==
Language in India www.languageinindia.com ISSN 1930-2940 Vol. 14:10 October 2014
==

Divergence Issues in English-Punjabi Machine Translation

Deepti Bhalla, M.Tech.
Nisheeth Joshi, Ph.D.
Iti Mathur, M.Sc.

==

Abstract

Machine Translation is emerging research area in the field of computer science. Language divergence problem is the key concern in machine translation. Divergence issues need to be identified carefully for their appropriate categorization. This paper discusses various types of translation divergence between a pair of natural languages i.e. English and Punjabi. The field of linguistic divergence is miscellaneous and need to be explored more for identifying its further classification. In this paper, we take DORR's classification (1993-1994) to identify different types of translation divergence in our language pair.

Keywords: Translation Divergence, Syntactic divergence, Lexical Divergence, English-Punjabi MT.

Introduction

The issue of translation divergence is a very considerable topic. Translation divergence can be defined as a problem which occurs during translation between two languages of different origin. It mainly occurs when the output translation of one language into another language is very much different from the novel translation. It is very difficult to collect the precise translation of an input sentence without analyzing the nature of translation divergence. The syntax is the one which can be used to realize the semantics of the sentence. If a sentence in one language has its corresponding translations in another language and if both these translations differ in their roles then divergence is likely to occur.

Divergence issues mainly arise due to the incongruous nature of source and target language. This paper focuses on the congruity and discrepancies that we observe while translating our input text from English to Punjabi language. In this paper we are examining the English-Punjabi language pair with emphasis on identifying the language specific divergences. Our primary goal is to identify different types of translation divergence between English and Punjabi language.

Some commonly identified divergences are as follows:

I She is feeling **sleepy**

ਉਸਨੂੰ **ਨੀਂਦ** ਆ ਰਹੀ ਹੈ |

Usnu **neend** aa rahi hai

II In this the adjective "**sleepy**" is mapped to"**ਨੀਂਦ**" *neend* which is noun in Punjabi.

You should not **cry**.

ਤੁਹਾਨੂ **ਰੋਣਾ** ਨਹੀ ਚਾਹੀਦੀ ਹੈ

Tuhanu **rona** nhi chahida hai

In this the noun in English "**cry** " is mapped to the verb "**ਰੋਣਾ**" *rona* in Punjabi.

The rest of the paper is organized as follows: Section 1 gives introduction portion. Section 2 describes the related work. Section 3 describes the script of our language pair. Section 4 describes the Dorr's classification of translation divergence. Section 5 describes the syntactic divergence. Section 6 describes the lexical-semantic divergence. In Section 7 we conclude the paper.

Related Work

B.J Dorr [1] [2] noted that certain types of translation divergence are universal as they exist among almost all language pairs where as certain types of translation divergence are specific to a particular language.

Shukla et al. [3] have identified and categorized various patterns that exist between English, Hindi, Urdu and Punjabi translations.

Goyal and Sinha [4] have discuss translation pattern between English-Sanskrit and Hindi-Sanskrit of various constructions to identify the divergence in English-Sanskrit-Hindi Language pairs. Through this the authors were able to come up with strategies to handle these situations and also come up with correct translations. The base has been the classification of translation divergence presented by Dorr.

Sinha and Thakur [5] studied different patterns of translation divergences both from Hindi to English and English to Hindi keeping in view the classification of translation divergence proposed by Dorr. They have observed that there are a number of areas in Hindi-English translation pair that fall under translation divergence but cannot be accounted for within the existing parameters of classification strategy.

Dave et al. [6] have studied the language divergence between English and Hindi and its implication to machine translation between these languages using the Universal Networking Language (UNL). UNL has been introduced by the United Nations University, Tokyo, to facilitate the transfer and exchange of information over the internet. The criteria for deciding the effectiveness of an interlingua are that (a) the meaning conveyed by the source text should be apparent from the interlingual representation and (b) a generator should be able to produce a target-language sentence that a native speaker of that language accepts as natural.

Mishra et al. [7] proposed a method to detect and implement the adaptation rules for the divergence in English to Sanskrit machine translation. They have performed a novel method that uses rules and ANN technique to detect and implement the adaptation rules for the divergence in English to Sanskrit machine translation. The work in this paper is the first work of translation divergence in English to Sanskrit translation.

Gupta et al. [8] presented adaptation for English-Hindi EBMT. They have discussed the issue of adaptation, in general, with special emphasis to divergence. Their work looks at adaptation of EBMT between English and Hindi. They have given special attention to the study of divergence by recognizing six different categories of divergence and providing schemes for identifying them.

Script and Writing System

Punjabi is one of the Indo-Aryan languages which is mainly used in the state of Punjab in India and other regions within India. It is also used in Pakistan. Punjabi is used in all the continents by the migrants from Punjab. In India the script used for Punjabi is Gurumukhi which is based on Devanagri and is written from left to right. According to Gurumukhi script, Punjabi language has 38 consonants and 19 vowels. Some independent vowels can be constructed using the three basic characters Ura (ੳ), Aira (ਅ) and Iri (ੲ). Punjabi is one of the constitutional languages of India and is the official language of Punjab. Along with Punjab it is also spoken in neighboring areas such as Haryana, Delhi and Himachal.

English, on the other hand, is based on Roman Script. It is one of the six international languages of United Nations. Almost all the countries use it as an official business language.

Dorr’s Classification of Translation Divergence

According to Dorr’s [1] [2] classification there are two areas of divergence namely syntactic & lexical-semantic divergence. Syntactic divergence deals with syntax of both the language. Lexical semantic divergence deals with those features that can be determined lexically. Their further classification is as follows:

Syntactic Divergence can be classified into sub-categories as: i) Constituent order divergence, ii) Adjunction divergence, iii) prepositional divergence, iv) Movement divergence, v) Dative divergence, and vi) Pleonastic divergence.

Lexical divergence can be classified in to sub-categories as: i) Thematic divergence, ii) promotional divergence, iii) Demotional divergence, iv) Structural divergence, v) Categorical divergence, vi) Lexical divergence, and vii) Conflational and Inflational divergence.

Syntactic Divergence

i. Constituent-order Divergence

This type of divergence concerns with word ordering from source language to target language. In English language word order is SVO (Subject Verb Object) while in Punjabi language it is SOV (Subject Object Verb).

Ram is running his private clinic

(SVO)

ਰਾਮ ਆਪਣਾ ਨਿਜੀ ਦਵਾਖਾਨਾ ਚਲਾ ਰਿਹਾ ਹੈ|

Ram apna niji dwakhana chala riha hai

(SOV)

In this example source sentence in English follows SVO word ordering while in Punjabi word ordering is SOV.

ii. Adjunction Divergence

Adjunction divergence deals with the positioning of adjuncts.

The book lying on the table is mine

ਜੋ **ਕਿਤਾਬ** ਮੇਜ ਉੱਤੇ ਪਈ ਹੈ ਉਹ ਮੇਰੀ ਹੈ|

ਉਹ ਮੇਰੀ **ਕਿਤਾਬ** ਹੈ ਜੋ ਮੇਜ ਉੱਤੇ ਪਈ ਹੈ|

The relative clause in English is at the initial position while in Punjabi it is at the middle position. In Punjabi these clauses can be moved to sentence middle or beginning position while in English they cannot be modified.

In English the relative clause book is at initial position while in Punjabi, **ਕਿਤਾਬ** (kitab) can be at the initial position or at the middle position.

iii. Prepositional Divergence

English is the language comprising of prepositions while in Punjabi they are postpositions.

In which room they entered?

They entered **in** which room?

ਉਨ੍ਹਾਨੇ ਕਿਸ ਕਮਰੇ **ਵਿਚ** ਪ੍ਰਵੇਸ਼ ਕੀਤਾ|

In English the prepositional part (in) is at the initial position or can be at the middle position while its corresponding in Punjabi ਵਿਚ (vich) is always at the middle position.

iv. Movement Divergence

Movement divergence is caused due to the displacement property of two different languages.

Ram took cat ਰਾਮ ਬਿੱਲੀ ਲੈ ਗਿਆ|

Ram billi le gia

Cat took ram ਬਿੱਲੀ ਰਾਮ ਲੈ ਗਿਆ |

Billi ram le gia

In this example meaning is preserved in Punjabi translation despite any movement, in both cases whether it is (Ram took cat) or (Cat took ram) the meaning remains same i.e. ਰਾਮ ਬਿੱਲੀ ਲੈ ਗਿਆ (Ram billi le gia) while in English the whole meaning changes due to movement of words.

In Punjabi subject and object can change their positions without changing the meaning of the source sentence but in English, subject and object cannot change their positions. They have a fixed and rigid structure.

v. Null Subject Divergence

Null subject divergence occurs whenever we are concerned with the presence of noun phrase subject. In English the reference of noun phrase subject cannot be left blank, its presence is must but in

In Punjabi it can be left blank while preserving the grammatical meaning of the sentence.

I will sleep ਸੋ ਜਾਵਾਂਗੀ

So jawangi

She is climbing ਕੁੱਦ ਰਹੀ ਹੈ

Kud rahi hai

In this example there is no corresponding translation for ‘**I**’ in I will sleep and for ‘**she**’ in she is climbing.

vi. Dative Divergence

In English subject NP cannot occur in the dative case form while in Punjabi subject NP can be marked with the dative case form.

Sita hates Gita ਸੀਤਾ ਗੀਤਾ ਤੋਂ ਨਫ਼ਰਤ ਕਰਦੀ ਹੈ|

Sita gita ton nafrat kardi hai.

In this the word hates mapes to ਨਫਰਤ ਕਰਦੀ ਹੈ (nafrat kardi hai) in Punjabi.

vii. Pleonastic Divergence

This type of divergence occurs due to the syntactic constituents having no sementic content.

It is raining ਮੀਹ ਪੈ ਰਿਹਾ ਹੈ

Mih pe riha hai

There is no jwellery in the almira ਅਲਮਾਰੀ ਵਿਚ ਜ਼ੇਵਰ ਨਹੀ ਹਨ

Almari vich zewar nhi han

English language comprises of pleonastic subject i.e. the position of the subject is filled with some dummy structure like ‘it’, ‘there’ etc which does not have any semantic content.

Lexical Semantic

i. Thematic Divergence

Thematic difference occurs when there is difference in structure of a verb.

Ram goes with **shyam** **ਰਾਮ** ਦੇ ਨਾਲ **ਸ਼ਜਾਮ** ਜਾਂਦਾ ਹੈ|

Ram de nal **shyam** janda hai

ਰਾਮ ਸ਼ਜਾਮ ਦੇ ਨਾਲ ਜਾਂਦਾ ਹੈ|

Ram shyam de naal janda hai

In the Punjabi translation of English sentence the subject NP occurs in dative case where as in English subject NP is in nominative case.

ii. Promotional and Demotional Divergence

Promotional divergence occurs when the a phrase in the source language is mapped to a higher order phrase in the target language. In promotional divergence there is promotion of the category i.e. the syntactic constituent promotes to higher position.

Play is on ਖੇਡ ਚਲ ਰਿਹਾ ਹੈ

Khed chal riha hai

In this the adverb '**on**' is realized as a main verb in Punjabi.

iii Demotional Divergence

Demotional divergence occurs when there is demotion of syntactic constituent from source language to the target language. Demotional divergence mainly occurs when the main verb phrase is mapped to the adverbial phrase.

We do not get such type of divergence while translating from English to Punjabi.

iv. Structural Divergence

This divergence mainly occurs when the verbal object is identified as one syntactic constituents in the source language, and as another constituent in target language.
There are various levels of grammatical differences that trigger structural divergence in a pair of translation languages. In Punjabi the passive counterpart does not have its passive counterpart in English rather they have their active counterpart.

Shippu attended **the marriage**

This example is translated as:

ਸ਼ਿੱਪੂ **ਵਿਆਹ ਵਿੱਚ** ਮੌਜੂਦ ਸੀ

Shippu **viaah vich** moujud si.

In English “*the marriage”* is a noun phrase while in Punjabi it becomes prepositional phrase “ਵਿਆਹ ਵਿਚ” *viaah vich* .

v. Categorical Divergence

Categorical divergence occurs due to the mismatch between part of speech of two languages.

The family was very **cruel** to the girl

ਉਹ ਟੱਬਰ ਕੁੜੀ ਲਈ ਬਹੁਤ **ਕਠੋਰ** ਸੀ|

Uh tabbar kudi lai bahut **kathor** si

The word **cruel** which denoted adjective in English maps to Punjabi word **ਕਠੋਰ** (kathor), which is a noun.

vi. Lexical Divergence

Lexical divergence occurs due to unavailability of corresponding mapping for a word in source language to the target language. Lexical divergence is not a separate class of divergence and usually it overlaps with conflational and inflational divergence.

She **broke** in to tears ਉਹ **ਫੋੱਟ ਫੋੱਟ ਕੇ** ਰੋਣ ਲੱਗੀ|

Uh **fut fut ke** ron lagi

In this example the English counterpart broke maps in to (ਫੋੱਟ ਫੋੱਟ ਕੇ) *fut fut ke* in Punjabi.

vii. Conflational and Inflational Divergence

A Conflational divergence occurs when two or more words in the source language maps to one word in the another language.

She **clicked** **it** ਉਸਨੇ ਇਹ ਤਸਵੀਰ ਲਈ ਸੀ

Uhne eh **tasveer lai si**

The reverse case can be considered for inflational divergence.

I **will take** this ਮੈ ਇਹ ਲਵਾਂਗੀ|

Mai eh **lawangi**

Other Issues

i. Determiner Mapping Problem

This is a type of divergence for which no proper categorization is possible. In English some of the articles such as a/an/the mark the presence of noun phrase manifestly. For example **ਕੁੜੀ ਸੁੱਤੀ** *kudi sutti* will have the corresponding translation in English (The/A girl slept). The gap between their grammar causes this kind of divergence to occur.

ii. Replicative Words

Punjabi language comprises of various replicative words for which no corresponding translations exist in English language. This translation gap results in variation of syntactic part of a sentence. For Example the English sentence “Kumar got tired of crying” will be translated as **ਕੁਮਾਰ ਰੋਂਦੇ ਰੋਂਦੇ ਥਕ ਗਿਆ** *Kumar ronde ronde thak giya* (Kumar cry cry tired got). In Punjabi “cry cry” is an adverbial phrase which occurred instead of “of crying” which is called gerundive prepositional phrase in English.

iii. There and It Sentence

In English ‘there’ and ‘it’ constructs are used for asserts that mark the existence or non-existence of something that is existential sentences. They are often used to denote dummy subjects.

In Punjabi these constructs can be realized by changing the position of noun phrase.

For example:

There is a girl in the house

ਘਰ ਵਿਚ ਇਕ ਕੁੜੀ ਹੈ

ghar vich ik kudi hai

House in a girl is

In this no corresponding counterpart of ‘there’ is found while translating from English to Punjabi.

iv. Divergence in case of Voice

The sentences in English are perfectly translated to Punjabi with their active voice. For example in English “Meera loves sachin” and its Punjabi translation will be **ਮੀਰਾ ਸਚਿਨ ਨੂ ਪਿਯਾਰ ਕਰਦੀ ਹੈ |** “*meera sachin nu pyar kardi hai*” but when it is “sachin is loved by meera” there will be no direct translation for this. For ‘to love’ there is no passive form of verb in Punjabi.

Conclusion

In this paper we have analyzed on some of the divergence patterns from English to Punjabi which are based on DORR’s [1] [2] classification of translation divergence. We have also explained that what kind of issues can complicate the translation pattern of two languages. From this discussion of divergence pattern we can figure out the variation and complexity of literature of two different languages. This work has been done as a part of ongoing research, where we are building English-Punjabi MT system. This analysis of translation divergence from English to Punjabi can prove to be very helpful in translation. In order to obtain accurate translation we will explore this work more in coming future.

===

References

[1] B.J. Dorr, Machine Translation Divergences: A Formal Description and Proposed Solution, *ACL Vol. 20*, pp 597-633, 1994.

[2] B.J. Dorr, *Machine Translation: A View from the Lexicon* (The MIT Press, USA, 1993)

[3] V. N. Shukla, Renu Balyan, Pattern Identification for English to Hindi, Urdu and Punjabi Translation Proceedings of CDAC, Noida, India, ASCNT-2011.

[4] Pawan Goyal and R.M.K. Sinha, Translation Divergence in English-Sanskrit-Hindi Language Pairs, *Sanskrit Computational Linguistics, A. Kulkarni and G. Huet (Eds.), LNCS 5406*, pp. 134–143, Springer-Verlag Berlin Heidelberg 2009.

[5] R.M.K. Sinha and Anil Thakur, Translation Divergence in English-Hindi MT, *EAMT Conference Proceedings* 2005.

[6] Dave, Shachi, Jignashu Parikh, and Pushpak Bhattacharyya. "Interlingua-based English–Hindi Machine Translation and Language Divergence." *Machine Translation* 16.4 : 251-304, 2001.

[7] Mishra, Vimal, and R. B. Mishra. "Divergence patterns between English and Sanskrit machine translation." *INFOCOMP Journal of Computer Science* 8.3 (2009): 62-71.

[8] Deepa Gupta and Niladri Chatterjee, Identification of Divergence for English to Hindi EBMT, *In Proceeding of MT Summit-IX*, pp. 141-148, 2003.

===

Deepti Bhalla, M.Tech.
deeptibhalla0600@gmail.com

Nisheeth Joshi, Ph.D.
nisheeth.joshi@rediffmail.com

Iti Mathur, M.Sc.
mathur_iti@rediffmail.com

Department of Computer Science
Apaji Institute
Banasthali University
P.O. Banasthali Vidyapith - 304022
Rajasthan
India