A MORE ACCURATE PUNJABI TO ENGLISH MACHINE TRANSLITERATION SYSTEM FOR PROPER NOUNS

Manpreet Kaur

Department of Computer Science, Punjabi University, Patiala.

ABSTRACT

Machine Transliteration has come out to be an emerging and very important part of NLP, which is not only concerned with representation of sounds in original, the characters, but optimally accurately and explicitly. Transliteration means to represent the characters of source language by the characters of target language, although keeping the action reversible. The main target of transliteration is to perpetuate the linguistic structure of the words. Appropriate transliteration of name entities play integral role in upgrading the charactertics of machine translation. Language Transliteration plays a significant role in various research areas like machine transliteration (MT) and cross-language information retrieval (CLIR) processes. The design of transliteration model is such like the articulate structure of words should be conserved as closely as possible. Two contrasting languages are considered in this case, one is Punjabi and another is English. There is numerous machine transliteration models used for Transliteration. Each model has peculiar requirements for implementation. After studying number of work done by various researchers in the area, we have developed algorithm based on statistical machine transliteration (pattern matching) from Punjabi to English and the accuracy appeared to be approximately 95.82%.

Keywords: Rule based approach, Statistical Machine Transliteration approach (SMT), Transliteration, Natural Language Processing (NLP).

I. INTRODUCTION

The communication among the human beings is done by language. The world is the collection of diverse culture, societies and people that communicate with one another using different languages. Thus, we need a technology to cross these language barriers. Our main desideratum is to construct and build software that will not only understand, analyze and generate languages that individuals use naturally, but also able to address our computer same way as we are addressing another person, but this is not facile to reach. The language transliteration is a necessary area in natural language processing. The software converts words written in Punjabi text to English text using Hybrid Approach of Machine Transliteration. The most intermittent problem with translators is to translate the proper names and technical terms in user input. The major defy is the transliteration of out of vocabulary (OOV) words appearing in the user input Manikrao Dhore, Kumar Shantanu (2012). The number of characters in, both English and Punjabi character sets varies in both the language that makes the transliteration process difficult. Punjabi Language is written from left to right using Gurumukhi script and it consists of consonants, vowels, halant , punctuation and numerals. Punjabi being an official language of Punjab can be easily understood or read by the person who knows Punjabi. Opposite to it, English is an international Language. The people who are not familiar with Punjabi can convert the file Written in Punjabi into English using Punjabi to English transliteration system.

Transliteration is sometimes muddled with Translation but both are different. Transliteration is the changeover of a word from one language to another without falling in its phonological characteristics and spelling equivalents of another language, particularly used to translate proper names and technical terms from languages while Translation is the action of interpretation of the meaning of a text i.e. a process that communicates the same message in another language. Therefore, if we need to read text in another language, and are more engrossed in pronouncing it than understanding it, we need transliteration. However, if we want to know what exactly it means, we need translation. For e.g.:- In translation "ਨਤੀਜਾ" will be translated into "result". Whereas in transliteration; "ਨਤੀਜਾ" will be transliterated into "ntija". Machine transliteration can be implemented by two ways. Transliteration of word from the origin language to foreign language is called Forward Transliteration while Transliteration of word from the foreign language back to the language of its origin is called Backward Transliteration.

Existing Approaches of MT

Accuracy and speed are the two main measures to evaluate the performance of machine transliteration tools. However, various approaches to machine transliterations have been proposed and each of this approach has its major benefits and drawbacks. The various approaches are:

The Direct Approach

The direct approach to machine transliteration is considered to be the most primordial or the primary approach of all, carrying out replacement of the words in the source language with words in the target language , carried out in particular sequence without much linguistic analysis and processing. The major substitute used in direct approach is a bilingual dictionary and that's why it is named as dictionary-driven machine transliteration Venkateswara Prasad T et al (2013).This approach is a unidirectional approach and access exclusive language pair at a time. The main process of transliteration in this approach is done by means of DBMT. It carries out word by word transliteration with the help of a bilingual dictionary usually followed by some syntactic rearrangement and re-ordering. DBMT system works in two junctures. In first step it starts with morphological analysis which removes morphological inflections from the words to get the target word from the source language words. The second step is the bilingual dictionary lookup in which a bilingual dictionary is looked up to get the target language words corresponding to the source-language words Amarpreet Kaur et al (2014).

1.1. The Rule-Based Approach (RBMT)

Rule-based approach involves the utilization of morphological, syntactical and semantic rules in the analysis of the source language text and the synthesis of the target-language text. In this transliteration system, a database of transliteration rules is used to transliterate the text from source to destination language Nakul Sharma (2011).

At any time, a sentence matches one of the rules or examples, it is translated directly using a bilingual dictionary. It uses contrastive knowledge of two languages. RBMT approach can be further branched into transfer-based and Interlingua machine transliteration respectively. Transfer based systems are more flexible and it can be elongated to language pairs in a multilingual environment. The Interlingua based systems can be used for multilingual transliteration **Kamaljeet et al (2010).**

1.2. The Corpus -Based Approach

The essential goal of CB approach is to analyze the systematic patterns of variation and to use pre-defined linguistic features. The corpus-based approach is classified into two approaches: - Statistical Machine Transliteration (SMT) and Example-Based Machine Transliterations (EBMT). The first approach, SMT is a MT paradigm in which the transliterations are achieved on the basis of statistical models. These statistical models specifications are derived from the analysis of bilingual text corpora Latha et al (2012). Statistical-based MT use entirely statistical based methods to alienate the words and to generate the texts. SMT is based on the aspect that every sentence in a language has a possible transliteration in another language. This approach requires minimal human effort and can be created for any language pair that has enough training data. The second approach, EBMT use previous transliteration examples to generate transliterations for an input provided by user. When an input sentence is presented to the system, it retrieves a similar source sentence from the example-based and its transliteration Vishal et al (2010)

1.3. Knowledge-Based Approach

The approaches require huge knowledge base that includes both ontological and lexical knowledge. Knowledge Based(KB) approach uses vast semantic and pragmatic knowledge and has the ability to reason about concepts which are highly modular and multilingual ,difficult to produce a universal representation of all languages and is only capable of demonstrating prototypes Nithya et al (2013).

II. LITERATURE SURVEY

Haque et al. (2009) developed English to Hindi Transliteration system based on the phrase-based statistical method. PB-SMT models have been used for transliteration by translating characters rather than words as in character-level translation systems. They have modelled translation in PB-SMT as a decision process, in which the translation a source sentence is chosen to maximize. They used source context modelling into the state-of-the-art log-linear PB-SMT for the English—Hindi transliteration task. An improvement of 43.44% and 26.42% has been reported respectively for standard and larger datasets.

Jia et al. (2009) developed Noisy Channel Model for Grapheme-based Machine Transliteration. They have experimented this model on English-Chinese. Both English-Chinese forward transliteration and back transliteration has been studied. The process has been divided into four steps: language model building, transliteration model training, weight tuning, and decoding. In transliteration model training step, the alignment heuristic has been grown diag-final, while other parameters have default settings. When decoding, the parameter distortion-limit has been set to 0, meaning that no reordering operation is needed. Lehal et al. (2010) developed Shahmukhi to Gurmukhi transliteration system based on corpus approach. In this system, first of all script mappings has been done in which mapping of Simple Consonants, Aspirated Consonants (AC), Vowels, other Diacritical Marks or Symbols are done. This system has been virtually divided into two phases. The first phase performs pre-processing and rule-based transliteration tasks and the second phase performs the task of post-processing. Bi-gram language model has been used in which the bi-gram queue of Gurmukhi tokens has been maintained with their respective unigram weights of occurrence. The Output Text Generator packs these tokens well with other input text which may include punctuation marks and embedded Roman text. The overall accuracy of system has been reported to be 91.37%. Malik et al. (2009) developed Punjabi Machine Transliteration (PMT) system which is rule-based. PMT has been used for the Shahmukhi to Gurmukhi transliteration system. Firstly, two scripts have been discussed and compared. Based on this comparison and analysis, character mappings between Shahmukhi and Gurmukhi scripts have been drawn an transliteration rules are formulated. The primary limitation of this system is that this system works only on input data which has been manually edited for missing vowels or diacritical marks which practically has limited use.

Sumita Rani et al.(2013) presented various techniques for transliteration from Punjabi language to Hindi Language. Most of the characters in Punjabi language have their same matching part present in a Hindi language. There are some characters exist in Hindi which is double sounds but no such characters are available for Punjabi. In this paper, transliteration system described is built on statistical techniques this system can be developed with minimum efforts. Knight et al. (2005) presented English-Japanese Transliteration system. This system is a phoneme based as they converted English word to English sounds and then into Japanese sound. Japanese frequently imports vocabulary from other languages, primarily from English. It has a special phonetic alphabet called Katakana, which is used primarily to write down foreign names and loanwords. In process of transliteration, first step is to generate scored word sequences. The idea behind is that ice cream should score higher than ice crème, which should score higher than ace Kareem. In the second step, converts English word sequences into English sound sequences. Gurpreet Singh Josan & Jagroop Kaur (2011) discussed that the fundamental activity of any MT application is to handle the vocabulary .Transliteration is the general choice for these words. In this paper, there are described our transliteration system based upon statistical techniques. This system can be developed with smallest amount efforts. There are many issues in machine transliteration left for further improvement and compared it with other potential algorithms.

Bhalla et al. (2013) have proposed rule based transliteration scheme for English to Punjabi. Some rules have constructed for syllabification. Syllabification is the process to extract or separate the syllable from the words. In this probabilities are calculated for name entities (proper names and location). For those words which do not come under the category of name entities, separate probabilities are being calculated by using relative frequency through a statistical machine translation toolkit known as MOSES. Using these probabilities the transliterating of input text from English to Punjabi is done. They have performed their experiment using Statistical Machine Translation tool. The average transliteration accuracy of the system is 92.75%

Rani and Luxmi (2013) have proposed Direct Machine Translation System from Punjabi to Hindi for Newspapers headlines Domain. The similarity between Punjabi and Hindi languages is due to their parent language Sanskrit. Punjabi and Hindi are closely related languages with lots of similarities in syntax and vocabulary. Ambiguity is major problem in machine transliteration. Direct transliteration does not provide any solution for ambiguity. Another methodology or rules are developed for such types of problems. From the accuracy analysis total number of accurate sentence are calculated and it has given accuracy of 97%.

Antony et al. (2010) have developed English to Kannada transliteration system that is based on statistical method. Their work addresses the problem of transliterating English to Kannada language using a publically available transliteration tool that is known as Statistical Machine Translation (SMT). The purpose of statistical transliteration method is to find the transliteration of source language word into target language with a specific probability. The word in the target language with the highest probability is chosen that indicates the best transliteration. The tool named MOSES is used for English to Kannada transliteration. It is a complete statistical machine transliteration toolkit. The other open source tolls named SRILM and GIZA++ are used for creating language and transliteration model. The proposed system achieved exact Kannada transliteration for 89.27% of English names.

Kumar and Kumar (2013) have presenting Statistical Machine Translation system to transliterate proper nouns written in Punjabi language into its equivalent English language. They have presented a statistical machine translation based approach to transliterate proper nouns of Gurumukhi script into its English equivalent names. The system is tested on various names belongs to various regions and overall accuracy of the system is very good. Accuracy of the system is depends on correctness of data stored into the database. The system can be furthered improved by adding unique names to the database. The proposed system is tested on more than 1000 names and system given as accuracy of 97%.

Das et al. (2009) have proposed transliteration system that is based on news corpus. They have trained the transliteration system using the English-Hindi datasets obtained from the NEWS 2009 Machine Transliteration Shared Task. English named entities are divided into transliteration units. The proposed system also considers the English and Hindi contextual information in order to calculate the probability of transliteration from each English unit to various Hindi candidate transliteration units and chooses the one with maximum probability. They have also devised some post processing rules to remove the errors.

II. METHODOLOGY USED

We have developed Punjabi → English transliteration system using hybrid approach (pattern matching) that benefits in transliterating Punjabi proper nouns into English proper nouns. For transliteration to be done a graphical user interface with the help of which input can be directly entered through keyboard. In our system, we have used the large parallel corpus which contains mapped words or proper Nouns. Source text has to be passed through various phases to get target text.

2.1 Preparation of database

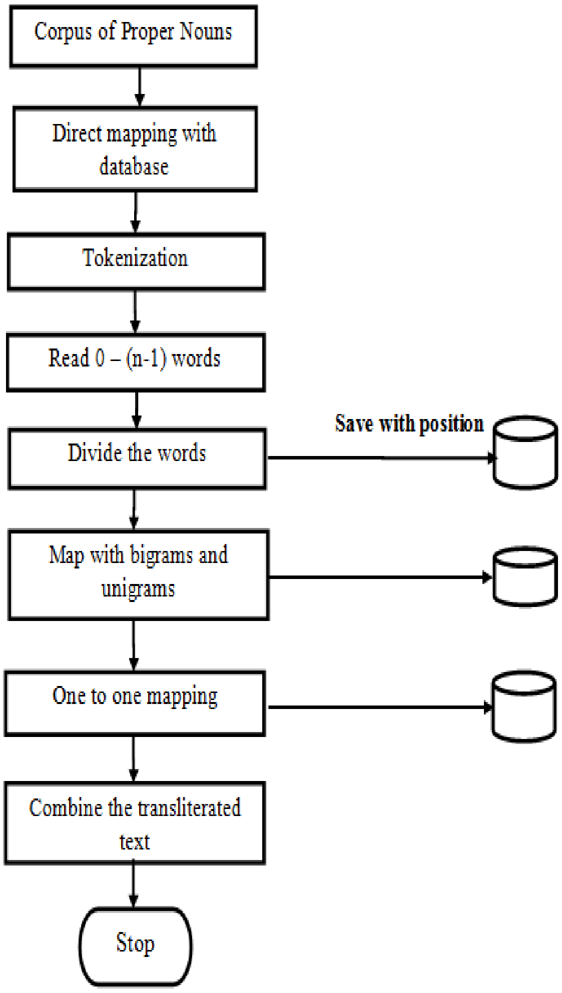
- **Collection of data:** Preparation of database includes corpus collection and setting of data in the form acceptable by the system. At the first step, we collected large data. The data is in the form of parallel corpus of English and Punjabi mapped words or proper nouns. The statistics of the corpus are as described here, 52,600 proper nouns from various sources were collected. The corpus contains person and place names from various online sites. And after removing identical names, around 46,000 names were left. These collected proper nouns are accurate. If the data in corpus is inaccurate, it causes the malfunctioning of the system. We made such large corpus as our system accuracy depends on the corpus collected. Large the size of corpus more will be the accuracy.
- **Normalization of data:** After collection of data, the corpus data was normalized so that it was brought to the same format to maintain equality. English text was converted to lower case letters. Special characters were removed from both English and Punjabi text and duplicate data was removed manually.
- **Alignment of data:** After the normalization, English and Punjabi text was aligned. Simultaneous checks were performed on parallel corpus to ensure the proper alignment of both Punjabi and English text.

2.2 Architecture of System

The system is developed by using Statistical transliteration approach rather than rule based approach. This is machine learning approach. In this system, there is a direct mapping of Punjabi words with equivalent English words. The architecture of the system contains two phases: Training and Testing.

2.2.1 Training Phase

During the training phase, system is trained by using the corpus. As we have collected a large parallel corpus of English and Punjabi proper nouns. So that corpus is used to train the system. The flow chart of training phase is as shown:



Flowchart1: Training Phase of System

In the training phase, initially we have a corpus of proper nouns. Any word which is inputted by user is first matched with corpus. If the word is matched then the output is directly transliterated to English equivalent word. But, if the results are not found in corpus then tokenization of words occurs. In tokenization, words are divided into tokens and then these tokens are read from start to end. After reading the words, we divide the words based on their position and also save those words along with their position in an array.

For E.g. RAMA is a word in English and after dividing it becomes RA and MA, so RA and MA are saved along with their transliteration part in database at their position i.e. RA and its transliteration at the 0th position and Ma and its transliteration at the 2nd position.

After positioning, we map the tokens with bigrams table without position. If the token pattern matched in table then it is saved along with their transliteration to the database. E.g. Sham is a word after dividing it becomes SH and AM, so here we save these bigrams without position with their transliteration in the database. And in the end if the results are not matched then finally there is one to one mapping of words i.e. words are mapped according to one to one mapping and then saved them in the database. So in this way the system learns from the parallel corpus and becomes trained.

2.2.2 Testing Phase

After the training phase there is testing phase. In the testing phase, the system performs the transliteration. The system takes Punjabi text as input and gives appropriate output. The procedure of this is as follows:

1. Input the Punjabi text to the system.
2. Divide the text into tokens and store the tokens in an array.
3. Extract the Punjabi word from an array.
4. Find that text in the database.

4.1 If matched, give its transliteration as output.

4.2 If not matched, otherwise further processing will be done

4.3 Find the length of the word.

5. Divide the word into 2 equal parts i.e. part1 and part2
6. Get all the possible transliterations from bigrams table
7. For part1:

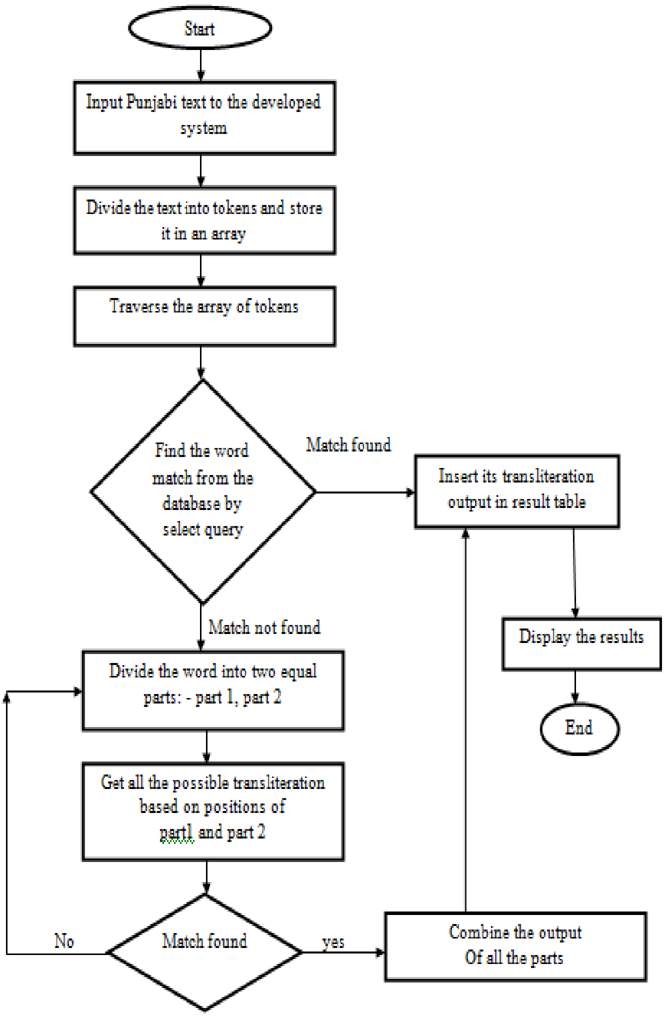
7.1 If the transliteration is present in database then get the most frequent transliteration as output of part1.

7.2 Else if further divide the part1 into tokens and match it with unigrams table and store the result in database.

7.3 Else go to step 3.

8. Similar procedure is followed for part2.
9. Combine the output for part1 and part2.
10. End

In this phase, initially the system takes an input Punjabi named entities. The inputted word is then firstly found in the database. If the word is present in the database then the system will return the transliterated word as an output. But in case if the match is not present in the database then further processing of the word will be done. In that case the segmentation of word is performed on the text. In segmentation, the word gets divided into 2 equal parts. After dividing, get all possible transliteration based on position of both parts. For the first part, if the transliteration of that part based on position is found then get the most frequently occurred transliteration of the part as output. Otherwise if the transliteration of the part is not found then we again divide the part into sub parts and perform the same procedure i.e. after sub dividing get the transliteration of both sub parts and if transliteration is found then gives the output otherwise again sub divided into parts. In this way, we get the transliteration of first part. Similar procedure is followed to get the transliteration of second part.


Start
Input Punjabi text to the developed system
Divide the text into tokens and store it in an array
Traverse the array of tokens
Find the word match from the database by select query
Match found
Insert its transliteration output in result table
Display the results
End
Match not found
Divide the word into two equal parts: - part 1, part 2
Get all the possible transliteration based on positions of part1 and part 2
No
Match found
yes
Combine the output Of all the parts

Flowchart2: Testing Phase of System

III. EXPERIMENTS AND RESULTS

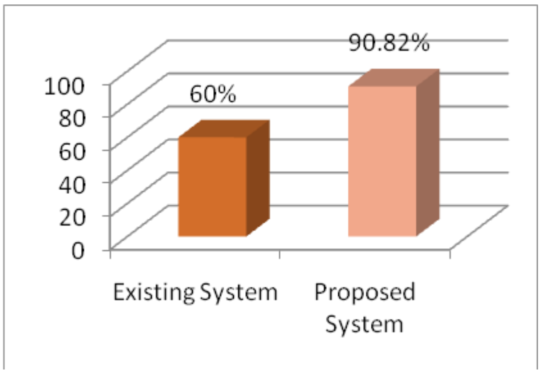
We have 24600+ entries in our database for Punjabi words and we have tested our software on 5000 Punjabi names. The given system has been tested to convert a doc file containing Punjabi names into its English equivalent. The system has option to upload doc file containing Punjabi names. It transliterates those names into English. System testing is done by taking names from magazines, telephone directories, newspaper, university records etc. This prospective system is accurate for the Punjabi words but not for the foreign words. Some names are not correctly translated by our system. These names are:

ਪਪਨਦੀਪ , ਮਣਨਾਲੀ , ਅੱਦਤਯ , ਇਕਲਵਯ , ਏਲਾ , ਯਤੇਂਦਰ , ਵਣਮਾਲੀ , ਪ੍ਰਚੂਰ , ਅਲੈਜੇਂਡਰ , ਅਲੈਕਸ , ਇਆਨ.

To calculate the performance of the system, transliteration accuracy rate is used. Accuracy Rate is calculated as the percentage of correct transliterations to the total generated transliterations by the system.

Accuracy Rate = (Correct Transliterations Generated / Total Transliterations Generated) * 100

The overall accuracy of our system is 95.69%.The subsequent graph shows the Comparison between Existing System and Proposed System.



Results generated by our system are:-

Punjabi Word	Rule Base Mapping	Statistical Approach
ਸਮਤਾਰਾਨੀ	samata rani	samta rani
ਨਰੇਸ਼ਕੁਮਾਰ	nereshkumar	naresh kumar
ਸ਼ਿਵਾਅਨੀ	shivaaani	shivani
ਬਾਵਲੀ ਕੌਰ	bavalikaur	bavli kaur
ਅਮਰਜੀਤ ਕੌਰ	amarjitkaur	amarjeet kaur
ਕੁਲਜੀਤ ਕੌਰ	kuljitkaur	kuljeet kaur
ਸੁਖਦੀਪ ਕੌਰ	sukhdipkaur	sukhdeep kaur
ਗੀਤਾ ਰਾਨੀ	gita rani	geeta rani
ਰਜਨੀ	rajnee	rajni
ਦੀਪਕ ਕੁਮਾਰ	dipak kumar	deepak kumar
ਪਿੰਕੀ	pinki	pinky
ਅਨੀਤਾ ਦੇਵੀ	anyta dewi	anita devi
ਨਵਰੀਤ ਕੌਰ	navrit kaur	navreet kaur
ਗਰੀਮਾ	gareema	garima
ਸ਼ਾਮ੍ਬੂਦਿਆਲ	shamboo dyal	shambhu dayal
ਰੋਮਿਕਾ	romica	Romika
ਇਕਬਾਲਸਿੰਘ	ikbal singh	iqbal singh

ਕਾਰਾਵੀਰ ਕੌਰ	caraveer kaur	karaveer kaur
ਜਸਵੀਰਕੌਰ	jasweer kaur	jasveer kaur
ਨਿਰਵਿੰਦਰ ਕੌਰ	niravnder kaur	nirwinder kaur
ਰੂਚਿਤਾ	roochita	ruchita
ਰਿੰਕੂਕੁਮਾਰ	rinkoo kumar	rinku kumar
ਮੰਜੂ ਰਾਨੀ	manjoo rani	manju rani
ਸ਼ੈਲੀ	shaily	shelly
ਗੌਰੀ	gowrie	gauri
ਮਨੇਤ ਰਾਮ	manget ram	manet ram
ਦਿਕਸ਼ਾ ਨਿਹਾਲਿਆ	diksha nihalaya	diksha nihalia
ਸ਼ੁਵਾਮ	shuam	shuvam
ਅਜੇ ਜਨਗਰਾ	ajey jagra	ajey Jangra
ਪਲਕਪ੍ਰੀਤ ਕੌਰ	pakapreet kaur	palakpreet kaur
ਬਨੀਤਾ ਦੇਵੀ	banita dewi	banita devi

IV. ERROR ANALYSIS

The overall performance and accuracy test of the system is quite good. But, there are several reasons for the miscues in the output.

4.1. Multiple Transliterations

In some case when any name is pronounced in Punjabi it corresponds too many English words, in that case system fails to interpret the correct word for transliteration.

4.2. Input wrong Words

Some time user does not get the expected results because of wrong input entered.

For e.g. **ਕਾਵ੍ਜ** in this word sometimes halant is used as such but we know it is used to write half

letter. The affect of this is shown below:

ਕ – Ka

ਕ੍ – K

4.3 Character Gap

This error usually arise due to character set variations between both English and Punjabi language, which makes the transliteration process very difficult. The numbers of vowels are 5 and 20 and numbers of consonants are 21 and 41, in both English and Punjabi language, so there is character gap in both the languages that leads to problems in transliteration process.

4.4. One-to-Multimapping Problem

This problem arises when a word in source language has multiple mappings in target language i.e. a single character in one script transform to multiple characters in another script. For e.g., ਵ character of Punjabi language can be mapped with two characters in English ‘v’ or ‘w’. Sly, ਫ character of Punjabi can be mapped to ‘ph’ or ‘f’. Thus to fix this problem an effective algorithm is required to select different mapped words at particular situations.

used to train the data. This system is giving promising results and can be further used by the researchers working on English and Punjabi Natural Language Processing tasks. Proper nouns from the State govt. English Documents, English Literature and other documents in English of one’s interest can be transliterated into Punjabi for use on the click on a button. Now, as future work, this approach is based on corpus as results always depends on training data thus, database can be improved by including more names to improve the accuracy.

V. CONCLUSION

In this paper a machine transliteration system is developed using statistical approach and further furthermore segmentation and mapping are

VI. REFERENCE

[1.] Abiola O.B, Adetunmbi A.O and Oguntimilehin. A(2015), “*A Review of the Various Approaches for Text to Text Machine Translations*”, International Journal of Computer Applications (0975 – 8887) Volume 120 – No.18.

[2.] Antony, P., Ajith, V. and Soman, K. (2010), “*Statistical Method for English to Kannada Transliteration*”, Communications in Computer and Information Science, Vol. 70, pp 356-362.

[3.] AbdulJaleel, N. and Larkey, L. (2003), “*Statistical transliteration for English-Arabic cross language information retrieval*”, Proceedings of the 12th international conference on information and knowledge management, pp: 139-146.

[4.] Bhalla, D. and Joshi, N. (2013), “*Rule Based Transliteration Scheme For English To Punjabi*”, International Journal on Natural Language Computing, Vol. 2, No. 2, pp. 67-73.

[5.] Devinder Kaur and Rishamjot Kaur(2015), "*English To Punjabi Script Converter System For Proper Nouns Using Hybrid Approach*", International Journal of scientific research and management (IJSRM), Volume 3, Issue 3, pp.2415-2420.

[6.] Dhore, M., Dixit, S. and Dhore, R. (2012), "*Hindi and Marathi to English NE Transliteration Tool using Phonology and Stress Analysis*", Proceedings of COLING 2012: Demonstration Papers, pp. 111-118.

[7.] Das A., Ekbal A., Mandal T. and Bandyopadhyay S. (2009), "*English to Hindi Machine Transliteration System at NEWS*", Proceedings of the 2009 Named Entities Workshop: Shared Task on Transliteration. Association for Computational Linguistics, pp. 80-83.

[8.] Debasis Mandal, D., Dandapat, S., Gupta, M., Banerjee, P. and Sarkar, S. (2007),"*Bengali and Hindi to English CLIR Evaluation*", Cross- Language Evaluation Forum CLEF, pp. 95-102.

[9.] Deep, K. and Goyal, V. (2011), "*Development of a Punjabi to English Transliteration System*", International Journal of Computer Science and Communication, Vol. 2, No. 2, pp. 521-526.

[10.] Finch, A. and Sumita, E. (2009), "*Phrase-based Machine Transliteration*", Proceedings of the 2009 Named Entities Workshop, pp. 52-56.

[11.] Gurpreet Singh Lehal and Tejinder Singh Saini, "*Development of a Complete Urdu-Hindi Transliteration System*", Proceedings of the COLING 2012: Posters, Mumbai, pp. 643-652.

[12.] Gurpreet Singh Lehal and Tejinder Singh Saini, "*Development of a Complete Urdu-Hindi Transliteration System*", Proceedings of the COLING 2012: Posters, Mumbai, pp. 643-652.

[13.] Gurpreet Singh Josan and Gurpreet Singh Lehal, "*A Punjabi to Hindi Machine Transliteration System*", Computational Linguistics and Chinese Language Processing Vol. 15, No. 2, June 2010, pp. 77-102.

[14.] Hermjakob, U., Knight, K. and Daume, H. (2008), "*Name Translation in Statistical Machine Translation Learning When to Transliterate*", Proceedings of Association for Computational Linguistics, pp. 389–397.

[15.] Haque, Dandapat, Srivastava, Naskar and Way(2009),"*English—Hindi Transliteration Using Context-Informed PBSMT: the DCU System for NEWS 2009*" ,Proceedings of the 2009 Named Entities Workshop, ACL-IJCNLP 2009, pages 104–107,Suntec, Singapore, 7 August 2009. ACL and AFNLP.

[16.] Hamadou, A. and Piton, O. (2011), "*Recognition and translation Arabic- French of Named Entities: case of the Sport places*", International conference series Finite-State Methods and Natural Language Processing, pp. 134-142

[17.] Jia, Zhu, and Yu, "*Noisy Channel Model for Grapheme-based Machine Transliteration*", Proceedings of the 2009 Named Entities Workshop, ACL-IJCNLP 2009, pages 88–91.

[18.] Joshi, H., Bhatt, A. and Patel. H. (2013), "*Transliterated Search using Syllabification Approach*", Forum for Information Retrieval Evaluation.

[19.] Josan, G. and Kaur, J. (2011), "*Punjabi To Hindi Statistical Machine Transliteration*", International Journal of Information Technology and Knowledge Management, Vol. 4, No. 2, pp. 459-463.

[20.] Jeong, K., Myaeng, S., Lee, J. and Choi, K. (1999), "*Automatic identification and back transliteration of foreign words for information retrieval*", In Proceedings of Information Processing and Management, Vol. 35, pp. 523-540.

[21.] Kumar, P. and Kumar, V. (2013), "*Statistical Machine Translation Based Punjabi to English Transliteration System for Proper Nouns*", in Engineering & Management, Vol. 2, Issue 8, pp. 318-321.

[22.] Knight, K. and Graehl, J. (1998), "*Machine transliteration*", Proceedings of the 35th annual meetings of the Association for Computational Linguistics, pp. 128-135.

[23.] Khantonthon, N., Kawtraku, A. and Poovarawan, Y. (2000), "*An Enhancement of Thai Text Retrieval Efficiency by Automatic Backward Transliteration*", WAINS 7: E-Business for the new Millennium, Bangkok, Thailand.

[24.] Kang, B. and Choi, K. (2000), "*Automatic Transliteration and Back transliteration by Decision Tree Learning*", In Proceedings of the 2nd International Conference on Language Resources and Evaluation, Athens, Greece.

[25.] Kamal Deep, Dr.Vishal Goyal, 2011, "*Development of a Punjabi to English Transliteration System*", International Journal of Computer Science and Communication, Vol. 2, No. 2, July-December 2011, pp. 521-526.

[26.] Knight, Graehl (2005), "*English-Japanese Transliteration system*", Computational Linguistics, Volume 24, Number 4, pp.599-612.

[27.] Lee, J. and Chang, S. (2003), "*Acquisition of English-Chinese Transliterated Word Pairs from Parallel-Aligned Texts using a Statistical Machine Transliteration Model*", Proceedings of the HLT-NAACL 2003 Workshop on Building and using parallel texts: data driven machine translation and beyond, Vol. 3, pp. 96-103.

[28.] Lee, J. and Choi, K. (1998), "*English to Korean Statistical transliteration for information retrieval*", Computer Processing of Oriental Languages, pp. 17-37.

[29.] Malik, "Punjabi Machine Transliteration System", In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL (2006), pp. 1137-1144.

[30.] Oh, J. and Choi, K. (2002), "*An English-Korean Transliteration Model Using Pronunciation and Contextual Rules*", Proceedings of the 19th international conference on Computational linguistics, Vol. 1, pp. 1-7.

[31.] Paul, M., Finch, A. and Sumita, E. (2009), "*Model Adaptation and Transliteration for Spanish-English SMT*", Proceedings of the 4th EACL Workshop on Statistical Machine Translation, pp. 105-109.

[32.] International Journal of Application or Innovation Rani, S. and Luxmi, V. (2013), "*Direct Machine Translation System from Punjabi to Hindi for Newspapers headlines Domain*", International Journal Of Computers & Technology, Vol. 8, No. 3, pp. 908-912.

[33.] Rama T. and Gali K. (2009), "*Modeling Machine Transliteration as a Phrase Based Statistical Machine Translation Problem*", Proceedings of the 2009 Named Entities Workshop, ACL-IJCNLP, pp. 124-127.

[34.] Sumita Rani, Vijay Laxmi, "*A Review on Machine Transliteration of related language: Punjabi to Hindi*" International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 3, March 2013.

[35.] Saini, T. and Lehal, G. (2008), "*Shahmukhi to Gurmukhi Transliteration System: A Corpus based Approach*", Advances in Natural Language Processing and Applications Research in Computing Science, pp. 151- 162.

[36.] Sharma S., Bora N. and Halder M. (2012), "*English-Hindi Transliteration using Statistical Machine Translation in different Notation*", International Conference on Computing and Control Engineering.

[37.] Stalls, B. and Knight K. (1998), "*Translating Names and Technical Terms in Arabic Text*", COLING ACL Workshop on Computational Approaches to Semitic Languages, pp. 34-41.

[38.] Verma, "*A Roman-Gurmukhi Transliteration system*", proceeding of the Department of Computer Science, Punjabi University, Patiala, 2006.

[39.] Vijaya, VP, Shivapratap and KP CEN, "*English to Tamil Transliteration using WEKA system*", International Journal of Recent Trends in Engineering, May 2009, Vol. 1, No. 1, pp: 498- 500.

[40.] Yan, Q., Grefenstette, G. and Evans, D. (2003), *"Automatic transliteration for Japanese-to-English text retrieval"*, Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 353-360.