

Translation Divergence in English-Hindi MT

R. Mahesh K. Sinha and Anil Thakur

Department of Computer Science and Engineering
Indian Institute of Technology, Kanpur 208016
INDIA

E-mail: {rmk, anilt}@iitk.ac.in

Abstract. Divergence related to mapping patterns between two or more natural languages is a common phenomenon. The patterns of divergence between two languages need to be identified and strategies devised to handle them to obtain correct translation from one language to another. In the literature on MT, some attempts have been made to classify the types of translation divergence between a pair of natural languages. However, the issue of linguistic divergence is such a complex phenomenon that a lot more need to be done in this area to identify further classes of divergence, their implications and inter-relatedness as well as the approaches to handle them. In this paper, we take Dorr's (1994) classification of translation divergence as base and examine the translation patterns between Hindi and English to locate further details and implications of these divergences. We attempt to identify the potential topics that fall under divergence and cannot directly or indirectly be accounted for or accommodated within the existing classification. Our primary goal is to identify different patterns of translation divergence from Hindi to English and the vice versa and, on the basis of that, suggest an augmentation in the classification of translation divergence.

1. Introduction

Hindi-English language pair for (machine) translation presents a rich case of divergence at different grammatical as well as extra-grammatical levels. It is important to identify the different types of divergences to obtain correct translation for Hindi sentences to English and the vice-versa. The translation divergences has been examined in the literature on MT from different theoretical perspectives for the purpose of their proper classification and handling (Dorr 1990a, 1990b, 1993, 1994, Barnett et al 1991a, Barnett et al 1991b, see Dorr 1994 for a brief review of them). The issue of translation divergence has not been discussed in detail in the context of machine translation between English and Hindi. However, the issue has recently caught attention of scholars working in the area of machine translation in English and Hindi (including other Indian languages) but the existing works (Dave et al, 2001; Gupta et al, 2003; Goyal et al, 2004;

Sinha and Thakur, 2004) are only a beginning in the vast and complex area of translation divergence. Dave et al (2001) discuss some of the major classes of translation divergence as proposed in Dorr (1993) and outline a UNL-based interlingua approach for the handling of some of the translation divergences between English and Hindi. Gupta et al (2003) discuss some of the translation divergences for English-Hindi MT, based on the classification proposed in Dorr (1994) and suggest a unified approach for their identification and resolution. However, the issue of translation divergence is complex one and a number of significant divergence issues have remained out of the scope in the existing works on the topic. We also notice that the classification of translation divergence as proposed in Dorr (1993, 1994), which has been the basis of discussion in these works, cannot accommodate a number of topics of divergences observed with respect to English and Hindi MT. In this work, our primary task has been to identify the different types of

translation divergences in the context of English and Hindi MT with a view to classify them according to the well-defined theoretical framework as proposed in the existing literature. We also suggest relevant modification in the present classification to accommodate new categories of translation divergences wherever an appropriate class is not available.

In this paper, we examine the different areas of translation divergences both from Hindi to English and English to Hindi machine translation perspectives. We take Dorr's classification of translation divergence as the point of departure to examine the topic of divergence in Hindi and English language pair. In Section 2, we discuss classification of translation divergence as proposed in Dorr (1994) and present relevant examples from Hindi-English translation pair to examine as to what extent the present classification can be adopted for these cases and to what extent we need further classes/categories of divergence to account for the examples of translation divergences we encounter in Hindi-English and English-Hindi MT language pairs. In section 3, we examine further topics of divergence between Hindi and English MT. We discuss the translation divergences under different topics of grammar and present our observations on their classification. We also present a brief outline of the proposal for the modification in the existing classification of translation divergence to account for new categories. In section 4, we conclude the paper.

2. Dorr's Classification and Divergence in English and Hindi MT

2.1. Dorr's Classification

Dorr (1994) has identified seven classes of translation divergences. These classes are: (i) Thematic Divergence, (ii) Promotional Divergence, (iii) Demotional Divergence, (iv) Structural Divergence, (v) Conflational Divergence, (vi) Categorical Divergence, and (vii) Lexical Divergence. The classes of translation divergence have been defined to account for different types of translation divergences found in a pair of translation

languages. She points out that the translation divergences arising out of idiomatic usage, aspectual knowledge, discourse knowledge, domain knowledge, or world knowledge remain out of the scope of her paper (Dorr 1994). In this paper, too, we have largely concentrated on the translation divergences arising out of grammatical aspects of the translation languages. However, we have also pointed out a few examples/types from socio-cultural aspects of language with a view to present their case as a potential translation divergence that need to be addressed in any study pertaining to the issue. In the following section, we discuss the main classes of translation divergences as proposed in Dorr (1994) with some illustrative examples from English and Hindi.

2.2. Divergence in Hindi-English and English-Hindi MT

2.2.1. Thematic Divergence

Thematic divergence refers to those divergences that arise from differences in the realization of the argument structure of a verb. The Hindi 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 type of thematic divergence (1).

(1) John likes Mary.

=> i. *jOn mErii-ko pasand karataa hE.*

{John Mary-ACC like do be.PR}

ii. *jOn-ko merii pasand aaii.*

{John-DAT Mary like came}

iii. *jOn-ko mErii pasand hE.*

{John-DAT Mary like be.PR}

However, we can also observe that for English to Hindi MT, there is a choice among three options in Hindi where the Hindi sentence in (i) presents no divergence, the ones in (ii-iii) present divergence. It is also a question of lexical choice as whether the English verb 'like' is treated as a transitive-active verb in Hindi with a lexical entry *pasand karanaa* or like an inchoative verb with a lexical entry *pasand aanaa* or a stative verb with a lexical entry *pasand honaa*. The option in (iii) is the most difficult one to obtain because to get *pasand hE* from *pasand honaa* involves complex

procedure. In both options (ii) and (iii), divergence arises because the lexical entry of the inchoative and stative verb *pasand aanaa* and *pasand honaa* respectively select a dative subject NP rather than a nominative subject NP. Besides the phenomenon of the dative subject constructions in Hindi, some other types of data that can be included under this class of divergence are the asymmetry between active and passive constructions in Hindi and English, the causative verbs in Hindi and their realization in English. However, they may also overlap with other classes of divergence.

2.2.2. Promotional and Demotional Divergence

Promotional and demotional divergences or Head-swapping divergences arise where the status (lower or higher) of a syntactic constituent in one language is affected in another language. For instance, when an adverbial element in one language is realized by a verbal element, it constitutes a case of promotional divergence and an opposite case will result in demotional divergence. Dave et al (2001) discuss an example such as “the play is on” => *khel chal rahaa hE*. {play on PROG be.PR} as an example of this type of translation divergence in English-Hindi MT. The status of divergence with respect to this example, too, depends on the choice of the lexical entry of 'on' as it can be categorized either as an adverb or an adjective. In the later case, the example does not involve any divergence: “the play is on” => *khel caalu hE*. {play on be.PR}.

2.2.3. Structural Divergence

Structural divergences are examples where an NP argument in one language is realized by a PP adjunct/oblique NP in another language. The verb ‘enter’ in an English sentence such as “he entered the room” => *vah kamare meN paravesh kiyaa* {he room in enter did} takes an NP argument ‘the room’ whereas its Hindi counterpart *pravesh karanaa* takes a PP adjunct *kamare meN* {room in} (Dave et al, 2001). We notice that ‘enter’ is an intransitive verb and takes a locative adjunct which in English is optionally an NP or PP and in Hindi it is obligatorily a PP. In Hindi, most of the arguments as well as adjuncts NPs of the verb

are realized with an overt postposition whereas in English, they are bare NPs. A potential linguistic question remains whether to categorize such divergences under structural divergence or whether they belong to some kind of morphological gaps in the system of languages. Further, some of the passive constructions in Hindi which are used to denote a number of grammatical functions can come both under structural divergence and examples of some kind of inflectional gaps. For instance, the impersonal passive constructions in Hindi have active counterparts in English which should certainly come under structural divergence. However, at the same time, these structural gaps also indicate that whereas Hindi realizes a certain kind of mood by the use of a type of passive construction, such grammatical device is not available in English (2).

(2) *raam se calaa nahiiN jaataa*.
 {Ram by walk not PASS}
 => Ram cannot walk.

Thus the example in (2) presents not only a type of structural divergence but also a type of morphological gaps between the grammatical systems of the two languages, which needs to be identified and classified separately.

2.2.4. Conflational and Inflational Divergence

A conflational divergence results when two or more words in one language are translated by one word in another language. The opposite case is referred to by inflational divergence. The English verb ‘stab’ is generally cited as an example of this class of divergence. The verb ‘stab’ incorporates the instrumental adjunct which in other languages is realized by overt use of the relevant instrumental adjunct. For instance, in Hindi, as in (3) ‘stab’ is mapped by overt use of both the instrumental adjunct *chuuraa* ‘knife’ and verb *bhONkanaa* ‘to insert’ (see also Dave et al, 2001).

(3) *cor-ne raam-ko chuuraa bhONk-kar maar daalaa*.
 {thief-ERG Ram-ACC knife insert-CPP kill dropped}
 => The thief stabbed Ram to death.

The status of divergence in this example, too, depends on the choice of the lexical entry.

For instance, for English to Hindi MT, the meaning of the verb ‘stab’ can be entered in the lexicon as {*chuuraa bhONkanaa*}. However, for the reverse translation such a choice is quite restricted because *chuuraa* ‘knife’ can be used in another sense also with the verb *bhONkanaa* ‘to insert’. Thus there is a need to examine such cases in detail for their classification as translation divergence in Hindi and English MT.

2.2.5. *Categorial Divergence*

Categorial divergences are located in the mismatch between parts of speech of the pair of translation languages. Dorr (1994) presents an example from English-German pair (“I am hungry”=> *Ich habe Hunger*) where an adjective in English ‘hungry’ is realized by a nominal element in German *Hunger* ‘hunger’. Another often cited example is the English adjective ‘jealous’ as in the sentence “she is jealous of me” which in many languages has a verbal mapping. The Hindi counterpart of this English sentence can be either (i) *vah mujhase irshyaa karatii hE* {she me-with jealousy do} or (ii) *usako mujhase irshyaa hE* {she-DAT me-from jealousy be.PR}. We notice that in Hindi, ‘jealous’ is realized by a conjunct verb (NV) construction where *irshyaa* ‘jealousy’ is a noun and *karanaa* ‘do’ is a verb in (i), and by a dative subject construction in which *irshyaa* ‘jealousy’ is a noun that does not constitute a conjunct verb. This example also presents a case where several classes of divergence seem to be involved.

2.2.6. *Lexical Divergence*

Lexical divergence arises out of the unavailability of an exact translation map for a construction in one language into another language. An example cited in Dorr (1994) of this type of divergence is where the Spanish counterpart of an English phrasal verb ‘break into’ is realized not by a parallel phrasal verb but by a different word. The Hindi equivalent of the Spanish sentence presents an identical case (“John broke into the room”=>*John kamare meN jabardastii ghusaa.*). We notice that in Hindi, the English phrasal verb ‘break into’ is realized by a different verb *ghusanaa* ‘enter’ which takes an adverbial element *jabardastii* ‘by

force’. The example shows that the divergence pattern not only involves differences in lexical mapping but also in structural mapping between the two languages. Besides, the domain of this type of translation divergence is far from clear. Most of the conflation and inflation as well as some other types of divergences can also overlap with this category. This shows that this category of translation divergence is not well defined in a sense to account for the relevant types of divergence in an exact way.

In the above discussion, we have briefly discussed the different classes of translation divergences as proposed in Dorr’s classification of translation divergence and examined the various types of translation divergences in Hindi and English MT. In the next section, our discussion shows that a number of translation examples in English-Hindi MT present divergence which cannot be clearly accounted for within the existing classification.

3. Divergence in English and Hindi MT

3.1. Further Topics in Divergence

3.1.1. *Word Order and its Implications*

Although some of the word order related divergences can be handled within the existing classification in the form of structural divergences, we may notice that many other such divergences need more exploration. For instance, the interpretation of the question particle *kyaa* in Hindi is dependent of the word order facts of English and Hindi. *kyaa* can be used both as a marker of interrogative pronoun in content question sentences and as a question particle in yes-no question sentences. Besides certain other factors such as the category of the verb (transitive vs. intransitive), it is the position of occurrence of *kyaa* that indicates its interpretation one way or the other. When *kyaa* occurs in the sentence-initial and sentence-final positions, it is generally interpreted as question particle whereas when it occurs in the sentence-medial position, it is an interrogative pronoun, as is illustrated in (4).

- (4) a. *aap kyaa khaa rahe hEN?*
 {you IP eat PROG be.PR}
 => What are you eating?

- b. *kyaa aap khaa rahe hEN?*
 {QP you eat PROG be.PR}
 c. *aap khaa rahe hEN kyaa?*
 {you eat PROG be.PR QP}
 => Are you eating?

The classification presented in Dorr (1994) does not seem to discuss the implications of the word order patterns of a language in triggering divergence. Dave et al (2001) discuss only obvious word order difference between Hindi and English without examining its implications for divergence. The subtle implications of the word order fact in the translation languages is a potential source of divergence that may fall in different classes of divergence.

3.1.2. Replicative Words

Hindi, like most of the South Asian languages, has replicative words for which it is difficult to find an exact counterpart in European languages such as English. Almost all kinds of words can be replicated to denote a number of different functions in Hindi. A noun-noun replication such as *bachchaa bachchaa* {'child child'} is used to denote quantification of the noun and can be expressed in English by a quantifier phrase 'every child'. A verb-verb replication such as *dekhate-dekhate* {'see see'} can be used to denote different types of functions that are mapped onto English in various ways depending on a number of factors. For instance, *dekhate dekhate saberaa ho gayaa* {'see see morning happened'} literally means 'It became morning while we were watching'. However, this is not the right English translation of the Hindi sentence. A closer translation will be 'In the meanwhile, it became morning.' In another case, the same verbal replication as in *cinemaa dekhate dekhate vah bor ho gayaa* {'movie see see he bored became'} functions as a reason adverbial which in English is mapped by a prepositional phrase with a gerund head as in 'He got bored by (repeatedly) watching movies'. While almost all the verbal lexical item can occur in replicative form, only a restricted class of nominal lexicon can be replicated. There has been debate whether this phenomenon of replication is amenable to grammatical rules or not (Abbi 1992). However, there is no doubt that this

presents a potential area of divergence from Hindi to English MT. The same cannot be said to be true regarding English to Hindi MT. For instance, The English counterparts of the Hindi examples of a noun-noun replication mentioned above can be obtained without resorting to replication of the relevant items. The use of replication is an optional device, which has other extra-grammatical functions such as focus/emphasis or style. In (5), the Hindi counterpart of the English sentence can be formed either with or without resorting to the phenomenon of replication

- (5) a. We know every child.
 => i. *ham pratyek bachche ko jaanate hEN.*
 {we every child ACC know be.PR}
 ii. *ham bachche bachche ko jaanate hEN.*
 {we child child ACC know be.PR}

This shows that the nature and types of translation divergences are different in the case of reverse translation. There are other dimensions of the replication phenomenon in Hindi which are the source of different types of translation divergences between Hindi and English. One cannot study this phenomenon under lexical divergence for the simple reason that the mapping patterns involved for these replicated words reflects a number of syntactic as well as semantic aspects.

3.1.3. Determiner System

English has (in)definite articles that mark the (in)definiteness of the noun phrase overtly. Hindi lacks an overt article system and different devices are used to realize the (in)definiteness of a noun phrase in Hindi. For instance, mapping of a bare NP in Hindi onto an NP with an article 'a-an/the' in English is dependent on a detailed syntactic and semantic analysis of the noun phrases in both the languages (e.g. *laRakaa aayaa* => the/*a boy came.). This gap in the systems of the grammar of these languages is the cause of divergence that cannot be properly categorized within the existing classification of translation divergence.

Another related point of divergence between Hindi and English is regarding the mapping of there- and it-sentences in English. In English, there- and it-constructions are used to denote

existential sentences (besides others). Hindi does not have a pleonastic subject construction and the contrast between existential and non-existential (mostly definite) sentences is realized by several other ways such as the movement of the noun phrase from its canonical position and the use of demonstrative elements.

(6) *jāngal meN sher hE.*
 {forest in lion be.PR}
 => There is a lion in the forest.

(7) *sher jāngal meN hE.*
 {lion forest in be.PR}
 =>The lion is in the forest.

We notice that the English interpretation of the Hindi sentence in (6) is not possible for the Hindi sentence in (7). However, the only difference between these two Hindi sentences is the respective positions of the subject NP and the (place) adverbial phrase. This type of divergence is related to more than one aspect of grammar such as the word order, lexical and structural gaps in languages, etc. and need to be examined in detail to categorize the type of divergence it represents.

3.1.4. Morphological Gaps

Another important area related to divergence that has not been properly recognized in the existing literature is the strategies the different natural languages adopt to denote different modality and aspectual properties of the verb. For instance, Hindi uses a certain type of passive construction that marks a certain kind (non-volitionality) of modality function. The English counterparts of such Hindi sentences are only partially able to express the exact meaning.

(8) *raam se shiishaa TuuT gayaa.*
 {Ram by glass break PASS}
 => i. The glass got broken by Ram.
 ii. Ram broke the glass unintentionally.

(9) *raam se galatii ho gaii.*
 {Ram by mistake happen PASS}
 => i. Ram made a mistake.
 => ii. *The mistake got made by Ram.

We notice that the Hindi sentences in (8) and (9) have identical structure but they cannot be realized identically in English. The second

English interpretation of (9) is not possible whereas in (8) the second translation is closer to the intended sense of the Hindi sentence. The possible English counterpart of the Hindi sentence (9) is far from the actual sense in which the Hindi impersonal passive has been used. The literal sense will be somewhat like this: 'Ram made a mistake unintentionally'. Thus we notice that to capture the exact meaning of an impersonal passive sentence in Hindi, English not only uses an active sentence but also has to resort to other devices (such as lexical insertion) to fill the gaps.

Another instance of gaps between the systems of the Hindi and English grammars can be located in the case of the realization of certain tense and aspectual properties of the verb. In Hindi, they are realized by verbal inflection whereas English resorts to different non-inflectional ways such as phrasal verb or an adverbial element or the prepositional phrase with gerund as the head, to realize these aspectual properties of the verb. For instance, the Hindi sentences in (10) and (11) are identical except the difference in tense. The habitual aspect of the tense is reflected by inflectional morphology on the verb in both the tenses. However, this habitual aspect in English is realized by the use of a phrasal verb in the case of the past tense (10) and by the use of an adverbial word 'often' in the case of the present (and future) tense (11). Thus the adverbial element in Hindi is optional whereas the one in English is not. In (12), we notice that a certain type of non-terminative/continuative aspect is realized by the verbal morphology in Hindi whereas English uses two structures; a verb 'keep' and a prepositional phrase 'on speaking' with gerund as head, to realize the same aspectual feature.

(10) *vah yahaaN aayaa karataa thaa.*
 {he here come.HAB do.IMP}
 => He used to come here.

(11) *vah yahaaN (aksar) aayaa karataa hE.*
 {he here (often) come.HAB do.IMP
 be.PR}
 => He *(often) comes here.

(12) *vah bolataa rahaa.*
 {he speak.IMP PROG}
 => He kept on speaking.

The existing classification is not quite clear as to where and how to account for these aspects of divergence. We need to examine this issue in detail to further identify different implications as well as the handling patterns for these divergences.

In this class we can also discuss cases of divergence related to transitivity and causativity in Hindi and English.

(13) *raam haNsaa.*

{Ram laughed}

=> Ram laughed.

(14) *raam-ne siitaa-ko haNsaayaa.*

{Ram-ERG Sita-ACC laugh-TRS}

=> Ram made Sita laugh.

(15) *raam-ne siitaa-ko mohan se*

haNsavaayaa.

{Ram-ERG Sita-ACC Mohan-by laugh-CAUS}

=> Ram got Mohan make Sita laugh.

We notice from examples in (13-15) that, in Hindi, there are three forms of a verb (in this case *haNsanaa* 'laugh') that are morphologically derived. (*haNsanaa* => *haNsaanaa* => *haNsavaanaa*). The English counterparts of these sentences show that in English, there is only one lexical verb 'laugh' and the other form are realized by syntactic processes (such as resorting to extended verbal constructions). In Hindi, *haNsaanaa* is a transitive verb which does not have a lexical counterpart in English (English has only the intransitive form as a lexical item). In English, it is realized by using two verbs 'make' and 'laugh'. *haNsavaanaa* is a lexical causative verb in Hindi which in English is realized by using three verbs 'get', 'make' and 'laugh', with separate argument structure of their own. The English counterpart of the Hindi example in (15) appears to be a forced translation. In certain cases, it is quite difficult to obtain an exact translation of a common Hindi ditransitive verb. For instance, in (16), the English counterpart of the transitive verb *piinaa* is 'drink'. However, Hindi also has a ditransitive verb *pilaanaa* derived from *piinaa*. English does not have a counterpart of this ditransitive verb.

(16) *maan-ne bachche-ko paanii pilaane kii koshish kii*

{mother-ERG child-ACC water drink-DIT of try did}

=> Mother tried to "feed" water to the child.

Gaps of this type are quite common between Hindi and English and can be identified as different types of morphological gaps.

3.1.5. *Stative Verb*

A class of verb such as 'sit', 'stand', 'sleep', and 'wake up' exhibit divergence with respect to the realization of their aspectual and participial forms. For these verbs in English, there is no distinction between the progressive aspectual form and the participial form (adjectival). For instance, 'sitting' can mean either *bETHaa (huua)* or *bETHa rahaa* in Hindi. However, in this case, the reverse translation also causes divergence. For instance, in (18), 'is sleeping' in English can be mapped both by *so rahaa hE* and *soyaa hE* in Hindi. In Hindi, they are distinguished by different lexical form of the relevant verb.

(17) *raam kursii par bETHaa hE.*

{Ram chair on sitting be.PR}

=> Ram is sitting on a chair.

(18) John is sleeping.

=> *jOn so rahaa hE. or, jOn soyaa hE.*

{John sleep PROG be.PR}, {John asleep be.PR}

These instances of divergence need to be separately categorized because of their different behavior.

3.1.6. *Conjunctions, Particles and Punctuation Marks*

The use of different conjunctions, punctuation marks, and particles in Hindi, are another source of divergence. Some of these particles such as *ki*, *na*, *yaa* and *vaalaa* have functional roles in Hindi that are mapped in English by different means than can be identified on the basis of the syntactic structure. *ki* is mainly used as a sentence complementizer, but can also be used to indicate alternate conjunction in an affirmative sentence (19) and an interrogative sentence (20) in Hindi.

(19) *raam siitaa se pyaar karataa hE naa ki gitaa se.*

{Ram Sita with love do be.PR not

- or Gita with}
 => Ram loves Sita not Gita
 (20) *raam siitaa se pyaar karataa hE ki*
gita se?
 {Ram Sita with love do be.PR or
 Gita with}
 => Does Ram love Sita or Gita?

Similarly, *yaa* 'or' is a conjunction particle in Hindi that conjoins two clauses or phrases. However, it can denote a different function in Hindi depending on the punctuation mark used in the relevant sentence. Notice the contrast between (21-22).

- (21) *vah dilli gayaa hE yaa kolkata.*
 {he Delhi went be.PR or-AFF
 Kolkata}
 => He has gone either to Delhi or Kolkata.
 (22) *vah dilli gayaa hE yaa kolkata?*
 {he Delhi went be.PR or-INT}
 => Has he gone to Delhi or Kolkata?

Certain clausal conjunctions in Hindi does not have structural counterparts in English where they are realized by different devices such as modal or auxiliary verbs.

- (23) *ho naa ho vah kahiiN gayaa ho.*
 {May be he somewhere went
 be.SUBJ}
 => He might have gone somewhere.

The particle *vaalaa* is another source of divergence in Hindi-English MT. It is used to denote a number of functions which are realized in English by different linguistic devices. (24) presents an example where it is used as a participial modifier in Hindi and is realized by a relative clause in English.

- (24) *dillii meN rahane-vaale log acche hote*
hEN.
 {Delhi in live-of people good
 be.IMP be.PR}
 => The people who live in Delhi are good.

Thus the divergence related to the use of different conjunctions, particles, and punctuation marks need further exploration for their proper classification.

3.1.7. Expressive and Echo Words

Expressive and echo words are source of divergence between Hindi and English. There is no (exact) parallel available for these lexical items in English. This may be related to the socio-cultural and even anthropological aspects of a natural language.

- (25) *patte kharakharaa rahe the.*
 {leaves 'kharakhara' PROG be.PST}
 => The leaves were making a *kharakhara* sound.

In (25), we notice that an expressive word in Hindi has been used as verb which does not have an equivalent in English. Thus we need to reproduce the same Hindi word in English translation, too.

Another related typological feature exhibited by all Indian languages is the occurrence of echo words where a lexical word is partially replicated to denote a wide range of meanings with subtle semantic constraints. The examples in (26-27) are illustrative.

- (26) *aap caay vaay pii kar jaaiye.*
 {you tea etc. take CPP go.IMPR}
 => Have some snacks before going.
 (27) *ise Thiik se jaaNc vaaNc lijiyegaa.*
 {it.OBL properly check 'vaaNc'
 take.IMPR}
 => Please examine it properly.

Echo words have no lexical status in the lexicon of the language. In (26), the use of the echo word *vaay* along with the main word *caay* 'tea' gives the sense of light refreshment. However, this is not a possible sense in which an echo word is used in (27). Here the main verb *jaaNcaanaa* 'examine' occurs with an echo word that has only an emphatic (or extension) function but it cannot be exactly expressed in English.

These are typical characteristic features of all the Indian languages and a translation system has to take into account the divergence associated with these phenomena.

3.1.8. Gerunds and Participle Clauses

Another significant level of divergence in Hindi and English can be located in the way the various clausal complements and adjunct (such

as verbal participles) in Hindi are realized in English.

(28) *vah yahaan aakar khush hE.*

{he here come.CPP happy be.PR}

=> He is happy to come here.

(29) *ham aapase milane (ke liye) aaye hEN.*

{we you see.GER.OBL (for)
come.PRF be.PR}

=> We have come to see you.

(30) *ham vahaan jaane meN samarth hEN.*

{we there go.GER.OBL in able
be.PR}

=> We are able to go there.

(31) *ham paDhanaa caahate hEN.*

{we read-GER want be.PR}

=> We want to study.

In the Hindi sentences in (28-31), both the different types of adjunct verbal clauses and complement verbal clauses are realized by different structures. In English, on the other hand, they all are realized by an infinitival clause. Thus, to capture exact mappings of these sentences for English to Hindi MT is a far more complex task (Sinha and Thakur, 2004) and is a difficult topic of divergence to handle.

3.1.9. Honorific

In Hindi, honorific features are expressed by several linguistic markers including the use of plural pronoun and plural verbal inflections. This feature is not available in a European language such as English in a similar way. This causes a complex topic of divergence in the sense that it interacts both with the grammatical and socio-cultural aspects of the language.

(32) *unake pitaa aaye hEN.*

{his father come be.PR.PL}

=> His father has come.

(33) *unakaa nOkar aayaa hE.*

{his servant come be.PR.SG}

=> His servant has come.

In (32), the subject *pitaa* 'father' is an honorific noun which is reflected by the use of plural inflectional elements on the agreeing elements such as verb and the genitive noun. On the other hand, in (33), *nOkar* 'servant' is a non-honorific noun and no plural inflectional element is used in the sentence. In the existing

classification, the translation divergences of this kind have not been pointed out.

3.2. A Proposal

The topics of divergence discussed in the previous section are reflective of the complexity one encounters in Hindi to English and English to Hindi MT. The classification of divergence proposed in the existing literature (Dorr 1993, 1994) and some of the later works that examine divergence for Hindi and English MT (Dave et al, 2001; Gupta et al, 2003) are reflective of only some of the obvious divergence issues. On the basis of the discussion above, we propose some modifications in the existing classification of translation divergences. We present a sample of them with a view to give a direction of the modifications needed.

- The class of structural/syntactic divergence need to be widened by proposing new subclasses to include topics of divergence such as Word Order and its Implication (3.1.1), Determiner System (3.1.3), some parts of Morphological Gaps (3.1.4), Gerunds and Participial Clauses (3.1.8).
- A new class (Morphological Gaps) needs to be created to account for most of the topics in Morphological Gaps (3.1.4).
- A new well-defined class (Lexical Mapping) needs to be created to accommodate topics such as Replicative Words (3.1.2), Expressive and Echo words (3.1.7), also some part of Honorific (3.1.9).
- A new class (Socio-cultural Gaps) needs to be created to accommodate topics such as Honorific (3.1.9) and some part of Expressive and Echo Words (3.1.7).
- A new class (Particles and Punctuation Marks) need to be created to accommodate topics under Conjunctions, Particles and Punctuation Marks (3.1.6).

4. Concluding Remarks

In this paper, we have examined the issue of classification of translation divergence for MT between English and Hindi. We have taken into account the classification of translation divergence as proposed in Dorr (1994) and

shown that to capture the various types of translation divergence between Hindi and English, we need to further modify the classification and augment it by new categories and subtypes. We have pointed out the various point of divergence between Hindi and English which have not been directly or indirectly discussed in the existing literature on divergence. However, due to constraints on space, it has not been possible to include more detailed discussions on issues of divergence related to different topics in section 3.1. We have presented an outline of the proposal for the modification in the existing classification in section 3.2.

On the basis of the discussion presented in this paper, we have shown that the translation divergence between Hindi and English machine translation is more varied and complex than the works in the existing literature can accommodate and account for. To obtain correct translation, we need to examine the different grammatical as well as some of the extra-grammatical characteristics of both Hindi and English to exhaustively identify the types of translation divergence in this pair of translation languages. Some of the topics, particularly those related to socio-cultural aspects of language need further exploration in light of the complexity in their formalization.

Abbreviations

ACC: Accusative Case, AFF: Affirmative, CAUS: Causative, CPP: Conjunctive Participle, DAT: Dative Case, DIT: Ditransitive, ERG: Ergative Case, Fu: Future Tense, GER: Gerund, HAB: Habitual Aspect, HON: Honorific Marker, IMP: Imperfective Aspect, IMPR: Imperative Mood, INT: Interrogative, PASS: Passive Participle, PL: Plural, PR: Present Tense, PST: Past Tense, SG: Singular, SUBJ: Subjunctive Mood, TRS: Transitive, VPRT: Verbal Participle.

References

ABBI, Anbita (1992). *Reduplication in South Asian Languages: An Areal, Typological and Historical Study*. New Delhi: Allied Publishers.

BARNETT, Jim, Inderjeet Mani, Paul Martin, Elaine Rich (1991a). 'Reversible Machine

Translation: What to do When Languages Don't Line Up'. In *Proceedings of the Workshop on Reversible Grammar in Natural Language Processing, ACL-91*, 61-70, University of California, Berkeley, CA.

BARNETT, Jim, Inderjeet Mani, Elaine Rich, Chinatsu Aone, Kevin Knight, Juan C. Martinez (1991b). 'Capturing Language-Specific Semantic Distinctions in Interlingua-Based MT'. In *Proceedings of Machine Translation Summit*, 25-32, Washington, DC.

DAVE, Sachi, Jignashu Parikh, and Pushpak Bhattacharyya (2001). 'Interlingua-based English-Hindi Machine Translation and Language Divergence'. *Machine Translation 16 (4)*, pp. 251-304.

DORR, Bonnie (1990a). 'A Cross-Linguistic Approach to Machine Translation'. In *Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, 13-32, Linguistic Research Center, The University of Texas, Austin, TX.

DORR, Bonnie (1990b). 'Solving Thematic Divergence in Machine Translation'. In *Proceedings of the 28th Annual Conference of the Association for Computational Linguistics*, 127-134, University of Pittsburg, Pittsburg, PA.

DORR, Bonnie (1993). *Machine Translation: A View from the Lexicon*. Cambridge, Mass: The MIT Press.

DORR, Bonnie (1994). 'Classification of Machine Translation Divergences and a Proposed Solution'. *Computational Linguistics, Vol. 20 (4)*, pp. 597-633.

DORR, Bonnie, Necip Fazil Ayan, and Nizar Habash (2004). 'Divergence Unraveling for Word Alignment of Parallel Corpora'. *Natural Language Engineering, 1 (1)*, pp. 1-17.

GOYAL, Shaily, Deepa Gupta and Niladri Chatterjee (2004). 'A Study of Hindi Translation Patterns for English Sentences with "Have" as the Main Verb'. In *Proceeding of iSTRANS, Vol. 1*, pp. 46-51.

GUPTA, Deepa, and Niladri Chatterjee (2003). Identification of Divergence for English to Hindi EBMT. In *Proceeding of MT Summit-IX*, pp. 141-148.

SINHA, R.M.K. and Anil Thakur (2004). 'Synthesizing Verb Form in English to Hindi Translation: Case of Mapping Infinitive and Gerund in English to Hindi'. In *Proceeding of iSTRANS, Vol. 1*, pp. 52-55.