and varied approaches utilized in developing these systems.

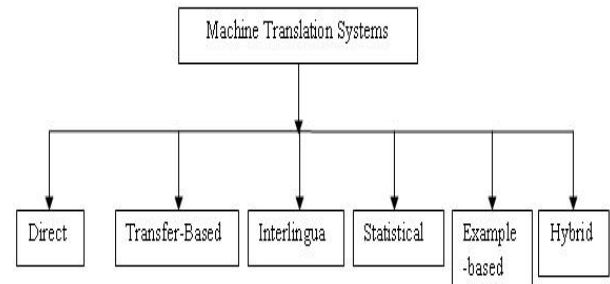


Fig. 1. Machine Translation Systems

4. CONCLUSION

In this paper given Translator represent English to Hindi translation because it is that the most known on-line machine translators. The aim of this study is to make understanding concerning the various performances of the 2 on-line translation services because of identical procedures. a completely automatic top quality AI system to induce may be a troublesome task. This paper represented MT techniques in a very longitudinal and angular distance method with a stress on the MT development for Indian languages also as non-Indian languages. During this paper, describe the web translation systems for English to Hindi translation and its translation quality.

5. REFERENCES

- [1] Bhojraj singh dhakar, sitiesh kumar sinha, krishna kumar pandey "A Survey of Translation Quality of English to Hindi Online Translation Systems (Google and Bing)", International Journal of Scientific and Research Publications, Volume 3, Issue 1, January 2013
- [2] Kanika Gupta, Monojit Choudhury, Kalika Bali "Mining Hindi-English Transliteration pair from Online HindiLyrics"<http://www.google.com/intl/en/press/zeitgeist2010/regions/in.html>
- [3] Sanjay Dwivedi and Pramod Sukhadeve "Translation Rules for English to Hindi Machine Translation System: Homoeopathy Domain", The International Arab Journal of Information Technology, Vol. 12, No. 6A, 2015
- [4] Atul Kr. Ojha, Akanksha Bansal, Sumedh Hadke and Girish Nath Jha "Evaluation of Hindi-English MT Systems", <http://translate.google.co.in/>
- [5] Prashant Mathur "Automatic Translation of Noun Compounds from English to Hindi", Language Technology Research Center International Institute of Information Technology Hyderabad - 500032, INDIA October, 2011
- [6] Akshar Bharati, Amba Kulkarni" English from Hindi viewpoint: A Paanianian Perspective" Satyam Computer Services Limited, [Presented at Platinum Jubilee conference of Linguistic Society of India, held at CALTS, University of Hyderabad, Hyderabad, during 6th8th [Dec 2005]
- [7] R.M.K. Sinha , A. Jain "An English to Hindi Machine-Aided Translation System" Indian Institute of Technology, Kanpur, India.
- [8] Kunal Sachdeva, Rishabh Srivastava, Sambhav Jain, Dipti Misra Sharma "Hindi to English Machine Translation:Using Effective Selection in Multi-Model SMT"Language Technologies Research Center, International Institute of Information Technology,

Hyderabad.

- [9] Charles Schafer, David Smith “An Overview of Statistical Machine Translation” [Johns Hopkins University] <http://www.nist.gov/speech/tests/mt>” Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics”
- [10] Jonathan H. Clark Chris Dyer Alon Lavie Noah A. Smith “Better Hypothesis Testing for Statistical Machine Translation: Controlling for Optimizer Instability “[Language Technologies Institute Carnegie Mellon University Pittsburgh, PA 15213, USA]
- [11] Kishore Papineni, Salim Roukos, Todd Ward, and Weijing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proc. of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pages 311–318, Philadelphia, PA, July.
- [12] Latha R. Nair David Peter S, “Machine translation systems for Indian language”, IJCA Journal 2012, Volume 39 - Number 1
- [13] LDC. 2005. Linguistic data consortium Chinese training data sources. <http://www ldc.upenn.edu/Projects/TIDES/mt2005cn.htm>.
- [14] Malcolm Williams "The Application of Argumentation Theory to Translation Quality Assessment"
- [15] Mary Hearne, Andy Way” Statistical Machine Translation: A Guide for Linguists and translator “[School of Computing, Dublin City University].
- [16] Rafal S Uzar “A corpus methodology for analyzing translation “[University of Lodz] .
- [17] Riccardo Schiaffino, Franco Zerao “Translation Quality Measurement in Practice”, 46th ATA Conference, Seattle. Aliquantum ..., 2005
- [18] G V Garje and G K Kharate “survey of machine translation systems India”, International Journal on Natural Language Computing (IJNLC) Vol. 2, No.4, October 2013.
- [19] Yongwei Yang, James Harter, Eldin J. Ehrlich “A Methodological Analysis of Translation Quality”, Journal of Cross-Cultural Psychology, Vol. 37 No. 5, September 2006.
- [20] B. Hettige, A. S. Karunananda “Developing Lexicon Databases for English to Sinhala Machine Translation”, University of Sri Jayewardenepura, Sri Lanka.
- [21] Niladri Chatterjee, Anish Johnson, Madhav Krishnab “Some Improvements over the BLEU Metric for Measuring Translation Quality for Hindi”, Indian Institute of Technology, Hauz Khas, New Delhi.
- [22] Neeraj Tomer, Deepa Sinha “Evaluating Machine Translation Evaluation’s BLEU Metric for English to Hindi Language Machine Translation”, Volume 1, No. 6, August 2012.
- [23] Ondřej Bojar, Vojtěch Diatkey, Pavel Rychlý, Pavel Straňák “HindEnCorp – Hindi-English and Hindi-only Corpus for Machine Translation”, Charles University in Prague.
- [24] Nisheeth Joshi, Iti Mathur “Design of English-Hindi Translation Memory for Efficient Translation”, National Conference on Recent
- [25] Raji P., “Reordering Approach in English-Malayalam Statistical Machine Translation,” *Master’s Thesis*, Coimbatore, India, 2010.
- [26] Sukhadeve P. and Dwivedi S., “Advancement of Clinical Stemmer,” available at: <http://languageinindia.com/may2011/kommaluricomplete.pdf#page=51>, last visited 2013.
- [27] Sukhadeve P. and Dwivedi S., “Developing Hindi POS Tagger for homoeopathy Clinical language,” in *Proceedings of the 2nd International Conference Advances in Computer Science and Information Technology*, Bangalore, India, pp. 310-316, 2012.
- [28] Sukhadeve P. and Dwivedi S., “Enlargement of Clinical Stemmer in Hindi Language of Homoeopathy Province,” in *Proceedings of the 2nd International Conference Advances in Computer Science and Information Technology*, Bangalore, India, pp. 239-248, 2012.
- [29] The Stanford Natural Language Processing Group., available at: <http://nlp.stanford.edu/software/tagger.html>, last visited 2013.
- [30] Unnikrishnan P., Antony P., and Soman K., “A Novel Approach for English to South Dravidian Language Statistical Machine Translation System,” *the International Journal on Computer Science and Engineering*, vol. 2, no. 8, pp. 2749- 2759, 2010.