# Myanmar Phrases Translation Model with Morphological Analysis for Statistical Myanmar to English Translation System

Thet Thet Zin[a], Khin Mar Soe[b], and Ni Lar Thein[a]

[a] University of Computer Studies,Yangon,Myanmar
thetthetzin.ucsy@gmail.com
[b] Natural Language Processing Laboratory
University of Computer Studies, Yangon, Myanmar

**Abstract.** This paper presents Myanmar phrases translation model with morphological analysis. The system is based on statistical approach. In statistical machine translation, large amount of information is needed to guide the translation process. When small amount of training data is available, morphological analysis is needed especially for morphology rich language. Myanmar language is inflected language and there are very few creations and researches of corpora in Myanmar, comparing to other language such as English, French, and Czech etc. Therefore, Myanmar phrases translation model is based on syntactic structure and morphology of Myanmar language. Bayes rule is also used to reformulate the translation probability of phrase pairs. Experiment results showed that proposed system can improve translation quality by applying morphological analysis on Myanmar language.

**Keywords:** Morphological Analysis, Statistical Machine Translation, Bayes rule, Syntactic Structure.

## 1 Introduction

Machine translation (MT) is the task of automatically translating a text from one natural language into another. There exist different approaches to address the problem of machine translation. This paper presents translation model of Myanmar phrases for statistical Myanmar to English machine translation system. Target and source language model based on N-gram (trigram) and translation model based on Bayes' rule to reformulate translation probability P (f| e). N-gram method (trigram) based source language model is used to extract phrases for segmented Myanmar sentence.

Myanmar language likes other Southeast Asia languages that do not place spaces between words. Therefore, the system used Myanmar Word Segmenter (MWS) which is implemented in UCSYNLP Lab which is available for research purpose. Firstly, the system used translation probabilities without additional morphology analysis of Myanmar language. There are many unknown words in this process. Some of the unknown words occur due to the inflective nature of Myanmar Language. Languages may be divided into three broad categories: isolating, agglutinative and inflective languages. Isolating languages, such as Chinese, have little or no morphology and thus do not benefit from morphologically analysis. Agglutinative languages, also known as agglomerative or compounding languages, are those in which basic roots and words can be combined to make new words. These languages, such as Turkish or Finnish, tend to have many morphemes. Inflectional morphemes are used to modify a word to reflect information such as tense.

Myanmar language may be agglutinative language and inflective language because Myanmar word can be combined to make new word. When a form of a word does not occur in the training data, the system is unable to translate it. According to experimental result, the Out-of-Vocabulary (OOV) rate exceeds 50% for tested dataset with 2000 training sentences, which means that half of the words in test set are not present in the training set. Most of the OOV words appear in proper nouns, verb and noun phrases. Regardless of sparseness of the data, the statistical-based approaches have some difficulties with specific natural language processing tasks whereas the rule-based approaches have the advantage of providing analyses of Myanmar language. Therefore, translation model used some rules for syntactic structure and morphological analysis of Myanmar language to improve in translation direction and to reduce the number of unknown words in translation. The rest of this paper is organized as follows: In Section 2, previous works in statistical machine translation is presented. Section 3 presents analysis of Myanmar language. The proposed system is presented in section 4. Finally, Section 5 and 6 discusses translation results and conclusion.

## 2 Related Work

In this section, previous works in Statistical machine translation on different languages are reviewed. Various researchers have improved the quality of statistical machine translation system by using different methods on different language. Probabilistic, models is created for simulating the translation process, in the models using bilingual corpora and then decoding a test sentence by searching (Brown et al., 1990). In 1993, they took the translation process as a noisy-channel model. In terms of modeling Berger et al., 1996 appended context-based information based on the Maximum Entropy principle to enrich the word-based models. Alignment model which is based on phrase structure is firstly proposed by Wang and Waible in 1998, which was automatically acquired from parallel corpus. Och et al., 1999 used beam search algorithm, which could make use of pruning strategies for balancing efficiency and accuracy. In 2002, Och and Ney first introduced the log-linear model into SMT. In 2004 Koehn suggested using features of lexical weighting. In this year, the famous phrase-based decoder, Pharaoh, was released to be a free SMT toolkit by Philipp Koehn and further updated to Moses by Koehn et al., 2007. In 2003 Koehn, Och and Marcu, used noisy channel based translation model and beam search decoder. They achieved fast decoding, while ensuring high quality. They presented experiential result on many languages (English-German, French-English, Swedish-English, and Chinese-English). Log-linear based statistical machine translation model is proposed by Zens and Ney in 2004. They solve search problem using dynamic programming and beam search with three pruning methods. A comparison with Moses showed that the presented decoder is significantly faster at the same level of translation quality.

A few researches investigated the use of morphology to improve translation quality. If source language is morphology rich language (such as German, Spanish, Czech), phrase-based model has limitations. When a form of a word does not occur in the training data, current systems are unable to translate it. Data sparseness problem can be overcome by using large training data or morphology analysis of source or/and target languages. In 2005 Goldwater and McClosky used morphological analysis of Czech to improve a Czech-English statistical machine translation system. This system solve data sparse problem caused by the highly inflected nature of Czech. Their combine model achieved high BLEU score of development and test set. Nguyen and Shimazu, 2006 proposed morphological transformational rules and Bayes' formula based transformational model to translate English to Vietnamese. The score of their system is better than baseline score.

An ideal system for machine translation would take advantage of both empirical data and linguistic analysis. Different authors have different objectives that they attempt to achieve high translation precision on many languages. Our translation model aims is to get correct translation phrases with very limited bilingual corpus for Statistical Myanmar to English machine translation.

Because of the lack of prior research on this task, we are unable to compare to our results to those of other researches; but the results do seem promising.

## 3 Analysis of Myanmar Language

The Myanmar Language is the official language of Myanmar. It is also the native language of the Myanmar and related sub-ethnic groups of the Myanmar, as well as that of some ethnic minorities in Myanmar like the Mon. Myanmar Language is spoken by 32 million as a first language and as a second language by 10 million, particularly ethnic minorities in Myanmar and those in neighbouring countries. Myanmar language is a tonal and pitch-register, largely monosyllabic and analytic language, with a Subject Object Verb (SOV) word order. The language uses the Myanmar script, derived from the Old Mon script and ultimately from the Brāhmī script.

The language is classified into two categories. One is formal, used in literary works, official publications, radio broadcasts, and formal speeches. The other is colloquial, used in daily conversation and spoken. This is reflected in the Myanmar words for "language": စာ sa refers to written, literary language, and စကား sa-ka refers to spoken language. Therefore, Myanmar language can mean either "မြန်မာစာ mran-ma-sa" (written Myanmar language), or "မြန်မာစကား mran-ma-sa-ka:" (spoke Myanmar language). Much of the differences between formal and colloquial Myanmar language occur in grammatical particles and lexical items. Different particles (to modify nouns and verbs) are used in the literary form from those used in the spoken form. For example, the postposition after nouns is၌ hnai: at in formal Myanmar language and မှာ hma: at in colloquial Myanmar language. The proposed system focuses on written Myanmar language.

Example: အမေသည်အိမ်၌ရှိသည်။ Mother is *at* home. (Formal form)
အမေအိမ်မှာရှိတယ်။ Mother is *at* home. (Spoken form)

## 4 Proposed System

Myanmar language does not place space between words. Thus, the proposed translation model use Myanmar Word Segmenter and phrase align Myanmar-English Bilingual corpus. The system created phrases by using N-gram method for input segmented sentence to search in the corpus. In this case, one segmented word is assumed one word. Example:

Myanmar Sentence: ဆရာများသည် သူတို့၏ တပည့်များ ကိုပညာရှိအောင်လုပ်ကြသည်။

"The teachers make their pupil wise."

After word Segmentation: ဆရာများ_သည်_သူတို့_၏_တပည့်များ_ကို_ပညာရှိ_အောင်_လုပ်_ကြ_ သည်_။_

This sentence contains 11 words. ဆရာများ sa-yar-myar 'teachers' is one word. Left-to-right trigrams on segmented input sentence are used to create phrases for translation. If all trigram phrases have not been observed in the corpus, bigrams and unigram phrases are used. If unigram and trigram phrases have the same meaning, longer n-grams is selected. Therefore, the system generally gets less and less number of phrases. Phrases for input sentence according to the longest N-gram method are: ဆရာများ၊ သည် ၊ သူတို့၏ ၊ တပည့်များကို ၊ ပညာရှိအောင် ၊ လုပ်ကြသည်

The system used Bayes' rule to reformulate the translation probability for translating Myanmar phrases into English phrases. Among all possible target language phrases, we will choose the phrases with the highest probability:

$$E = \arg\max_{e} \{\Pr(e_1^I \mid f_1^J)\} \qquad (1)$$

$$= \arg\max_{e} \{ \Pr(e_1^I) . \Pr(f_1^J \mid e_1^I) \} \qquad (2)$$

This allows for a language model Pr (e) and a separate translation model Pr (f |e). The system are not focus on English phrase reordering. Rearranging the English phrases is implemented in separate part as a subsystem of statistical Myanmar to English translation system.

## 4.1 Problems in the System

No additional morphology for Myanmar language is applied to translation model. The system is unable to learn translations of words that do not occur in the data, because they are unable to generalize. Translation model knows nothing of morphology therefore it fail to connect different word forms. Myanmar verbs can have many suffixes and some suffixes have the same meaning by attaching the main verbs. This is difficult for translation. For instance; စားသည် sar-ti: ‘eat’; စား၏ sar-ei: ‘eat’; စားပါသည် sar- par-ti: ‘eat’ have different verb suffixes သည်: ti; ၏: ei; ပါသည်: par-ti. But they have the same meaning to translate English language. The root verb is စား sar: ‘eat’ and they are present tenses.

Unknown words can be reduce by using large training data or morphology analysis of source or/and target language. The large scale Myanmar Corpus is unavailable at present. Morphology analysis is complex and computationally expensive method. Therefore, we analyzed OOV words category for morphology analysis process. The system separately takes 215 parallel sentences as testing datasets and 12827 sentences is used as the training dataset. There is no overlap of parallel sentences between training and testing datasets. The effect of the rate of out-of-vocabulary (OOV) words on translation quality, the training dataset is divided into several different smaller sizes. Figure 1 shows the OOV rate of Myanmar-English testing dataset. We measure OOV based on types (each word in the vocabulary) as well as tokens (each word in the text).X-axis and Y-axis represent number of training sentences and OOV rate (%) respectively.

According to the figure 1, the OOV rate increases as the number of training sentences decreases.

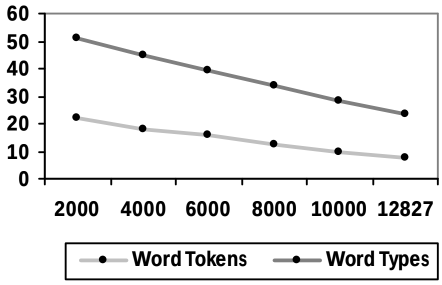


**Figure 1:** OOV Rate on Myanmar-English Test Set

Most of the unknown words occur in proper noun, noun and verb phrases. To reduce unknown words in noun and verb phrases, the system considers morphology analysis on number category of noun phrases, suffixes and particles of verb phrases for Myanmar language. Morphological analysis is applied on pre-processing phrase of translation process.

When translation model has learned multiple possible translations for a particular word or phrase, the choice of which translation to use is guided by conditional probability rather than by linguistic information. Sometimes linguistic factors like case marker, tense, or number categories of noun phrases are important determinants for what translation ought to be used in a particular context. Because phrase-based approaches lack linguistic information they do not have an appropriate means of choosing between alternative translations. The system selects one word according to conditional probability. Therefore, sometime translation result is incorrect.

Some postpositional markers have ambiguous meanings in translation. One way of helping the disambiguation of ambiguous words is use syntactic structure of language and to annotate words with their part-of-speech (POS). Example of ambiguous in postpositional markers (PPM) is shown in table 1.

**Table 1:** Ambiguous in Postpositional Markers

| Myanmar Sentences | Postpositional Markers | English translation |
|---|---|---|
| မျက်မှန်နှင့်လူကလေးသည်တော်သည်။ 'The boy with glasses is clever.' | နှင့် | နှင့် (with) |
| အမေနှင့်မမသည်ဈေးမသို့သွားသည်။ 'Mother and Ma Ma go to the market'. | နှင့် | နှင့်(and) |

In the first sentence, PPM နှင့် nint is used-PPM. Its meaning is နှင့် nint 'with. But in the second sentences, PPM နှင့် nint is similar-PPM. Its meaning is နှင့် nint "and". Therefore, translation of PPM is depended on sentence structure. We annotated Myanmar POS tags manually. We appended the Myanmar and English POS tags in training and test corpus to compare with the system without applying Myanmar POS.

## 4.2 Morphology of Myanmar Verb Phrases

The roots of Myanmar language verbs are almost always suffixed with at least one particle which conveys such information as tense, intention, politeness, mood, etc. These verb suffixes make us difficult in translation of Myanmar to English. Because some suffixes have the same tense and the same meaning. However, Burmese verbs are not conjugated in the same way as most European languages; the root of the Burmese verb always remains unchanged and does not have to agree with the subject in person, number or gender. The most commonly used verb particles and their usage are shown below with an example verb root ကစား ka-sa 'play'. The statement ကစား ka-sa 'play' is imperative.

- The suffix သည် ti (literary form) can be viewed as a particle marking the present tense and/or a factual statement: ကစားသည် ka-sa-ti 'play'
- The suffix ခဲ့ hkai denotes that the action took place in the past. The suffix သည် ti in this case သည် ti denotes a factual statement rather than the present tense: ကစားခဲ့သည် ka-sa-hkai- ti 'played'
- The particle နေ nay is used to denote an action in progression. It is equivalent to the English '-ing': ကစားနေသည် ka-sa-nay ti 'playing'
- The particle မည် mai (formal form: မည်) is used to indicate the future tense or an action which is yet to be performed: ကစားမည် ka-sa- mai 'will play'

Verbs are negated by the particle မ ma, which is prefixed to the verb. When the corpus contains only imperative verb ကစား ka-sa 'play', we can generally decide Myanmar verb tense by looking verb particles ခဲ့ hkai 'past tense', နေ nay 'continuous tense', မည် mai 'future tense'. Myanmar verb can be divided into three main categories: Individual Verb, Compound Verb and Adjective Verb. For example: individual verb: စားသည် sar-ti: 'eat'; compound verb: ပြေးဖက်သည် pye-pet-ti 'run and hug'; Adjective Verb: ပျော်သည် pw-ti 'is happy'. Some verbs can be used to support other verbs. For example: ပြောသည် pyaw-ti 'tell' and ပေးသည် pay-ti 'give' are individual verbs and can be used as main verbs in sentences. But in this verb ပြောပေးသည် pyaw-pye-ti 'tell', ပေး pye 'give' is not the main verb. It behaves particle to support the main verb ပြော pyaw 'tell'.

More than two individual English verbs can include in Myanmar compound verb. For example: three individual verbs: ကြွရောက် kyw-rwet ‘come’, အားပေး ar-pye ‘encourage’, ချီးမြှင် khee-hmyint ‘award’ include in compound verb : ကြွရောက်အားပေးချီးမြှင့်သည် kyw-rwet-ar-pye-khee-hmyint-ti ‘come and encourage and award’. ‘ကြွရောက်အားပေးချီးမြှင့်သည်’ is Myanmar Compound verb. It has three English individual verbs “come, encourage and award”. Verb particle ကြ kyat can be omitted in the sentence. For example: ကျောင်းသားများ ကစားနေ ကြသည်။ ‘Students are playing.’ and ကျောင်းသားများ ကစားနေ သည်။ ‘Students are playing’. Compound Verbs pose special problems to the robustness of a translation method, because the word itself must be represented in the training data: the occurrence of each of the components is not enough.

## 4.3 Verb Phrases Detection for Morphological analysis

Different languages may differ in their syntactic structure in general: for instance the placement of the verb in sentence or the use of postpositional markers in the sentences. Currently, no mature deep analysis that has been worked done is available for Myanmar language. The proposed system detects verb phrases in Myanmar sentence by using syntactical structure of sentence. Myanmar language is SOV pattern. Verb suffixes are at the end of Myanmar sentences and Myanmar verb (stem) is very complex to define. Example of Myanmar verbs are shown in Table 2.

**Table 2:** Examples of Myanmar Verbs

| Myanmar Verbs | Main Verbs | Suffixes | English meaning |
|---|---|---|---|
| လုပ်အားပေးနေသည်။ | လုပ်အားပေး | နေသည် | contributing |
| အားပေးနေသည်။ | အားပေး | နေသည် | encouraging |
| ပေးနေသည်။ | ပေး | နေသည် | giving |
| နေသည်။ | နေ | သည် | live |

Verb suffixes are mined from any Myanmar sentences by using N-gram method in the system. Main verb is in front of suffix. We will first provide some examples to illustrate this concept and conclude this section with a formal definition. The following sentences are collected from Myanmar grammar books. The main verbs are marked by parentheses and suffixes are marked by italics. Example:

(1)မောင်မြသည်သူမအားပန်းများကို(ပေး)*သည်*။(Mg Mya gives flowers to her.)

(2)မောင်လှသည်ကျောင်းသို့မြန်မြန်(သွား)*ခဲ့သည်*။(Mg Hla went to school quickly.)

(3)မောင်ဘသည်စာဖတ်ခြင်းကိုအလွန်(ကြိုက်နှစ်သက်)*သည်*။ (Mg Ba likes reading very much.)

According to syntactic structure of Myanmar language, generally verb phrases are always end of the sentence. Firstly verb suffixes are extracted and then define main verb. According to analysis, post positional markers or adverb phrases are in front of main verbs. In above examples post positional marker ကို:ko , adverb phrases မြန်မြန် myan-myan ‘quickly’, အလွန်:ar-lwan ‘very’ are in front of main verbs. The system defined five types of adverb, seventeen types of postpositional markers and thirteen types of verb particles according to Myanmar grammar rules to detect verb phrases in the sentences. The system does not consider complex sentences structure with the conjunction words. Some verb phrases are same in main verb category but different in suffixes category. But they have the same meaning in translation. The system solved this problem by defining possible verb suffixes groups.

## 4.4 Inflectional Form of Number Category of Noun

In Myanmar language, the particles "များ၊ တို့၊ တွေ" makes plural form of noun according to Myanmar grammar. In English, we don't want to explicitly store the plural of every noun, since there are mostly very predictable. The relatively few exceptions can be stored separately and rules used to generate the rest. The same applies to verb, where the addition of ***s, ing,*** and ***ed*** should be handled by morphological rules (although exceptions are somewhat more common). The regular plural of nouns (eg. book, class) is formed by adding ***s*** or ***es***. Example for one regular rule is:

Rule 1: Singular words which end in vowel + *y* add *s*

Examples: boy/boys, key/keys

Rule: *Vy*→*Vys* (V is any vowel.)

In either case, words not covered by the rules can be marked in the lexicon as having irregular plurals (eg. ***man/men***). Moreover some verbs have irregular forms (such as past tense of "**read**" is also "**read**"). Morphological rules cannot handle irregular verb forms. We also used lexicon as having irregular verb forms to handle irregular verb forms.

# 5 Translation Results

## 5.1 Corpus Statistics

For experiments, the corpus contains sentences from Myanmar text books, grammar books and websites. Corpus statistics are shown in table 3 and 4. Zawgyi-One Myanmar font is used for Myanmar Language. The system separately take 215 parallel sentences as testing datasets, and the remaining is used as the training dataset. There is no overlap of parallel sentences between training and testing datasets.

**Table 3:** Corpus Statistics

| Myanmar-English | | |
|---|---|---|
| Sentences Pairs | 13042 | |
| Language | Myanmar | English |
| Total Word | 61824 | 56263 |
| Vocubary Size | 2713 | 2405 |
| Average Sentence Length (Word) | 18 | 10 |

**Table 4:** Statistics of the Myanmar-English datasets

| Sentence Pairs of Datasets | | Total Words | | Vocabulary Size | |
|---|---|---|---|---|---|
| | | Myanmar | English | Myanmar | English |
| Train | 12827 | 60805 | 55335 | 2168 | 1965 |
| Test | 215 | 1019 | 928 | 545 | 440 |

## 5.2 Evaluation Criteria

Machine translation can be evaluated using the well-known measures precision, recall, and the F-measure. The F-measure has significantly higher correlation with human judgments than recently proposed alternatives. In this paper, we measure evaluation of the translation system in term of the standard measure of precision, recall and F-measure in equation 3, 4 and 5. Only single references are used in this measure. These reference sentences are manually translated. Our system does not consider word order of Myanmar and English language. Therefore, we ignore the word order of candidate and reference sentences.

$$\Pr eision \quad (C \mid R) = \frac{|C \cap R|}{C} \qquad (3)$$

$$\operatorname{Re} call \quad (C \mid R) = \frac{|C \cap R|}{R} \qquad (4)$$

$$F - measure = \frac{2 * (precision * recall)}{(precision + recall)} \quad (5)$$

C=set of candidate phrases
R=set of reference phrases

Table 5 shows the results for Myanmar-English translation with varying sizes of training sentences. According to the table, the proposed method begins to get some improvements over the corresponding baseline. When the size of training data sentences is less than 10000 sentences, morphology analysis method has good compared with the corresponding baselines. In proposed system, most errors occur in postpositional markers. Postpositional markers have ambiguous meaning in translation. One way of helping the disambiguation of ambiguous words is use syntactic structure of language and to annotate words with their part-of-speech (POS). Therefore, we annotated Myanmar POS tags manually. We appended the Myanmar and English POS tags in training and test corpus to compare with baseline system. By using POS tags, the system reduced ambiguous in postpositional markers.

**Table 5:** Translation Results

| Corpu ssize (sent:) | Baseline (only probability) (%) | | | Baseline+ morpho (%) | | | Baseline+ morpho+POS (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precisi on | Recall | F-Score | Precisi on | Recall | F-Score | Preci sion | Recall | F-Score |
| 5000 | 67.3 | 68.3 | 67.8 | 69.2 | 68.5 | 68.8 | 70.1 | 71.8 | 70.9 |
| 10000 | 70.3 | 69.6 | 69.9 | 71.6 | 70.1 | 70.8 | 77.3 | 74.8 | 76 |
| 12827 | 73.2 | 71.5 | 72.34 | 76.1 | 71.6 | 73.8 | 79.7 | 80.7 | 80.2 |

The best results got by adding morphology and POS of Myanmar language to baseline system. We also analyzed OOV words in proposed system. The system reduced OOV words in noun and verb phrases. Compound verbs and proper nouns pose problems to the robustness of a translation method and increased unknown words rate in translation. OOV words reduction is shown in table 6.

**Table 6:** OOV reduction rate

| Category of OOV words | OOV% | OOV words |
|---|---|---|
| Nouns | 22.8 | 89 |
| Verbs | 22.3 | 87 |

## 5.3 Error Analysis

Compound verbs and proper nouns pose problems to the robustness of a translation method. For example: compound verb သွားစားသည် twe-sar-ti 'go and eat' has two meaning. သွားသည် thwr-ti; 'go', စားသည် sar- ti 'eat' and meaning of သွားစားသည် twe sar ti is 'go and eat'. Althoug the corpus contain သွားသည် thwr-ti 'go' and စားသည် sar-ti 'eat', we have difficult to translate these words သွားစားသည် twe-sar-ti 'go and eat' to get correct translation. Because the word itself must be represented in the training data: the occurrence of each of the components is not enough. Some errors occurred in adjective. Myanmar adjectives vary according to sentence patterns. There are 95 errors in tested sentences. The causes in detail are:

- Unknown words: The foreign word did not occur in the training corpus, so translation was not possible at all.
- Unknown translation: The word occurred in the training corpus, but fails to translate: fail to align the word to its correct translation, which often happens for rare words.
- Segmentation Error: Word Segmenter output is not suitable for correct translation result.
- Detecting verb phrases Error: Errors in finding verb phrases in the input sentence especially when input sentence is too long and include conjunction words.
- Untranslatable: Some phrases are not translatable into English phrase correctly.
- Others: missing English particle in noun phrases and so on.

## 6 Conclusion

We have shown that Myanmar-English phrase-based SMT can be improved by combining the syntactic structure, POS and morphological analysis of Myanmar Language. By adding these three features the system can achieve a better result than can be obtained with each individually. This improvement was primarily due to a reduction of the sparse data problem caused by the highly inflected nature of Myanmar language. An alternative method for reducing this problem is to use a larger parallel corpus. However, the large scale Myanmar Corpus is unavailable at present. For that reason, we believe that the approach presented in this paper is a promising one. In the future, we would like to apply other Myanmar morphological features in translation model and to test in more training data and domain specific corpus.

## References

Brown Peter F., John .Cocke, Stephen.A.Della Pietra, Vincent J.Della Pietra, Fredrick. Jelinek, John.D. Lafferty, Robert.L. Mercer, and Paul.S. Roossin.1990. *A statistical approach to machine translation.* Computational Linguistics, 16(2), 79-85.

Brown Peter F., Vincent. J. Della Pietra, Stephen. A. Della. Pietra, and Robert. L. Mercer. 1993. *The mathematics of statistical machine translation: parameter estimation*. Computational Linguistics, 19(2), 263-311.

Berger, A. L., S. A. D. Pietra, and V. J. D. Pietra, *A maximum entropy approach to natural language processing,* Computational Linguistics 22(1), 39-72, 1996.

Department of the Myanmar Language Commission, Ministry of Education Yangon, 1998. *Myanmar-English Dictionary*.

Department of the Myanamar Languge Commission, Ministry of Education, Union of Myanmar. 2005. *Myanmar Grammar.*

Goldwater Sharon and McClosky David. 2005. *Improving Statistical MT through Morphological Analysis*. Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pages 676-683, Vancouver, October.

Htay Hla. H., G.Bharadwaja.Kumar, Kavi .N.Murthy. 26 march 2007. *Statistical Analyses of Myanmar Corpora,* Technical report, Department of computer and Information Science, University of Hyderabad. 26 march 2007.

Koehn Philipp. 2004. *Pharaoh: A beam Search Decoder for Phrase-Based Statistical Machine Translation Models.* In proceedings of the Sixth Conference of the Association for Machine Translation in the Americas..pp 115-124.

Koehn Philipp, Marcello, Cowan Brooke, Zens Richard, Dyer Chris, Bojar Ondrej, Constantin Alexandra and Herbst Evan.2007. *Moses: Open Source Toolkit for Statistical Machine Translation.* In proceedings of the ACL 2007 Demo and Poster Sessions, PP 177-180.

Koehn Philipp and Hoang Hieu 2007 June. *Factor Translation Models*. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, PP.868-876, Pragun.

Koehn, P., F. J. Och, and D. Marcu, *Statistical phrase-based translation*, In proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL), Edomonton, Canada,2003.

Marcu Daniel and Wong William. 2002 Jul. *A phrase-based joint probability model for statistical machine translation.* In proceeding of EMNLP, pp 133-136.

Minkov, E., Toutanova, K., and Suzuki, H. 2007. *Generating complex morphology for machine translation*. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics. Prague, Czech Republic, 128-135.

NieBen Sonja and Ney Hermann. 2004. *Statistical machine translation with scarce resources using morpho-syntactic information.* Computational Linguistics 30, 2, 181-204.

Nguyen Thai Phuong and Shimazu Akira. 2006 Augest. *Improving Phrase-Based SMT with Morpho-Syntactic Analysis and Transformation*. Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, pages 138-147, Cambridge.

Och. Franz. J., Tillmann Christoph, and Ney Hermann. *Improved alignment models for statistical machine translation*. In proceedings of the Conference on Empirical Method in Natural Language Processing and Very Large Corpora, University of Maryland, College Park, MD, pp. 20-28.

Och.Franz. J and Ney. Hermann. *Discriminative training and maximum entropy models for statistical machine translation*. 2002. in proceedings of ACL, pp. 295-302.

Pa Pa Win, Thein Ni Lar. *Disambiguation in Myanmar Word Segmentation.* February 2009. Proceedings of the seventh international conference on computer applications, Pages 1-4.

Richard Zens and Hermann Ney, *Improvements in Phrase-Based Statistical Machine Translation*, 2004.

Thumana Ashin 2004. *New Method English Grammar*. Second Edition. Yangon, Myanmar.

Wang Ye-Yi and Waibel Alex.1998. *Modeling with structures in statistical machine translation*. In proceedings of COLING/ACL, Montreal, Quebec, Canada, pp. 1357-1363.