Hybrid Approach for Translation of Common English Phrases to Punjabi

A DISSERTATION

SUBMITTED BY

AMAN AGRAWAL(10906136)

To

Department of Lovely Faculty of Technology and Sciences

In partial fulfillment of the Requirement for the

Award of the Degree of

Master of Technology In Computer Science And Engineering

Under the Guidance of

Mr. Rohit Sethi

(Asst. Professor, C.S.E)

ABSTRACT

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. This concept helps those people who knows and even in reality most of the people don't know more than two or three languages at a same time. But using this system and the system which we can design in future where we just enter the standard english phrases and by selecting the choice of language we get two different types of results in which one contain typical language output in according to selected choice of selected language and other which is most important in general english converted language based content.

ACKNOWLEDGEMENT

DECLARATION

I hereby declare that the dissertation proposal entitled, "**Hybrid Approach for Translation of Common English Phrases to Punjabi**" submitted for the M.Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Date: -

Aman Agrawal

10906136

CERTIFICATE

This is to certify that Aman agrawal (10906136) has completed M.tech dissertation proposal titled "Hybrid Approach for Translation of Common English Phrases to Punjabi" under my guidance and supervision. To the best of my knowledge , the present work is the result of his original investigation and study. No part of the dissertation proposal has ever been submitted for any other degree or diploma. Dissertation proposal is fit for the submission and the   partial   fulfillment of   the condition for the award M.tech computer Science and Engineering.

Date:

NAME : ROHIT SETHI

UID:

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my advisor and my mentor, Mr. Rohit Sethi (Asst. Professor, C.S.E), for his excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research. He always support me for doing research in this dissertation proposal. Whenever I face some problem my mentor help and guide me. I would also like to thank my parents and my brother. They were always supporting me and encouraging me with their best wishes.

AMAN AGRAWAL

10906136

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

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. It is a technique which takes a string or an alphabet of source language as input and generates the corresponding string or alphabet in the target language as output. The source language in language transliteration is the original character of the language 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 , it split the source language words into units and then it mapped these source language units with their corresponding units in the target language. 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' , 'ੳਾ', 'ਜ' , 'ੳਾ' , 'ਤ' and finally these transliterated units results into target language word 'ਰਜਤ'.

1.1 Transliteration:-Transliteration is a way of converting alphabets of a source language with their approximate phonetic or spelling equivalents alphabets in target language.Here, we focus on Punjabi language as the system will work for the transliteration of Punjabi phrases to English.

There is a simplified transliteration given by Wikipedia for the transliteration of Punjabi alphabets into English. They defined English alphabets corresponding to particular Punjabi alphabets for both vowels and consonants. The process of Transliteration faces some while transliterating the source language to target language. The main problem is the lack of letters present in one language and not in the other. For solving this problem a superset of letters from all the Indian Languages is formed. Even after creating a superset of letters from all Indian languages, it turns

out that for complex Aksharas combinations of the unique Roman letter are not easily identified. The large number of vowels in Punjabi scripts are 21 which lead to the complexity in transliteration.

The table 1 gives the Punjabi vowels, their transliteration defined by ISO 15919 and the simplified translietartion defined by google.

Table No. 1 (Gurumukhi vowels and their transliteration in English)

Gurmukhi	ISO 15919	Simplified
ਅ	A	A
ਆ	Ā	A
ਇ	I	I
ਈ	Ī	I
ਉ	U	U
ਊ	Ū	U
ਏ	Ē	E
ਐ	Ai	Ai

ੳ	Ō	O
ਐ	Au	Au

Similiarly, Table No. 2 gives the Punjabi Consonants, their transliteration defined by ISO 15919 and the simplified translietartion defined by google.

Table No. 2(Gurumukhi consonants and their transliteration in English)

Gurmukhi	ISO 15919	Simplified
ਕ	K	K
ਖ	Kh	Kh
ਗ	G	G
ਘ	Gh	Gh
ਙ	ṅ	N
ਚ	C	Ch

ਛ	Ch	Chh
ਜ	J	J
ਝ	Jh	Jh
ਞ	Ñ	N
ਟ	ṭ	T
ਠ	ṭh	Th
ਡ	ḍ	D
ਢ	ḍh	Dh
ਣ	ṇ	N
ਤ	T	T
ਥ	Th	Th

ਦ	D	D
ਧ	Dh	Dh
ਨ	N	N
ਣ	n̲	N
ਪ	P	P
ਫ	Ph	Ph
ਬ	B	B
ਭ	Bh	Bh
ਮ	M	M
ਯ	Y	Y
ਰ	R	R

ੜ	r̲	R
ਲ	L	L
ਲ਼	ḷ	L
ਵ	V	V
ਸ਼	Ś	Sh
ਸ	S	S
ਹ	H	H
ਕ਼	Q	Q
ਖ਼	k̲h̲	Kh
ਗ਼	Ġ	G
ਜ਼	Z	Z

ੜ	ṛ	R
ੜੑ	ṛh	Rh
ਢ	F	F
ਞ	ẏ	Y
ਟ	ṭ	T
ਹ	ḥ	H
ਵ	W	W

The process of transliteration is further divided into two parts based on the conversion of languages i.e source language to target language or target language to source language :

1. Forward Transliteration
2. Backward Transliteration

Forward Transliteration:- Forward transliteration is the process of converting the source language into target language. 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'

Backward Transliteration :- Backward Transliteration is the process of converting the target language into source language. 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.

Translation is a process of converting the character or text of one language into another language. The meaning of the text of source language must be fully stores in the target language for either human translation or automated translation. The process of Translation is not only mapping one word of the source language its corresponding word in the target language. A good and accurate translator system , it must explain and examine all the elements in the text and know how each word of the text with their effect on another word and also requires large scale expertise in grammar, sentence structure, meanings, etc., in the source and target languages both.

1.2 Translation :- Translation is a way to convert the alphabet or text of one language into another language with same meaning. For example, sentence in Punjabi language" ਮੈਂ ਠੀਕ ਹਾਂ " 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 gives the input and second is target language in which he get the output. Like, English to Spanish,here English is the source language in which user is giving the output and Spanish is the target language in which he/she expects the output. Similairly, In Punjabi to English Translation , Punjabi will be the source language and English will be the target language.

Translation process are of two types:

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. 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.

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. 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.

1.3 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.

Table No. 3: Gurumukhi alphabets

ਗੁਰਮੁਖੀ	ਪੰਜਾਬੀ	English	Pronunciation
ੳ	ਉੁੜਾ	Oo'rhaa	Root, Rough
ਅ	ਐੜਾ	Ai'rhaa	Hard, harry
ੲ	ਈੜੀ	Ee'rhee	Engine, Easy
ਸ	ਸੱਸਾ	sas'saa	Some, Sad
ਹ	ਹਾਹਾ	haa'haa	Holly, Horror
ਕ	ਕੱਕਾ	Kak'kaa	King, Kingdom
ਖ	ਖੱਖਾ	khakh'khaa	Khanna, Khurana
ਗ	ਗੱਗਾ	gag'gaa	Good, Gold
ਘ	ਘੱਘਾ	ghag'ghaa	'Ghuman, Ghostier
ਙ	ਙੰਙਾ	Ngan'ngaa	England, English
ਚ	ਚੱਚਾ	chach'chaa	Charity, Chunk
ਛ	ਛੱਛਾ	chhachh'chhaa	cheddar, Chopper
ਜ	ਜੱਜਾ	jaj'jaa	Jurassic, Joy
ਝ	ਝੱਝਾ	jhaj'jhaa	harder stressed 'j' as in jolt

ਞ	ਞੰਞਾ	Njan'njaa	There is no English equivalent of this Sound
ਟ	ਟੈਂਕਾ	tain'kaa	Tomato, Tiger
ਠ	ਠੱਠਾ	thath'thaa	Thames, Thomas
ਡ	ਡੱਡਾ	ddad'daa	Dirty, Drill
ਢ	ਢੱਢਾ	dhad'daa	Daddy, Address
ਣ	ਣਾਣਾ	nhaa'nhaa	English 'n' with tip of tongue upwards touching palette of mouth
ਤ	ਤੱਤਾ	tat'taa	in Tim, Tie
ਥ	ਥੱਥਾ	thath'thaa	Thailand
ਦ	ਦੱਦਾ	dad'daa	Softer th' like in Then, Thought
ਧ	ਧੱਧਾ	dhad'daa	Stressed 'th' like in s Seventh, Fourth
ਨ	ਨੱਨਾ	nan'naa	Never, Night
ਪ	ਪੱਪਾ	pap'paa	Potato, Personal

ਫ	ਫੱਫਾ	phaph'phaa	Philosophy, Philips
ਬ	ਬੱਬਾ	bab'baa	Baby, Borrow
ਭ	ਭੱਭਾ	bhab'baa	Bha combined together like Bharat, Bhakti
ਮ	ਮੱਮਾ	mam'maa	Monkey, Mouse
ਯ	ਯੱਯਾ	yay'yaa	Yield, Yard
ਰ	ਰਾਰਾ	ra'raa	Risk, Right
ਲ	ਲੱਲਾ	lal'laa	Loan, Light
ਵ	ਵੱਵਾ	vav'vaa	Victory, Van
ੜ	ੜਾੜਾ	rhar'rhaa	hard rr like in arrow, arrange
ਸ਼	ਸ਼ੱਸ਼ਾ	shash'shaa	Shade, Shade
ਖ਼	ਖ਼ੱਖ਼ਾ	kha'khaa	Laddakh, Charkha
ਗ਼	ਗ਼ੱਗ਼ਾ	gag'gaa	Gold, Gun
ਜ਼	ਜ਼ੱਜ਼ਾ	Zaz'zaa	Zebra, Zippy
ਫ਼	ਫ਼ੱਫ਼ਾ	faf'faa	Ph in philosopher
ਲ਼	ਲ਼ੱਲ਼ਾ	lal'laa	L in Land, Lead

1.4 English Language: Arose in the Anglo-Saxon kingdoms of Kingdom English Language is written in Roman script. 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].

Table No. 4: English alphabets table

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], [ɛ]	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	[ɾ], [r]	Russia, Rabbit
S	[s]	Skip, Snake
T	[t]	Tomato, Tiger
U	[u]	Ultimate, Umbrella
W	W	Wheel, Word
V	[b]	Vehicle, Van
X	[ks]	Xerox, Xenon
Y	Y	Year, Yak
Z	[z]	Zebra, Zippy

CHAPTER 2

SCOPE OF STUDY

The dissertation proposal "hybrid approach for translation of common Punjabi phrases to English" will work for translation of common Punjabi phrases to English. This proposal will be helpful for two types of user:

1. Suppose, a user see a sentence written in Gurumukhi letter" ਮੈ ਤੁਹਾਡੇ ਨਾਲ ਆਵਾਂਗਾ " 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 into English. The translated sentence will be "I will come with you".

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.

CHAPTER 3

OBJECTIVES

The objective of the proposed system is to translate the Common Punjabi Phrases into English. After studying research papers related to "Transliteration of Punjabi to English" and "Translation system", we proposed the system. To achieve the objective of proposed dissertation following steps are required :-

Step 1:- First step is to make the database for the proposed system. The database will have 2 dictionaries: First for some Common Punjabi Phrases and its corresponding transliteration in English and Second is for transliterated Common Punjabi Phrases and its corresponding translation in English.

Step 2:- Second step is to define Unicode of Gurumukhi letter and also some rules for transliteration of Common Punjabi Phrases.

Step 3:- Third step is the Transliteration of Common Punjabi Phrases into English and then translate them in English by applying some grammar rules

Step 4:- Fourth step is the testing of system outcomes and calculating the standard measures for language models to check the accuracy of the system.

CHAPTER 4

LITERATURE REVIEW

The process of Transliteration is very extensive and has been studied for several different language pairs, and many techniques have been proposed. Basically there are two techniques for transliteration: Grapheme based and Phoneme based. In Grapheme based technique transliteration is the mapping of grapheme sequence from a source language to target language ignoring the phoneme-level processes. While, in phoneme based approach transliteration they consider about the pronunciation or the source phoneme rather than spelling or the source grapheme.

Research papers related to **transliteration**:-

A Review of Transliteration System for Punjabi to English (Devinder et al. , 2014)- In this paper, author first discussed about the existing approaches for the Punjabi to English transliteration like, direct mapping approach , rule-based approach and statistical machine translation approach. They observed that the maximum accuracy of the existing system is 63% which needs further improvement. The main problem they observed for the transliteration of Punjabi to English is dealing with the multiple mapeed character. They also observed that most common faced problem in translating system is to translating the proper names and technical terms.There is difference between the translation and transliteration for language pairs that has different alphabets , such as Punjabi/Hindi. Phonetic translation across these pairs is called Transliteration. Transliteration helps the user who don't know a particular language but want to communicate in that language.So it convert the alphabets of a source language to the target language .To achieve this goal the process of transliteration has complex methods or say conventions for dealing with those alphabets which are present in the source language but don't have their correspond in the target language. Example, the number of consotants in Punjabi language are 40 while in English language these are only 21.Their is another method called Transcription for coverting the alphabets from one source language to target .The process of Transliteration is different from the process of   transcription.In transcription we mapped the sound of source language to alphabets of target language

while in Transliteration we mapped the alphabets of source language to alphabets of target language.

A Statistical Machine Transliteration Based Punjabi to English Transliteration System for Proper Nouns (Pankaj et al. , 2013)**-** In this paper, author proposed a statistical machine transliteration system for transliteration of Punjabi proper nouns to English.The Transliteration system works in two phases: Training phase and Transliteration phase. Training phase consist of five stages: Extract Bi-Gram, Extract Tri-Gram, Extract Four-gram, Extract Five-Gram, Extract Six-Gram. Transliteration phase consist of two stages: First, direct mapping of source word with database, Second , apply N-Gram approach of training phase.. In First phase i.e, System training phase , they construct a method which gives training to the system on the basis of data stored in the database. Here , they stored 15000 uniques names in the database , based on which the training is given to the system. And in Second phase i.e, transliteration phase , system first tries to find the word enter by the user directly from the database and if word is present in the database then system give its corresponding output but if the word is not present in the database then with the help of generated tables , system can transliterate the word into target language. The accuracy of the system is depends on the data stores in databases and has been tested with more than 1000 names and gives an accuracy of 97%.

A Rule Based Transliteration Scheme for English to Punjabi (Deepti et al. , 2013) -In this paper, author proposed a rule based transliteration scheme for English to Punjab. They construct some rules for syllabification. Syllabification is the process to extract or separate the syllable from the words. In this we are calculating the probabilities 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 we are transliterating our input text from English to Punjabi. **Algorithm for transliteration system is :**

1. Enter input string in English.
2. Identify Vowels and Consonants.
3. Identify Vowel-Consonant combination and consider it as one syllable.
4. Identify Consonants followed by Vowels and consider them as separate syllables.

5. Identify Vowels followed by two continuous Consonants as separate syllable.

6. Consider Vowel surrounded by two Consonants as separate syllable.

7. Transliterate each syllable into Punjabi.

A Survey of Transliteration System for Indian Languages(Antony et al. , 2012)- In their survey, they discuss about the natural language parsers and different developments of morphological analyzer and generator. From surveying the morphological analyzer, generator and parsers for Indian Languages they found that almost all approaches were applied in different morphological analyzer and generator. They also found that either statistical approach or hybrid approach is applied for almost all existing Indian Language parsers. The main effort and challenge behind each and every development is to design the system by considering the morphological rich features of language.They do survey on the following pairs of languages:

1. English to Hindi Machine Transliteration.

2. English to Tamil Language Machine Transliteration.

3. English to Kannada Language Machine Transliteration.

4. English to Malayalam Language Machine Transliteration.

5. English to Telugu Language Machine Transliteration.

6. English to Indian Language Machine Transliteration.

A Rule Based Transliteration System from Hindi to Urdu (Bushra et al. , 2012) - In this paper they have considered the problem of transliteration of proper nouns from Hindi to Urdu. In this paper, they overcome the shortcomings when proper nouns are transliterated using mappin.They observed some challenges in transliteration from Hindi to Urdu. First problem is Ambiguous Character means one alphabet in Hindi has multiple alphabet mapped in Urdu which creates a problem in transliteration. Second problem that they observed is Non existence of nukta Symbol in Hindi script. Third problem that they observed is Transliteration of Proper Nouns. As some Urdu words are written in different ways. Therefore we have tried to formulate some rules based on our observations of outputs of our experiments. For some very specialized spellings**.** However , they solved these problems by introducing mapping tables and

some rules. In the paper discussed by Bushra and Tafseer. Transliteration system was presented with a good accuracy.

Statistical Approach for Transliteration from English to Punjabi (Jasleen et al. , 2011)- In this paper, author presents a transliteration system from English to Punjabi based on Statistical approach using MOSES. They observed the following problems in transliteration from Punjabi to English or English to Punjabi . First Problem is the Character Gap between the languages. As we know the number of constants and the number of vowels in the Punjabi language are 41 and 20 respectively while in English language the number of constants and the number of vowels are 21 and 5 respectively. So, it means some alphabets in the Punjabi language does not have their coreesponding in the English language which makes difficult while we transliterate the words from Punjabi to English. For Example, for Punjabi alphabet 'ੜ' there is no

corresponding alphabet in English. Second problem is the One-to-Multi mapping Problem which means the a single alphabet in the source language may convert to more than one alphabets in target language. For example, for English alphabet 't' there are two corresponding alphabets in the Punjabi language 'ਟ' and 'ਤ' which

means the system can convert the alphabet 't' of English language to either of them which also create a problem in transliteration.Third problem that they observed is - Multi-to-One map problem which means that their may be a single alphabet in the target language for multiple alphabets in the source language or there may be a combination of alphabets in the target language for the single alphabet in the source language. For example, in following name, combination of two characters 'ch' in English language forms the single character in Punjabi language 'ਚ'. To solve the

problem they represent a Transliteration System which consists of three phases: Preprocessing, Transliteration unit and Post processing. They apply some Transliteration rules at phase 3 to improve the accuracy. They calculate the efficiency of the Transliteration system in two ways: first they manually calculate the efficiency as well as using Bilingual Evaluation Understudy metrics i.e. BLEU. The accuracy of the system has been tested with 3844 English names and 3844 Punjabi names. The system produced transliteration in English to Punjabi with an accuracy of 63.31%.

Hybrid Approach for Punjabi to English Transliteration System(Kamal et al. , 2011)-In this paper, author developed a hybrid approach for Punjabi to English Transliteration System. The approach is hybrid because it uses the 3 steps for the Transliteration. Three stages of the system are: Tokenization ,Rules and Direct Mapping and last English words with dictionary. Passing through these three stages language text has converted into target language. The system is divided into two test cases: first for person names and second for City names, State names and River names. The accuracy of the system was very low using direct mapping. The accuracy of test case 1 is 95% and the accuracy for test case 2 is 91.40%. The system gives overall 93.22% accuracy for Punjabi words not for foreign words.

English-korean named entity transliteration using statistical substring based and rule based approaches (Yu-chun et al. , 2011) In this paper, auhtor comprises three approaches for the English- Korean named entity transliteration. First approach is grapheme substring based , second approach is phoneme substring based and the third approach is rule based methods. The first approach i.e, substring based approach is further comprises three approaches.First they pre-process the string then the second step is Substring alignment and after that third step is CRF training. There standard and non-standard runs achieves 0.43 and 0.332 in top-1 accuracy which were ranked as the best for the English-Korean pair. In this paper, we adopt the substring-based transliteration approach with CRF model for English-Korean named entity transliteration. The characters in the source and target language are aligned in bi direction and then group into substrings to generate the substring mappings from the source language to the target language. Then, the transliteration is formulated as a sequential tagging problem to tag the substrings in the source language with the substrings in the target language. The CRF algorithm is used to deal with this tagging problem. For English substring generation, we create two types of substrings. One is based on the English orthography, and the other is based on the phonemic symbols from the CMU pronouncing dictionary. In addition, we also construct a rule-based transliteration system based on the Korean writing method of loanwords from the National Institute of Korean language. From the evaluation results, the substring-based method based on the English orthography performs better than other runs. For future work, we plan to add more phonetic features for the CRF training and try to

integrate the CRF-based statistical based method and the rule-based methods to improve the transliteration performance. We also try to apply the re-ranking techniques from the web data to get better transliteration results.

Constraint Based Hybrid Approach to Parsing Indian Languages (Akshar et al. , 2009)- In this paper , author describes a constraint based hybrid approach to parsing Indian languages. The hybrid approach is two stage constraint based for dependency parsing. There are many problem which the system faced by data driven parsing in the overall parsing of the system. In first stage only intra clausal dependencies are handled and later in the second stage the inter clausal dependencies are identified. The proposed parser by them is still being improved at various fronts. At first stage a pharse "mai hospital gayaa kyonki mai bimaar thaa" will be first translated word by word " 'I' 'hospital' 'went' 'because' 'I' 'sick' 'was' " .Later, after applying the grammer rules the phrase will be translated as   'I went hospital because I was sick'. They decribed their future scope by handling some unhandled constructions to enrich the verb frame and also by adding the prioritization mechanism The proposed parser is still being improved at various fronts..

 English to Korean Statistical Transliteration for Information Retrieval (Jae Sung et al. , 2009)- A language independent Statistical Transliteration Model (STM) was presented by them in this paper. In order to generate transliteration variations, system automatically learn rules from word aligned pairs. The system compared two methods based on STM for the transliteration of English language to Korean language: First is the pivot method and Second is the direct method. The first is a two steps method: First it converts the alphabets or words of English language into their corresponding pronunciation symbols using the STM and after that it converts these pronumciation symbols in to their corresponding Korean words using the Korean standard conversion rule. In second method,system directly converts the English alpabets or words into their corresponding Korean words using the STM without intermediate steps i.e, converting the English words into pronunciation symbols.. The system gives an accuracy of 45.1% and 78.5% for standard run and best nonstandard run respectively.

A Transliteration System Based on Character Sequence Modeling (Manoj et al. , 2009) -In this research paper, author discuss about the various factor of character sequence modeling and their impact on the transliteration accuracy. They described four different factors that have impact on the transliteration accuracy. Those factors are granularity, smoothing technique, corpus variation and word origin. They also show that one can achieve the transliteration accuracy by proper control or make use of the monolingual resources. Through this result they conclude that when the resources are insufficient as per the demand then a a reasonable transliteration system can be built which lack large parallel corpora by employing monolingual resources. However, there were some error in their system because of the following reasons: multiple transliterations, origin misclassification,lack of context sensitive mapping rules and there were some schewa related problems also.

Research papers related to **Translation**:-

Improved Domain Adaption for Statistical Machine Translation (Wei-wang et al. , 2012)- In this paper, author described that phrase features are independent to domain in case of Improved Domain Adaption. Experiments for the System are carried out for English and 10 European Languages. The most general ten European Languages are: Italian, Spanish, French, Portuguese, German, Swedish, Finnish, Turkish, Danish, and Dutch. As the system coverts these 10 languages to English, they Construct 20 language pairs from the English and European Language in both directions ex: English to Spanish and Spanish to English. There are two domains in the proposed model: First is generic and Second is patent. The data size for first domain is 250 million words for each pair pair of language. The data for the second domain is the European Patent Office. There is a generic development set of 3400 sentences in the MERT tuning set and there are 2000 sentences in the patent domain development set. For testing purpose, the test case for first domain contains 5000 sentences and each patent domain test set contains 2000 sentences. For each language pair, they use a 4-gram generic language model for the training purpose. They also train a 4-gram patent LM from the target side of the patent parallel data. LM data overlap with the development set and the test set is removed.The improved patent translation accuracy for the Transliteration system is 0.35 BLEU.

A Systematic Comparison of Phrase Based, Hierarchical and Syntax Augmented Statistical Machine Translation (Andreas et al. , 2008)- In this paper , author calculate the value of Probability of synchronous context-free grammar approaches built upon the state of phrase systems. There are three data configurations for the system for the experiments. The data available for the NISt 2008 track translation task was the first data configuration.Single trained language model was used for the second data configuration on the target side only . 10% bilingual training data of Simulation of a low data scenerion was used for the third data configuration. Average Test result for Chineese-English System is 34.4 for system and 53.3 for Arabic-English system.

Large Language Models in Machine Translation (Thorsten et al. , 2007) – In this paper , author presents a large language models in machine translation.The model has a distributed architecture for training and applies large-scale language models to machine translation. They trained 5-gram language models because of varying in amount of text from 13 million to 2 trillion. The data is divided into four parts: target data obtained from English side of Arabic-English parallel data, ldcnews data obtained from concatenation of several English news data sets provided by ldc, webnews data obtained from over several years upto December 2005,and web general data. They use Perlexity for measuring the quality of language model. Perplexity ranges 280.96 to 228.98 for target data, ranges from 351.97 to 210.93 for ldcnews data, ranges from 221.85 to 164.15 for webnews data. BLEU score for Arabic-English translation system was 0.4535.

Empirical Study on Computing Consensus Translations from the Multiple Machine Translation Systems (Wolfgang et al. , 2007)- In this paper , author represent an Empirical Study on Computing Consensus Translations from the Multiple Machine Translation Systems. Two data sets was created for conducted the Experiments of the translation for translation of Chinese- English text. A enhanced verison for word sausage newtroks and a two-pass serach algorithm which determines the words and re-orders their bags that build the parts of the final hypothesis was used for computing the Composite translations in the system.The

model that is used for computing the the composite transaltion is BLEU matrices. All measures of system performance is in terms of IBM-BLEU score. BLEU score for combination of 1,2,3 is 32.10 for primary system ,32.97 for consensus and 38.54 for oracle.

An Improving Word Alignment with Bridge languages (Shankar et al. , 2007) - In this paper , author represent An Improving Word Alignment with Bridge languages. They use multi-lingual, parallel, sentence aligned corpora in several bridge languages to increase or improve statistical machine translation. The approach consist of two components. By performing Experiments on the translation system they found that multilingual, parallel text in Spanish ,French , Russian, and Chinese can be utilized in this framework to improve translation performance on an Arabic-to-English task.

CHAPTER 5

METHODOLOGY

The dissertation proposal "Hybrid Approach For Translation of Common Punjabi Phrases to English" will work for translating the common Punjabi phrases to English. There will be two outputs for the proposed system: first for transliteration and second for translation.

First of all, there will be a database of three dictionary. First dictionary will store the Common Punjabi Phrases in Gurumukhi, Second dictionary will store the transliterated Common Punjabi Phrases stored in dictionary 1 in English and third dictionary will store the translation of Common Punjabi Phrases of dictionary 2 in English. When a user gives an input, the system will first check database ,whether the phrase is stored or not. If yes, then the system will give its corresponding transliterated and translated phrases as output but if the phrase is not present in the database then the system will apply transliteration and translation process.

Transliteration system of common Punjabi phrases will use Unicode of the Punjabi alphabets and some rules to transliterate the Common Punjabi Phrases into English. Unicode is a standard for the consistent encoding, representation and handling of text expressed in most of the world's writing systems.

Table No. 3: Unicode font List :-

ਗੁਰਮੁਖੀ	UNICODE
ੳ	U0A13
ਅ	U0A05
ੲ	U0A73
ਸ	U0A38

ਹ	U0A39
ਕ	U0A15
ਖ	U0A16
ਗ	U0A17
ਘ	U0A18
ਙ	U0A19
ਚ	U0A1A
ਛ	U0A1B
ਜ	U0A1C
ਝ	U0A1D
ਞ	U0A1E
ਟ	U0A1F
ਠ	U0A20
ਡ	U0A21
ਢ	U0A22

ਣ	U0A23
ਤ	U0A24
ਥ	U0A25
ਦ	U0A26
ਧ	U0A27
ਨ	U0A28
ਪ	U0A2A
ਫ	U0A2B
ਬ	U0A2C
ਭ	U0A2D
ਮ	U0A2E
ਯ	U0A2F
ਰ	U0A30
ਲ	U0A32
ਵ	U0A35
ੜ	U0A5C

ਸ਼	U0A36
ਖ਼	U0A59
ਗ਼	U0A5A
ਜ਼	U0A5C
ੜ	U0A5E
ਲ਼	U0A33
ਅ	U0A05
ਆ	U0A06

Stage 1: Consider a sentence "ਮੈ ਤੁਹਾਡੇ ਨਾਲ ਆਵਾਂਗਾ". By applying Unicode and Rule based approach this sentence will be transliterate to " Mai tuhade naal awanga".

Similarly, the sentence "ਤੁਸੀਂ ਕੀ ਕਿੱਤਾ?" will be transliterated to " tussi ki kitta?"

Stage 2: Second phrase will translate the Common Punjabi Phrases obtained from stage 1 to English. This approach will also work in two steps.

First step is to translate the Common Punjabi Phrases word by word and then apply the grammar rules for making it a complete sentence.

Some English grammar rules are:-

Rule 1: To find the subject and verb from a sentence first find the verb and then ask who or what performed the verb, from which we can find the subject.

Rule 2: Sentences can have more than one subject, verb or both.

Rule 3: If a verb follows to, then the verb is main verb and is called an infinite.

Rule 4: If we request or command like Silence please or Stop! , the subject is "You" because if we want to know who is to keep silent or stop , the subject is clearly understood i.e. "you".

Rule 5: A subject will come before a phrase beginning with of. This is a key rule for understanding subjects.

Rule 6: Two singular subjects connected by or, either/or, or neither/nor require a singular verb. Like two singular subjects negative and positive are connected by: either positive or negative, positive or negative, neither positive nor negative.

Rule 7: The verb in an or, either/or, or neither/nor sentence agrees with the noun or pronoun closest to it.

Rule 8: As a general rule, use a plural verb with two or more subjects when they are connected by and.

Example 1.:- Consider a phrase " mere dil da ranjha".

The system will first translate the phrase word by word:

Mere : my ;   Dil : heart ;   Ranjha : Romeo

And then by applying grammar rules the phrase will be translated as "My hearts romeo".

Example 2:- Phrase "Tussi ki kitta?" will be translated word by word:

Tussi: you   ;   Ki: what   ;   Kitta: do

And then after applying rules the translated phrase will be "what did you do? "

Example 3:- Consider a phrase "kiwen ho tusi"

The system will first translate the phrase word by word:

Kiwen : how   ;   ho: are ;   tusi: you

And , then by applying rules the translated pharse will be "How are you"

The system will work as per following steps:

First, User Enter a Engish sentence that he/she wants to translate and transliterate in Punjabi. Second, there is an option to choose whether to translate the sentence without splitting it into words or by splitting the sentence first into words then translate them in English. Based on this , If the user wants to translate the phrase without splitting the pharse into words then the system will first check the dictionary of pharses for the translating the phrase without splitting it into words and if the phrase is stored in the dictionary , then corresponding output will be generated. Second, if the user wants translate the phrase by splitting the phrase into words then system will first find the corresponding translations of each words and based on the grammer rules, the output will be generated.
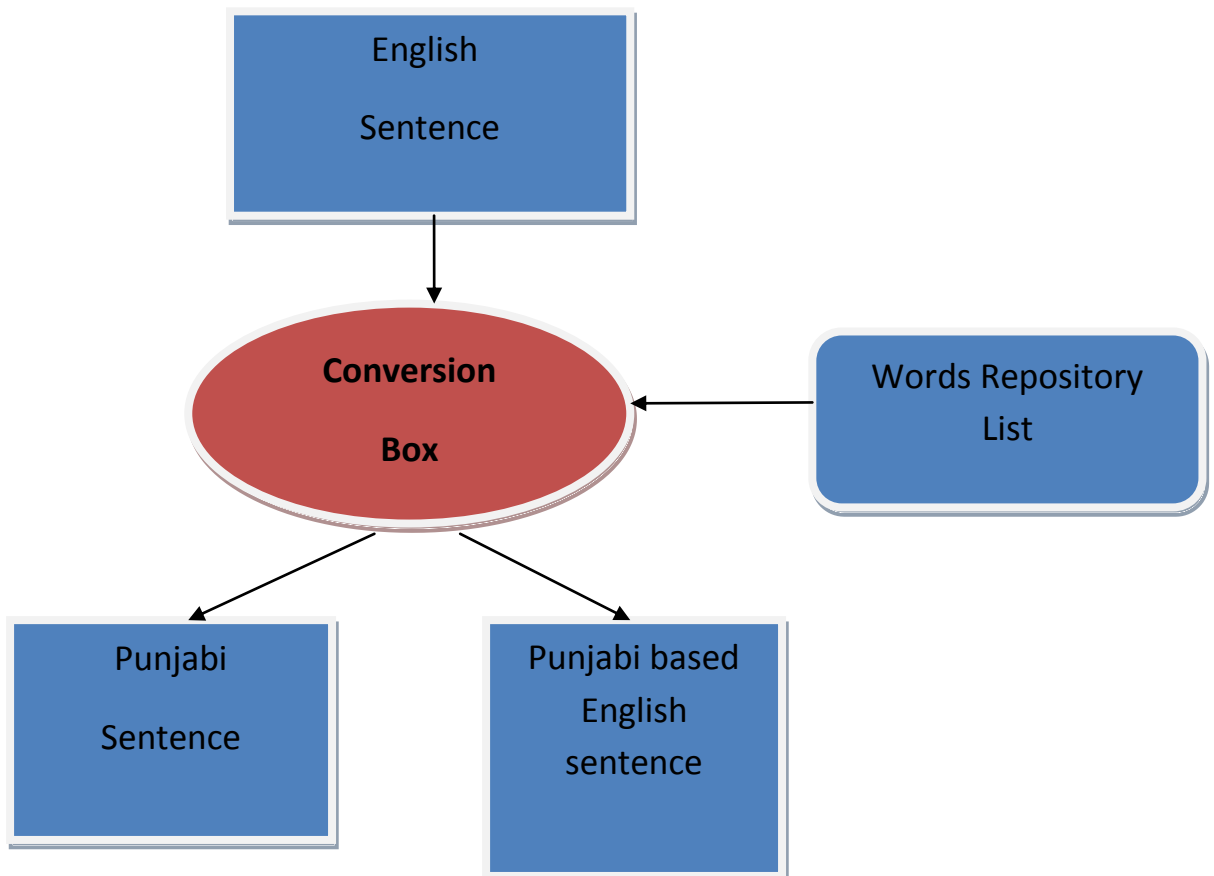

Fig. 5.1(Flow Chart of System)

Pseudo code for conversion from English to Punjabi and English based Punjabi common sentences.

MAIN PROCEDURE

Enter English sentence to variable English

enter option to split in words or take full sentence to Full

initialize array of Tokens[] containing words

initialize object of Converter class as Obj

if (English ! = NULL)

if (Full == true)

set Token[0] to English

else

split English w. r. t. spaces and store words as tokens to Tokens[] array

end if

//iterate all the tokens from the array of tokens

foeach Tokens[] as token

word = Obj.checkWord(token) //find the word of token w. r. t. Obj (Converter)

if (word != NULL)

PRINT word

End if

endfor

end if

//PROCEDURE checkWord FROM CONVERTER CLASS

Procedure: checkWord(token)

Initialize array of English2Punjabi[] words //contains Unicode words in Punjabi from English words

Initialize flag to false

Foreach English2Punjab as Word

If (Word == token)

Create object OutputWord containing Punjabi word

flag = true

break

End if

End for

If (flag == false)

Create object OutputWord containing NULL word

End if

Return OutputWord to Main Procedure

CHAPTER 6

RESULTS AND DISCUSSION

As already discussed, the proposed system will work for Transliteration of Common Punjabi Phrases as well as for Translation of Common Punjabi Phrases.. So, Whenever a user gives a Common Punjabi Phrase as an input to the system, the system will give its corresponding transliterated and translated phrase in English.

Using this proposed system named "Hybrid Approach for Translation of Common English Phrases to Punjabi", we have implemented the required work to convert the general phrases of english sentences to its corresponding punjabi sentences and achieved the success as well.

To do the same thing we have designed the complete system into Advance Java programming language using JSP and Servlets techniques.

Fig. 6.1(Interface of System)

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

Kaun Ho Tusi ।

English to Punjabi Time Taken: 229 (micro secs)	English to Punjabi English Time Taken: 39 (micro secs)

Total Time Taken: 268 (micro secs)

Developed By: Aman Aggarwal
Institute: Lovely Professional University
Reg No: 10906136
Thesis Title: Hybrid Approach for Translation of Common English Phrases to Punjabi

As shown in the above result window in which three different sections as listed below

1. English Sentence Box (Pink background shaded)
2. Punjabi Convereted Sentence Box (Blue background shaded)
3. Punjabi Based English Sentence Box (Yellow background shaded)

So we just need to enter the the english general sentence in the english box like we entered "who are you?" and the correspondig result generated for punjabi language is "ਕੋਨ ਹੋ ਤੁਸੀ ।" and further the Punjabi based English conversion is Kaun Ho Tusi ।.

The system will work in two ways :-

First the translation of sentence can be done without splitting it into words.

Second , by splitting the sentence into words.

Fig. 6.2(Interface of system)

Hybrid Approach for Translation of Common English Phrases to Punjabi

English Content

who are you

Convert   Complete Sentence Based Conversion ? YES ☑

Punjabi Conversion

ਕੋਨਹੋਤੁਸੀ ।

Punjabi Based English Conversion

Kaun ho tusi ।

| English to Punjabi Time Taken: 26940 (micro secs) | English to Punjabi English Time Taken: 5460 (micro secs) |

Total Time Taken: 32401 (micro secs)

Developed By: Aman Aggarwal
Institute: Lovely Professional University
Reg No: 10906136
Thesis Title: Hybrid Approach for Translation of Common English Phrases to Punjabi

If you want to translate the sentence without splitting it into words then simply click the check button "YES" and if you want to translate the sentence after splitting it into sentence , uncheck the "YES" check button.

Further the window shows the time taken to generate the sentence in two different manners.

1. From english to punjabi based conversion time taken
2. From english to punjabi based english conversion time taken

Like in the above example for sentence "Who are you?" the conversion time taken 229 micro seconds into punjabi and 29 micro seconds time taken for conversion to punjabi based english conversion.

Now , suppose we are translting the above phrase without splitting it into words then the interface will be like

Figure 6.3(Implementation)
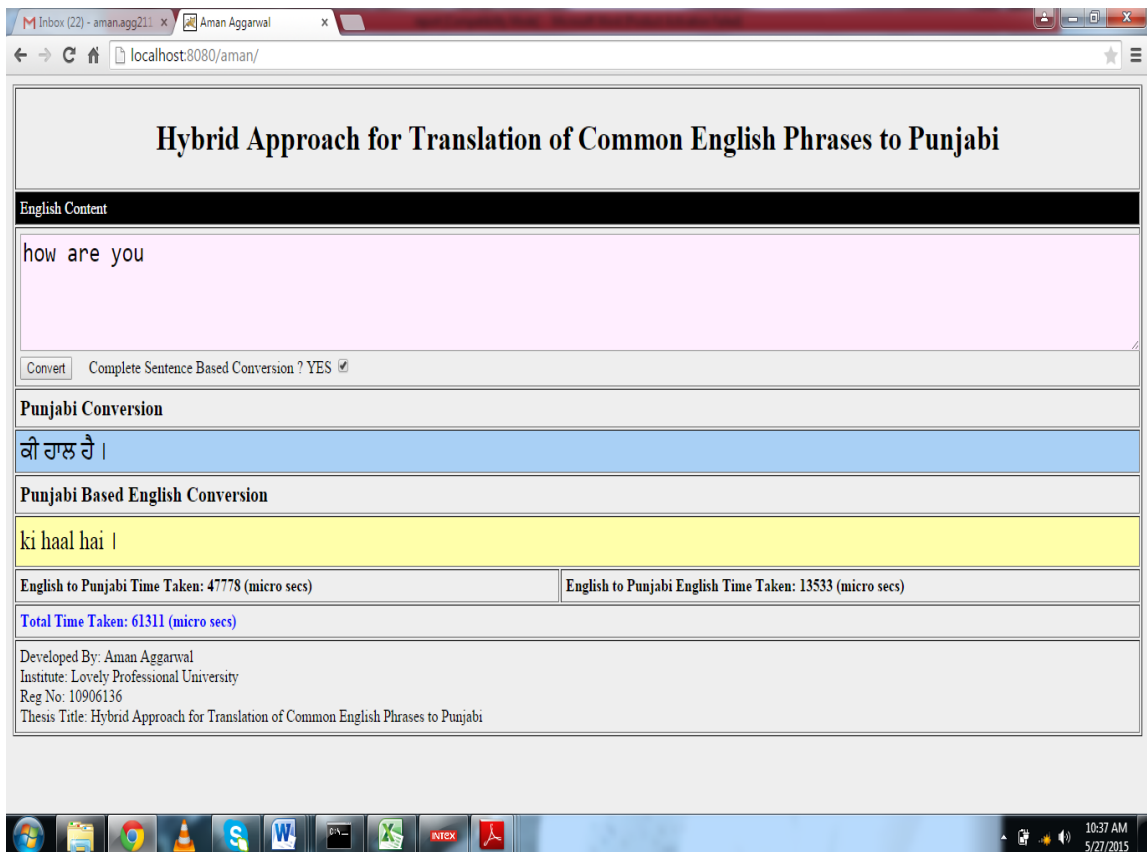
When we translate the pharse without splitting it into words then, the system will check the database for the entered phrase and if the phrase is in the database then, its corresponding translation and transliteration will come as output on the given interface.

Ex:- Enter pharse "how are you? "

There can be two options for the transliteration of the given phrase.First to translate it without splitting it into words and second by splitting it into words. The database has follwing options for the given pharse.

The pharse "How are you" has stored it self in the databse with its corresponding translaiteration "ki haal hai". If we translate the phrase by first splitting it into words then it wil work like :

First check for the word "How"

In database, the word "How" is available with its corresponding transliteration "Kiwen ". The word "are " is also available with its transliteration "hai". Similairly, the word "you" is also available in the database with its transliteration "tusi". So the output of the given pharse will be "kiwen ho tusi".

The complete experimental work is to be carried out using Java Platform The dissertation proposal "Hybrid Approach For Translation of Common Punjabi Phrases to English" will work for translating the common English phrases to Punjabi. There will be two outputs for the proposed system: first for translation and second for transliteration.

First of all, there will be a database of three dictionaries. First dictionary will store the Common Punjabi Phrases in Gurumukhi, Second dictionary will store the transliterated Common Punjabi Phrases stored in dictionary 1 in English and third dictionary will store the translation of Common Punjabi Phrases of dictionary 2 in English. When a user gives an input, the system will first check database, whether the phrase is stored or not. If yes, then the system will give its corresponding transliterated and translated phrases as output but if the phrase is not present in the database then the system will apply transliteration and translation process.

The time taken to convert the sentence depends upon the following factors.
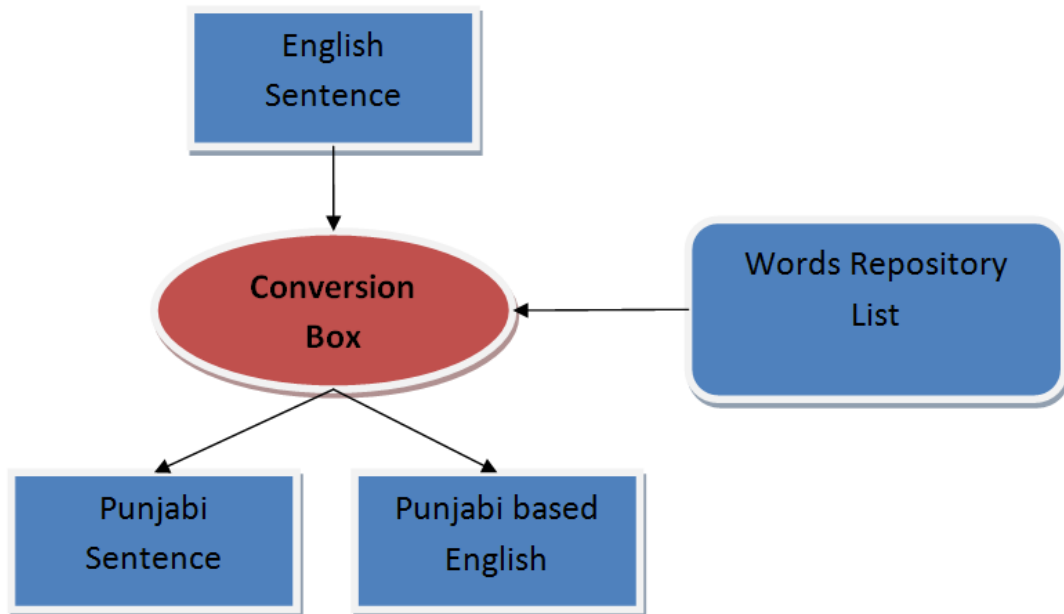
Transliteration system of common Punjabi phrases will use Unicode of the Punjabi alphabets and some rules to transliterate the Common Punjabi Phrases into English. Unicode is a standard for the consistent encoding, representation and handling of text expressed in most of the world's writing systems. In the latest version of Unicode there are more than 110,000 characters covering 100 scripts and multiple symbol sets.

The following unicode character set is being used to generate the reqired output.

The white background characters are assigned with the code values so if we want any code value then we have to do the sum up of left side code value to the top side code values together to generate a unicode character value. The combinations of different code values togethere generate the words in punjabi language which is the main concept used to form the dictionary of the transaltion backup.

Base	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F
Gurmukhi: 0A00 – 0A7F ref																
0A00		◌ਁ	◌ਂ	◌ਃ		ਅ	ਆ	ਇ	ਈ	ਉ	ਊ					ਏ
0A10	ਐ			ਓ	ਔ	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ	ਟ
0A20	ਠ	ਡ	ਢ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ		ਪ	ਫ	ਬ	ਭ	ਮ	ਯ
0A30	ਰ		ਲ	ਲ਼		ਵ	ਸ਼		ਸ	ਹ			◌਼		ਾ	ਿ
0A40	◌ੀ	◌ੁ	◌ੂ					◌ੇ	◌ੈ			◌ੋ	◌ੌ	◌੍		
0A50										ਖ਼	ਗ਼	ਜ਼	ੜ		ਫ਼	
0A60							੦	੧	੨	੩	੪	੫	੬	੭	੮	੯
0A70	◌ੰ	◌ੱ	ੲ	ੳ	ੴ											
Gujarati: 0A80 – 0AFF ref																
0A80		◌ઁ	◌ં	◌ઃ		અ	આ	ઇ	ઈ	ઉ	ઊ	ઋ	ઌ	ઍ		એ
0A90	ઐ	ઑ		ઓ	ઔ	ક	ખ	ગ	ઘ	ઙ	ચ	છ	જ	ઝ	ઞ	ટ
0AA0	ઠ	ડ	ઢ	ણ	ત	થ	દ	ધ	ન		પ	ફ	બ	ભ	મ	ય
0AB0	ર		લ	ળ		વ	શ	ષ	સ	હ			◌઼	ઽ	◌ા	◌િ
0AC0	◌ી	◌ુ	◌ૂ	◌ૃ	◌ૄ	◌ૅ		◌ે	◌ૈ	◌ૉ		◌ો	◌ૌ	◌્		
0AD0	ૐ															
0AE0	ૠ	ૡ	◌ૢ	◌ૣ		૦	૧	૨	૩	૪	૫	૬	૭	૮	૯	
0AF0		૱														

Color key:	Unassigned code point	Non-printable control character	Selected search character

Unicode of Gurumukhi Alphabets

The following snap shot is shown in which the java class is prepared to store the punjabi words with its corresponding english word and further transliteration punjabi based english word like tusi.

```java
package com.model;

public class Word {

    public String engWord, punWord, pingWord;

    public int order;


    public Word(){

        engWord = "";

        punWord = "";

        pingWord = "";

        order = 0;

    }


    public Word(String eng, String pun, String ping, int ord){

        engWord = eng;

        punWord = pun;

        pingWord = ping;

        order = ord;

    }

}
```

The above class file shows the storage procedure of a english word.

The four main attributes are indicated to store a word.

1. engWord: English word
2. punWord: unicode character based punjabi word stream contains punjabi unicode character set.
3. pingWord: punjabi based english word contains english unicode character set
4. order: this is integer value show the order to appear for a word from value 1 to 10 in which if two words need to be displayed then order will be che

The way is shown above to store the punjabi, english and punjabi based english words into the list with order number.

The word class structure is already shown above in which the constructor takes four different parameters in which the first word is english word, second word is unicode chain of punjabi word, third is punjabi based transliteration of english word and fourth is order of the word to be displayed.

Now the following class shown the conversion procedure which is again the java based code system shown below.

The above code shown is activated when the sentence is to be entered from the main screen and covert button hits the code starts get executing and first splits the entire sentence into an array of strings of words and the loop iteration or words take place in which each word is checked into the available list of words one by one and further the result is sent back to the calling main procecure.

Experiental time calculation values:

Sentence in english to be taken: "who are you?"

Table 6. Time calculated (unit: 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

Methodology to calculate time interval for conversion:

In java there is one command is available using that command the time of the system is to be converted into nano seconds and stored into a variable start and when the word searching get completed again the time taken from the system in nano seconds format and stored into the end varilables and then difference values are to be calculated.

Long start, end;

Start = System.nanoTime();

//Algorithm works

End = System.nanoTime();

Long difference = end – start;

//display the result availabe in difference variable.

Graphical representation of results shown.

The graph respresentation to be done with the help taken from online tool available chartgo.com website.

Time graph for conversion from english word to punjab words.

The graph shows the time taken by the system for the translation of "How are you " English phrase to Punjabi . As we see, if we again translate this phrase , it will take less time because when we translate this pharse first time, it will get stored in the priority array and if we again translate it, it will first search the priority array instead of searching the complete array.Consider a example of translating the pharse "How are you " If we translate the pharse without splitting it into words then the time taken for the first transaltion is 97 and if we translate it again , it will take less time 62 microseconds and so on.
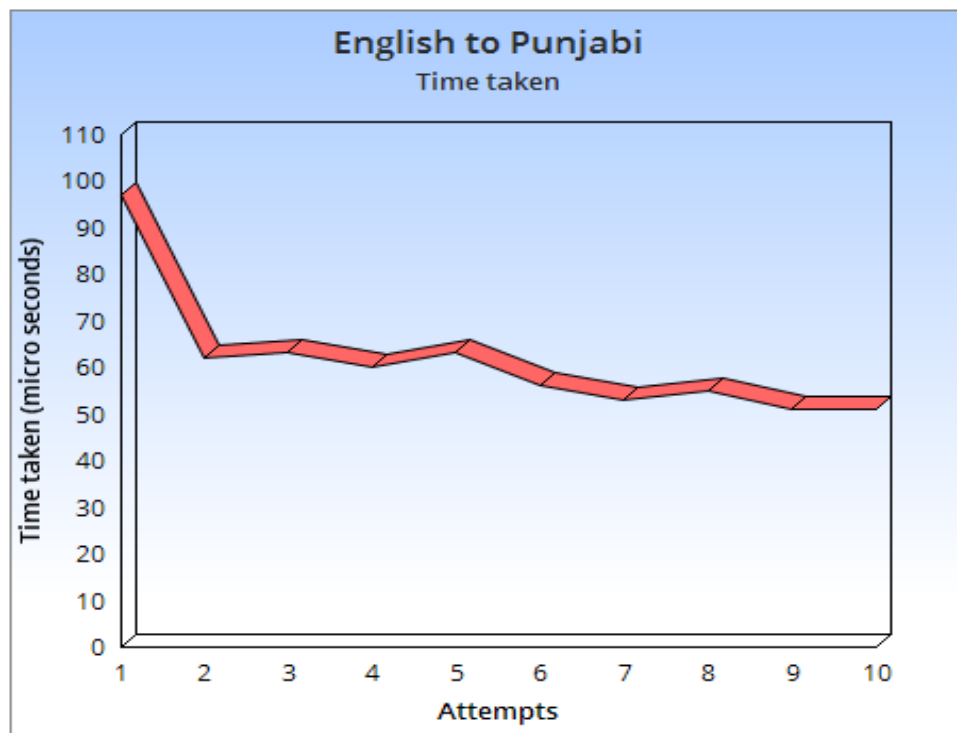
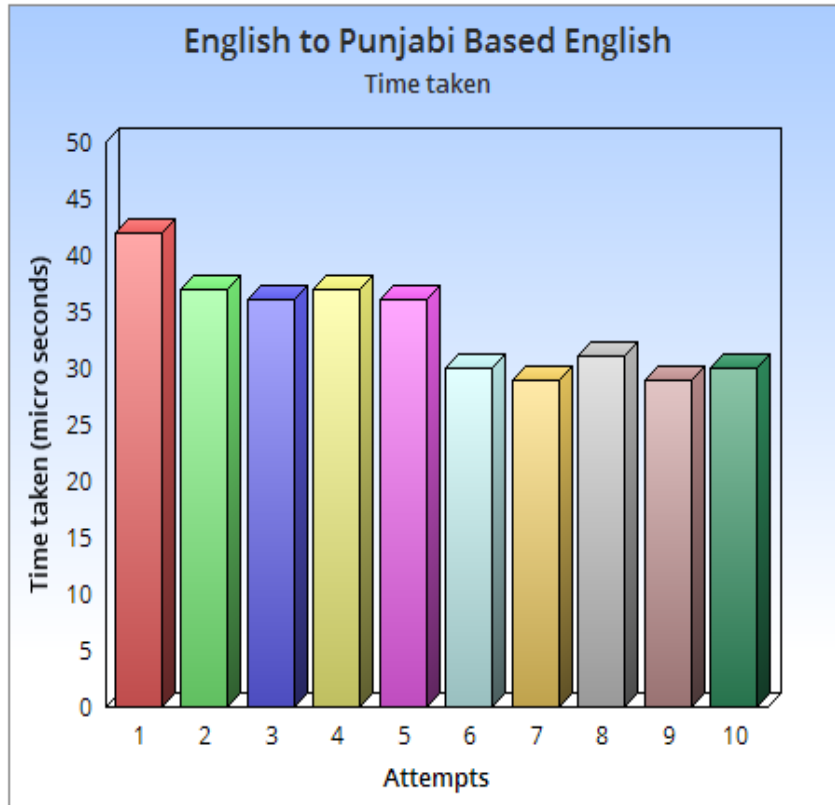This graph shows the time take by the system for the transaltion of pharse "How are you?".
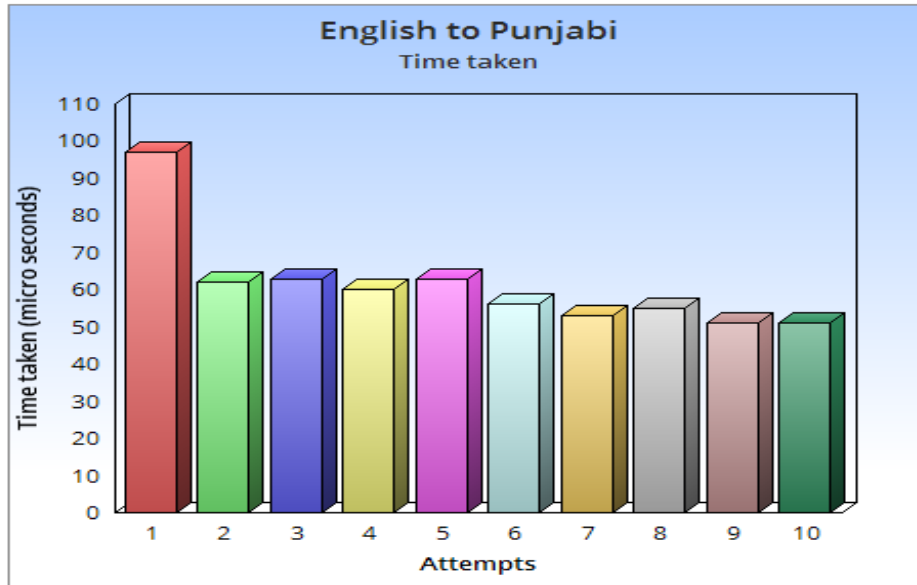
This graph shows the translartion of the same phrase "How are you?" while first spillting the phrase first into words.

Conversion from english to punjabi based english words

This graph shows the time taken by the system for the transliteration of the "How are you" English Phrase. If the transliteration is performed for the first time, system will first serach the complete dictionary and store that pharse in the priority array. So, we again enter that pharse for the transliteration , it will first search the priority array instead of searching the complete dictionary.

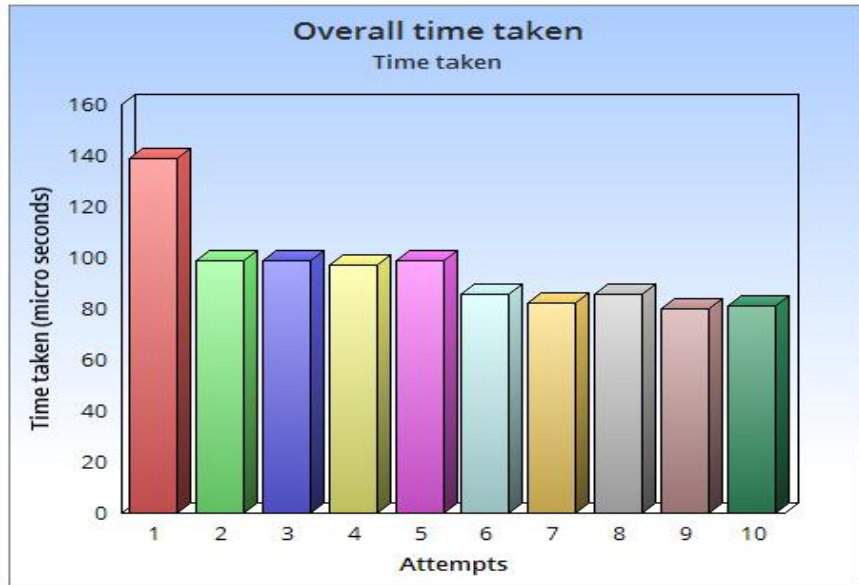**English to Punjabi Based English**
Time taken
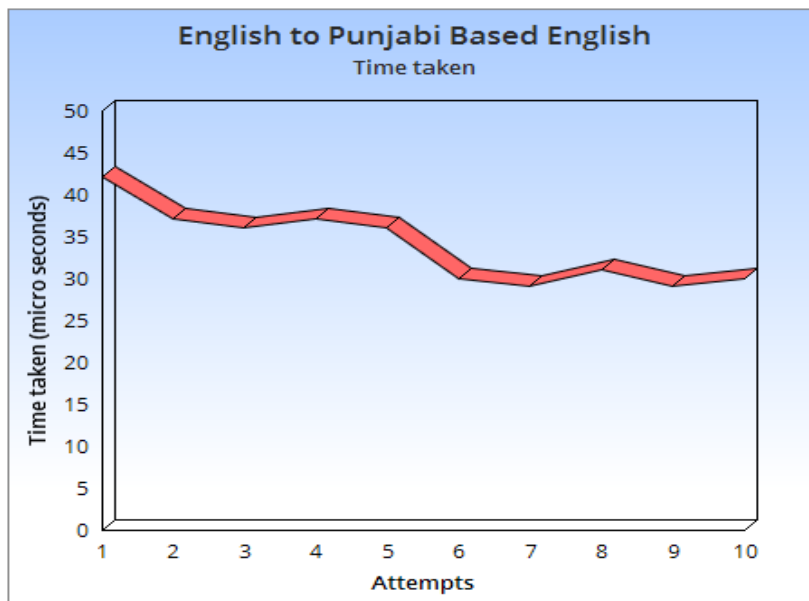
This graph show the transliteration of the phrase"How are you" with first spilliting into words.

Overall time taken for conversion

The following graph shows tha overall time taken by the system for the transaltion as well as transliteration of the common English Pharse.
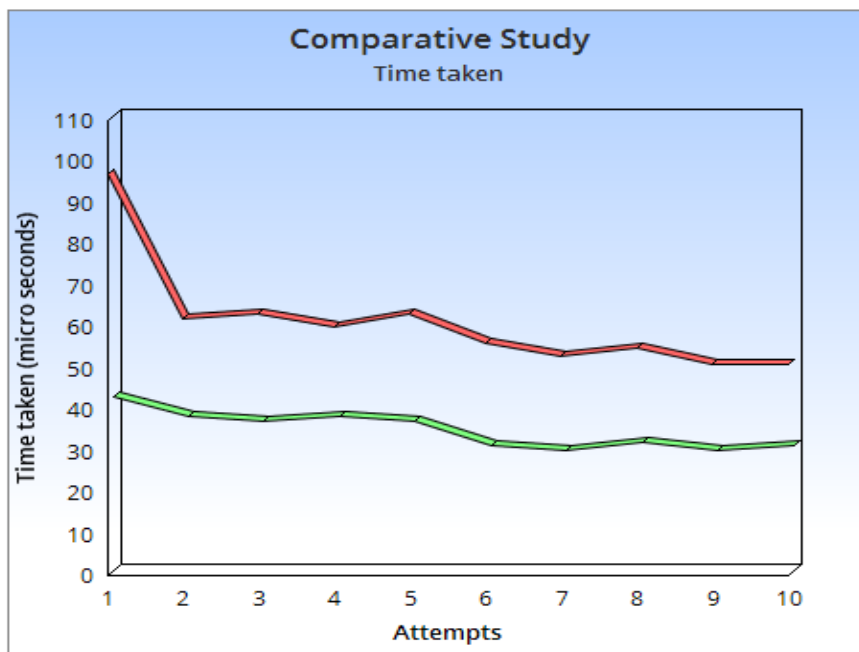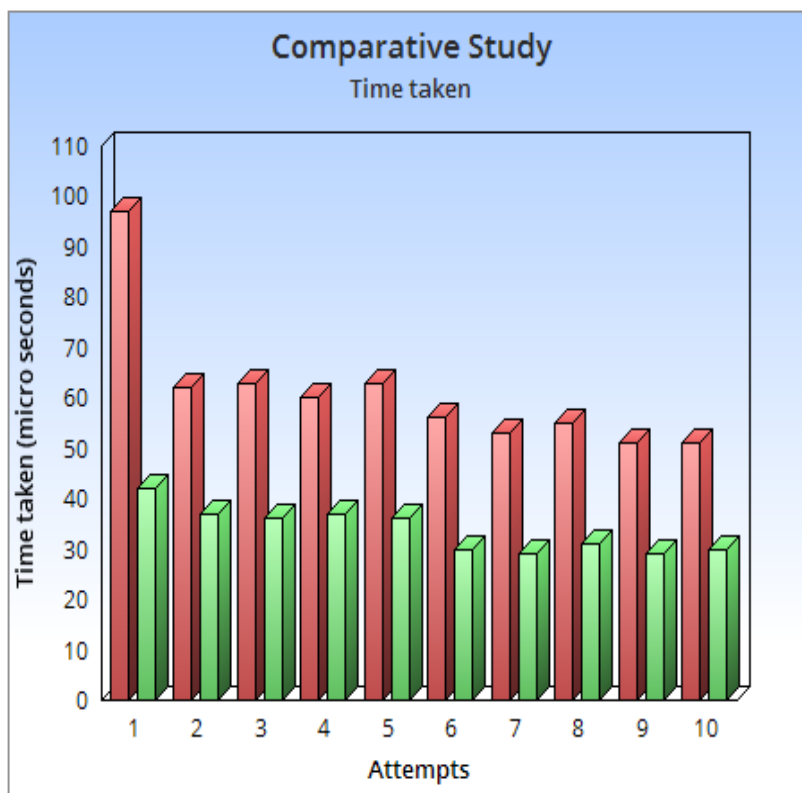

Overall time taken
Time taken
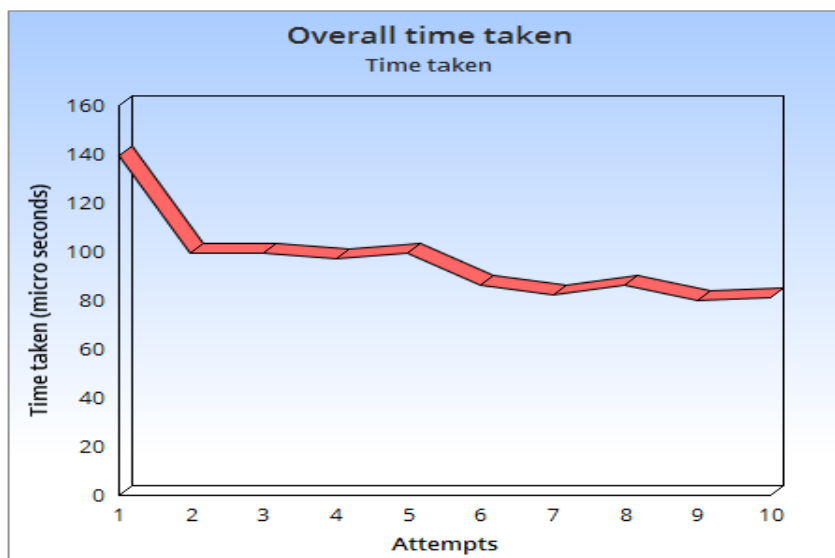
The above shows the overall time taken by the system for the translation and transliteration of the phrase "How are you" by first splitting it into words.

Comparasion between both types of conversions

The following two graph shows the comparative study between the transaltion and trasnliteration of the Common English Phrase. By storing the pharse in the priority array based on its search count, the time taken to translate as well as transliterate the Common English Pharse reduce.

Comparative Study
Time taken

The graph above shows the comparision of the time taken by the system for both translation and transliteration of the pharse "How are you?". Red bar shows the time taken by the system without splitting the phrase into words and the green bar shows the time taken by the system by first splitting the pharse into words.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

In this proposed thesis work 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. The scope in the future for the existing thesis work is quite brighter in which we can try to convert the more general phrases of english into punjabi by storing more phrases in database , also by applying some more techniques like n-gram. We can also try to convert the phrase not just into punjabi but to some other available languages too like hindi, marathi, telegu, bengali etc. This concept helps those people who knows and even in reality most of the people don't know more than two or three languages at a same time. But using this system and the system which we can design in future where we just enter the standard english phrases and by selecting the choice of language we get two different types of results in which one contain typical language output in according to selected choice of selected language and other which is most important in general english converted language based content.

SUMMARY

The dissertation proposal "hybrid approach for translation of Common Punjabi Phrases to English" will work for translation of Common Punjabi Phrases to English. The approach for the proposed system is hybrid because it will use dictionaries and apply some rules for Transliteration as well as for Translation of the Common Punjabi Phrases. The proposed system will work first for transliteration of Common Punjabi Phrases by using Unicode of the Punjabi alphabets and also apply some rules and then translate them into English. The system is proposed to have the sufficient accuracy for transliteration of Common Punjabi Phrases to English for the users who don't know about the Gurumukhi letter and also for the Translation of converted Common Punjabi Phrases by transliteration to English for the users who also want to know the meaning of that Phrase.

REFERENCES

RESEARCH PAPER

Akshar Bharati, Samar Hussian, Meher Vijay, Kalyan Deepak, Dipti Mishra Sharma, and Rajeev Sangal (2009), "Constraint Based Hybrid Approach to Parsing Indian Languages" Workshop on Lexical Resources for Natural Language Processing, 5 - 8 January held at IIIT- Hyderabad.

Andreas Zollmann, Ashish Venugopal, Franz Och and Jay Ponte (2008), "A Systematic Comparison of Phrase Based, Hierarchical and Syntax Augmented Statistical Machine Translation", 22nd International Conference on Computational Linguistics (Coling 2008), pages 1145–1152 Manchester, August.

Antony P J and K P Soman (2012), "Computational Morphology and Natural Language Parsing for Indian Languages: A Literature Survey", International Journal of Scientific & Engineering Research, Volume 2, Issue 12, December.

Bushra Baig , M.Kumar and Sujoy Das(2011) , "A Rule Based Transliteration System from Hindi to Urdu", International Journal of Scientific & Engineering Research, Volume 2, Issue 12, December.

Davinder Brar and Er. Rishamjot Kaur(2014) ," A Review of Transliteration System for Punjabi to English",   International Journal of Advanced Research in Computer Science and Software Engineering,   Volume 4, Issue 7, July.

Deepti Bhalla, Nisheeth Joshi and Iti Mathur(2013)," Rule Based Transliteration Scheme for English to Punjabi", International Journal on Natural Language Computing (IJNLC) Vol. 2, No.2, April.

Jae Sung Lee and Key Sun Choi(2009),” English to Korean Statistical Transliteration for Information Retrieval” , 19 th international conference on computational linguistics-volume 1, pages 1-7, March.

Jasleen Kaur, Gurpreet Singh Josan(2011),”Statistical Approach to Transliteration from English to Punjabi”, International Journal on Computer Science and Engineering (IJCSE) proceedings, Vol. 3 No. 4 April.

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

Manoj Kumar Chinnakotla and Om P. Damani(2009), “Character Sequence Modeling for Transliteration”, 7th International Conference on Natural Language Processing Macmillan Publishers, India, May

Pankaj Kumar and Vinod Kumar(2013), “A Statistical Machine Transliteration Based Punjabi to English Transliteration System for Proper Nouns, International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 8, August.

Shankar Kumar, Franz Och and Wolfgang Macherey(2007), “Joint Conference on Empirical Methods in Natural Language Processing and Computational Natutal Language Learning”, pp. 42-50, January.

Thorsten Brants, Ashok C.Popat, Peng Xu, Franz J. Och, Jeffery Dean(2007),” Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning” ,pp.858-867, April.

Wei Wang , Klaus Macherey, Wolfgang Macherey , Franz Och and Peng Xu (2012), proceedings of Google Inc. , 1600 Amphitheatre Pkwy,Mountain View, USA.

Wolfgang Macherey , Franz Josef Och (2007), "Joint Conference on Empirical Methods in Natural Language Processing and Computational Natutal Language Learning", pp.986-995, September.

WEBSITES

1. Transliteration" , Internet Source:- http:/ / en.wikipedia. org/ wiki/ transliteration acessed on jan,2011.
2. http://shodhganga.inflibnet.ac.in:8080/jspui/handle/10603/4456.
3. http://mylanguages.org/punjabi_alphabet.php.
4. http://en.wikipedia.org/wiki/Translation#Modern_translation.
5. http://en.wikipedia.org/wiki/Translation.
6. http://en.wikipedia.org/wiki/Transliteration.
7. http://ijmes.chass.ncsu.edu/docs/TransChart.pdf.
8. http://dictionary.reference.com/browse/transliterate.
9. http://www.merriam-webster.com/dictionary/transliterate.
10. ftp.unicode.org