Lexical Resources to Enrich English Malayalam Machine Translation

Sreelekha. S, Pushpak Bhattacharyya

Dept. of Computer Science & Engg., IIT Bombay, Mumbai, India

{sreelekha, pb}@cse.iitb.ac.in

Abstract

In this paper we present our work on the usage of lexical resources for the Machine Translation English and Malayalam. We describe a comparative performance between different Statistical Machine Translation (SMT) systems on top of phrase based SMT system as baseline. We explore different ways of utilizing lexical resources to improve the quality of English Malayalam statistical machine translation. In order to enrich the training corpus we have augmented the lexical resources in two ways (a) additional vocabulary and (b) inflected verbal forms. Lexical resources include IndoWordnet semantic relation set, lexical words and verb phrases etc. We have described case studies, evaluations and have given detailed error analysis for both Malayalam to English and English to Malayalam machine translation systems. We observed significant improvement in evaluations of translation quality. Lexical resources do help uplift performance when parallel corpora are scanty.

Keywords: Lexical Resources, Statistical Machine Translation, English-Malayalam Machine Translation

1.   Introduction

Each Machine processing of Natural (Human) Languages has a long tradition, benefiting from decades of manual and semi-automatic analysis by linguists, sociologists, psychologists and computer scientists among others. Development of a full-fledged bilingual Machine Translation (MT) system for any two natural languages with limited electronic resources and tools is a challenging and demanding task. Since India is rich in linguistic divergence there are many morphologically rich languages quite different from English as well as from each other, there is a large requirement for machine translation between them. Development of efficient machine translation systems using appropriate methodologies and with limited resources is a challenging task. There are many ongoing attempts to develop MT systems for Indian languages (Antony, 2013; Kunchukuttan et al., 2014; Sreelekha et al., 2014; Sreelekha et al., 2015) using both rule based and statistical approaches. There were many attempts to improve the quality of Statistical MT systems such as; using Monolingually-Derived Paraphrases(Marton et al., 2009), Using Related Resource-Rich languages (Nakov and Ng, 2012) Considering the large amount of human effort and linguistic knowledge required for developing rule based systems, statistical MT systems became a better choice in terms of efficiency. Still the statistical systems fail to handle rich morphology.

Consider the English sentence,

He has been sent to the mosque for opening the door

The English–Malayalam SMT system translated it as,

Malayalam-   അവൻ അയച്ചു mosque വാതിൽ തുറന്നു
          {avan ayachu mosque vathil thurannu}
          {He sent   mosque door opened}
          {He sent   mosque opened door}

Here the system fails to translate the verb phrase "*has been sent to*" together and it translated a part of the phrase "*sent*" as "*അയച്ചു*"*{ayachu}{sent},* which is wrong in the context. The same way another verb phrase *"for opening the door"* has been translated partly as "*തുറന്നു*" *{thurannu} {opened}.* Also, the system has deficiency in vocabulary and it couldn't translate the English word "*Mosque*". In these kinds of situations in order to learn various inflected forms and verb phrases, lexical resources can play a major role. In this paper we discuss the usage of various lexical resources and how it can be used for improving the translation quality with a detailed analysis about various linguistic phenomena.

2.   Challenges in English –Malayalam Machine Translation

Major design challenges in Machine Translation (MT) are the syntactic structural transfer, which is the conversion from a syntactic analysis structure of the source language to the structure of the target language and the ambiguities.

2.1 Challenge of Ambiguity

There are three types of ambiguities: structural ambiguity lexical ambiguity and semantic ambiguity.

2.1.1.   Lexical Ambiguity

The Words and phrases in one language often have multiple meaning in another language.

For example, the English sentence,

English-   *He picked the photo*
Malayalam-   അവൻ ഫോട്ടോ എടുത്തു
          {avan photo eduthu}

Here in the above sentence "*picked*", has ambiguity in meaning. It is not clear that whether the word "*picked*", is used as the "clicked the photo" (എടുത്തു *{eduthu}* in

620

Malayalam) sense or the "took" sense. However this is a good example where both in source language and target language ambiguity is present for the same word. This kind of ambiguity has to be identified from the context.

2.1.2. Structural Ambiguity

The In this case, due to the structural order, there will be multiple meanings. For example,

Malayalam- അവിടെ പൊക്കമുള്ള പെണ്കുട്ടികളും
ആണ്കുട്ടികളും ഉണ്ടായിരുന്നു
*{avide pokkamulla penkuttikalum ankuttikalum
undayirunnu}*
English - *There were tall girls and boys there*

Here from the words *"പൊക്കമുള്ള പെണ്കുട്ടികളും
ആണ്കുട്ടികളും" {pokkamulla penkuttikalum aankuttikalum}
{tall girls and boys}* it is clear that, girls are tall but it is not clear that boys are tall, since in Malayalam to represent tall girls and boys only one word *"പൊക്കമുള്ള"*
{pokkamulla} {tall} is being used. It can have two interpretations in English according to its structure.

{There were tall girls and boys there}
or
{There were tall girls and fat boys there}

One of the big problems in Machine Translation is to generate appropriate Machine Translations by handling this kind of structural ambiguity.

2.1.3. Semantic Ambiguity

The In this case, due to the understanding of the semantics, there will be multiple translations. For example, consider the English sentence,

I travel with bag and umbrella
I travel with my kids

Here this English sentence can be translated in Malayalam as,

ഞാന് ബാഗും കുടയും കൊണ്ടാണ് സഞ്ചരിക്കാറുള്ളത്
{njan bagum kudayum kontanu sancharikkarullathu}
{I bags umbrella with travel}
or
ഞാന് എന്റെ കുട്ടികളോടൊപ്പമാണ് സഞ്ചരിക്കാറുള്ളത്
{njan ente kuttikalodoppamanu sancharikkarullathu}
{I travel with my kids}

Here, in the two English sentences "*with*" gets translated to കൊണ്ടാണ് *{kontanu} {with}* and ഒപ്പമാണ് *{oppamanu}*
{with} respectively. This disambiguation requires knowledge to distinguish between "*bag- umbrella*" and "*kids*".

2.2 Structural Differences

There are word order differences between English and Malayalam such as, English language follows Subject -Verb- Object (SVO) and Malayalam language follows Subject- Object-Verb (SOV).

Consider an example for word ordering,

English- *Gita went to market*
(S) *(V)* *(O)*
Malayalam- ഗീത ചന്തയില് പോയി
{Gita chanthayil poyi}
(S) *(O)* *(V)*

In addition, Malayalam is morphologically very rich as compared to English, wherein there are a lot of post-modifiers in the former as compared to the later.

For example, the word form *"കടലില്" {kadalil}* {in the sea} is derived by attaching *"ില്" {il}* as a suffix to the noun *"കടല്"{kadal}{sea}* by undergoing an inflectional process. Malayalam exhibits agglutination of suffixes which is not present in English and therefore these suffixes has equivalents in the form of pre positions. For the above example, the English equivalent of the suffix *"ില്"*
{il} is the pre position "in the" which is separated from the noun "sea". This kind of structural differences have to be handled properly during MT.

2.3 Vocabulary Differences

Languages differ in the way they lexically divide the conceptual space and sometimes no direct equivalents can be found for a particular word or phrase of one language in another.

Consider the sentence,

നാളെ കളഭാഭിഷേകം ഉണ്ട്
{ nale kalabhabhishekam undu}

Here the word, "കളഭാഭിഷേകം" *{kalabhabhishekam}* as a verb has no equivalent in English, and this word have to be translated as "the pooja which will cover the idol with sandalwood". Hence the sentence will be translated as,

Malayalam- നാളെ കളഭാഭിഷേകം ഉണ്ട്
{ nale kalabhabhishekam undu}
English- *Tomorrow, the pooja which will cover
the idol with sandalwood is there.*

Translating such language specific concepts pose additional challenges in machine translation. .

3. Experimental Discussion

We now describe our experiments and results on phrase based baseline SMT system[1] for English- Malayalam and Malayalam – English, specifically with the usage of lexical resources. We use Moses (Koehn et al., 2007) and Giza++[2] for learning the statistical models (Och 2001). There are structural differences between Malayalam and English and in the generation of word forms due to the morphological complexity. In order to overcome this

---

[1] http://www.cfilt.iitb.ac.in/SMT-EM
[2] http://www.statmt.org/

difficulty and make the machine to learn different morphological word forms, lexical resources can play a major role. Different word forms such as verb phrases, morphological forms prepositional phrases etc can be used. Moreover the SMT system lacks in vocabulary due to the small amount of parallel corpus. Comparative performance studies conducted by Och and Ney (2003) have shown the significance of adding lexical words into corpus and the improvement in the translation quality. We have used lexical words, IndoWordnet (Bhattacharyya, 2010), verb phrases etc. to increase the coverage of vocabulary. We have done many experiments to improve the quality of machine translation by augmenting various lexical resources into the training corpus. The statistics of lexical resources used are shown in Table 1 and the results are shown in Tables 2, 3, 4 and 5.   Our experiments are listed as below:

Sl. No	Corpus Source	Training Corpus [Manually cleaned and aligned]	Corpus Size [Sentences]
1	ILCI	Tourism	23750
2	ILCI	Health	23750
3	Joshua	Tourism	29518
	Total		77018
Sl. No	Lexical Resource Source	Lexical Resources in Corpus	Lexical Resource Size [Words]
1	CFILT, IIT Bombay	Indo Wordnet Synset words	25341
2	CFILT IITB, Joshua, Olam	Lexical words	144505
3	CFILT IITB	Verb Phrases	200544
	Total		370390
Sl. No	Corpus Source	Tuning corpus(MERT) [Manually cleaned and aligned]	Corpus Size [Sentences]
1	ILCI	Tourism	250
2	ILCI	Health	250
	Total		500
Sl. No	Corpus Source	Testing corpus [Manually cleaned and aligned]	Corpus Size [Sentences]
1	ILCI	Tourism	1000
2	ILCI	Health	1000
	Total		2000

Table 1: Statistics of Corpus and Lexical resources used

3.1 SMT system with an unclean corpus

The learning of proper grammatical structures was prevented by the stylistic constructions, misalignments, wrong and missing translations etc. present in the unclean corpus (Refer Table-1). This reduced the translation quality. For example, consider a sentence from the corpus where the translation is wrong,

English: Tus is located on the banks of the Berach river near Udaipur and the Sun temple here has an important place in the study of sculpting tradition .

Equivalent Malayalam Translation (wrong)

ബേടച്ച് നദിയുടെ തടത്തില് സ്ഥിതി ചെയ്യുന്ന റ്റൂസ് ഉദയ്പുരിന്റെ സൂര്യക്ഷേത്രത്തില് ശില്പകലയുടെ മഹത്വമുണ്ട്.

{*bedach nadiyude thadathilu sdhithi cheyyunna tuse udaipurinte soorykhsethrathile silpakalayude mahathwamundu*}

{*berach river's bank located Tus Udaipur's Sun temple sculpting importance*}

{*Tus located on berach river's bank udaipur's sun temple important sculptin tradition*}

The comparative performance results of cleaned corpus over uncleaned corpus were shown in the Tables 2, 3, 4 and 5..

3.2 SMT system with cleaned corpus

In order to improve the quality of translation, we have removed the stylistic constructions, unwanted characters, wrong translations from the parallel corpus. We have corrected the grammatical structures, missing translations, wrong phrases, misalignments between parallel sentences which improves the learning of word to word alignments. Consider the sentence discussed in section 3.1, which has both source side and target side translation errors. We have corrected the translations as,

English : Tus located on the banks of the Berach river near Udaipur and the Sun temple have an important place in the study of sculpting tradition.

Malayalam : ഉദയ്പുരിന്റെ അടുത്ത് ബേടച്ച് നദിയുടെ തടത്തില് സ്ഥിതി ചെയ്യുന്ന റ്റൂസിനും സൂര്യക്ഷേത്രത്തിനും ശില്പ കലയുടെ പഠനത്തിന് പ്രത്യേക സ്ഥാനമുണ്ട്.

{*udaipurinte aduthu bedahcu nadiyude tadathil sthiti cheyyunna toosinum sooryakshethrathinum silpakalyude padanathinu pratyeka sthanamundu*}

{*Udaipur's near Berach river's bank located Tus and Sun temple sculpting tradition's for study important place* }

{*Tus located on the banks of the Berach river near Udaipur and the Sun temple have an important place in the study of sculpting tradition.*}

After cleaning, the translation quality improved to more than 14 times compared to system with unclean corpus. We observed during error analysis that, the machine lacks

in sufficient amount of vocabulary and hence we investigated on the usage of lexical words to improve the quality of machine translation.

3.3 Corpus with lexical words

We have extracted a total of 437832 parallel English-Malayalam lexical words from parallel corpus. We have also used dictionary words available from Joshua[3] corpus and Olam[4] dataset after manual validation. For example, while translating the English sentence,

He reached early on for the movie
SMT system translated it in Malayalam as,

അവൻ സിനിമയ്ക്ക് എത്തി
{avan cinemaykku ethi}
{He cinema reached}
{He reached cinema}

Here the system failed to translate the meaning of the multiword ***early_on.*** In lexicalWord list it has the following equivalent,

early_on : മുൻപുതന്നെ*{munputhanne}*

We have augmented the extracted parallel lexical words into the training corpus. After training, the above sentence is translated correctly as,

അവൻ സിനിമയ്ക്ക് മുൻപുതന്നെ എത്തി
{avan cinemaykku munputhanne ethi}
{He for the cinema early on reached}
{He reached early on for the movie}

Since the lexical words are extracted from the same corpus, it helped in improving the translation quality to a great extent. During the error analysis we observed that even though the machine translation system is able to give considerably good quality translation, it faces difficulties in translating different words and its concepts. Hence we investigated the usage of word-synsets to make the system learn the words with its concepts.

3.4 Corpus with IndoWordnet synsets

We have used an algorithm to extract the bilingual words from IndoWordnet according to its semantic and lexical relations (Bhattacharyya 2010). Bilingual mappings are generated using the concept-based approach across words and synsets (Kumar et.al, 2008). We have considered all the synset word mappings for a single word and generated that many entries of parallel words. For example, the word ***beautify*** has the following equivalent synset words in the IndoWordnet.

beautify: ശോഭിക്കുക അലങ്കരിക്കുക സജ്ജീകരിക്കുക
മനോഹരമാക്കുക മോടി_പിടിപ്പിക്കുക
{beautify: shobhikkuka alnkkarikkuka sajjeekarikkuka manoharikkuka modi_pidippikkuka}
{beautify: shining decorating arranging beautify

---

making_up }

Consider an English sentence,
Decorations should beautify the occasion
The SMT system translated it in Malayalam as,

അലങ്കാരം ചെയ്യു അവസരം
{alankaram cheyyu avasaram}
{decorations do occasion}

Here the system fails to translate the meaning of "*beautify*" correctly. After augmenting the synsets of beautify to the corpus, SMT system was able to translate the equivalent English meaning in *Malayalam as,*

അലങ്കാരം അവസരത്തിനെ മനോഹരമാക്കണം
{alankaram avasarathine manoharamakkanam}
{decoration occasion should beautify }
{ Decorations should beautify occasions }

Since the synsets covers all common forms of a word, the augmentation of extracted parallel synset words in to the training corpus not only helped in improving the translation quality to a great extent but also, helped in handling the word sense disambiguation well. But we observed during error analysis that the system fails in handling case markers and inflected forms and further we investigated on handling it.

3.5 Corpus with verb phrases

In order to overcome the verbal translation difficulty we have programmatically extracted English - Malayalam parallel verbal forms and their translations which contain various phenomena with a frequency count. In addition we have used pos-tagged corpus to extract verbal phrases. We have augmented the manually validated 200544 entries of verbal translations into the training corpus.

Consider an English sentence,
English: He took the decision for being alive
SMT system translated it in *Malayalam as,*

അവൻ തീരുമാനം എടുത്തു ജീവൻ
{ avan theerumanam eduthu jeevan}
{ He decision took live}

Here the system fails to translate and convey the importance of verb phrase *"for being alive" in this sentence.* After augmenting the corpus with the equivalent meaning of English - Malayalam verb phrase pair,

for being alive: നിലനിൽക്കാൻ വേണ്ടി
{ nilanilkkan vendi }
{being alive for }
the system translated the sentence correctly as,

അവൻ നിലനിൽക്കാൻ വേണ്ടി തീരുമാനം എടുത്തു
{avan nilanilkkan vendi theerumanam eduthu}
{He being alive for decision took}
{He took the decision for being alive}

English-Malayalam Statistical MT System		BLEU score	MET- EOR	TER
With Unclean Corpus	Without Tuning	1.26	0.110	96.06
	With Tuning	2.52	0.117	93.55
With Cleaned Corpus	Without Tuning	25.06	0.193	84.05
	With Tuning	29.07	0.206	80.92
With Lexical words	Without Tuning	31.78	0.262	78.91
	With Tuning	32.94	0.269	74.30
With Wordnet synsets	Without Tuning	35.21	0.351	73.79
	With Tuning	36.05	0.353	70.06
With verb Phrases	Without Tuning	38.15	0.355	67.94
	With Tuning	39.90	0.358	64.48

Table 2 : Results of English- Malayalam SMT BLEU score, METEOR, TER Evaluations

Malayalam-English Statistical MT System		BLEU score	MET-EOR	TER
With Unclean Corpus	Without Tuning	1.16	0.105	95.32
	With Tuning	1.80	0.119	91.52
With Cleaned Corpus	Without Tuning	22.01	0.187	86.32
	With Tuning	25.22	0.190	83.7
With Lexical words	Without Tuning	28.65	0.210	79.56
	With Tuning	30.54	0.226	76.30
With Wordnet synsets	Without Tuning	32.20	0.263	73.19
	With Tuning	34.46	0.283	71.19
Corpus with verb Phrases	Without Tuning	36.10	0.299	68.36
	With Tuning	37.90	0.355	63.88

Table 3: Results of Malayalam-English SMT BLEU score, METEOR, TER Evaluations

Malayalam-English Statistical MT System		Adequacy	Fluency
With Unclean Corpus	Without Tuning	12.87%	16.3%
	With Tuning	15.56%	19.65%
With Cleaned Corpus	Without Tuning	51.01%	61.21%
	With Tuning	54%	65.32%
With Lexical words	Without Tuning	59.08%	71.21%
	With Tuning	62.01%	75.04%
With Wordnet synsets	Without Tuning	67.36%	79.21%
	With Tuning	69.1%	81.68%
Corpus with verb Phrases	Without Tuning	72.01%	84.32%
	With Tuning	74.89%	85.34%

Table 4: Results of Malayalam-English SMT Subjective Evaluation

English-Malayalam Statistical MT System		Adequacy	Fluency
With Unclean Corpus	Without Tuning	10.6%	14.8%
	With Tuning	17.8%	24%
With Cleaned Corpus	Without Tuning	51.6%	62.3%
	With Tuning	55%	66%
With Lexical words	Without Tuning	61.4%	72%
	With Tuning	63.6%	76.2%
With Wordnet synsets	Without Tuning	67%	80.34%
	With Tuning	69.99%	82%
With verb Phrases	Without Tuning	73.01%	86.01%
	With Tuning	77.23%	87%

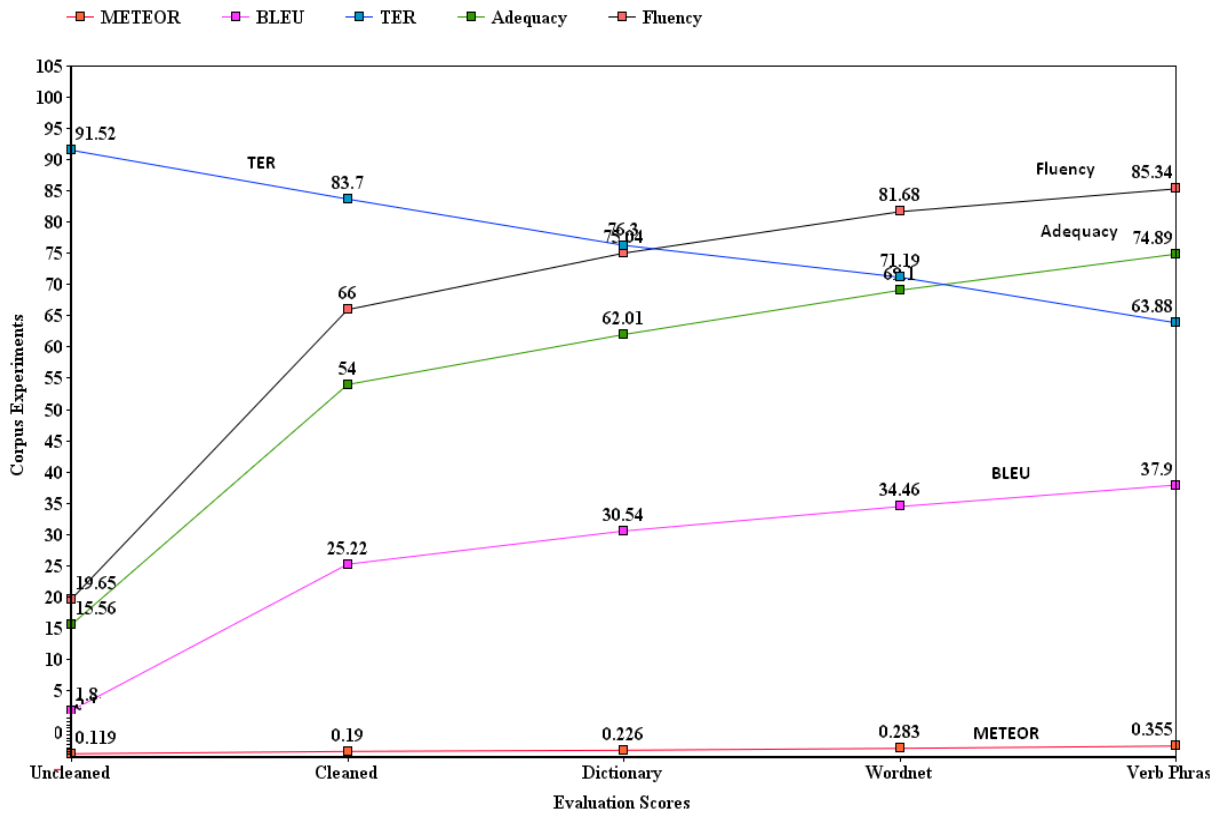Table 5: Results of English- Malayalam SMT Subjective Evaluation

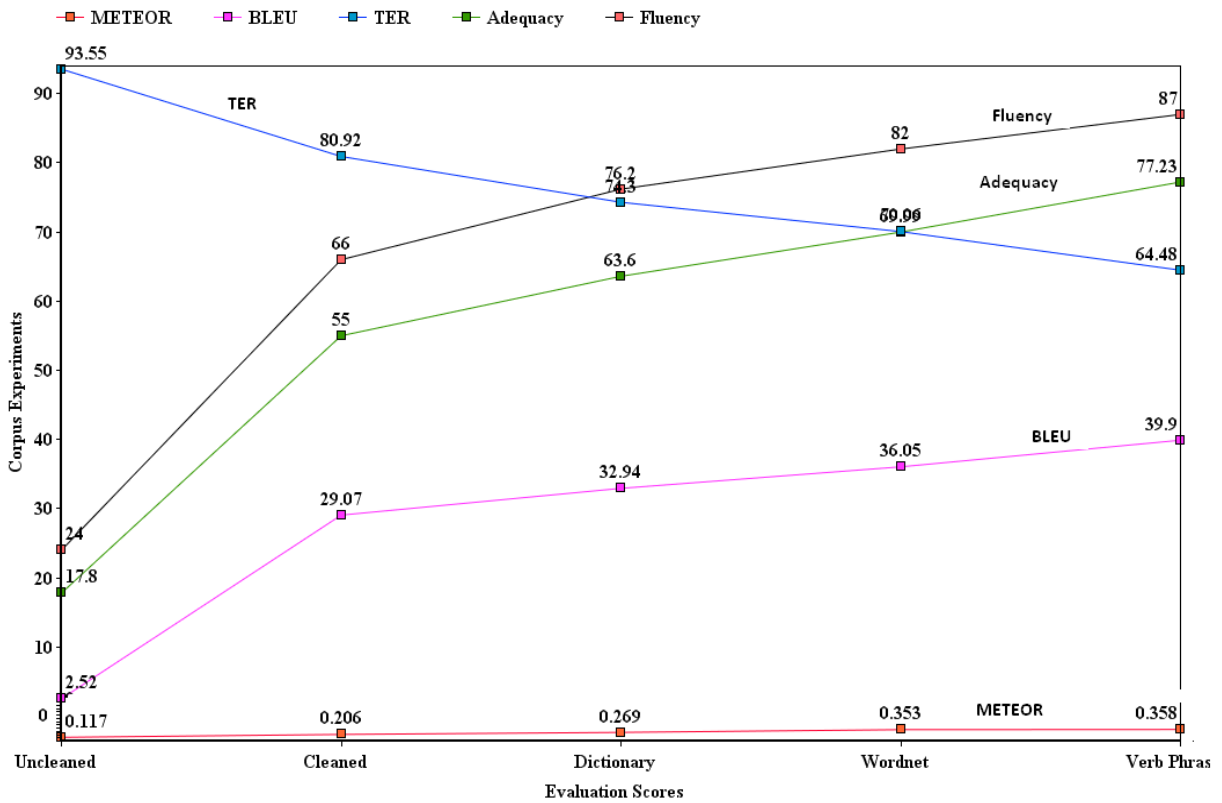Figure 1. Malayalam - English SMT Analysis

Figure 2. English - Malayalam SMT Analysis

The error analysis has shown that the verb phrase augmentation helped in translating verbal inflections correctly and hence the quality of the translation has been improved drastically.

4. Evaluation & Error Analysis

We have tested the translation system with a corpus of 2000 sentences taken from the 'ILCI tourism, health' corpus as shown in Table 4. In addition we have used a tuning (MERT) corpus of 500 sentences as shown in Table 3. We have evaluated the translated outputs of both Malayalam to English and English to Malayalam SMT systems in all 5 categories using various methods such as subjective evaluation, BLEU score (Papineni et al., 2002), METEOR and TER (Agarwal and Lavie 2008). The results of BLEU score, METEOR and TER evaluations are displayed in Tables 2 and 3. We gave importance to subjective evaluation to determine the fluency (F) and adequacy (A) of the translation, since for morphologically rich languages subjective evaluations can give more accurate results compared to other measures. We have followed the subjective evaluation procedure with the help of linguistic experts as described in Sreelekha et.al.(2013) and the results are given in Table 4 and Table 5. Fluency is an indicator of correct grammatical constructions present in the translated sentence whereas adequacy is an indicator of the amount of meaning being carried over from the source to the target. For each translation we assigned scores between 1 and 5 depending on how much sense the translation made and its grammatical correctness.

We have observed that the quality of the translation is improving as the corpus is getting cleaned and more lexical resources are being used. Hence, there is an incremental growth in adequacy, fluency, BLEU score, METEOR score and reduction in TER score. In addition, we were able to handle the one-to-many mapping of phrases to a great extend by increasing the frequency of occurrence with the usage of linguistic resources. The performance comparison graph is shown in figure 1 and figure 2. The fluency of the translation is increased up to 85.34% in the case of Malayalam to English and up to 87% in the case of English to Malayalam, which is 4 times more than the baseline system results.

5. Conclusion

In this paper we have mainly focused on the usage of various lexical resources for improving the quality of Machine Translation for low resource languages. We have discussed the comparative performance of phrase based Statistical Machine Translation with various lexical resources for both Malayalam – English and English - Malayalam. As discussed in the experimental Section, Statistical Machine Translation quality has improved significantly with the usage of various lexical resources. Moreover, the system was able to handle the rich morphology to a great extend. We can see that there is an incremental growth in both the systems in terms of

BLEU-Score, METEOR and a decrement of TER evaluations, which shows the translation quality improvement. Also, our subjective evaluation results show promising scores in terms of fluency and adequacy. This leads to the importance of utilizing various lexical resources for developing an efficient Machine Translation system for morphologically rich languages.

Our future work will be focused on investigating more effective ways to handle the rich morphology and hence to improving the quality of Statistical Machine Translation.

6. Acknowledgements

7. Bibliographical References

Agarwal, A., Lavie, A. (2008), *Meteor, M-Bleu, M-ter Evaluation matrics for high correlation with Human Ranking of Machine Translation output*, Proceedings of the Third Workshop on Statistical Machine Translation,, pages 115–118, Columbus, Ohio, USA, Association for Computational Linguistics.

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

Bhattacharyya, P. (2010). *IndoWordnet*, LREC.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., Herbst, E. (2007) *Moses: Open Source Toolkit for Statistical Machine Translation*, Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Czech Republic.

Kumar, R., Mohanty, Bhattacharyya, P., Kalele, S., Pandey, P., Sharma, A., Kapra, M. (2008) *Synset Based Multilingual Lexical: Insights, Applications and Challenges,* Global Wordnet Conference.

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*, Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing(EMNLP), Volume 1- Pages 381-390.

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.

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

Och F. J. and Ney, H. 2001. *Statistical Multi Source Translation.* MT Summit.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). *BLEU: a Method for Automatic Evaluation of Machine Translation*, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics

(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, 10th International conference on NLP.

Sreelekha, Bhattacharyya, P., Malathi, D. (2014). *Lexical Resources for Hindi Marathi MT, WILDRE proceedings*.

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*, International conference on NLP-ICON 2015.