

# New Phrase Chunking Algorithm for Myanmar Natural Language Processing

Dr. Myintzu Phyo Aung

Assistant Lecturer

Research and Development Department

University of Computer Studies, Mandalay, Myanmar

# Outline

---

- Abstract
- Introduction
- Proposed Myanmar Phrase Chunking Algorithm
- Myanmar Grammar
- Chunk Markers of Myanmar Language
- Evaluation and Discussion
- Conclusion and Future Work

# Abstract

---

- Chunking is the subdivision of sentences into non recursive regular syntactical groups.
- The chunker can operate as a preprocessor for Natural Language Processing systems.
- This study aims to propose new phrase chunking algorithm for Myanmar natural language processing.
- The developed new algorithm accepts Myanmar tagged sentence as input and generates chunks as output.

# Introduction

---

- The determination of a correct syntactic tree structure of an input sentence is crucial in Machine Translation.
- Phrase chunking or shallow parsing is a Natural Language Processing (NLP) task which divides a text into phrases, non-overlapping and syntactically correlated parts of words, based on linguistic properties and useful in several tasks
- Myanmar is a very productive inflectional, agglutinative and grammatical regularity lack language, it is less computerized and lack in computational resources.

## Cont'd

---

- The developed new phrase chunking algorithm accepts Myanmar tagged sentence as input and generates chunks as output.
- Input Myanmar sentence is split into chunks by using chunk markers such as postpositions, particles and conjunction and define the type of chunks.
- The algorithm was evaluated with POS tagged Myanmar sentences based on three measure parameters.

# Motivations

---

- To be successful at any NLP applications of Myanmar language, some amount of functional analysis is necessary to be feasible grammatical relations.
- In the previous Burmese Phrase Segmentation system used the CRF++ tool to identify phrase boundaries with markers such as သည့်<sub>+</sub> ,က ,က<sup>o</sup> <sub>l</sub> and tested on four types of corpora.
- However, it cannot solve phrases with zero markers and cannot also differentiate homonyms.

# Proposed Myanmar Phrase Chunking Algorithm

---

- Input: POS tagged Myanmar Sentence
- Output: Myanmar Chunks with Chunk Type
- Begin
- Convert POS tagged Myanmar sentence into POS and word string arrays
- While( $i < \text{length of POS string array}$ )
  - Read POS[i] of word [i]
  - If POS[i] is chunk marker
    - String array for word is segments before and after word[i]
  - Else
- Increment i
- End While
- Generate Myanmar chunks with corresponding type of chunks for input sentence
- End

# Myanmar Language Grammar

---

- Grammar, the study of rules behind the languages, which does not concern the meaning directly is called the syntax of that language.
- Myanmar can be defined as postpositional language because of its use of postposition (Wi Bat),
- Myanmar Language Syntax: Subject Object Verb- SOV
- English is prepositional language, (syntax: Subject Verb Object- SVO) because of its use of preposition.



# Example Myanmar Sentence and Constituent Phrases

Sentence	<p style="text-align: center;">ဒေ ငှဲ အ ဘ ငှဲ ဆဲ နဲ ဝဲ စ ဝဲ ဟဲ ကညဲ သညဲ                  ပျဲ ငှဲ မဲ ဝဲ ခဲ ဝဲ မဲ ဝဲ ရဲ ဝဲ နဲ ဝဲ ဟဲ ဝဲ ဟဲ ဝဲ ကဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ                  ရဲ ရဲ ဝဲ ဝဲ သညဲ</p>							
English meaning	<p style="text-align: center;">Daw Aung San Su Kyi obtained Peace Noble Prize in 1991.</p>							
Phrase	Noun Phrase		Verb Phrase					
			Noun Phrase		Noun Phrase		Verb Phrase	
Word	ဒေ ငှဲ အ ဘ ငှဲ ဆဲ ဝဲ ဝဲ စ ဝဲ ဟဲ ကညဲ	သညဲ ဝဲ	ပျဲ ငှဲ မဲ ဝဲ ဝဲ ခဲ ဝဲ မဲ ဝဲ ရဲ ဝဲ နဲ ဝဲ ဝဲ ဟဲ ဝဲ	ကဲ ဝဲ ဝဲ	ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ ဝဲ	ဝဲ ဝဲ	ရဲ ရဲ ဝဲ ဝဲ ဝဲ ဝဲ	သညဲ ဝဲ

# Noun Chunk Markers of Myanmar Language

Myanmar Phrase Chunk Marker	English Meaning	Example
ကတံ ဘ ဝ ဝိဘတံ	Nominative	သညံ၊ က၊ မံ ဘ
ကံ ဝ ဝိဘတံ	Objective	ကံ
ကြ ကံ ဘ ရံ ဝ ဝိဘတံ	Departure	မံ၊ က
ရေဝံ ဝ ဝိဘတံ	Direction	သံ
ဆံ ကံ ဝ ဝိဘတံ	Arrival	သံ
အသံ ဝ ဝိဘတံ	Accusation	ပျံ ဝ ဝိဘတံ
အေ ဝ ဝိဘတံ	Reason	ပေ ဝ ဝိဘတံ
လကံ ဝ ဝိဘတံ	Acceptance	အံ ဝ
အခံ ဝ ဝိဘတံ	Time	မံ ဝ၊ ကြံ ဝ၊ ဝ ဝိဘတံ
နေ ရံ ဝ ဝိဘတံ	Place	မံ ဝ၊ ကြံ ဝ၊ ဝ ဝိဘတံ
လံ ကံ ဝ ဝိဘတံ	Agreement	အလံ ကံ၊ အရံ

# Verb Chunk Marker of Myanmar Language

---

Myanmar Phrase Chunk Marker	English Meaning	Example
နက် <sup>၀</sup> ယာဝ <sup>၀</sup> ဝိဘတ <sub>+</sub>	Verb Postposition	သည <sub>+</sub>

# Adjective Chunk Marker of Myanmar Language

Myanmar Phrase Chunk Marker	English Meaning	Example
နှစ်က မဝင် ဝိသေသနာ ပါးစပ် ချိတ် ပါးစပ် ချိတ် ဝိသေသနာ ပါးစပ် ချိတ်	Adjective Phrase Change Particle	ဝိသေသနာ ပါးစပ် ချိတ် မပါးစပ် ချိတ်

# Types of Chunks Identified by the Algorithm

---

No.	Chunk Type	Description
1.	NC	Noun Chunk
2.	VC	Verb Chunk
3.	AdjC	Adjective Chunk
4.	AdvC	Adverb Chunk
5.	CC	Conjunction Chunk

# Results and Discussion

---

- The algorithm is tested with six types of Myanmar sentences
- The algorithm uses the POS tags of chunk markers instead of words.
  - it can solve the homonyms.
  - It can not segment the chunks correctly for the sentences with no makers and sentences containing more than one chunk markers continuously.
- The algorithm is tested with 1000 Myanmar POS tagged sentences and results are shown with three measures:  
Precision, Recall and F-measure

## Cont'd

---

$$\textit{Precision} = \frac{\text{No. of correctly identified chunks}}{\text{No. of identified chunks}} * 100\%$$

$$\textit{Recall} = \frac{\text{No. of correctly identified chunks}}{\text{No. of total chunks in testing sentences}} * 100\%$$

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

# Calculation of Precision, Recall and F-measure

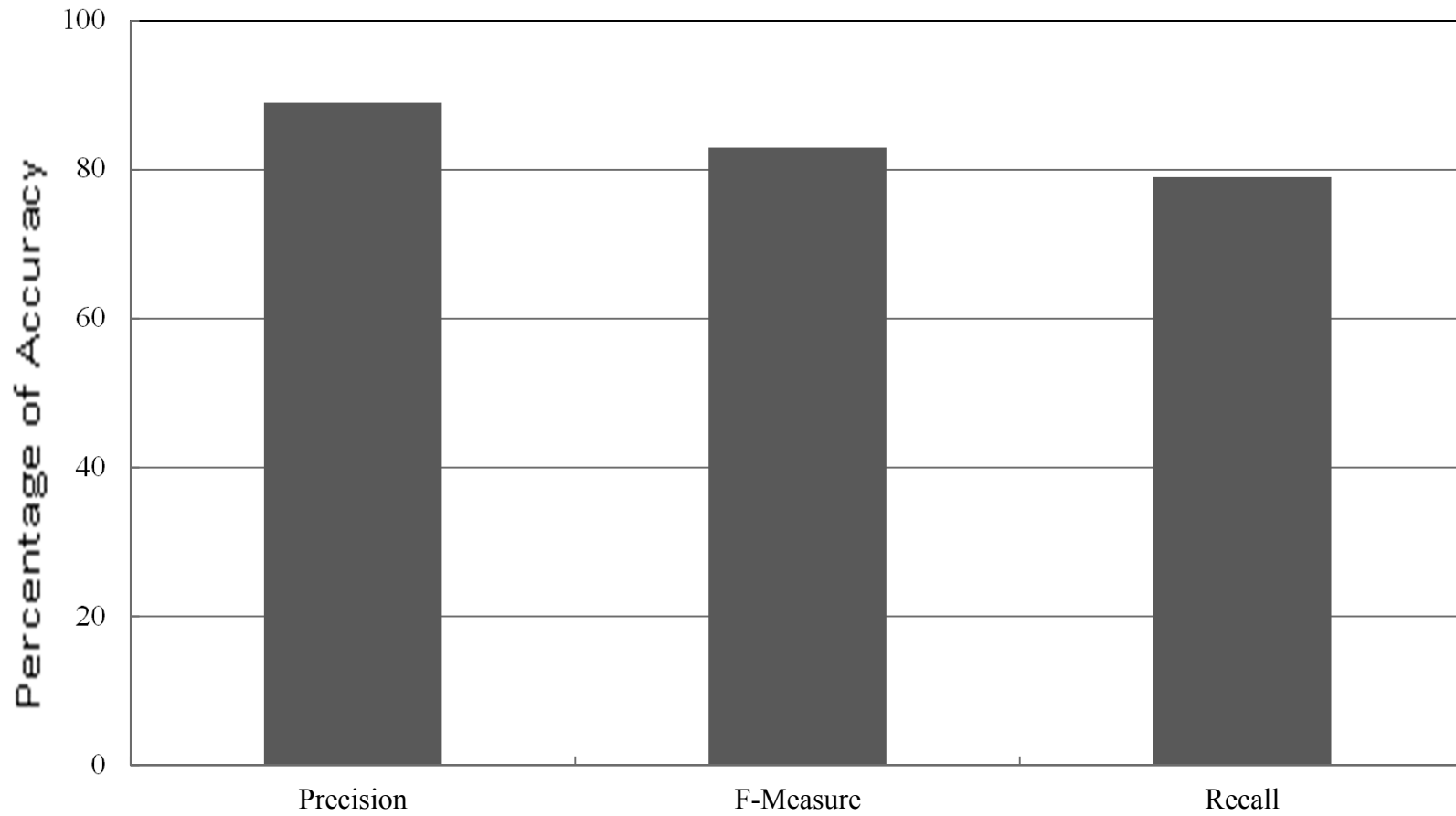
---

Correctly identified chunks	Identified chunks	Total chunks in testing sentences	Precision	Recall	F-measure
2357	2667	2981	89%	79%	83%



# Comparison of percent accuracy for three measures

---



## Conclusion

---

- Phrase chunking is one of the promising solutions for the ambiguities and produce at certain level of grammatical information.
- The new Myanmar phrase chunking algorithm is implemented. It segments the Myanmar sentence into chunks by using the postpositions, particles and conjunctions.
- The algorithm is tested with POS tagged Myanmar sentences and results are obtained 89% 79% and 84% in Precision, Recall and F-measure respectively.
- This new algorithm is effective and feasible for Myanmar Natural Language Processing applications.

## References

- [1] F. Oliveira, et al, “Systematic Noun Phrase Chunking by Parsing Constraint Synchronous Grammar in Application of Portuguese Chinese Machine Translation”, the 6<sup>th</sup> International Conference on Information Technology and Applications, ICITA, 2009.
- [2] E. F. T. K. Sang, S. Buchholz, “Introduction to the ConLL-2000 Shard Task: Chunking”, Proceedings of CoNLL-2000 and LLL-2000, pages 127-132, Lisbon, Portugal, 2000.
- [3] W. W. Thant et al, “Function Tagging for Myanmar Language”, International Journal of Computer Applications, Volume 26, No2, July 2011, pp 34-41.
- [4] M. P. Aung, K. T. Lynn, “Constuction of Finite State Machine for Myanmar Noun Phrase”, Proceedings of MJIT\_JUC Joint International Symposium 2013, Hiratsuka, Japan, 2013.
- [5] M. T. Win et al, “Burmese Phrase Segmentation”, Proceedings of Conference on Human Language Technology for Development, Alexandria, Egypt, May 2011, pp 27-33.

**Thank You!**