

Linguistic Divergence Pattern in English and Magahi Machine Translation

Shweta Chandra

Amity University Patna

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 are 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 to identify and resolve the divergent linguistic features between English as source language and Magahi as target language pair. In Dorr (1993, 1994) an attempt has been made to classify different types of translation divergence. However, the issue of linguistic divergence is such a complex phenomenon that a lot more needs to be done in this area to identify further categories of translation divergence, their implications as well as the approaches to handle them. In this study, I take Dorr's classification of translation divergence as the point of reference to examine the different types of TDs that are encountered in Magahi-English and English-Magahi MT. We attempt to identify the types of TDs in these pairs of translation languages that we find cannot be accounted for within the existing categorization. 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 other 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 Magahi. 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 Magahi

Magahi 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 Magahi.

Like most of the Indo-Aryan languages, Magahi (Verma 2003, 1991, 1985) is also a morphologically rich and non-configurational language, unlike English. In addition, English applies expletives, existential subjects and no verbal honorific agreement. Besides, Magahi 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 CLASSIFICATION AND PATTERNS OF TDs IN ENGLISH AND Magahi MT

2.0 Dorr's Classification

Dorr's (1993) classification divides TDs between two broad classes: (A) Syntactic Divergences, and (B) Lexical-semantic Divergences. These two classes of TDs have been further categorized into different types, as follows:

(A) Syntactic Divergences

- i. Constituent order divergence
- ii. Adjunction divergence
- iii. Preposition-stranding divergence
- iv. Movement divergence
- v. Null subject divergence
- vi. Dative divergence
- vii. Pleonastic divergence

(B) Lexical-semantic Divergences

- i. Thematic divergence
- ii. Promotional divergence
- iii. Demotional divergence
- iv. Structural divergence
- v. Conflational divergence
- vi. Categorial divergence
- vii. Lexical divergence

Dorr (1994) has examined the structure of the lexical-semantic divergences and proposed a LCS-based approach to handle them in MT. This classification takes into account various sources of differences between a set of translation language and captures a large sets of translation divergences. The classification is also broadly based on the Government and Binding framework (Chomsky 1986, Jackendoff 1990) of linguistic theory which assumes a deep structure to capture the surface structure variations. The classification and the resolution of the translation divergence have been discussed largely from the perspective of the universal grammar. The classification claims to capture major grammatical issues in translation divergence across languages. However, it also misses a number of points that pertain to a particular set of translation languages. The issue of divergence between a set of languages is associated with a number of factors ranging from linguistic to socio- and psycho-linguistic aspects of the languages involved. Although Dorr's classification takes into account many of the major linguistic factors associated with TDs, there still remain a number of points related to both linguistic and extra-linguistic factors that may exist in different sets of translation languages. Furthermore, the parameters of the classification does not take into account subtle semantic/pragmatic factors to the extent they are relevant for the classification of translation divergence in various languages. Without going into a detailed discussion of the different classes and categories of translation divergence as proposed in Dorr (1993, 1994), we discuss the types of translation divergence in English-Magahi and Magahi-English MT, particularly that which are not amenable to the existing classification of translation divergence. Below we discuss the various types of English-Magahi and Magahi-English translation divergence in light of the classification of translation divergence in Dorr (1993).

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.

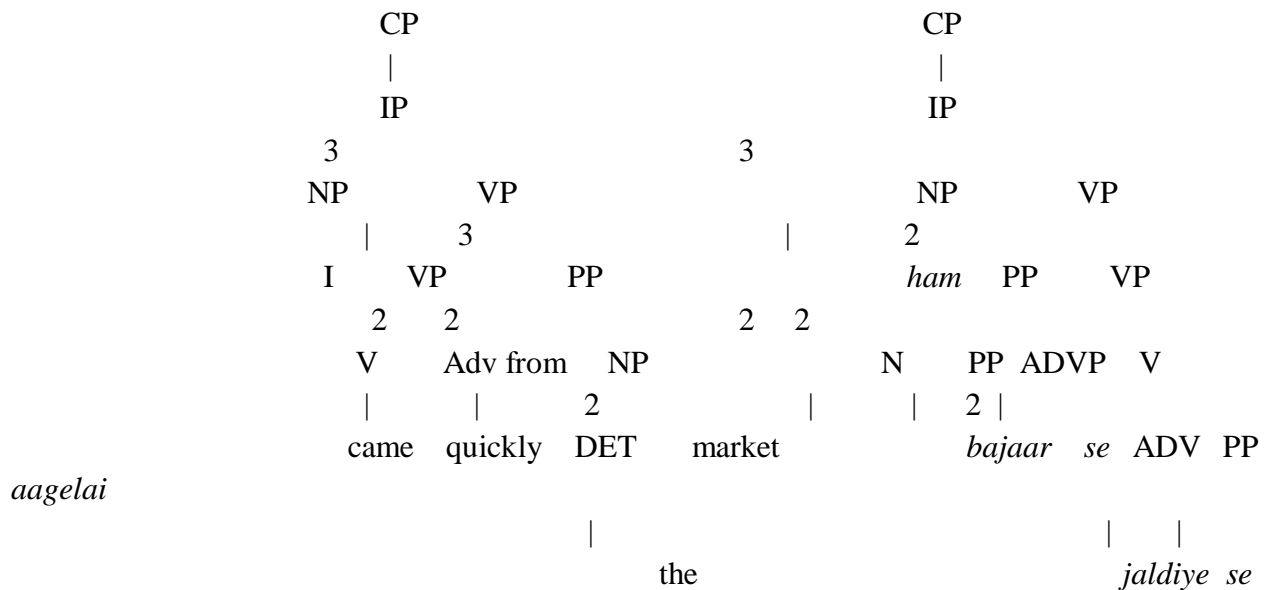
[CP [IP [NP I]

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

[CP [IP [NP *ham*]

[VP [PP [N *bazaar*] [PP *se*]] [VP [ADVP [ADV *jaldiya*] [PP *se*]] [V *aagelai*]]]]]

In this above instantiated example, [V *aagelai*] is the syntactic head of the sentence whereas [NP *ham*] is the syntactic subject and [ADV *jaldiya*] is the adverb which makes an ADVP including an adposition [PP *se*] in Magahi. 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.



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]...[T(Z'n) Z'n] [T(Q'1) Q'1]...[T(Q'm) Q'm])]

[Event COMELoc

([Thing I],

[Path FROMLOC ([Position ATLOC [Thing I] [Location THE

MARKET])])]

[manner QUICKLY])]

Where COMELoc is the head of LCS, 'I' pronominal is the LCS subject, FROMLOC 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 COMELoc

([Thing X],

[Path FROM/TOLoc ([Position ATLoc ([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 ([GOLoc] ⇔ [V came])
- S' ⇔ S ([RAHIM] ⇔ [NP I])
- O' ⇔ O ([TOLOC] ⇔ [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

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.

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.

3.1.1 Syntactic Divergence

i. Constituent Order Divergence:

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, Magahi 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 Magahi. Some examples presented to illustrate the point.

(1) He eats a mango. [S-V-O]

u aam khaa hai. [S-O-V]

he mango eat bePres

u aam khaa hai. [S-O-V]

he mango eat bePres

aam u khaa hai. [O-S-V]

mango he eat bePres

ii. Adjunction Divergence: Adjunction divergence relates to the difference in the placement of adjuncts (modifiers) between source and target languages. It is a well-known fact that natural languages differ with respect to the placement of adjunct (modifier) elements in a sentence. In some languages, modifiers precede the head element that it modifies whereas in some others they follow the head. English and Magahi show some differences on this account. The difference is mainly located in adverbial modifiers at the level of a clause. The placement of various adjectival modifiers in nominal phrase in both English and Magahi follow largely a similar pattern. That is, they precede the head noun that they modify. However, there are some differences between English and Magahi with respect to the placement of adjectival modifiers, too. For instance, the relative clause in Magahi can occur in sentence-initial, sentence-medial and sentence-final positions whereas in English the relative clause cannot occur in the sentence-initial position. We can look at the examples in (2) to illustrate this divergence.

(2) *je laikaba uhaan khaRa hai u hamar bhai hai*

{RP boy there standing bePres he my brother bePres }

=> i. ‘The boy who is standing there is my brother.’

ii. *‘Who boy is standing there he is my brother.’

In (2), the relative clause in Magahi occurs in the sentence-initial position. The English sentence (i) shows the relative clause in the sentence-medial position. The relative clause is not allowed in the sentence-initial position as is evident from the ungrammaticality of the sentence in (ii). We can notice that the divergences caused by differences in the word order of different modifier elements are actually part of the broad word order differences between English and Magahi.

iii. Preposition-stranding divergence: English has prepositions (precede the head noun) whereas Magahi has postpositions (follow the head noun). In English, there are certain cases where the head noun moves from its canonical position to some other position in a sentence leaving the preposition stranded in the original position. This phenomenon is called preposition-stranding is quite alien to most of the South-Asian languages which includes in Magahi. The examples in (3) illustrate the point.

(3) Where are you coming from?

=> *tu kahaaN se aabit ha?*

you where from come PROG bePres Hon

In English, the interrogative clause is obtained by the use of a wh-word which fronts to the sentence-initial position. For instance, in (4a) the affirmative sentence contains a place adverbial PP (prepositional phrase) with preposition *from*. When the sentence is transformed into an interrogative clause (4b), the wh-word moves to the sentence-initial position and can optionally leave the preposition in the original place. The option of “postposition-stranding” does not arise in Magahi.

(4) a. He is coming from the market.

b. Where is he coming from?

c. From where is he coming?

(5) a. *u baajaar se aabit hai.*

b. *u kahaaN se aabit hai?*

Thus, we notice that the preposition-stranding divergence also broadly comes under word order differences between English and Magahi.

iv. Movement divergence: Natural languages show displacement property that causes translation divergences between the source and the target languages. For instance, in Magahi the subject and the object NPs in certain cases exchange their canonical positions, as in (6-7). Another case of movement-related difference between Magahi and English is interrogative sentences. They are formed simply by substituting an NP by the relevant interrogative pronoun whereas in English the interrogative pronoun further obligatorily moves to the sentence-initial position. We can see the example in (9-10) for illustration.

(6) a. *ham okara se pyar kara he.*

I he Poss from love do be.Pres

=> 'I love him.'

b. *okara se ham pyaar kara he*

he Poss from I love do be.Pres} .

=> 'I love him.'

(7) a. I love him.

b. *Him I love.

As we notice, the subject and the object NP can exchange their respective position in Magahi, as in (6) without any alternation in the meaning (except some sort of definiteness/emphasis marking). This type of movement/scrambling results in the ungrammaticality of the respective construction in English, as in (7b).

v. Null subject divergence: Natural languages may differ on whether the subject NP position is overtly filled. In English, the subject NP position cannot be left empty whereas in Magahi (and many other Indian languages) the subject NP is optionally dropped, as in (8-9).

(8) *aa jaaibo.*

{come go Fut}

=> 'I will come.'

(9) *khaanaa khaait hai.*

{food eat PROG be.PR}

=> 'He/She is eating food.'

(10) * is eating food.

The examples in (8-9) clearly show that the subject NPs are not overtly present in the sentence and still the sentence is grammatical in Magahi. The absence of the subject NP makes the English sentence, as in (10), ungrammatical.

vi. Dative divergence: In Magahi, the subject NP may be marked with the Dative case form whereas in English the subject NP cannot occur in the Dative case form.

(11) *hamara bukhaar ho.*

I Dat fever came

=> 'I have fever.'

(12) *hamara okara se nafrat ho.*

I Dat hePoss from hate be.Pres

=> 'I hate him.'

(13) *okara gosaa aailo.*

he Dat anger came}

=> 'He got angry.'

vii. Pleonastic divergence: The divergence caused by the occurrence of a pleonastic subject NP in one language and the absence of the same in the other has been termed pleonastic divergence

(Dorr 1993). English has pleonastic subject (in which the subject position is filled by a dummy NP which is without semantic content). The Magahi counterparts of such English sentences are realized differently. For instance, in (14), *it* occurs in the subject NP position but does not have any semantic content of its own. The Magahi counterpart of this sentence is formed by a subject with semantic content.

(15) It is raining.

=> *buni paRit hai.*

(rain fall Prog be.Pres)

In this type of divergence, the English sentence contains a dummy subject NP and a verb phrase. Thus, the verb {*rain*} in English is translatable in Magahi by *paani paRanaa* ‘snow’. However, in Magahi, *paRanaa* ‘fall’ is a verb and *buni* ‘rain’ is a noun. Thus, in the construction *buni paRanaa*, *buni* can be potentially taken as the subject NP and *paRanaa* its verb. Thus, there can be two English translations of the Magahi sentence *buni paRit hai*: i. It is raining, and ii. The rain is falling.

Another type of example that comes under this divergence is a type of “there-sentences” in English the Magahi counterpart of which is realized quite differently.

(16) *sabji meN namak na hai.*

tea in salt not be.Pres

=> ‘There is no salt in the vegetable.’

(17) There was a monk in India.

=> *bhaarat meN ego sadhu halai.*

{India in a monk be.Pst}

In the English sentences in (16-17), *there* occurs in the subject position without a semantic content of its own. In the Magahi counterparts of these sentences, the subject position is filled with the respective NP.

3.1.2 Lexical-semantic Divergences

i. Thematic Divergence: Thematic divergences refer to divergences that arise because of the differences in the argument structure of a verb. The Magahi counterpart of an English example in which the subject NP occurs in the dative case whereas the subject NP in English is in the nominative case can be cited as a case of thematic divergence, as in (18).

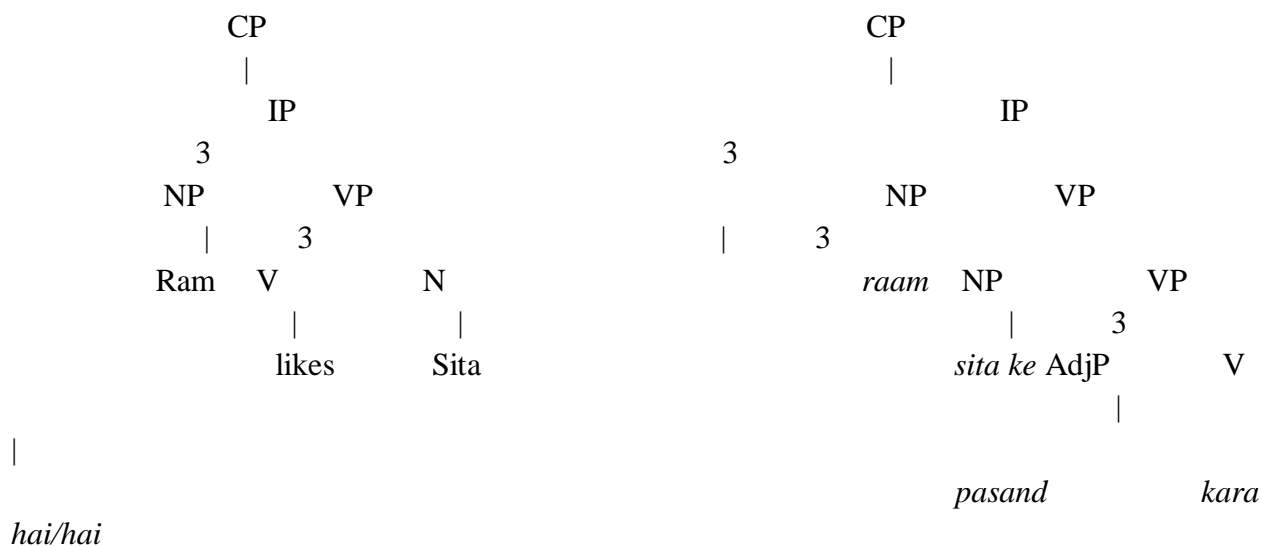
(18) Ram likes Sita.

=> i. *raam sita-ke pasand kara hai.*

Ram Sita-Acc like do be.Pres

ii. *raam-ke sita pasand hai.*

Ram-Dat Sita like be.Pres



We notice that for English to Magahi MT, there is a choice between two options in Magahi where the Magahi sentence in (i) presents no divergence, the one in (ii) presents divergence. It is also a question of lexical choice as whether the English verb *like* is treated like a transitive-active verb in Magahi with a lexical entry as *pasand karal* or like a stative verb with a lexical entry *pasand hoil*. Only in the latter case, divergence (thematic) exists. Besides the phenomenon of the dative subject constructions in Magahi, some other types of data that can be included under this class of divergence are the asymmetry between active and passive constructions in Magahi and English, the causative verbs in Magahi and their realization in English. However, they may also overlap with other classes of divergence. For instance, all the cases where the “have-sentences” in English are mapped by dative-subject constructions cannot be included under thematic divergence. These constructions can reflect different types of TDs, some of which may not be amenable to the existing classification (we discuss some of them in the following section).

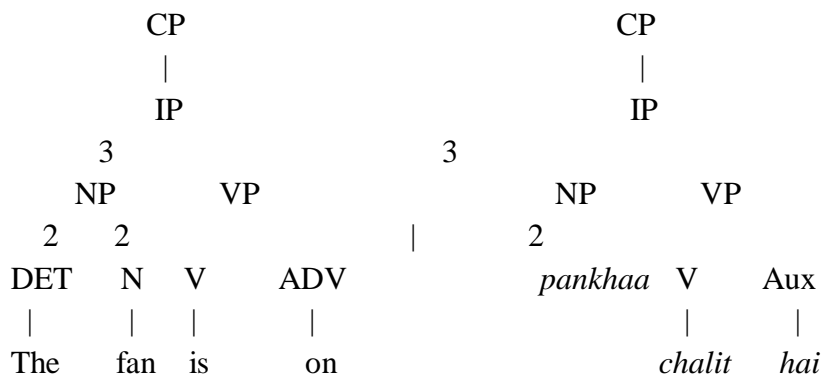
ii. Promotional Divergence: Promotional divergences arise when the position of a syntactic constituent in the source language undergoes a change to a higher up position in the target language. For instance, when an adverbial element in one language is realized by a verbal element, it constitutes a case of promotional divergence. Dorr cites an example from English-Spanish translation pair, as in (19) where an adverbial element *usually* in English is mapped by a verb in Spanish.

- (19) John usually goes home. [English]
 => Juan suele ir a casa. [Spanish]
 ‘John tends to go home.’ (Literal English translation)

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 (*chalit hai*) in Magahi counterpart.

(20) The fan is on.

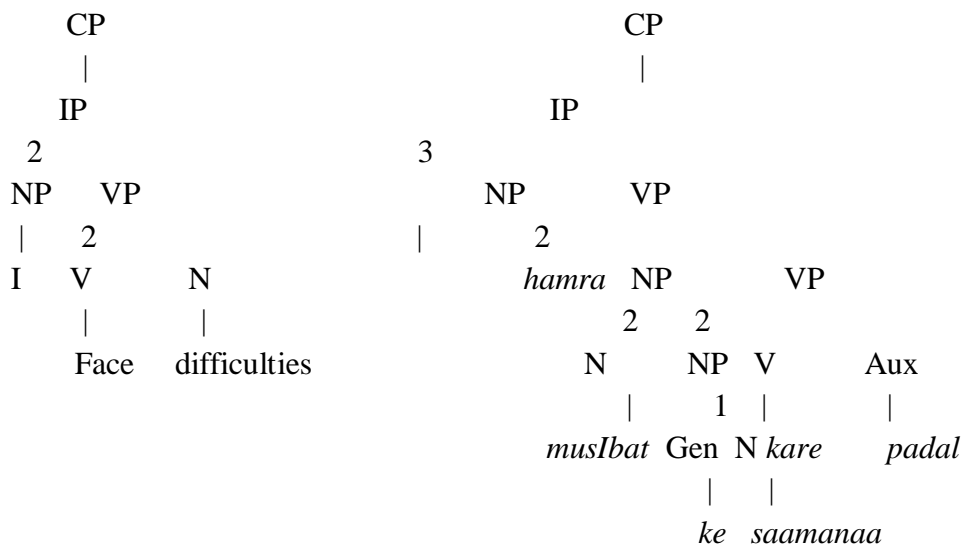
pankhaa chalit hai.
 Fan run.Prog be.Pres



iii. Structural Divergence: 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 (*musIbata ke saamanaa*) in Magahi. For instance,

21. I face difficulties.

hamara musIbata ke saamanaa kare(ke) paDal.
 I Dat difficulty of face do have.Pst.Prf



iv. Conflational and Inflational Divergence: Conflational divergence results when two or more words in the source language are translated by a single word in the target language. The opposite of this case is referred to by inflational divergence. The English verb *stab* has been cited as an example in this respect. The verb *stab* incorporates the instrumental adjunct which in Spanish is realized by the overt use of the relevant instrumental adjunct, as in (22).

(22) I stabbed John. [English]

=> Yo le di punaladas a Juan. [Spanish]

‘I gave knife-wounds to John’ (Literal English translation)

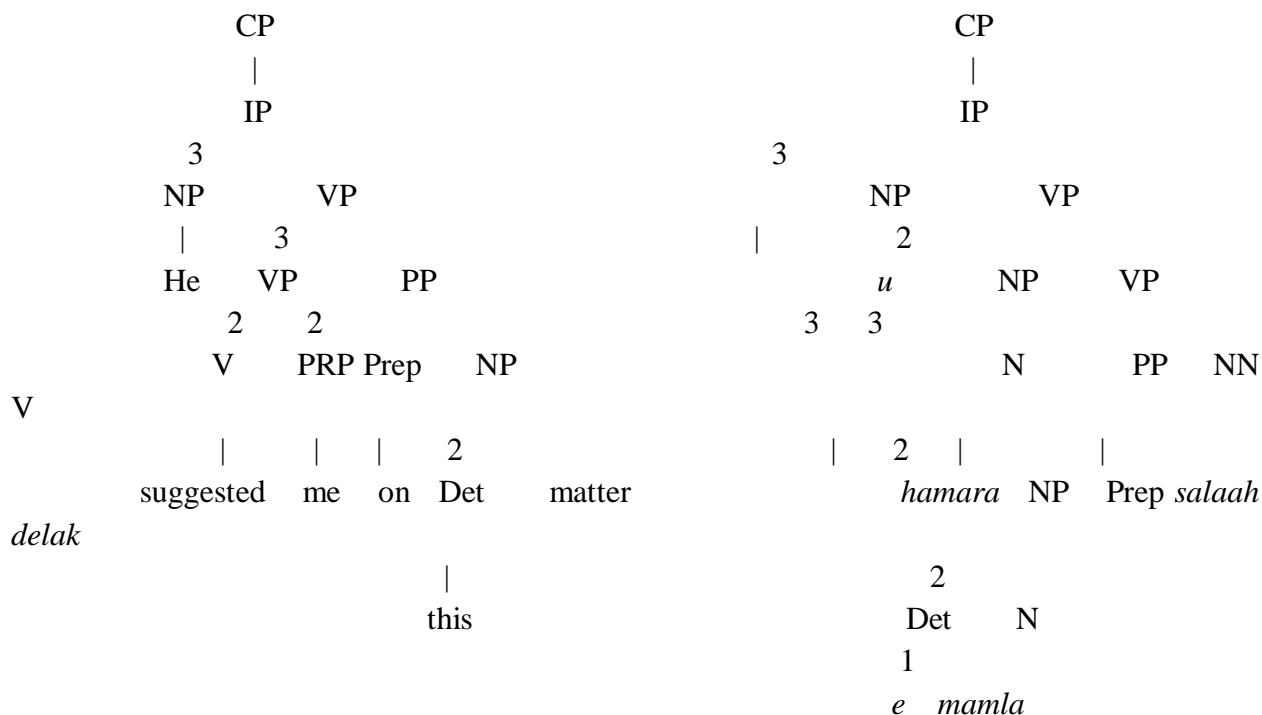
In the inflational example below, the SL ‘suggested’ gets realized in the TL as having two elements ‘*salaah delak*’. In the conflational instance, phrasal verb “looked for” gets translated in Magahi as having only one word i.e., ‘*khojalak*’.

(23) He suggested me on this matter.

u hamaraa e maamalaa meN salaah delak.
 heNom3 me Dat.Obj. this matter inObl advise give Pst Prf 3

(24) He looked for a room.

u ego makaan khojalak
 he Nom 3 a house search Pst Prf 3

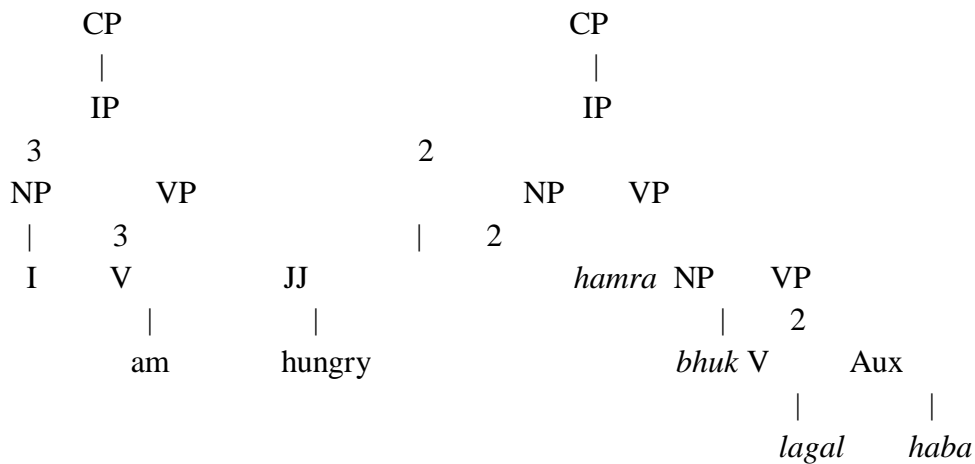


v. Categorial Divergence: Categorial divergences occur because of the mismatch between parts of speech between the source and the target languages. When there is a change in the very grammatical category of a linguistic element in TL it is known as the categorial divergence. In the example instantiated below the predicative adjective ‘hungry’ in English gets translated as the nominal phrase in Magahi. Thus, there is a change in the parts of speech categories from adjective to noun.

25. I am hungry.

hamaraa bhukh lagal haba

I Dat3 hunger seem Prf be Pres



vi. Lexical Divergence: Lexical divergence is caused when a lexical item in the source language does not have an exact map in the target language. Most of the conflational and inflational as well as some other types of divergences overlap with lexical divergence. Dorr (1994) presents an example from English-Spanish MT where the phrasal verb *break into* in English is mapped by different word/phrase in Spanish, as in (26).

(26) John broke into the room. [English]

=> Juan forzo la entrada al cuarto. [Spanish]

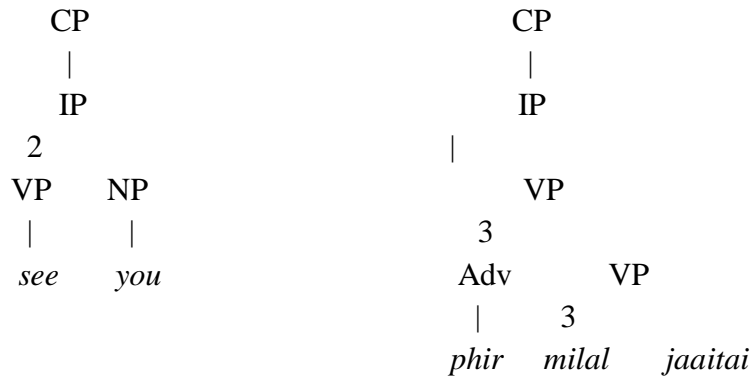
‘John forced (the) entry to the room.’ (Literal English translation)

Thus, in the following example, there is no equivalent translation for the expression ‘see you’ in Indian languages like Magahi (except Bangla where it can be exactly mapped ‘*dekha hoibe*’). There is fair amount of lexical divergences between a pair of languages that are responsible for creating divergence issues in MT.

(27) See you!

phira milal jaaitai

again meet goFut Pl 1 IPFV



In this section, we have briefly outlined the various types of TDs that have been proposed in Dorr (1993, 1994) with examples from Magahi-English and English-Magahi MT. This has been done as a background for the investigation of further issues/types in TDs in the context of Magahi-English and English-Magahi MT pairs.

4.The Identification and Proposed Resolution Procedure

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

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.

[CP [IP [NP Ram] [VP [V likes] [N Sita]]]]
 \Leftrightarrow [State BEI_{Ident} ([Thing Sita],
 [Position AT_{Ident} ([Thing Ram], [Thing like]),
 [manner SEVERELY]])]
 \Leftrightarrow [CP [IP [NP Ram] [VP [N Sita] [V loves]]]]

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.

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.

[CP [IP [NP [DET The] [N fan]] [VP [V is] [ADV on]]]]
 ⇔ [State BEIdent ([Thing THE FAN],
 [manner OFF])]
 ⇔ [CP [IP [NP *pankhaa*] [VP [V *chalit*] [AUX *hai*]]]]

4.2 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_{Loc}
 ([Thing I],
 [Path TO_{Loc} (Position IN_{Loc} ([Thing I], [Thing DIFFICULTIES]))]])
 ⇔ [CP [IP [NP *hamra*] [VP [NP [N *musIbata*] [NP [GEN *ke*] [N *saamanaa*]]] [VP [V *kare*]
 [AUX *paDal*]]]]]

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

4.3 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. “*salaah delak*” in Magahi 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_{Loc}
 ([Thing HE],
 [Path TO_{Loc} (Position IN_{Loc} ([Thing HE], [Thing ME], ([Thing THE MATTER]))]])
 ⇔ [CP [IP [NP *u*] [VP [NP [N *hamara*] [PP [NP [DET *e*] [N *maamala*]]] [Prep *meN*]]] [VP [NN
salaaha] [V *delak*]]]]]

4.4 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]_{Ident}

([Thing I],[Position AT]_{Ident} ([Thing I], [property HUNGRY]))

↔[CP [IP [NP *hamra*] [VP [NP *bhukh*] [VP [V *lagal*] [AUX *haba*]]]]]]

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 Magahi counterpart.

4.5 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 *milal*] [AUX [*jaaitai*]]]]]]]]

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 Magahi language pair is to bring out various types of divergent, incompatible or incongruent features. 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 111 between English and Magahi language pair can prove to be fruitful for any Indo Aryan 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.

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.

Pinkey 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. 1990. Semantic Structures. MIT Press, Cambridge.

Verma Manindra K. (1991) “Exploring the Parameters of Agreement: The Case of Magahi.” *Language Sciences* 13(2): 125-143.

Verma, S. *The structure of the Magahi verb*. 1985. New Delhi: Manohar Publication.

Verma, S. “Magahi”. 2003. In *The Indo- Aryan Languages*, edited by George Cardona and Dhanesh Jain, 498- 514. London: Routledge.