ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 3, March 2013

A Review on Machine Transliteration of related languages: Punjabi to Hindi

Sumita Rani, Dr.Vijay laxmi

Abstract— Transliteration is the process of mapping source language into target language. The reverse process is known as back transliteration. It takes a character string in a source language and generates the same character string in the target language. Punjabi and Hindi are the languages which are originated from Sanskrit language (which is one of the oldest languages) so Hindi and Punjabi are most closely related languages so there are lots of similarities in syntax and vocabulary in both languages. It is comparatively easy to develop than the system between very different language-pairs like Hindi and English or Punjabi and English. The fundamental activity in any machine translation application is to handle the vocabulary words. This paper presents a new approach to improve Punjabi to Hindi transliteration by combining a basic character to character mapping approach with Statistical Approach.

Index Terms— Transliteration, Punjabi, Hindi, Statistical Approach.

I. INTRODUCTION

The present system involves Punjabi as a source language and Hindi as a target language. Both languages are very closely related languages such as syntax and vocabulary similarities. The classification of Punjabi-Hindi language pair as Indo-Iranian languages. The both languages belong to the same family but they have lot of differences between them. Both languages are not mutually comprehensible. Mutual comprehensible of the languages depends on the factors like degree of phonetical, morphological, syntactical and lexical similarities. In written terms, Punjabi and Hindi are not mutually comprehensible but in spoken term, both are mutually comprehensible. This relation is reversible with the speakers of Punjabi more able to understand Hindi but reverse is very difficult. Punjabi is the mother tongue of the Indian state of Punjab. On the other hand, Hindi and English has been official languages of all over India. There is a great demand for transliteration of documentation from Punjabi to Hindi. Requirement for every NLP (Natural Language Processing) application is to deal with complexity of vocabulary words.

Ms. Sumita Rani, *C.S.E. Department, Guru Kashi University, Talwandi Sabo, Bathinda, India, Mobile No 9463897669.*

Dr. Vijay Laxmi*, C.S.E. Department, Guru Kashi University, Talwandi Sabo, Bathinda, India, Mobile No. 9417570894.*

II. USE OF MACHINE TRANSLITERATION IN RELATED LANGUAGES

The languages are to be closely related if the structure of the grammar of both languages is very close (almost same) and both the languages have same semantics and lexicon. Generally such languages are originated from the same source (Sanskrit language) and both are related to Indo Aryan family. The languages are not related according to their surface (exterior) similarity. They can be derived from a common parent language. If both languages found regular sound connection between them then they are related according to Linguists.

Punjabi and Hindi both are two Indo-Iranian languages so there is a great interaction of the history of Punjabi and Hindi language. Both have many common features which have closely related these languages. In written form, both languages are distinct but related and similar scripts. Most of the characters in source languages (Punjabi language) have their corresponding matching part present in a target language (Hindi language).There are some characters exist in Hindi which are double sounds त्र ,क्ष ,ज्ञ (tr, ksh, gya) but no such characters are available for Punjabi.

Punjabi and Hindi are structurally same. There are both languages have same types of Genders, Nouns, Numbers, Person Tense. There are many common words in both languages which express the same meaning in both languages.

III. DIRECT CHARACTER MAPPINGS

Direct mapping of Punjabi Consonants and Vowels into Hindi Consonants and Vowels is explained. It is the base of the transliteration process. It is also known as character to character mapping.

Gur muk	Dev anag	Gur muk	Dev anag	Gur muk	Dev anag
hi	ari	hi	ari	hi	ari
ੳ	-	ਾ	ा	ਕ	क
ਅ	अ	ਿ	ि	ਖ	ख
ੲ	-	ੀ	ी	ਗ	ग
ਆ	आ	ੁ	ु	ਘ	घ
ਇ	इ	ੂ	ू	ਙ	ङ
ਈ	ई	ੇ	े	ਚ	च
ਉ	उ	ੈ	ै	ਛ	छ
ਊ	ऊ	ੋ	ो	ਜ	ज
ਏ	ए	ੌ	ौ	ਝ	झ
ਐ	ऐ	-	ऋ	ਞ	ञ
ਓ	ओ	-	ॠ	-	ॡ
ਔ	औ	-	ॐ	ੰ	॰

Fig. 1

Gur muk	Dev anag	Gur muk	Dev anag	Gur muk	Dev anag	Gur muk	Dev anag
hi	ari	hi	ari	hi	ari	hi	ari
ਟ	ट	ੜ	ड़	ਲ	ल	ਗ਼	ग़
ਠ	ठ	ਪ	प	ਲ਼	ळ	ਜ਼	ज़
ਡ	ड	ਫ	फ	-	ऴ	ੜ	ड़
ਢ	ढ	ਬ	ब	ਵ	व	ਢ਼	ढ़
ਣ	ण	ਭ	भ	ਸ਼	श	ਫ਼	फ़
ਤ	त	ਮ	म	-	ष	ਯ਼	य़
ਥ	थ	ਯ	य	ਸ	स	ਰ਼	ऱ
ਦ	द	ਰ	र	ਹ	ह	ਸ਼	-
ਧ	ध	ਰ	ऱ	ਕ਼	क़	ਹ਼	ह़
ਨ	न	-	ऱ	ਖ਼	ख़	ਵ਼	व़

Fig. 2

IV. LITERATURE SURVEY

Hindi - Punjabi Machine Transliteration System (For Machine Translation System) is proposed by Vishal Goyal and Gurpreet Singh Lehal. Transliteration is a process that takes a character in a source language and generates it into target language. Some words are basically not translated rather are transliterated. In this paper, there has taken Hindi as source language and Punjabi as target language. Thus, Hindi words will be transliterated into Punjabi words. Hindi and Punjabi both are closely related languages and hence it is very easy to develop than the system between very different language pairs like Hindi and English or Punjabi and English. There are implemented many complex rules for transliteration between Hindi and Punjabi languages. [1]

A Punjabi to Hindi Machine Transliteration System is proposed by Gurpreet Singh Josan and Gurpreet Singh Lehal. [Computational Linguistics and Chinese Language Processing , June 2010]. Transliteration is the general choice for handle the vocabulary words in any MT application. Transliteration is the process of mapping source language into target language. This paper presents a new approach to improve Punjabi to Hindi transliteration by character to character mapping approach with rule based and Soundex based approach. A fully accurate

Transliteration system is not possible due to the missing corresponding letters in two scripts. It is possible to transliterate

across the scripts due to the basic sounds of the source language, the word in the source script may be pronounced differently in the target script. [2]

Punjabi to Hindi statistical machine transliteration is proposed by Gurpreet Singh Josan & Jagroop Kaur [International Journal of Information Technology and Knowledge Management, 2011]. 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. [3]

Hybrid Approach for Punjabi to English Transliteration System is proposed by Kamal Deep and Dr.Vishal Goyal.[International Journal of Computer Applications,2011]. Language transliteration is the vital area in natural language processing. Accurate transliteration plays an important role in the machine translation. The transliteration model must be based upon the phonetic structure of words that are very closely. There has developed hybrid (statistical +rules) approach based transliteration system. [4]

Evaluation of Direct Machine Translation System For Punjabi To Hindi is proposed by Gurpreet Singh Josan and Gurpreet Singh Lehal[IJCSI International Journal of Computer Science Issues, 2009]. The Direct MT system is based upon enlargement of syntactic similarities between more or fewer related natural languages. Hindi and Punjabi are closely related languages with lots of similarities in syntax and vocabulary. Our system achieved the accuracy of the translation. Word to word translation might also be a solution for language pair of Punjabi and Hindi [5]

Direct Approach for Machine Translation from Punjabi to Hindi is proposed by Gurpreet Singh Josan and Gurpreet Singh Lehal [CSI Journal of Computing, 2012]. The Direct MT system is based upon consumption of syntactic similarities between related natural languages. These similarities make Direct Translation methodology for Punjabi-Hindi language pair. In this paper there are discussed about the Punjabi –Hindi direct translation system. The accuracy of the translation justifies that the simple word to word translation along with n-gram based word sense disambiguation approach and rule based approach provides a high accuracy and simple solution for language pair of Punjabi and Hindi. [6]

Evaluation of Hindi to Punjabi Machine Translation System is proposed by Vishal Goyal and Gurpreet Singh Lehal[IJCSI International Journal of Computer Science Issues, 2009]. Punjabi University, Patiala, India has developed Punjabi to Hindi Machine translation system with high accuracy of about 92%.Thus, this paper presents the evaluation results of Hindi to Punjabi machine translation system. From the above analysis, it is concluded the overall accuracy of Hindi to Punjabi machine translation system is found to be 95.12%. The accuracy can be improved by improving and extending the multilingual dictionary. [7]

V. CONCLUSION

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 are double sounds त्र ,क्ष ,ज्ञ (tr, ksh, gya) but no such characters are available for Punjabi. The major inaccuracies in the transliteration are due to poor word selection. In this paper, there have described the transliteration system build on statistical techniques. This system can be developed with minimum efforts. There are many issues left for further improvement. The system itself could be improved. Comparison with another many potential algorithms is also on future schedule.

VI. REFERENCES

[1] Gurpreet Singh Josan, Jagroop Kaur, "Punjabi to Hindi Statistical Machine Transliteration," International Journal of Information Technology and Knowledge Management July-December 2011, Volume 4, No. 2, pp. 459-463.

[2] Vishal Goyal, Gurpreet Singh Lehal, "Hindi-Punjabi Machine Transliteration System (For Machine Translation System)".

[3] Kamal Deep and Dr.Vishal Goyal, "Hybrid Approach for Punjabi to English Transliteration System," International Journal of Computer Applications (0975 – 8887) Volume 28– No.1, August 2011.

[4] Gurpreet Singh Josan, Gurpreet Singh Lehal, "Evaluation of Direct Machine Translation System For Punjabi To Hindi," IJCSI International Journal of Computer Science Issues, 2009.

[5] Vishal Goyal and Gurpreet Singh Lehal, "Evaluation of Hindi to Punjabi Machine Translation System," IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1, 2009.

[6] Gurpreet Singh Josan and Gurpreet Singh Lehal, "Direct Approach for Machine Translation from Punjabi to Hind,i" [CSI Journal of Computing | Vol. 1 • No.1, 2012].

[7] 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.

Ms. Sumita Rani , I have received my B-Tech degree in computer Engineering from Yadavindra College of Engineering, Talwandi Sabo (Bathinda) in 2011 and pursuing M-Tech degree in computer science & Engineering from Guru Kashi University, Talwandi Sabo (Bathinda). My Research area is Natural Language Processing and Machine Translation.

Dr Vijay Laxmi received her B-Tech degree in computer science & Engineering from SLIET Longowal in 2003, and Ph.D degree in Computer Science & Engineering in year 2012. Her research area is Grid Computing & OCR. She has published 25 research papers in various National/International Conferences and Journals. At present, she is engaged in Guru Kashi University, Talwandi Sabo, and Punjab as Dy. Dear Research and an Associate Professor in Computer Science Engineering & Information Technology department.