

# Hybrid Approach for Translation of Common English Phrases to Punjabi

<sup>1</sup>Aman Aggarwal, <sup>2</sup>Rohit Sethi

<sup>1,2</sup>Dept. of CSE, Lovely Professional University, Phagwara, Punjab, INDIA

<sup>1</sup>Assistant Professor, Dept. of CSE, Lovely Professional University, Phagwara, Punjab, INDIA

## Abstract

This paper belongs to general conversion of one language phrases to another language phrases using translation techniques.

Transliteration is the process of converting alphabets in source language with their approximate phonetic or spelling equivalents alphabets in target language. In this case we have taken two different languages, one is Punjabi and another is English. We have done the work regarding the conversion of common English phrases to Punjabi language phrases. Also one more attempt made to convert Punjabi Language phrases to its corresponding English slangs with same pronunciation means whosoever reads the English written words or phrases the meanings and language will be spoken Punjab itself. So this thing makes a unique idea for those who just know English language but don't understand and know Punjabi language but still able to convey the message in Punjabi language without concern about the meaning and understanding to other people.

## Keywords

Translation, English to Punjabi, Punjabi Phrases, Transliteration.

## I. Introduction

Machine Transliteration is a way to convert character or words of one alphabetical system of source language to the corresponding character or words in another alphabetical system of target language [1]. Transliteration is a technique which takes a string of source language as input and generates the corresponding string in the target language as output. The source language in language transliteration is the original character of the language which the system will transliterate and the target language is the language in which the system will transliterate the original language. The process of Transliteration is a 2 step process:

First is the separating of the source language word into units and Second is relating the source language units with units in the target language, by resolving different combinations of alignments and unit mappings. For example, consider a Transliteration system for English to Punjabi. The Transliteration process will work as: first English word „Rajat“ will be divided into source language units as ‘r’ ‘a’ ‘j’ ‘a’ ‘t’ and then these units are transliterated into target language units ‘ਰ’, ‘ਐ’, ‘ਜ’, ‘ਐ’, ‘ਤ’ and finally these transliterated units results into target language word ‘ਰਜਤ’.

## A. Transliteration

Transliteration is the process of converting alphabets in source language with their approximate phonetic or spelling equivalents alphabets in target language [23].

Transliteration are of two types:

### 1. Forward Transliteration

Forward transliteration is the process of converting the source language into target language [4]. Source language S, is the original language of the alphabets and target language is the language in which the language is transliterated. Now suppose, Given a name

pair (S, T) then forward transliteration is converting the S to T. For example, given S is the English Language and T is the Punjabi Language then the English Language word S= "Aman" will be transliterated to T in Punjabi Language as "ਅਮਨ"

## 2. Backward Transliteration

Backward Transliteration is the process of converting the target language into source language [4]. Similarly as in forward transliteration, in backward transliteration also Source language S is the original language and Target language T is the language in which the system transliterate the language. For example, Given a name pair (S, T) where T is the Punjabi language and S is the English language. Then transliteration of Punjabi word "ਮਨਜੀਤ" in English word "manjeet" is the backward transliteration.

## II. Proposed Work

Translation is a process of converting the character or text of one language into another language. For translation either human translation or automated, the meaning of a text in the original language must be fully restored in the target language [21]. Translation is not a mere word-for-word substitution. A translator system must explain and examine all the elements in the text and know how each word may effect another and also requires large scale expertise in grammar, sentence structure, meanings, etc., in the source and target languages both.

## A. Translation

Translation is a way to convert the alphabet or text of one language into another language with same meaning[19]. For example, sentence in Punjabi language "ਮੈਂ ਓਕ" will be translated into English language as "I am ok" and in Indonesian language as "Saya ok". Basically Translation process is a pair of two language: One is source language, in which the user is the input and second is target language in which he the output. Like, English to Spanish, Punjabi to English.

Translation process are of two types:

### 1. Rule-Based Translation

Rules-based systems for translation is a 2 step process, the system will first uses the dictionary of common words and then apply language and grammar rules [19]. To focus on certain production or rule, the system requires special dictionaries. Rules-based systems gives standard translations of one language to another with specialist dictionaries.

### 2. Statistical Translation

Statistical Translation system "learn" to covert the one language word to another word by interpreting large amounts of data for each language i.e. for the language in which the user is giving input and the language in which the user gets its output [19]. Statistical translation systems have no knowledge of language or grammar rules. The statistical system can be trained for specific production or rule using additional data relevant to the sector needed.

## (I). Punjabi Language

Punjabi Language is written in Gurumukhi Script. Gurumukhi means from the mouth of Guru (Spiritual Leader). Gurumukhi script was derived from the Sharada script and standardized by Guru Angad Dev ji (Second Sikh Guru) in the 16th century. There are 35 distinct letters in Gurumukhi (or Punjabi) alphabet. In addition to Gurumukhi alphabets, there are six consonants created by placing a dot (bindi) at the foot of the consonant and there are nine dependent vowel signs used to create the independent vowels with three bearer characters [3].

Our designed system works to take every Punjabi alphabet from the table and converts it into corresponding Unicode character value. Unicode is a computing industrial standard used to express the world's different language character sets over the internet and software systems. All the world's famous languages are coded into Unicode character set values. Unicode uses different encoding schemes mainly which are popular and used are UTF-8, UTF-16 and ASCII encoding techniques, all of these schemes are having the same coding values which is of one byte size of data.

Table 1: Gurumukhi Alphabets [3]

ੳ	ੳੳ	Oo'rhaa	R in Root, Rough
ਅ	ਅੳ	Ai'rhaa	H in Hard, harry
ੲ	ੲੳ	Ee'rhee	E in Engine, Easy
ਸ	ਸੳ	sas'saa	S in Some, Sad
ਹ	ਹੳ	haa'haa	H in Holly, Horror
ਕ	ਕੳ	Kak'kaa	K in King, Kingdom
ਖ	ਖੳ	khakh'khaa	'Kh' in Khanna, Khurana
ਗ	ਗੳ	gag'gaa	G in Good, Gold
ਘ	ਘੳ	ghag'ghaa	'gh' in Ghuman, Ghostier
ਙ	ਙੳ	Ngan'ngaa	'ng' in England, English
ਚ	ਚੳ	chach'chaa	'Ch' in Charity, Chunk
ਛ	ਛੳ	chhachh'chhaa	ch' in cheddar, Chopper
ਜ	ਜੳ	jaj'jaa	J in Jurassic, Joy
ਝ	ਝੳ	jhaj'jhaa	harder stressed 'j' as in jolt
ਞ	ਞੳ	Njan'njaa	There is no English equivalent of this Sound
ਟ	ਟੳ	tain'kaa	T in Tomato, Tiger
ਠ	ਠੳ	thath'thaa	'th' in Thames, Thomas
ਡ	ਡੳ	ddad'daa	D in Dirty, Drill
ਢ	ਢੳ	dhad'daa	dd in Daddy, Address
ਣ	ਣੳ	nhaa'nhaa	English 'n' with tip of tongue upwards touching palette of mouth
ਤ	ਤੳ	tat'taa	like t in Tim, Tie
ਥ	ਥੳ	thath'thaa	'th' as in Thailand
ਦ	ਦੳ	dad'daa	Softer th' like in Then, Thought
ਧ	ਧੳ	dhad'daa	Stressed 'th' like in s Seventh, Fourth

ਨ	ਨੳ	nan'naa	N in Never, Night
ਪ	ਪੳ	pap'paa	P in Potato, Personal
ਫ	ਫੳ	phaph'phaa	ph in Philosophy, Philips
ਬ	ਬੳ	bab'baa	B in Baby, Borrow
ਭ	ਭੳ	bhab'baa	Bha combined together like Bharat, Bhakti
ਮ	ਮੳ	mam'maa	M in Monkey, Mouse
ਯ	ਯੳ	yay'yaa	Y in Yield, Yard
ਰ	ਰੳ	ra'raa	R in Risk, Right
ਲ	ਲੳ	lal'laa	L in Loan, Light
ਵ	ਵੳ	vav'vaa	V in Victory, Van
ੜ	ੜੳ	rhar'raa	hard rr like in arrow, arrange
ਸ਼	ਸ਼ੳ	shash'shaa	Sh in Shade, Shade
ਖ਼	ਖ਼ੳ	kha'khaa	Kh in Laddakh, Charkha
ਗ਼	ਗ਼ੳ	gag'gaa	G in Gold, Gun
ਜ਼	ਜ਼ੳ	Zaz'zaa	Z in Zebra, Zippy
ਫ਼	ਫ਼ੳ	faf'faa	Ph in philosopher
ਲ਼	ਲ਼ੳ	lal'laa	L in Land, Lead

English Language: English Language is written in Roman script. English language arose in the Anglo-Saxon kingdoms of England. It is one of six official languages of the United Nations. In India English is the second spoken language. There are 26 letters in English. Out of which 21 are consonants [B,C,D,F,G,H,J,K,L,M,N,P,Q,R,S,T,V,W,X,Y,Z] and 5 are Vowels [A,E,I,O,U]. [4]

Table 2: English Alphabets

English Alphabet	English Sound	Pronunciation Example
A	[a]	Arrow, Apple
B	[b]	Borrow, Boy
C	[θ],[k]	Crow, Cat
D	[d]	Draft, Dog
E	[e],[ε]	English, Elephant
F	[f]	Future, Fortune
G	[g]	Giraffe, Gold
H	[h]	Humble, Hen
I	[i]	India, Ice
J	[j]	Jump, Jug
K	[k]	King, Kite
L	[l]	Love, Light
M	[m]	Monkey, Mouse
N	[n]	Naughty, Nest
O	[o],[ɔ]	Operation, Oracle
P	[p]	Pronunciation, Peacock
Q	[k]	Quarrel, Queen
R	[r],[ɹ]	Russia, Rabbit
S	[s]	Skip, Snake
T	[t]	Tomato, Tiger
U	[u]	Ultimate, Umbrella
W	[w]	Wheel, Word
V	[v]	Vehicle, Van
X	[ks]	Xerox, Xenon
Y	[y]	Year, Yak
Z	[z]	Zebra, Zippy

The paper title “hybrid approach for translation of common English phrases to Punjabi” will work for translation of common English phrases to Punjabi. This paper will be helpful for two types of user:

1. Suppose, a user see a sentence written in Gurumukhi letter” ਮ ਤ ਨ ” and did not know about the Gurumukhi letter. He know about the English alphabets, so it will transliterate this sentence into English “mai tuhade naal awanga”. The transliterated sentence in English and sentence written in gurumukhi letter both will have the same pronunciation. So, transliteration will help the user to read common Punjabi phrases written in Gurumukhi letter.
2. Suppose, the same user now want to know the meaning of this sentence. The second phase of the proposal system will help user. It will translate the transliterate sentence from English sentence. The original sentence would be “I will come with you” which will be translated to Punjabi sentence and further the Punjabi based sentence written into English alphabets.

So, the proposed system will help the user in two ways: first, it will make the user to read the common phrases written in Gurumukhi by transliterate them and then will also tell the meaning of the sentence by translating the common phrases into English.

### III. Research Implementation to be done in Advance Java

The research work is to be implemented in Advance Java language using JSP and Servlets. Java originally developed by Sun Microsystems Company but now overtaken by ORACLE company recently. The first version of java (java 1.0) was introduced on 1995 .The latest version of java released is java 1.7. Java is a platform which accepts the program written in other language and convert it into java programming language and runs that language on java virtual machine. Its main feature is that java writes the program once and hence runs that program on multiple operating system.

Java is platform independent and open source language. It supports huge number of readymade code so it is easy to implement any kind of complex research work. As per huge calculations are required in implementation of security algorithms so there is required high class APIs which are available in JAVA language. Results Analysis is complex in java , and take more time for result's analysis.

### IV. Research Analysis and Results

In this research the analysis to be taken out for time consumption for conversion for English to Punjabi and as well for English to Punjabi based transliteration words.

The sentence is to be taken “who are you?”

Output is taken: ਕੋਨ ਹੋ ਤੁਸੀ ।

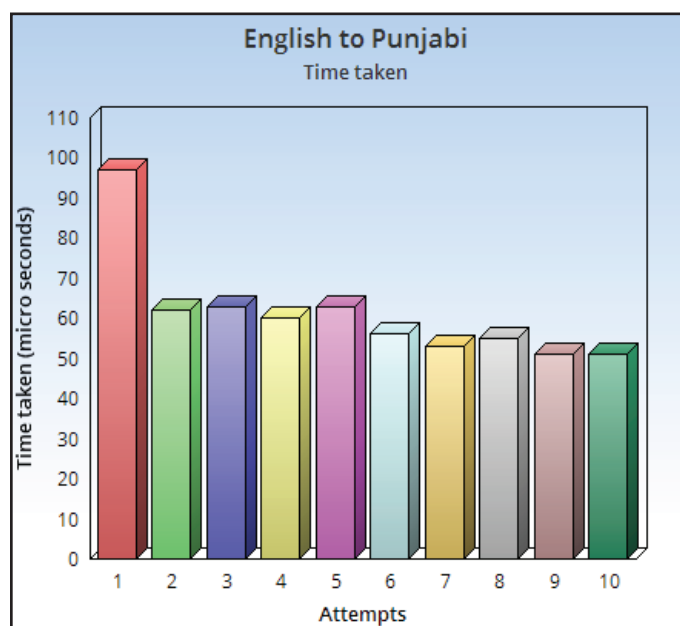
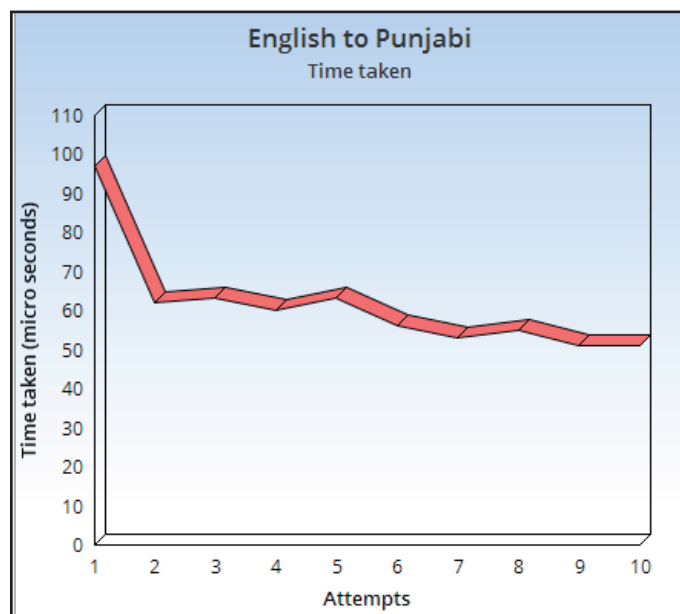
Hybrid Approach for Translation of Common English Phrases to Punjabi	
English Content	
who are you?	
Convert	
Punjabi Conversion	
ਕੋਨ ਹੋ ਤੁਸੀ ।	
Punjabi Based English Conversion	
Kam Ho Tusi ।	
English to Punjabi Time Taken: 51 (micro secs)	English to Punjabi English Time Taken: 30 (micro secs)
Total Time Taken: 82 (micro secs)	
Developed By: Annu Aggarwal	
Institute: Lovely Professional University	
Reg No: 10906136	
Thesis Title: Hybrid Approach for Translation of Common English Phrases to Punjabi	

Table 3: Time Consumption in Micro Seconds

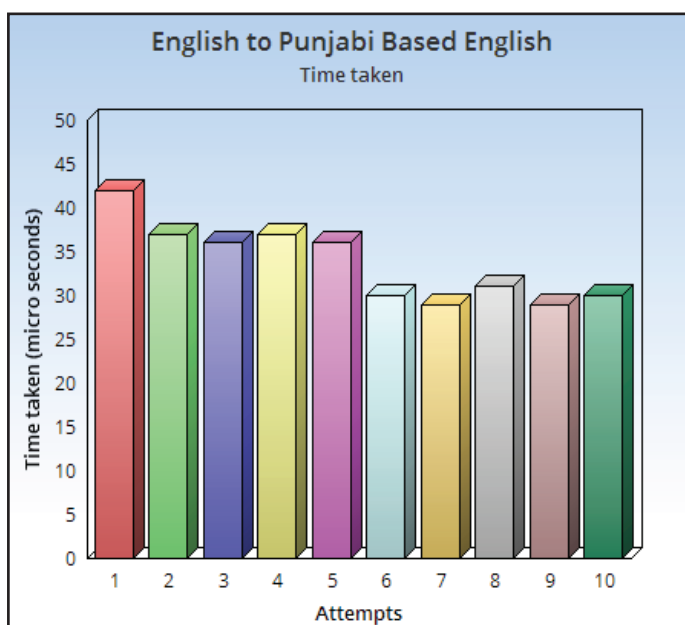
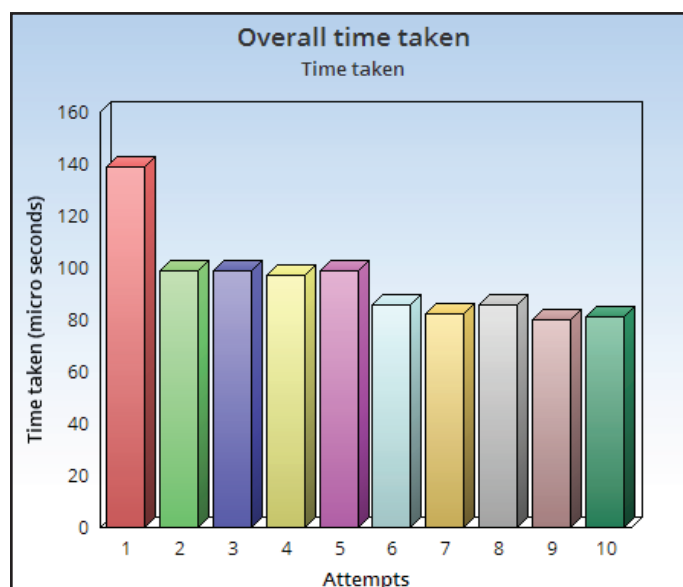
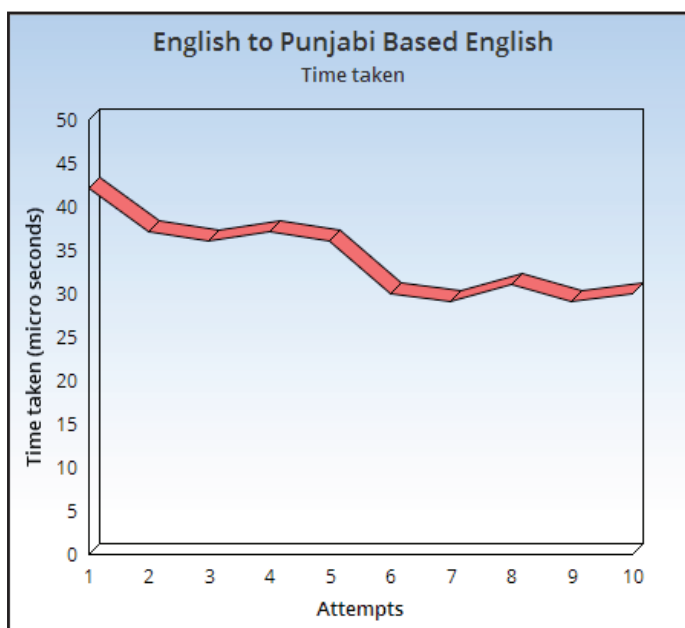
Punjabi conversion time taken	Punjabi based english conversion time taken	Total time taken
97	42	139
62	37	99
63	36	99
60	37	97
63	36	99
56	30	86
53	29	82
55	31	86
51	29	80
51	30	81

Graphical representation for the above time consumption of conversion of different categories.

#### A. For English to Punjabi Conversion



## B. For English to Punjabi Based English Conversion



## V. Findings

1. Punjabi is most useable language around the world related to one of the big religion of Sikhism.
2. Most people want to speak in Punjabi but don't know how to do due to not very much understanding and knowledge about the language.
3. System provides a platform where anybody can speak in Punjabi by just typing worldwide most useable English language based sentences and gets the corresponding sentences in English itself but contains Punjabi pronunciation and meaning with understanding.
4. Also system is faster because of the corresponding words and meanings are stored in two different architectures i.e. one is data structures using linked lists and other is database management system.

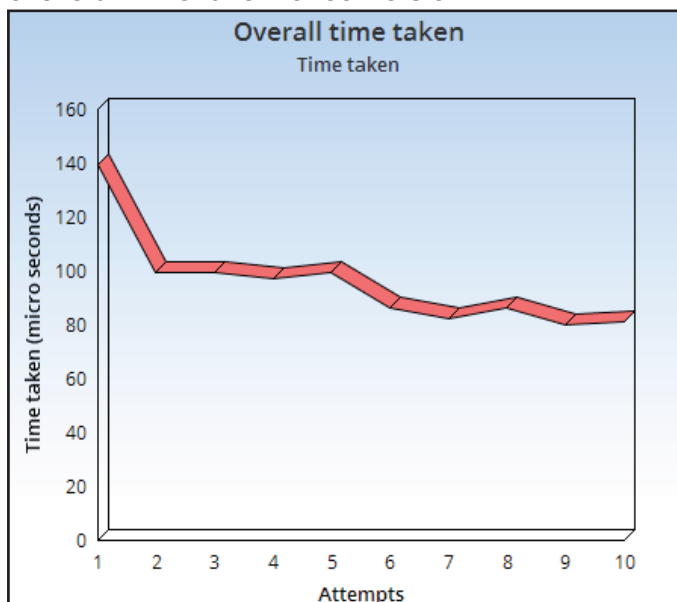
## VI. Conclusion

In this paper containing the proposed work of converting English general phrases to Punjabi based sentences makes the life easier for those who want to convey the message to those people who only knows Punjabi language but don't know other one. So the system makes it easy by translating the common English sentences to corresponding Punjabi based sentences and also to English sentences with full of Punjabi meaning and understanding. Using this system it does not matter whether the speaker know about Punjabi word meaning but the speaker does know the converted sentence has meaning originated from its original English sentence by which it was converted.

## References

- [1] Pankaj Kumar, Vinod Kumar, "Statistical Machine Translation Based Punjabi to English Transliteration System for Proper Nouns", Proceedings of Department of Computer Engineering, Guru Kashi University, Talwandi Sabo, 2013.
- [2] Antony P J, K P Soman, "Computational Morphology and Natural Language Parsing for Indian Languages: A Literature Survey", proceeding of Computational Engineering and Networking Research Centre, AMRITA Vishwa Vidyapeetham University, Coimbatore, 2012.
- [3] Jasleen Kaur, Gurpreet Singh Josan, "Statistical Approach to Transliteration from English to Punjabi", Proceedings of Department of Computer Engineering, Yadindra College

## C. Overall Time Taken for Conversion



- of Engineering, Talwandi Sabo, 2011.
- [4] Kamal Deep, Vishal Goyal, "Hybrid Approach for Punjabi to English Transliteration System", Proceedings of Department of Computer Science, Punjabi University, Patiala, 2011.
  - [5] Akshar Bharati, Samar Hussian, Meher Vijay, Kalyan Deepak, Dipti Mishra Sharma, Rajeev Sangal, "Proceedings of Language Technologies Research Centre", IIIT-Hydrabad, 2009.
  - [6] Jae Sung Lee, Key Sun Choi, "Proceedings of Department of Computer Science, Korean Advanced Institute of Science and Technology", 373-1 Kusong-dong, Yusong-gu, Taejon, 305-701, Korea.
  - [7] Wei Wang, Klaus Macherey, Wolfgang Macherey, Franz Och, Peng Xu, "Proceedings of Google Inc., 1600 Amphitheatre Pkwy, Mountain View, USA, 2012.
  - [8] Andreas Zollmann, Ashish Venugopal, Franz Och, Jay Ponte, "Licensed under the Creative Commons Attribution-Noncommercial- Share Alike 3.0 Unported license", 24, 2008.
  - [9] Thorsten Brants, Ashok C.Popat, Peng Xu, Franz J. Och, Jeffery Dean, "Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning", pp. 858-867, 2007.
  - [10] Wolfgang Macherey, Franz Josef Och, "Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natutal Language Learning", pp. 986-995, 2007.
  - [11] Shankar Kumar, Franz Och, Wolfgang Macherey, "Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natutal Language Learning", pp. 42-50, 2007.