Linguistic Divergence Patterns in English to Marathi Translation

S. B. Kulkarni
Assistant Professor
Dept. of CS & IT,
Dr. B. A. M. U.
Aurangabad

P. D. Deshmukh
Assistant Professor
Dept. of English,
Pandit J. Nehru College,
Aurangabad

M. M. Kazi
Research Student
Dept. of CS & IT,
Dr. B. A. M. U.
Aurangabad

K. V. Kale
Professor and Head
Dept. of CS & IT,
Dr. B. A. M. U.
Aurangabad

ABSTRACT

In machine translation system, the text is translated from one language known as source language into another language known as target language. The development of a machine translation system needs to identify the patterns of divergence between two languages. A detail study of divergence issues in machine translation is required for their proper classification and detection. The primary objective of this paper is to understand the types of divergence problems that operate behind English to Marathi translation. In this paper, the various divergence patterns between English-Marathi language pair are considered. This will enable us to come up with strategies to handle these situations and coming up with correct translation.

General Terms
Natural Language Processing, Machine Translation, Computational Linguistics

Keywords
Linguistic Divergence, Lexico-Semantic Divergence, Syntactic Divergence, English – Marathi Translation

1. INTRODUCTION

The issue of translation divergence is a complex topic in machine translation (MT). The translation divergence can be defined in terms of language-to-language differences in the respective grammar. Thus a divergence occurs when a sentence in language L1 translates into a sentence in L2 in a very different form [2, 8]. The topic has been studied from different perspectives and a number of approaches have been proposed to handle them. It is crucial for any MT system to identify the nature of translation divergences and resolve them so as to obtain correct translation [5, 6, 7]. The translation divergences occur at different levels and affect the quality of the translation according to the degree of complexity involved in a particular translation divergence. It has also been noted that certain types of translation divergences are universal in the sense that they exist across the languages whereas certain other types of translation divergences are specific to a pair of translation languages [11, 13, 14]. Therefore, the translation divergences need to be studied from both across-language and language-specific perspectives [3, 4]. This paper examines English and Marathi translation language pair largely from the perspective of identifying the language-specific divergences. English and Marathi differ in many respects and hence this translation language pair presents a rich source for the study of translation divergence in Machine Translation [1, 17, 18]. These languages also show significant differences from the point of view of syntactic, structural and socio-cultural perspectives that need to be properly examined [8, 9, 10]. This paper considers Dorr’s classification of translation divergence for the basic classification of divergence. The various other types of divergence with respect to English and Marathi language pair are focused [2, 16, 19]. Section 2, deals with the literature survey of classification of translation divergence proposed by Dorr [4, 5] and others. In section 3 various Lexico-semantic and syntactic divergences are considered. Some of the other types of commonly found divergence patterns in English-Marathi language pair are discussed in section 4 and the paper concluded in section 5.

2. LITERATURE SURVEY

The divergence study is with respect to language pair. Till date considerable amount of research have been carried out on the nature and patterns of translational divergence between English and Hindi [20, 21, 22]. Based on this studies efforts are made for developing automatic translation systems between this language pair: English and Hindi [23]. Also the work is found for the English to Bangali language pair [20]. The lexico-semantic divergence in Urdu-to-English language pair is done in Example Based Machine Translation approach [24, 25, 26]. Divergences occur at different levels and severely affect the quality of a translation. Dorr (1994) proposes ways to look into this aspect of translation in minute details between any two languages involved in translation. Based on issues related to problems that arise in translation between languages [2, 3, 4]. Based on large number of empirical examples found in English, Spanish, and German corpora, Dorr has classified divergences into two broad types and defined them accordingly:

(a) Syntactic Divergence, and
(b) Lexical-semantic Divergence

There are no such studies have been made yet to understand the nature and patterns of divergence between English and Marathi language pair Therefore, keeping this requirement in sight the paper explores and tries to understand the areas of linguistic divergence in translation between the languages.

3. DIVERGENCES IN ENGLISH TO MARATHI MACHINE TRANSLATION

3.1 Lexico-semantic Divergence

3.1.1 Thematic Divergence

Thematic divergence refers to those divergences that arise from differences in the realization of the argument structure of a verb. The verbal object in one language becomes as the subject of the main verb in other language. The Marathi 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. जॉन मेरीला पसंद करतो.
   (John-Mary-ACC like do be.PR)
   ja”ana maorIla psant krtoa.

ii. जॉनला मेरी पसंद पडली.
    (John-DAT Mary like came)
   ja’anala maorI psant pDlaal.

iii. जॉनला मेरी पसंद आहे.
     (John-DAT Mary like be.PR)
   ja’anala maorI psant Aaho

However, one can also observe that for English to Marathi MT, there is a choice among three options in Marathi where the Marathi 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 Marathi with a lexical entry pasant karane or like an inchoative verb with a lexical entry pasant padane or a static verb with a lexical entry pasant aasne.

3.1.2 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”

\[ \text{त्याने खेळावण्यास प्रवेश केला.} \]

\[ \text{tyaano KaalatmaQao p/vaolSa kola.} \]

[he room in enter did]
takes an NP argument ‘the room’ whereas its Marathi counterpart pravesh Kela takes a PP adjunct Kholti madhe [room in]. We notice that ‘enter’ is an intransitive verb and takes a locative adjunct which in English is optionally an NP or PP and in Marathi it is obligatorily a PP.

3.1.3 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. In case of English to Marathi the example is

“the play is on”

\[ \text{खेळ चालू आहे.} \]

\[ \text{khel calu ahaa.} \]

[play on PROG be.PR] as an example of this type of translation divergence in English-Marathi 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.

3.1.4 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

Marathi, as in (3) ‘stab’ is mapped by overt use of both the instrumental adjunct sura ‘knife’ and verb bhoskne ‘to insert’.

(3) चोराने रामाला सुरा ऑसकून मात्र टाकले.
   cora-ne raam-la suai bhOskun maaron takale
   caaoranaa ramala sura BaaSaauna maaSHAa Takla
   {thief-ERG Ram-ACC knife insert-CPF 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 Marathi MT, the meaning of the verb ‘stab’ can be entered in the lexicon as [ suri bhOskun ]. However, for the reverse translation such a choice is quite restricted because suria ‘knife’ can be used in another sense also with the verb bhOskun ‘to insert’. Thus there is a need to examine such cases in detail for their classification as translation divergence in Marathi and English MT.

3.1.5 Categorial Divergence

Categorial divergences are located in the mismatch between parts of speech of the pair of translation languages. In case of categorial divergence changes in category. For example, the predicate is adjectival in one language but nominal in other language. Consider the example as

I am feeling hungry.

मला भृक लागली आहे.

=> malaa Buk laagalal Aaho.

In English “hungry” is adjective and but in Marathi “bhukhi” (hunger) becomes the noun.

3.1.6 Lexical Divergence

Lexical divergence arises out of the unavailability of an exact translation map for a construction in one language into another language. Consider counterpart of an English phrasal verb ‘break into’ is realized not by a parallel phrasal verb but by a different word. The Marathi equivalent of this sentence presents an identical case

John broke into the room.

१. जॉन जबरदस्ती खोलीमध्ये शिरले.

=> ja’aa na jabardstalnaa Aat ISaarla.

We notice that in Marathi, the English phrasal verb ‘break into’ is realized by a different verb shirne ‘enter’ which takes an adverbial element jabardastine ‘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 conflational and inflational 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.

3.2 Syntactic Divergence in English-Marathi Translation

3.2.1 Constituent-order Divergence

Constituent-order divergence relates to the word-order distinctions between English and Marathi. Essentially, the constituent order describes where the specifier and the complements of a phrase are positioned. For example, in English the complement of a verb is placed after the verb and the specifier of the verb is placed before. Thus English is an SOV language. Marathi, on the other hand, is an SOV
language. Example (I) shows the constituent-order divergence between English and Marathi.

(I) Ram is playing tennis.

S V O
राम टेनिस खेळतो आहे.
RAM TENNIS PLAYING IS.

Also, in Marathi, the qualifier of the complement succeeds the verb whereas in English, it succeeds the complement, consider example (II)

(II) He saw a girl whose eyes were blue.

S V O Q
त्याने एका मुळेला पर्यंते जिथे डोळे निघते होते.
HE ONE GIRL SAW WHOSE EYES BLUE WERE.

3.2.2 Adjunction Divergence

Syntactic divergences associated with different types of adjunct structures are classified as Adjunction divergence. Hindi and English differ in the possible positioning of the adjective phrase. In Marathi a PP can be placed between a verb and its object or before the object, while in English it can only be at the maximal level (i.e., not between the verb and its object (II)).

(I) a. He called me [to his house].
b. He called [to his house] me.

(II) a. त्याने मला त्याच्या घरी बोलले.

HE TO ME HIS HOUSE CALLED.

b. त्याने त्याच्या घरी मला बोलले.

HE HIS HOUSE TO ME CALLED.

3.2.3 Null-subject Divergence

In Marathi, unlike in English, the subject of the sentence can be left implicit as in (I), for example.

(I) Long ago, there was a king.

खूप पूर्वी एक राजा होता.
LONG AGO ONE KING WAS

Marathi allows dropping of the subject where the subject is obvious as in (I), repeated here for convenience.

3.2.4 Pleonastic Divergence

Another related point of divergence between Marathi 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). Marathi does not have a pleonastic subject construction and the contrast between existential and nonexistential (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.

(I) जंगलानप्रमिध्ये वाघ आहे.

[forest in lion be PR]
There is a lion in the forest.

बाघ जंगलानप्रमिध्ये आहे
[lion forest in be PR]
The lion is in the forest.

We notice that the English interpretation of the Marathi sentence in (I) is not possible for the Marathi sentence in (II). However, the only difference between these two Marathi 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.

4. COMMONLY FOUND DIVERGENCE IN ENGLISH AND MARATHI MACHINE TRANSLATION

We found some commonly found divergence patterns in English and Marathi Language pair, which are not considered in the earlier classification. Some of them are given in the following section.

4.1 Replicative Words

Marathi, 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 Marathi. A verb-verb replication such as pahata-pahata (‘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.

पहाता पहाता सकाच झाली.

=> pahata pahata skal. Jaalal.

literally means ‘it became morning while we were watching’. However, this is not the right English translation of the Marathi sentence. A closer translation will be ‘In the meanwhile, it became morning.’

4.2 Determiner System

English has articles that mark the (in)definiteness of the noun phrase overtly. Marathi lacks an overt article system and different devices are used to realize the (in)definiteness of a noun phrase in Marathi. For instance, mapping of a bare NP in Marathi onto an NP with an article ‘the/a-anithe’ in English is dependent on a detailed syntactic and semantic analysis of the noun phrases in both the languages for instance

मूलगा आला
Maulagaa Aalaa.

=> 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.

4.3 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, Marathi uses a certain type of passive construction that marks a certain kind (non-volitionality) of modality function. The English counterparts of such Marathi sentences are only partially able to express the exact meaning.

(I) रामा कढून काच फुटला.

{Ram by glass break PASS}

i. The glass got broken by Ram.

ii. Ram broke the glass unintentionally.

(II) रामा कढून चूक झाली.

{Ram by mistake happen PASS}

i. Ram made a mistake.
ii. *The mistake got made by Ram.
We notice that the Marathi sentences in (4) and (5) have identical structure but they cannot be realized identically in English. The second English interpretation of (5) is not possible whereas in (4) the second translation is closer to the intended sense of the Marathi sentence. The possible English counterpart of the Marathi sentence (5) is far from the actual sense in which the Marathi 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 Marathi, English not only uses an active sentence but also has to resort to other devices (such as lexical insertion) to fill the gaps.

4.4 Honorific Difference
In Marathi, 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.

(6) त्याचे वडील आले आहेत.
*His father come be.PR.PL*  
=> His father has come.

In (6), the subject vadil ‘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. In the existing classification, the translation divergences of this kind have not been pointed out.

4.5 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 kaya in Marathi is dependent of the word order facts of English and Marathi. kaya 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 generally interpreted as question particle when it occurs in the sentence medial and sentence final positions, it is generally interpreted as interrogative pronoun, as is illustrated in (1).

(1) a. आपण काय खात आहात?
What are you eating?

b. काय आपण खात आहात?
{QP you eat PROG be.PR}

c. आपण खात आहात काय?
{You eat PROG be.PR QP}
Are you eating?

5. CONCLUSION
Some important types of divergence related to translation from English to Marathi are discussed in this paper keeping in mind both manual and machine translation. In this paper, we have examined the various patterns of translation divergence in English and Marathi language pair. We have taken into account the classification of translation divergence considering the Lexico-sematic is summarized in Table 1.

<table>
<thead>
<tr>
<th>Divergence Type</th>
<th>English Sentence</th>
<th>Marathi Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thematic Divergence</td>
<td>John likes Mary</td>
<td>जॉन मेमीला पसंत करतो</td>
</tr>
<tr>
<td>Structural Divergence</td>
<td>He entered the room</td>
<td>त्याने बॉलायच्या प्रवेश केला.</td>
</tr>
<tr>
<td>Promotional and Demotional Divergence</td>
<td>The play is on</td>
<td>खेळ चालू आहे.</td>
</tr>
<tr>
<td>Conflation and Inflationary Divergence</td>
<td>The thief stabbed Ram to death.</td>
<td>घोराले रामाला सूर्य भोसलून मारून दाखले.</td>
</tr>
<tr>
<td>Categorial Divergence</td>
<td>I am feeling hungry.</td>
<td>मला भूख असली आहे.</td>
</tr>
<tr>
<td>Lexical Divergence</td>
<td>John broke into the room</td>
<td>जॉन जबरदस्तीने खोलायच्ये शिरिला.</td>
</tr>
</tbody>
</table>

syntactic divergence is shown in table 2 and the commonly found divergence patterns in English and Marathi language pair is summarized in Table 3. We have considered the various point of divergence between Marathi and English which have not been directly or indirectly discussed in the existing literature on divergence. To obtain correct translation, we need to examine the different grammatical as well as some of the extra-grammatical characteristics of both Marathi and English to exhaustively identify the types of translation divergence in this pair of languages. This will enable us to come up with strategies to handle these situations and coming up with correct translation.

<table>
<thead>
<tr>
<th>Divergence Type</th>
<th>English Sentence</th>
<th>Marathi Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constituent-order Divergence</td>
<td>Ram is playing tennis.</td>
<td>राम टेनिस खेळल.</td>
</tr>
<tr>
<td>Adjunction Divergence</td>
<td>He called me [to his house].</td>
<td>त्याने मला त्याच्या घरी बोलावले.</td>
</tr>
<tr>
<td>Null-subject Divergence</td>
<td>Long ago, there was a king.</td>
<td>खूप पूरी एक राजा होता.</td>
</tr>
<tr>
<td>Pleonastic Divergence</td>
<td>There is a lion in the forest.</td>
<td>जंगलामध्ये वाघ आहे.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Divergence Type</th>
<th>English Sentence</th>
<th>Marathi Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicative Words</td>
<td>It became morning while we were watching</td>
<td>पाहून पाहून सकाळ झाली.</td>
</tr>
<tr>
<td>Determiner System</td>
<td>The/*a boy came.</td>
<td>मुलगां आला.</td>
</tr>
</tbody>
</table>
6. ABBREVIATIONS

7. ACKNOWLEDGMENTS

The authors would like to acknowledge and extend our heartfelt gratitude to the UGC who have funded for development of UGC SAP (II) DRS Phase-I F.No-3-42/2009 major research project.

8. REFERENCES


9. AUTHOR’S PROFILE
Mrs. Sonali B. Kulkarni, SMIEEE, SMACM, MIETE, Assistant Professor, Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) - India.

Dr. Pradnya D. Deshmukh, Assistant Professor, Department of English, Pandit Jawaharlal Nehru College, Aurangabad (MS) - India.

Mr. M. M. Kazi, Research Student, is full time PhD Student working in the Department of Computer Science and IT, DR. B. A. M. University, Aurangabad (MS) – India 431004.

Dr. K. V. Kale, Professor & Head, M.Sc, MCA Ph.D. FIETE, Presently working as a professor and head, Department of Computer Science and Information Technology, Dr Babasaheb Ambedkar Marathwada University, Aurangabad, MS-India. He is a fellow of IE TE, Life member of CSI, IAPR, ISCA, IEEE and elevated to senior member of IEEE.