

Manish Kumar

Error Recognition in Translation of Computer Related Text Though General Purpose Hindi to Punjabi Machine Translation System

Manish Kumar

(Department of Computer Science, Baba Farid College, Bathinda)
manish21578@gmail.com

Abstract

A machine translation system can be a general machine translation system or it can be a domain specific machine translation system as well. A domain specific machine translation system is such a system which is developed to translate the text related to some particular domain from one language to another language. While developing a machine translation system which is specific to a particular domain then there are certain challenges related to that domain only. In this paper we have studied the different kind of errors which are generated while we translate a technical text through general machine translation system.

Introduction

Accuracy level of any general machine translation system decreases while it translates the text of some specific domain.

Main reason for this is specific terminology used in that particular domain is some time creates ambiguity while integrated with the general text of the source or target language. Since, long time researchers have contributed by developing various domain based systems to minimise the ambiguity in translation of technical terms of each domain [1-21].

Index Terms – Machine Translation System, Technical ambiguity, Non-Technical ambiguity, Inflection errors, Word out of vocabulary error, disambiguation.

2. Error Recognition

Different kinds of errors are recognised while translating the computer related text through general purpose Hindi to Punjabi Machine Translation System

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



Manish Kumar

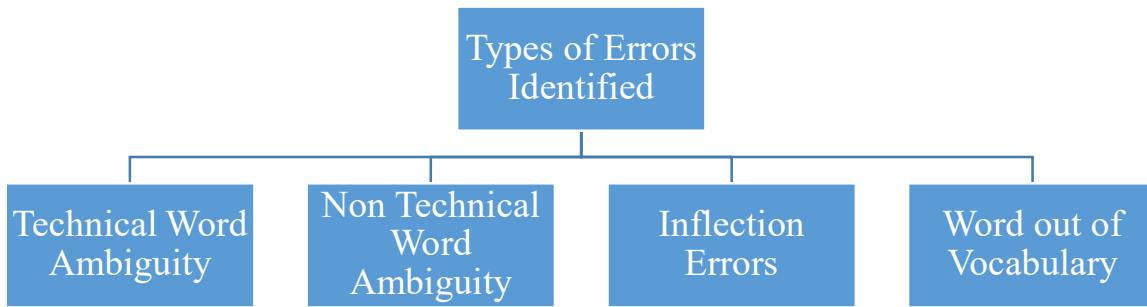


Figure 1
Categories of Identified Errors

2.1 Technical Word Ambiguity

An ambiguity is a word, phrase or sentence which has more than one meaning. In technical translation there are certain words, phrases which may not be ambiguous otherwise but in the technical sense. Handling any kind of ambiguity is the most difficult part of developing a translation system. During process of evaluating the translation done by general purpose Hindi to Punjabi machine translation system

many technical ambiguous terms were identified which if not rectified, changes the semantics of the sentence absolutely. This section describes various technical ambiguous words, their different appearance in text and also the wrongly translated text by existing Hindi to Punjabi Machine Translation system. Following are examples of three most ambiguous words of technical translation.

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



Manish Kumar

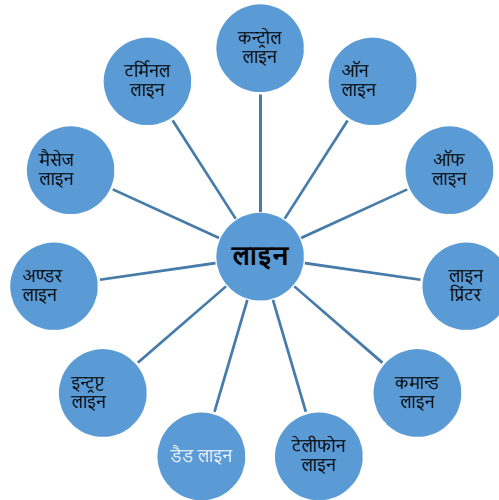


Figure 2

Different Variants of लाइन (lāin)

Table 1 shows the actual translation of the word by Hindi to Punjabi Machine Translation system. Hindi to Punjabi

Machine translation system is translating the word instead of transliteration. [22-25]

Table 1
Translation of लाइन (lāin) by baseline MT System

Hindi Text Input	Roman Script of Hindi Text	Translation by Hindi to Punjabi Translation System	Roman Script of Translated Test
ऑन लाइन	<i>ān lāin</i>	ਆਨ ਲਕੀਰ	<i>ān lakīr</i>
ऑफ लाइन	<i>āph lāin</i>	ਆਫ ਲਕੀਰ	<i>āph lakīr</i>
लाइन प्रिंटर	<i>lāin priṅṭar</i>	ਲਕੀਰ ਪ੍ਰਿੰਟਰ	<i>lakīr priṅṭar</i>
टेलीफोन लाइन	<i>ṭēlīphōn lāin</i>	ਟੇਲੀਫੋਨ ਲਕੀਰ	<i>ṭēlīphōn lakīr</i>
डैड लाइन	<i>ḍaiḍ lāin</i>	ਡੈਡ ਲਕੀਰ	<i>ḍaiḍ lakīr</i>

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



Manish Kumar

कन्ट्रोल लाइन	<i>kanṭrōl lāin</i>	ਕੰਟਰੋਲ ਲਾਈਨ	<i>kanṭrōl lakīr</i>
कमान्ड लाइन	<i>kamāṇḍ lāin</i>	ਕਮਾਂਡ ਲਾਈਨ	<i>kamāṇḍ lakīr</i>
इन्ट्रप्ट लाइन	<i>inṭrapṭ lāin</i>	ਇੰਟਰਪਟ ਲਾਈਨ	<i>inṭrapaṭ lakīr</i>
अण्डर लाइन	<i>aṇḍar lāin</i>	ਅੰਡਰ ਲਾਈਨ	<i>aṇḍar lakīr</i>
मैसेज लाइन	<i>maisēj lāin</i>	ਮੈਸੇਜ ਲਾਈਨ	<i>maisēj lakīr</i>
टर्मिनल लाइन	<i>ṭarminal lāin</i>	ਟਰਮਿਨਲ ਲਾਈਨ	<i>ṭarminal lakīr</i>



Figure 3
Different Variants of की (*kī*)

Table 2 shows the translation of word की (*kī*) done by Hindi to Punjabi translation system. System was translating word की (*kī*) as ਦੀ(*dī*). When word की (*kī*) is

prefixed or post fixed with some other technical terms it should have been transliterated as **ਕੀ** (*kī*).

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



Manish Kumar

Table 2
Translation of की (kī) by baseline MT System

Hindi Text Input	Roman Script of Hindi Text	Translation by Hindi to Punjabi Translation System	Roman Script of Translated Test
की-पैड	<i>kī-paiḍ</i>	ਦੀ-ਪੈਡ	<i>dī-paiḍ</i>
की-बोर्ड	<i>kī-bōrḍ</i>	ਦੀ-ਬੋਰਡ	<i>dī-bōraḍ</i>
की-फील्ड	<i>kī-phīlḍ</i>	ਦੀ - ਫੀਲਡ	<i>dī - phīlaḍ</i>
की-एट्रीब्यूट	<i>kī-ēṭrībīyūṭ</i>	ਦੀ-ਏਟਰੀਬਿਊਟ	<i>dī-ēṭrībīyūṭ</i>
प्राइमरी-की	<i>prāimrī-kī</i>	ਪ੍ਰਾਇਮਰੀ-ਦੀ	<i>prāimrī-dī</i>
की-कन्स्ट्रेंट	<i>kī-kanṣṭrēṅṭ</i>	ਦੀ-ਕੰਸਟਰੇਂਟ	<i>dī-kaṁṣṭarēṅṭ</i>
फॉरेन-की	<i>phārēn-kī</i>	ਫਾਰੇਨ- ਦੀ	<i>phārēn- dī</i>
की-वैल्यू	<i>kī-vaiḷyū</i>	ਦੀ - ਵੈਲਿਊ	<i>dī - vaiḷiū</i>

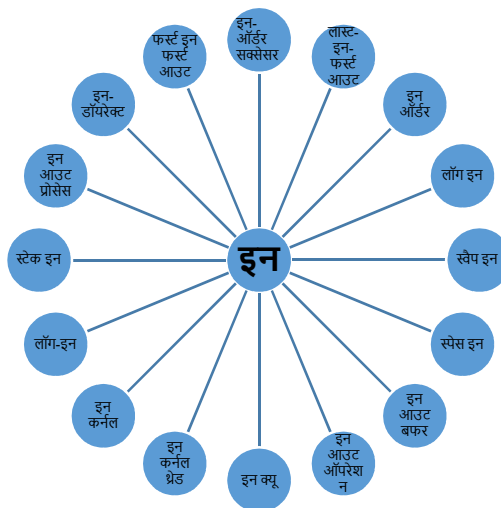


Figure 4

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),

ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



Manish Kumar

Different Variants of इन (in)

Hindi to Punjabi Machine translation system was translating this word as ਇਸ (is), which is correct while we talk about translation of general Hindi text. But just like word की when word इन (in) is

prefixed or post fixed with some other technical word it should be transliterated instead of translation. Table shows the incorrect translation by Hindi to Punjabi Machine Translation system.

Table 3
Translation of इन (in) by baseline MT System

Hindi Text Input	Roman Script of Hindi Text	Translation by Hindi to Punjabi Translation System	Roman Script of Translated Test
इन ऑर्डर	<i>in ārdar</i>	ਇਸ ਆਰਡਰ	<i>is ārdar</i>
ਲਾੱਗ इन	<i>lāg in</i>	ਲਾਗ ਇਸ	<i>lāg is</i>
स्वैप इन	<i>svaip in</i>	ਸਵੈਪ ਇਸ	<i>savaip is</i>
स्पेस इन	<i>spēs in</i>	ਸਪੇਸ ਇਸ	<i>sapēs is</i>
इन आउट बफर	<i>in āuṭ baphar</i>	ਇਸ ਆਉਟ ਬਫਰ	<i>is āuṭ baphar</i>
इन आउट ऑपरेशन	<i>in āuṭ āprēshan</i>	ਇਸ ਆਉਟ ਆਪਰੇਸ਼ਨ	<i>is āuṭ āprēshan</i>
इन क्यू	<i>in kyū</i>	ਇਸ ਕਿਊ	<i>is kiū</i>
इन कर्नल थ्रेड	<i>in karnal thrēḍ</i>	ਇਸ ਕਰਨਲ ਥਰੇਡ	<i>is karnal tharēḍ</i>
इन कर्नल	<i>in karnal</i>	ਇਸ ਕਰਨਲ	<i>is karnal</i>
ਲਾੱਗ-इन	<i>lāg-in</i>	ਲਾਗ - ਇਸ	<i>lāg - is</i>
स्टेक इन	<i>sṭēk in</i>	ਸਟੇਕ ਇਸ	<i>saṭēk is</i>
इन आउट प्रोसेस	<i>in āuṭ prōsēs</i>	ਇਸ ਆਉਟ ਪ੍ਰੋਸੇਸ	<i>is āuṭ prōsēs</i>
इन-ਡਾਇरेक्ट	<i>in-ḍāyṛēkṭ</i>	ਇਸ - ਡਾਇਰੇਕਟ	<i>is - ḍāirēkaṭ</i>

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



2.2 Non-Technical Word Ambiguity

Other than technical ambiguities discussed in previous section, there were several non technical ambiguous words were also found during translation of computer text using baseline Hindi to Punjabi Machine Translation System. Non technical ambiguities were not specific to some particular text or subject but these were some very basic terms which contain more than one sense when we translate them from Hindi to Punjabi.

बड़ा (*barā*) : This word is non technical word which has two meaning. One is **बहुत** (*bahut*) which means ‘Very’ and another meaning is **विशाल** (*vishāl*) which means ‘large’. In certain cases system was not translating the word as required.

Input : आकार **बहुत बड़ा** हुआ करता था
(*kār bahut barā huā karatā thā*)

Output : **सबूब बहुत बहुत** होइआ करदा सी
(*arūp bahut bahut hōiā karadā sī*)

आप (*āp*): Word **आप** (*āp*) is another non technical ambiguous word. In general word **आप** (*āp*) means **तुसी** (*tusī*) which is correct translation. But when word **आप** (*āp*) is pre-fixed with word **अपने** (*apnē*) it makes a composite word **अपने आप** (*apnē āp*). Correct translation of this composite word should be **आपਣੇ आप** (*apnē āp*) and not as **आपਣे तुसी** (*āpāṇē tusī*).

Input: सिस्टम में प्री प्रोसेसर नामक प्रोग्राम अपने **आप ही** ट्रांसलेशन वाले चरण के शुरू होने से पहले क्रियान्वित होता है
(*sisṭam mēṃ prī prōsēsar nāmak prōgrām apnē āp hī trāṃslēshan vālē caraṇ kē shurū hōnē sē pahlē kriyānvit hōtā hai*)

Output: सिस्टम ਵਿੱਚ ਪ੍ਰੀ ਪ੍ਰੋਸੇਸਰ ਨਾਮਕ ਪ੍ਰੋਗਰਾਮ **ਆਪਣੇ ਤੁਸੀ** ਹੀ ਟਰਾਂਸਲੇਸ਼ਨ ਵਾਲੇ ਪੜਾਅ ਦੇ ਸ਼ੁਰੂ ਹੋਣ ਵਲੋਂ ਪਹਿਲਾਂ ਕਿਰਿਆਵਿੰਘ ਹੁੰਦਾ ਹੈ
(*sisṭam vīcc prī prōsēsar nāmak prōgrām āpāṇē tusī hī trāṃslēshan vālē paṛāā dē shurū hōṇ valōṃ pahilām kiriāvīngh hundā hai*)



2.3 Inflection Error

Inflection is a grammatical process which combines words and affixes to produce alternative grammatical forms of words. Inflections does not form any different lexical entry but it just shows the inflectional variant of the same word. For example, word ‘books’ is an inflectional variant of word book but it is not a new lexical entry.

Both Hindi and Punjabi are inflection based languages where some special characters represent the inflectional variant of the word. During translation from Hindi to Punjabi every term must be translated according to the inflection rule defined in the system. Although system was handling the inflection issues for all the general Hindi text, but it was not able to handle the inflection related issues while translating technical terms from Hindi to Punjabi.

Table 4 displays various inflectional errors observed in the translated text of Existing Hindi to Punjabi Machine Translation system.

Table 4
Output of Inflection Errors by baseline system

Hindi Input	Roman Script of Hindi Input	Punjabi Output	Roman Script of Punjabi Output
निर्देशों	<i>nirdēshōṃ</i>	ਨਿਰਦੇਸ਼ੋਂ	<i>nirdēshōṃ</i>
इकाइयों	<i>ikāiyōṃ</i>	ਇਕਾਇਯੋਂ	<i>ikāiyōṃ</i>
आकड़ों	<i>ākṛōṃ</i>	ਆਕੜੋਂ	<i>ākṛōṃ</i>
कैरेक्टरों	<i>kairēkṛōṃ</i>	ਕੈਰੇਕਟਰੋਂ	<i>kairēkṛōṃ</i>
पिक्सेलों	<i>piksēlōṃ</i>	ਪਿਕਸੇਲੋਂ	<i>piksēlōṃ</i>
मॉनीटरों	<i>mānīṭrōṃ</i>	ਮਾਨੀਟਰੋਂ	<i>mānīṭrōṃ</i>
मशीनें	<i>mashīnēṃ</i>	ਮਸ਼ੀਨੋਂ	<i>mashīnēṃ</i>
यूजरों	<i>yūjrōṃ</i>	ਯੂਜਰੋਂ	<i>yūjrōṃ</i>
आउटपुटों	<i>āuṭpuṭōṃ</i>	ਆਉਟਪੁਟੋਂ	<i>āuṭpuṭōṃ</i>
प्रोग्रामरों	<i>prōgrāmṛōṃ</i>	ਪ੍ਰੋਗਰਾਮਰੋਂ	<i>prōgrāmṛōṃ</i>

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



2.4. Word out of Vocabulary

Transliteration module of the Hindi to Punjabi Machine translation system handles all the word which are not part the bilingual corpus. All those words are transliterated. System should transliterate only those words which do not have word to word replacement from Hindi to Punjabi. But existing system was transliterating some those words as well which has an equivalent entry in the Punjabi corpus as well.

References

- [1] Hutchins, W.J., Somers, L., (1992), "An Introduction to Machine Translation," Academic Press, London, from web site <http://www.hutchinsweb.me.uk/> IntroMT-Contents.pdf assessed on 1/6/2013.
- [2] Hutchins, J., (1992), " Machine Translation: General Overview," from web site <http://www.hutchinsweb.me.uk/Mitkov-2003.pdf> accessed on 1/6/2013.
- [3] Rao, Durgesh,(2001), "Machine Translation in India: A Brief Survey, " SCALLA conference, Bangalore. from web site <http://elda.org/en/proj/scalla/SCALLA2001/SCALLA2001Rao.pdf> accessed on 1/6/2013.
- [4] Hutchins, W.J., (1978), "Machine Translation and Machine-aided translation," Journal of Documentation, 34(2), pp 119-159.
- [5] Hutchins, W.J. (1986), "Machine Translation: Past, Present, Future," Ellis Horwood Series in Computers and their Applications, Chichester, New York, pp 382.
- [6] Douglas, A., Lorna, B., Siety M., Lee, R., Sadler, L., (2001), "Machine Translation: An Introductory Guide", from web site <http://promethee.philo.ulg.ac.be/engdep1/download/bacIII/Arnold%20et%20al%20Machine%20Translation.pdf> accessed on 12/10/2015.
- [7] Goyal, V., (2010), "Development of a Hindi to Punjabi Machine Translation System," Ph.D. Thesis, Department of Computer Science, Punjabi University, Patiala.



Manish Kumar

- [8] Allen, J., (1995), Natural Language Understanding. Second Edition. Benjamin Cummings, Chap 9.
- [9] Bharati, A., Chaitanya, V., Sangal, R., (2004), Natural Language Processing: A Paninian Perspective, PHI. , Chap 11.
- [10] Jurafsky, D., Martin, J.H., (2000), "Speech and Language Processing. Second Edition," PHI, Chap 8.
- [11] Dostert, L.E., (1955), "The Georgetown-IBM experiment. In Locke & Booth :1955," pp.124-135.
- [12] Hutchins, J., (1992), "General Introduction and brief History," from web site <http://www.hutchinsweb.me.uk/IntroMT-1.pdf> accessed on 12/10/2015.
- [13] ALPAC, (1966), "A Report of the Automatic Language Processing Advisory Committee," Division of Behavioral Sciences, National Academy of Sciences, National Research Council Publication 1416, Washington, D.C. accessed on 01/6/2013.
- [14] George and Kumar, (2002), "Machine Translation - An Evolving Platform For Social Change," In proceedings of Symposium on Translation Support Systems, IIT Kanpur, pp. 130-137.
- [15] Hutchins, W.J., (1997), "Looking back to 1952: the first MT conference TMI-97," in proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation, July 23-25, St. John's College, Santa Fe, New Mexico, USA, pp.19-30.
- [16] Hutchins, W. J., (1994), "The Georgetown-IBM Demonstration," MT News International, pp.19-30.
- [17] Ping, K., (1996), "A Socio-semiotic Approach to Meaning in Translation," BABLE (42) 2, pp 74-83.
- [18] Reifler, E., (1961), "MT linguistics and MT lexicography," Kent, University of Washington, pp. 841-852.
- [19] Jurafsky, D., & Martin, J., (2003), "Speech and Language Processing: An

Research Cell: An International Journal of Engineering Sciences,

Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com> ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.



Manish Kumar

Introduction to speech recognition, computational linguistics and natural language processing," Prentice-Hall, New Jersey 2003.

- [20] John., H., (2009), "Multiple uses of Machine Translation and computerized translation tools," ISMTCL: International Symposium on Data Mining and Sense Mining, Machine Translation and Controlled Languages, and their application to emergencies and safety critical domains. July 1-3, Centre Tesnière, University of Franche-Comté, Besançon, France, pp. 13-20.
- [21] Toma, P., (1974), "Computer translation: in its own right," in *Kommunikations for schung and Phonetik* (Hamburg: Buske), pp.155-164.
- [22] Fujii, A., Ishikawa, T., (2001), "Japanese/English Cross language Information retrieval : Exploration of query translation and Transliteration," *Computer and the Humanities*, 35(4), pp. 389-420.
- [23] Sakai, T., Kumano, A., Manabe, T., (2002), "Generating Transliteration rules for cross language Information retrieval from Machine Translation Dictionaries," *IEEE 2002*, pp 555 – 565.
- [24] Madhvi, G.,(2005), "OM : One tool for many (Indian) Languages," *Journal of Zhejiang University Science*, 6(A), pp. 1348-1353.
- [25] Hettige, B., Karunananda, A.S., (2007), "Transliteration system for English to Sinhala Machine translation," *Second International conference on Industrial and Information Systems, ICIIS*, pp 8-11.

Research Cell: An International Journal of Engineering Sciences,

*Special Issue November 2017(ETME-17), Vol. 25, Web Presence: <http://ijoes.vidyapublications.com>ISSN: 2229-6913(Print),
ISSN: 2320-0332(Online), UGC Approved Journal (S.No.63019)*

© 2017 Vidya Publications. Authors are responsible for any plagiarism issues.

