

# Development of domain based Hindi to Punjabi Machine Translation System - Scope and Challenges

Manish Kumar  
Research Scholar  
Punjab Technical University  
Jalandhar  
[manish21578@gmail.com](mailto:manish21578@gmail.com)

Vishal Goyal  
Assistant Professor  
Department of Computer Science  
Punjabi University, Patiala  
[vishal.pup@gmail.com](mailto:vishal.pup@gmail.com)

Jaskaran Bhullar  
Associate Professor  
MIMIT, Malout  
[bhullarjaskarn@gmail.com](mailto:bhullarjaskarn@gmail.com)

## Abstract

Technical education plays vital role in the development of any state or country. Language also plays the major role in the educational development of any region. India is a Multi-lingual country where every region has their own prominent local language, which is the primary mode of communication and is also mode of basic education. On the other hand English plays the dominating role for providing Technical education in Punjab. In this paper we study the scope and challenge to develop a domain based Machine Translation system which can translate technical terms from Hindi to Punjabi to break the language barrier.

**Keywords:** Machine Translation, Optical Character Recognition, Higher Education, Hindi, Punjabi Technical Ambiguity, Non-Technical Ambiguity

## 1. Introduction

Machine translation these days plays a vital role in breaking the language barrier for accessing information. Computer education is quite crucial part of an education system as it is need of the day. But being a multilingual country, students are getting their basic education in their mother tongue. Condition is same in case of state of Punjab where large number of students are getting their basic education in their mother tongue i.e. Punjabi. It is essential to have a system which can translate the computer related text from one language to another. There are many Hindi to Punjabi Machine translation systems. In our study we are evaluating the Hindi to Punjabi Machine Translation system developed by Goyal and Lehal [1] for translating text of computer subjects. This system is based on direct machine translation. This system is designed to translate the general text and showing very good results. Evaluation of the system shows that system gives 95% accuracy for daily news text, 80.5% for articles, 90.3%

Official Language quotes, 78.5% for blogs and 85.4% for literature[2] Accuracy level is quite good while translating general kind of Hindi text and inputs from web based data as it is web based Machine Translation System[3]. The same system is giving lower accuracy when computer related text is given to the system for translation. In this paper we would discuss the scope and various challenges for developing a domain based Machine Translation system which can translate the computer related text material from Hindi to Punjabi with higher level of accuracy.

## 2. Scope of the System

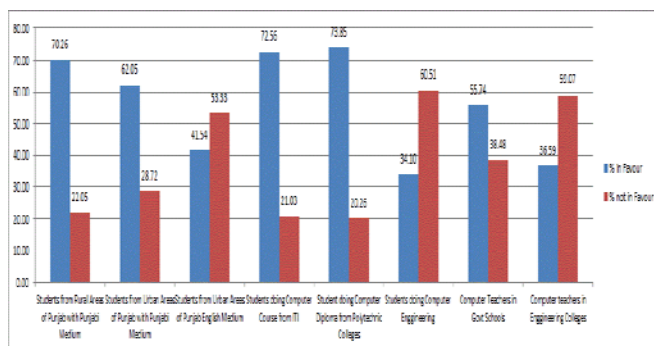
Scope of the system which can provide the computer related subject material in Punjabi through Machine Translation can be judged by taking a feedback from the two main stakeholders of any education system i.e. students and teachers. Primary data collected from both the stake holders for such purpose. Table 1 shows the feedback in terms of percentage of respondents who were in



favour and against providing computer education in Punjabi.

<b>Category</b>	<b>% in Favour</b>	<b>% not in Favour</b>
Students from Rural Areas of Punjab with Punjabi Medium	70.26	22.05
Students from Urban Areas of Punjab with Punjabi Medium	62.05	28.72
Students from Urban Areas of Punjab English Medium	41.54	53.33
Students doing Computer Course from ITI	72.56	21.03
Student doing Computer Diploma from Polytechnic Colleges	73.85	20.26
Students doing Computer Engineering	34.10	60.51
Computer Teachers in Govt. Schools	55.74	38.48
Computer teachers in Engineering Colleges	36.59	59.07

Figure 1 shows the comparison chart of percentage of students and teachers who are in favour and not in favour for providing computer education in Punjabi.



**Figure 1**

Above comparison clears that students and teachers are in favour of providing computer education in their mother tongue i.e. Punjabi. Especially students who are taking their education in Punjabi Medium are more invigorated for the same.

Other than encouraging response from the students and computer teachers, developing such a

system will defiantly lead to interest of students in getting skilled education which will further lead to more employment opportunities.

Countries like Japan, has been translating the technical text from English to Japanese for long time for providing technical education to their students in their local language[4]. At the same time advanced countries like Germany, France, Russia, China where engineering education in most of the areas is in one's mother tongue [5]. Many states of India has already initiated the process of imparting technical knowledge in their regional languages like Maulana Azad National Urdu University (MANUU), established in 1998 as a Central University with an all India Jurisdiction by an Act of Parliament with the mandate to promote and develop Urdu language and to impart higher, technical and vocational education in Urdu[6]. Tamil Nadu was one of the first states to offer degree in their mother tongue[7]. There is a possibility that in the state of Punjab also technical education at higher level could be provided in Punjabi. But in that case there will be a problem regarding availability of study material related to various computer subjects in Punjabi. Although in study material is available in Hindi in various Hindi Spoken states like Utter Pradesh, Rajasthan.

Looking into all the above factors, it can be concluded that there is a good scope of a Machine Translation System which can translate the computer related subject material from Hindi to Punjabi.

### 3. Challenges

A domain based Machine Translation system has their own challenges which can reduce the accuracy level of a generalised machine Translation System. There are certain problems which arise when text related to some particular

domain is entered as input to a Machine Translation system. In this part of study we would present various challenges which ascend to develop a machine Translation system which can translate text related to computer subjects from Hindi to Punjabi.

### 3.1 Digitisation of text related to Computer Subject

First problem arises when developing a domain based system is to arrange the text related to that particular domain in digitised form. Non availability of digitised data of computer subjects leads to arranging books of computer based subjects in Hindi and digitising them using either of two given methods.

#### 3.1.1 Manual Typing

Manual typing is quite time consuming work and encompasses high cost as well. Though it can produce more accurate data reliant on the performance of the person involved in typing, but typing such a enormous data manually is not a very feasible idea for digitising Hindi text.

#### 3.1.2 Through Hindi OCR (Optical Character Recognition)

Another way is using Hindi OCR to digitise the text. There are some OCRs available in the market which can digitise the Hindi image files to Hindi text. For our study we use Hindi OCR developed

by a German company Ind.senz[8] . As a test data input in the form of bmp images for four computer subjects was given to the OCR. There are few problems in using OCRs as well for digitising the text.

- a. OCR cannot digitised any word which is written in English.
- b. It leaves all the mathematical symbols or equations for digitisation.
- c. Images cannot be digitised using Hindi OCR
- d. Accuracy level of the digitised text through OCR is not 100%. Using Hindi OCR.

After creating text files we have to manually rectify the errors, which again involve efforts from developer.

Due to above mentioned reasons it is a big challenge to first get the computer related subject in digitised form, as it requires lots of efforts from developer.

### 3.2 Errors During Translation

Once the data has been digitised, then we can translate the text available in Hindi to Punjabi using Hindi to Punjabi Translation system. In this part of study we discovered the following errors after translation of text.

#### 3.2.1 Technical Terms Ambiguity

Every domain have their own set of terms which gives some specific meaning relate to that particular domain only, but in natural language it might have some other meaning also. It is termed as Technical Terms Ambiguity. For example

**Table 2**  
(Percentage Error in Digitisation through OCR)

Subject	Total words	Error words	% Errors
C Language	20984	512	2.43%
Data Structure	36767	326	.88%
Fundamental of Computers and IT	37325	626	1.67%
Operating System	46044	768	1.71%

- a. **Input:** इस प्रकार के एट्रीब्यूट को की-एट्रीब्यूट कहते हैं

(isa prakāra kē ēṭrībīyūṭa kō kī-ēṭrībīyūṭa kahatē hai)

**Output:** इस प्रकार के एट्रीब्यूट नੂੰ दी-एट्रीब्यूट कहिंदे है

(is prakār dē ēṭrībīyūṭ nūṁ dī - ēṭrībīyūṭ kahindē hai)

Word की-एट्रीब्यूट should have been transliterated as वी-एट्रीब्यूट, but system has translated the word की (kī) as दी (dī).

- b. **Input:** OR (और) नामक एकसक्यूसिव बिटवाइज ऑपरेटर का प्रयोग

(OR (aura) nāmaka ēksakyūsiva biṭavāija ôparēṭara kā prayōga)

**Output** :OR (अउरे) नामक ऐक्सक्यूसिव बिटवाइज ऑपरेटर का प्रयोग

(OR (atē) nāmaka ēksakiūsiv biṭvāij āprēṭar dā prayōg)

Here word और represents a bitwise operator OR which should be translated as अउरे (aur).

- c. **Input:** एक कार्ड पर पंच(Punch) रहते हैं

(ēka kārḍa para pañca (Punch) rahatē haī)

**Output** :एक कार्ड उत्ते पंच रहिंदे हन

(ikk kārḍa uttē pañj rahindē han)

Word पंच should have been transliterated as पंच (pañc) not as the literal meaning as पंज

- d. **Input:** यूजर तथा सिस्टम के मध्य एक ऑन-लाइन(On-Line)कम्यूनिकेशन स्थापित करता है

(yūjara tathā sistama kē madhy aēka ôṅ a-lāina (On-Line) kamyūnikēśana sthāpita karatā hai)

**Output:** यूजर अउरे सिस्टम के विचकार ईक आन - लकीर कंमिउनिक्शेसन सथापत करदा है

(yūjar atē sistam dē vickār ikk ān – lakīr kammiūnikēshan sathāpat karadā hai)

System translates the word ऑन-लाइन as आन - लकीर which is not correct while using word ऑन-लाइन as a technical term. In this case correct transliteration is आन – लाईन(ān – lain)

- e. **Input:** बार कोड रीडर (bār kōḍa rīḍara )

**Output:** वार वेड रीडर(vār kōḍ rīḍar)

वार is the Punjabi translation of word बार.

But in this sentence word बार is a technical term which should be transliterated as वार(bār)

- f. **Input** : मास(Mass)-स्टोरेज डिवाइस के उदाहरण हैं-टेप व डिस्क (māsa (Mass)-stōrēja ḍivāisa kē udāharaṇa haī-ṭēpa va ḍiska)

- g. **Output** :महिना - सटोरेज डिवाइस के उदाहरण हं - टेप अउरे डिस्क

(mahīnā - saṭōrēj ḍivāis dē udāharaṇ hām - ṭēp atē ḍisak)

System is translating मास as महीना, but here it should be transliterated as मास(mās)

- h. Input:** रिसोर्स एलोकेशन जब सिस्टम में मल्टीपल यूजर लॉग इन(Log In) करते हैं  
(*risōrsa ēlōkēšana jaba sistama mē malṭīpala yūjara lôga ina (Log In) karatē hai*)

**Output:** रिसोर्स एलोकेशन जदें सिस्टम विंच मल्टीपल यूजर लारा इस करदे हन

(*risōras ēlōkēshan jadōm sistam vicc malṭīpal yūjar lāg is karadē han*)

इस is the Punjabi meaning of word इन.

But here लॉग इन is a technical word which should have been transliterated as लारा-इन्(lāg-in)

- i. Input:** प्रोग्रामर उसमें गलती ढूँढता है और बग(Bug) को सही करता है

(*prōgrāmara usamē galatī dhūḍhatā hai aura бага (Bug) kō sahī karatā hai*)

**Output:** प्रोग्रामर उस विंच गलती चुँदता है अउे बगल्ला नुँ ठीक करदा है

(*prōgrāmar usvicc galatī dhūḍhtā hai atē baglā nūḡ ṭhīk karadā hai*)

बग is a technical term which means error so it should be transliterated as बग(bag)

- j. Input:** एक मल्टी वे(Multi-Way) सर्च ट्री का ऑर्डर होता है

(*ēka malṭī vē (Multi-Way) sacra ṭrī kā ôṛḍara hōtā hai*)

**Output :** ईक मल्टी उह सरच टरी दा आरडर हुँदा है

(*ikk malṭī uh sarac ṭarī dā ārḍar hundā hai*)

In above case correct transliteration of मल्टी वे is मल्टी वे (malṭīvē) not मल्टी उह

### 3.2.2 Non-Technical Terms Ambiguity

Other than technical ambiguities, system gives vagueness in translation of some non-technical words as well. For example:

- a. Input:** शक्तिशाली कम्प्यूटर के आविष्कार तक का समय पाँच भागों में बांटा गया है

(*śaktiśālī kampyūṭara kē āviṣkāra taka kā samaya pāca bhāgō mē bāṅṭā gayā hai*)

**Output :** शक्तिशाली कंप्यूटर के खेज उँक दा सभां पंज भुँजिआ विंच बाँटा गिआ है

(*shaktiśālī kampiūṭar dē khōj takk dā samāḡ pañ bhajjiā vicc bāṅṭā giā hai*)

Word भागों in this sentence means “parts” and not “running”. So correct translation is भागां(bhāgām)

- b. Input:** एक ग्राहक व उसके खाते के बीच सम्बंध तभी बनेगा

(*ēka grāhaka va usakē khātē kē bīca sambandha tabh ībanēgā*)

**Output:** ईक ग्राहक अउे उसदे खाँदेदे विंच संबंध उँदें बनेगा

(*ikk gāhak atē usdē khāndē dē vicc sambandh udōḡ baṅēgā*)

Word खाते has been misunderstood as “eating” instead of “account”. So it has been translated as खाँदे, but correct translation is खाँदे(khātē)

c. **Input** : सामान्य खाताबही (sāmānya khātābahī)

**Output:** ਇੱਕੋਜਿਰੇ ਖਾਤਾਬਹੀ (ikkōjihē khātābhī)

Word सामान्य has been translated as ਇੱਕੋ ਜਿਰੇ, but it should have been translated as आम(ām)

d. **Input:** सिस्टम में प्री प्रोसेसर नामक प्रोग्राम अपने आप ही ट्रांसलेशन वाले चरण के शुरू होने से पहले क्रियान्वित होता है

(sistama mẽ prīprōsēsara nāmaka prōgrām aapanē āpahī trāsalēsana vālē caraṇa kē śurū hōnē sē pahalē kriyānvita hōtā hai)

**Output:** ਸਿਸਟਮ ਵਿੱਚ ਪ੍ਰੀ ਪ੍ਰੋਸੇਸਰ ਨਾਮਕ ਪ੍ਰੋਗਰਾਮ ਆਪਣੇ ਚੁੱਸੀ ਹੀ ਟਰਾਂਸਲੇਸ਼ਨ ਵਾਲੇ ਪੜਾਅ ਦੇ ਸ਼ੁਰੂ ਹੋਣ ਵਲੋਂ ਪਹਿਲਾਂ ਕਿਰਿਆਨਵਿੱਧ ਹੁੰਦਾ ਹੈ

(sistam vicc prīprōsēsar nāmaka prōgrām āpaṇē tusī hī tarāṁslēshan vālē paṛāa dē shurū hōṇ valōṁ pahilāṁ kiriāvingh hundā hai)

Word आप here does not mean “You”, so translation as ਚੁੱਸੀ is wrong. Here आप has been used with word अपने which means “Automatically” so it should be translated as आप(āp)

e. **Input:** पीढ़ी के कम्प्यूटरों का आकार बड़ा होता (pīrhī kē kampyūṭarō kā ākāra barā hōtā)

**Output:** ਪੀੜ੍ਹੀ ਦੇ ਕੰਪਿਊਟਰਾਂ ਦਾ ਸਰੂਪ ਬਹੁਤ ਹੁੰਦਾ

(pīrhī dē kampiūṭrām dā sarūp bahut hundā)

In this case, word बड़ा means “Big” and not as “Very”. So correct translation should be ਵੱਡਾ (vaḍḍā)

### 3.2.3 Inflection Errors

Some inflection errors are observed in the system, while translating a technical noun word in plural form Hindi to Punjabi. For example:

a. **Input** : निश्चित संख्या के पिक्सेलों के ब्लॉक (niścita saṁkhyā kē pikselō kē blōka)

**Output:** ਨਿਸ਼ਚਿਤ ਗਿਣਤੀ ਦੇ ਪਿਕਸੇਲੋਂ ਦੇ ਬਲਾਕ (nishcit giṅtī dē pikselōṁ dē balāk)

b. **Input** : इसका प्रयोग हाथों द्वारा प्रिंटेड कैरेक्टरों को सीधे कम्प्यूटर

(isa kā prayōga hāthō dvārā priṅṭēḍa kairēkṭarō kō sīdhē kampyūṭara)

**Output:** ਇਸਦਾ ਪ੍ਰਯੋਗ ਹੱਥਾਂ ਦੁਆਰਾ ਪ੍ਰਿੰਟੇਡ ਕੈਰੇਕਟਰੋਂ ਨੂੰ ਸਿੱਧੇ ਕੰਪਿਊਟਰ

(is dā prayōg hatthāṁ duārā priṅṭēḍ kairēkṭarōṁ nūṁ siddhē kampiūṭar)

c. **Input:** निर्देशों के समूह-प्रोग्राम को संग्रहित किया जा सकता था

(nirdēśō kē samūha-prōgrāma kō sangrahīta kiyā jā sakatā thā)

**Output:** ਨਿਰਦੇਸ਼ੋਂ ਦੇ ਸਮੂਹ - ਪ੍ਰੋਗਰਾਮ ਨੂੰ ਸੰਗਰਹੀਤ ਕੀਤਾ ਜਾ ਸਕਦਾ ਸੀ

(nirdēshōṁ dē samūh – prōgrām nūṁ sangrahīt kītā jā sakdā sī)

d. **Input:** उनका उपयोग गणितज्ञों वैज्ञानिकों तथा इंजिनियरों तक ही सीमित था

(*unakā upayōga gaṇitajñō vaijñānikō tathā iñjiniyarō taka h īsīmīta thā*)

**Output:** ਉਨ੍ਹਾਂ ਦਾ ਵਰਤੋ ਗਣਿਤਗਿਆਵਾਂ ਵਿਗਿਆਨੀਆਂ ਅਤੇ ਇੰਜਿਨਿਅਰੋਂ ਤੱਕ ਹੀ ਸੀਮਿਤ ਸੀ

(*unhām dā vartō gaṇitgiāvām vigiānīām atē iñjinirōm takk hī sīmit sī*)

e. **Input:** एक कम्प्यूटर की सभी डिवाइसों का नियंत्रण करता है

(*ēka kampyūṭara kī sabhī ÷ivāisō kā niyantraṇa karatā hai*)

**Output:** ਇੱਕ ਕੰਪਿਊਟਰ ਦੀ ਸਾਰੇ ਡਿਵਾਇਸੋਂ ਦਾ ਕਾਬੂ ਕਰਦਾ ਹੈ

(*ikk kampiūtar dī sārē ÷ivāisōm dā kābū karadā hai*)

f. **Input:** यह कन्सਟ੍ਰੇन्ट दो रिलेशनों के बीच निर्धारित होता है

(*yaha kansatrēṇṭa dō rilēśanō kē bīca nirdhārita hōtā hai*)

**Output:** ਇਹ ਕੰਸਟਰੇਂਟ ਦੋ ਰਿਲੇਸ਼ਨੋਂ ਦੇ ਵਿੱਚ ਨਿਰਧਾਰਤ ਹੁੰਦਾ ਹੈ

(*ih kaṁṣṭarēṇṭ dō rilēshnōm dē vicc nirdhārat hundā hai*)

g. **Input:** उन कर्मचारियों के नाम को प्राप्त कीजिये जो कि उन सारे प्रोजेक्टों

(*una karmacāriyō kē nāma kō prāpta kījiyē jō ki una sārē prōjēktō*)

**Output:** ਉਨ੍ਹਾਂ ਕਰਮਚਾਰੀਆਂ ਦੇ ਨਾਮ ਨੂੰ ਪ੍ਰਾਪਤ ਕਰੋ ਜੋ ਕਿ ਉਨ੍ਹਾਂ ਸਾਰੇ ਪ੍ਰੋਜੈਕਟੋ

(*unhām karamcāriām dē nām nūm prāpat karō jō ki unhām sārē prōjēktō*)

System is translating technical noun terms ending with sound ੋ as ੋ, but it should be

translated as ੋ. So पिकसेलों should be translated as ਪਿਕਸੇਲਾਂ(**piksēlām**), ਕੈਰੇਕਟਰੋਂ as ਕੈਰੇਕਟਰਾਂ(**kairēkaṭrām**) and so all other terms.

### 3.2.4. Word out of Vocabulary

In a Machine Translation system, only technical terms or noun entries should be transliterated. It is observed that system is transliterating some normal Hindi words as well into Punjabi as these words are not found in the bi-lingual corpora.

a. **Input :** कई दशक पहले आविष्कारित ये कम्प्यूटर

(*kaī daśaka pahalē āviṣkārita yē kampyūṭara*)

**Output:** ਕਈ ਦਸ਼ਕ ਪਹਿਲਾਂ ਆਵਿਸ਼ਕਾਰਿਤ ਇਹ ਕੰਪਿਊਟਰ

(*kaī dashak pahilām āvishkārit ih kampiūṭar*)

Word आविष्कारित should be translated as ਖੋਜੋ(khōjē)

b. **Input :** अपनी दीप्ति जल्दी-जल्दी खो देते हैं

(*apanī dīpti jaldī-jaldī khō dētē haī*)

**Output:** ਆਪਣੀ ਦੀਪਤੀ ਜਲਦੀ - ਜਲਦੀ ਖੋਹ ਦਿੰਦੇ ਹਨ (*āpṇī dīpatī jaldī – jaldī khōh dindē han*)

Word ਦੀप्ति should be translated as ਚਮਕ(camak)

c. **Input:** एकसिंगल यूजर ऑपरੇटिंग सिस्टम

(*ēka singala yūjar aōparēṅga sistama*)

**Output:** ਏਕ ਸਿੰਗਲਯੂਜਰ ਆਪਰੇਟਿੰਗ ਸਿਸਟਮ

(*ēk siṅgal yūjar āprēṭiṅg sistam*)

Word एक should be translated as ਇੱਕ  
(ikk)

d. **Input:** संसाधनों का आबंटन

(*saṁsādhanō kā ābaṅṭana*)

**Output:** ਸੰਸਾਧਨਾਂ ਦਾ ਆਬੰਟਨ

(*saṁsādhnāṁ dā ābṅṭan*)

Word आबंटन should be translated as ਵੰਡ  
(vaṅḍ)

e. **Input:** इस डाटाबेस को वर्णित करने के लिए

सर्वप्रथम हमें प्रत्येक फाइल के रिकॉर्डों की  
बनावट निर्धारित करनी चाहिए।

(*isa ḍāṭābēsa kō varṇita karanē kē liē*  
*sarvaprathama hamē pratyēka phāila kē*  
*rikōṛḍō kī banāvaṭa nirdhārita karanī*  
*cāhiē*)

**Output:** ਇਸ ਡਾਟਾਬੇਸ ਨੂੰ ਵਰਣਿਤ ਕਰਣ ਲਈ

ਸਰਵਪ੍ਰਥਮ ਸਾਨੂੰ ਹਰ ਇੱਕ ਫਾਇਲ ਦੇ ਰਿਕਾਰਡਾਂ

ਦੀ ਬਣਾਵਟ ਨਿਰਧਾਰਤ ਕਰਣੀ ਚਾਹੀਦੀ ਹੈ

(*is ḍāṭābēs nūṁ varṇit karaṅ laī*  
*sarvaprtham sānūṁ har ikk phāil dē*  
*rikāṛḍāṁ dī baṅāvaṭ nirdhārat karṇī*  
*cāhīdī hai*)

Word सर्वप्रथम should be translated as ਸੱਭ

ਤੋਂ ਪਹਿਲਾਂ(sabbhṭōṁpahilāṁ)

f. **Input:** यद्धपि रिलेशनल एलजेब्रा में यह

बिल्कुल सीधा है

(*yaddhapi rilēśanala ēlajēbrā mē yaha*  
*bilkula sīdhā hai*)

**Output:** ਯੱਧਪਿ ਰਿਲੇਸ਼ਨਲ ਏਲਜੇਬਰਾ ਵਿੱਚ

ਇਹ ਬਿਲਕੁੱਲ ਸਿੱਧਾ ਹੈ

(*yaddhapi rilēśnal ēljēbrā vicc ih bilkull*  
*siddhā hai*)

Word यद्धपि should be translated as  
ਹਾਲਾਂਕਿ (hālāṅki)

### 3.2.5 Other Translation Errors

System gives some other translating errors as well like spelling errors, wrong entry in the lexicon. For example:

a. **Input:** सॉफ्टवेयर में सुधार

(*sōphṭavēyara mē sudhāra*)

**Output:** ਸਾਫਟਵੇਯਰ ਵਿੱਚ ਸੁਧਾਰ

(*sāphṭa vēyar vicc sudhār*)

There is spelling problem in this case word  
ਸੌਫਟਵੇਯਰ should be written as ਸਾਫਟਵੇਅਰ  
(sāphṭavēar) in Punjabi

b. **Input:** क्रांतिकारी आविष्कार से सम्पूर्ण

सीपीयू का परिपथ

(*krāntikārī āviṣkāra sē sampūrṇa sīpīyū kā*  
*paripatha*)

**Output:** ਕ੍ਰਾਂਤੀਵਾਦੀ ਖੋਜ ਵਲੋਂ ਸੰਪੂਰਣ ਸੀਪੀਊ ਦਾ  
ਪਰਿਪਥ

(*krāntivādī khōj valōṁ sampūraṅ sīpīūdā*  
*paripath*)

There is spelling problem in this case word  
ਸੀਪੀਊ should be written as ਸੀਪੀਯੂ(sīpīyū)  
in Punjabi

c. **Input:** डाटा और प्रोग्राम एक धारा में

क्रियान्वित करने की पारम्परिक

(*ḍāṭā aura prōgrāma ēka dhārā mē*  
*kriyānvita karanē kī pāramparika*)

**Output:** ਡਾਟਾ ਅਤੇ ਪ੍ਰੋਗਰਾਮ ਇੱਕ ਧਾਰਾ ਵਿੱਚ

ਕਿਰਿਆਵਿੱਧ ਕਰਣ ਦੀ ਰਿਕਾਇਤੀ



(*dātā atē prōgrām ikk dhārā vicc  
kiriāvīṅgh karan dī hikāitī*)

Wrong translation of पारम्परिक. it should  
have been translated as रिवायती(rivāitī)

**d. Input:** फ़लो ऑफ़ इनफॉर्मेशन

(*phlō ōpha inaphōrmēsana*)

**Output:** ਫਲਾਂ ਆਫ ਇਨਫਾਰਮੇਸ਼ਨ

(*phalām āph inphārmēshan*)

फ़लो should have been translated as  
ਫਲੇ(phalō)

**e. Input:** होटलों के पते आदि

(*hōṭalō kē patē ādi*)

**Output:** ਹੋਟਲਾਂ ਦੇ ਪਦੇ ਆਦਿ

(*hōṭlām dē padē ādi*)

Correct translation for पते is पते(pate) and  
not पदे

**f. Input:** फ़्लॉपी या हार्डडिस्क से हटाना है

(*phlōpī yā hārḍa ḍiska sē haṭānā hai*)

**Output:** ਫਲਾਪੀ ਜਾਂ ਹਾਰਡਡਿਸਕ ਵਲੋਂ ਵਿਡਾਰਨ  
ਚੈ

(*phalāpī jāṃ hārḍa ḍisak valōṃ viḍāran  
hai*)

Translation for word हटाना should be  
ਹਟਾਨਾ (haṭānā)

**g. Input:** यदि स्वैप कैश एन्ट्री नॉन-जीरो है तो

(*yadi svaipa kaiśa ēṅṭrī nōna-jīrō hai tō*)

**Output:** ਜੇਕਰ ਸਵੈਪ ਕੈਸ਼ ਏਂਟਰੀ ਨਾਨ - ਜੀਰੋ ਹੈ  
ਤਾਂ

(*jēkar savaip kaish ēṅṭrī nān - jīrām hai  
tām*)

Word जीरो has been wrongly translated as  
ਜੀਰਾਂ. It should be translated as ਜੀਰੇ (jīrō)

All the above mentioned errors are needs to  
be fixed using different approaches. Some of  
the errors could be resolved by developing a  
domain based lexicon, as technical  
dictionaries are the principal aid to technical  
translation[9]. In some other kind of errors  
algorithms will be required to develop to  
remove the ambiguities.

#### 4 Limitations

Present study is based on evaluating the  
output of five computer subjects translated  
through Hindi to Punjabi machine Translation  
system developed by Goyal and Lehal [1]. if  
input of some other computer related subject  
is given or some other Hindi to Punjabi  
Machine translation system is used then there  
could be change in number and type of errors.

#### 5 Conclusion

After considering all the facts mentioned in  
this study, we can conclude that lots of efforts  
are required while developing a domain based  
translation system like developing a system  
for translating computer related text. There  
arespecific domain based problems which  
needs to be undertook very carefully.

Looking into the response from the  
stakeholders and future prospectus of  
promoting local language as mode of  
technical education, there is enormous scope  
of developing such systems.

**References:**

1. Goyal V, Lehal G.S. (2011) "Hindi to Punjabi Machine Translation System" Proceedings of the ACL-HLT 2011, pp 1-6
2. Goyal V, Lehal G.S. (2009) "Evaluation of Hindi to Punjabi Machine Translation System" International Journal of Computer Science, Vol. 4 No 1.
3. Goyal V, Lehal G.S. (2009) "Web based Hindi to Punjabi Machine Translation System" Journal of Emerging Technologies in web Intelligence, Vol 2, No. 2.
4. Nagao M. et all (1982) "An English Japanese Machine Translation System of the Titles of Scientific and Engineering Papers",Coling 82: proceedings of the Ninth International Conference on Computational Linguistics, Prague, July 5-10, 1982,pp 245-252
5. [http://en.wikipedia.org/wiki/Medium\\_of\\_instruction#cite\\_note-20](http://en.wikipedia.org/wiki/Medium_of_instruction#cite_note-20) Accessed on 10/06/12
6. The Maulana Azad National Urdu University Act, 1996 (No. 2 of 1997) Extracted from website [http://www.manuu.ac.in/University\\_Act.pdf](http://www.manuu.ac.in/University_Act.pdf) accessed on 15/06/12
7. [www.tamiltribune.com/99/0903.html](http://www.tamiltribune.com/99/0903.html) [Accessed on 15/06/12]
8. <http://www.indsenz.com/int/index.php> [Accessed on 04/07/12]
9. Oubine I. I., Tikhomirov B. D.(1982) "Machine Translation System and Computer Dictionaries in the Information Service. Ways of their Development and Operations" Collings 1982.

