



Transliteration System To Aid Learning Pronunciations For Non-Native Speakers Using A Hybrid Approach

¹Radha Mogla, ²C. Vasantha Lakshmi, ³Niladri Chatterjee

¹Research Scholar, ²Professor, ³Professor

¹Department of Physics & Computer Science, Faculty of Science, Dayalbagh Educational Institute, Dayalbagh, Agra, Uttar Pradesh, 282 005, India

² Department of Physics & Computer Science, Faculty of Science, Dayalbagh Educational Institute, Dayalbagh, Agra, Uttar Pradesh, 282 005, India

³Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas New Delhi, Delhi 110016, India

Abstract : India is a linguistically rich country and extensive literature is available in many languages. Many people want to read and admire literature available in these languages, but language becomes a barrier. Some people may understand a particular language, and yet find the script difficult to read. In learning a new language, one of the first steps is learning the pronunciation of the various words. To make learning Hindi pronunciation easy for a non-native speaker, the present work uses a UTF-8-based transliteration method to transliterate from Hindi to Telugu by considering the distinct nuances pronunciation of both languages. Another hybrid transliteration system is presented to transliterate English lexical words to Telugu using Hindi as a bridge language even without any English-Telugu parallel database. The results for English to Telugu show that without using any new parallel database, transliteration systems can be developed for English to other Indian languages by using Hindi as a bridge language. These systems can help in learning the pronunciation of a new language and also allow the users to peruse the vast literature available in other languages.

Index Terms - Transliteration, Phrase based machine translation, UTF-8 based transliteration.

I. INTRODUCTION

Natural Language Processing (NLP) is a vast research area. Works related to NLP may be broadly classified into two categories: (a) development of language independent technologies and (b) development of language-specific technologies. In the first category various techniques are developed to aid NLP problems based on mathematical or statistical modelling and open-source toolkits are also developed like Moses for Statistical Machine Translation (Koehn et al, 2007) and Sequitur grapheme to phoneme converter using Join sequence model (Bisani and Ney,2008). The second category deals with language specific monolingual or bilingual NLP tasks in which language dependent tools are developed for NLP problems like transliteration of English words to Korean (Oh and Choi,2002), Hybrid approach for English to Hindi machine translation (Gupta and Chatterjee,2017) to name a few.

India is a linguistically rich country and research work involving Indian languages has been pursued diligently. Computer processing of Bangla language (Sengupta and Chaudhuri, 1993), Factor-based evaluation Machine-aided translation (Sinha et al., 1995), English-Punjabi Machine Translation (Jindal et al., 2017), Factor-based evaluation for English to Hindi translation (Balyan and Chatterjee, 2018),Rule-Based Machine Translation system for Marathi to English (Kharate and Patil, 2019) and Neural machine translation between Indian language pairs (Dewangan et al., 2021) are some of the prominent works in this concept.

The work presented in this paper belongs to the second category in which a language dependent transliteration system is developed for transliterating Hindi lexical words to Telugu by considering the phonetic gaps between both the languages. Transliteration is the task of converting a text written in the source language, to the target language, by preserving the phonetic properties of the source language while following the pronunciation restrictions of the target language.

Learning pronunciation of a new language is one of the main steps in learning a new language. If a reader knows Telugu, then having the Hindi words transliterated into Telugu can make the reading process easy as the reader will get the pronunciation for every Hindi word in Telugu. This system will be helpful in transcribing literature available in Hindi and people trying to read a new language. There are text to speech converters which can also help in this but due to the phonetic restrictions of the target language there will always be touch of source language while learning pronunciation using text to speech tool. If a transliteration system is developed by considering the phonetic gaps between the source and the target language and pronunciations of the source language's lexical words is given in the readers native language then better pronunciation can be learned. This idea leads to the present work where a script specific transliteration system is developed for transliterating English to Telugu using Hindi as a bridge language where Hindi words are converted to Telugu using UTF-8 based encoding.

II. LITERATURE REVIEW

Many researchers in the past have attempted transliteration between English and Indian languages, but transliteration between Indian languages is rare. Some relevant work done for transliteration between Indian languages including English is discussed in this section:

Kunchukuttan et al., (2021) used a LSTM based Encoder-decoder model for multilingual transliteration between English and Indian languages based on orthographic similarities between Indian languages. In multi-lingual transliteration they observed an increase of 6-7% when transliterating from English.

Dhindsa and Sharma, (2017) used a combination-based approach for transliteration of Named Entities from English to Hindi. Their approach consists of mainly two steps: searching the words in a pronunciation dictionary and if the word is not found in the dictionary then they transliterated the words based on some predefined rules.

Kunchukuttan et al., (2015) developed Brahmi-Net for transliteration between Indian languages using Phrase-Based Machine Translation for training the transliteration system. They observed improved translation accuracy when using transliteration for out of vocabulary words. They quoted the translation accuracy in terms of BLUE scores for 8 languages pairs which was between 6.58 to 71.26 and observed an increase in BLUE score of about 0.5 to 3 for different language pairs when using translation with transliteration.

Srivastava and Bhat, (2013) used a Statistical Machine Translation method using Moses and Giza++ to train the transliteration model for Hindi to seven other Indian languages. They evaluated the results on two datasets and observed the word level accuracy of 27% and 34% for Hindi to Telugu transliteration. However they claimed that human evaluators gave an average score of 3.6 out of 4 for Hindi to Telugu transliteration system.

Sowmya and Varma, (2009) proposed a transliteration-based text input method for Telugu based on edit distance. They used English to Telugu transliteration for typing Telugu text.

Goyal and Singh, (2009) developed a rule-based transliteration system with 57 rules for Hindi to Punjabi transliteration. They claimed that 100000 words from various domains were transliterated accurately.

III. HINDI AND TELUGU

Hindi and Telugu are two distinct languages written using two very different scripts. Hindi is written using Devnagari Script, whereas Telugu is written using Telugu script. Phonetically both languages are very similar, and most of the letters of both scripts can be matched one to one. For Hindi letters, UTF-8 based encoding is e0 a4 ** and e0 a5 **, and for Telugu letters, it is e0 b0 ** and e0 b1 **. Here ** is the combination of English letters and numbers which represent a unique code for a particular letter of Hindi and Telugu. Figure 1. shows some Hindi letters, their UTF-8 code, and their Telugu equivalents.

Letters	Hindi Letters	UTF-8	Telugu Letters	UTF-8
Letter A(vowel)	अ	e0 a4 85	అ	e0 b0 85
Letter AA (vowel)	आ	e0 a4 86	ఆ	e0 b0 86
Letter KA (consonant)	क	e0 a4 95	క	e0 b0 95
Letter KHA (consonant)	ख	e0 a4 96	ఖ	e0 b0 96
Vowel sign (Modifier) AA	ा	e0 a4 be	ా	e0 b0 be
Vowel sign (Modifier) I	ि	e0 a4 bf	ి	e0 b0 bf
Vowel sign (Modifier) U	ु	e0 a5 81	ు	e0 b1 81
Vowel sign (Modifier) UU	ू	e0 a5 82	ూ	e0 b1 82

Fig.1. One to one matching of Hindi and Telugu letters along with their UTF-8 code

There are some phonetic gaps between Hindi and Telugu, which are as follows:

There are few extra vowels and vowels modifiers in Telugu which are not there in Hindi, so some of the Hindi vowels had to be matched with the equivalent Telugu letters.

Letters with “Nukta” (a diacritic which changes the pronunciation of a letter and is represented by a dot below the letter) in Hindi are used to write loan words. In Hindi, some of the loan words are written by adding “Nukta”(a dot below the consonant) these letters are to be matched with the equivalent sounding letters in Telugu as “Nukta” is not present in Telugu.

In Telugu, when the last letter of a word is a consonant, it is pronounced by adding a “long a” sound to it. This affects the pronunciation of a transliterated Hindi word in Telugu. So in the transliterated Telugu output if the last letter of the word is a consonant it should be converted to its equivalent half consonant.

These gaps are identified between Hindi and Telugu, and the Hindi words are transliterated by considering these pronunciation gaps between the two languages

IV. METHODOLOGY USED

In the present work a hybrid system is presented for English lexical words to Telugu transliteration. For this purpose, first English to Hindi transliteration is done using a Phrase Based Statistical Machine Translation model trained for English lexical words to Hindi transliteration using the same database and settings described by Mogla et al, (2021). The transliterated Hindi words are further converted to Telugu using UTF-8 based encoding. By using Hindi as a bridge language between English and Telugu. English to Telugu transliteration is performed without using English-Telugu parallel database.

English to Hindi Transliteration: English is a language having diverse rules for pronunciation for letter and grapheme. Same letter be it a vowel or a consonant often have different pronunciations when they appear in different words. Some times same letter is pronounced differently when it appears more than once in a word. For example, in the word “doctor” (dɒktə(r)) the “letter o” appears twice with two different pronunciations “ɒ” and “ə”. Mogla et al, 2021 created an English-Hindi parallel data base of around 2000 English lexical words by considering the diversity in English pronunciation and trained a transliteration system using Phrase-Based Statistical Machine translation system using Moses (Koehn et al., 2007) and Giza++ (Och and Ney, 2003) for English to Hindi transliteration using a small and diverse training data. They segmented the words letter wise for training and testing and after some script specific post processing used another parallel database of 100 most frequent English words for direct transliteration of the test set. Their transliteration system was tested on five short stories. In the present work for English to Telugu transliteration, instead of developing a parallel database for English to Telugu transliteration, Hindi is used as a bridge language. The outputs generated for the English to Hindi transliteration system for two stories given by Mogla et al. (2021) are further converted to Telugu using UTF-8 based encoding.

Hindi to Telugu Transliteration: For Hindi to Telugu transliteration UTF-8 based conversion is used. Some of the Hindi letters and modifiers are matched to their equivalent sounding characters rest are matched with their equivalent UTF-8 code by changing a4 to b0 and a5 to b1 in the UTF-8 code. Figure 2. shows the letters of Hindi and their equivalent similar sounding Telugu letters considered during this work for transliteration.

<i>Hindi Letter</i>		<i>Considered Equivalent Telugu letter</i>	
Letter FA	फ़	Letter PHA	ఫ
Letter ZA	ज़	Letter JA	జ
Letter O	ओ	Letter O	ఓ
Letter AU	औ	Letter OO	ఔ
Letter Chandra O	ऑ	Letter AU	ఌ
Modifier O	ो	Modifier O	ొ
Modifier AU	ौ	Modifier OO	ౌ
Modifier Chandra O	ॉ	Modifier AU	ౌ
Letter DDDHA	ड़	Letter DDA	డ
Purna Viram (full stop)		Full stop	.

Fig.2. Hindi letters and their similar sounding letters considered during transliteration

V. RESULTS AND CONCLUSION

Two Hindi paragraphs are taken from the Hindi Wikipedia page (Hindi Wikipedia) and are transliterated to Telugu using UTF-8 based conversion. The outputs generated are given to Human evaluators who are familiar with both languages and are asked to evaluate the transliteration accuracies. The evaluators returned around 100% transliteration accuracy with the phonetic restrictions of the target language and considering the phonetic gaps between the source and target languages. Figure 3 shows the input Hindi paragraphs and their Telugu transliteration.

दिल्ली सल्तनत के उदय से पहले आगरा का इतिहास स्पष्ट नहीं है। १७ वीं शताब्दी के एक वृत्तान्त में सिकंदर लोदी (१४८८-१५१७) के समय से पहले आगरा को एक पुरानी बस्ती के रूप में बुलाया था, जो महमूद गजनवी द्वारा इसके विनाश के कारण महज एक गाँव था। ११ वीं सदी के फ़ारसी कवि मासूद सलमान ने आगरा के किले पर गजनवी के आक्रमण का उल्लेख किया है, जो तब राजा जयपाल के शासनाधीन था। जयपाल के आत्मसमर्पण के बावजूद, महमूद ने किले को लूट लिया था। १५०४ में सिकंदर लोदी ने अपनी राजधानी को दिल्ली से आगरा स्थानांतरित किया था। उनके काल में किले में कई महल, कुएँ और एक मस्जिद का निर्माण किया गया। १५२६ में पानीपत के प्रथम युद्ध में हार के बाद यह मुगल शासन के अंतर्गत आया। १५४० और १५५६ के बीच, शेरशाह सूरी ने इस क्षेत्र पर शासन किया। यह १५५६ से १६४८ तक मुगल साम्राज्य की राजधानी रहा। आगरा पर बाद में मराठों का अधिपत्य रहा, जिनके बाद यह ब्रिटिश राज के अंतर्गत आ गया।

आगरा अपनी कई मुगलकालीन इमारतों के कारण एक प्रमुख पर्यटन स्थल है, विशेषकर ताजमहल, आगरा किला और फतेहपुर सीकरी के लिये, जो सभी यूनेस्को विश्व धरोहर स्थल हैं। दिल्ली और जयपुर के साथ आगरा गोलडन ट्राइंगल टूरिस्ट सर्किट में शामिल है; और लखनऊ और वाराणसी के साथ यह उत्तर प्रदेश राज्य के एक पर्यटक सर्किट, उत्तर प्रदेश हेरिटेज आर्क का हिस्सा है। सांस्कृतिक रूप से आगरा ब्रज क्षेत्र में स्थित है। आगरा २७.१८° उत्तर ७८.०२° पूर्व में यमुना नदी के तट पर स्थित है। समुद्र-तल से इसकी औसत ऊँचाई करीब १७१ मीटर (५६१ फीट) है। यह यमुना एक्सप्रेसवे के माध्यम से दिल्ली से और आगरा-लखनऊ एक्सप्रेसवे के माध्यम से लखनऊ से जुड़ा हुआ है।

दिल्ली सल्तनत के उदय से पहले आगरा का इतिहास स्पष्ट नहीं है। १७ वीं शताब्दी के एक वृत्तान्त में सिकंदर लोदी (१४८८-१५१७) के समय से पहले आगरा को एक पुरानी बस्ती के रूप में बुलाया था, जो महमूद गजनवी द्वारा इसके विनाश के कारण महज एक गाँव था। ११ वीं सदी के फ़ारसी कवि मासूद सलमान ने आगरा के किले पर गजनवी के आक्रमण का उल्लेख किया है, जो तब राजा जयपाल के शासनाधीन था। जयपाल के आत्मसमर्पण के बावजूद, महमूद ने किले को लूट लिया था। १५०४ में सिकंदर लोदी ने अपनी राजधानी को दिल्ली से आगरा स्थानांतरित किया था। उनके काल में किले में कई महल, कुएँ और एक मस्जिद का निर्माण किया गया। १५२६ में पानीपत के प्रथम युद्ध में हार के बाद यह मुगल शासन के अंतर्गत आया। १५४० और १५५६ के बीच, शेरशाह सूरी ने इस क्षेत्र पर शासन किया। यह १५५६ से १६४८ तक मुगल साम्राज्य की राजधानी रहा। आगरा पर बाद में मराठों का अधिपत्य रहा, जिनके बाद यह ब्रिटिश राज के अंतर्गत आ गया।

आगरा अपनी कई मुगलकालीन इमारतों के कारण एक प्रमुख पर्यटन स्थल है, विशेषकर ताजमहल, आगरा किला और फतेहपुर सीकरी के लिये, जो सभी यूनेस्को विश्व धरोहर स्थल हैं। दिल्ली और जयपुर के साथ आगरा गोलडन ट्राइंगल टूरिस्ट सर्किट में शामिल है; और लखनऊ और वाराणसी के साथ यह उत्तर प्रदेश राज्य के एक पर्यटक सर्किट, उत्तर प्रदेश हेरिटेज आर्क का हिस्सा है। सांस्कृतिक रूप से आगरा ब्रज क्षेत्र में स्थित है। आगरा २७.१८° उत्तर ७८.०२° पूर्व में यमुना नदी के तट पर स्थित है। समुद्र-तल से इसकी औसत ऊँचाई करीब १७१ मीटर (५६१ फीट) है। यह यमुना एक्सप्रेसवे के माध्यम से दिल्ली से और आगरा-लखनऊ एक्सप्रेसवे के माध्यम से लखनऊ से जुड़ा हुआ है।

Fig. 3 Telugu Transliterated Output for Hindi Input paragraph

Two of the stories transliterated in Hindi by Mogla et al., (2021) are used for English to Telugu transliteration. Their outputs are further converted to Telugu using UTF-8 based transliteration. Telugu outputs, along with the English stories, are given to experts familiar with English and Telugu. The outputs are evaluated by considering two categories of transliterated words: wrong transliteration and phonetically similar transliteration.

Table 1 shows the accuracy of the English to Telugu transliteration system. Figure 4. shows the Telugu output of the second English story (from 'Gutenberg E-Book of Aesop's Fables'), and the wrong transliteration is highlighted by yellow color. Accuracies for Telugu outputs are the same as the accuracies reported by Mogla et al., (2021) for English to Hindi transliteration.

Table1: Accuracy for English to Telugu transliteration using a hybrid approach

Story	Wrong transliteration	Phonetically similar or correct transliteration	Accuracy
Story 1	17	134	88.74%
Story 2	23	194	89.40%

The results obtained by using an English to Hindi transliteration system for English to Telugu transliteration show good accuracies. The same system can be used to transliterate English lexical words to other Indian languages using Hindi as a bridge language and by identifying the pronunciation gaps between Hindi and the target languages without developing any new parallel database. As English is developing as one of the most popular second language for people having different mother tongue, a transliteration system from English to Indian languages can help in learning the pronunciation of English lexical words for better learning of the language.

Pronunciation gaps between Indian languages other than English can also be identified to develop a transliteration system between them using UTF-8 based transliteration. These systems will enable non-native speakers to learn the pronunciation of a new language and admire the literature of a language which a user can understand but cannot read.

<p>THE HARE AND THE TORTOISE</p> <p>A HARE one day made himself merry over the slow pace of the Tortoise, vainly boasting of his own great speed in running.</p> <p>The Tortoise took the laughing and boasting in good part. "Let us try a race," she said; "I will run with you five miles for five dollars, and the Fox out yonder shall be the judge."</p> <p>The Hare agreed, a course was arranged, and away they started together. True to his boasting the Hare was out of sight in a moment.</p> <p>The Tortoise never for a moment stopped, but jogged along with a slow steady pace, straight to the end of the course. Full of sport, the Hare first outran the Tortoise, then fell behind. Having come midway to the goal, he began to nibble at the young herbage, and to amuse himself in many ways. After a while, the day being warm, he lay down for a nap, saying: "She is behind me now. If she should go by, I can easily enough catch up."</p> <p>When the Hare awoke, the Tortoise was not in sight; and running as fast as he could, he found her comfortably dozing at their goal.</p> <p>Those who are very quick are apt to be too sure. Slow and steady often wins the race.</p>	<p>ద్ హెరర్ వున్ ద్ టోటస్</p> <p>ఐ హెరర్ వన్ డే మేడ్ హిమ్సెల్ఫ్ మెరి ఒవర్ ద్ స్లో పేస్ డ్ టోటస్ , వేగ్ బాస్టింగ్ డ్ హిస్ ఓన్ గ్రేట్ స్పీడ్ ఇన్ రనింగ్ .</p> <p>ద్ టోటస్ టుక్ ద్ లాఫింగ్ అండ్ బాస్టింగ్ ఇన్ గుడ్ పార్ట్ . "లెట్ అస్ ట్రి ఐ రేస్ , " షి సెడ్ ; "ఆయ్ విల్ రన్ విత్ యూ ఫివ్ మిల్స్ ఫార్ ఫివ్ డాల్లర్స్ , వున్ ద్ ఫాక్స్ ఆజిట్ యోన్డర్ షాల్ బి ద్ జిడ్జి . "</p> <p>ద్ హెరర్ అగ్రీడ్ , ఐ కౌస్ వాజ్ అరేంజ్డ్ , వున్ వే దే కౌర్స్ టు ద్ టెండ్ . లూ టూ హిజ్ బాస్టింగ్ ద్ హెరర్ వాజ్ ఆజిట్ డ్ సాయ్స్ ఇన్ ఐ మొమెంట్స్ .</p> <p>ద్ టోటస్ నెవర్ ఫర్ అ మొమెంట్ స్టాప్డ్ , బట్ జాగ్డ్ అలొంగ్ విత్ అ స్లో , స్టేడీ పేస్ , స్ట్రేట్ టూ ద్ టెండ్ గోల్ . ఫుల్ ఆఫ్ స్పోర్ట్ , ద్ హెరర్ ఫిర్స్ట్ ఓవ్రాన్ ద్ టోటస్ , డెన్ బింద్ బింద్ బింద్ . హెవింగ్ కమ్ మిడ్వే టూ ద్ గోల్ , హి బిగన్ నెబిల్ ద్ యంగ్ హెర్బేజ్ , అండ్ టు అమ్యూజ్ హిమ్సెల్ఫ్ ఇన్ మెన్యూ వేయ్స్ . ఆఫర్ అ వైల్ , ద్ డే బింగ్ వామ్ , హి లై డౌన్ ఫర్ అ నాప్ , సేయింగ్ : "షి ఐ బింద్ మి నోవ్ . ఐఫ్ షి షుల్డ్ గో బై , ఐ కన్ టేస్ట్ లిట్టెల్ ఓవర్ డ్ హి . "</p> <p>వెన్ ద్ హెరర్ అవోకైడ్ , ద్ టోటస్ వాజ్ నోట్ ఇన్ సైట్ ; అండ్ రనింగ్ అజ్ ఫాస్ట్ అజ్ హి కౌల్డ్ , హి ఫౌండ్ హిర్ కంఫర్టబిల్ డోయింగ్ అట్ ద్ గోల్ .</p> <p>దోజ్ వాజ్ ఆర్ వరీ క్వీక్ ఆర్ అప్ట టు బి టూ సూర్ . స్లో అండ్ స్టేడీ ఓఫెన్ విన్ ద్ రేస్ .</p>
--	--

Fig.4 Telugu transliteration of English story

REFERENCES

[1]. Anoop Kunchukuttan, Siddharth Jain, and Rahul Kejriwal. 2021. A Large-scale Evaluation of Neural Machine Transliteration for Indic Languages. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 3469–3475, Online. Association for Computational Linguistics.

[2]. Dhindsa, B., K., and Sharma, D., V., English to Hindi Transliteration System Using Combination-Based Approach. International Journal of Advanced Research in Computer Science, vol. 8, no. 8, pp. 609–613. 2017. <https://doi.org/10.26483/ijarcs.v8i8.4801>

[3]. Gupta, Susmita, and Niladri Chatterjee. "A hybrid approach using phrases and rules for Hindi to English machine translation." International Journal on Natural Language Computing 6, no. 3 (2017): 57-73.

[4]. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E., 2007, June. Moses: Open source toolkit for statistical machine translation. In Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions .177-180. <https://doi.org/10.3115/1557769.1557821>

[5]. Kunchukuttan, A., Puduppully, R., Bhattacharyya, P., Brahmi-Net: A transliteration and script conversion system for languages of the Indian subcontinent , Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies: System Demonstrations . 2015

[6]. M. Bisani and H. Ney. "Joint-Sequence Models for Grapheme-to-Phoneme Conversion". Speech Communication, Volume 50, Issue 5, May 2008, Pages 434-451

[7]. M.K. Sinha, K. Sivaraman, A. Agrawal, R. Jain, R. Srivastava, & A. Jain (1995). ANGLABHARTI: a multilingual machine aided translation project on translation from English to Indian languages. In 1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century. IEEE. <https://doi.org/10.1109/icsmc.1995.538002>

[8]. Namrata G Kharate, & Dr.Varsha H. Patil. Handling Challenges in Rule Based Machine Translation from Marathi to English. International Journal on Natural Language Computing, 8(4), pages39–49. 2019. <https://doi.org/10.5121/ijnlc.2019.8404>

[9]. Och, F.J. and Ney, H., A systematic comparison of various statistical alignment models. Computational linguistics, 29(1), 19-51. 2003. <https://doi.org/10.1162/089120103321337421>

[10]. Oh, Jong-Hoon, and Key-Sun Choi, "An English-Korean transliteration model using pronunciation and contextual rules" in Proceedings of the 19th international conference on Computational linguistics-Vol. 1, Association for Computational Linguistics, pp. 1-7. 2002

[11]. R. Mogla, C. V. Lakshmi and N. Chatterjee, 'A Systematic Approach for English- Hindi Parallel Database Creation for Transliteration of General Domain English Words,' 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-5, doi: 10.1109/GUCON50781.2021.9573841.

[12]. Renu Balyan, & Niladri Chatterjee. Factor-based evaluation for English to Hindi MT outputs. Language Resources and Evaluation, 52(4), pages 969–996. 2018. <https://doi.org/10.1007/s10579-018-9426-y>

- [13]. Rishabh Srivastava, & Riyaz Ahmad Bhat, Transliteration systems across indian languages using parallel corpora. In Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27), pages 390-398. 2013
- [14]. Sengupta, & B B Chaudhuri. Natural Language Processing in an Indian Language (Bengali)-I: Verb Phrase Analysis. IETE Technical Review, 10(1), pages 27–41. 1993. <https://doi.org/10.1080/02564602.1993.11437284>
- [15]. Shishpal Jindal, Vishal Goyal, & Jaskarn Singh. Building English-Punjabi Parallel corpus for Machine Translation. International Journal of Computer Applications, 180(8), pages 26–29.2017. <https://doi.org/10.5120/ijca2017916036>
- [16]. Shubham Dewangan, Shreya Alva, Nitish Joshi, & Pushpak Bhattacharyya. Experience of neural machine translation between Indian languages. Machine Translation, 35(1), pages 71–99.2021. <https://doi.org/10.1007/s10590-021-09263-3>
- [17]. Sowmya, V. B., & Vasudeva Varma. (2009, March). Transliteration based text input methods for telugu. In International Conference on Computer Processing of Oriental Languages, pages 122-132. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-00831-3_12
- [18]. (Gutenberg E-Book of Aesop's Fables')<http://www.gutenberg.org/files/49010/49010-0.txt>
- [19]. (oald) <https://www.oxfordlearnersdictionaries.com/>
- [20]. (Hindi wikipedia) <https://hi.wikipedia.org/wiki/%E0%A4%86%E0%A4%97%E0%A4%B0%E0%A4%BE>

