



INTERNATIONAL JOURNAL OF
RESEARCH IN COMPUTER
APPLICATIONS AND ROBOTICS
ISSN 2320-7345

**LINGUISTIC TO SOCIO-AND-PSYCO LINGUISTIC
ASPECTS IN ENGLISH-TO-MARATHI LANGUAGE
TRANSLATION**

S. B. Kulkarni¹, P. D. Deshmukh², M. M. Kazi³, K. V. Kale⁴

^{1,3,4}Dept. of Comp. Sci. and IT, Dr. B. A. M. University, Aurangabad (MS) – **India**

²Second Department of English, Pundit Jawaharlal Nehru College, Aurangabad (MS) - **India**

Abstract

The translation divergence is a challenging problem in the area of machine translation. A detailed study of divergence issues in machine translation is required for their proper classification and detection. The language divergences between English and Marathi can be considered as representing the divergences between SVO (Subject - Verb - Object) and SOV (Subject - Object - Verb) classes of languages. This paper discusses translation patterns between English and Marathi to identifying the potential topics of translation divergences. The typical type (specific to language pair such as English and Marathi) of divergence is based on different aspects such as linguistic to socio-and psycho-linguistic, grammatical differences and socio-cultural aspects.

Keywords: linguistic divergence, socio-cultural 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 (Dorr, 1993, Sinha *et. al.*, 2005b). 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 (Gupta *et. al.* 2003, Masud *et. al.* 2003, Khan *et. al.* 2003). 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 (Lavanya *et. al.*, 2005, Saboor *et. al.*, 2010, Remya *et. al.*, 2009). Therefore, the translation divergences need to be studied from both across-language and language-specific perspectives (Dorr, 1994, Sinha, 2005). In this paper, we examine 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 (Menno *et. al.*, 2009, Antony, 2013, Nair *et. al.*, 2012). These languages also show significant differences from the point of view of syntactic, structural and socio-cultural perspectives that need to be properly examined (Sinha *et. al.*, 2005b, Zang *et. al.*, 2008, Ellison *et. al.*, 2006). We have considered Dorr's classification of translation divergence for the basic classification of divergence. Also, we have considered the various other types with respect to English and Marathi language pair (Dorr, 1993, Tidke *et. al.*, 2013, Dave *et. al.*, 2001). Section 2, deals with the literature survey of classification of translation divergence proposed by Dorr (Sinha, 2005, Gupta *et. al.*, 2003). In Section 3 various aspects such as linguistic to socio-and psycho-linguistic, grammatical differences and socio-cultural divergence is considered. Some of the other types of commonly found divergence patterns in English-Marathi language pair are discussed in Section 4 and we have concluded the paper 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 (Dash, 2013, Gupta *et. al.*, 2003, Sinha *et. al.*, 2005c). Based on this studies efforts are made for developing automatic translation systems between this language pair: English and Hindi (Dash, 2004). Also the work is found for the English to Bangali language pair (Dash, 2013). The lexico-semantic divergence in Urdu-to-English language pair is done in Example Based Machine Translation approach (Saboor *et. al.*, 2010). 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 arises in translation between languages (Dorr, 1993, Dorr, 1994, Sinha, 2005). 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

We do not found 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 we have tried to explore and understand the areas of linguistic divergence in translation between the languages.

3. Socio- Cultural Divergence patterns in English to Marathi Translation

3.1 Honorificity Marker

In Marathi, the honorific feature is marked by the pluralization of the verb and the use of plural pronominal elements whereas in English it is not the case.

Ex. A. a. राष्ट्रपती आला आहे, तो आता भाषण करणार आहे.
{president come CPT be.PR. he now lecture deliver.FU}
=> The President has arrived. He will deliver a lecture now.
<=> राष्ट्रपती आले आहेत, ते आता भाषण करणार आहेत.
{president come CPT be.PR. he now lecture deliver.FU}

In both English-Marathi and Marathi- English Machine Translation, the divergence caused by this socio-cultural aspect of the respective language arises.

3.2 Role of Kinship Relation

It is found that all relatives whether he is the brother of one's father or brother of one's mother, is called uncle in English but in the case of Marathi there are different expressions for these relatives. 'kAkA' is used to denote paternal uncle, 'mAmA' is used for maternal uncle. The same is true in case of 'aunt'. It is very difficult to recognize proper relation when translating from English to Marathi.

3.3 Mappings of Time

Usually, people's perception of different objects in the world is dependent upon several socio-cultural beliefs. For instance, time is conceptualized in the Indian culture differently than that is done in the Western culture. These concepts are expressed through our respective languages and difference in concepts manifests itself in the language that is the source of translation divergence. For instance, in English, the concept of a.m. vs. p.m. cannot be exactly mapped in Marathi. The Marathi counterpart of a.m. and p.m. denote only a small part of time and the other parts of time is denoted by different other terms. The example in (*Ex. A. a.*) shows that the time at the 5 o'clock in the morning is denoted by a.m. in English but the exact translation of a.m. in Marathi does not produce an appropriate Marathi expression. However, in (*Ex. A. b.*), we notice that the time at the 11 o'clock in the morning which is expressed in English by a.m. can also be expressed in Marathi by the exact translation of the term a.m. A similar situation is noticeable with respect to the mapping of p.m. in examples (*Ex. A. c.-e.*).

Ex. A. a. He arrived at 5 a.m.

=> तो सकाळी ५ वाजता आला.

{he five o'clock a.m. came}

<=> He arrived at 5 o'clock in the morning.

Ex. A. b. He arrived at 11 a.m.

<=> तो सकाळी ११ वाजता आला.

{he eleven o'clock a.m. at came}

Ex. A. c. He arrived at 3 p.m.

<=> तो दुपारी ३ वाजता आला.

{he three o'clock p.m. came}

Ex. A. d. He arrived at 5 p.m.

=> तो संध्याकाळी ५ वाजता आला.

{he five o'clock evening at came}

<=> He arrived at 5 o'clock in the evening.

Ex. A. e. He arrived at 7 p.m.

=> तो रात्री ७ वाजता आला.

{he seven o'clock night at came}.

<=> He came at 7 o'clock in the night.

3.4 Optative Sentences

The verb in the subordinate clause in the sentences of optative mood in Marathi occurs in different forms depending on the gender, number and person of the subject NP of the subordinate clause whereas its English counterpart remains constant (root form of the verb) in all the cases. In this case, the verb form does not indicate tense. The divergence is triggered by the similarity of the form of the verb with other tense forms in the case of English whereas it is not the case in Marathi as shown in (*Ex. D. a.*)

Ex. D. a. आम्हाला वाटते की राम यशस्वी होवो.
{ we want be.PR that Ram successful be.OPT }
=> We want that Ram succeed.

This type of divergence can be resolved by taking into account the semantic type of the verb.

3.5 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,

pahata-pahata sakal zali.
पाहता पाहता सकाळ झाली.
=> pahta pahta skaL JaalaI.

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'.

3.6 Expressive Elements

Expressive words exist in all natural languages and pose difficulty in processing, particularly in mapping onto another language. The reason is that these words do not have exact parallel in another language. Thus the word धपकन/धाडकन is only distantly mapped by 'bump' in English, as in (Ex. F. a.).

(Ex. F. a.) ती धपकन पडली.
{she 'dhaRaam' with fell}
=> She fell with a 'bump'.

The expressive words usually originate from the sound associated with the semantics of the action verb and can be adverbial or verbalized action-verbs. One may argue this to be just a lexical gap but indeed it is not so. However, some of these words can be handled in the lexicon but as in many cases the mapping also involves structural changes, the issue involves a wider scope of interpretation.

4. Other aspects in English to Marathi Translation

4.1 Let-sentences

The Marathi permissive-sentences are mostly translated into English by "let-sentences", as in (Ex. A. a.). However, there are certain wish-sentences that also occur in the form of a let-sentence (Ex. A. b.).

(Ex. A. a.) Let him go.
<=> त्याला जाऊ दया.

(Ex. A. b.) चला, जेवण करुया.
{go food eat.OPT}
=> Let us go and eat now.
आपण सगळे जाऊया आणि खाऊया.
{us now eat let}

The difference in the English sentences between (*Ex. A. a.*) and (*Ex. A. b.*) is only in the use of a pronoun. The use of first person plural pronoun 'us' in (*Ex. A. b.*) makes the sentence a wish-sentence rather than a permissive-sentence. Thus the reverse translation in (*Ex. A. a.*) does not involve translation divergence whereas in (*Ex. A. b.*), a translation divergence occurs. The nature of this translation divergence again pertains to the gaps in the realization of different verbal inflections and functions between English and Marathi.

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 'a-an/the' in English is dependent on a detailed syntactic and semantic analysis of the noun phrases in both the languages for instance in example (*Ex. B. a.*)

(*Ex. B. a.*) **Maulagaa Aalaa.**

मुलगा आला.

=> 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.

Consider the examples (*Ex. C. a.*) and (*Ex. C. b.*)

(*Ex. C. a.*) => **रामाकडून काच फुटला.**

{Ram by glass break PASS}

=> i. The glass got broken by Ram.

=> ii. Ram broke the glass unintentionally.

(*Ex. C. b.*) **रामाकडून 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 (*Ex. C. a.*) and (*Ex. C. b.*) have identical structure but they cannot be realized identically in English. The second English interpretation of (*Ex. C. b.*) is not possible whereas in (*Ex. C. a.*) the second translation is closer to the intended sense of the Marathi sentence. The possible English counterpart of the Marathi sentence (*Ex. C. b.*) 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 News Headings

The news headings in English and Marathi follow different grammar rules . In English, generally the present tense form of the verb is used whereas Marathi uses past tense form of the verb. See example (*Ex. D. a.*)

(*Ex. D. a.*) सुनामीमध्ये लाखो लोक मेले.
{tsunami in millions died}
=> millions **die** in tsunami.
<=> सुनामीमध्ये लाखो लोक मरत आहेत.
{tsunami in million people die}

In this case, translation in both the Marathi-English and English-Marathi cases involves divergence.

4.5 Indirect Speech

The indirect speech sentences in Marathi and English differ in both the form of tense and the use of pronominal elements. Consider the following example (*Ex. E. a.*)

(*Ex. E. a.*) राम म्हणाला की, मी जाणार नाही.
{Ram-ERG told that I not go-FU}
=> Ram said that I/he would not go.
<=> राम म्हणाला की, मी/तो नाही जाणार
{Ram-ERG told that he not go-FU}

(*Ex. E. b.*) Ram said that I was coming.
=> राम म्हणाला की मी येणार आहे.
{Ram- ERG said that I come PROG be.PR}
=> *Ram said that I **am** coming.

The use of the pronoun मी 'I' in (*Ex. E. a.*) is ambiguous and can be translated either by 'I' or 'he' in English. The example in (*Ex. E. b.*), shows that the tense in the English indirect speech sentences is past but must be mapped by present tense in the Marathi sentence.

4.6 Had-Counterfactual Clause

In Marathi, the counterfactual conditional clause is marked by a conjunction जर / तर ('if') which in English can be realized either by a had-clause or an if-clause. In the former case, translation divergence occurs. Consider the following example (*Ex. F. a.*)

(*Ex. F. a.*) जर तू इथे असता तर मी ही आलो असतो.
{if you here be.SUBJ then we also come-SUBJ}
<=> **Had** you been here we would have also come.

In reverse translation for the English sentence, the divergence remains the same.

4.7 Stative and Progressive Aspect

English seems to lack an exact counterpart of Marathi stative verb/adjective which is realized by the progressive aspect marker. In English, there is no distinction between the progressive aspect denoting sentence and its stative counterpart. The English verbs such as *sit*, *stand*, *sleep*, and *wake* fall in this category. In Marathi, they are distinguished by different lexical form of the relevant verb as shown in example (*Ex. G. a.*)

(Ex. G. a.) राम खुर्चीवर बसलेला आहे.

{Ram chair on sitting be.PR}

=> Ram is **sitting** on a chair.

<=> राम खुर्चीवर बसत आहे.

{Ram chair on sitting PROG be.PR}

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 from English to Marathi. We have examined the various patterns of translation divergence in English and Marathi language pair. In this paper, we described the typical type of divergences in English to Marathi machine translation which is based on different aspects such as linguistic to socio-and psycho-linguistic and socio-cultural aspects. We have taken into account the classification of translation divergence considering the Socio-cultural aspects which is summarized in Table 1 and some other types of divergence patterns are shown in Table 2. 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.

Abbreviations

ACC: Accusative Case, CPP: Conjunctive Participial, Particle, DAT: Dative Case, ERG: Ergative Case, Fu: Future Tense, HON : Honorific Marker, IMP: Imperfective Aspect, PASS: Passive Particle, PL: Plural, PR: Present Tense, PST: Past Tense, SG: Singular, SUBJ: Subjunctive Mood, TRS: Transitive.

Acknowledgment

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.

Table 1. Socio-Cultural aspects in English to Marathi Translation

Divergence Type	Marathi Sentence	English Sentence
Honorificity Marker	(I) राष्ट्रपती आला आहे, तो आता भाषण करणार आहे. (II) राष्ट्रपती आले आहेत, ते आता भाषण करणार आहेत.	The President has arrived. He will deliver a lecture now.
Role of kinship relation	(I) काका (II) मामा	Uncle
Mappings of Time	तो सकाळी ५ वाजता आला.	He arrived at 5 a.m.
Optative Sentences	आम्हाला वाटते की राम यशस्वी होवो.	We want that Ram succeed
Replicative words	पाहता पाहता सकाळ झाली.	It became morning while we were watching
Expressive Elements	ती धपकन पडली.	She fell with a 'bump'.

Table 2. Other type of Divergence in English to Marathi Translation

Other	Marathi Sentence	English Sentence
Let-sentences	त्याला जाऊ दया.	Let him go.
Determiner System	मुलगा आला.	the/*a boy came.
Morphological Gaps	रामाकडून काच फुटला.	The glass got broken by Ram.
News Headings	सुनामीमध्ये लाखो लोक मेले.	Millions die in tsunami.
Indirect Speech	रामने म्हटले की मी जाणार नाही.	Ram said that I/he would not go.
Had-Counterfactual Clause	जर तू इथे असता तर मी ही आलो असतो.	Had you been here we would have also come.
Stative and Progressive Aspect	राम खुर्चीवर बसलेला आहे.	Ram is sitting on a chair.

REFERENCES

- Menno Zaanen and Colin Higuera, 2009, Grammatical Interface and Computational Linguistics, *In the proceedings of EACL 2009 workshop, Greece, ACL*
- B. Dorr, 1993, Machine Translation: a View from the Lexicon, *The MIT Press, Cambridge, Mass.*
- B. Dorr, 1994, Classification of Machine Translation Divergences and a Proposed Solution, *Computat. Linguistics 20(4)*, pp. 597-633.
- R. M. K. Sinha and Anil Thakur, 2005, Translation Divergence in English-Hindi MT, *In the Proceeding of EAMT Xth Annual Conference, Budapest, Hungary, 30-31 May*
- Gupta, Deepa, and Niladri Chatterjee, 2003, Identification of Divergence for English to Hindi EBMT, *In Proceeding of MT Summit-IX*, pp. 141-148
- Md. Abu Nisar Masud, Md. Munasir Mamun, 2003, A General Approach to Natural Language Generation, *In Proceeding of IEEE, INMIC*.
- S. Khan, Z. Parvez, 2003, An Expert System Driven Approach to generating Natural Language in Romanized from English Documents, *In Proceeding of IEEE, INMIC*
- R. M. K. Sinha and Anil Thakur, 2005b, Handling ki in Hindi for Hindi-English MT, *In the Proceeding of MT Summit X, Bangkok, 12-16 September*
- Min Zang, Hongfei Jiang, 2008, Grammar comparison study for Translation Equivalence Modeling and Statistical Machine Translation, *In the Proceeding of the 22nd International Conference of Computational Linguistics*, pp. 1097-1104
- T. Mark Ellison, Simon Kirby, 2006, Measuring Language Divergence by Intra-Lexical Comparison, *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pp. 273-280.
- Prahallad Lavanya, Prahallad Kishore, Ganapathiraju Madhavi, 2005, A simple approach for building transliteration editors for Indian languages, *J Zhejiang Univ SCI, 6A(11)*, pp. 1354-1361
- Debasri Chankrabarti, Hameng Mandalia, Ritwik Priya, 2008, Hindi Compound Verbs and their Automatic Extraction, *Coling-2008, Manchester*, pp. 27-30
- Saboor, A.; Khan, M.A., 2010, Lexical-semantic divergence in Urdu-to-English Example Based Machine Translation, *Emerging Technologies (ICET), 6th International Conference on*, pp. 316-320
- Remya Rajan, Remya Sivan, Remya Ravindran, K.P. Soman, 2009, Rule Based Machine Translation from English to Malayalam, *International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pp. 439-441.
- Hatem, A.; Omar, N.; Shaker, K., 2011, Morphological analysis for rule based machine translation, *Semantic Technology and Information Retrieval (STAIR), 2011 International Conference on*, pp. 260-263
- Charugatra Tidke, Shital Binayakya, Shivani Patil, Rekha Sugandhi, Inflection Rules for English to Marathi Translation, *International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 4, April 2013*, pg.7 – 18
- Antony P. J., 2013, Machine Translation Approaches and Survey for Indian Languages, *Computational Linguistics and Chinese Language Processing Vol. 18, No. 1, ACL*, pp. 47-78

- Latha R. Nair and David Peter S, 2012, Machine Translation Systems for Indian Languages, *International Journal of Computer Applications* (0975 – 8887) Volume 39– No.1
- Dave, S., Parikh, J., & Bhattacharyya, P., 2001, Interlingua based English-Hindi machine translation, *Journal of Machine Translation*, 16(4), <http://dx.doi.org/10.1023/A:1021902704523>, pp. 251-304
- Niladri Sekhar Dash, 2013, Linguistic Divergences in English to Bengali Translation, *International Journal of English Linguistics*; Vol. 3, No. 1
- Gupta, D., & Chatterjee, N., 2003, Identification of divergence for English to Hindi EBMT, *Proceeding of MT Summit-IX*, pp. 41-48
- Sinha, R. M. K., & Thakur, A., 2005c, Divergence patterns in machine translation between Hindi and English, *Proceeding of MT Summit X. Phuket, Thailand*, pp. 346-353
- Dash, N. S., 2004, Issues involved in the development of a corpus-based machine translation system, *International Journal of Translation*, 16(2), pp. 57-79
- Abdus Saboor, Mohammad Abid Khan, 2010, Lexical-Semantic Divergence in Urdu-to-English Example Based Machine Translation”, *6th International Conference on Emerging Technologies (ICET) @ 2010 IEEE*

A Brief Author Biography

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. Majharoddin. 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 IETE, Life member of CSI, IAPR, ISCA, IEEE and elevated to senior member of IEEE