ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 11, November 2023

Chinese Hindi Machine Translation: A Comparative Study

[1] D.S. Rawat, [2] D.K. Lobiyal

[1] [2] [3] [4] [5] Centre for Chinese and South East Asian Studies, School of Language, Literature and Cultural studies
School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi
Corresponding Author Email: [1] dsrawat.jnu@gmail.com, [2] lobiyal@gmail.com

Abstract*— In this paper, machine translation of Chinese sentences to Hindi has been implemented using Moses which is a statistical machine translation system. Moses has been trained by using a small self-generated corpora of 500 sentence pair of Chinese-Hindi. This model has been tuned using 100 different sentences of Chinese language. All 500 sentences of both Chinese and Hindi language have been pre-processed before they are used for training of the model. These steps include tokenization, truecasing and cleaning of the corpus. The model has been tested for 35 sentences of Chinese which are translated to Hindi. Further, the same sentences of Chinese are also tested using online Google translator and Microsoft Bing translator. Finally, BLEU score has been calculated for each of the translators and their results are compared. The results of Chinese-Hindi sentence pairs show an accuracy of 67%. Also it gives a BLEU score of 0.5. The accuracy obtained by Google translator and Microsoft Bing are 62% and 36%, respectively.

I. INTRODUCTION

Statistical Machine Translation (SMT) is one of the translation approaches developed by researcher due to the limitation of rule-based machine translation systems. SMT poses a demand for large sized parallel corpora of language pairs. Development of corpora itself requires huge efforts and time from the language experts. Machine translation transforms a text from one language to another without the loss of the semantics of the text [1]. There are three main approaches for MT. These are rule based, interlingua based and statistical approaches [2,3]. In the rule based approach sentence in source language is analyzed using morphological, syntactic and semantic rules. Then corresponding target language sentences are generated using natural language generation rules. In interlingua based approach a sentence from the source language is transformed into an intermediate language independent representation from where a target language sentence is generated. However, it is difficult to find a language independent representation. To overcome the limitations of both the approaches, statistical based MT approach were used since 1990 [3]. In this approach, a large parallel corpus is required which is used to determine probability of possible translation in target language. A target sentence of maximum probability is considered as the translation of a given source language sentence. In this paper, Statistical MT based translation approach is used for translating Chinese Hindi sentences and the results of this approach are compared with other approaches.

II. RELATED WORK

Most of the research in China is carried out for Chinese to English MT system. Jinshan, Tongyi, Mingtai and Yaxin are some of the translation systems developed by the researchers[4]. Further, the present status of machine translation research in China is presented in [5]. A Chinese-English Machine Translation System is developed by authors of [6]. This system is developed based on Micro-Engine Architecture. A hybrid approach based, i.e. rules based and statistical Chinese-English Machine Translation System was developed by authors of [7]. A syntax based reordering with automatically derived rules for improved statistical machine translation system is developed by authors of [8]. Authors in [9] developed a translation system for Chinese-English for disambiguating the particles “DE” in Chinese.

In India from 1990 a large number of MT system are developed. Most of these systems are developed for translation between English and Indian languages or among the Indian language. Some of the systems developed for Indian Languages include Anglabharti, Anubharti, Anusaaraka, Anuvaadak, Matra, Mantra, Shakti, UNL based English-Hindi MT system, etc. A survey of these systems is given by authors in [10]. Further, authors in [11] have presented a survey of Indian languages SMT.

III. METHODOLOGY AND DATASET

We developed a small corpus of 500 Chinese-Hindi sentences to train the Moses. However, Google translator and Microsoft Bing are already trained. This parallel corpus consisting of self-developed 500 sentences is used as training data set. A tuning data set of 100 sentences was also developed and used for tuning. Before training the translation system, pre-processing is performed. This includes tokenization, truecasing and cleaning of sentences. However, truecasing is not required for Chinese and Hindi. Cleaning of sentences was also not required as we have limited the length of sentences in both the languages to 80 words only. Further, all the necessary procedures required to train Moses are performed.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 11, November 2023

IV. EXPERIMENTAL RESULTS

To evaluate the performance of translation using three different translators, Moses, Google translator and Microsoft Bing, same 35 Chinese sentences after cleaning are tested. The translation of these Chinese sentences using these translators are given in tables I-III below along with corresponding human translation. For Google translator and Microsoft Bing sentences are tested one-by-one online. But for Moses all these sentences are tested using configuration file.

Table I: Translation of Chinese sentences using Moses.

S.No.	Chinese Sentences	Hindi Sentences (Moses)	Hindi Sentences (Human Understanding)
1.	他們有自己所選擇的水果	जिस क़िस ् म के मेवे पसन	उनके पास खुद चुने हुए फल हैं।
2.	你們究竟否認你們的主的哪一件恩典呢 ？	किस किस नेअमत को झुठलाओगे	तुम आख़िर अपने ईश्वर के किस उपकार को नकारोगे?
3.	曾教授 《 古蘭經 》 ，	उसी ने क़ुरान की तालीम फरमाई	प्रोफेसर चंग “क़ुरान”
4.	有接受勸告的人嗎	तो कोई है जो नसीहत हासिल करे ?	क्या कोई सलाह को स्वीकार करने वाला है?
5.	使它沒有缺陷的 ？	शिगाफ	क्या उसे दोषमुक्त करता है?
6.	山岳逝去之日	पहाड़ उड़ने लगेंगे	मृत पहाड़ के दिन
7.	那對於你們是一樣的	तुम ् हारे लिए	वह तुम लोगों के लिए एक ही है।
8.	是至睿的 。	हिमत वाला	दूरदर्शी
9.	大半是不明理的 。	ऐ रसूल	अधिकांश तर्कसंगत नहीं है।
10.	這等人	यही वह लोग	इस तरह के लोग
11.	能使人生	जिलाता है	जीवन को
12.	那末	अगर	इस प्रकार
13.	獲此美德	जो बड़े नसीबवर हैं	नैतिक उत्कृष्टता प्राप्त करते हैं।
14.	在那日	जिस दिन	उस दिन
15.	不然	बल ् कि	अन्यथा
16.	除獨一至尊的真主外	यकता क़हार खुदा	सर्वोपरि अल्लाह के अलावा
17.	他們就滅亡了	बुझ गए	उनका अन्त हुआ।
18.	你只是一個警告者 。	डराने वाले हो	तुम केवल एक सचेतक हो।
19.	除真主外	ख़ुदा के सिवा	अल्लाह के अलावा
20.	將召集我們	सबको इकट ् ठा करेगा	हमारा आह्वान करेंगे
21.	至於悖逆者	जिन लोगों ने बदकारी की	जहाँ तक विश्वासघातियों की बात है
22.	倆目	सलाम	दो आँखें
23.	至慈的主	ज़ाहिर	कृपालु ईश्वर
24.	謹守拜功	नमाज़ अदा करते हैं	सख़्ती से नियमों का पालन करना।
25.	犯罪的人	कीमिया	अपराधी
26.	如今在那裡呢 ？	शरीक	अभी भी वहाँ है।

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 11, November 2023

S.No.	Chinese Sentences	Hindi Sentences (Moses)	Hindi Sentences (Human Understanding)
27.	真主說	फ़रमाया	अल्लाह कहते हैं
28.	我是印度人。	मैं भारतीय हूं।	मैं भारतीय हूँ।
29.	然後	फिर	बाद में
30.	這	ये	यह
31.	曾否認使者	दूत को झुठलाया	दूत को इनकार किया
32.	應當服從我 。	मेरी इताअत करो	मेरी आज्ञा का पालन करना चाहिए
33.	園圃和源泉 。	मै तो यक़ीनन तुम पर	उद्यान और स्रोत
34.	也沒有忠實的朋友	न कोई दिलबन	न ही वफ़ादार दोस्त है
35.	他將使我死	वह वही हैै जो मुझे मार डालेगा	वह मुझे मार डालेगा।

Table II: Translation of Chinese sentences using Google

S.No.	Chinese Sentences	Hindi Sentences (Google Translator)	Hindi Sentences (Human Understanding)
1.	他們有自己所選擇的水果	फलों के मामले में उनकी अपनी पसंद होती है	उनके पास खुद चुने हुए फल हैं।
2.	你們究竟否認你們的主的哪一件恩典呢 ？	तुम अपने रब की किस कृपा को झुठलाते हो?	तुम आख़िर अपने ईश्वर के किस उपकार को नकारोगे?
3.	曾教授 《 古蘭經 》 ，	प्रोफेसर ज़ेंग "कुरान"	प्रोफेसर चंग “कुरान”
4.	有接受勸告的人嗎	क्या कोई है जो सलाह मानता हो?	क्या कोई सलाह को स्वीकार करने वाला है?
5.	使它沒有缺陷的 ？	इसे दोषमुक्त बनाएं?	क्या उसे दोषमुक्त करता है?
6.	山岳逝去之日	जिस दिन पहाड़ मर गये	मृत पहाड़ के दिन
7.	那對於你們是一樣的	आपके लिए भी यही बात है	वह तुम लोगों के लिए एक ही है।
8.	是至睿的 。	सबसे बुद्धिमान है	दूरदर्शी
-9.	大半是不明理的 。	उनमें से अधिकांश अनुचित हैं	अधिकांश तर्कसंगत नहीं है।
10.	這等人	इस तरह के लोग	इस तरह के लोग
11.	能使人生	जीवन बना सकते हैं	जीवन को
12.	那末	तब	इस प्रकार
13.	獲此美德	इस पुण्य को प्राप्त करें	नैतिक उत्कृष्टता प्राप्त करते हैं।
14.	在那日	उस दिन	उस दिन
15.	不然	अन्यथा	अन्यथा
16.	除獨一至尊的真主外	सिवाय एक और केवल अल्लाह के	सर्वोपरि अल्लाह के अलावा
17.	他們就滅亡了	और वे नष्ट हो गये	उनका अन्त हुआ।

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 11, November 2023

S.No.	Chinese Sentences	Hindi Sentences (Google Translator)	Hindi Sentences (Human Understanding)
18.	你只是一個警告者 。	आप सिर्फ एक चेतावनी हैं	तुम केवल एक सचेतक हो।
19.	除真主外	सिवाय अल्लाह के	अल्लाह के अलावा
20.	將召集我們	हमें बुलाएंगे	हमारा आह्वान करेंगे
21.	至於悖逆者	जहाँ तक अवज्ञाकारियों का प्रश्न है	जहाँ तक विश्वासघातियों की बात है
22.	倆目	दो आँखें	दो आँखें
23.	至慈的主	परम दयालु प्रभु	कृपालु ईश्वर
24.	謹守拜功	प्रार्थनाओं का ध्यान रखें	सख़्ती से नियमों का पालन करना।
25.	犯罪的人	आपराधिक	अपराधी
26.	如今在那裡呢 ？	यह अब कहाँ है	अभी भी वहाँ है।
27.	真主說	अल्लाह कहता है	अल्लाह कहते हैं
28.	我是印度人。	मै भारतीय हूँ।	मैं भारतीय हूँ।
29.	然後	तब	बाद में
30.	這	यह	यह
31.	曾否認使者	एक बार दूत को मना कर दिया	दूत को इनकार किया
32.	應當服從我 。	मेरी बात माननी चाहिए	मेरी आज्ञा का पालन करना चाहिए
33.	園圃和源泉 。	उद्यान और फव्वारे	उद्यान और स्रोत
34.	也沒有忠實的朋友	कोई वफादार दोस्त नहीं	न ही वफ़ादार दोस्त है
35.	他將使我死	वह मुझे मार डालेगा	वह मुझे मार डालेगा।

Table III: Translation of Chinese sentences using Microsoft Bing

S.No.	Chinese Sentences	Hindi Sentences (Micrsoft Bing)	Hindi Sentences (Human Understanding)
1.	他們有自己所選擇的水果	उनके पास फलों की अपनी पसंद है।	उनके पास खुद चुने हुए फल हैं।
2.	你們究竟否認你們的主的哪一件恩典呢 ？	आप अपने रब की किस कृपा से इनकार करते हैं?	तुम आख़िर अपने ईश्वर के किस उपकार को नकारोगे?
3.	曾教授 《 古蘭經 》 ，	कुरान पढ़ाया जाता है	प्रोफेसर चंग “कुरान”
4.	有接受勸告的人嗎	क्या कोई है जो सलाह स्वीकार करता है?	क्या कोई सलाह को स्वीकार करने वाला है?
5.	使它沒有缺陷的 ？	इसे निर्दोष बनाएं	क्या उसे दोषमुक्त करता है?
6.	山岳逝去之日	जिस दिन पहाड़ मर गया	मृत पहाड़ के दिन

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 11, November 2023

S.No.	Chinese Sentences	Hindi Sentences (Micrsoft Bing)	Hindi Sentences (Human Understanding)
7.	那對於你們是一樣的	यह आपके लिए भी ऐसा ही है।	वह तुम लोगों के लिए एक ही है।
8.	是至睿的 。	यह सबसे शानदार है	दूरदर्शी
9.	大半是不明理的 。	इसमें से अधिकांश अनुचित है।	अधिकांश तर्कसंगत नहीं है।
10.	這等人	ऐसे लोग	इस तरह के लोग
11.	能使人生	जीवन बना सकते हैं	जीवन को
12.	那末	बस	इस प्रकार
13.	獲此美德	इस पुण्य को प्राप्त करें	नैतिक उत्कृष्टता प्राप्त करते हैं।
14.	在那日	उस दिन	उस दिन
15.	不然	नहीं तो	अन्यथा
16.	除獨一至尊的真主外	अद्वितीय सर्वोच्च अल्लाह के अलावा	सर्वोपरि अल्लाह के अलावा
17.	他們就滅亡了	वे मर गए।	उनका अन्त हुआ।
18.	你只是一個警告者 。	आप सिर्फ एक चेतावनी देने वाले हैं	तुम केवल एक सचेतक हो।
19.	除真主外	सिवाय अल्लाह के।	अल्लाह के अलावा
20.	將召集我們	हमें तलब किया जाएगा।	हमारा आह्वान करेंगे
21.	至於悖逆者	विद्रोहियों के लिए	जहाँ तक विश्वासघातियों की बात है
22.	倆目	दोनों	दो आँखें
23.	至慈的主	परम दयालु प्रभु	कृपालु ईश्वर
24.	謹守拜功	गुणों का निरीक्षण करें	सख़्ती से नियमों का पालन करना।
25.	犯罪的人	अपराध करने वाले लोग	अपराधी
26.	如今在那裡呢 ？	अब यह वहाँ है	अभी भी वहाँ है।
27.	真主說	अल्लाह ने कहा:	अल्लाह कहते हैं
28.	我是印度人。	मैं भारतीय हूं।	मैं भारतीय हूँ।

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 11, November 2023

S.No.	Chinese Sentences	Hindi Sentences (Micrsoft Bing)	Hindi Sentences (Human Understanding)
29.	然後	उसके बाद	बाद में
30.	這	यहन	यह
31.	曾否認使者	उसने दूत से इनकार कर दिया।	दूत को इनकार किया
32.	應當服從我 。	मेरे प्रति आज्ञाकारी बनो	मेरी आज्ञा का पालन करना चाहिए
33.	園圃和源泉 。	बगीचे और फव्वारे	उद्यान और स्रोत
34.	也沒有忠實的朋友	कोई वफादार दोस्त भी नहीं हैं।	न ही वफ़ादार दोस्त है
35.	他將使我死	वह मुझे मरने पर मजबूर कर देगा।	वह मुझे मार डालेगा।

We calculated Bilingual Evaluation Understudy (BLEU) score online for the results obtained from all the three translators for the 35 sentences for Chinese-Hindi translation using Interactive BLEU Score Evaluation. The results of BLEU score are given below in the tables IV-VI. The accuracy achieved using Moses, Google translator and Microsoft Bing are 67%, 62% and 36%. Microsoft Bing gives the poorest results as it has used large language model. However, it may not have been trained well for Chinese language text. Google translator also give results closed to Moses. Google translator is trained using Neural Machine translation. But training for Chinese language text may be a limitation. However, Moses gives best results as it is trained by the Chinese-Hindi corpus developed by us. But the results obtained are of moderate quality as the corpus size is small.

Table–IV: BLEU score for Moses translation

BLEU 0.67 Precision X Brevity = 0.67 X 100				
Type	1-gram	2-gram	3-gram	4-gram
Individual	4.32	0.53	0.37	0.28
Cumulative	4.32	1.51	0.89	0.67

Table–V: BLEU score for Google translation

BLEU 0.62 Precision X Brevity = 0.62 X 100				
Type	1-gram	2-gram	3-gram	4-gram
Individual	4.26	0.47	0.32	0.23
Cumulative	4.26	1.41	0.86	0.62

Table–VI: BLEU score for Microsoft Bing translation

BLEU 0.36 Precision X Brevity = 0.36 X 100				
Type	1-gram	2-gram	3-gram	4-gram
Individual	2.72	0.35	0.19	0.10
Cumulative	2.72	0.98	0.56	0.36

V. CONCLUSION

This is an attempt to investigate the performance of Chinese- Hindi translation using Moses, Google translator and Microsoft Bing. The result obtained are of moderate quality as these languages belong to different families. Chinese belongs to Sino- Tibetan and Hindi to Indo-European language families and they also use different type of scripts. However, this comparison is quite encouraging as Moses and Google gives closer accuracy. However, this translation is carried out considering a small parallel corpus of Chinese-Hindi sentence pairs. In future, we will investigate the accuracy of Moses translation for large sized corpus and the results will be compared with other translators. Also we will make attempts to include language divergences for improving the quality of Moses translation.

RÉFÉRENCES

[1] Phillip Koehn, Machine Translation in Daniel M.Bikel and Imed Zitouni (ed.), *Multilingual Natural Language Processing Application: from Theory to Practice*, IBM Press, copyright 2012. P. 331.

[2] Jun-ichi Tsujii, Machine Translation: Future Aspects in Makoto Nagao(ed.), *Language and Artificial Intelligence: Proceeding of an International symposium on Language and*

Artificial Intelligence held in Kyoto, Japan, 16-21 March, 1986, Elsevier Science Publishers B.V., 1987(Copyright), PP.265-282.

[3] Chirstopher D. Maning and Hinrich Schutze, *Foundation of Statistical Natural Language Processing*, MIT Press, 1999.

[4] Hu Miaodan(胡妙丹), Pre-editing in Chinese to English Machine Translation: A case study of the Google Machine Translation Engine(汉译英机器翻译的议前编辑-以谷歌翻译为例), M.A Thesis, Zhejiang nor42mal University,May2012.

[5] Wang Lei (王蕾), Analysis on History and Present Status of Machine Translation Research (机器翻译研究的历史和现状分析), 研士学位论文, 外国语言学及饮用语言学, 湖壮工紫大学, 2013.

[6] Liu Qun, A Chinese-English Machine Translation System Based on Micro- Engine Architecture, in proceedings of 3rd Conference of the Association for Machine Translation in the Americas, AMTA '98, Langhorne, PA, USA, October 28-31, 1998.

[7] Qun Liu and Shiwen Yu, Trans Easy: a Chinese-English Machine Translation System Based on Hybrid Approach, in proceedings of 3rd Conference of the Association for Machine Translation in the Americas, AMTA '98, Langhorne, PA, USA, October 28-31, 1998.

[8] Karthik Visweswariah, et al., Syntax Based Reordering with Automatically Derived Rules for Improved Statistical Machine Translation, *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), Beijing, August2010,*PP.1119-1127.

[9] Pi-Chuan Chang, et al., Disambiguating "DE" for Chinese-English Machine Translation, in proceedings of fourth Wsssorkshop on Statistical Machine Translation, Athens Greece, March 03-31, 2009.

[10] A Godase, S Govilkar, "Machine translation development for indian languages and its approaches", *International Journal on Natural Language Computing (IJNLC)*, Vol. 4, No.2, 2015.

[11] B.N.V.N. Raju and M.S.V.S. Bhadri Raju, " Statistical Machine Translation System for Indian Languages", *6th International Conference on Advanced Computing*, IEEE, 2016.