Morphology Injection for English-Malayalam Statistical Machine Translation

Sreelekha S, Pushpak Bhattacharyya
IIT Bombay, Mumbai, India
{sreelekha, pb} @cse.iitb.ac.in

Abstract

Statistical Machine Translation (SMT) approaches fails to handle the rich morphology when translating into morphologically rich language. This is due to the data sparsity, which is the missing of the morphologically inflected forms of words from the parallel corpus. We investigated a method to generate these unseen morphological forms. In this paper, we analyze the morphological complexity of a morphologically rich Indian language Malayalam when translating from English. Being a highly agglutinative language, it is very difficult to generate the various morphological inflected forms for Malayalam. We study both the factor based models and the phrase based models and the problem of data sparseness. We propose a simple and effective solution based on enriching the parallel corpus with generated morphological forms. We verify this approach with various experiments on English-Malayalam SMT. We observes that the morphology injection method improves the quality of the translation. We have analyzed the experimental results both in terms of automatic and subjective evaluations.

Keywords: Morphology Injection, Statistical Machine Translation, English-Malayalam Machine Translation

1. Introduction

Malayalam, Telugu, Kannada and Tamil are being most prominent out of 17 languages in the southern Indian family of Dravidian languages; about 95 per cent of the South Indian population speaks one of these four languages. Out of these over 38 million people is speaking Malayalam primarily in the state of Kerala. Throughout its gradual evolution of the present day Malayalam, the influence of Sanskrit is evident in the alphabet, phonology and vocabulary and to a lesser extent in the morphology of Malayalam. Malayalam is a highly agglutinative and inflectionally rich language with a free word order. The semantic and syntactic relations between the verbs and other constituents in a sentence are represented by the case endings of the words. V*aachakam*, which denotes the matter, action and quality, and *dyootakam,* which denotes the relationships, are the two types of "*Sabdam*", a combination of sounds with a meaning. *Naamam* (noun), *kriyaa* (verb) and *bheedakam* (modifier) are the three types of *Vaachakam*. *gati* (preposition), *ghatakam* (conjunction) and *vyaakseepakam* (interjection) are the three types of *dyootakam* (Varma, 2000). Malayalam has a strong postpositional inflections with highly agglutinative suffixes (Namboodiri, 1998). These inflections carry information about tense, mood and aspect for verbs and cases (accusative, dative, etc.), gender, number, person information for nouns.

Most approaches to Statistical Machine Translation, i.e., phrase based models (Koehn, Och and Marcu, 2003), syntax based models (Yamada and Knight 2001) do not allow incorporation of any linguistic information in the translation process. The introduction of factored models (Koehn and Hoang, 2007) provided this missing linguistic touch to the statistical machine translation.

Factored models (Koehn and Hoang, 2007) treat each word in the corpus as vector of tokens. Each token can be any linguistic information about the word which leads to its inflection on the target side. Hence, factored models are preferred over phrase based models (Koehn, Och and Marcu, 2003) while translating from morphologically poor language to morphologically richer language. There were many attempts to improve the quality of SMT systems such as; using Monolingually-Derived Paraphrases(Marton et al., 2009), Using Related Resource-Rich languages (Nakov and Ng, 2012), (Minkov et. al., 2007) . In this paper, we study SMT models and the problem of sparseness and morphological complexity in the context of translation to a highly agglutinative, morphologically rich language Malayalam from English. There are many ongoing attempts to develop MT systems for Indian languages (Antony, 2013; Bharathi et. al., 1996; Kunchukuttan et al., 2014; Nair et. al., 2012; Sreelekha et al., 2013; Sreelekha et al., 2015; Sreelekha et al., 2015; Sreelekha et al., 2016; Sreelekha et al., 2018) using both rule based and statistical approaches. Even though there were many attempts to develop Machine Translation systems between English and Malayalam, the complexity of morphology, especially the word compounding phenomena and the various derivation morphology forms makes the translation quality worse. In this paper we propose a simple and effective solution to handle the morphological complexity which is based on enriching the input with various morphological forms of words. The flow of the paper is as follows: Section 2 describes Morphological Phenomena in Malayalam; Section 3 describes Morphology Generation technique; Section 4 describes the experimental discussion and section 5 describes the conclusion.

2. Morphological Phenomena in Malayalam

Malayalam poses many morphological variations due to the Inflections, Derivations and Word compounding features. Noun or verb is attached to suffixes to generate words of the same category in inflectional morphology. On the other hand noun or verb with a suffix attached to it generates a word of new category in Derivational morphology. New words are formed by combining a noun and a noun, noun and adjective, verb and noun, adverb and verb, adjective and noun and in some cases all the words of an entire sentence to reflect the semantics of the sentence in word compounding (Jurafsky 2003).

2.1 Nouns, Pronouns and Verbs Characteristics

GNP, the Gender (Masculine, feminine and neuter), Number (singular or plural) and Person (person, second person or third person) information is used for the processing of nouns and pronouns. For proper nouns and abstract nouns there is no plural. Nominative, accusative, sociative, dative, instrumental, genitive and locative are the seven cases in Malayalam. Inflections of verbs can be generated by considering tense, mood and aspect. There are two main groups of verb classification: *kaaritam*, simple verbbases with 'ക്കു' *(kku)* (eg.ചിരിക്കുക *(chirikkuka/ to laugh)*; *akaaritam*, simple verb-bases which do not contain 'ക്കു'*(kku) (e.g. ചാടുക (chaaduka/ to jump))*. Past finite, present finite, future finite, negative past, negative present, negative future, or infinitive are the Tense forms. There are direct imperatives (singular and plural), indirect imperatives and negative imperatives. Participles can be verb participle (positive and negative), conditional participle (positive and negative), concessive participle (positive and negative), relative participle (past, present, future, and negative). Aspect can be habitual, trial, completive, durative, reflexive, or perfective. Mood can be expressed as possibility (positive and negative), obligatory (positive and negative), inceptive, ability (past, present, future), or causative (past, present, future). Past tense markers in Malayalam is based on the verb base (*kaaritam* or *akaaritam*) and on the phoneme quality of the last character in the root such as palatal, labial, *vyanjana* or *chil* (Varma, 2000).

2.2 Derivational Morphology for nouns

Adjectives, adverbs and verbs can be derived from nouns by adding proper suffixes. Modifiers are qualifiers and of three types: *naamavisheshanam* (adjective), when it modifies a noun; *kriyaavisheshanam* (adverb), when it modifies an adverb; *bhedakavisheshana* (modifier of modifier), when it modifies a modifier. The modifiers can be pure modifiers (such as, determinative adjectives, superlative adjectives, interrogative adjectives, temporal adverbs, special adverbs and adverbs of manner) or those derived from nouns and verbs. The Table 1 shows the inflectional and derivational morphology for nouns commonly found in Malayalam.

2.3 Derivational morphology for Verbs

A new category of word is generated by attaching noun or verb with a suffix. The derivations considered are:

i) participles (verbal participle, conditional participle, concessive participle and relative participle.

For example, കഴിഞ്ഞ *(kazhinja* / over-relative participle), കഴിഞ്ഞാൽ (*kazhinjaal* / if over-conditional participle).)

ii) infinitives (The suffix taken by infinitives is "*aan*". For example, *വര് (vaRz/ come) + ആൻ (aan / to) = വരാൻ (vaRaan / to come)).*

Table 2 shows the inflection generation forms for a verb "varuka".

		type	Suffix replacement	Rule(end chatr)	Example	
					root	Inflected form
Inflections	Plural		/kal	i	/kutti	/kuttikal
	Case	Nominative	No suffix	-	/krishnan	/krishnan
		Accusative	/e	/n	/krishnan	/krishnane,
			/ine	/u	/indu	/induvine
		Dative	/odu	/i	/kavi	/kaviyodu
			/inodu	/u	/tanu	/tanuvinodu
		Sociative	/ikku	/i	/rathi	/rathikku
			/inu	/u	/indu	/induvinu
		Instrumental	/aal	/i	/tadi	/tadiyaal
			/inal	/u	/indu	/induvinal
		Genitive	/inte	/a	/kavita	/kavitayude
			/ude	/n	/mohan	/mohante
Derivation	adjective	Quality	/aaya	/n	/nallavan	/nallavanaya
			/ulla	/i	/bhangi	/bhangiyulla
		Place	/ile	/m	/maram	/marathile
	Adverb	Manner	/aayi	/i	/bhangi	/bhangiyayi
		Direction	/ekkz	/u	/kadz	/kattilekkz

Table 1 : Inflection generation rules for Noun

2.4 The complexity in Word compounding

Type		Imperative forms	Suffix	Example Inflection
Tense		Past	/um	/varum
		Present	/unnu	/vannu
Aspect	Perfect	Present	/ittundu	/vannittundu
			/irikkunnu	/vannirikkunnu
		Past	/ikkazhinjirikkunnu	/vannethikazhinjirikkunnu
			/itittundayirunnu	/vannittundayirunnu
		Future	/irikkum	/vannirikkum
			/ittundakum	/vannittundakum
	continuous	Present	/kontirunnu	/vannukontirunnu
		Past	/iriikukkayayirunnu	/vannukontirikkukayayirunnu
	Imperfective		/unnundu	/varunnundu
			/ukayanau	/varukayanu
	Ingressive		/aarayi	/vararayi
	Other Auxiliaries		/poyi	/vannu poyi
			/kalanju	/vannukalanju
Mood	optative		/atte	/varatte
	Intentional		/aam	/varaam
	Debititive		/anam	/varanam
			/Etheeru	/vannetheeru
	Debititive (-ve)		/anta	/varanta
			/ikkooda	/varakkoda
			/aanpadilla	/varanpadilla
	Ability		/aam +dative subject	/varaam
			/aankazhinju	/varankazhinju
			/aansadhichu	/varan sadhichu
			/aanothu	/varanothu
	Permission (+ve)		/am	/varam
			/ate	/varatte
			/oloo	/vannoloo
	Permission (-ve)		/aruthu	/vararuthu
	Degree of certainity		/Ekkaam	vannekkam
			/umayirikkam	varumayirikkam
	Authority for assertion		/athre	/varumathre
			/ennukettu	/varumennukettu

Table 2: Inflection generation forms for a verb "varuka"

Word compounding (*sandhi*) is called the sound changes when two words or suffixes join and it is common in Malayalam. Consider an example to understand it's complexity, were five words are joined as a single word;

Malayalam: *ഞാനിന്നൊരാളെപ്പറ്റിച്ചു*

{njaaninnoraleppattichu}

English: *I fooled one person today.*

Panini has classified the word compounding according to the position in which the compounding occurs, such as:

- word_medial (padamadhyam) occurs between a stem and a suffix;
- word final (padaanta) occurs between two words;
- hybrid (ubhaya), both word medial and word final involves.

Malayalam compounding rules are also classified as:

1) Vowel sandhi*: വാഴ (vazha/ rain)+അല്ല (alla/not)= വാഴയല്ല (vazhayalla)*

2) Vowel- consonant sandhi: *താമര (thaamara/lotus)+ കുളം (kuLam /pond)= താമരക്കുളം (thaamarakkulam / lotus pond)*

3) Consonant -consonant sandhi:- *വേനൽ (venal*/summer) + *അവധി* (*avadhi*/leave)= *വേനലവധി* (*venalavadhi / summer leave).*

Word 1st end	Word-2nd beg	Substitution	Example
/am	Vowel(v2)	(1,1,/ma+ss(v2))	varam+alla = varamalla
/am	Vowel(v2),word class=casemarker	(1,1,/tha+ss(v2))	varam+e= varathe
/N	/da, /tha, /na	(1, 1, /nta, NNa resp.)	TaN+taar=tanN Taar
/am	/ka,/cha,/da, /tha, /pa	/nka, /ncha, /nda, /nta, /mpa, /nga	varam+kal = varangal
/N, n, L, I, r	Vowel(v2)	(-, 1, /Na, /na, /La, /la, /ra	Aval + il = avalil

Table 3: Substitution (Aadesa Sandhi)

Word-1stend	Word-2ndbeg	Substitution	Example
Vowel	Vowel	(-,1,ya+ss(v2))	tara+odu = tarayodu
/U, /uu	Vowel	(-,1,/va+ss(v2))	rhitu+aayi=rhituvayi
Word-1=/a, /I, /e	Vowel	(-, 1, va, ss(v2)	e+ ir = ivar
/a	consonant	(-, -, ss(aa))	kala+mELa = kalaamELa

Table 4: Addition Rules (Aagama Sandhi)

Word-1st -end	Word-2nd begining	Substitution	Example
/z	Vowel	(1,1,ss(v2))	veedz+il = veettil
Word-1= /alla, /illa	Vowel	(-,1,ss(v2))	alla+ennz=allennz
Word-1=/oru	Vowel	(1,1, ss(v2)	oru+aal = oraal
Word-1=/aayi, /pOyi	Vowel	(-,1, s(v2))	poyi+ennu = poyennu
Word-1=/u,/um, Word-1 category = verb	Vowel	(1,1, ss(v2))	Pokum + illa = pokilla

Keralapanini (Varma, 2000) has classified the compounding rules based on the changes occur during the compounding as:

- *lOpa sandhi* (elison), one of the sounds is lost;
- *aagama sandhi* (addition), new sound is added;
- dvitva sandhi (germination or reduplication),

Table 5: Elision Rules (Lopa Sandhi)

one of the sounds geminates;

- *aadEsa sandhi* (displacement or substitution), one of the sounds is displaced by another sound.

The rules which we have created for compounding in each category are given in Tables 3, 4, 5 and 6; Table 3 shows the Substitution Rules (AAdesa Sandhi), Table 4 shows the Addition Rules (Aagama Sandhi), Table 5 shows the Elision Rules (Lopa Sandhi). The morphophonemic changes at the boundary depend on the ending vowel or consonant, the category of the first or the second word and the beginning vowel or consonant of the second word.

2.5 Sanskrit compounding

Word 1st end	Word-2nd -beg	Substitution	Example	Sandhi
/a	/a, /aa	(-1,ss(aa))	Padya+ avasanam = padyaavasanam	Deergha Sandhi
/i	/I, /ii	(1,1,ss(ii))	Kavi +iisvaran = kaviisvaran	
/u	/u, /uu	(1, 1,ss(/uu))	Guru+upakaram = guruupakaram	
/A	/i, /u	(,1,ss(ee/oo))	Sara+upadesam = saaroopdaesam	Guna Sandhi
/aa	/E	(1,1,ss(/ya))	Sada+eevam =sadaivam	Vridhi Sandhi
/i	Vowels except / e	(1,1,ss(/ya))	Athi+aavasyam = Athyavasyam	
/u	/a, /aa	(1,1,ss(/va))	Uru+aagamanam = gurvagamanam	YaNa Sandhi

Table 6: Sanskrit Compounding Rules

Malayalam and Sanskrit shares many of the compounding rules since Malayalam is derived from Sanskrit. Compounding in Sanskrit are:

- vowel compounding *(swarasandhi), joining of* two vowels;
- consonant compounding *(vyanjanasandhi),* consonants join.

D*eerghasandhi, guNasandhi, vridhisandhi and yaNsandhi are the further classifications of Vowel Sandhi*. Table 6 shows the Sanskrit compounding rules.

3. Morphology Generation Technique

The SMT systems face the problem of data sparsity; the data does not have enough inflectional forms when translating from a morphologically poor language to a morphologically rich language. Another case is that data sparseness arises only when using factored models. To handle this, we need to generate all combinations of the factors used. We have used a Morphology injection method that generates various morphological forms of noun and verb entities by classifying them and augments the training data with newly generated morphological forms of nouns. The basic algorithm of the Morphology injection method can be described as below:

1. Find out the noun/verb entity pairs (Eng-Mal)
2. Categorize Malayalam nouns/verbs into classes
3. Generate new morphological forms of the nouns using the rules
4. Augment the training data with the generated inflected forms

We have created rules for handling the inflections in noun and verb. Moreover, for handling the word compounding phenomena in Malayalam, we have created rules for elision, substitution, addition and for Sanskrit compounding and are presented in Tables 1, 2, 3, 4, 5 and 6. Then we have generated the respective inflected forms using the created rules with the help of a parallel dictionary of root words between English and Malayalam.

3.1 General Factored model for handling morphology

Factored translation models allow additional annotation at the word level by considering word as a vector of tokens. Factored translation models can be seen as the combination of several components (language model, reordering model, translation steps, and generation steps). These components define one or more feature functions that are combined in a log linear model [Koehn and Hoang, 2007]:

$$p(e|f) = \frac{1}{Z}\sum_{i=1}^{n} \lambda_i h_i(e,f) \quad (1)$$

From equation (1), each h_i is a feature function for a component of the translation, the λ_i values are weights for the feature functions and Z is the normalization constant.

Figure 1 shows a general factored model approach for translation from a morphologically poor language to a morphologically rich language. On the source side we have: Surface word, root word, and set of factors *S* that affect the inflection of the word on the target side. On the target side, we have: Surface word, root word, and suffix (can be any inflection). The model has the following mapping steps:

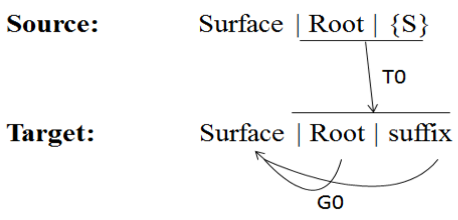

Figure 1: Factored model setup to handle inflections

- **Translation step (T0):** Maps source root word and factors in *S* to target root word and target suffix

  Generation step (G0): Maps target root word and suffix to the target surface word. Note that the words which do not take inflections have *null* as values for the factors in *S*.

Figure 2 shows the factored model setup to handle nominal inflections in Malayalam.

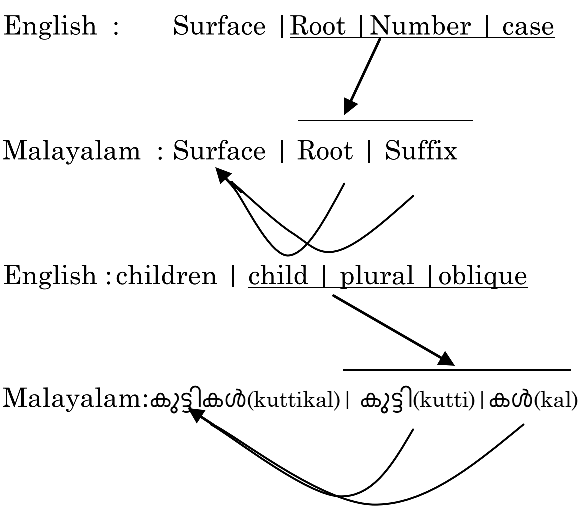

Figure 2: Factored model setup to handle nominal inflections

4. Experimental Discussion

We performed the experiments on ILCI (Indian Languages Corpora Initiative) English-Malayalam dataset. Domain of the corpus is health and tourism. We used 46,000 sentence pairs for training and 3000 sentence pairs for testing. The inflected-form dictionary was created using the Malayalam word lexicon. It consisted of 50,000 noun forms and 150,000 verb forms of Malayalam. The generated verb and noun forms have been validated manually over a period of 6 months with an English-Malayalam bilingual expert who is having a Master's degree in Malayalam Literature. Table 7 shows the statistics of the corpus used for training, testing, tuning and the generated word-form dictionary. *Moses*[1] toolkit was used for training and decoding. Language model was trained on the target side corpus with *IRSTLM*[2].

Sl. No	Corpus Source	Training Corpus	Corpus Size [Parallel Sentences]
1	ILCI	Health	23000
2	ILCI	Tourism	23000
Total			46000
Sl. No	Corpus Source ILCI	Tuning Corpus (MERT) size	Testing Corpus Size [Parallel Sentences]
1	Tourism	500	1500
2	Health	500	1500
Generated inflected form dictionary			Parallel DictionarySize
Noun			1,00,000
Verb			1,50,000
Total			2,50,000

Table 7: Statistics of the corpus used

For our experiments, we compared the translation output of the following systems:

- Phrase-based (unfactored) model (**Phrase**);
- Basic factored model for solving noun and verb morphology (**Fact**);
- Phrase-based model trained on the corpus used for *Phr* augmented with the word form dictionary for solving noun and verb morphology (**Phrase-Morph**);
- Factored model trained on the corpus used for *Fact* augmented with the word form dictionary for solving noun and verb morphology (**Fact-Morph**).

---

[1] http://www.statmt.org/moses/
[2] https://hlt.fbk.eu/technologies/irstlm-irst-languagemodelling-toolkit

With the help of syntactic and morphological tools, we extract the number and case of the English nouns and number, person, tense, aspect and modality of the English verbs. We have followed both the automatic evaluation (BLEU score (Papineni et al., 2002)) and subjective evaluation procedure with the help of linguistic experts as described in (Sreelekha et.al.(2013) for evaluating the systems. Table 8 shows the experimental results in terms of BLEU score evaluation and Table 9 shows the experimental results in terms of subjective evaluation (Fluency and Adequacy). For evaluation, we randomly chosen 250 translation outputs from each system were manually given adequacy and fluency scores. The scores were given on the scale of 1 to 5 going from worst to best, respectively. The BLEU score and subjective evaluations shows promising improvements in terms of the improvement of translation quality for both the Phrase and Factor based models.

Morph Problem	Model	Adequacy	Fluency
		En-Ml	En-Ml
Noun	Fact	28%	35%
	Fact-Morph	56%	67.32%
Verb	Fact	31.43%	45.09%
	Fact-Morph	58.89%	71.23%
Noun & Verb	Fact	35.67%	46.02%
	Fact-Morph	51.36%	65.34%
Noun	Phrase	26.87%	36.21%
	Phrase-Morph	50.56%	64.67%
Verb	Phrase	27.87%	37.12%
	Phrase-Morph	52.56%	64.87%
Noun & Verb	Phrase	33.87%	36.12%
	Phrase-Morph	57.56%	68.12%

Table 9: Subjective Evaluation Results-Morphology Injection

Morph Problem	Model	BLEU Score	
		Without Tuning	With Tuning
		En-Ml	En-Ml
Noun	Fact	26.17	28.23
	Fact-Morph	32.42	33.45
Verb	Fact	26.54	28.82
	Fact-Morph	36.54	38.30
Noun & Verb	Fact	24.01	26.08
	Fact-Morph	31.56	32.65
Noun	Phrase	26.78	29.01
	Phrase-Morph	31.30	33.12
Verb	Phrase	26.98	29.17
	Phrase-Morph	32.41	35.56
Noun & Verb	Phrase	27.51	29.92
	Phrase-Morph	35.03	37.73

Table 8: BLEU score Evaluation of Morphology Injection

5. Conclusion

SMT approaches suffer due to the morphological complexity when translating into a morphologically rich language. We solve this problem by enriching the original data with the missing morphological forms of words in Malayalam. Morphology injection performs very well and improves the translation quality. We observe huge improvement in BLEU score, adequacy and fluency of the translation outputs. We observe up to 38.30 improvements in BLEU score, up to 58.89% improvement in adequacy and up to 71.23% improvement in fluency. This method is more effective when used with factored models than the phrase-based models. A possible future work is to generalize the approach of morphology generation and verify the effectiveness of morphology injection on more morphologically complex languages.

6. Acknowledgments

The authors would like to thank Dr. Suman Mary Idicula and Dr. Latha R. Nair of CUSAT for their helps during this work. This work is funded by Department of Science and Technology, Govt. of India under Women Scientist Scheme-WOS-A with the project code-SR/WOS-A/ET-1075/2014.

7. Bibliographical References

Antony P. J. 2013. *Machine Translation Approaches and Survey for Indian Languages*, The Association for Computational Linguistics and Chinese Language Processing, Vol. 18, No. 1, March 2013, pp. 47-78

Bharathi A. and Sangal, V.R. (1996) *Natural Language Processing – A Paninian Perspective*. New Delhi: PHI.

Jurafsky D., Martin H.M., "Speech and natural language processing", Prentice Hall, 2003.

Koehn, Philipp, Franz Josef Och, and Daniel Marcu. 2007. *Statistical phrase-based translation*. NACAL on Human Language Technology- Volume 1. ACL.

Koehn, Philipp and Hieu Hoang. 2007. *Factored Translation Models*. In Proceedings of the 2007 Joint Conference on EMNLP and CoNLL, pages 868–876.

Kunchukuttan, A., Mishra, A., Chatterjee, R., Shah, R., and Bhattacharyya, P. (2014) *Shata-Anuvadak: Tackling Multiway Translation of Indian Languages*, LREC 2014, Rekjyavik, Iceland, 26-31.

Marton, Y., Callison-Burch, C. and Resnik, P. (2009) *Improved Statistical Machine Translation Using Monolingually-derived Paraphrases*, EMNLP, Volume 1- Pages 381- 390.

Minkov, E., Toutanova, K., and Suzuki, H. 2007. Generating complex morphology for machine translation. ACL 07, pages 128–135.

Nair, Latha R. and Peter S. David 2012. *Machine Translation Systems for Indian Languages*, International Journal of Computer Applications (0975 – 8887), Volume 39– No.1, February 2012.

Nakov, P. I. and Ng, H. T. (2012). Improving Statistical Mahcine Translation for a Resource-Poor Language Using Related Resource-Rich Languages, Journal of AI Research, Volume 44, pages 179-222.

Namboodiri, E.V.N. (1998) *Kerala Bhasha Vyakaranam*. Kottayam: DC Books., Kerala, India.

Och, F. J. and Ney, H.. *A Systematic Comparison of Various Statistical Alignment Models.* Computational Linguistics, 2003.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). *BLEU: a Method for Automatic Evaluation of Machine Translation*, ACL Philadelphia, pp. 311-318.

Sreelekha, Dabre, R., Bhattacharyya, P. (2013). *Comparison of SMT and RBMT, The Requirement of Hybridization for Marathi–Hindi MT,* ICON.

Sreelekha. S, Bhattacharyya, P., Malathi.D. (2018). “Statistical vs. Rule Based; A Case Study on Indian Language Perspective”, Springer Journel of Advances in Intelligent Systems and Computing Vol. 632, DOI 10.1007/978-981-10-5520-1_60.

Sreelekha, Bhattacharyya, P., Malathi, D. (2015). *A Case Study on English- Malayalam Machine Translation*, iDravidian Proceedings.

Sreelekha, Dungarwal, P., Bhattacharyya, P., Malathi, D. (2015). *Solving Data Spasity by Morphology Injection in Factored SMT*, ACL-Anthology, International conference on NLP - ICON 2015.

Sreelekha, Bhattacharyya, P.. (2016). *Lexical Resources to enrich English Malayalam Machine Translation*, LREC .

Varma, A.R.R.R. (2000) *Kerala Paaniniiyam*. Kottayam: DC Books, kerala, India.

Yamada, Kenji. And Knight, Kevin. 2001. *A Syntax-based Statistical Translation Model,* ACL.