Text Summarization System Using Myanmar Verb Frame Resources

May Thu Naing
Information Techology Supporting and Maintainence Department,
University of Computer Studies, Taunggyi, Myanmar
E-mail: maythu.tgi@gmail.com

Aye Thida
Faculty of Computer Science, Artificial Intelligence Lab
University of Computer Studies, Mandalay, Myanmar
E-mail: ayethida.royal@gmail.com

Abstract—In today's era, when the size of information and data is increasing exponentially, there is an upcoming need to create a concise version of the information available. Until now, humans have tried to create "summaries" of the documents. Especially, Myanmar Natural Language Processing does not have computerized text summarization. Therefore, this paper presents a summary generation system that will accept a single document as input in Myanmar. In addition, this work presents analysis on the influence of the semantic roles in summary generation. The proposed text summarization system involves three steps: first, the sentences are parsed using Part of Speech tagger with Myanmar Language Tool Knowledge Resource (ML2KR); secondly, pronouns in the original text are resolved using Myanmar Pronoun Resolution Algorithm (MPAR); thirdly, the sentences are labeled with semantic roles using Myanmar Verb Frame Resource (MVF), finally, extraction of the sentences containing specific semantic roles for the most relevant entities in text. After that, the system abstracts the important information in fewer words from extraction summary from single documents.

Index Terms—Text summarization, Myanmar Language Tool Knowledge Resource, Pronoun resolution, Semantic roles, Myanmar Verb Frame Resource, Summary generation system.

I. INTRODUCTION

Text summarization is a hard problem of Natural Language Processing because, to do it properly, one has to really understand the point of a text. This requires semantic analysis, discourse processing, and inferential interpretation (grouping of the content using world knowledge). Document summarization aims to automatically create a concise representation of a given document that delivers the main topic of the document. Automatic document summarization has drawn much attention for a long time because it becomes more and more important in many text applications. The following are the important reasons in support of automatic text summarization:

- A summary or abstract saves reading time.
- A summary or an abstract facilitate document selection and literature searches.
- It improves document indexing efficiency.
- Machine generated summary is free from bias.
- Customized summaries can be useful in question-answering systems where they provide personalized information.
- The use of automatic or semi-automatic summarization by commercial abstract services may allow scaling the number of published texts they can evaluate.

Input to a summarization process can be one or more text documents. When only one document is the input, it is called single document text summarization and when the input is a group of related text documents, it is called multi-document summarization. The underlying assumption is that the topic-related documents can provide more knowledge and clues for single summarization of the specified document. From human's perception, users would better understand a document if they read more topic-related [1].

Generally, there are two approaches to automatic summarization: extraction and abstraction. The first approach in creating summaries (most common) is based on identifying important words in texts by using their frequencies, and determining those sentences that contain a bigger number of important words. These sentences are extracted from the original text, and taken to constitute the summary. In this paradigm, the summarization is performed through sentence extraction: the summary is a subset of the sentences in the original text. In contrast, abstractive methods build an internal semantic representation and then use natural language generation techniques to create a summary that is closer to what a

human might generate.

II. Literature Review

Automatic Text Summarization is an important and challenging area of NLP. The task of a text summarizer is to produce a synopsis of any document or a set of documents submitted to it. The early work in summarization dealt with single document summarization [1] where systems produced a summary of one document, whether a news story, scientific article, broadcast show or lecture.

A. Ashari, M. Riasetiawan [20] show how to implement document summarizing system uses TextRank algorithms and Semantic Networks and Corpus Statistics. The quality of the summaries influenced by the style of writing, the selection of words and symbols in the document, as well as the length of the summary output of the system.

The natural language processing community has recently experienced a growth of interest in semantic roles, since they describe WHO did WHAT to WHOM, WHEN, WHERE, WHY, HOW etc. for a given situation, and contribute to the construction of meaning. If for text analysis, semantic roles have gained their way into natural language analysis systems they are rarely used at their full potential for text generation. Christopherson [2] was among the first to investigate the usefulness of semantic roles in summaries. More recently, Suanmali et al. [3] used semantic roles and WordNet to compute the semantic similarity of two sentences in order to decide if the sentences are to be kept or not in the summary.

Dana Tranadabat [4] proposed text summarization method which is combining semantic roles and named entity for sentence extraction. Summarization task was initiated with the thought in mind of getting a summary of the document which will2 not be based on extraction of informative sentences from the document, but the idea of generation of sentences. To create summaries, we should be able to represent the document in a model from which it's possible to generate sentences. Also, we needed the semantics to come into play, while creating the summaries. So, the idea of generation of sentences comes from compressing a given sentence. This is a sentence or avoiding adverbs [5].

S. Narayanan, S. Harabagiu [6] show the first to stress the importance of semantic roles in answering complex questions. Their system identified predicate argument structures by merging semantic role information from PropBank and FrameNet. Expected answers are extracted by performing probabilistic inference over the predicate argument structures in conjunction with a domain specific topic model. The Berkeley FrameNet database consists of frame-semantic descriptions of more than 7000 English lexical items, together with example sentences annotated with semantic roles [7]. PropBank is a bank of propositions. A "proposition" is on the basic structure of a sentence [8], and is a set of relationships between nouns and verbs, without tense, negation, aspect and modal modifiers. Arguments which belong to propositions are annotated by PropBank with numbered role labels (Arg0 to Arg5) and modifiers are annotated with specific ArgM (Argument Modifiers) role labels. Each verb occurrence in the corpus receives also a sense number, which corresponds to a roleset in the frame file of such verb. A frame file may present several rolesets, depending on how many senses the verb may assume. In the roleset, the numbered arguments are "translated" into verb specific role descriptions. Arg0 of the verb "sell", for example, is described as "seller". Thus, human annotators may easily identify the arguments and assign them the appropriate role label. There is currently no frame semantic representation of Myanmar.

III. Nature of Myanmar Verb

A verb is a word used to express an action or state in English. A verb is one class of lexical units whose characteristic syntactic role is as a predicate or predicator and which is characteristically that of words denoting actions or processes. A verb is a word that expresses action or defines a state of being. The verb is the most essential element of a sentence and the only part of speech that can stand alone. The form (spelling) of a verb changes to indicate certain properties, such as mood, tense, and voice, that reflect how it is used in a sentence. A verb also shares the characteristics of person and number with the noun or phrase that functions as its subject. In Myanmar, a word that stands for doing, being or having is called a verb.

Example: သွား (go)၊ လာ(come)၊ စား (eat) ၊ သောက် (drink) ၊ ပျောက် (lose)

A. Kind of Myanmar Verb

There are three kinds of meaning in Myanmar Verb:

❖ Verb to do (ပြုခြင်းပြ ကြိယာ)
❖ Verb to be (ဖြစ်ခြင်းပြ ကြိယာ)
❖ Verb to have (ရှိခြင်းပြ ကြိယာ)

And the following list is three kinds of structure in Myanmar verb.

❖ Simple Verb (ပင်ကိုယ်ကြိယာ)
❖ Compound Verb (ပေါင်းစပ်ကြိယာ)
❖ Attribute Verb (ဂုဏ်ရည်ပြ ကြိယာ)

B. Verbs to Do of Myanmar

A verb that denotes doing or action is called a 'verb to do' in Myanmar.
Example:

ရေး (write), ဖတ် (read), စား (eat), ပြော (talk)

C. Verbs to Be of Myanmar

A state of being connected with the animate or the inanimate refers to a verb to be.

Example: ဖြစ် (သည်) (be), ပြို (သည်) (collap), ပျော် (သည်) (be happy)

D. Verbs to Have of Myanmar

A verb which refers to the state of having or location of a person, a place or a thing is called a 'verb to have' in Myanmar.

Example:

ရှိ (သည်) (have), တည်ရှိ (သည်) (locate)

E. Original Verbs of Myanmar

Original verbs of Myanmar are verbs which cannot be divided and formed grammatically and they indicate the state of doing, being or having.

Example:

စား (eat), သွား (go), ကျက် (be cooked)

F. Compound Verbs of Myanmar

Compound verbs of English are verbs affixed by a verbal postposition on the formation of two words.

Example:

နှုတ်ခွန်းဆက်သသည် (greet)

G. Attribute Verbs of Myanmar

Attributive verbs of Myanmar are formed by attributive words which are followed by verbal postpositions.

Example: ရဲရင့်သည် (be brave), မှောင်မိုက်၏ (be dark), အေးသည် (be cold)

IV. MYANMAR VERB FRAME RESOURCE

Myanmar Language has not delimiter or word boundary. Therefore, tokenization and word segmentation is important for further Myanmar NLP application. The tokenization in Myanmar 3 algorithm and segmentation algorithm using lexicon is reliable and useful resource for NLP application. Part-Of-Speech Tagging is very important for Myanmar syntactic analysis. Pronominal anaphora resolution algorithm for Myanmar can take the free words order and grammatical role in pronoun resolution in Myanmar. The role of subject and object in Myanmar are found to have significant impact on anaphora resolution for reflexive and possessive pronouns.

Myanmar verb frame files built together with example sentences annotated with semantic roles following PropBank guidelines [9]. But, this system could not reproduce the same experience of PropBank. This system interested in designing Myanmar Verb Frame files in relatively independent modules to facilitate the collaborative construction of this resource. Once PropBank guidelines and PropBank frames files are available for consultation, it is design to adopt a different approach: instead of firstly building frames files and Annotator´s Guidelines, Myanmar Verb Frame is start by annotating a corpus using English frames files and guidelines as model. Therefore, unlike PropBank, in this first phase it annotated only semantic role labels and not verb senses. In this way, the difficulties of the task were experienced, identified language-specific aspects of SRL for Myanmar language, and generated a corpus that used as base to build frames files.

A. Frame File

Each Myanmar verb frame file has included:

(a) Description of the verb:

In the description, the information is given; *name of the Myanmar verb*, *name of the English verb*; and *its sense id* is given according to the number of senses a verb has in Propbank, example sentence of the verb with semantic roles and the verb frame.

(b) Verb Frame

The frames are based on simple present tense indicates habitual acts taking it as default. Some Myanmar verb have the same English verb. To construct 1100 Myanmar verb frame files, 750 English verb frame files from Prop3bank was used. For example, "သိတင်းသုံး" is used for Monks in Myanmar. "ဓနေ" is used for normal people. But the meaning of these two Myanmar verb is (stay). Therefore, we develop all different Myanmar verb frames for the same English verb.

```
သွား
EnglishRel: go
RolesetId: go.01
Arg0: ,null
Arg1: entity-in-motion,PREP_NOM
Arg2: instrument,
PREP_ACCURATION(PREP_REASON)
Arg3: start point,PREP_DEPATURE
Arg4: end point,PREP_ARRIVAL
Example: [မောင်ဖြိုးဝေလင်းသည်]-[Arg1]/[အိမ်မှ]-[Arg3]/
[ကျောင်းသို့]-[Arg4]/ [ဆိုင်ကယ်ဖြင့်]-[Arg2]/ [သွားသည်]-[Rel]#
```

Fig.1. Example of Myanmar Verb Frame File

(c) Example sentence:

As this example shows, the arguments of the verbs are labeled as numbered arguments: Arg0, Arg1, Arg2 and so on.

B. Core Arguments of Frame File

Table 1 shows the core argument list using in Myanmar Verb Frame. Frame files provide verb-specific description of all possible semantic roles, as well as illustrate these roles by examples. The Arg0 label is assigned to arguments which are understood as agents, causers, or experiencers.

Table 1. List of Core Arguments in Myanmar Verb Frame

Tag	Description
Arg0	Agent(usually the subject of a transitive verb)
Arg1	Patient(usually its direct object or the subject of a intransitive verb)
Arg2	instrument, benefactive, attribute
Arg3	starting point, benefactive, attribute
Arg4	ending point

Arg0 arguments (which correspond to external arguments) are the subjects of transitive verbs and a class of intransitive verbs called unergatives.

John (*Arg0*) sang the song.
John (*Arg0*) sang.

Semantically, external arguments Proto-Agent properties, such as

- volitional involvement in the event or state
- causing an event or change of state in another participant
- movement relative to the position of another participant

The Arg1 label is usually assigned to the patient argument, i.e. the argument which undergoes the change of state or is being affected by the action. Internal arguments (labeled as Arg1) are the objects of transitive verbs and the subjects of intransitive verbs called unaccusatives:

John broke *the window* (*Arg1*) .
The window (Arg1) broke.

These arguments have Proto-Patient properties, which means that these arguments

- undergo change of state.
- are causally affected by another participant .
- are stationary relative to movement of another participant.

Every sentence does not include Arg2 , Arg3 and Arg4. They depend on the meaning of the sentence. If an argument satisfies two roles, the highest ranked argument label should be selected, where **Arg0 >> Arg1 >> Arg2>>… .**

C. Modifier Arguments of Frame File

Table 2 shows the types of modifier arguments in Myanmar Verb Frame Resources.

Locative (Argm-loc)
Locative modiers indicate where some action takes place.

မောင်လင်းလင်းသည်မန္တလေးမြို့တွင်နေထိုင်သည်။
Mg Lin Lin lives in Mandaly.
Arg0: မောင်လင်းလင်းသည် Mg Lin Lin
Rel: နေထိုင်သည် lives
Argm-loc: မန္တလေးမြို့တွင် in Mandalay

Temporal (Argm-tmp)
Temporal Argms show when an action took place, such as `in 1987', `last Wednesday', `soon' or `immediately'.

[မောင်မျိုးလှသည်]-[Arg1]/[ရန်ကုန်မြို့တွင်]-[Arg2]/
[ယယ်စဉ်ကတည်းက]-[Argm-tmp]/ [နေသည်။]-[Rel]#
မောင်အောင်ထက်သည် ယခုနှစ်တွင် အောင်မြင်လာသော အဆိုတော်တစ်ယောက်ဖြစ်သည်။
Mg Aung Htet is a successful singer this year.
Arg1: မောင်အောင်ထက်သည် Mg Aung Htet
Rel: ဖြစ်သည် is
Arg2: အောင်မြင်လာသောအဆိုတော်တစ်ယောက် a successful singer
Argm-tmp: ယခုနှစ်တွင် this year.

Manner (Argm-mnr)
Manner adverbs specify how an action is performed. For example, `works well' is a manner.

မုတ်သုန်လေသည်ပြင်းထန်စွာတိုက်ခတ်နေသည်။
Monsoon is heavily blowing.
Arg1: မုတ်သုန်လေသည် Monsoon
Rel: တိုက်ခတ်နေသည် is blowing
Argm-mnr: ပြင်းထန်စွာ heavily

Direction (Argm-dir)
Directional modifers show motion along some path. 'Source' modiers are also included in this category.

မောင်လှမျိုးသည်ရန်ကုန်မှပြန်လာသည်။
Mg Hla Myo come back from Yangon.
Arg1: မောင်လှမျိုးသည် Mg Hla Myo
Rel: ပြန်လာသည် come back
Argm-mnr: ရန်ကုန်မှ from Yangon.

Table 2. List of Modifier Arguments in Myanmar Verb Frame

Tag	Description	Example
Argm-loc	Locative	The museum
Argm-tmp	Temporal	Now, by next summer
Argm-mnr	Manner	Heavily, clearly, at a rapid rate
Argm-dir	Direction	To market

V. DOMAIN SPECIFIC IN MYANMAR SUMMARIZER

The domain text of proposed summarization system are documents which are about human achievements, the extremes of the natural world, events and items so strange and unusual that readers might question the claims. These texts can be gotten from many books and Myanmar websites such as "Treasure Layout Magazine", "Mingalar Mg Mal Issue" and "Thu Ta Sw Sone Magazine". They are monthly magazines in Myanmar. The Treasure Layout Magazine and Mingalar Mg Mal issue are intended for children to give knowledge about education, religion, health and many sections. We use texts from the title "The World's outstanding people" in Treasure Layout Magazine. This author of this title is Mg Kae Tun. For this magazine, he wrote two or three texts for this title every month from 2012 to 2015. We also use texts from "The Miracle World" that is written by "Aung Hein Htet" and "The Rich Knowledge for Children" that is written by "Min Win" in Mingalar Mg Mal Issue. In addition, we use many texts from Thu Ta Sw Sone Magazine and Pyi Myanmar Journal. A lot of news of unusually things and people are described this Magazine and journals. They are written by many writers. However, Myanmar writers translate news from "Ripley believe it or not" and "Guinness: World Records" books in English to Myanmar. The following Table 3 describes list of news for using proposed Myanmar Text Summarization System.

Table 3. Categories of Input Text

Magazine	Numbers of Text	Title	Writer	Total Sentences
Treasure Layout Magazine	32	The World's outstanding people	Mg Kae Tun	235
Mingalar Mg Mal	19	The Miracle World, The Rich Knowledge for Children	Aung Hein Htet, Min Win	196
Thu Ta Sw Sone Magazine	24	Many Titles	Many writers	133

VI. PROPOSED MYANMAR TEXT SUMMARIZATION SYSTEM

In this section, we discusses about our proposed Myanmar Text Summarization System. The overall architecture of proposed system is presented in Figure 2. The input domain text of this system explained the section III.

A. Word Segmentation and Part of Speech Tagging

For preprocessing of proposed system, word segmentation is the first stage. Without a word segmentation solution, no NLP application (such as Part-of-Speech (POS) tagging and translation) can be developed. Words can be combined to form phrases, clauses and sentences. Thus, in proposed system, word segmentation is performed with Myanmar Word Segmenter [10]. This segmenter generated all possible segmented sentence of phase pattern as their score for further processing. It can handle the unkown case of word and it does not only depend on lexicon. For the next step of the preprocessing stage which is Part of Speech (POS) Tagging, rule based POS tagging of Myanmar language. [10] is used. This tagging used the context-free grammer (CFG) as rules which parsing is start with sentence and left to right parsing structure to define the POS of each word.

B. Pronominal Anaphora Resolving

In order to identify the semantic role a specific entity express, the pronoun must be first identified in the input text. This is the task of pronoun anaphora resolution. For the next step for our summary generation system, this system uses resolving method for anaphoric references in POS tagging sentences. A rule-based system creates an anaphoric link between the pronoun and its antecedent based on Hobbs algorithm. This system applies Myanmar Pronominal Anaphora Resolution Algorithm (MPAR) [11] to resolve pronoun in input text.

C. Semantic Role Labeling

For the next step, we perform semantic role labeling on pronominal resolving sentences. In labeling system of Myanmar semantic roles, semantic role is indicated by a particular syntactic position (e.g. object of a particular preposition). The preposition in sentences are very important to label semantic roles. Because semantic role resource for this system is Myanmar verb frame resource that is following the Preposition Bank Guidelines. The following is how to relate semantic roles and prepositions in sentences.

- ❖ Agent: subject (PREP_NOM)
- ❖ Patient: direct object (PREP_OBJ)
- ❖ Instrument: object of "with/by" (PREP_ACC)
- ❖ Source: object of "from" (PREP_DEPT)
- ❖ Destination: object of "to" (PREP_ARRIVAL)

For semantic role labeling, predicate argument identification algorithm and mapping arguments with semantic roles algorithm [12] was applied in this system.

D. Generation of Summary

The final step of the summary generation system implies two kinds of summary. The first one is extractive summary which is sentence selecting, among the list of sentences from which summaries can be generated, the ones in which the entity has core semantic roles. The second one is using the combination rules on the sentences of extractive summary.

E. Generation of Extractive Summary

There have four main stages:

1. Identifying the main character (most frequent Noun).
2. Selection of sentences containing main roles of main characters.
3. Generation of extractive summary.
4. Generation of abstractive summary.

F. Generation of Abstractive Summary

The second step is "extractive" to "abstractive" step in which the extracted information will be mentally sorted into a pre-established format and will be "edited" using heuristics techniques.

The editing of the raw material ranges from minor to major operations. [13] describes the rules for abstracting and states that redundancy; repetition and circumlocutions are to be avoided. And it gives a list of linguistic expressions that can be safely removed from extracted sentences or re-expressed in order to gain conciseness. These include expressions such as it was concluded that X, to be replaced by X, and it appears that, to be replaced by Apparently. Also, some transformations in the source material are allowed, such as concatenation, truncation, phrase deletion, voice transformation, paraphrase, division and word deletion. It mentions the inclusion of the lead or topical sentence and the use of active voice and advocates conciseness. But in face, the issue of editing in text summarization has usually been neglected, notable exceptions being the works by [14].

F. Moawad, and M. Aref [15] presented to create an abstractive summary for a single document using rich semantic graph reducing technique. This approach summaries the input document by creating a rich semantic graph for the original document, reducing the generated graph and then generating the abstractive summary from the reduced graph. For reducing semantic graph, they used heuristics rules. By following their concept, [16] proposes an idea to create a semantic graph for the original document and relate it semantically and by using several rules reduce the graph and generate the summary from reduced graph.
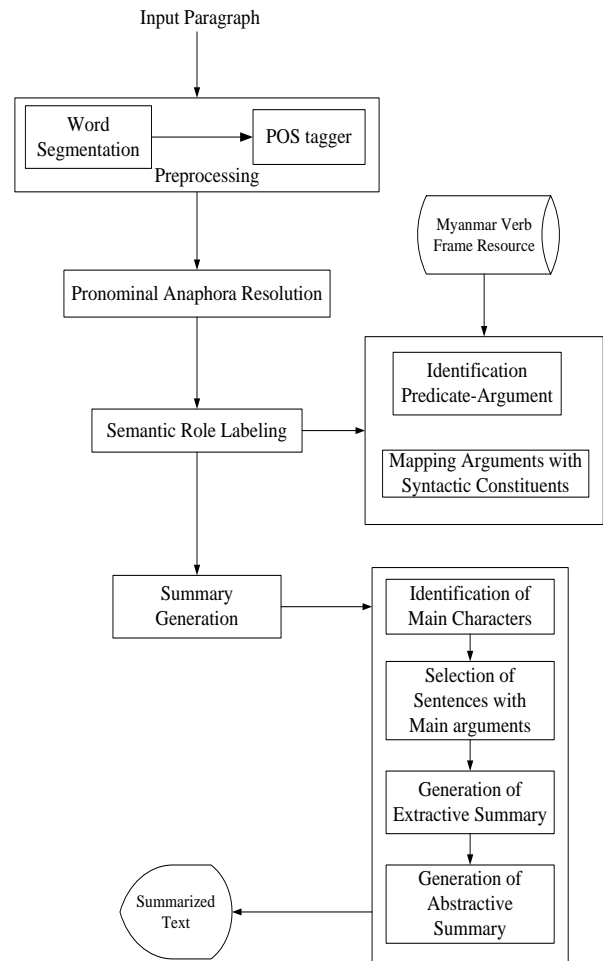

Fig.2. Architecture of Proposed System

Therefore, we use the concept of reducing heuristics rules for semantic graph and cut-and-paste approach to addressing the text generation problem in single-document summarization. This approach goes beyond simple extraction, to the level of simulating the revision operations to edit the extracted sentences. [17] proposed cut-and-paste approach for abstractive summarization. It has two revision operations: sentence reduction and sentence combination. Since this approach generates summaries by extracting and combining sentences and phrases from the original text, they call it the cut-and-paste approach. While extraction-based approaches mostly operate at the sentence level, and occasionally at the documents or clause level, the cut-and-paste approach often involves extracting and combining phrases. This cut-and-paste approach addresses only the text generation problem in summarization; it does not address the document understanding problem in summarization.

For reducing the extractive sentence, a set of heuristic rules are applied on the part of speech structures to reduce it by merging, deleting, or combining the sentences etc. Figure 3 presents summarization algorithm that can be applied on the POS tag of two simple sentences:

Sentence1= [SubN1, ObjN1, Mverb1]
Sentence2= [SubN2, ObjN2, Mverb2]

Each sentence is composed of three tags: Subject (SubN1), Object (ObjN) and Main Verb (Mverb). With the help of such rules, the summarized text can get beyond the extractive summarization.

Begin

    Input: Extracted sentences from Original Document.
    Output: Final[]
    Initalization:
    Arg[]={},New_arg={},New_argList[]={},i=0

Step 1:   For each extracted sentences
        Add subjects in every sentences to
        ListSubject[];
    End for
Step 2:   For each ListSubject[] do
        If (ListSubject[i] is Equal to
        ListSubject[i+1]) Then
          Replace connective word to the end of
          the first sentence.
          Remove subject and conjuction words
          from the second sentence.
          Final[]+=Merge the first sentence and
        the second sentence.

          i=i+1;
        Else
          Final[]+=Sentence of ListSubject[i].
        End if
    End for
End

Fig.3. Summarization Algorithm

VII. Evaluation of Summarization System

Evaluating summaries and automatic text summarization systems is not a straightforward process. What exactly makes a summary beneficial is an elusive property.

A. Compression Ratio (CR)

Generally speaking there are at least two properties of the summary that must be measured when evaluating summaries and summarization systems: the Compression Ratio (how much shorter is than the original) [18]:

$$CR = \frac{lengthofSummary}{lengthofFullText} \qquad (1)$$

Table 4. Compression Ratio in Summaries

Total Documents	Total Syllable	Total Sentences	Total Syllable in Summary	Total Sentences in Summary	Compression Ratio
75	14074	564	8635	269	61%

B. Precision and Recall

The common information retrieval metrics of precision and recall can be used to evaluate a new summary [19]. A person is asked to select sentences that seem to best convey the meaning of the text to be summarized and then the sentences selected automatically by a system are evaluated against the human selections. Recall is the fraction of sentences chosen by the person that were also correctly identified by the system.

$$\mathrm{Re}\,call = \frac{|\, system - humanchoiceoverlap\,|}{|\, sentencesc\,hosenbyhuman\,|} \qquad (2)$$

Precision is the fraction of system sentences that were correct

$$\mathrm{Pr}\,ecision = \frac{|system - humanchoiceoverlap|}{|sentencesc\,hosenbysy\,stem|} \qquad (3)$$

Table 5. Precision and Recall in Summaries

Total Sentences system-human choice overlap in Summary	185
Total Sentences Chosen by System in Summary	325
Total Sentences Chosen by Human in Summary	221
Precision	83%
Recall	57%

VIII. Conclusion

Pronominal Anaphora Resolution Algorithm, Semantic Role Labeling System and Myanmar Verb Frame Resource do not exist in Myanmar language before. This research is the first progress of a Myanmar language resource system for Myanmar NLP applications. Semantic and pragmatic problem solved in the NLP system, they do not seem to be easy to solve in the near future, because its research fields are wide, so expensive and longtime plan to investigate the semantic. Current Myanmar Verb Frame Resource does not cover all verbs

in Myanmar lexicons because resource is used just only two lexicons: Myanmar-English Computational lexicon and Lexique Pro lexicon. Therefore, Myanmar Verb Frame Resource needs to add more Myanmar verbs like phrasal verbs of English.

This paper presented how Myanmar text summarization system with semantic roles in detail. The importance of pronominal resolution and semantic role in text summarization is disccussed. Moreover, the extractive summarization and abstractive summarization are explained. The results of summarization system for 75 documents is about 61 percent. Their precision and recall is 83 % and 57% by comparing humman summary and system summary. Therefore, by performing text summarization system consider main semantic role for sentences selection and combination sentences, the system produce more meaningful summaries. The system can identify target verb of any sentences (spoken and written style sentences), the performance of summarization system are more accurate. Although this system use combination rules in summarization algorithm, these rules does not cover all Myanmar sentences types. If summarization rules are added to the summarization algorithm, summarized texts will be more meaningful.

Therefore, if other different types of text like factual and scientific papers or news contains reference pronouns or frequent nouns, our summarization system seems to perform well on these texts. This research has been done in the motivation to produce meaningful summaries, which has been exemplified by summarization results. In the future work, more several case studies using several documents with different sizes, and hence assess the results of this summarization system.

REFERENCES

[1] X. Wan, J. Yang, and J. Xiao, " *Single Document Summarization with Document Expansion*", Proceedings of AAAI2007, 2007.

[2] Christopherson, and L. Steven, "*Effects of Knowledge of Semantic Roles on Summarizing Written Pose"*, Contemporary Educational Psychology, Vol 6, No 1, Jan 1981, p. 59-65.

[3] L. Suanmali, N. Salim, and M.S. Binwahlan, "*SRL-GSM : A Hybrid Approach based on Semantic Role Labeling and General Statistic Method for Text Summarization*" Journal of Applied Sciences, 2010.

[4] D. Trandabăţ , "*Using semantic roles to improve summaries"*, Proceedings of the 13th European Workshop on Natural Language Generation ENLG2011, Nancy, France, 2011, p. 164-169.

[5] H. Jing, "*Sentence Reduction for Automatic Text Summarization*", Proceedings of the 6th Applied Natural Language Processing Conference, Seattle, Washington, USA, 2000, p. 310–-315.

[6] S. Narayanan, S. Harabagiu, "*Question Answering based on Semantic Structures*", In Proceedings of the 19th COLING, 2004, pp. 184–191.

[7] C.F. Baker, C.J. Fillmore, and J.B. Lowe, "*The Berkeley FrameNet Project*", In Proceedings of the COLING-ACL, Montreal, Canada, 1998.

[8] M. Palmer, D. Gildea, and P. Kingsbury. "*The Proposition Bank: An annotated corpus of semantic role*", Association for Computational Linguistics, 2005.

[9] M.T. Naing and A. Thida, "Myanmar Proposition Bank: Verb Frame Resource and An Annotated Corpus of Semantic Roles", in Proceedings of ASEAN Community Knowledge Networks for the Economy, Society, Culture, and Environmental Stability, May 2014, pp. 42.

[10] S.L Phue, "*Development of Myanmar Language Lexico-Conceptual Knowledge Resources*", PhD Thesis, Univesity of Computer Studies, Mandalay, December 2012.

[11] M.T. Naing and A. Thida, "*Pronominal Anaphora Resolution Algorithm in Myanmar Text*", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) [Volume 3, Issue 8], August 2014, pp. 2795-2800.

[12] M.T. Naing and A. Thida, "*A Sematic Role Labeling Approach in Myanmar Text*", in Proceedings of the ICGEC, International Conference on Genetic and Evolutionary Computing (Yangon, Myanmar) August 26,27,28, 2015.

[13] H. Saggion and G. Lapalme, "*Generating Indicative-informative Summaries with SumUM,*" Computational Linguistics, , 2002, Vol. 28, p. 497-526.

[14] H. Jing, "*Sentence Reduction for Automatic Text Summarization*", Proceedings of the 6th Applied Natural Language Processing Conference, Seattle, Washington, USA, 2000, p. 310–-315.

[15] F. Moawad, and M. Aref, "*Semantic Graph Reduction Approach for Abstractive Text Summarization*", Seventh International Conference in Computer Engineering & Systems (ICCES), 2012.

[16] S.S. Govilkar, "*Conceptual Framework for Abstracive Text Summarization*", International Journal on Natural Languae Computing (IJNLC), Vol. 4, No.1, February 2015.

[17] H. Jing and K. McKeown, "*Cut and Paste Based Text Summarization*", Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics, Seattle, Washington, USA, 2000, p. 178–-185.

[18] M. Hassel, "Summaries and the Process of Summarization from Evaluation of Automatic Text Summarization – A practical implementation", Licentiate Thesis, KTH NADA.

[19] A. Nenkova and K. McKeown, "Automatic Summarization", Foundations and Trends® in Information Retrieval: Vol. 5: No. 2–3, 2011, pp 103-233.

[20] A. Ashari, M. Riasetiawan, "Document Summarization using TextRank and Semantic Network", International Journal Intelligent Systems and Applications, Nov 2017, pp 26-33 (http://www.mecs-press.org/).

Authors' Profiles

May Thu Naing
She is an Assistant Lecturer of Information Techology Supporting and Maintainence Department at University of Computer Studies, Taunggyi, Myanmar. She is a member of Natural Language Processing Project. Her research interests include Data Mining and Natural Language Processing. She is currently working Myanmar to English, Paoah, Shan Translation System Project in Taunggyi, Myanmar. She received B.C.Sc(Hons) degree from Computer University (Mandalay) in 2008, M.C.Sc degree in 2010 and her Ph.D degree in Computer Science from the University of Computer

Studies, Mandalay(UCSM), Myanmar in 2016.

Aye Thida
Aye Thida is a Professor, at Faculty of Computer Science, Artificial Intelligence Lab), University of Computer Studies, Mandalay (UCSM), Myanmar. Her research interests include in Machine Translation, Text to Speech System and Big Data Management. She is currently working NLP researches. She received B.Sc(Hons:), Maths degree from the Mandalay University, Myanmar and her M.I.Sc and Ph.D degrees in Computer Science from the University of Computer Studies, Yangon(UCSY), Myanmar.