

# Proposed Framework of Rule-based Grammar Checker for Myanmar Language

Tin Myo Latt, Aye Thida  
University of Computer Studies, Mandalay  
tinmyolatt.tml@gmail.com, ayethida.royal@gmail.com

## Abstract

*Natural language processing is normally used to describe the function of computer system which analyze or synthesize spoken or written language. One area of Natural language processing is concerned with creating proofing systems, such as grammar checkers. Many researchers have been worked for Grammar checker of Asian Languages. However, Myanmar Grammar Checker has not still well developed yet. This paper develops the grammar checker which uses to detect grammatical errors in the formal texts written in Myanmar language. The aim of this paper is to develop the Grammar Checker for detecting grammatical errors in Myanmar texts and resulting from the lack of agreement, order of words in various phrases. Rule-based approach will be used for Grammar Checking System. The proposed framework of this paper is to describe the overview of Myanmar grammar checker.*

**Keywords:** *Natural language processing, grammar checker, Rule-based approach*

## 1. Introduction

Grammar is also considered as natural language since it is a language that is spoken, written or signed by humans for general purpose communication. Natural Language Processing (NLP) which is a branch of artificial intelligent are designed to enable ordinary user to communicate with computer. Some of the applications of NLP include machine translation where it translates human language text to another system. The easiest tasks syntax; a more difficult task is determining the semantic meaning of a sentence. [13]

Grammar (or syntax) refers to a system of rules describing what correct sentences have to look like. Somehow these rules exist in people's minds so that for the vast majority of sentences people can easily decide whether a sentence is correct or not.

A grammar checker looks for grammatical errors and, in many cases, suggests possible corrections.

Grammar checking is one of the most widely used tools within language engineering. Spelling, grammar and style checking for English has been an integrated part of common word processors for some years now. Grammar checker determines the syntactical correctness of a sentence. Grammar checking is mostly used in word processors and compilers. Grammar checking for application like compiler is easier to implement because the vocabulary is finite for programming languages but for a natural language it is challenging because of infinite vocabulary. [2]

Myanmar language is the official language in Myanmar. It is a tonal and syllable-based language. The basic word order of the Myanmar language is subject-object-verb (SOV). The syntactic structure of Myanmar grammatical categories to be able to use in tagging Myanmar text with standard Part-of-Speech tags (POS). In Myanmar lexicon, all words are defined with basic tags and these words can be called as stem words or root words. There are nine POS classes for all Myanmar words. These are Noun, Pronoun, Verb, Adjective, Adverb, Postpositional Marker, Particles and Interjection.

## 2. Related Works

Rozana Kasbon et. al [11] tried to the development of a Malay Language Sentence Checker, which translates short form texting language to its complete and correct forms, checks whether the sentences have any grammatical and structural errors and suggests the type of correction to the sentences. From the functional testing all the system components work as required. The user testing is also conducted and it is found that the system is more suitable to be used by students from lower secondary and primary schools. The developed system is able to translate short form sentences to its longer and complete form, then categorizing the

sentences based on its lexicon and finally checking the grammar structure and suggesting the correction to be made to the sentences.

Debela Tesfaye [2] proposed for Afan Oromo Language, advanced tools have been lacking and are still in the early stages. A rule based grammar checker is presented. The rule base is entirely developed and dependent on the morphology of the language. Different 123 rules were constructed and used in order to identify grammatical error of the language. With the use of these carefully constructed error detection rules, the system can detect and suggest corrections for a number of grammatical errors in Afan Oromo texts. Afan Oromo grammar checker has been developed and tested on real-world errors.

Khaing Htet Win [8] considered Myanmar prepositions play an important role of Myanmar sentences because the percentages of preposition errors are the highest in Myanmar sentence. A Transformation Based Learning (TBL) Algorithm to the automatic correction of preposition errors in Myanmar Language. Myanmar Preposition Checking System (MPCS) which can handle missing preposition errors, misused preposition errors and unwanted preposition errors. This proposed system improves the quality of corrections for Myanmar prepositions errors in students and non-native writers. However it cannot be stated to correct conjunction and particle errors of Myanmar sentences which are ambiguous for poor readers and non-native learner.

Nay Yee Lin et. al [10] proposed Machine Translation systems that expect target language output to be grammatically correct within the frame of proper grammatical category. In Myanmar-English statistical machine translation system, the target language output (English) can often be ungrammatical. To solve this need, they propose an ongoing chunk-based grammar checker for translated English sentences. They intend to develop a grammar checker by using trigram language model and rule based model. It identifies the chunk types and generates context free grammar (CFG) rules for recognizing grammatical relations of chunks. They hope that it encourages improving the translation quality of Myanmar to English.

Win Win Thant et. al [15] Parsing is important in Linguistics and Natural Language processing to understand the syntax and semantics of a natural language grammar. Writing the grammar production for Myanmar language is a bit difficult because Myanmar is a relatively free word order,

morphologically rich and agglutinative language and has a strong case marking system. This paper presents a context free grammar (CFG) based top-down parsing for Myanmar sentences which is the output of function tagging system. They use the function tags that are proposed in our previous function tagging method for producing the grammar rules.

In this paper, they present successful parses of simple and complex Myanmar sentences. The context free grammar for Myanmar sentence parsing has been introduced. They described CFG because it is easier to maintain. They have used the function tags and developed some rules for function tags and phrases.

### 3. Morphology

Morphology is the study of the way words are built up from smaller meaning-bearing units, morphemes. A morpheme is often defined as the minimal meaning-bearing unit in a language. So for example, the word မြစ်(river) consists of a single morpheme (the morpheme မြစ်) while the word ကြောင်များ(cats) consists of two: the morpheme ကြောင်(cat) and the morpheme များ(-s).

As this example suggests, it is often useful to distinguish two broad classes of morphemes: stems and affixes. The stem is the “main” morpheme of the word, supplying the main meaning, while the affixes add “additional” meaning of various kinds. Affixes are further divided into prefix, suffixes, infixes, and circumfixes.

There are many ways to combine morphemes to create words. Four of these methods are common and play important roles in speech and language processing: inflection, derivation, compounding, and cliticization. [4]

#### 3.1. Morpheme for Myanmar

The morpheme is the lowest compositional level of grammatical meaning. Typically, this unit corresponds in Burmese/Myanmar to the syllable of the phonological dimension. Most morphemic units in Burmese/Myanmar also correspond to a basic orthographic unit to which spelling rules apply in Formal Burmese.

Morphemes are either free or bound forms, with the free forms corresponding to word level units and the bound forms to a closed class of grammatical affixes.

There are only productive bound morphemes in Burmese, both are prefixes. The first is the unstressed **ā-** prefix (အ), which prototypically functions as a deverbal nominalizer ( **ā-paing** အပိုင် Nom + ‘possess’> ‘possession’. The second bound morpheme is the negative prefix **mā-** (မ-), which usually is found on the first verb in the final clause of a negated sentence. The prefix **mā-** (မ -) is an immediate constituent of the verb, which is the head of the word construction as in: **mā-swa:** မ -သွား ‘not go’; **mā-kaung:** မ -ကောင်း ‘not good’. [3]

### 3.2. Inflectional Morphology

Inflection is the combination of a word stem with a grammatical morpheme, usually resulting in a word of the same class as the original stem, and usually filling some syntactic function like agreement. For example, Myanmar language has the inflectional morpheme - ချား(-s) for marking the plural on noun, and the inflectional morpheme - ခဲ့ for marking the past tense on verbs. [4]

**Table 1. Noun Morpheme**

	Myanmar	English
Singular	ကြောင်	cat
Plural	ကြောင်များ	cats

**Table 2. Verb Morpheme**

Morphological From Classes	Inflected Verbs	Explanation
Stem	စား	စားသည်(eats)
-s form	ကြ	စားကြသည် (eat)
-ing participle	နေ	စားနေသည် (eating)
Past form	ခဲ့	စား ခဲ့သည် (ate)

## 4. Myanmar Language

Burmese or Myanmar language is the official national language of the Union of Myanmar, previously known as Burma. The Myanmar /Burmese language belongs to the Tibeto-Myanmar language group of the Sino-Tibetan family. It is also morphologically rich and agglutinative language. Myanmar words are postpositionally inflected with various grammatical features.

### 4.1. Myanmar Grammar

Grammar studies the rules behind languages. The aspect of grammar that does not concern meaning directly is called syntax. Myanmar (syntax: SOV), because of its use of postposition (wi.Bat), would probably be defined as a “postpositional language”, whereas English (syntax: SVO) because of its use of preposition would probably be defined as a “prepositional language”. There are really only two parts of speech in Myanmar, the noun and the verb, instead of the usually accepted eight parts (Pe Maung Tin 1956:195). Most Myanmar linguists [14] accepted there are eight parts of speech in Myanmar. Myanmar nouns and verbs need the help of suffixes or particles to show grammatical relations.

### 4.2. Parts of Speech (POS)

The syntactic structure of Myanmar grammatical categories to be able to use in tagging Myanmar text with standard Part-of-Speech tags (POS). In Myanmar lexicon, all words are defined with basic tags and these words can be called as stem words or root words. There are nine POS classes for all Myanmar words.

**Table 3. Parts of Speech Recognized by the Myanmar Language Commission**

Part of Speech	Burmese	Types
Noun	နာမ်	singular, plural, abstract, mass, natural, Compound, Verbal and attributive.

Pronoun	နာမ်စား	personal, demonstrative, interrogative and numeral
Verb	ကြိယာ	action, event/equative, quality, existence
Adjective	နာမ်ဝိသေသ သန	qualitative, demonstrative, enumerative, interrogative
Adverb	ကြိယာဝိသေ သသန	Manner
Division Particle	ဝိဘတ်	sentence final, phrase or clause final
Conjunction	သပ္ပန္န	coordinate and subordinate
Function Particles	ပစ္စည်း	plural, relative clause, restriction on an object, restriction on an action, demonstrates speaker's bravery, interrogative, negative

Exclamation/ Inter-jection	အာမေဇိုတ်	emotive words registering shock, pity, surprise, wonder, fear, unexpected events.
-------------------------------	-----------	---

### 4.3. Myanmar Noun

Nouns are three kinds which are common, proper and personal. Common nouns are names of whole species, proper nouns are names of individuals, as distinguished from others of the same species, and personal nouns are names of individuals, relatively considered, as speaking, spoken to, and spoken of. [1]

**Table 4. Examples of Myanmar Noun**

Noun	Myanmar	English
Common	အိမ်	House
	ကျောင်း	School
	လူ	person
Proper	ရန်ကုန်	Yangon(city-name)
	ဧရာဝတီ	Irrawaddy(river-name)
	မချောစု	Ma Chaw Su (person-name)
Personal	ကျွန်ုပ်,	I
	, ကျနော်,	
	ကျမ	
	သင်	
	သူ	he

Nouns in Burmese are pluralized by suffixing the particle *တွေ* [twe] in colloquial Burmese or *များ* [myar] in formal Burmese. The particle *တို့* [tou] which indicates a group of people or things, is also suffixed to the modified noun.

#### 4.4. Myanmar Verb

The roots of Burmese verbs are almost always suffixed with at least one particle which conveys such information as tense, intention, politeness, mood, etc. Many of these particles also have formal/literary and colloquial equivalents.

The most commonly used of verb particles and their usage are shown with an example verb root *စား*: *ca*: [sá] "to eat". Alone, the statement *စား* is imperative. The suffix *တယ်* *tai* [dɛ́] (literary form: *သည်* *sany* [ðì]) can be viewed as a particle marking the present tense and/or a factual statement: *စားသည်* *ca*: *sany* [sá ðì] "I eat". [7]

Verbs are of two kinds, transitive, which express actions that pass from the agent to the object and intransitive, **which** express being, or some state of **being**, or **an** action which is confined to the agent. [1]

**Table 5. Examples of Myanmar Verb**

Verb	Myanmaar	English
Transitive	ရိုက်သည်	to strike
	ချစ်သည်	to love
Intransitive	ဖြစ်သည်	to be
	နေသည်	<i>to</i> remain
	ကောင်းသည်	<i>to</i> be good
	ပျက်သည်	<i>to</i> be ruined or in a state of <i>ruin</i>

#### 4.5. Myanmar Adjective

Myanmar language does not have adjectives per sentence. Rather, it has verbs that carry the meaning "to be X", where X is an English adjective. These verbs can modify a noun by means of the grammatical particle *တဲ့* [tai] in colloquial and literary form *သော* [sau] which is suffixed as follows:

Colloquial: *ချောတဲ့လူ* *hkyau*: *tai. lu* [tɛʰó dɛ̀ lù]

Formal: *ချောသောလူ* *hkyau*: *so: lu*

Gloss: "beautiful" + adjective particle + "person" [7]

### 5. Materials and Methods

Three methods are widely used for grammar checking in a language; syntax-based checking, statistics-based checking and rule-based checking. [9]

#### 5.1. Syntax-based Checking

Number In this approach, a text is completely parsed, i.e. the sentences are analyzed and each sentence is assigned a tree structure. The text is considered incorrect if the parsing does not succeed [9].

#### 5.2. Statistics-based Checking

In the statistical approach the system is trained on a corpus to learn what is „correct,. In this approach, a POS-annotated corpus is used to build a list of POS tag sequences. Some sequences will be very common (for example determiner, adjective, noun as in the old man), others will probably not occur at all (for example determiner, determiner, adjective). Sequences which occur often in the corpus can be considered correct, whereas uncommon sequences might be considered as errors. This method has a few disadvantages. One of these is that it can be difficult to understand the error given by the system as there is not a specific error message. This also makes it more difficult to realize when a false positive is given. [9]

#### 5.3. Rule-based Checking

Using the rule-based approach to grammar checking involves manually constructing error detection rules for the language. These rules are then used to find errors in text that has already been analyzed, i.e. Tagged with a part-of-speech tagger. These rules often contain suggestions on how to correct the error found in the text. [9]

### 6. Overview of Proposed System

The proposed will use Rule-based approach to develop grammar checker using for Myanmar Language. The major objective of the proposed system is Myanmar sentence to check grammar

errors. To do this, the proposed system performs the following steps:

In step1, this system accepts the input Myanmar sentence to check grammatically errors.

In step 2, the sentence is passed through morphological analyzer which uses full form lexicon to assign each word it's all possible part-of-speech information.

In step 3, this system builds a POS tagging for the sentence. This tagging will use context free grammar for grammatical relation of the function tags.

And then step 4, it recognizes types of phrases within the input sentence, syntactic roles and features of each word. This system will also detect and analyze types of errors in input Myanmar sentence by rule matching.

Finally, the proposed system will display the grammatical errors of input sentence and the result sentence with correct grammar.

### 6.1. Sample Input and Output

This section provides sample Myanmar sentences that were given as inputs to the Myanmar grammar checking system along with the output generated by the system.

The input sentence is as below:

Myanmar: အထက(၁)ရန်ကင်း၊ စ  
စတုတ္ထတန်း(ခ) မှ ဦးပိုးဖာဖာသား မော  
မောင်ပြည့်ဖြိုးပိုင်အား မောင်ပြည့်ဖြိုးပိုင်စိုးဟု ခေါ်ပါရန်

English: Pyae Phyo Paing , the son of U Poe Zar, who attends Grade-5 in Section-B at B.E.H.S(1) YanKin is renamed as Pyae Phyo Paing Soe.

This sentence gives wrong meaning as below:

Myanmar: ဦးပိုးဖာဖာသည် အထက(၁)ရန်ကင်း၊ စ  
စတုတ္ထတန်း(ခ)တွင် တက်ရောက်နေသူဖြစ်သည်။

English: U Poe Zar is attending Grade-5 in Section-B at B.E.H.S (1) YanKin.

This sentence is the meaningless because the meaning of the sentence is that the man U Poe Zar is attending in primary school.

The system will correct the grammar so the output sentence:

Myanmar: ဦးပိုးဖာဖာသားအထက(၁)ရန်ကင်း၊ စတုတ္ထတန်း(ခ) မှ  
မှ မောင်ပြည့်ဖြိုးပိုင်အား မောင်ပြည့်ဖြိုးပိုင်စိုးဟု ခေါ်ပါရန်။

English: The son of U Poe Zar who attends Grade-5 in Section-B at B.E.H.S(1) YanKin is renamed as Pyae Phyo Paing Soe.

This sentence is meaningful grammar correct:

Myanmar: မောင်ပြည့်ဖြိုးပိုင်သည်

အထက(၁)ရန်ကင်း၊ စ တုတ္ထတန်း(ခ)တွင် တက်ရောက်နေသူ ဖြစ်သည်။

English: Pyae Phyo Paing attends Grade-5 in Section-B at B.E.H.S(1) YanKin.

### 7. Conclusion

Grammar checker determines the syntactical correctness of a sentence. Grammar checking is mostly used in word processors and compilers. Grammar checking for application like compiler is easier to implement because the vocabulary is finite for programming languages but for a natural language it is challenging because of infinite vocabulary. Grammar checker is one of the most widely used applications in word processors, which itself is a very important tool for local language. Myanmar grammar checker will develop and test on real-world errors. Many people tend to ignore the importance of grammar, spelling, syntax abbreviation and punctuation in their daily usage of texting and speaking. And then, there is a very little amount of work done in natural language processing of Myanmar Language. So this research work proposes only rule-base grammar checker in NLP for Myanmar Language.

The ongoing research will be described other than rule-based grammar checker, a hybrid system combining both statistic-based and rule-based grammar checker for Myanmar language.

And then this research will be expressed the detail of morphological analysis, context-free grammar rules, error types (agreement errors, missing words, extra words, wrong word order) and algorithms.

### References

- [1] A. Judson, Grammatical Notices of the Burmese Language, Maulmain: American Baptist mission Press. 1842
- [2] D. Tesfaye, "A rule-based Afan Oromo Grammar Checker", *International Journal of Advanced Computer Science and Applications*, Vol. 2, No.8, 2011
- [3] D. Bay, S. Johnson, Burmese Grammatical Units
- [4] D. Jurafsky & James H. Martin, *An introduction to natural language processing, Computational linguistics, and speech recognition*, Speech and

- Language Processing, Copyright c 2006, All rights reserved. Draft of June 25, 2007.
- [5] Department of the Myanmar Language Commission, Myanmar-English Dictionary. Ministry of Education. Myanmar, 2006.
- [6] Department of Myanmar Language commission, Ministry of education, Union of Myanmar, မြန်မာစာ မြန်မာစကား, June 2007
- [7] [http://en.wikipedia.org/wiki/Burmese\\_Language](http://en.wikipedia.org/wiki/Burmese_Language)
- [8] K.H. Win, “Myanmar Preposition Checking System”, *International Journal of Computer Applications* (0975 – 8887) Volume 106 – No.8, November 2014.
- [9] Naber, D., “A Rule-Based Style and Grammar Checker”, Diplomarbeit. Technische Fakultät Bielefeld, 2003.
- [10] N.Y. Lin, K.M. Soe, and N.L. Thein, “Developing a Chunk-based Grammar Checker for Translated English Sentences”, *25th Pacific Asia Conference on Language, Information and Computation*, pages 245–254
- [11] R. Kasbon, N.A. Amran, E.M. Mazlan and S. Mahamad, “Malay Language Sentence Checker”, *World Applied Sciences Journal 12* (Special Issue on Computer Applications & Knowledge Management): ISSN 1818-4952, 19-25, 2011.
- [12] S.P. Soe (2010) မြန်မာဘာသာစကား: Aspects of Myanmar Language
- [13] Terry, H., 2010. PCAI-Where Intelligent Technology Meets the Real World, Retrieved February 2010 from the World Wide Web: [http://www.pcai.com/web/ai\\_info/natural\\_lang\\_proc.html](http://www.pcai.com/web/ai_info/natural_lang_proc.html).
- [14] T. Lwin, နည်းသစ်မြန်မာသဒ္ဒါ, 1978.
- [15] W.W. Thant, T.M. Htwe and N.L. Thein, “Context Free Grammar Based Top-Down Parsing of Myanmar Sentences”, *International Conference on Computer Science and Information Technology (ICCSIT'2011) Pattaya* Dec. 2011.