

Applying Rule-Based Maximum Matching Approach for Verb Phrase Identification and Translation (Myanmar to English)

Soe, Thae Thae¹, Thida, Aye²

¹University of Computer Studies, Mandalay, Myanmar

² University of Computer Studies, Mandalay, Myanmar

Abstract: *Phrase Identification is one of the most critical and widely studied in Natural Language processing (NLP) tasks. Verb Phrase Identification within a sentence is very useful for a variety of application on NLP. One of the core enabling technologies required in NLP applications is a Morphological Analysis. This paper presents the Myanmar Verb Phrase Identification and Translation Algorithm and develops a Markov Model with Morphological Analysis. The system is based on Rule-Based Maximum Matching Approach. In Machine Translation, Large amount of information is needed to guide the translation process. Myanmar Language is inflected language and there are very few creations and researches of Lexicon in Myanmar, comparing to other language such as English, French and Czech etc. Therefore, this system is proposed Myanmar Verb Phrase identification and translation model based on Syntactic Structure and Morphology of Myanmar Language by using Myanmar- English bilingual lexicon. Markov Model 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: Myanmar verb phrase identification and translation, morphological analysis, Rule-Based Maximum Matching

1. Introduction

Language plays an important role in human communication because it is used as a channel not only for expressing thoughts but also for exchanging information. In the age of Information Technology, The Internet has become a primary source for people to exchange their thoughts and information .It is simply and convenience for all people around the word. However, they have difficult to communicate among them because of different their native languages. Some people are familiar with two or more kind of Languages, spoken and written languages but most are not. Due to these difficulties and increased use of network, there is an increased need for language translation to facilitate among people in communication, publication and learning subjects. Attempts of language translation are almost as old as computer themselves. Machine Translation (MT) is the attempt to automate all or part of the process of translation between human languages and is one of the oldest large-scale applications of computer science. Developing a system that accurately produces a good translation between human languages is the goal of MT system.

Human Language translation is a difficult task for natural language because there has language ambiguity and it varies according to their features and nature. Myanmar word transformations are similar to other Asian Language including Indian, Japanese, Thai and Chinese Language. The problem of Machine Translation can be view as consisting of three phrases (i) analysis of the source language to choose appropriate target language lexical item (words or phrases) , (ii) reordering phrase where the chosen target language lexical items are reordering to produce a meaningful target language sentence and (iii) disambiguation of words senses where the correct meaning of words is chosen for translation.

The Myanmar-English MT system is developed by composing two main modules which are identification and translation. First, module, identify the Myanmar Verb Phrase from input of Myanmar Sentence. And then, second module, translate the Myanmar Verb Phrase into English Verb Phrase using Myanmar English Bilingual Lexicon. Each step in Machine Translation process is hard technical problem, to which the best known solutions are either not adequate, or good enough only in narrow application domains, falling when applied to other domains. The proposed system is concentrated on improving one of these two steps, namely identification and translation, while having in mind that some of the core techniques can be applied to other parts of a Machine Translation (MT).

There are many research fields in Natural Language Processing system and Machine Translation System. There is no one who has developed complete Machine Translation System for Myanmar to English language. Therefore, this research aims to emphasize and develop the identification and translation of Myanmar verb phrase which is a part of Myanmar-English Machine Translation System.

The rest of this paper is organized as follows: In section 2, previous works in phrase identification for machine translation is presented. Section 3, presents Nature of Myanmar Language and Myanmar sentence structure. The proposed system is presented in section 4. Section 5 presented types of Myanmar Verb Phrase and section 6 described morphological analysis for Myanmar Verb Phrase. Finally section 7, 8, 9 and .10 discusses about the results and error analysis of proposed system and conclusion.

2. Related Work

In this section, previous works in the structure of verb Phrase identification and machine translation on different language are reviewed. Various researchers have improved the quality of machine translation system by using different methods on different language. Wajid Ali et al proposed the structure of Urdu verb phrases, and detail a series of experiment to automatically tag them. A 100,000 words Urdu corpus is manually tagged with VP chunk tags. The corpus is then used to develop a hybrid approach using HMM based statistical chunking and correction rules [11]. The technique is enhanced by changing chunking direction and merging chunk and POS tags. . Kim, Changhyun et.al described [5] Korean-Chinese machine translation system. This system includes source language pattern part for analysis and a target language pattern part for generation. Basically used Pattern-based knowledge and translates Korean Verb Phrase into Chinese Verb Phrase. M. Selvame et al presented an improvement of Rule Based morphological Analysis and POS Tagging in Tamil Language [8] via Projection and Induction Techniques. Rule based approach is applicable to the languages which have well defined set of rules to accommodate most of the words with inflectional and derivational morphology. Fridah Katushemerwr [1] demonstrated the application of finite state approach in the analysis of Runyakitara verb morphology. Language specific knowledge and insight have been applied to classify and describe the morphological structure of the language, and quasi context-free and rewriting rules have been written to account for grammatical verbs of Runyakitara.

In 2005 Goldwater and McClosky [2] 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, [9] 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. Kamaijeetkaur Batra and GS Lehal, [6] presented rule-based machine translation of Noun Phrase from Punjabi to English. The system use transfer approach. The system had analysis, translation and synthesis component. In 2004, Koehn [4] suggested using features of lexical weighting. In this year, the famous phrase-bassed decoder, Pharaoh, was released to be a free SMT toolkit by Philipp Koehn and further updated to Mosses by Koehn et al, 2007. In 2006, Narayan Kumar Choudhary [10] presented about the

Developing a Computational Framework for the Verb Morphology of Great Andamanese.

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 phrase identification and translation model aims to get correct translation phrases with very limited bilingual lexicon for Myanmar to English machine translation.

3. Nature of Myanmar Language

The Myanmar Language is the official language if 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 neighboring 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 Brahmi 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 refers to written, literary language, and စကား refers to spoken language. Therefore, Myanmar language can mean either written Myanmar language or spoken Myanmar Language.

- စာအုပ် စားပွဲ ပေါ်မှာ ရှိတယ်။ (spoken language)
- စာအုပ်သည် စားပွဲ ပေါ်တွင် ရှိသည်။ (formal language)

3.1 Myanmar Sentence Structure

There are two kinds of sentences according to the syntactic structure of Myanmar language. They are simple sentence and complex sentence. Figure1: shows the syntactic structure of Myanmar language.

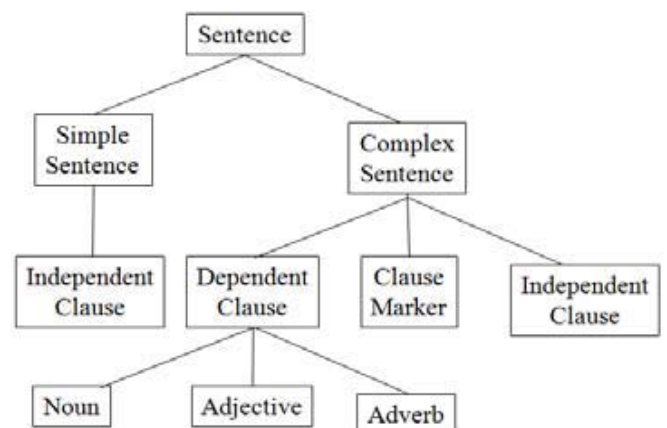


Figure 1: Syntactic Structure

3.1.1 Simple Sentence

The simple sentences are declarative, negative, and interrogative. It contains only one clause. There are two basic phrases such as subject phrase and verb phrase in a simple sentence.

For example:

သူ (Subject phrase) အိပ်နေသည် (Verb phrase)

However, a simple sentence can be constructed by only one phrase. This phrase may be verb phrase or noun phrase.

For example:

စားသွားသည် (Verb phrase)

Besides, a simple sentence can be constructed by two or three phrases.

For example:

ရန်ကုန်တွင် (Place phrase) နေသည် (Verb phrase)

Myanmar phrases can be written in any order as long as the verb phrase is at the end of the sentence.

For example:

ဦးဘသည် မန္တလေးမှ ပြန်လာသည် (Subject, Place, Verb)
မန္တလေးမှ ဦးဘသည် ပြန်လာသည် (Place, Subject, Verb)

A simple sentence can be extended by placing many other phrases between subject phrase and verb phrase. All of the following are simple sentences, because each contains only one clause. It can be quite long.

For example:

ဦးဘသည် မန္တလေးမှ ရန်ကုန်သို့ မီးရထားဖြင့် ပြန်လာသည်။
(U Ba comes back from Mandalay to Yangon by train.)

It is also constructed by adding noun phrases such as subject phrase, object phrase, time phrase and verb phrase. These added noun phrases are called emphatic phrases.

For example:

ပါမောက္ခ ဦးဘသည် သား မောင်မောင်နှင့်အတူ အထက် မန္တလေးမှ မြို့တော် ရန်ကုန်သို့ အမြန် မီးရထားဖြင့် မနေ့ နံနက်က ရော့ရော့မော့မော့ ပြန်လာသည်။

Professor U Ba and his son Mg Mg came back safely from upper Mandalay to capital Yangon by express train in yesterday morning.

3.1.2 Complex Sentence

A complex sentence consists of two or more independent clauses (or simple sentences) joined by postpositions, particles or conjunctions. There are at least two verbs or more than two verbs in a complex sentence. There are two kinds of clause in a complex sentence called independent

clause(IC) and dependent clause (DC). DC is in front of IC. A complex sentence contains one independent clause and at least one dependent clause. DC is the same as IC but it must contain a clause marker (CM) in the end. A clause maker may be post positions, particles or conjunctions. There are three dependent clauses depending on the clause marker.

(1)Noun DC (joined by postpositions such as မှာ(က)ကို)
မမ ဈေးသို့ သွားသည် ကို ကျွန်မ မြင်သည်။

I see that Ma Ma goes to the market.

Noun DC : မမ ဈေးသို့ သွားသည် ကို
IC : ကျွန်မ မြင်သည်။

(2)Adjective DC (joined by particles such as သော ၊ သည် ၊ မည်)
မမ ပေးသော စာအုပ် ကို ကျွန်မ ဖတ်သည်။

I read the book that is given by Ma Ma.

Adjective DC :မမ ပေးသော (စာအုပ်)
IC :စာအုပ် ကို ကျွန်မ ဖတ်သည်။

(3)Adverb DC (joined by conjunctions such as သောကြောင့် ၊ လျက် ၊ သဖြင့်)
မိုးရွာနေ သောကြောင့် ကျွန်မဈေးသို့ မသွားပါ။

I do not go to the market because it is raining.

Adverb DC : မိုးရွာနေ သောကြောင့်
IC:ကျွန်မဈေးသို့မသွားပါ။

3.1.3 Negative Sentence

Generally the negative sentence is ending with “ပါ” and its roots word has prefix “မ” such as “မ..... ပါ”. It also depends on the tense type and modality. For example:

(i) သူသည် (Subject Phrase) ကျောင်းသို့(Noun Phrase) မသွားပါ။
(Verb Phrase)

He doesn’ t go to school.

(ii) လှလှသည် (Noun Phrase) ဒီနေ့ လာလိမ့်မည် မဟုတ်ပါ။
(Verb Phrase)

Hla Hla will not come today?

(iii) စာအုပ်သည် (Noun Phrase) မထူပါ။ (Verb Phrase)

This book is not thick.

Normally, negative meaning of verb is adding prefix “မ” in front of the root verb word. But some verbs have non-linear structure such as “work”. This positive meaning is “အလုပ်လုပ်”, the negative meaning is “အလုပ်မလုပ်” . In this case “မ” is placed within the root words.

3.1.4 Interrogative Sentence

There are two types of questions, yes/no question. Yes/no questions area mentioned in auxiliary verb. In wh-questions, the WH feature identifies the class of Phrase which is signaled by words such as who, what, when, where, why and how (as in how many, how much, how careful). These words fall in several different categories, who, whom, and what can appear as pronouns and can be used to specify simple NPs, what and which appear as determines in NPs, where and when appear as prepositional phrases, how acts as an adverbial modifier to adjective and adverbial phrases and whose acts as a possessive pronoun. The wh-words also can act in different roles such as relative clause. In Myanmar Language question ending format is fixed. The suffix if the yes/no question is "လား" and wh-question is "လဲ" "နည်း".

For example:

(i) မင်းဘုရားပွဲကို (Subject phrase) သွားမလား။ (Verb Phrase)

Will you go to Pagoda festival?

(ii) မင်းစာမေးပွဲ (Subject Phrase) အောင်သလား။ (Verb Phrase)

Do you pass the exam?

(iii) ဤခရီး (Subject Phrase) နီးသလား။ (Verb Phrase)

Is this trip is near?

4. The Proposed System

In Natural Language Processing, some results have already been obtained, however, a number of important research problems have not been solved yet. This section explains the details of Myanmar Verb Phrase identification and translation process by using Rule-Based Maximum Matching Approach. This process accepts the segmented Myanmar words with parts of speech to the system (example မောင်မောင် / NPR /ကျောင်း /NCCS/ သို့ / PODIR /လျှင်မြန်စွာ/ ADVPM /ခပ်သုပ်သုပ် /ADVM/ သွား/ HV/ နေ / PAVPC/ သည်/ POVP). The longest maximum matching method scans the input text by sequentially reading each word from the input text and match the predefined Myanmar Grammar Rule. To identify the Myanmar Verb Phrase, firstly extract the root verb in a given sentences and then, consider the morphological analysis of prefix, suffix and tense particle of the root verb. Finally, the system translates the Myanmar Verb Phrase into English words by using Myanmar-English Bilingual Lexicon. The system's output is လျှင်မြန်စွာ ခပ်သုပ်သုပ်သွားနေသည်။

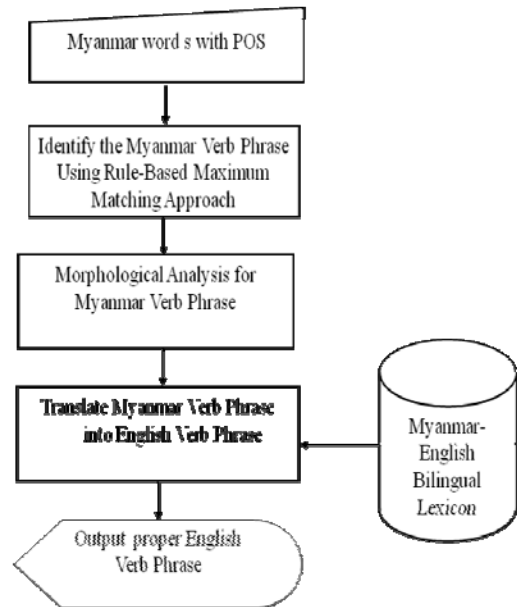


Figure 2: Overview of Proposed System

5. Verb Phrase in Myanmar Language

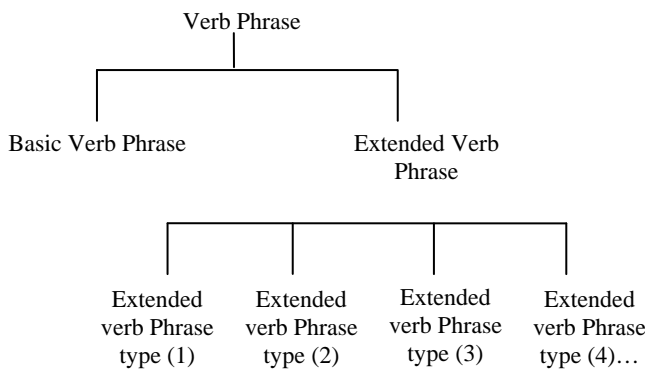
Verb Phrase consists of some adverbial modifiers followed by the head verb or root verb and its complements. Every verb must appear in one of the five possible forms: base, simple present, simple past tense, present participle and past participle. The auxiliary and modal verbs usually take a verb phrase as a complement, which produces a sequence of verbs to form a tense system.

The root form of *be* and *have* and the modal auxiliary such as present and past forms of *do*(*did*), *can*(*could*), *may*(*might*), *shall*(*should*),*will*(*would*), *must*, *need* and *dare* are the auxiliary verbs. In this case, “be” and “have” can be either auxiliary or main verb. These two forms are separate properties. The auxiliary *be* requires a present –participle form or in the case of passive form (past-participle form) of verb phrase to follow it, whereas the verb *be* requires a noun phrase complement or preposition phrase or adjective phrase or adverb phrase. The auxiliary *have* requires a noun phrase complement. English sentences typically contain a sequence of auxiliary verbs followed by a main verb. Auxiliary verbs can be used in declarative sentence, negative sentences and yes/no questions. The structure of Myanmar verb phrase is: ဦးဆောင်ကြိယာ (Root Verb) + ကြိယာဝိဘတ် (Verb Preposition) [3].

Example: သွား သည်။ (go)

Myanmar Verb Phrase can be divided into two types:

- I. အခြေခံကြိယာပုဒ် (Basic Myanmar Verb Phrase)
- II. တိုးချဲ့ကြိယာပုဒ်(Extended Myanmar Verb Phrase)



5.1 Basic Myanmar Verb Phrase

The basic Verb Phrase consists of a Root Verb and Verb Preposition. The Root Verb may be either Action or State or Compound Verb.

Example: မောင်မောင်ကျောင်းသွားသည်။
 အခြေခံကြိယာပုဒ် → ဦးဆောင်ကြိယာ+ကြိယာဂိဘတ်
 BV → HV +POVP
 သွား သည်။ (go)

5.2 Extended Myanmar Verb Phrase

Extended Verb Phrase is based on basic Verb Phrase and it is extended with verb modifiers. There are four types of extended Verb Phrase.

တိုးချဲ့ကြိယာပုဒ် → ကြိယာအထူးပြု+/- လေးအနက်ပြု+/-
 အငြင်းဝါဘတ်+ ဦးဆောင်ကြိယာ+/- ကြိယာထောက်
 တစ်ခု/တစ်ခုထက်ပို + ကြိယာဂိဘတ် တစ်ခု/တစ်ခုထက်ပို

Extended Myanmar Verb Phrase type (1)

In extended verb phrase type (1), one or more adverbs are before the head verb and one or more verb prepositions are after the head verb.

တိုးချဲ့ကြိယာပုဒ် → ကြိယာအထူးပြု + ဦးဆောင်ကြိယာ+
 ကြိယာဂိဘတ်
 EVP → ADV + HV + POVP
 ချောချောမောမော ရောက် သည်။(arrive at safely)

Extended Myanmar Verb Phrase type (2)

In extended verb phrase type (2), one or more verb particles and verb propositions are after the head verb.

တိုးချဲ့ကြိယာပုဒ် → ဦးဆောင်ကြိယာ+ကြိယာထောက်ပစ္စည်း+
 ကြိယာဂိဘတ်
 EVP → HV + PAVP+POVP
 စားနေသည်။ (is eating)

Extended Myanmar Verb Phrase type (3)

According to the extended verb phrase type (3), the negative particle can be included before the head verb. If verb particle is after the head verb, the negative particles may be between

the head verb and verb particle. Then, one or more verb prepositions can be following.

တိုးချဲ့ကြိယာပုဒ် → အငြင်းဝါဘတ်+ ဦးဆောင်ကြိယာ+/-
 ကြိယာထောက် တစ်ခု/တစ်ခုထက်ပို + ကြိယာဂိဘတ်
 တစ်ခု/တစ်ခုထက်ပို
 EVP → PANEG +HV+ PAVPS +/-POVP
 မ အိပ် ချင်သေး ဘူး။ (don't want to sleep)

Extended Myanmar Verb Phrase type (4)

In extended verb phrase type (4), one or more verb modifiers are before the head verb and one or more verb preposition are after the head verb.

တိုးချဲ့ကြိယာပုဒ် → အလေးအနက်ပြု+ ဦးဆောင်ကြိယာ+
 ကြိယာဂိဘတ် တစ်ခု/တစ်ခုထက်ပို
 EVP → ADVM + HV+ POVP
 အရမ်း ကောင်းသည်။ (is very good)

6. Translation with Morphological Analysis for Myanmar Verb Phrase

In Myanmar, Verb does not change its form based on the gender of the subject/object rather it changes with respect to tense, aspect, modality and number only. Including different spelling, there are 38 inflected forms of the root verb in Myanmar. Table 4.1 list the tense suffixes for these different forms. As stated before, Myanmar Verb morphology has some non-linear characteristics. Often, the root changes its form when certain suffixes are added to it based on tenses and on many occasion, it varies non-linear. For example, the verb စား (eat) when followed by suffix “နေ သည်” (present continuous), become “စားနေသည်” , whereas when followed by the suffix “ခဲ့ သည်” (simple past tense) becomes “ စားခဲ့ သည်” , the suffixes “ပြီးပြီ” (future tense) becomes စားပြီးပြီ . The negative meaning of prefix “မ” becomes (does/do not work) “မစားပါ”, suffix “ကြ” (plural of subject) becomes “စားကြသည်” respectively. Similarly, the verb “အလုပ်လုပ်” (work) when followed by suffix nay (present continuous) becomes “အလုပ်လုပ်နေသည်”. But, the negative meaning of prefix ma becomes (does/do not work) “အလုပ်မလုပ်ပါ”. Thus, the addition of the prefix “မ” changes the root forms of “ အလုပ်လုပ်” to “အလုပ်မလုပ်” , which is an indication of non linearity.

Myanmar verb can be divided into three main categories: Individual Verb, Compound Verb and Adjective Verb. For example: individual verb: စားသည် ‘eat’; compound verb: ပြေးဖက်သည် ‘run and hug’; Adjective Verb: ဖျော်သည် pw-ti ‘is happy’. Some verbs can be used to support other verbs. For example: ပြောသည် ‘tell’ and ပေးသည် ‘give’ are individual verbs and can be used as main verbs in sentences. But in this verb ပြောပေးသည် ‘tell’, ပေး ‘give’ is not the main

verb. It behaves particle to support the main verb ပြော 'tell'. More than two individual English verbs can include in Myanmar compound verb. For example: three individual verbs: ကြွရောက် 'come', အားပေး 'encourage', ချီးမြှင့် 'award' include in compound verb : ကြွရောက်အားပေးချီးမြှင့်သည် 'come and encourage and award'. 'ကြွရောက်အားပေးချီးမြှင့်သည်' is Myanmar Compound verb. It has three English individual verbs "come, encourage and award". Verb particle ကြ t 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.

7. Markov Model

Markov Model has been widely used in several of Natural Language Processing tasks. (such as POS tagging, Spell Checking, Machine Translation, Automatic Text Summarization, Information Retrieval (IR), Automatic Text Extraction and so on. This system developed a Markov Model to identify Myanmar Verb Phrases based on predefined Myanmar Grammar Rule-Based Maximum Matching Approach of totally 200 rules. This model constructed both Simple and complex sentences of nearly 2000 sentences.

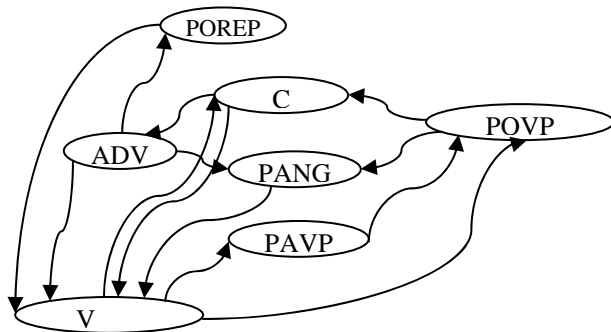


Figure 3: Markov Model for MVP

- PANG={မ}
- V= {သွား, စား, အိပ်, ချက်ပြုတ်, ကနီ, ခွဲ,..}
- C= { ချို, နှင့်,..}
- POREP= {တွင်, မှာ, နဲ့,..}
- ADV= {လျင်မြန်စွာ, ခင်ခင်မင်မင်, အပြေးအလွှား,.. }
- PAVP= {ခွဲ, လိမ့်, နေ,..}
- POVP= {သည်, မည်, ဘူးလား, ဘူး, ဝါ, ဖီ,..}

8. Algorithm for Myanmar Verb Phrase Identification

```

Input: A= {word1, word2... wordn} // Set Segmented words with Part of Speech (Myanmar Sentence.)
Output: Myanmar Verb Phrase into English Proper English Verb Phrase //Translate Verb Phrase using Myanmar- English Bilingual Lexicon.
Begin
Steps:
1. Read input sentence A.
2. Set i =0;
3. Input [i] =A. next token (); //Read input sentence A and tokenized the words by "/" and set to array [i].
4. For(s=0; s<=i: s++)
    4.1 Find VAC, VST or VCP from Input [i]. // where VAC is Act
    On verb, VST is State Verb and VCP is Compound Verb.
    4.2 If (input[i] == "VAC" || input [i] == "VST" || input[i] == "VCP") then
        k=s; // set k to s
        EndIf
    End//for
    if (input[k+2] == "POVP" || input[k+4] == "POVP" )then
        {Input[k] = "HV";
        Identify myanmar verb phrase;
        }
    EndIf
ENDIF
5. DISPLAY MYANMAR VERB PHRASE
END.
    
```

9. Experimental Results

The proposed system, there are nearly 2000 training sentences and 1500 testing sentences. Myanmar 3 font is used for Myanmar Language. The sentences consist of 5 to 35 words. We divided sentences into simple sentences and complex sentences. The simple sentences are declarative, negative and interrogative. Three types of complex sentences are joined with particles, adjective and adverb respectively. The accuracy of verb phrase identification is calculated by using well-known measure precision; recall and F-measure in equation (1), (2) and (3). This system ignore the words order. We have a little limitation in some simple and complex sentences.

$$precision = \frac{\text{no of correctly identified verb phrase}}{\text{no of tested sentences}} * 100\%$$

$$recall = \frac{\text{no of correctly identified verb phrase}}{\text{no of all sentences}} * 100\%$$

$$F - Measure = \frac{2 * (Precision + Recall)}{(Precision + Recall)} * 100\%$$

Table 1: Evaluation Results for Verb Phrase Identification

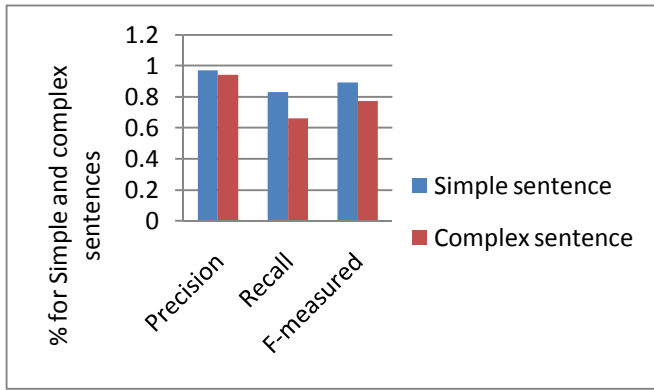


Table 2: Evaluation results for Verb Phrase identification

Type of sentences	No of Sentences	Precision	Recall	F-measured
Simple Sentences	1000	0.97	0.83	0.89
Complex Sentences	500	0.94	0.66	0.77

9.1 Error Analysis

Errors in proposed system are as follow. Compound verb has two meaning. သွားခဲ့ and စားခဲ့သည် and meaning of သွားစားခဲ့သည် is (went and ate). Although our system can translate it as သွားသည်: go) and (စားခဲ့သည် ate), we have difficulty to translate (သွားစားခဲ့သည်) : went and ate) to get correct translation. Some verb support to previous verb: ပြောပေးသည် give), correct translation is “talk”. Beside then in the negative inflection of verb has more error because negative particle of Myanmar “မ” can take as prefix or middle of stem verb such as (“မပြောဘူး”:not tell) and (“နားမထောင်ဘူး”:not listen). In the latter case (“နားမထောင်ဘူး”) is analyzed as “နား” and “မထောင်ဘူး” which as (ear and not stand).

In adjective, we have same error like negative verb inflection like (ရိုသေ respectful) of negative form as (“မရိုသေ”: not respectful) or (“မရိုမသေ”: not respectful). Although the word of “မရိုသေ” is not problem in analyzer, the word “မရိုမသေ” has error occurs.

10. Conclusion

In Natural Language Processing, Phrase identification is one of the most critical and widely used as research area. Verb Phrase identification within a sentence is very useful for a variety of application in Natural Language Processing (NLP). In this paper, Myanmar Verb Phrase identification Algorithm is proposed by developing a Markov Model to show statistical results. In experimental result, a proposed algorithm shows the efficient results with precision, recall and F-measure in simple sentences and complex sentences.

As a future work, after identifying the Myanmar Verb Phrase, translates to English Verb Phrase by using Myanmar-English Bilingual Lexicon. The design and algorithm of the Myanmar Verb Phrase identification and translation system developed in this research can be extended in further research directions in the fields of NLP and IR such as text categorization, document summarization, question answering, query processing and document ranking in search engine development etc.

References

- [1] Fridah Katushemererwr et al, “Finite State Methods in Morphological Analysis of Runyakitara Verbs” Nordic Journal of African Studies.
- [2] 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, pages 676-683, Vancouver , October
- [3] J.Okeli, a.Allot, “Burmese/Myanmar dictionary of grammatical forms”.
- [4] Koehn , P.F.J. Och et al, “ Statistical Phrase-Based Translation”, Processing of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, PP868-876, Prapun.
- [5] Kim, Changhyun et al, “Verb Pattern Based Korean-Chinese Machine Translation System”
- [6] KamaijeeKaur Batra and GS Lehal, “Rule-Based Machine Translation of Noun Phrase from Punjabi to English”, International Journal of Computer Sciences Issue, 2010
- [7] Md. Musfique Anwaar, Mohammad Aabed Anwar et al “Syntax Analysis and Machine Translation of Bangla Sentences”, Dept. of Computer Science & Engineering, Jahangirnagar University, Bangladesh
- [8] M. Selvam et al “ Improvement of Rule-Based Morphological Analysis and POS Tagging in Tamil Language via Projection and Induction Techniques.” INTERNATIONAL JOURNAL OF COMPUTERS, Issue 4, Volume 3, 2009
- [9] Nguyen et al, “Improving Phrase-Based SMT with Morpho-Syntactic Analysis and Transformation. Proceeding of the conference on Empirical Method in Natural Language Processing and Very Large Corpora, University of Maryland, College Park, MD, pp 20-28
- [10] N. K. Choudhary, “Developing a Computational Framework for the Verb Morphology of Great Andamanese”, Centre for Linguistics in India, JNU, 2006.
- [11] Wajid Ali et al, “A hybrid approach to Urdu Verb Phrase Chunking.”, Department of the Myanmar Language Commission, Ministry of Education, Union of Myanmar 2005

Author Profile



Thae Thae Soe received the B.C.Sc. and M.C.Sc. degrees in University of Computer Studies, Mandalay, Myanmar in 2004 and 2008, respectively. I am also an assistance lecturer and

a Ph.D candidate of University of Computer Studies,
Mandalay. My research field is Natural Language Processing
(NLP). I am very interested in NLP.