

A Study on Divergence in Malayalam and Tamil Language in Machine Translation Perceptive

Jisha P Jayan

Virtual Resource Centre for
Language Computing
IITM-K, Trivandrum

jisha.jayan@iiitmk.ac.in

Elizabeth Sherly

Virtual Resource Centre for
Language Computing
IITM-K, Trivandrum

sherly@iiitmk.ac.in

Abstract

Machine Translation has made significant achievements for the past decades. However, in many languages, the complexity with its rich inflection and agglutination poses many challenges, that forced for manual translation to make the corpus available. The divergence in lexical, syntactic and semantic in any pair of languages makes machine translation more difficult. And many systems still depend on rules heavily, that deteriorates system performance. In this paper, a study on divergence in Malayalam-Tamil languages is attempted at source language analysis to make translation process easy. In Malayalam-Tamil pair, the divergence is more reported in lexical and structural level, that is been resolved by using bilingual dictionary and transfer grammar. The accuracy is increased to 65 percentage, which is promising.

Keywords- Translational divergence; semantic; syntactic; lexical;

1 Introduction

The problem with divergence in machine translation in a complex topic, which 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. For any MT system, this topic is very crucial as to obtain an accurate translation, it is very much needed to resolve the nature of translational divergence. This divergence can be seen at different levels. Based on the complexity that occur in the specific translation, divergence affects the translation quality. Some translational divergences are universal in the sense that they occur across the languages while certain

others are specific with respect to the language pair (Lavanya et al., 2005; Saboor and Khan, 2010). Hence, the divergence in the translation need to be studied both perspectives that is across the languages and language specific pair (Sinha, 2005). The most problematic area in translation is the lexicon and the role it plays in the act of creating deviations in sense and reference based on the context of its occurrence in texts (Dash, 2013).

Indian languages come under Indo-Aryan or Dravidian scripts. Though there are similarities in scripts, there are many issues and challenges in translation between languages such as lexical divergences, ambiguities, lexical mismatches, re-ordering, syntactic and semantic issues, structural changes etc. Human translators try to choose the correct wording by using knowledge from various sources, and the factors like phonology, orthography, morphology as well as knowledge of the person, and cultural differences influences the translation. Therefore, it is hard to get a translation of one person as same as other translator. MT is a complex and challenging research area because language translation itself is very difficult. While human processes language understanding and translation on many levels, but a machine processes data, with its linguistic form and structure, it is difficult to get the sense. This requires more of cognitive and intelligent systems in NLP, rather than considering MT development only in linguistic point of view. Many works have been performed based on linguistic and lexical level, but MT across the languages is a challenging task for several reasons like, the difference in the structure of source and target languages, ambiguity, multiword units like idioms, phrases and tense generation and many more. In this paper, we have considered two Dravidian languages Malayalam and Tamil and various challenges and issues in semantic and syntactical in both the languages are discussed.

Malayalam and Tamil belong to Dravidian language family. Malayalam and Tamil are closely related to each other in grammar with a rich literary tradition. However, Malayalam is highly influenced by Sanskrit language at lexical, grammatical and phonemic levels where as Tamil is not. The Noun morphology is same in both the languages as the word may contain the root alone or root with suffixes attached to it. Agglutination is widely seen in Tamil and Malayalam. In both languages, the case markers are found to be attached to the nouns and pronouns. Post-positions are also seen to be attached to these. Morphology includes inflection, sandhi, and derivation. The Tamil verbs inflect for person, number and gender whereas Malayalam verbs do not. Hence the gender marking of the noun is not a relevant feature when Malayalam is considered.

Language divergence in most cases result in the ambiguities in translation. The divergence issue across a language is associated with many factors ranging from linguistic, cultural, and societal to psychological aspects of the languages. Syntactic and lexico-semantic divergence is the two board categories of divergence proposed by Dorr. Sentence level ambiguities are referred as syntactic while at the word level is semantic. A hybrid approach to develop the Malayalam to Tamil MT system comprising paradigm, rule and machine learning methods are proposed. The system deals with the analysis, transfer, and generation process. The issues being raised in the various stages in the development of the Machine Translation are discussed here. The next section deals with the related works carried out in this area. In section 3, various translational divergences are considered with respect to Malayalam and Tamil languages. Some other types of divergences found while translation are discussed in next section. Fifth sections give the methods for handling the divergence and finally the paper is concluded in section 6.

2 State of Art

Dorr (1994) gives a systematic solution to the problem of divergence derived from the formalization of two different information namely the linguistic ground on which the lexical and semantic divergence are based and the technique to solve these problems. The paper explains mainly seven types of divergence with examples with re-

spect to English, Spanish and German. The paper focuses on Thematic, promotional, demotional, structural, conflational, categorical and lexical divergences. Barnett et al. (1994) divide distinctions between source and target languages into two categories mainly translation divergences and translation mismatches. The information conveyed in source and target language remain same while the structure of the sentence differ in translational divergence (1990). In definition of lexical-semantic representation and translation mappings is described. The paper discussed on the justification for distinguishing promotional and demotional divergence, the limits imposed on the range of repositioning possibilities, notion of full coverage in context of lexical selection and resolution of interacting divergence types. The paper concludes with a brief description of UNITRAN, a system for translation across a variety of languages, which accommodates the divergence types.

Nizar and Dorr (2002) proposed a novel approach to handle divergence in translation in a Generation-Heavy Hybrid Machine Translation (GHMT) system. Deep symmetric knowledge of source and target language is required for these approach. Various examples are illustrated to show the interaction between statistical and symbolic knowledge in GHMT system.

Dorr (1990) presented a mechanisms for mapping an underlying lexical-conceptual structure to a syntactic structure used by the UNITRAN. Also explains the ways to solve the problem of thematic divergences in machine translation. The solution is implemented in the bidirectional system for English, Spanish, and German. The two types of thematic divergence namely the reordering of arguments for a given predicate and reordering of predicates with respect to arguments or modifiers is explained. They presented three mechanisms to solve the thematic divergences with a set of general linking routines

Zhiwei (2006) describes different types of translation divergence in Machine Translation. Even though, translation divergence occurs at all phases of MT, the author concentrated on the translation divergence in the transfer phase. The translational divergence that are found in lexical selection in target language, in tense in thematic relation, in head-switch, in structure, in category, in conflation is described. The ambiguity with respect to syntactical, semantic and contextual that relate with

the co-occurrence based approaches for the selection of translation equivalence. The author also suggests the use of feature vector to represent the co-occurrence cluster. The paper proposes some suggestions in Mt system.

Akeel and Mishra (2013) discussed about the language divergences and the ambiguities present in English to Arabic machine translation and the control methods used to resolve the same. The authors have implemented English to Arabic machine translation system using ANN and rule based approach. The proposed system is capable of handling conventional type of ambiguities like lexical and syntactic ambiguity and also structural divergences. The lexical ambiguities include category, homograph, transfer or translational, pronoun reference, gender and number. The word-order, agreement, tense and aspect are various structural divergence explained in the paper.

Dave et al. (2001) have studied the language divergence between English and Hindi. This paper also studied the implication of language divergence in machine translation using Universal Networking Language (UNL), introduced by United Nations University, to facilitate the transfer and exchange of information over the internet. The language divergence between these languages is considered as divergence between SOV and SVO classes of languages. Two criteria are considered for deciding the effectiveness of an interlingua. The first criteria is that the meaning conveyed by source text should be apparent from the interlingual representation. The next is a generator should be able to produce a target language sentence that a native speaker of that language accepts as natural. The use of lexical resources in constructing a semantically rich dictionary semi-automatically with an overview of major differences between English and Hindi is described. The syntactic and lexical-semantic divergences between Hindi and English from a computational linguistic perspective is explained.

Kulkarni et al. (2014) studied various divergence pattern that occurred between English and Marathi language pair translation. Their study involved the divergences based on lexico-semantic and syntactic. They further classified lexico-semantic divergence into thematic divergence, structural divergence, promotional and demotional divergence, conflation and inflation divergence, categorical divergence and lexical diver-

gence. Syntactic divergence include constituent-order divergence, adjunction divergence, null-subject divergence and pleonastic divergences. They also focused on divergence that occurred in English and Marathi machine translation that are common. These include divergence found in replicative words, morphological gaps, determiner systems, honorific differences, word related divergences. All the translation divergences are explained with relevant examples.

Jayan et al. (2012) conducted a study to know the pattern of the source and target languages and the morphophonemic changes occurring during the translation of a sentence. The paper focused on the different types of divergence that occur among this language pair. The non-configurational nature of Malayalam is being explained with various examples. Author discusses the complexity of translational divergence among their language pair and provides a solution to classify and resolve the divergence problem. The paper gives a brief idea about the English Malayalam MAT system based on the Anglabharathi, which is an interlingua based approach.

Goyal and Sinha (2009) discusses the translation pattern between English-Sanskrit and Hindi - Sanskrit of various constructions to identify the divergence in these language pairs. Through this the authors come up with strategies to handle these situations and also with correct translations. The base of their classification of translation divergence is presented by Dorr (1994).

Sinha and Thakur (2005) studied different patterns of translation divergences both from Hindi to English and English to Hindi keeping in view the classification of translation divergence proposed by Dorr. They have observed that there are a number of areas in Hindi-English translation pair that fall under translation divergence but cannot be accounted for within the existing parameters of classification strategy.

Mishra and Mishra (2009) proposed a method to detect and implement the adaptation rules for the divergence in English to Sanskrit machine translation. The divergences in language between English and Sanskrit can be considered as representing the divergences between SVO (Subject - Verb - Object) and SOV (Subject - Object - Verb) classes of languages. The type of divergence detected is based on different aspects like linguistic to socio- and psycho-linguistic, role of conjunctions and

particles, participle, gerunds and socio-cultural aspects. They have performed a novel method that uses rules and ANN technique to detect and implement the adaptation rules for the divergence in English to Sanskrit machine translation.

Gupta and Chatterjee (2003) presented adaptation for English-Hindi EBMT. They have discussed the issue of adaptation, in general, with special emphasis to divergence. Their work looks at adaptation of EBMT between English and Hindi. They have given special attention to the study of divergence by recognizing six different categories of divergence and providing schemes for identifying them

3 Translational Divergence

The divergence study is mainly dealt with languages. Divergence is a language dependent phenomenon and is not necessary that same set of divergences will occur across all the languages. Lexico-semantic translation divergences are accounted for by means of parameterization of the lexicon (Dorr, 1993). The reason behind divergence mainly lies with the incompatibility of a language with another, a common phenomenon for almost all natural languages. There are situations where a sentence in the source language needs to be translated in the target language in entirely different forms, making the task of both translations a difficult process.

3.1 Conflational and Inflational Divergence

Conflational divergence occurs when the sense conveyed by a single word is expressed by two or more words in one of the languages. The opposite case of conflational divergence is referred to by inflational divergence. The problem due to the lexical gap, an instance where there is an absence of a word to express a specific concept of the source word in the target language during the translation process. Not all the words in one language have equivalent words in another language. In some cases a word in one language is to be expressed by a group of words or phrases in another. For examples:

വല (vala) : കിണற்றെറുമുറ്റം_മரச்சட்டം (kiNaR-Rai_muTuTum_maraccaTTam)

പേട (peeta) : പാലിൽ_ചെയ്യப்படും_ഇണിப்பு (paalil_ceyyappaTum_inippu)

കാൽവെള്ള (kaal~veLLa): പാതத்தിൻ_அடிப்பகுதி (paattatin_aTippakuti)

3.2 Lexical Divergence

This type of divergences arises mainly due to the lack of an exact lexical map for a word or construction in one language into another. Most of the conflational and inflational divergence overlap with lexical divergence.

3.3 Syntactic Ambiguity

A word may belong to more than one POS category, resulting in the syntactic ambiguity. This is a problem with the parsing resulting in the source language. There is a lack of one-to-one correspondence of parts of speech between two languages. For examples: can be INTF or PRP, can be NN or PSP, can be PSP or VM etc.

അവൻ/PRP അത്രയും/QF ദൂരം/NN ബസിൽ/NN നിന്ന~/VM പോയി/VM ./SYM

(avan~/PRP atrayuM/QF duuraM/NN basil~/NN ninn~/VM pooyi/VM ./SYM)

അവൻ/PRP അവിടെ/NST നിന്ന~/PSP ഇപ്പോൾ/RB ഇറങ്ങി/VM ./SYM

(avan~/PRP aviTe/NST ninn~/PSP ippool~/RB iRangngi/VM ./SYM)

In the first sentence, word “നിന്ന” refers to “stood” while in the second sentence the same word means “from”. Another example

3.4 Null Subject Divergence

In Malayalam subject dropping can be seen frequently. But Tamil has restrictions in subject dropping. If the verb occurring in Tamil is an action at the end of a sentence, then subject dropping is not possible. But if it occurs in a statement, null subject construction is possible.

3.5 Agreement Divergence

Given two languages may have completely different structures. Malayalam is free-word order language while Tamil has SOV structure, so to identify the phrase performing the function of the subject in the sentence is a challenging process. Gender issues in translation are a subject that serves to point out the limitations of machine translation. The agreement between the subject and verb is one of the problematic issues in MT. The Tamil verbs depend on the PNG (Person Number Gender) information about nouns. Malayalam verbs do not depend on this information. Hence the gender marking of the noun is not a relevant feature when Malayalam is considered.

| | Past | Present | Future |
|------------------------|--------------------------------------|--|--------------------------------------|
| 1st singular | வந்தேன் (<i>vandteen</i>) | வருகிறேன் (<i>varukiReen</i>) | வருவேன் (<i>varuveen</i>) |
| 1st plural | வந்தோம் (<i>vandtoom</i>) | வருகிறோம் (<i>varukiRoom</i>) | வருவோம் (<i>varuvoom</i>) |
| 2nd singular | வந்தாய் (<i>vandtaay</i>) | வருகிறாய் (<i>varukiRaay</i>) | வருவாய் (<i>varuvaay</i>) |
| 2nd singular honorific | வந்தீர் (<i>vandtiir</i>) | வருகிறீர் (<i>varukiRiir</i>) | வருவீர் (<i>varuviir</i>) |
| 2nd plural | வந்தீர்கள் (<i>vandtiirkaL</i>) | வருகிறீர்கள் (<i>varukiRiirkaL</i>) | வருவீர்கள் (<i>varuviirkaL</i>) |
| 3rd singular male | வந்தான் (<i>vandtaan</i>) | வருகிறான் (<i>varukiRaan</i>) | வருவான் (<i>varuvaan</i>) |
| 3rd singular female | வந்தாள் (<i>vandtaaL</i>) | வருகிறாள் (<i>varukiRaaL</i>) | வருவாள் (<i>varuvaaL</i>) |
| 3rd singular honorific | வந்தார் (<i>vandtaar</i>) | வருகிறார் (<i>varukiRaar</i>) | வருவார் (<i>varuvaar</i>) |
| 3rd plural | வந்தார்கள் (<i>vandtaarkaL</i>) | வருகிறார்கள் (<i>varukiRaarkaL</i>) | வருவார்கள் (<i>varuvaarkaL</i>) |
| 3rd singular neutral | வந்தது (<i>vandtatu</i>) | வருகிறது (<i>varukiRawu</i>) | வரும் (<i>varum</i>) |
| 3rd plural neutral | வந்தன (<i>vandtana</i>) | வருகின்றன (<i>varukinRana</i>) | வரும் (<i>varum</i>) |

Table 1: Finite forms of verb vaa (வா)

From the above Table 1, for a single word “வருக” we have present tense “வருகുന്നു”, past tense “വന്നു” and future tense “വരും” in Malayalam, there have eleven forms for each tense in Tamil depending the agreement of a verb with the subject.

3.6 Semantic Divergence

Word Sense Disambiguation (WSD) or Lexical Ambiguity Resolution is a fundamental task, which processes to identify the sense of a word in a given sentence. Words can have more than one meaning and sometimes group of words or whole sentence may have more than one meaning in a language resulting in the ambiguities. Words having multiple meanings are called polysemy. Word disambiguation is another major challenge in translation where a same word can act differently based on the context. For examples:

രസം (*rasaM*) - അനുഭൂതി, കരി, ഇഷ്ടം, രുചി, അഭിരുചി, മെരിക്കുറി

(*anubhuuti, kaRi, ishTaM, ruci, abhiruci, mer~kkuRi*)

ലളിതമായ ചേരുവകൾ ചേർത്തുള്ള കുരുമുളക് രസം കഴിക്കുന്നത് ദഹനത്തെ സഹായിക്കും.

(*laLitamaaya ceeruvakaL~ceerituLLa kurumuLak*

rasaM kazhikkunnat dahanatte sahaayikkum.)
അവന് അടുത്തിടെയായി എന്നോട് അത്ര രസം ഇല്ല.
(*avan~aTuttuTeyaayi ennooT atra rasaM illa.*)

The word in the above two sentences clearly distinguishes the difference in meanings respectively “curry” and “realization”. Other examples:

നട (*naTa*) - ക്രിയ, നടക്കുക, പടി
(*kRIya, naTakkuka, paTi*)

അടി (*aTi*) - ചുവടുവെട്ട്, പാദം, തല്ല്
(*cuvaTaLav, paadaM, tall*)

വാണം (*vaanaM*) - മാണം, അഭിമാണം
(*maanaM, abhimaanaM*)

ഉത്തരം (*uttaraM*) - മറുപടി, ചോദ്യോത്തരം, താങ്ങ
(*maRupaTi, coodyoottaraM, taangng*)

From the above examples here the word “നട” can act as Noun as well as Verb depending on the context. This is applicable for both the languages. Thus the differences are either because of the absence of a concept or due to the differences at the level of identifying concepts. The case where there is an absence of a concept is comparatively easier to handle than the cases where the domains overlap with other domains.

Unless the correct meaning is captured, there are more chances for incorrect translation.

3.7 Structural Divergence

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

അവൻ ഇനലെ സിനിമയ്ക്ക് പോയി. ഞാനും.
(*avan~innale sinimaykk pooyi. njaanum.*)

அவன் நேற்று திரைப்படத்திற்கும் போனான். நானும் போனேன்.

(*avan ndeeRR tiraippTattindkum poonaan. ndaanum pooneen.*)

In the above example, in second sentence, its not necessary to specify “പോയി” *poyi* with “ഞാനും” *njaanum* as its understood from the sentence construction. But in Tamil, this construction is not seen.

4 Other issues in Translation

4.1 Idiom Translation

An idiom can be defined as a phrase or sentence whose meaning is not specific from meaning of its individual units and that must be learnt as a whole unit. These are used in written as well as verbal communication, giving a stylistic way of delivering a message. While translating idioms from Malayalam to Tamil, some phrases can only be translated with correct sense, but some others can only be transliterated only.

മിനുനതെല്ലാം പൊന്നല്ല.
(*minnunnatellaaM ponnalla.*)

மின் நுவதெல்லாம் பொன்னல்ல
(*minnuvatellaam ponnalla.*)

In the above example, the correct translation is possible but in below example literal translation of the phrase gives an incorrect translation.

കൈയാല പുറത്തെ തേങ്ങ പോലെ.
(*kaiyaala puRatte teengga poole.*)

மதில் மேல் பூநை பொல
(*matil meel puunai poola*)

4.2 Divergence in Copula Transfer

A copula is a verb or verb like word, capable of functioning as a main verb in a sentence. These are different from action verbs grammatically and semantically. In Malayalam, “ആക്” (*aak*) and “ഉണ്ട്” (*uNT*) functions as copula while Tamil lacks copula.

രാമൻ സീതയെ ഇഷ്ടം ആണ്.

(*raaman siitaye ishTaM aaN.*)

இராமனுக்கு சீதையை பிடிக்கும்
(*iraamavukku ciitaiyai pitikkum*)

While translating from Malayalam to Tamil, copula is not needed as the translation without copula gives a meaningful sentence in Tamil .

4.3 Stylistic Variations

All languages have its own style. There are sentences in Malayalam for which direct translation in Tamil is not possible. There occur such sentences where many to one mappings can be seen from Malayalam to Tamil. The example below depicts this nature:

നീ എവിടെ പോകുന്നു.

(*nii eviTe pookunnu*)

നീ എങ്ങോട്ട് പോകുന്നു.

(*nii enggooTT pookunnu*)

நீ எந்த்தை போகிராய்

(*nii efkai pookiraay*)

നീ എവിടെയാണ് പോകുന്നത്.

(*nii eviTeyaaN pookunnu*)

நீ எந்த்தை போகிராய்

(*nii efkai pookiraay*)

നീ എങ്ങോട്ടാണ് പോകുന്നത്.

(*nii enggooTTaaN pookunnu*)

நீ எந்த்தை போகிராய்

(*nii efkai pookiraay*)

For the all four sentences shown above in Malayalam, there is only one equivalent translation in Tamil. In Malayalam, there is usages like “കാണുന്നുണ്ട്” (*kaaNunnuNT*) , “പോകുന്നുണ്ട്” (*pookunnuNT*). Tamil doesn’t have these structure, which makes error in translation.

ഞാൻ നാളെ അവനെ കാണുന്നുണ്ട്.

(*njaan~naaLe avane kaaNunnuNT.*)

நான் நாளை அவனை பார்ப்பேன்

(*ndaan ndaaLai avanee paarpppeen.*)

நான் நாளை அவனை பார்க்கவிருக்கிறேன்

(*ndaan ndaaLai avanai paarkkavirukkiReen.*)

Other difference mostly found is the dropping of a word from a sentence in Tamil which is a must in Malayalam. In the examples given below, “ഒന്ന്” (*onn*) from first sentence and “ഒരു” (*oru*) from second sentence is dropped. The following examples demonstrate this typical characteristics.

ഞാൻ ഒന്ന് പറഞ്ഞോട്ടെ.

(*njaan~onn paRanjnjooTTe.*)

நான் சொல்லட்டுமா

(*ndaan collaTTumaa.*)

അവനെ ഒരു നോക്ക് കാണാൻ കൊതിയാകുന്നു.

(*avane oru nookk kaaNaan kotiyaakunnu.*)

அவனை பார்க்க ஆசைபடுகிறேன் .

(*avanai paarkka aacaipaTukikeen.*)

5 Handling Divergences

The divergence found in the MT system cannot be handled completely by the system itself as sometimes there may not be word meanings that can be replaced exactly. Various modules are built for bridging the divergence encountered while translating the text from Malayalam to Tamil. The linguistics modules are very crucial for any machine translation. A combination of linguistic as well as statistical machine learning techniques are used to build translation capability. Various modules included in the MT system are Malayalam morphological analyser which is a paradigm based, parts of speech tagger and chunker statistical based, lexical transfer comprising the bilingual dictionary, morphological generator for Tamil with some additional minute modules. The architecture of the system is based on the analyse-transfer-generate.

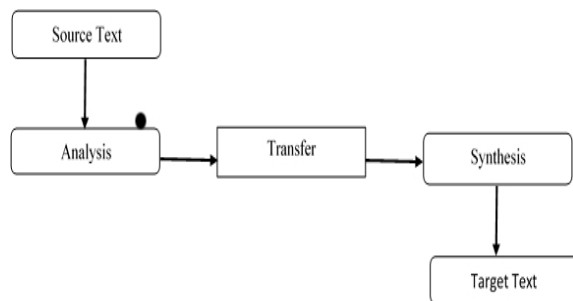


Figure 1: System Architecture

In the present study, the divergence issues are solved at the analysis part only ie at the source side. The analysis of the source language side deals with different modules. At each level various grammatical rules are incorporated to get a much better result. The Figure 2 depicts the increase in the accuracy when certain rules are applied.

The test was carried out for 100 sentences from various domains, When rules are not given the accuracy obtained ranges from 47-55%. At the same time when certain rules were included in different modules for rectifying the errors , the

accuracy increased to 53-65%. The divergences at the lexical are solved using bilingual dictionary. The structural transformations are carried out at the Transfer grammar module.

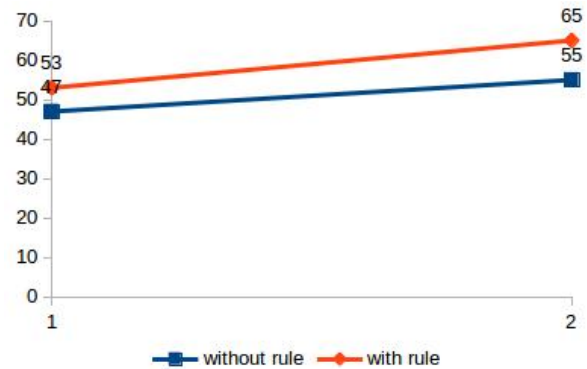


Figure 2: Accuracy Graph

The structural transformations are carried out at the Transfer grammar module. The Lexical divergences are solved using bilingual dictionary. Many rules are incorporated at various modules for rectifying the errors occurring during the translation process.

6 Conclusion

Identifying the patterns of divergence in any languages is very crucial to obtain a quality machine translation system. The translational divergence requires combinations of manipulations both at the lexical and structural level. Some important types of divergence related to translation based on Dorr's classification from Malayalam to Tamil are discussed in this paper. Both manual and machine translation among this language pair is taken in view. The divergence found can be semantic or syntactic types. Various patterns of translation divergence found at different phases of machine translation is examined.

References

- Abdus Saboor and Mohammad Abid Khan. 2010. *Lexical-semantic divergence in Urdu-to-English Example Based Machine Translation*, In Emerging Technologies (ICET), 6th International Conference on IEEE. 316-320
- Bonnie J. Dorr. 1994. *Machine translation divergences: A formal description and proposed solution*, *Computational Linguistics* , 20(4):597-633.

- Bonnie J. Dorr. 1990. *A cross-linguistic approach to machine translation*, In *Proceedings, Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Linguistics Research Center, The University of Texas, Austin, Texas, 13-32.
- Bonnie Dorr. 1993. *Machine Translation: A View from the Lexicon*, MIT press.
- Bonnie Dorr. 1990. *Solving thematic divergences in machine translation*, In the *Proceedings of the 28th annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics. 127-134.
- Deepa Gupta and Niladri Chatterjee. 2003. *Identification of divergence for English to Hindi EBMT*, In *Proceeding of MT Summit-IX*, 141-148.
- Feng Zhiwei,. 2006. *Translational divergence in Machine Translation*, Eafterm Symposium, Haikou.
- James Barnett, Inderjeet Mani and Elaine Rich . 1994. *Reversible machine translation: What to do when the languages dont match up*, *Reversible Grammar in Natural Language Processing*, Springer US. 321-364
- Jayan V, Sunil R, Sulochana Kurambath G and R Ravindra Kumar . 2012. *Divergence patterns in Machine Translation between Malayalam and English*, In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*. ACM. 788-794.
- Marwan Akeel and R.B. Mishra. 2013. *Divergence and Ambiguity Control in an English to Arabic Machine Translation*, In the *International Journal of Engineering Research and Applications*, 3(6):1670-1679.
- Niladri Sekhar Dash. 2013. *Linguistic Divergences in English to Bengali Translation*, *INFOCOMP International Journal of English Linguistics*, 3(1):31-40
- Nizar Habash, Bonnie Dorr . 2002. *Handling translation divergences: Combining statistical and symbolic techniques in generation-heavy machine translation*, Springer Berlin Heidelberg.
- Pawan Goyal and R. Mahesh K. Sinha . 2009. *Translation divergence in English-Sanskrit-Hindi language pairs*, In *Sanskrit Computational Linguistics*, Springer Berlin Heidelberg. 134-143.
- Prahallad Lavanya, Prahallad Kishore, and Ganapa Thiraju Madhavi . 2005. *A simple approach for building transliteration editors for indian languages*, *Journal of Zhejiang University Science A*, 6(11):1354-1361.
- R. Mahesh K. Sinha and Anil Thakur . 2005. *Translation Divergence in English-Hindi MT*, In the *Proceeding of EAMT Xth Annual Conference*, Budapest, Hungary. 30-31.
- Shachi Dave, Jignashu Parikh and Pushpak Bhat-tacharyya . 2001. *Interlingua-based English-Hindi Machine Translation and Language Divergence*, *Machine Translation*, 16(4):251-304.
- S. B Kulkarni, P.D. Deshmukh, M. M. Kazi and K. V. Kale . 2014. *Linguistic Divergence Patterns in English to Marathi Translation*, *International Journal of Computer Applications(0975 8887)*, 87(4):21-26.
- R. Mahesh K. Sinha and Anil Thakur. 2005. *Translation Divergence in English-Hindi MT*, In the *Proceeding of EAMT Xth Annual Conference*, Budapest, Hungary. 245-254.
- Vimal Mishra and R.B. Mishra. 2009. *Divergence patterns between English and Sanskrit machine translation*, *INFOCOMP Journal of Computer Science*, 8(3):62-71.