

# Dealing with Linguistic Divergences in English-Bhojpuri Machine Translation

Pitambar Behera<sup>\*</sup>, Neha Maurya<sup>1</sup> & Vandana Pandey<sup>2</sup>

<sup>\*</sup>Centre for Linguistics & <sup>12</sup>Dept. of Linguistics

<sup>\*</sup>Jawaharlal Nehru University & <sup>12</sup>Banaras Hindu University

New Delhi, Banaras India

[pitambarbehera2, neha.mourya8, vandanapandey732]@gmail.com

## Abstract

In Machine Translation, divergence is one of the major barriers which plays a deciding role in determining the efficiency of the system at hand. Translation divergences originate when there is structural discrepancies between the input and the output languages. It can be of various types based on the issues we are addressing to such as linguistic, cultural, communicative and so on. Owing to the fact that two languages owe their origin to different language families, linguistic divergences emerge. The present study attempts at categorizing different types of linguistic divergences: the lexical-semantic and syntactic. In addition, it also helps identify and resolve the divergent linguistic features between English as source language and Bhojpuri as target language pair. Dorr's theoretical framework (1994, 1994a) has been followed in the classification and resolution procedure. Furthermore, so far as the methodology is concerned, we have adhered to the Dorr's Lexical Conceptual Structure for the resolution of divergences. This research will prove to be beneficial for developing efficient MT systems if the mentioned factors are incorporated considering the inherent structural constraints between source and target languages.

## 1 Overview

The terminology 'divergence' refers to the concept of structural or 'parametric variation' between a source language (SL) and a target language (TL) pair in Machine Translation (MT). In other words, it emerges when the decoded output content lacks 'well-formedness' because of the inherent linguistic constraints. According to Dorr (1993), "translation divergence arises when the natural translation of one language into another results in a very different form than that of the original." Therefore, it is pertinent for the identification of divergences as it facilitates and builds a blueprint towards the architectural design and implementation of MT platforms (Parameswari, 2015). So far, the availability of literature in divergence is meagre with regard to the less-resourced languages like Bhojpuri. In English-Indian languages, research on divergence has been conducted in around 9 languages: Sanskrit (Shukla et al., 2010), Hindi (Gupta & Chatterjee, 2003; Sinha & Thakur, 2005; Sinha & Thakur, 2005a), Urdu (Saboor & Khan, 2010; Muzaffar et al., 2016), Marathi (Dave et al., 2001; Kulkarni et al., 2013), Punjabi (Bhalla, 2014), Bengali (Das, 2013), Hindi-Nepali (Manger, 2014), Telugu (Ithagani, 2014), & Sindhi (Nainwani, 2015).

Dorr (1993) has classified various divergences broadly into two primary categories: syntactic and lexical-semantic. Dorr's classification of divergences is based on the Government and Binding framework as proposed and explained by (Chomsky, 1981; Jackendoff, 1990) of linguistic theory which attempts at capturing surface structure variations by deep structure. The deep structure provides a background as the universal structure which may possibly be applicable to all languages. Therefore, it can however be posited that both the classification and resolution of translation divergences are explained from the perspective of the universal grammar formalism.

### 1.1. The Areal Features of English and Bhojpuri

Bhojpuri belongs to the Indo-Aryan or Indian language family whereas English owes its origin to the Germanic family. There are a lot of incompatible, divergent and linguistically-grounded features pertaining to morphology, syntax and semantics (Muzaffar et al., 2016) between English and Bhojpuri.

---

This work is licensed under a Creative Commons Attribution 4.0 International License.  
License details: <http://creativecommons.org/licenses/by/4.0/>

Like most of the Indo-Aryan languages, Bhojpuri (Ojha et al., 2015; Singh, 2014; Singh, 2015; Singh, 2015a) is also a morphologically rich and non-configurational language, unlike English. In addition, English applies expletives, existential subjects and no verbal honorific agreement. Besides, Bhojpuri as a South Asian language has some atypical constructions: complex predicates, serial verb constructions, non-nominative subjects, conjunctive participle and so on (Subbārāo, 2008 & 2012).

## 2 Dorr's LCS for Dealing with Divergences

The Lexical Conceptual Structure (LCS)<sup>2</sup> is the semantic representation of predicate argument structures through decomposition of their features.

Give: [CAUSE (x, [GO (y, [TO (z)])])]

In the theoretical specification demonstrated above, the verb 'give' can be decomposed as having three predicates viz. CAUSE, GO, and TO, as per the intuition that a sentence i.e. 'Rohit gave Sita a pencil' means that Rohit (which equals to x) caused the pencil (=y) to go to Sita (=z). In other words, agent is Rohit, patient is the pencil and the beneficiary is Sita. Hence, the LCS theoretical specification can be fit into any natural language having this type of structural specifications. As a result, it becomes language independent in nature and the issue of divergences can be addressed applying this concept. Divergences can be approached from two points of views: syntax (the syntactic structure) and semantics (lexical semantics).

### 2.1 The Syntactic Structure

Constituents in the sentence are grouped on the basis of their relevance and position in the hierarchy. The convention applied in the bracketing is based partially on the Government and Binding theory with some modifications in notation for simplification of the concepts.

- CP: it is the complementizer phrase such as 'that' in English and 'kI' in Hindi which augments a subordinate clause in a sentence.
- IP: it stands for the inflectional phrase which encapsulates the auxiliaries (modal and be verbs) in English.
- Some other notation conventions are nominal phrase (NP or DP), verbal phrase (VP), prepositional phrase (PP), adverbial phrase (ADVP), adjectival phrase (ADJP) etc.

The instance "I came quickly" is considerable here. The structural representation of the given sentence syntactically is provided below.

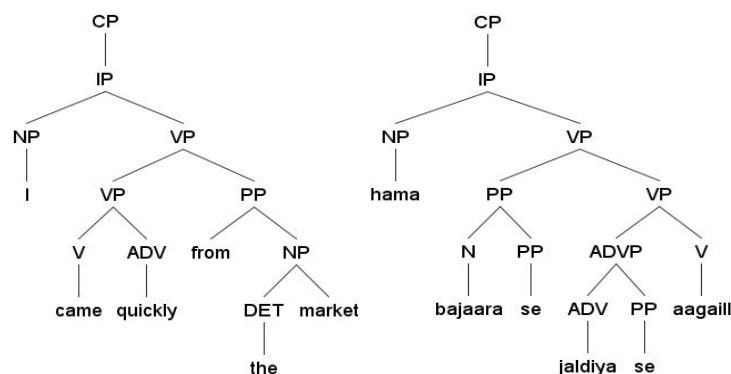
[CP [IP [NP I]

[VP [VP [V came] [ADV quickly]] [PP from [NP [DET the] [N market]]]]]

[CP [IP [NP hama]

[VP [PP [N bajaare] [PP se]] [VP [ADVP [ADV jaldiya] [PP se]] [V aagail]]]]]

In this above instantiated example, [V aagail] is the syntactic head of the sentence whereas [NP hama] is the syntactic subject and [ADV jaldiya] is the adverb which makes an ADVP including an adposition [PP se] in Bhojpuri. On the other hand, English sentence contains [V came] as the head of the sentence, [NP I] as the subject and [ADV quickly] as the ADVP.



<sup>2</sup> For more basics on notation convention please refer to Dorr, 1994; Gupta and Chatterjee, 2003 and Muzaffar et al., 2016

## 2.2 The Lexical Semantics

The syntactic constituents are analyzed for providing an intermediate representation in a form known as the LCS. The LCS may be acquired with the unification of Root Lexical Conceptual Structure (RLCS) of the constituent words in the given sentence. It is a modified and adapted version of the representation as proposed by Jackendoff (1983, 1990) which conforms to the following form:

$$[_{T(X)} X' ([_{T(W)} W'), [_{T(Z'1)} Z'1] \dots [_{T(Z'n)} Z'n] [_{T(Q'1)} Q'1] \dots [_{T(Q'm)} Q'm])]$$

In addition, this representation is compositional with decompositional features, language independent in nature and provides a theoretical framework for the representation of a sentence with the help of semantics. The sentence “I came from the market quickly” is represented in the LCS as the following.

$$[_{Event} COME_{Loc} ([_{Thing} I], [_{Path} FROM_{Loc} ([_{Position} AT_{Loc} [_{Thing} I] [_{Location} THE MARKET])]) [_{manner} QUICKLY])]$$

Where COME<sub>Loc</sub> is the head of LCS, ‘I’ pronominal is the LCS subject, FROM<sub>Loc</sub> is the LCS object, QUICKLY is the LCS modifier. The Root Lexical Conceptual Structure (RLCS) is ‘an uninstantiated LCS’ (Dorr, 1994) which is associated with the definition of a word in the lexicon. For instance, the RLCS of the verb ‘come’ is as follows.

$$[_{Event} COME_{Loc} ([_{Thing} X], [_{Path} FROM/TO_{Loc} ([_{Position} AT_{Loc} ([_{Thing} X], [_{Thing} Z])])])]$$

To get a composed (CLCS) we unify RLCSs for ‘come’ and ‘I’. Generalized Linking Routine (GLR) correlates the constituent words of the syntactic representations to those of the LCS by the mappings as demonstrated in the following.

- ❑ V' ⇔ V ([GO<sub>Loc</sub>] ⇔ [V came])
- ❑ S' ⇔ S ([RAHIM] ⇔ [NP I])
- ❑ O' ⇔ O ([TO<sub>Loc</sub>] ⇔ [PP from ...])
- ❑ M' ⇔ M ([FAST] ⇔ [ADV quickly])

Lastly, the lexical-semantic items are related in a systematic manner to their corresponding syntactic categories by applying Canonical Syntactic Realization (CSR): For instance:

LCS Types	Syntactic Categories
Event, State	V (verb)
Thing	N (noun)
Property	Adj (adjective)
Path, Position	P (preposition)
Location, Time, manner, Intensifier and Purpose	ADV (adverbial)

Table. 1 The LCS Types and Notation Conventions

## 3 Categorization of Divergences

Dorr (1993) has classified various divergences broadly into two primary categories: syntactic and lexical-semantic. Furthermore, each of the classes has been sub-categorized and the corresponding instances have been drawn as in the following.

### 3.1 Syntactic Divergences

This set of divergences, which are based on the syntax of concerned languages, has been sub-categorized into seven lower-level types: constituent order, adjunction, preposition-stranding, movement, null-subject, dative subject and pleonastic. These are some of the universal parametric variations atypical to English as an SL and any natural language as the TL.

### 3.1.1 Constituent Order

This divergence pertains to the word-ordering of the concerned SL and TL languages. It emerges when there is mismatch between the word order patterns of SL and TL. On one hand, English is a configurational language which follows a rigid pattern (SVO) and is unmarked. On the other hand, Bhojpuri being an Indic language allows relatively free word order patterns viz. SOV (unmarked), SVO and OVS (marked). However, both the types of patterns are acceptable syntactically in Bhojpuri.

For instance,

The boys are playing Cricket.

<i>laikana</i>	<i>krikeTa</i>	<i>khela-taaDana.</i>
Boys.M.PL.3.NOM	cricket	play.PRS.PL.IPFV.M.
S	O	V

<i>laikana</i>	<i>khela-taaDana</i>	<i>krikeTa.</i>
Boys.M.PL.3.NOM	play.PRS.PL.IPFV.M.	cricket
S	V	O

<i>krikeTa</i>	<i>khela-taaDana</i>	<i>laikana.</i>
Cricket	play.PRS.PL.IPFV.M.	Boys.M.PL.3.NOM
O	V	S

### 3.1.2 Adjunction

Adjunction divergence concerns with the difference of mapping in complements (prepositional, non-finite verbal complements etc.) and adjuncts (prepositional phrases, participial constructions etc.) between two languages. In the English input sentence, the infinitival adjunction is translated as prepositional complement ‘badanaama kare ke kosIs’ in Bhojpuri.

He tried to defame me.

<i>u</i>	<i>hamake</i>	<i>badanaama</i>	<i>kare ke kosIs</i>	<i>kailasa</i>
he.NOM.SG.3.NHON	me.DAT	defame	do of try	do.PST.PRF.NHON.3

He came here after having food.

<i>u</i>	<i>khaanaa</i>	<i>khaile</i>	<i>ke baada</i>	<i>ihaaN aayal.</i>
he.NOM.SG.3.NHON	food	eat.PST.PRF.	after	here come.PST.PRF.NHON.3

### 3.1.3 Preposition-stranding

Preposition-stranding, otherwise called as P-stranding, is one of the syntactic constructions where the preposition occurs somewhere in the sentence (generally at the end) other than its canonical position; adjacent to its object. This construction is quite alien to most of the South-Asian languages which includes Bhojpuri. As a result parametric variation emerges between a pair of languages.

For instance,

Where are you coming from?

<i>kahaaN tu</i>	<i>aava-ta</i>	<i>hauaa</i>	<i>se?</i>
where You.	come.PRS.	be.PROG.PRS.3.	from

### 3.1.4 Movement

When we try to move certain constituents in English input sentence, they cannot be moved as freely as Bhojpuri. This accounts for the fact that Indian languages are relatively free so far as the process of scrambling is concerned. If we shuffle the word order of the following input sentence, it becomes grammatically acceptable but semantically not well-formed. Because ‘the book’ semantically cannot buy ‘Ram’. In other words, the inanimate object cannot play the role of an animate subject which is logical.

For Example,

Ram purchased a book. \*A book purchased Ram

<i>raama</i>	<i>ekhe/ekthe kitaaba</i>	<i>kharidalana</i>
--------------	---------------------------	--------------------

Ram.NOM.PST.3.HON a book buy.PST.IPFV.3.HON  
*ekhe/ekthe kitaaba raama kharidalana.*  
 a book Ram.NOM.PST.3.HON buy.PST.IPFV.3.HON

### 3.1.5 Null-subject

When the position of the subject is either left implicit or attributed by some pronouns such as ‘there’ in English it is called a null-subject. When there is a covert subject the agreement features are generally marked with the verb when there is S-V agreement. So, in the example mentioned below there is no equivalent translation for the existential ‘there’ in Bhojpuri. Consequently, this divergence crops up which proves to be a barrier in MT.

Example,

There was a lion in the forest.

*jangala meN ekhe baagha rahala.*

Forest in.LOC a tiger be.PST.PRF.3.

### 3.1.6 Dative Subject

This construction is otherwise known as the non-nominative construction which is atypical to South Asian languages. The psychological predicates such as ‘hunger’, ‘thirst’ and so on are expressed with the addition of a dative subject postpositional marker in Indian languages (here ‘ke’).

I have fever.

*ham-ke bokhaara hava/baa*

I.DAT.SG.1. fever be.PRS.1.IPFV

### 3.1.7 Pleonastic

Pleonastic pronoun or dummy pronoun is a pronoun which lacks meaning and is used when the argument is irrelevant, non-existent or redundant. In the input sentence, ‘it’ is considered to be impersonal semantically and intransitive syntactically. On the other hand, with the absence of the pleonastic subject, ‘water’ becomes the subject of the intransitive verb ‘rain’ /barasata/.

It is raining.

*paanii barasata hava*

water rain.PRS.PROG be.PRS.

## 3.2 The Lexical-Semantic Divergences

On the basis of lexical-semantics of the given languages, the lexical-semantic divergences have been sub-classified into seven sub-divisions: thematic, promotional, structural, inflational, conflational, categorial and lexical.

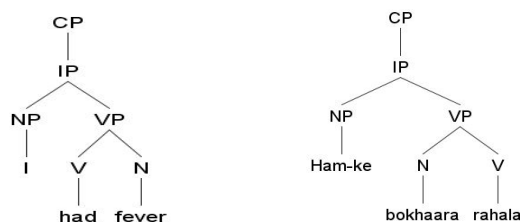
### 3.2.1 Thematic

This categorization is based on the principle of the thematic roles of the arguments of the verbs. For instance the role of the agent (Ham-ke DAT) is realized as the dative subject in Bhojpuri whereas the English counterpart has nominative case (I-NOM) marker on the agentive subject and accusative on the other argument. As a result owing to the fact that Bhojpuri allows an oblique subject where the object gets co-indexed with the verb which is not true so far as English is concerned.

I had fever.

*ham-ke bokhaara rahala*

I.DAT. fever be.PST.

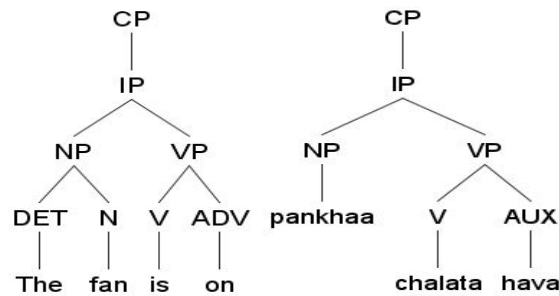


### 3.2.2 Promotional

This divergence occurs when one constituent in a given language having a lower position in the hierarchy gets promoted to a higher position in the target language. In this case, the category of adverb ('on' in English) which has a lower position (modifier of the verb) in the hierarchy gets promoted to the higher status of verb (chalata hava) in Bhojpuri counterpart.

The fan is on.

*pankhaa chalata hava.*  
 Fan run.PROG be.PRS



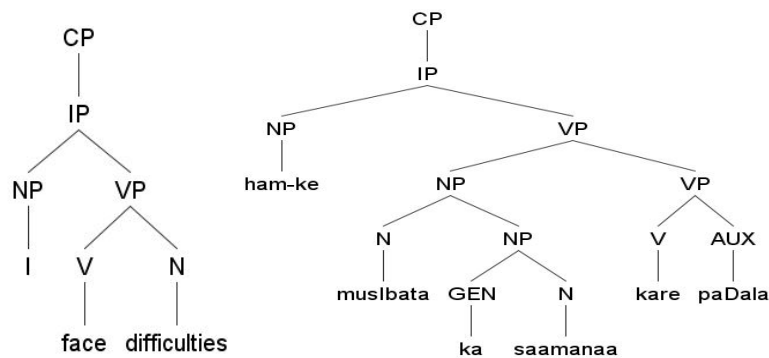
### 3.2.3 Structural

It occurs when there is difference between languages on the basis of structure or syntax. For instance, the nominal phrase argument in English is translated as the prepositional adjunct in the target language. This divergence generally originates when there is phrase-level parametric variations which becomes a barrier for MT. So, the NP, which is an argument of the verb 'face' in English, is translated as an adjunctive PP (muslbata kaa saamanaa) in Bhojpuri.

For instance,

I face difficulties.

*hamke muslbata ka saamanaa kare(ke) paDala.*  
 I.DAT difficulty of face do have.PST.PRF.



### 3.2.4 Inflational & Conflational

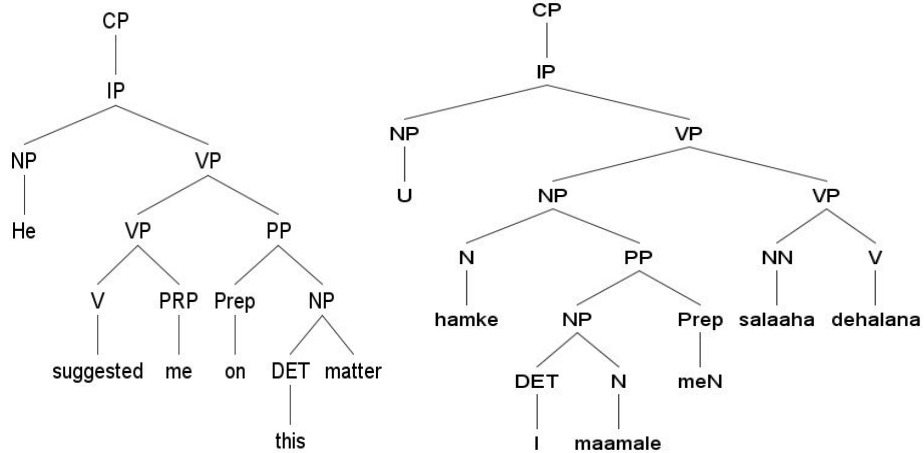
On one hand, when one linguistic element in SL is inflated as the realization of more than one element in the TL it is called as the inflational divergence. On the other hand, when two or more linguistic elements in SL are conflated to be realized as one word in TL it is known as conflational divergence. In the inflational example below, the SL 'suggested' gets realized in the TL as having two elements 'salaaha dehalana'. In the conflational instance, phrasal verb "looked for" gets translated in Bhojpuri as having only one word i.e. 'khojalasa'.

He suggested me on this matter.

*u hamake I maamalaa meN salaaha dehalana.*  
 he.NOM.3 me.DAT.OBJ. this matter in.OBL advise give.PST.PRF.3

He looked for a room.

*u ekhe makaana khojalasa*  
 he.NOM.3 a house search.PST.PRF.3

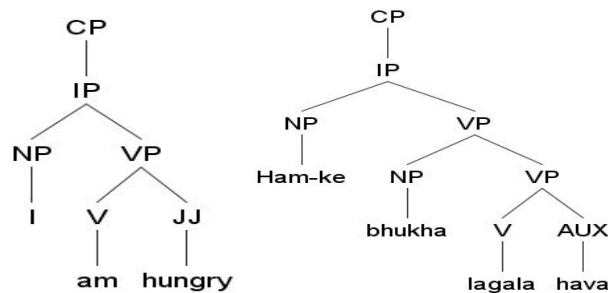


### 3.2.5 Categorial

When there is a change in the very grammatical category of a linguistic element in TL it is known as the categorial divergence. In the example instantiated below the predicative adjective ‘hungry’ in English gets translated as the nominal phrase in Bhojpuri. Thus there is a change in the parts of speech categories from adjective to noun.

I am hungry.

*ham-ke bhukha lagala hava*  
 I.DAT.3 hunger seem.PRF be.PRS.



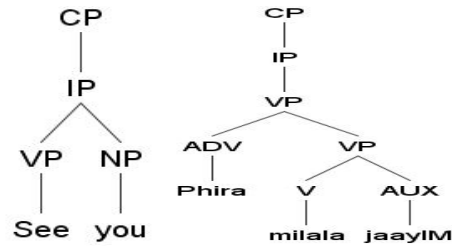
### 3.2.6 Lexical

This divergence occurs when two or more divergence types combine or because of the unavailability of the exact equivalent translation. Thus, in the following example, there is no equivalent translation for the expression ‘see you’ in Indian languages like Bhojpuri. There is fair amount of lexical divergences between a pair of languages that are responsible for creating divergence issues in MT.

See you!

*phira milala jaayIM*

Again meet go.FUT.PL.1.IPFV



#### 4 The Identification and Resolution Procedure

This section provides the systematic method for the identification and probable solution of the lexical-semantic divergences between English and Bhojpuri.

##### 4.1 Thematic

This divergence emerges when the GLR invokes the following steps of relation (Dorr, 1990a).

Firstly, one needs to relate the syntactic object with the LCS subject  $\Rightarrow O' \Leftrightarrow S$

Secondly, one needs to have the relation between the syntactic subject to the LCS object  $\Rightarrow S' \Leftrightarrow O$

The syntactic structure and the corresponding CLCS are provided in the following.

$$\begin{aligned}
 & [CP [IP [NP I] [VP [V had] [N fever]]]] \\
 & \Leftrightarrow [State BEI_{Ident} ([Thing I], \\
 & \quad [Position AT_{Ident} ([Thing I], [Thing FEVER]), \\
 & \quad \quad [manner SEVERELY]])] \\
 & \Leftrightarrow [CP [IP [NP Ham-ke] [VP [N bokhaara] [V rahala]]]]
 \end{aligned}$$

In the above instantiated example, the subject gets the thematic roles of a nominative agentive subject in English and concedes the role of impersonal and non-agentive subject with dative case marker.

##### 4.2 Promotional

In this divergence the GLR augments in the following manner.

One needs to consider the following steps:

1. One needs to relate the LCS verb with the syntactic object  $\Rightarrow V' \Leftrightarrow S$
2. Promote the LCS modifier (adverb) position to the position of verb  $\Rightarrow M' \Leftrightarrow V$

The syntactic structure and the respective CLCS are demonstrated below.

$$\begin{aligned}
 & [CP [IP [NP [DET The] [N fan]] [VP [V is] [ADV on]]]] \\
 & \Leftrightarrow [State BEI_{Ident} ([Thing THE FAN], \\
 & \quad [Position AT_{Ident} ([Thing THE FAN], \\
 & \quad \quad [manner OFF]])] \\
 & \Leftrightarrow [CP [IP [NP pankhaa] [VP [V chalata] [AUX hava]]]]
 \end{aligned}$$

##### 4.3 Structural

This divergence is quite different from the above two types of divergences in so far as the alternation of the position of the constituents is concerned. But it changes the nature of the relation between various positions.

The syntactic structure and CLCS are provided below.



[CP [IP [NP I] [VP [V face] [N difficulties]]]]

⇔ [Event GO<sub>Loc</sub>

([Thing I],

[Path TO<sub>Loc</sub> (Position IN<sub>Loc</sub> ([Thing I], [Thing DIFFICULTIES]))]]])

⇔ [CP [IP [NP ham-ke] [VP [NP [N muslbata] [NP [GEN kaa] [N saamanaa]]] [VP [V kare] [AUX paDala]]]]]]

One of the arguments of the verb in English is translated as the prepositional phrase in Bhojpuri which creates complexity for automatic translation.

#### 4.4 Inflational & Conflational

In the bracketing representation demonstrated following, it is quite obvious that the one-word token verbal element i.e. ‘suggested’ is translated as having two tokens i.e. “salaaha dehalana” in Bhojpuri and is a quintessential example of inflational divergence. It is completely based on the economy of usage of strings in both the concerned languages. When the economically inflated expressions in the SL are reduced to a conflated expression in the TL counterpart, it is called as conflational divergence. The syntactic structure and the CLCS are demonstrated in the following.

[CP [IP [NP He] [VP [VP [V suggested] [PRP me]] [PP [Prep on] [NP [DET this] matter]]]]]]

⇔ [Event GO<sub>Loc</sub>

([Thing HE],

[Path TO<sub>Loc</sub> (Position IN<sub>Loc</sub> ([Thing HE], [Thing ME], ([Thing THE MATTER]))))]])

⇔ [CP [IP [NP U] [VP [NP [N hamke] [PP [NP [DET I] [N maamalaa]] [Prep meN]]] [VP [NN salaaha] [V dehalana]]]]]]

#### 4.5 Categorial

In this divergence, there is no identity in the relationship between the syntactic category and the concerned lexical-semantic item.

[CP [IP [NP I] [VP [V am] [JJ hungry]]]]

⇔ [State BE<sub>Ident</sub>

([Thing I],[Position AT<sub>Ident</sub> ([Thing I], [property HUNGRY])])

⇔ [CP [IP [NP Ham-ke] [VP [NP bhukha] [VP [V lagala] [AUX hava]]]]]]

In the examples mentioned above the divergence owes to the fact that adjectival parts of speech category in English is translated into a nominal category in Bhojpuri counterpart.

#### 4.6 Lexical

This divergence is considered to be one of the by-products of any of the above-described combination of divergences. In addition, the unavailability of the proper translation in the target language is also encapsulated in this category.

[CP [IP [VP See] [NP [you]]]]

[CP [IP [VP [ADV Phira] [VP [V milala] [AUX [jaayIM]]]]]]

### 5 Conclusion

The successful implementation of any MT platform solely depends upon how well an instance of translation is retrieved from a plethora of data and modified to cater to the demand of the desired translation output. Although the heuristic linguistic rules are capable of dealing with several errors but are not sufficient for tackling the exceptional cases of linguistic divergences (Gupta & Chatterjee, 2003). Therefore, our rationale for dealing with the divergence patterns between English and Bhojpuri language pair is to bring out various types of divergent, incompatible or incongruent features.

In this study, we have focused light on classifying various divergences between English and Bhojpuri translations. The theoretical framework for classification is based on Dorr’s classification of divergences from syntactic and lexical-semantic points of views. So far as the identification and resolution are concerned, we have adhered to the LCS schema. This analytical study on divergence

between English and Bhojpuri language pair can prove to be fruitful for any IA language in general and less-resourced languages in particular to develop efficient and qualitative Machine Translation platforms.

## References

- Abdus Saboor & Mohammad Abid Khan. 2010. *Lexical-semantic divergence in Urdu-to-English example based Machine Translation*. 6th International Conference on Emerging Technologies (ICET), vol., no., 316, 320.
- Atul Ku. Ojha, Pitambar Behera, Srishti Singh, and Girish N. Jha. 2015. Training and Evaluation of POS Taggers in Indo-Aryan Languages: A Case of Hindi, Odia and Bhojpuri. In *Language Technology Conference-2016*.
- Bonnie J. Dorr. 1990b. Solving thematic divergence in Machine Translation. In *28th Annual Conference of the Association for Computational Linguistics*, 127-134: Pittsburg, PA University of Pittsburg.
- Bonnie J. Dorr. 1993. *Machine Translation: A View from the Lexicon*. The MIT Press, Cambridge, Mass.
- Bonnie J. Dorr. 1994. Classification of Machine Translation divergences and a proposed solution. *Computational Linguistics*, 20(4): 597-633.
- Bonnie J. Dorr. 1994a. Machine translation divergences: A formal description and proposed solution. *Computational Linguistics*, 20(4), 597-633.
- Deepa Gupta and Niladri Chatterjee. 2003. Identification of divergence for English to Hindi EBMT. In *MT Summit-IX*, 141-148.
- Deepti Bhalla, Nisheeth Joshi, and Iti Mathur. 2014. Divergence Issues in English-Punjabi Machine Translation. *Language in India*, 14(10).
- K. Parameswari. 2015. Development of Telugu-Tamil Transfer-Based Machine Translation system: With Special reference to Divergence Index. In *1st Deep Machine Translation Workshop*, 48-54.
- K. M. Manger. 2014. Translation Divergences in Hindi-Nepali Machine Translation. *Language in India*, 14(5).
- Kārumūri V. Subbārāo. 2008. Typological characteristics of South Asian languages. *Language in South Asia*, pages 49-78.
- Kurumari V. Subbarao. 2012. *South Asian languages: A syntactic typology*. Cambridge University Press, England.
- Niladri S. Das. 2013. Linguistic divergences in English to Bengali Translation. *International Journal of English Linguistics*; Vol. 3, No. 1; 2013.
- Noam Chomsky. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Pinky Nainwani. 2015. *Challenges in Automatic Translations of Natural Languages- A Study of English-Sindhi Divergence*. Doctoral Dissertation, Centre for Linguistics, Jawaharlal Nehru University, New Delhi, India.
- Preeti Shukla, Devanand Shukl, and Amba Kulkarni. 2010. Vibhakti divergence between Sanskrit and Hindi. In *Sanskrit Computational Linguistics*. Springer Berlin Heidelberg, 198-208.
- Rai M. K Sinha & Anil Thakur. 2005. Translation Divergence in English-Hindi MT. In *EAMT Xth Annual Conference*, Budapest, Hungary.
- Rai M. K Sinha & Anil Thakur. 2005a. Divergence Patterns in Machine Translation between Hindi and English. In *MT Summit X*. Phuket, Thailand, 12-16th September, 346-353.
- Ray S. Jackendoff. 1983. *Semantics and Cognition*. MIT Press, Cambridge.
- Ray S. Jackendoff. 1990. *Semantic Structures*. MIT Press, Cambridge.
- S. B Kulkarni, P. D. Deshmukh, and K. V Kale. 2013. Syntactic and structural divergence in English-to-Marathi Machine Translation. *IEEE 2013 International Symposium on Computational and Business Intelligence*, August 24-26, 2013, New Delhi, 191-194, (doi: 10.1109/ISCBI.2013.46).
- Shachi Dave, Jignashu Parikh, and Pushpak Bhattacharya. 2001. Interlingua-based English-Hindi Machine Translation. *Journal of Machine Translation*, 16(4), 251-304. (<http://dx.doi.org/10.1023/A:1021902704523>.)
- Srishti Singh and Esha Banerjee. 2014. Annotating Bhojpuri Corpus using BIS Scheme. In *Proceedings of 2nd Workshop on Indian Language Data: Resources and Evaluation (WILDRE-2), Ninth International Conference on Language Resources and Evaluation, LREC (2014)*, Reykjavik, Iceland, May 26-31, 2014.
- Srishti Singh. 2015. *Challenges in Automatic POS Tagging of Indian Languages- A Comparative Study of Hindi and Bhojpuri*. M. Phil. Dissertation. Jawaharlal Nehru University, New Delhi.
- Srishti Singh and Girish N. Jha. 2015a. Statistical Tagger for Bhojpuri (employing Support Vector Machine). In *4th International Conference on Computing, Communication and Informatics (ICACCI, 2015)*.
- Sharmin Muzaffar, Pitambar Behera, and Girish N. Jha. 2016. Classification and Resolution of Linguistic Divergences in English-Urdu Machine Translation. In *WILDRE-3 (LREC-2016)*, Portoroz, Slovenia.

- Sharmin Muzaffar, Pitambar Behera, and Girish Nath Jha. 2016. A Pāniniān Framework for Analyzing Case Marker Errors in English-Urdu Machine Translation. *Procedia Computer Science (Elsevier)*, 96, 502-510.
- Venkanna Ithagani. (2014). Linguistic Convergence and Divergence in Telugu-Urdu Contact Situation: A Study with Special Reference to Telangana Dialect.