

Unexplored aspects in the transliteration of ENGLISH to PUNJABI

Er.Sukhnandan Kaur¹, Ms.Rupinderdeep Kaur²

¹(Student (Pursuing M.Tech.-CSE), Swami Vivekanand Institute of Engineering & Technology, Banur, India)

²(Lecturer (CSED), Thapar University, Patiala, India)

ABSTRACT:

Transliteration is meant to preserve the sounds of the syllables in words. Transliteration is helpful in situations where one does not know the script of a language but knows to speak and understand the language nevertheless. The following attempt deals with the unexplored aspects of famous Indian language named as Punjabi. Professional people whose native language is not Punjabi but need to do some work in Punjabi then the transliteration helps them out. Most of the efficient Transliteration software systems do not provide much facilities apart from the basic that meet the requirements of multilingual people. So, to make any language's transliteration effective, we need to focus on all the approachable corners of that particular language.

Keywords: Abbreviations, Dual language, Numeral Transliteration.

1. Introduction:^[1]

Transliteration is used to phonetically translate proper names and technical terms especially from languages in Roman alphabets to languages in non-Roman alphabets such as from English to Korean, Japanese, and Chinese. Because transliterations are usually representative index terms for documents, proper handling of the transliterations is important for an effective information retrieval system. However, there are limitations on handling transliterations depending on dictionary lookup, because transliterations are usually not registered in the dictionary. For this

reason, many researchers have been trying to overcome the problem using machine transliteration.

Transliteration is the process of mapping a written word from a language-script pair to another language-script pair. For example, punjabi word ਪਰਮ and ਸ਼ੇਅਰ corresponds to English words param and share respectively.

1.1 Forward Transliteration:

Transliterating a word from the language of its origin to a foreign language is called Forward Transliteration.

As in the case of English to Punjabi transliteration.

Param -> ਪਰਮ

1.2 Backward Transliteration:

Transliterating a loan-word (is a word borrowed from one language and incorporated into another) written in a foreign language back to the language of its origin is called Backward Transliteration. For Example:

ਤਬਲਾ → TABLA

2. Transliteration by using AnmolLipi:

It tells the way how different letters of the English are mapped onto the Punjabi letters. There are number of softwares which are helpful in mapping Punjabi letters corresponds to the English. This is shown in TABLE1.

Table1. Transliteration

Aa ਅ, ਓ	Bb ਭ, ਬ	Cc ਛ, ਚ	Dd ਧ, ਦ
Ee ਓ, ਏ	Ff ਢ, ਫ	Gg ਘ, ਗ	Hh “ , ” , ਹ
Ii ੀ, ਿ	Jj ਝ, ਜ	Kk ਖ, ਕ	Ll ਲ, ਲ
Mm ਮ, ਮ	Nn ਨ, ਨ	Oo ਓ, ਓ	Pp ਫ, ਪ
Qq ਥ, ਤ	Rr ਰ, ਰ	Ss ਸ਼, ਸ	Tt ਠ, ਠ
u ਊ, ਊ	Vv ਝ, ਵ	Ww ਵਿ, ਵ	Xx ਯ, ਣ
Yy ਯ, ਯ	Zz ਗ਼, ਜ਼		
^ ਖ਼	& ਫ਼		~ ਝ

Transliteration takes place as the example given below. As one wants to type ਨਾਨਾ in Punjabi then he needs to type nwnw. As transliteration is case sensitive. So, great care of words in caps is to be taken while typing. For Example:

nwnw
ਨਾਨਾ

3. Some of the unexplored aspects:

Numerous softwares are providing the facility of transliteration, but still much of the work has to be done to make it more fruitful to the users. In this paper, we have discussed many such aspects under which research is needed to be done. Some of them are as follows:

3.1 Numeral Number Transliteration:

Numerals are symbols also representing numbers. In mathematics there are other

meanings and definitions of numbers, over the different stages of the history of science. The symbols used in various modern Indian scripts for the numbers from zero to nine are shown in Table2.

Table2. Numeral transliteration

੦	ਸਿਫਰ	sifar	0
੧	ਇੱਕ	ikk	1
੨	ਦੋ	do	2
੩	ਤਿੰਨ	tinn	3
੪	ਚਾਰ	chā	4
੫	ਪੰਜ	punj	5
੬	ਛੇ	che	6
੭	ਸੱਤ	satt	7
੮	ਅੱਠ	ath	8
੯	ਨੌਂ	naun	9
੧੦	ਦੱਸ	dass	10

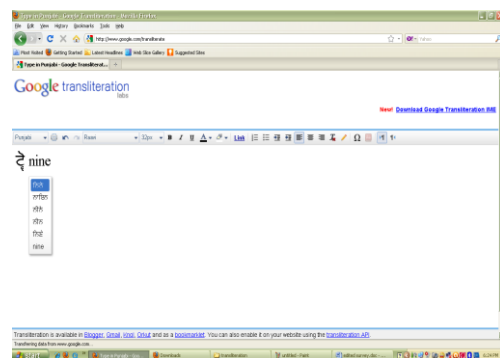


Fig.1. Google Transliteration

As we tried to type the numerals by using the current available software such as Google transliteration, then as in the above Fig.1, we typed nine and the possible transliterations were as in Fig.2.

nine

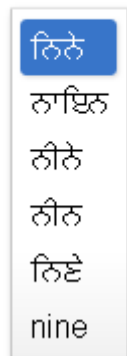


Fig.2 Transliteration of Nine

But as per our requirement we want to transliterate nine as :

੯ ਨੌਂ naun 9

So, our problem of the exact transliteration of the numerals are still there.

3.2 Abbreviation:

Simply stated, an abbreviation is any shortened form of a word or phrase. These are used to speed up the whole process of dealing with any language is enhanced by using abbreviations .In the Punjabi Transliteration, the facility for abbreviations is not provided which if provided then enhance its functionality .

Like:

P.U. – Panjabi University
ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ

If we type P.U. in any of the software which gives the facility of transliteration, it gives us as:

ਪ.ਉ.

In the above, the software does not take the letter with dot(P.), it takes P itself to give us the output. But functionality can be enhanced if it takes dot with any of the letter. By doing so we can add abbreviations in the database.

3.3 Name of Punjabi Months:

As every regional language contains its own data. People often use their own language for communication. Same happens with the

people who speak Punjabi language. Table 3 shows the names of Punjabi months.

Table 3. Name of Punjabi Months

English	Punjabi
March/April	ਚੇਤਰ
April/May	ਵੈਸਾਖ
May/June	ਜੇਠ
June/July	ਰਾੜ
July/August	ਸਾਵਣ
August/September	ਭਾਦਰੋ
September/October	ਅਸੂ
October/November	ਕਤਕ
November/December	ਮਘਰ
December/January	ਪੋਹ
January/February	ਮਾਘ
February/March	ਫਗਣ

When we checked it by typing june in google transliteration software window as in Fig.3:

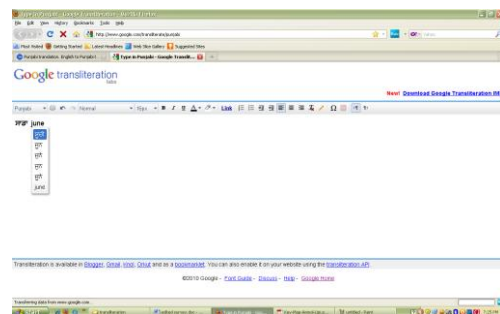


Fig.3 Months Transliteration by Google

we get the following as shown in Fig. 4 options :

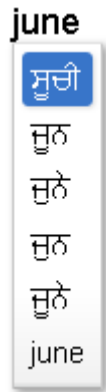


Fig.4 . Transliteration of June

But we want the transliteration of June as ਜੂਨ, then it is not possible with the existing software as shown in Fig.4.

3.4 Switching between different languages:

In linguistics, switching is the concurrent use of more than one language, or language variety, in conversation. Multilingual - people who speak more than one language - sometimes use elements of multiple languages in conversing with each other. Thus, we need to plant this facility.

As an quintessential, **ਅਦਾਜ਼**

This feature is tested on many softwares but all are unable to do this. As shown in Fig.5.



Fig.5 Transliteration By Tamil cube

If we type saada, it gives us the complete word in Punjabi as :

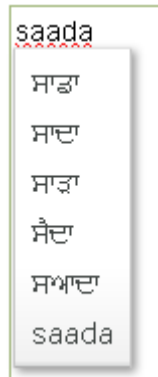


Fig.6. Transliteration of word Saada

but if we want to type the first letter in English and the rest in Punjabi. The above mentioned task can be done by a space between the first character and the rest. But doing so we have virtually two words instead of one word. From this, we conclude that the switching between different languages under one word is still not in light. The example of the above is given in Fig.6.

4.Difference between dictionary lookup and machine Transliteration:

As it is completely impossible to add all the words used in transliteration, so the alternative method of using machine transliteration apart from the Dictionary Lookup. These two softwares are different from each other from the manner as shown in Table4.

Both of these methods are widely used by the variety of softwares.

Table 4. Difference between Dictionary lookup and Machine Transliteration.

Dictionary Lookup	Machine transliteration
1.It is the direct method of transliteration of the whole word in one attempt.	1.In machine transliteration, each letter is mapped with the predefined set of corresponds.
2. It is fast and convenient to use.	2.It takes more time than that of dictionary lookup.
3.In this all the words are arranged in an ordered form.	3.Machine Transliteration does not follow any

	ordered form.
--	---------------

5.Conclusion:

To recapitulate all the above, it is quite vivid that not all features of Gurmukhi (particularly the older Gurmukhi used in the Bani) are represented yet. However, this is not a problem for modern Punjabi. This paper described a way to enhance web search for Romanized transliterations of Punjabi words. The evaluation of the of survey was done by taking ample examples with the Punjabi language. It was desirable to use some of the existing transliteration softwares . the above defined task able to take textual context into account, or at least be able to return a ranked list of possible English translations.

6.References:

1. Manoj Kumar Chinnakotla and Om P. Damani, Character Sequence Modeling for Transliteration. In the Proceedings of 7th International Conference on Natural Language Processing,(ICON) 2009.
2. Vijay John ,A Method for Enhancing Search Using Transliteration of Mandarin Chinese, Sai Ram Enggineering College Staff Journal ,2008-2009.
3. Jong-Hoon Oh , Key-Sun Choi, An ensemble of transliteration models for information retrieval, Information Processing and Management: an International Journal archive,Volume 42 Issue 4, July 2006
4. Shilei Zhang, Zhiwei Shuang and Yong Qin, Automatic, Pronunciation Transliteration for Chinese-English Mixed Language Keyword Spotting, Pattern Recognition (ICPR), 20th International Conference, 2010.
6. Kevin Knight, Jonathan Graehl, Machine Transliteration, Computational Linguistics, 1998.