Linguistic Divergence in various Language pair in Machine Translation Perceptive

Simran N. Maniyar[1], Sonali B. Kulkarni[2,] Pratibha R. Bhise [3]

[1](Departemet of Computer Sciece and Information Techonology, Dr.Babasaheb Ambedker Marathwada University,Auragabad, India)
[2](Departemet of Computer Sciece and Information Techonology, Dr.Babasaheb Ambedker Marathwada University,Auragabad, India)
[3](Departemet of Computer Sciece and Information Techonology, Dr.Babasaheb Ambedker Marathwada University,Auragabad, India)

Abstract:
Machine translation is a big domain in research area. In machine translation we translate source language to target language. That time their occur divergence. Divergence is the big problem for machine translation. In this paper, we review on translation divergences between two Languages pair. In Machine translation various types of divergence are present. Lexical divergence, symentatic divergence, structural divergence, thematic divergence etc. We study all the previous work done in language divergence and examine the types of divergence. Our goal is identify the different types of divergence problem from Marathi to English and the vice-versa. This paper hopes to open up new avenues for future research in the area of translation divergence between two language pair not yet been studied in close details.
Key Word: *Machine Translatio; Lingustic divergence; Artificial Neweral Network.*

I. Introduction

Machine Translation is a domain of research that combines ideas and techniques from Linguistics, Computer Science, Artificial Intelligence, Translation theory and Statistics for automating the process of translation from one language to another. In Machine Translation use various method of translate source language to target language between two language pair. Most of use a first language is English to other language. Indian constitution has 22 'scheduled' or national languages and almost 2000 dialects. Though only about 5% of the world's population speaks English as a first language. English is very widely used in media, commerce, science, technology, and education in India. In such a situation, there is an obvious large market for translation between English and the various Indian languages. The divergence is a major issue in translation between to natural languages. Divergence can be defined as the differences that occur in language with respect to the grammar. The divergence mainly occurs when these occur a translation from a source language to the target language. Marathi-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 Marathi sentences to English and the vice-versa.

In any Machine Translation system, this topic is very much needed as to obtain an accurate translation, it is important to resolve the nature of translational divergence . Divergence can be seen at different levels. Based on the complexity that occur in the specific translation, divergence affects the language translation quality. According to Dorr [1993], "translation divergence occurs when the common translation of one language into another results in a very different form than that of the original". Due to online resources, Machine Translation between two languages is more in demand. A lot of work has been done in Machine Translation for Indian languages but the success rate has been poor due to poor handling of divergence and other issues between the source language and target language which cause different kinds of errors and comprehensibility issues. 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 following table shows.

II. Literature survey

Paper Name	Author	year	Language pair	Rule based	Technique	Types of divergence	Future scope
A Study on Divergence in Malayalam and Tamil Language in Machine Translation Perceptive [1]	Jisha P Jayan, Elizabeth Sherly	-	Malayalam and Tamil	certain rules	statistical machine learning techniques	The paper focuses on Thematic, promotional, demotional, structural, conflflational, categorical and lexical divergences.	-
Divergence patterns between English and Sanskrit machine translation [2]	Vimal Mishra 1 R.B.Mishra 2	-	English and Sanskrit.	detection rules, adaptation rules	ANN	conjunctions and particles related divergence, participle related divergence and gerunds related divergence	Our future work is carrying to perform case based reasoning in combination of rule based and ANN model for this purpose
Linguistic Divergences in English to Bengali Translation [3]	Niladri Sekhar Dash	2013	English to Bengali	well-investigated	rule-based MT	Syntactic Divergence, and ,Lexical-semantic Divergence.	We can also think of initiating similar studies between Hindi and Bengali to recognize as well as understand what kinds of problem, divergence, and convergence exist between the languages. Once we succeed to address these problems, then developing a translation system from Hindi to Bengali (or vice versa) cannot be a distant dream.
Statistical Vs Rule Based Machine Translation; A Case Study on Indian Language Perspective [4]	Sreelekha S.		English-Indian langugae	morphology, syntax, lexical selection and transfer and generation	SMT, RBMT	-	-
Classification and Resolution of Linguistic Divergences in English-Urdu Machine Translation [5]	Sharmin Muzaffar, Pitambar Behera & Girish Nath Jha	2016	English-Urdu	-	syntactic structure and lexical-semantics	lexical-semantic and syntactic divergences	
Lexical Resources for Hindi Marathi MT [6]	Sreelekha. S, Pushpak Bhattacharyya		Marathi to Hindi and Hindi to Marathi	Morphology	SMT Subjective Evaluation method	-	Our future work will be focused on investigating more lexical resources for improving the quality of Statistical Machine Translation systems and there by develop an accurate MT system for both Marathi-Hindi and Hindi-Marathi Machine Translation.
Role of	SREELE	2017	Indian	-	Phrase-	.	A possible future

Morphology Injection in SMT: A Case Study from Indian Language Perspective [7]	KHA S and PUSHPA K BHATTA CHARYY A		Languag e		based Statistical Machine Translation		work is to generalize the approach of morphology generation and verify the effectiveness of morphology injection on more morphologically complex languages
Comparison of SMT and RBMT, The Requirement of Hybridization for Marathi – Hindi MT [8]	Sreelekha .S, Raj Dabre Pushpak Bhattacha ryya	2013	Marathi – Hindi	-	SMT and RBMT	Lexical Ambiguity, Structural Ambiguity	Our future work will be focused on the in tegration of Rule-Based system components namely the Morphological Analyses into the Statistical Machine Translation system and there by develop a Hybridized MT system for Marathi- Hindi Machine Translation.
Translation Divergences in Chinese–English Machine Translation: An Empirical Investigation [9]	Dun Deng, Nianwen Xue	2016.	Chinese –English	syntax-based translation rules	SMT, Hierarchical	translation divergences,	-
Dealing with Linguistic Divergences in English-Bhojpuri Machine Translation [10]	Pitambar Beh, Neha Maurya 1 & Vandana Pandey 2	2016	English-Bhojpuri	heuristic linguistic rules	Dorr's Lexical Conceptual Structure	lexical-semantic and syntactic divergence	-
Divergence Identification and Handling for English-Maithili Machine Translation [11]	Ritu Nidhi, Tanya Singh	2019	English-Maithili			lexical semantic divergences	
A Novel Approach for English to South Dravidian Language Statistical Machine Translation System [12]	Unnikrish nan P, Antony P J, Dr. Soman K P	2010	English to South Dravidia n	transfer rules	phrase-based statistical machine translation.	-	-
Translation Divergence in English-Hindi MT [13]	R. Mahesh K. Sinha and Anil Thakur	2005	English-Hind		UNL-based interlingua, Dorr's	(i)ThematicDiverg ence, (ii) PromotionalDiverg ence, (iii) Demotional Divergence, (iv) Structural Divergence, (v) ConflationalDiver gence, (vi) Categorial Divergence, and (vii) Lexical Divergence.	-
Divergence	R.	-	Hindi	-	Dorr's,	Syntactic	-

Patterns in Machine Translation between Hindi and English [14]	Mahesh K. Sinha, Anil Thakur		and English		UNL-based Interlingua, empirical base MT,Rule-Based MT	Divergences, ,Lexical-semantic Divergences.	
Study of Divergence for Example Based English-Hindi Machine Translation [15]	*Deepa Gupta Niladri Chatterjee*	2001	English-Hindi	syntactic rules,transfer rule	EBMT	Syntactic divergence ,Lexical – Semantic divergence	-
Linguistic Divergence of Sinhala and Tamil languages in Machine Translation [16]	W.S.N. Dilshani, S. Yashothara, R.T.Uthayasanker, S. Jayasena	2018	Sinhala – Tamil	-	Dorr's Classificaton,SMT	Lexical semantic divergence	the English language is used as the linking language in Sri Lanka, this research opens up new research areas in order to find the divergences among Sinhala- English as well as Tamil-English languages
Applying Divergence Principles to French to English Machine Translation [17]	Violaine PRINCE, Jacques CHAUCHE	2009	French to English	transformation rules	transfer-based translation	Syntactic divergence	-
Identification of Divergence for English to Hindi EBMT [18]	Deepa Gupta and Niladri Chatterjee	2003	English to Hindi	-	Example-Based Machine Translation	Identify all types divergence	-
Interlingua-based English–Hindi Machine Translation and Language Divergence [19]	SHACHI DAVE, JIGNASHU PARIKH and PUSHPAK BHATTACHARYYA	2001	English–Hindi	morphological, syntactic and semantic	Interlingua and transfer-based approaches	lexical-semantic divergence	The L-UW dictionary has to be enriched enormously both in terms of the UW content and the semantic attributes so as to capture the word and world knowledge
Handling Structural Divergences and Recovering Dropped Arguments in a Korean/English Machine Translation System [20]	Chung-hye Han1, Benoit Lavoie2, Martha Palmer1, Owen Rambow3, Richard Kittredge2, Tanya Korelsky2, Nari Kim4, and Myunghee Kim2	2000	Korean to English	Transfer rules based on syntactic lexico-structural transfer	Transfer rules based	Structural Divergences	-
Divergence and Ambiguity Control in an English to Arabic Machine Translation [21]	Marwan Akeel*, R. B. Mishra	2013	English to Arabic	morphological, syntactic	ANN and rule based machine translation system	Lexical, syntactic ambiguity, and the structural divergences	-
Linguistic Divergence	S. B. Kulkarni,	2014	English to	-	-	Identify all types	-

Patterns in English to Marathi Translation [22]	P. D. Deshmukh, M. M. Kazi, K. V. Kale		Marathi				
SOLVING THEMATIC DIVERGENCES IN MACHINE TRANSLATION [23]	Bonnie Doff*	-	Spanish, English, and German.	transfer rules.	direct approach	Thematic divergence	-
Lexical-Semantic Divergence in Urdu-to-English Example Based Machine Translation [24]	Abdus Saboor, Mohammad Abid Khan	2010	Urdu-to-English	-	Example Based Machine Translation	Lexical-Semantic Divergence	-
Translation Divergence in English-Sanskrit-Hindi language pairs [25]	Pawan Goyal1 and R. Mahesh K. Sinha	-	English-Sanskrit-Hindi	-	-	Identify all types	-

III.   Types of Divergence

Thematic Divergence: Thematic divergence refers to those divergence that arise from differences in the relation of argument structure of verb. The verbal object is one language convert as the subject of the main verb in another language

1.   ram belives krushana.

--राम कृष्णांवर विश्वास ठेवतो.

ram krushna ver vishwas tevto

--राम ला कृष्णा वर विश्वास आहे.

ram la krushana ver vishvas aahe.

Here In English sentence main verb is belive. This main verb convert of the static verb of Marathi language(vishvas thavto). Here presrnt the themantic divergence.

 Conflational and Inflational divergence: Conflational divergences exist when the sense transfer by a single word is communicating by two or more words in one of the languages. The contradictory case of conflational divergence is indicating to by inflational divergence.

The Ajay stabbed tiger to death.

अजयने वाघाला सुरा भोसकून मारून टाकले.

When reverse translation bhOskun 'to 'to insert' here one word manning is different.

Lexical Divergence: This type of divergences occurs mainly due to the lack of an exact lexical map for a word or development in one language into another. Most of the conflational and inflational divergence overlaps with lexical divergence

Thief broke into bank.

चोर जबरदस्तीन बँकेत शिरले.

We notice that in Marathi, the English phrasal verb 'break into' is realized by a different verb shirla 'enter' which takes an adverbial element jabardastii by 'force'. Here structural mapping between two languages. Here the conflational and inflational divergence overlaps 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.

 Semantic Divergence: Word Sense Disambiguation (WSD) or Lexical Ambiguity Resolution is a basic task, which mechanism to analyze the sense of a word in a given sentence. Words can have more than one meaning and sometimes group of words or full sentence may have more than one meaning in a language resulting in the uncertainty. Words having various meanings are called polysemy. Word disambiguation is another major task in translation where a same word can act differently based on the context

a novel approach to solving a problem.

समस्येचे निराकरण करण्यासाठी नवीन दृष्टीकोन.

Here we notice the 'novel' there are two meaning. Here we use novel means navin(नवीन) but this word another meaning is kadmbri(कादंबरी) . here occurred the semantic divergence.

Structural Divergence: Both the languages have almost same grammatical structures. But there occurs some minor variations in the construction of the sentences

he read the book

त्याने पुस्तक वाचले

Takes an NP argument 'the book' whereas its Marathi counterpart Vachle takes a pp adjuct . we notice that 'read' 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.

Categorical Divergence: Categorical divergence is described by the use of different parts-of-speech in the source language as well as in the target language. It occurs when part-of-speech of a word in the source language changes during the process of translation to generate an entirely different lexical category of words in the target language. In such a case, the mismatch in translation normally occurs due to the change in the parts-of-speech of words of the concerned languages.

I am feeling happy

मला आनंद होत आहे.

In English "happy" is adjective and in Marathi "Aanand" (happy) becomes the noun.

Promotional Divergence and Demotional Divergence: Promotional divergence refers to the promotion of a logical modifier of the source language to the head position in translation in the target language. Due to this reason, it is called Head Swapping Divergence in which a modifier is realized as an adverbial phrase in the source language but as a main verb in the target language. It implies that an adverbial element in the source language is exposed by a verbal element in the target language and opposite case is demotional divergence.
Ex.

The lecture is on.

व्याख्यान चालू आहे.

Here the choice of the lexical entry of 'on' as it can be categorized either as an adverb or an adjective.

IV. Other types of divergence:

Replicative Words: In Marathi language most of replactive words are used but in english language we are not use replactive words. A verb verb repeated such as bagata bagata {see,see} can be used to denote different type of functions that are mapped into English in various ways depending on number of factors.

चित्रपट बगता बगता तो बोर झाला.

Means 'He got bored while watching the movie' this is not correct translation of Marathi sentence. This type of divergence present in Marathi to English translation.

Determiner System: 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.

1.वनामध्ये वाघ आला
{Forest in lion be.PR}
=> There is a lion in the forest.

2.वाघ वनामध्ये आला
{Lion forest in be.PR}
=>The lion is in the forest

We notice that the English interpretation of the Marathi sentence in (1) is not possible for the Marathi sentence in (2). 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.

Morphological Gaps:
In certain cases, it is quite difficult to obtain an exact translation of a common marathi ditransitive verb. For instance, in (16), the English counterpart of the transitive verb piinaa is 'drink'. However, Hindi also has a ditransitive verb pajne derived from pinai. English does not have a counterpart of this ditransitive verb.

आईने बाळाला पानी पाजन्या प्रयत्न चा केला.

{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 marathi and English and can be identified as different types of morphological gaps.

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 basne or bslela in Marathi. However, in this case, the reverse translation also causes divergence. For instance, in (2), 'is sleeping' in English can be mapped both by zopt aahe and zoplela aahe in Marathi. In Marathi, they are distinguished by different lexical form of the relevant verb.

1. राम खुर्ची वर बसलेला आहे.

{Ram chair on sitting be.PR}

=> Ram is sitting on a chair.

2. sham is sleeping.

=> शाम झोपत आहे. or, शाम झोपला

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

V. Conclusion

In this paper, we have identified the problem of translation divergence for MT between English and Marathi. Review on existing work for divergences for different languages national and international language pair. Basis of review we show that the translation divergence between Marathi to English it is varied and complex than the word in existing literature. We need to study the different grammatical and some of the extra-grammatical attribute of both Marathi and English to exhaustively identify the types of translation divergence in this pair of translation languages. This will implement us to come up with approach to handle these situations and coming up with correct translation.

References

[1]. Jisha P Jayan, Elizabeth Sherly," A Study on Divergence in Malayalam and Tamil Language in Machine Translation Perceptive".
[2]. Vimal Mishra, R.B.Mishra," Divergence patterns between English and Sanskrit Machine Translation",2014
[3]. Niladri Sekhar Dash," Linguistic Divergences in English to Bengali Translation", International Journal of English Linguistics, Vol. 3, No. 1; E-ISSN 1923-8703,2013.
[4]. Sreelekha S.," Statistical Vs Rule Based Machine Translation; A Case Study on Indian Language Perspective".
[5]. Sharmin Muzaffar, Pitambar Behera & Girish Nath Jha," Classification and Resolution of Linguistic Divergences in English-Urdu Machine Translation",2016
[6]. Sreelekha. S, Pushpak Bhattacharyya," Lexical Resources for Hindi Marathi MT".
[7]. SREELEKHA S and PUSHPAK BHATTACHARYYA," Role of Morphology Injection in SMT: A Case Study from Indian Language Perspective",2017
[8]. Sreelekha .S, Raj Dabre, Pushpak Bhattacharyya," Comparison of SMT and RBMT, The Requirement of Hybridization for Marathi – Hindi MT",2019
[9]. Dun Deng∗, Nianwen Xue," Translation Divergences in Chinese–English Machine Translation: An Empirical Investigation".
[10]. Pitambar Behera* , Neha Maurya1 & Vandana Pandey," Dealing with Linguistic Divergences in English-Bhojpuri Machine Translation" Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing,2016
[11]. Ritu Nidhi, Tanya Singh," Divergence Identification and Handling for English-Maithili Machine", Pramana Research Journal, Volume 9, Issue 2, ISSN NO: 2249-2976,2019
[12]. Unnikrishnan P, Antony P J, Dr. Soman K P," A Novel Approach for English to South Dravidian Language Statistical Machine Translation System", International Journal on Computer Science and Engineering Vol. 02, No. 08, 2010, 2749-2759,2010
[13]. R. Mahesh K. Sinha and Anil Thakur," Translation Divergence in English-Hindi MT",2005
[14]. R. Mahesh K. Sinha, Anil Thakur," Divergence Patterns in Machine Translation between Hindi and English".
[15]. Deepa Gupta Niladri Chatterjee," Study of Divergence for Example Based English-Hindi Machine Translation"2001
[16]. W.S.N. Dilshani, S. Yashothara, R.T.Uthayasanker, S. Jayasena," Linguistic Divergence of Sinhala and Tamil languages in Machine Translation"IEEE,2018
[17]. Violaine PRINCE, Jacques CHAUCHE," Applying Divergence Principles to French to English Machine Translation", Proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational Science,2009
[18]. Deepa Gupta and Niladri Chatterjee," Identification of Divergence for English to Hindi EBMT",2003
[19]. SHACHI DAVE, JIGNASHU PARIKH and PUSHPAK BHATTACHARYYA," Interlingua-based English–Hindi Machine Translation and Language Divergence", Machine Translation,2003
[20]. Chung-hye Han1, Benoit Lavoie Martha Palmer1, Owen Rambow3, Richard Kittredge2, Tanya Korelsky2, Nari Kim4, and Myunghee Kim," Handling Structural Divergences and Recovering Dropped Arguments in a Korean/English Machine Translation System", Springer-Verlag Berlin Heidelberg ,2000
[21]. Marwan Akeel*, R. B. Mishra," Divergence and Ambiguity Control in an English to Arabic Machine Translation", Journal of Engineering Research and Applications, Issue 6, Vol. 3, Nov-Dec 2013
[22]. S. B. Kulkarni, P. D. Deshmukh, M. M. Kazi, K. V. Kale," Linguistic Divergence Patterns in English to Marathi Translation", International Journal of Computer Applications, Volume 87 – No.4, 2014.
[23]. Bonnie Doff*," SOLVING THEMATIC DIVERGENCES IN MACHINE TRANSLATION".
[24]. Abdus Saboor, Mohammad Abid Khan," Lexical-Semantic Divergence in Urdu-to-English Example Based Machine Translation", 6th International Conference on Emerging Technologies,IEEE,2010
[25]. Pawan Goyal1 and R. Mahesh K. Sinha," Translation Divergence in English-Sanskrit-Hindi language pairs".
[26]. Prof. Goraksh V.Garje, Manisha Marathe, Urmila Adsule," Translation of Simple English Interrogative Sentences to Marathi Sentences" 2014
[27]. Naila Ata, Bushra Jawaid, Amir Kamran," Rule Based English to Urdu Machine Translation

[28].   G V Garje, G K Kharate, Harshad Kulkarni," Transmuter: An Approach to Rule-based English to Marathi Machine Translation", International Journal of Computer Applications, Volume 98 , No-21, 2014.
[29].   K. Parameswari," Development of Telugu-Tamil Transfer-Based Machine Translation system: With Special reference to Divergence Index," Proceedings of the 1st Deep Machine Translation Workshop, 2015.
[30].   Melamed, R. Green, and J. Turian, Precision and Recall of Machine Translation, In Proceedings of the HLTNAACL 2003, Short Papers: 61–63, Edmonton, Canada, 2003
[31].   Dave, Sachi, Parikh, Jignashu, Pushpak Bhattacharya., Interlingua-based English– Hindi Machine Translation and Language Divergence, Machine Translation, Vol. 16, pp. 251–304, Kluwer Academic Publishers, 2001.
[32].   Dorr, B. J., Machine Translation: A View from the Lexicon, MIT Press. Cambridge, MA, 1993