AnglaPunjabi: English to Punjabi MT system based on AnglaMT Paradigm

Abstract

This paper presents a system overview of English to Punjabi Machine-Aided Translation system (MAT) named AnglaPunjabi based on AnglaBharati methodology. AnglaBharati is a pseudo-interlingua rule-based translation methodology for translation from English to Indian languages. Based upon AnglaBharati Punjabi language was taken up as case study. It was found that the customization is successful and is standing equally among the Indo-Aryan family. This system was developed by funding from TDIL, DIT and claims to have nearly 65% accuracy for health and tourism domains. The main task in developing the system for Punjabi involves creating bilingual dictionary and generation of Punjabi text from Inter-lingua. Some target language dependent codes have also been modified for Punjabi.

Introduction

English is a highly positional language with rudimentary morphology and default SVO sentence structure. Indian languages are verb ending, free word-group order language with lots of structural similarity. Indian languages can be classified into four broad groups according to their origin. These are Indo-Aryan family (Hindi, Bangla, Assamiya, Punjabi, Marathi, Oriya, Gujrati etc.); Dravidian family (Tamil, Telugu, Kannada & Malayalam); Austro-Asian family and Tibetan-Burmese. Within each group the languages exhibit a high degree of structural homogeneity. We exploit this similarity to a great extent in our system. This paper describes the methodology for how AnglaBharati can be used to develop AnglaPunjabi and the key points required to handle in the system for adaptation.

This paper has been divided into following sections:

Section (A) is System Overview. This section talks about the architecture of AnglaPunjabi system along with the points of customization for Punjabi language. This has sub-sections describing modules of the system in detail.

Section (B) has details about the External Interface designed for AnglaMT engine and Section(C) concludes the paper.

(A) System Overview

AnglaBharati uses pseudo-interlingua approach for translation. It analyses English sentence and generates its intermediate form named PLIL (Pseudo Lingua for Indian Languages) with all the disambiguation possible. This PLIL has the word and word-group order as per the structure of the group of target languages. This intermediate representation is converted to Indian language using text-generator. For illustration consider following examples:

1. Boy is going .

   PLIL: <aff {sub_np (boy noun masculine singular third [human] [muzdA:m 10] [] []) } {toinf } {main_vp_active (go verb_5 normal is masculine singular third [jA] 3 [] []) } > . sviram

   Output: ਮੁੰਡਾ ਜਾ ਰਿਹਾ ਹੈ (muzdA jA rihA hE)

2. Boys are going.

PLIL: <aff {sub_np (boys noun masculine plural third [human] [muzdA:m 10] [] []) } {toinf } {main_vp_active (go verb_5 normal are masculine plural third [jA] 3 [] []) } } > . sviram

Output: ਮੁੰਡੇ ਜਾ ਰਹੇ ਹਨ (muzde jA rihe hana)

In the above examples, the SVO nature of English is mapped to SOV of Punjabi in PLIL form. The verb "go" in English is "jA rihA hE" in Punjabi for masculine singular third person and verb type 5 (ing form). Similarly for masculine plural third person it is "jA rihe hana" in Punjabi. Such patterns are generated though textgenerator module using PLIL information. These modules are explained in detail in later sections.

Let us first understand about the AnglaBharati technology and its various modules. A block schematic diagram of AnglaBharati methodology is shown in Figure1.The shaded boxes are the points of customization for Punjabi.

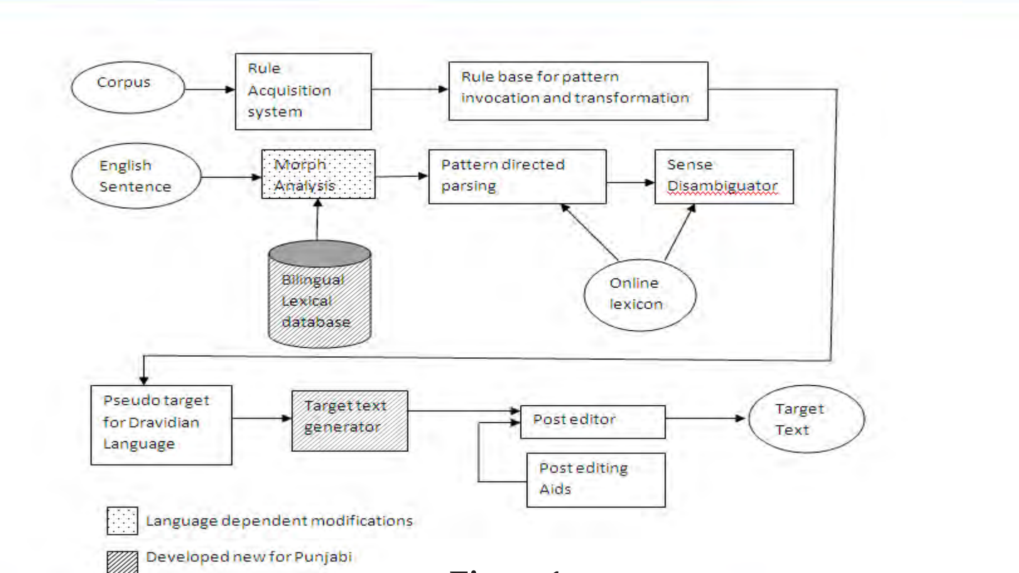


Figure 1

The AnglaBharati has following major modules:

(I) Rule base: This module contains hand crafted rules in Context Free Grammar like structures for corresponding target language structures applicable to a group of Indian Languages. These patterns have been developed by examining English corpus and target language sentences. For an input sentence, word information is combined to form patterns and these patterns are matched to the rules in the rule base. Top down approach is used for matching the rules. System tries all the possible combinations as a result more than one pseudo target is generated.

This intermediate representation is used in AnglaBharati system architecture as it exploits structural similarity of Indian languages.

This module is also responsible to disambiguate the word senses to the extent possible using interleaved semantic interpreter. Further disambiguation, correct right choice and lexical preferences are taken care of by Text generator module. AnglaPunjabi has used this rule base module as it is .Hence, no tuning done for

Punjabi language.

(II) Morphological Analyzer: Morphological analyzer takes as input the English sentence, reads each word, identifies the root word, and retrieves the necessary information from the lexical database about that word. Identification of different kinds of phrasal (group of words when used together exhibit entirely different meaning from their original meanings) is also done during morphological analysis. If a particular word is not found in the lexical database, system transliterates the word and also tries to guess the expected syntactic category of the word. An on-line lexical database is created, as we proceed with the process of analysis of the input text. Here, we store information about all the words encountered in the text, and the analyzer first examines the entries of this lexical database before searching in the large multi-lingual database. As a large number of words get repeated in a typical text, the on-line lexical database helps to reduce the search time to a very large extent.
Morphological analyzer makes use of the multi-lingual lexical database and different paradigm tables for English language to extract the root word from different forms of the word.

For Example: the word ’play’ as a verb has 5 forms (plays, play, played, played, playing), and as a noun it has two forms (play, plays). But, in the lexical database, only the root word ’play’ is stored with syntactic and semantic knowledge and for Punjabi only the root word meaning is stored.

The **Transliteration module** of Morphological Analyzer contains transliteration rules to get Punjabi from English word. If a word is not present in lexical database then system transliterates that word based upon rules. Though Hindi and Punjabi have similarities like GNP (Gender-Number-Person), some dissimilarity are there. Punjabi doesn’t’ have same conjuncts consonant. In Punjabi we have ਹ, ਰ , ਵ as conjunct consonants and Punjabi conjuncts are used as ਨ +ਹ = ਨ੍ਹ, ਪ + ਰ= ਪ੍ਰ, ਸ + ਵ =ਸ੍ਵ . Likewise, the transliteration rules have been framed for Punjabi. The English characters are converted into Roman notations which helps unambiguous representation of these words while converting to Punjabi. All words with capital letters will be treated as acronym and processed separately.
For example,
Crocin--> krosina
AICTE-->eAIsItII

As depicted in Figure-1, the following two modules have been developed new for Punjabi language:

i. Development of English-Punjabi language lexical database: This step involves entry of Punjabi words for each English word/phrase/word-groups for each semantic role and entry of paradigm numbers and other specific information for Punjabi.

ii. Development of target language text generator: The text generator converts the pseudo code in PLIL which is generated by the rule base of the system into Punjabi language. The main task in the module is to generate Punjabi words from the morphological information available in the PLIL. It also constructs sentences, such that the syntactic and semantic constraints of Punjabi language are met. Morphological and syntactic rules, which covers all the features of Punjabi language, are required as input to this module.

(III) English-Punjabi lexical database

The lexical database is the fuel to the translation engine. It contains various details needed for an English word along with its grammatical and syntactic information, like its rootword, part of speech it belongs to, semantic information and Punjabi meaning along with paradigm number. A root word may have multiple categories and/or multiple meanings. Current version of AnglaPunjabi system comprises of 57,000 words in lexical database.

For example,

```
play
16 verb
~G play;~G to play with instruments
[],[],[],[],[];[Tn,Tn.pr],[],[animate]
,[instrument],[]
Keda:10;vajA:1
##
3 noun
~G activity done for amusement esp by
children/a drama on stage before the
audience;~G play games
[activity];[activity]
nAtaka:m 4;Keda:m 4
***
```

Lexicon database module consists of two sub modules, namely: Lex entry and creation of domain specific lexicon. Lexicon entry means mapping English word with Punjabi meaning along with its paradigm number and other related information as explained above. If the meaning of any word does not specifically belong to Health or Tourism domain, then that word will remain in the general lexicon else corresponding entry will be copied to domain specific list. For example, English word "cancer" has two meanings, one related to <illness> and other to <zodiac>. In this case, the meaning related with illness will be moved to domain specific list. This is done to restrict the system to search the lengthy database if user is looking for domain translation only.

(IV) Punjabi Text Generator

This module decodes the PLIL information to generate Punjabi sentences. Text Generator is divided into following sub processes based on the language rules to be implemented for translation.

1. **Paradigm file generation**: With the help of paradigm files, root word is extracted from the original word and all the information about that word is retrieved. The paradigm number has to be generated for different word categories as per the requirement. The paradigm number is a unique id given to a set of those words that have similar word ending and generate similar forms by substituting suffix. We have paradigm files for nouns, verbs, adjectives and pronouns.

Root	Number	Case	Generated Form
muMd_A	singular	direct	muMd_A
muMd_A	plural	direct	muMd_e
muMd_A	singular	oblique	muMd_e
muMd_A	plural	oblique	muMd_iAz

Noun paradigms: In case of noun, the classification is based on gender, number, case marker and word ending. For words with similar word ending, same gender and same inflectional form will fall under one paradigm, ie. they will be given one unique id.

For example, root word boy(ਮੁੰਡਾ in Punjabi) will have 4 variations ਮੁੰਡਾ, ਮੁੰਡੇ, ਮੁੰਡੇ, ਮੁੰਡਿਆਂ as shown above :

For another word horse (ਘੋੜਾ in Punjabi), the variations are ਘੋੜਾ ,ਘੋੜੇ ,ਘੋੜੇ,ਘੋੜਿਆਂ. The suffix addition are same as boy (ਮੁੰਡਾ), hence both will have a single paradigm number.

Adjective Paradigms: Adjectives in Punjabi has variations only for words ending with "_A", so a single paradigm has been created for two genders. The forms are generated by considering direct/oblique forms along with gender and number information.

For example, adjective "good" (with root word ਚੰਗਾ) qualifying masculine noun will have 4 variations as ਚੰਗਾ ਚੰਗੇ ਚੰਗੇ ਚੰਗਿਆਂ. And adjective "good" qualifying feminine noun will have variations as ਚੰਗੀ ਚੰਗੀਆਂ ਚੰਗੀ ਚੰਗੀਆਂ

variations as ਚੰਗੀ ਚੰਗੀਆਂ ਚੰਗੀ ਚੰਗੀਆਂ

Root	Gender	Number	Case	Generated Form
caMgA	"m"	"s"	"d"	caMg_A
caMgA	"m"	"p"	"d"	caMg_e
caMgA	"m"	"s"	"o"	caMg_e
caMgA	"m"	"p"	"o"	caMg_iAM
caMgA	"f"	"s"	"d"	caMg_I
caMgA	"f"	"p"	"d"	caMg_IAM
caMgA	"f"	"s"	"o"	caMg_I
caMgA	"f"	"p"	"o"	caMg_IAM

Verb Paradigms: For Punjabi verbs the forms are inflected according to gender, number, person and tenses. Transitive and Intransitive nature of verbs are also considered. For example, root word KA(ਖਾ), i.e. to eat will be inflected as shown below:

Past Tense: ਖਾਧਾ ਖਾਧੇ ਖਾਧੀ ਖਾਧੀਆਂ

Subjunctive: ਖਾਵਾਂ ਖਾਵੇਂ ਖਾਵੇ ਖਾਈਏ ਖਾਵੋ ਖਾਣ

Future: ਖਾਵਾਂਗਾ , ਖਾਵੇਂਗਾ , ਖਾਵੇਗਾ , ਖਾਵਾਂਗੇ, ਖਾਓਗੇ/ ਖਾਵੋਗੇ , ਖਾਣਗੇ , ਖਾਵਾਂਗੀ , ਖਾਵੇਂਗੀ , ਖਾਵੇਗੀ / ਖਾਏਗੀ , ਖਾਵਾਂਗੀਆਂ , ਖਾਵੋਗੀਆਂ , ਖਾਣਗੀਆਂ

These forms will get generated automatically through program using verb paradigm file.

Pronoun paradigms: A pronoun is a kind of noun, but its function is different from noun. The pronouns in Punjabi are classified as follows:

Root	del char	Generated Form
wusIz	3	wusIz
wusIz	3	wuhAnUM
wusIz	3	wuhAde woz
wusIz	3	wuhAde viYca
wusIz	3	wuhAde 'we
wusIz	3	wuhAde
wusIz	3	wuhAdI
wusIz	3	wuhAde waYka

2. Postposition disambiguator

Punjabi postpositions are similar to prepositions in English. These link noun, pronoun, and phrases to other parts of the sentence. Some Punjabi postpositions are ਨੇ ne, ਨੂੰ nUM, ਉੱਤੇ uYwe 'over', ਦਾ xA 'of', ਕੋਲੋਂ koloz 'from', ਨੇੜੇ

nedZe ‘near’, ਲਾਗੇ lAge ‘near’ etc. In Punjabi, postpositions follow the noun or pronoun unlike English, where these precede the noun or pronoun, and thus termed prepositions. This module maps monosemous or polysemous prepositions in English with lexical postpositions in Punjabi.

Example1: A girl **with** beautiful eyes. < > sohaNI | KUbasUrawa ~ aVKa **vAlI** ^ ika | {} ~ ladZakI

Example 2: The kid is playing **with** letters. < > baccA aVKara **nAla** Keda rihA hE

Here, English preposition “with” has been disambiguated with two meanings in Punjabi, “vAlI” and “nAla”. Hence, suitable addendum has been substituted in Punjabi for “with” by comparing the English sentence with its Punjabi translation.

3. Generation of Verb forms: This module derives the verb forms in Punjabi using TAM (Tense, Aspect and Modality). This module is used to find out proper translation of the sentence with the proper suffix. The design has five fields:

13. Finiteness

14. Auxiliary Verb

15. Main verb type

16. Phrasal field

17. Suffix

Example: “normal”, “am”, verb_5,-1, “_ rihA_*hAz

In this “rihA_*hAz” is suffix. * before “hAz” means this word cannot change in Punjabi after translation.

English: I am playing.

Punjabi: ਮੈਂ ਖੇਡ ਰਹਿਾ ਹਾਂ (mEz Keda rihA hAz).

4. Disambiguation of to-infinitive: “to-infinitives” can be used as a noun equivalent to form the subject of the sentence. The “to-infinitive” in English has multiple mapping in Punjabi based upon the context. Eg.

(1)To retreat now would be a disgrace. <> ਪਛਿੱੇ ਹੱਟਣਾ ਬਦਨਾਮੀ ਹੋਏਗਾ (piYCe haYtaNA baxanAmI hoegA)

In this example, the pattern type is non-finite, masculine, singular, third person so in Punjabi _NA suffix is attached with verb,piYCe haYtaNA

As the object of the verb

Read the sentences given below:

(2) I like to read books. < > ਮੈਂ ਕਤਿਾਬਾਂ ਪੜ੍ਹਣਾ ਪਸੰਦ ਕਰਦਾ ਹਾਂ (mEz kiwAbAz paDNA pasaMxa karaxA hAz)

(3) They decided to send an application. <> ਉਹ ਅਰਜ਼ੀ ਭੇਜਣ ਦੇ ਲਈ ਨਰਿਣਾ ਕੀਤੇ (uha arajZI BejaNa xe laI niraNA kIwe)

In the sentences given above to read and to send are the objects of the verbs .

To infinitives as adjectives

Read the sentences given below:

That is a place to visit . <> ਉਹ ਦੇਖਣ ਦੇ ਲਈ ਇਕ ਥਾਂ ਹੈ (uha xeKaNa xe laI ika WAz hE)

It is time to go. <> ਇਹ ਜਾਣਾ ਸਮਾਂ ਹੈ (iha jANA samAz hE)

Here the infinitives function as adjectives qualifying the nouns place and time.

To infinitives as adverbs

To infinitives can modify verbs and adjectives.

They are eager to win. <> ਉਹ ^ ਜਿੱਤਣ ਦੇ ਲਈ | ਜਿੱਤਣਾ ~ ਉਤਸ਼ਾਹੀ ਹਨ (uha ^ jiYwaNa xe laI | jiYwaNA ~ uwaSAhI hana)

She is anxious to leave. <> ਉਹ ^ ਜਾਣ ਦੇ ਲਈ | ਜਾਣਾ ~ ਚਤਿਤਿ ਹੈ (uha ^ jANa xe laI | jANA ~ cizwiwa hE)

We are willing to go. <> ਅਸੀਂ ^ ਜਾਣ ਦੇ ਲਈ | ਜਾਣਾ ~ ਇੱਛਕ ਹਨ (asIz ^ jANa xe laI | jANA ~ iYCaka hana)

In the first two sentences the to-infinitives to win and to leave modify the adjectives eager and anxious respectively. In the last sentence to-infinitive “to go” modify the verb “willing”.

5.Symbol Mapping

This module does mapping of English symbols/ keywords to Punjabi keywords. It also assigns a paradigm number, which tells how the Punjabi meaning changes its form. Basically this module maps different notations used in intermediate form. Following convention for paradigm numbers have been used:

English Keyword	Punjabi keyword	Paradigm No.
# “had_been”	rihA_sI	

1

Consider a sentence, “Being paralytic, he did not go.”Its Punjabi translation is “aXarazga hoNa xe kAraNa uha nahIz giA”. Here “hoNa xe kAraNa” is the mapping pattern substituted for adjectice “being”.. like:

“adjbeing”,”hoNa xe kAraNa”,7

Another example,

(1) “adj_yet”: He is good **yet** unsuccessful. uha ^ caMgA | vaXIA ~ **Pera BI** asaPala hE

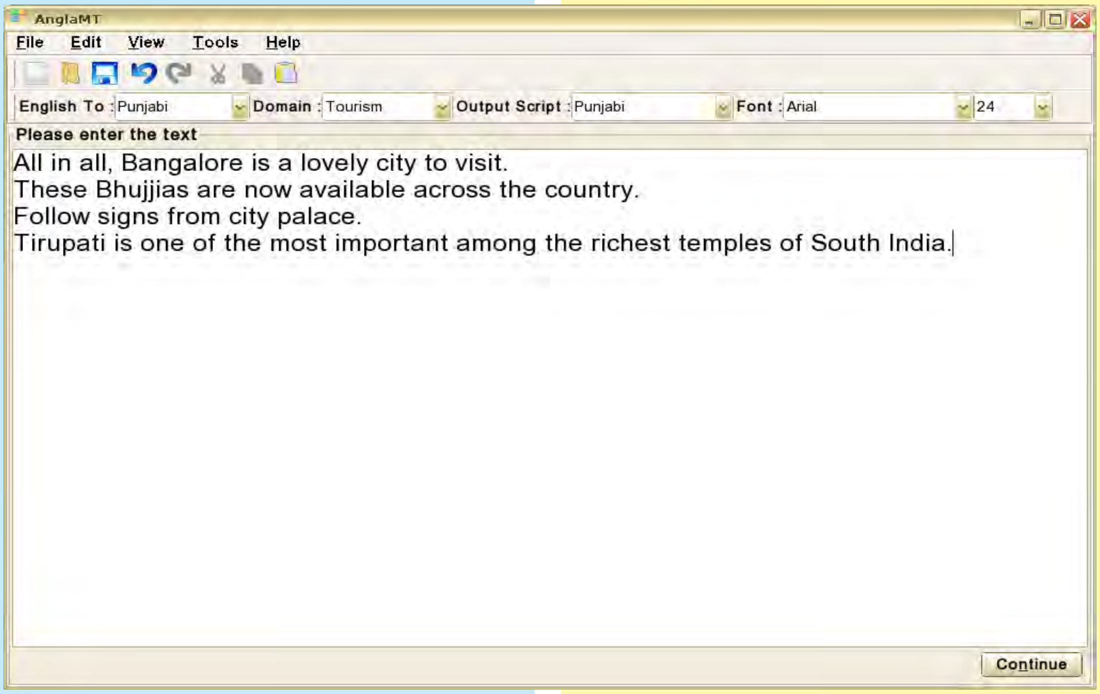
2. “had_been_never”: We **had been** never friends **asIz kaxe nahIz** xoswa rihe sana

(V) Raw Example base: AnglaPunjabi besides using all the modules of AnglaBharati also makes use of an abstract example base for translating frequently encountered noun phrases and verb phrasal. The example-base is statistically derived from the corpus. Ambiguities in the meanings of the verb phrasal are also resolved using an appropriate distance function in the example-base (Bhandari, Sinha and Jain,2002). The data files of Example base contain acronyms and headings. The appropriate Punjabi translation of acronyms and headings are incorporated in the data files.

Example1: To Err is human#manuYKa galawIAz xA puwalA hE
Example 2: VLSI#vIelaesaAI

(B) Interface: The AnglaPunjabi system aims to design, develop and deploy a Machine Translation (MT) System from English to Punjabi Language in Tourism and Health Domains. As a result, two versions of interfaces were developed (i) web version and (ii) desktop version. This paper explains Desktop interface only.

Desktop GUI: This GUI has been designed for users working on standalone PCs. After loading the system, an editor will be displayed (Fig.1). It contains many options for text editing and contains many user friendly menu items. In this GUI, user can input sentence(s) directly or open a file from the desired directory in the PC by using the "Open" menu item. Facility for English spell checking is also provided in the software. The spell checker can be used in real time. After inputting the sentence(s) or file for translation click on Translate menu item and sentences will be translated as shown in Fig.2. The resulting window will contain these sentences arranged in grid format . Fig.2 shows the translation window where all sentences will have their corresponding translations along with multiple alternatives, if any. From the multiple alternatives user will select the best possible sentence and still if not satisfied with the output can do the correction using Punjabi keyboard as shown in Fig.3. Translations can be saved in UNICODE format as shown in Fig.4

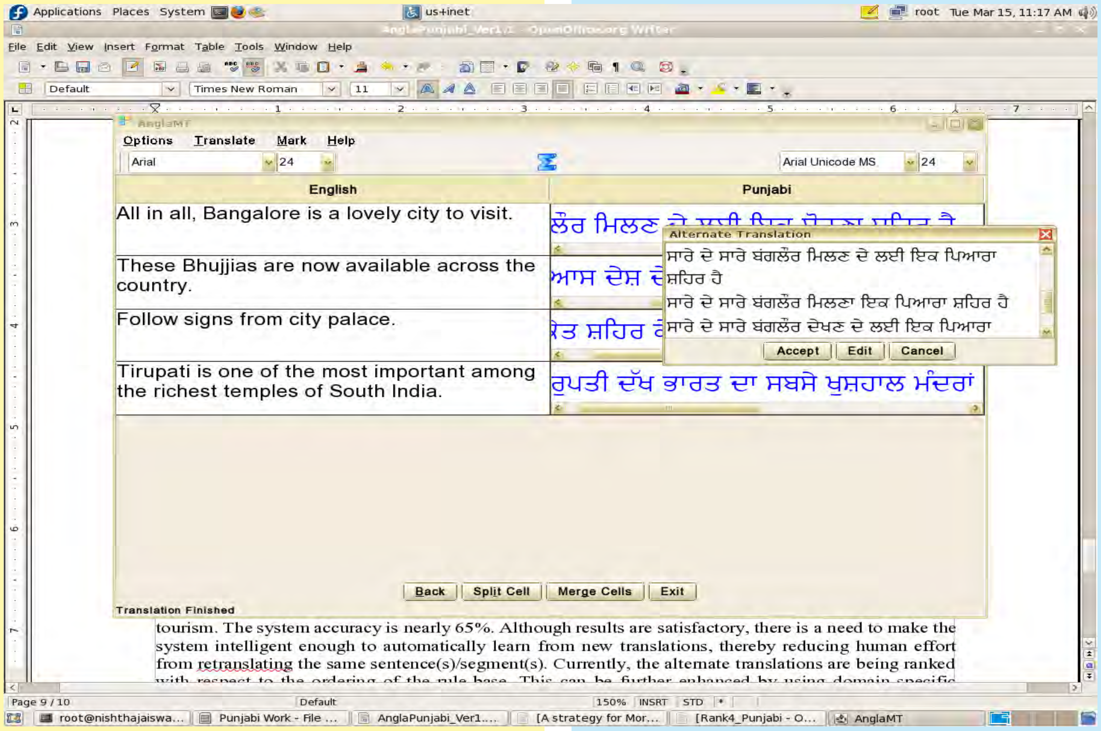


English To : Punjabi
Domain : Tourism
Output Script : Punjabi
Font : Arial
24

Please enter the text

All in all, Bangalore is a lovely city to visit.
These Bhujjias are now available across the country.
Follow signs from city palace.
Tirupati is one of the most important among the richest temples of South India.

Fig.1 Input Sentence(s) Screen


Applications Places System
us+inet
root Tue Mar 15, 11:17 AM
AnglaPunjabi_Ver1.1 - OpenOffice.org Writer
File Edit View Insert Format Table Tools Window Help
Default
Times New Roman
11
AnglaMT
Options Translate Mark Help
Arial 24
Arial Unicode MS 24
Alternate Translation
Accept Edit Cancel
Back Split Cell Merge Cells Exit
Translation Finished
Page 9 / 10
Default
150% INSRT STD *
root@nishthajaiswa... Punjabi Work - File ... AnglaPunjabi_Ver1.... [A strategy for Mor... [Rank4_Punjabi - O... AnglaMT

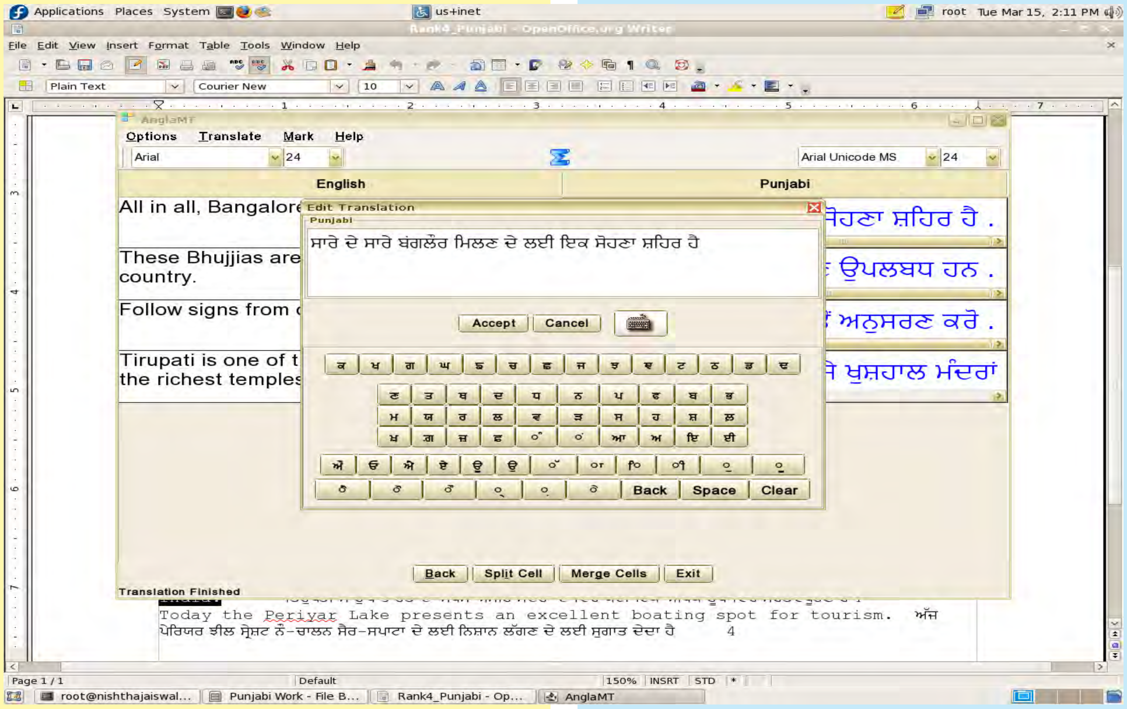
English	Punjabi
All in all, Bangalore is a lovely city to visit.	ਲੌਰ ਮਿਲਣ
These Bhujjias are now available across the country.	ਆਸ ਦੇਸ਼ ਦੇ
Follow signs from city palace.	ਰੇਤ ਸ਼ਹਿਰ ਕੇ
Tirupati is one of the most important among the richest temples of South India.	ਰੁਪਤੀ ਦੱਖ ਭਾਰਤ ਦਾ ਸਬਸੇ ਖੁਸ਼ਹਾਲ ਮੰਦਰਾਂ

Alternate Translation

ਸਾਰੇ ਦੇ ਸਾਰੇ ਬੰਗਲੌਰ ਮਿਲਣ ਦੇ ਲਈ ਇਕ ਪਿਆਰਾ ਸ਼ਹਿਰ ਹੈ
ਸਾਰੇ ਦੇ ਸਾਰੇ ਬੰਗਲੌਰ ਮਿਲਣਾ ਇਕ ਪਿਆਰਾ ਸ਼ਹਿਰ ਹੈ
ਸਾਰੇ ਦੇ ਸਾਰੇ ਬੰਗਲੌਰ ਦੇਖਣ ਦੇ ਲਈ ਇਕ ਪਿਆਰਾ

tourism. The system accuracy is nearly 65%. Although results are satisfactory, there is a need to make the system intelligent enough to automatically learn from new translations, thereby reducing human effort from retranslating the same sentence(s)/segment(s). Currently, the alternate translations are being ranked with respect to the ordering of the rule base. This can be further enhanced by using domain specific

Fig 2. Translation Screen


Applications Places System
us+inet
root Tue Mar 15, 2:11 PM
Rank4_Punjabi - OpenOffice.org Writer
File Edit View Insert Format Table Tools Window Help
Plain Text
Courier New
10
AnglaMT
Options Translate Mark Help
Arial 24
Arial Unicode MS 24
Edit Translation
Punjabi
Accept Cancel
ਕ ਖ ਗ ਘ ਙ ਚ ਛ ਜ ਝ ਞ ਟ ਠ ਡ ਢ
ਣ ਤ ਥ ਦ ਧ ਨ ਪ ਫ ਬ ਭ
ਮ ਯ ਰ ਲ ਵ ੜ ਸ ਹ ਸ਼ ਲ਼
ਖ਼ ਗ਼ ਜ਼ ਫ਼ ੰ ਂ ਆ ਅ ਇ ਈ
ਔ ਓ ਐ ਏ ਊ ਉ ੱ ਾ ਿ ੀ ੁ ੂ
ੌ ੋ ੈ ੍ ਼ ੇ Back Space Clear
Back Split Cell Merge Cells Exit
Translation Finished
Page 1 / 1
Default
150% INSRT STD *
root@nishthajaiswal... Punjabi Work - File B... Rank4_Punjabi - Op... AnglaMT

English	Punjabi
All in all, Bangalore	ੋਹਣਾ ਸ਼ਹਿਰ ਹੈ .
These Bhujjias are country.	ਣ ਉਪਲਬਧ ਹਨ .
Follow signs from c	ਾਂ ਅਨੁਸਰਣ ਕਰੋ .
Tirupati is one of t the richest temples	ੇ ਖੁਸ਼ਹਾਲ ਮੰਦਰਾਂ

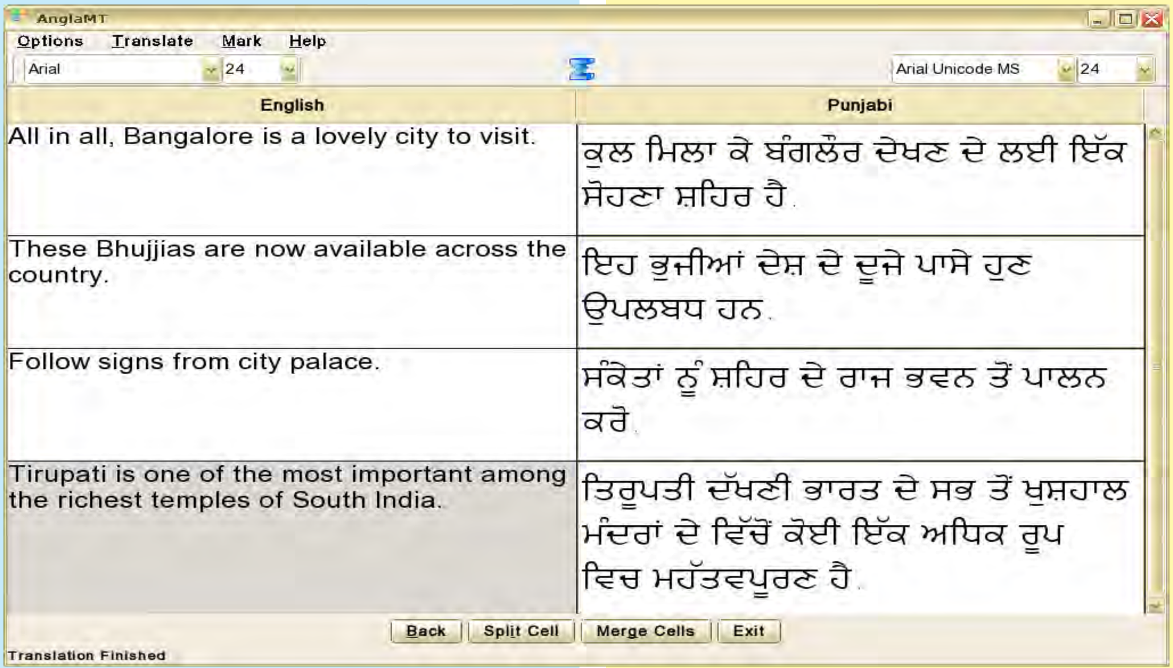
Edit Translation

Punjabi

ਸਾਰੇ ਦੇ ਸਾਰੇ ਬੰਗਲੌਰ ਮਿਲਣ ਦੇ ਲਈ ਇਕ ਸੋਹਣਾ ਸ਼ਹਿਰ ਹੈ

```
Today the Periyar Lake presents an excellent boating spot for tourism.  ਅੱਜ
ਪੇਰਿਯਰ ਝੀਲ ਸ੍ਰੇਸ਼ਟ ਨੌ-ਚਾਲਨ ਸੈਰ-ਸਪਾਟਾ ਦੇ ਲਈ ਨਿਸ਼ਾਨ ਲੱਗਣ ਦੇ ਲਈ ਸੁਗਾਤ ਦੇਦਾ ਹੈ        4
```

Fig. 3. Edit Translation



English	Punjabi
All in all, Bangalore is a lovely city to visit.	ਕੁਲ ਮਿਲਾ ਕੇ ਬੰਗਲੌਰ ਦੇਖਣ ਦੇ ਲਈ ਇੱਕ ਸੋਹਣਾ ਸ਼ਹਿਰ ਹੈ.
These Bhujjias are now available across the country.	ਇਹ ਭੁਜੀਆਂ ਦੇਸ਼ ਦੇ ਦੂਜੇ ਪਾਸੇ ਹੁਣ ਉਪਲਬਧ ਹਨ.
Follow signs from city palace.	ਸੰਕੇਤਾਂ ਨੂੰ ਸ਼ਹਿਰ ਦੇ ਰਾਜ ਭਵਨ ਤੋਂ ਪਾਲਨ ਕਰੋ.
Tirupati is one of the most important among the richest temples of South India.	ਤਿਰੂਪਤੀ ਦੱਖਣੀ ਭਾਰਤ ਦੇ ਸਭ ਤੋਂ ਖ਼ੁਸ਼ਹਾਲ ਮੰਦਰਾਂ ਦੇ ਵਿੱਚੋਂ ਕੋਈ ਇੱਕ ਅਧਿਕ ਰੂਪ ਵਿਚ ਮਹੱਤਵਪੂਰਣ ਹੈ.

Translation Finished

Fig 4. Output

(C) Conclusion

AnglaPunjabi System accepts unconstrained texts. The text may be made up of headings, texts under quotes/parenthesis, currencies, numerals, roman numbers. It has user friendly graphical interface for both desktop and web based. Current version of AnglaPunjabi has been tuned to two domain-health and tourism. The system accuracy is nearly 65%. Although results are satisfactory, there is a need to make the system intelligent enough to automatically learn from new translations, thereby reducing human effort from retranslating the same sentence(s)/segment(s). Currently, the alternate translations are being ranked with respect to the ordering of the rule-base. This can be further enhanced by using domain specific information and target language statistics. The alternate translations can be ranked based on hidden Markov model of Punjabi in the specific domain. For each alternate translation, the language model yields a figure of merit reflecting preferences for style and lexical choice. AnglaPunjabi system has been web enabled and is available at URL: **http://tdil-dc.in** for free translation.