An Improved Semantic Role Labeling for Myanmar Text

Zin Mar Kyu*, Naw Lay Wah

University of Computer Studies, Mandalay, Patheingyi, Mandalay 1001, Myanmar

ARTICLE INFO

ABSTRACT

The semantic role labeling plays a key role to extract semantic information from language text for the purpose of information retrieval, text summarization, plagiarism test, etc. This paper presents two main parts: MynNet and semantic role labeling. MynNet, which is a FrameNet for Myanmar language is proposed to extract the semantic relations and various kinds of lexical meanings of each individual word. Moreover, it is exploited to know how each of them depends on the others in what kind of semantic ways. The major job of semantic role labeling is to parse the arguments of a sentence from small words into combination of text chucks in the form of suitable arguments to be able to express the semantic meaning of that sentence. In labeling, different forms of segmented text phrases are considered, and it is affordable to categorize different relationship types of segmented Myanmar Text phrases such as core, non-core, etc. The proposed system is trained on roughly 70,000 words with 3000 hand-annotated semantic information for general domain. This system is tested with untrained domains with different kinds of sentences in order to test efficiency of our labeling ability for semantic roles. According to the experiment results, we achieve better results than other semantic role labelers of Myanmar Text.

1. INTRODUCTION

Linguists have always been trying to revolute the techniques of Natural Language Processing (NLP) in order to promote the automatic understanding of languages by the machines. In today's era, Internet revolution creates us to deal with the beauties of language such as various lexical usage, intricate semantic structures and meaning daily used by internet users as well as timely published by the articles or papers on the Internet.

Due to easily accessible resources on the Internet, it evolves many problems, that is, the users are copying the resources of others without asking the permission, the Internet cannot understand what the users want or search due to their incomplete or complex inputs, etc.

The abovementioned problems cannot solely be solved in the means of syntax learning in which only bases on lexical expressions such as (subject, verb). To understand and deduce the meaning of the lexical structures, the semantic learning becomes popular which considers agent, goal, recipient, etc. which play in states of affairs of a sentence or paragraph.

Apart from this, NLP has become an intrigued research area especially in semantic analysis [1]. To extract exact meaning of language text, semantic analysis becomes a hot topic in analyzing syntax and semantic roles of text in NLP. Semantic Role Labeling (SRL) is the process of identifying and labeling semantic roles of predicates such as noun, cause, purpose, etc. [2].

Semantic role labeling has been widely used in text summarization, classification, information extraction and similarity detection such as plagiarism detection, etc. Gildea and Jurafsky [3] have proposed a first SRL system developed with FrameNet corpus and targeted to English Text. Later, many researches have been developing for different languages such as Indian [1], Vietnamese [2], Swedish [4], etc.

In SRL labeling, we can see it into two perspectives such as technical advancement of SRL regardless of languages; that is; just for international language, English; and SRL for non-English language such as Thai, Myanmar, Chinese, etc. In the aspects of first perspective, there are some popular labeling techniques in SRL process such as statistical classifiers [1] and dependency parser [5], etc.

Although they have their own achievements, it still has gaps to fill with more advanced SRL techniques for more accurate SRL results. For second perspective for specific language, some languages have their own structures. They need specific text corpus such as Chinese [6], Myanmar while others are able to adapt and extend English based corpus such as Swedish [4].

Due to abovementioned challenges encountered in both perspectives, we propose an improved SRL process based on Level-wise Segmented Parser (LSP) for labeling semantic roles of predicates in a sentence. Meanwhile, this system develops a specific corpus for Myanmar Text called MynNet primarily based on English FrameNet design and structures [5,7]. SRL techniques developed for English language could not achieve expected accuracy result for Myanmar Text. Therefore, we develop a complete semantic relationships and dependencies for Myanmar Text and name it as MynNet which has structured according to English FrameNet 1.5 [5].

In this paper, we propose a design of FrameNet structure for verb, adjective and adverb part of speeches of Myanmar Text.

*Corresponding author. Email: zinmarkyu.ucsm@gmail.com

Our purpose of developing FrameNet design for Myanmar Text is to assist other linguistic processes for Myanmar Text such as similarity calculation between user inputs and pre-stored Myanmar text especially Web, plagiarism detection for copyright protection and semantic role labeling, etc.

In order to prove that the effectiveness of MynNet and proposed semantic role labeling, we perform a broad experiment with different evaluation metrices, which are famous in evaluation both the accuracies and performance of the system. They are evaluated with different experiment parameters such as trained or untrained domain, simple or complex sentences, and different kinds of sentences such as affirmative, negative and question, etc. The results show that our system achieves satisfactory results.

The rest of paper is organized as follows. The related literature is reviewed in Section 2. The underlying theories are preliminarily studied in Section 3 before going to present detailed explanation of proposed system and Myanmar FrameNet structure is discussed in Section 4. Semantic Role Labeling for Myanmar Text is discussed in Section 5. The experimental results are discussed in Section 6, and the paper is concluded in Section 7 by giving a glimpse of future work.

2. RELATED WORK

Li et al. [6] proposed a sentence similarity theory based on Chinese Dependency Graph using Chinese FrameNet. They computed three types of similarity of sentences namely word similarity, dependency graph similarity and external component similarity.

In a study, Pham et al. developed a semantic role labeling system based on candidate syntactic constituent extraction for Vietnamese text. They use their own ProBank corpus and tested their Labeling results with *F*-measure, one of well-known information retrieval methods [2].

Johansson et al. presented automatic extraction of semantic roles with the aid of Swedish FrameNet called SweFN. It is augmented on English version of FrameNet developed by Berkeley research group. Their SweFN frames and names are directly derived from English ones with some exceptions such as internal relations, frame elements selection and definition, etc. They used cross-frame generalization and cluster-based features methods to make their role label classifiers more robust [4].

Nomani presented a statistical system for identifying semantic roles and their relationships for two major Indian languages, Hindi and Urdu. They identify arguments pertaining to a given verb from an input sentence; they classified those arguments into semantic labels such as DOER, THEME, LOCATIVE, etc. using statistical classifiers [1].

Kshirsagar et al. [8] augmented standard FrameNet model with the derivation of features from FrameNet and ProBank and compares their annotated exemplars. They proved that FrameNet increases 4% in *F*-measure.

Chang et al. [9] presented a novel application of alternating structure optimization for semantic role labeling of noun predicates

in Chinese NomBank. In developing FrameNet for SRL process, Japanese [10], Spanish [11], Italian [12], Bulgarian [13], Thai [14], Chinese [6] and German [15] FrameNet have been developed and using for NLP purpose.

Ning et al. [16] proposed a method in extracting many syntactic features subjectively for semantic role labelling system. In order to improve SRL performance, they extract more effective features from subjective ones using genetic algorithm. Their algorithm optimizes syntactic features and help improving semantic labelling for Chinese text.

Semantic role labelling techniques for Chinese text are also proposed by Wang. This paper proposed a new method for SRL with two main subtasks called clustering and labeling. Clustering is used to replace syntactic parsing, during which similar sentences are clustered together. In the labeling step, artificial neural networks are planted as many as the number of clusters, each of which takes charge of summing up the features of chunks of a sentence and then labeling them with semantic roles [17].

Paul et al. proposed new methods for SLR and text classification for plagiarism detection for English language. They mainly focus on how to reduce the time of checking in plagiarism and time of labelling in roles for sentence level text [18].

Sentence reframing techniques using SRL is proposed by Sethi et al. [20]. Their technique is used for Hindi sentences by mainly using NLP techniques.

The semantic role labeling for Myanmar Text has been developed by Naing and Thida by proposing a method of assigning semantic roles on the structured trees of Myanmar Text using Myanmar Verb Frame. They used predicate-argument identification algorithm for Labeling. They used ProBank for Myanmar Text corpus. Their work could only perform for simple sentences and their ProBank need to be scaled up for more complete words Labeling [19].

Regards to language structure and design, FrameNet and ProBank are popular. The key feature of FrameNet is to derive generalizations about frame structure and grammatical organizations using corpus, and the describing the valances of target words using frame semantics [19].

FrameNet is a popular and mature tool in NLP for automatic language understander for English, Chinese, etc. Kshirsagar et al. [8] arguments standard FrameNet model with the derivation of features from FrameNet and ProBank and compares their annotated exemplars. They proved that FrameNet increases 4% in *F*-measure.

Although FrameNet is a mature approach for some languages such as English, Chinese, etc., Myanmar language still lacks of consideration of semantic structure with the aid of FrameNet. To address some limitations of them with the effectively exploration of semantic frames of FrameNet, we introduce FrameNet for Myanmar Text for semantic role Labeling of Myanmar Text.

3. PRELIMINARY STUDY

The underlying theories used in this paper are briefly described in this section in the aspects of FrameNet and SRL process.

3.1. FrameNet and Frame Semantics

FrameNet (FN) is a repository of lexical and semantical expressions of language. It describes predicational words such as verb, noun, adjective, etc. [21]. The standard FrameNet is intended for English language and the purpose is to analyze conceptual structure of linguistic meaning [7]. It works on semantic frames that defines dependencies between words and describes its roles among participants of a language [4].

The key feature of FrameNet is to derive generalizations about frame structures and grammatical organizations and describing of valances of target words using frame semantics [21]. FrameNet is modeled for conceptual knowledge extraction from subject matters that can be events, scenario, user inputs or articles from the Internet, etc.

The basic units of frame semantics are frames and its elements are called frame element (FEs) [5,22]. Each FE definitely comes from one frame and it may be associated with zero or more other frame elements. In association, there may be different kinds of association types such as inheritance, monotonic, subframes, etc. In addition, each frame can act different roles (core or non-core) depending on subject matters.

The major product of FrameNet is FrameNet lexical database, released by Berkeley FrameNet project, which contains more than 13,000 lexical units (LUs), fully annotated with 1000 hierarchically related semantic frames. There are more than 200,000 annotated sentences in corresponding semantic frames [5].

The lexical unit is word that describes the meaning of a word [5,22]. Typically, each sense of a word might have different semantic frames, depending on their concepts or combination with other words. In this case, semantic frames act like script structure for those words to describe a particular type of objects in some specific condition.

Generally, FrameNet annotations make the frame element realize the semantic meanings of the object contained in sentence, each consisting of a frame element (for example, the soldiers), a function organized grammatically (for example, object) and a phrase type such as noun, pronoun, etc. We can learn three types of annotation on each frame element which are systematically tagged. They are FE descriptions, frame-to-frame relations, and lexical entries summarizing the valence patterns for each lexical unit.

3.2. Semantic Role Labeling

Semantic role labeling, one of the NLP techniques, is a kind of semantic parsing the partial text of human languages such as English, Chinese, Myanmar, etc. Its purpose is to identify and labels semantic arguments of predicates in a sentence. SRL is conceived of as identifying the chunks of sentences with labels to indicate various relevant syntactic and semantic properties of words of a sentence.

Semantic annotation boosts language processing steps to be faster with more accurate results in the area of NLP. By labeling the arguments, NLP process can gain crucial information about the relations of particles/events of a language sentence [22]. Its major work is to specify semantic arguments associated with a verb contained in a sentence and other arguments into different roles such as noun, adjective, etc.

For clear understanding, let we give a sentence as an example for SRL process, e.g., "The soldiers scarify their lives for their countries." The verb of this sentence is "scarify". Depending on this verb, SRL labels the other part of speech as "The soldiers (Noun) (Who)" and "their lives (Noun) (What)" and "to their countries (Noun) (Whom)", etc. Intuitively according to the example, it determines not only subject and object like does in Parts of Speech (POS) tagging, it also determines the whole semantic roles of a sentence. It depends on the delineation of matters such as "who did", "what" to "whom" "how" to "when" and "where" with "which purpose", etc.

Therefore, it defines the sematic roles of each term based on semantic relationship between their predicates and terms based on the semantic relationship of individual words contained in a sentence [18].

Semantic role labeling plays a key role in automatic text-based processes such as question answering (Q&A), machine translation, text classification and information retrieval, etc. [18].

Although semantic role labeling has solved some obstacles of NLP in understanding human languages, it still has challenges for better role labelling with accurate results.

4. MYNNET DESIGN FOR MYANMAR TEXT

In FrameNet is modeled for conceptual knowledge extraction from the subject matter that can be events, scenarios or user inputs or articles from the Inter net. The basic units of frame semantics are frames and its elements named as FEs 5.5 Each FE definitely comes from one frame and it may be associated with zero or more other fame elements. In association, there may be different kinds of association such as inheritance, monotonic, subframes, etc. In addition, each frame can act different roles (core or non-core) depending on subject matters.

In this paper, we name our FrameNet for Myanmar text as MynNet. In our MynNet design, there will be one record in the database for each Myanmar word for each part of speech: verb, noun, adjective and adverb (eg. အပွင့်ပ‌ဝေးးသည်၌၊ လ‌ဝုဘသဝေတာ၊ လချွင်မပြန်စဝွတ, etc.). The reason why we construct MynNet for those four POSs is that they are the key parts of any language to make the language to be beautiful as well as complex with different semantic and syntax structures and expressions.

ProBank users mostly consider semantic annotations only based on verb without consideration of other POSs. According to the research of Myanmar Text and also our experiences, we strongly believe that not only verb can hold more semantic relations than other POSs. Therefore, our MynNet design considers all semantic dependencies of four most import POS. We called it as annotation table in this paper. In this case, we also consider lexical units for synonyms.

After structuring the annotation table, we will arrange the syntactic realization table for associated FEs (clauses) that might include/organize as one event or scenario. We organize all these kinds of syntactic realization is developed for each word so that the lexical expression, the semantic terms and relations and also their usual occurrences could be learnt in our MynNet design without using

any special knowledge extraction from semantic relationships like ProBank.

The detailed explanation of MynNet for verb, noun, adjective and adverb are given in the following section with examples. Our MynNet design is structured based on FrameNet 1.5.

In MynNet design, we consider only verb frame because a verb of a sentence is capable of identifying roles of other phrases included in that sentence. Regards to semantic relationships for each verb FE, we use annotated semantic information as well as their lexical units called synonyms as shown in Table 1.

Table 1 describes core-types, relations and relationship types of text element of an example word "အ�note:ပcေဝးသည် (give standing ovation)". This frame consists of various FEs (place, purpose, etc.) and some elements are denoted as core that plays necessary role of that frame while the others are non-core, peripheral, etc. depending on their significances regarding main verb.

Similarly, another example for a verb FrameNet is given in Table 2. It consists of corresponding frame elements which are syntactically or semantically related to the verb "donate". Using this table, it will be capable of understanding the semantic roles of the arguments in a sentence which contains a verb called "donate".

The core frame roughly corresponds to non-cores in an eventuality such as peripheral and extra-thematic FEs. Peripheral FEs refer to place, time, manner, means, which are semantically related with core FE whereas extra-thematic stands for describing causal connection between frames without conceptually belong to core FE [5,7,21].

4.1. Frame Relationships

Frames are grouped into semantic domains depending on their specific tasks such as Communication, motion, society, etc. [5,7,22]. When two frames are related, their frames elements can relate each other in unpredictable way, which is more than one relation between the frame elements.

Table 1 | Annotation table for a verb "Give standing ovation"

Frame element	Core type
Give standing ovation (အနote:ပcေဝးသည်)	Core
Means (Manner)	Peripheral
Where	Non-core
Who [Agent]	Core
To whom [Agent]	Core
Place	Peripheral
Purpose	Extra-thematic
Time	Peripheral

Table 2 | Annotation table for a verb "donate"

Frame element	Core type
I	Donor
Now	Time
The money	Theme
Charity	Recipient
Donate	Predicate

Each pair of FEs relates each other with exactly one relationship type, which may be directly or indirectly. In order to offer better understanding, we would like to briefly discuss what kinds of relationship type are used in our design and how they perform.

4.1.1. Inheritance

This relationship defines as full inheritance, with the possibility of multiple inheritances. This means that if Frame B derives from Frame A, there will be exactly one FE relationship corresponding to each FE in A and FE in B there will different name from that in E. Please see the example in end of this sub-section.

4.1.2. Monotonic

This relation performs semantic relationship type between parent and child with exact one way. That is a frame from parent A will exactly have in its child.

4.1.3. Subframe

The subframe relationship contain complex frame and explore indirect relationship among FEs. In subframe relation, many frames express concepts about natural, well-defined frames, such that, many complex events can be broken down into a series of smaller events, by defining their particular roles.

4.2. Syntactic Realization

The syntactic realization table for associated FEs (clauses) that might include/organize as one event or scenario is arranged to understand all possible combination of phrases, which will be organized with that main verb.

We organize all these kinds of syntactic realization for each verb word. The sample syntactic pattern is shown in Table 3. It helps to understand SRL process for lexical expressions, semantic terms and relations, and also their usual occurrences.

For a main verb, "အနote:ပcေဝးသည် (give standing ovation) [VP]" shown in Table 1, syntactic realization is analyzed in order to understand all syntax structures that can potentially organize for that main verb. Due to limited space of the paper, we just describe

Table 3 | Syntactic realization for "give standing ovations"

Number of annotations	Patterns
1	{Agent[NP,Subject]}
2	{Recipient[NP,Object]}
3	{Agent[NP,Subject],Recipient[NP,Object]}
4	{Agent[NP,Subject],Recipient[NP,Object], Means[Adv]}
5	{Agent[NP,Subject],Recipient[NP,Object], Means[Adv],Time[PP(Time), NP]}

NP, noun phrase; Sub, subject; Obj, object; PP, preposition; VP, verb phrase; Adv, adverb; PP, preposition.

only four patterns for that verb although each verb frame might have all possible forms of patterns depending on all of their possible actions.

Another syntactic realization structure is illustrated in Table 4. This time is for a verb "go". Depending on its possible syntactic structure for each different domain, different verb can create different number of syntactic structures as shown in Tables 3 and 4.

The overall design of MynNet is depicted in Figure 1. It shows connected table frames and their attributes as structured in MynNet. According to the diagram, each frame has its related frames with corresponding relation types. The frame elements (FEs) will derive the frame entity and will have lexical units and POS types depending on their syntax.

The design consists mainly of annotating sentences selected from a domain as examples of a particular lexical unit. It uses the table

Table 4 | Syntactic realization for "go"

| Number of annotations	Patterns
1	{Agent[NP,Subject]}
2	{Agent[NP,Subject], Direction[NP,Place]}
3	{Agent[NP,Subject], Source[NP,Place], Direction[NP,Place]}
4	{Agent[NP,Subject], Source[NP,Place], Direction[NP,Place], Instrument[NP,Object]}
5	{Agent[NP,Subject], Source[NP,Place], Direction[NP,Place], Instrument[NP,Object]}, Time[NP,Time]}
6	{Agent[NP,Subject],Recipient[NP,Object], Means[Adv],Time[PP(Time), NP], Purpose[PP(Purpose),VP,NP]}
7	{Agent[NP,Subject],Recipient[NP,Object], Means[Adv],Time[PP(Time),NP], Place[PP(Place),NP],Purpose [PP(Purpose),VP,NP]}
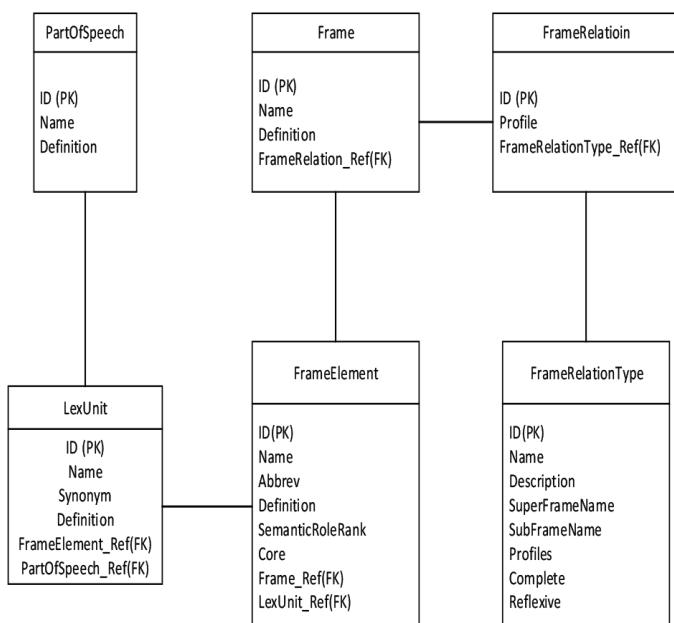
Figure 1 | Myanmar FrameNet corpus. PK, primary key; FK, foreign key.

to store the sentences and annotation on them that can reflect the structure of the concepts involved in the sentences.

In the following diagram, we embed the multiplicity and dependency types such as aggregation, composition, etc. for simplicity of diagram. The database is designated for particular domain and related sub domain. This database representation can create the semantic links between the sub-corpus while extracting the semantic meanings from the sentences.

There is a set of annotation layers for FEs, phrase types (POS), grammatical functions and lexical units in this database. Each set is presented with corresponding annotation tables to link to sentences, sub-corpus and LUs. This MynNet makes it possible for one sentence to be annotated more than once, with different kinds of words for different domains and annotation sets.

5. SEMANTIC ROLE LABELING FOR MYANMAR TEXT

Although many researchers have developed and implemented SRL improvement for English and other languages, ongoing development of SRL for Myanmar language challenges us to investigate more accurate sematic role labeling system without limitation of specific domain area.

We therefore introduce two main contributions in this paper. The first contribution is improving SRL process with detailed segmented Labeling parser. Another contribution is targeting to domain independent SRL process so as to release rigid annotations that can work well only in small ranges.

Apart from them, we are the first of introducing FrameNet for Myanmar Text by supplying required annotated information especially for Myanmar Text by basically learning the foundation structure of English FrameNet.

In this section, MynNet design for Myanmar text and role labeling process are explained in detail as follows.

5.1. Semantic Role Labeling

In Labeling semantic roles of predicates of a sentence, we mainly use LSP. It initially breaks down a sentence into individual words. In this case, English language puts the space between the words and each word can represent either a complete or partial meaning. However, in Myanmar Text, space is optional between words. Therefore, first of all, combination of words has to be found and translates their meaning just for individually independent meaning.

For example, a word "*COOK*" for English has many forms in Myanmar Text such as "ချက်ကင်သသည်" or "ချက်ကင်ပြုတ်တင်သသည်" or "ချက်ကင်ပြုတ်တင်ကဿာုဇဇာင်ဇင်လုဒုဇဿာင်ဇင်သသည်" etc. The work "cook" can be formed in different number of words in Myanmar text unlike one word in English.

They are just different usage form of general cooking word "*COOK*" but not like lexical units such as ပဝ္ဝကင်ပဝ္ဝကင်ဆဿဂုအဿဿာင်ဇင်တညင်သသည် (*simmer*), ကဇင်သသည် (*grill*), etc. Therefore, combining individual words to form a meaningful word is performed as the basic level by our LSP.

As another level, LSP tries to combine each word into chunks that can represent roles of a sentence such as agent, theme, purpose, cause, place, time, reason, purpose, etc. Each chunk can have different combination of words.

For example, agent can either be individual word such as "လူ့ (people)" or combination of words such as "လူမဝိၚးပၚေါ်တၚ်ၚ်ၚ် မၚျၚာၚးပၚြတၚးလၚ့သၚၚောၚလၚၚ့မၚျၚာၚး (A crowd of people on the street)". Similarity, other phrases such as place, purpose, etc. have their own corresponding combinations of words.

Figure 2 describes how SRL works in labelling Myanmar SRL process. To unlock the language barriers, the same sentence in English version is described in Figure 3 to give clear figures about the work of our SRL process. The NP phrases can be subject, object, place, time phrase, etc. We shortly write all those types of phrases as NP in Figure 3.

That sentence is level-wise labeled such that it labels each word at a time and it iteratively combines the corresponding words depending on semantic annotation of MynNet.

The variety of POS combination for a phrase is trained in MynNet such that Agent[Subject] can be composed in any possible combinations of phrases such as NP = N or NP = Adj + N or NP = Place + Adj + N, etc. Our LSP can accurately tag complex phrases into their corresponding labels.

It then finds syntactic realization structures from Table 2 so that certain syntactic patterns are explored which are believed (based on some prior knowledge obtained from trained domain corpus) to be important in current labelling sentence. It then compares against with all arrangements all possible combinations of phrases depending on targeted main verb.
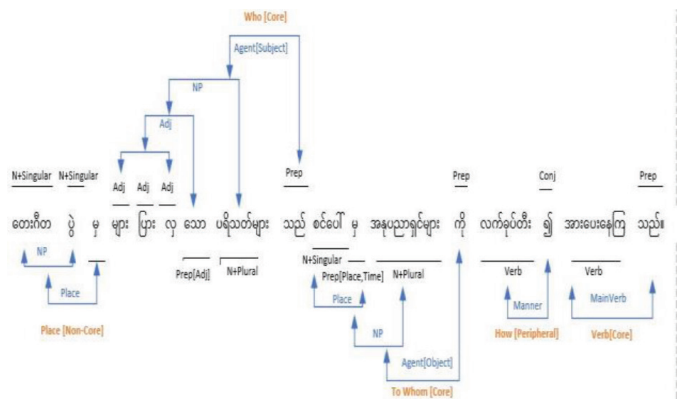

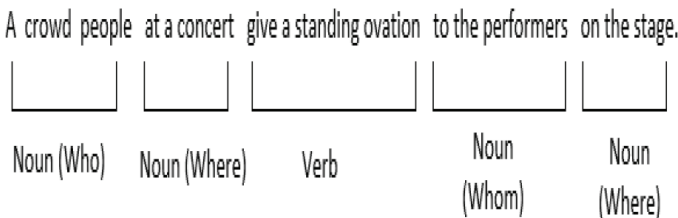**Figure 2** | SRL labelling example in Myanmar sentence.

Figure 3 | SRL labelling example in English sentence.

Most of SRL approaches [1–4,6,19] works only on Labeling without additionally consideration of their importance roles in a sentence. In our SRL process, it extensively labels the roles of predicates such as place, subject, etc. It afterward identifies their importance roles such as core, non-core, peripheral, etc. with relationship types defined by MynNet.

The final level of LSP identifies place, Agent[Subject], Recipient [Object], means and mainverb for the phrases found in an input sentence. It additionally tags each of them such as non-core, core, peripheral, etc. depending on their roles acted in that sentence. From the start of individual words of a sentence, LSP works level ups by organizing each word into meaningful words, phrases, etc. until a complete labelled sentence is obtained.

6. EXPERIMENTAL RESULTS

In this section shows experimental results and discusses their different results. In experimenting, we consider two different domains called trained and untrained domain. For trained domain preparation, we train MynNet with common words that usually use in general for all domains and their detailed semantic annotation and patterns. For untrained domain, we prepare MynNet with words included in such domains but just for POS without annotation of their semantic frames, types and relationships. We chose movie review and sport news for untrained domain.

To measure efficiency of our SRL process, we use three types of evaluation methods for all domain tests: namely precision, recall and F-measure.

Precision is to measure how many labelled words are correct from all labelled words done by SRL approaches as following Equation (1).

$$P = \frac{\text{Number of correct labels}}{\text{Number of all labels made by the system}} \qquad (1)$$

Recall is to measure how many words are correctly labelled from all words included in an input text file as Equation (2).

$$R = \frac{\text{Number of correct labels}}{\text{Number of all available labels}} \qquad (2)$$

F-Measure is calculated on the precision and recall results of discovered services within the range of values [0, 1] in which 1 means the best score and 0 means the worst score as following Equation (3).

$$F\text{-measure} = \frac{(2 * P * R)}{(P + R)} \qquad (3)$$

Figure 4 summarizes average values resulted by our LSP on all evaluation metrics. They are tested on roughly 1800 sentences for each domain. Each sentence has altered in at least five different forms of syntax and lexical units without changing their semantic meaning. According to the data in Figure 4, it clearly demonstrates that our LSP works better for trained domain than untrained domains.

In order to show efficiency of our LSP against with contemporary SRL approach for Myanmar Text developed by the study [19],

we perform another experiment for all evaluation metrics in the same routine.

In this time, we summarize all evaluation results and show overall average values in Figure 5. The results of Figure 5 show that our LSP outperforms other approach [19] due to our theory advancement of level-wise labeling process. Furthermore, semantic annotation and their possible patterns aid our LSP to significantly achieve better performance in all domains compared with them.
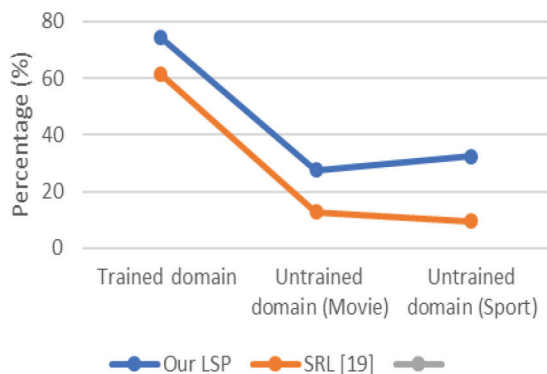
Apart from the abovementioned measures, we additionally test our proposed system with different kinds of sentences such as simple and complex sentences, etc. The complex sentences mean the sentence, which contains more than one sentence with the help of linking words. We use 3800 sentences for simple sentences 3400 sentences for complex sentences. All of these sentences contain variety of sentences types such as affirmatives, negatives and questions type sentences.

The remaining experiments are carried out to evaluate our system with different experimental parameters. The following Figure 6 shows the results of all measurements in each different sentence, simple and complex sentences. Intuitively, labeling the semantic roles for simple sentences gets far better results in all evaluation metrics than those for complex sentences.
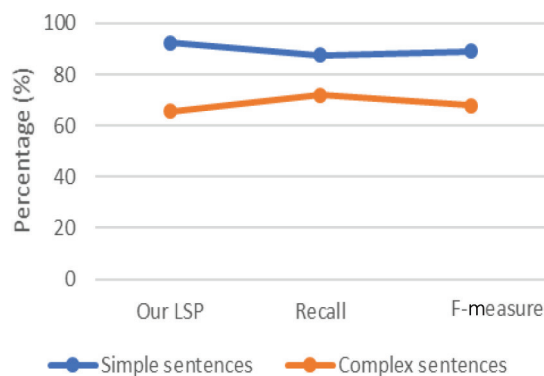
The next experiment is performed to evaluate which sentences can be correctly labeled as shown in Figure 7. According to the values tabularized in Figure 7, the higher results are achieved in affirmative
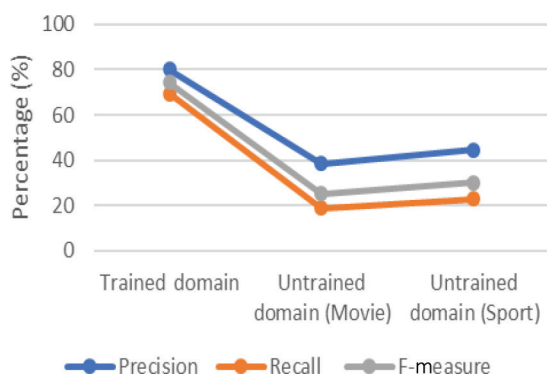
Figure 6 | Overall performance evaluation between simple and complex sentences.
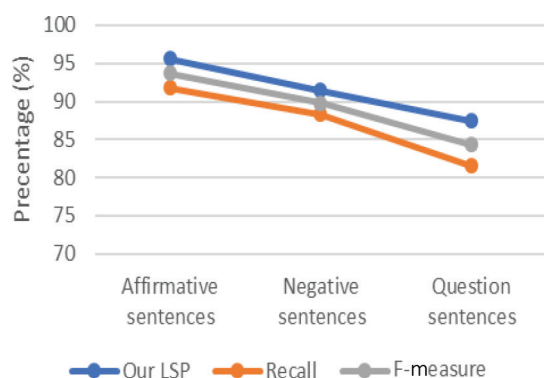
Figure 7 | Overall performance evaluation between affirmative, negative and complex sentences.

sentences for all kinds of measurements. The least value results in question sentences.

7. CONCLUSION

To conclude, this paper presents two main portions, MynNet design and proposed semantic role Labeling. We introduce Myanmar FrameNet structure to be able to extensively use for further NLP process such as plagiarism detection, semantic role Labeling, similarity matching between users' inputs and available web resources for Myanmar Text, etc. As standard FrameNet 5.1 except lemma and meta relationships, we consider all aspects of semantic roles of POS called verb, adverb, noun and adjective so that all semantic directions of Myanmar Text can be completely covered by this FrameNet. We leave lemma and meta data to add to our current MynNet as future work.

Regards to SRL, semantic role labeling is one of the solutions of understanding complex and subtle concepts of human languages. This paper has proposed an advanced level-wise semantic labeling approach for Myanmar Text. In addition, it has provided significant roles of phrases of a sentence with the help of frame element types and relationships. It can therefore be used for further NLP process such as information extraction, text summarization, plagiarism test, etc. In the future, we have plan to extend this work to achieve more accurate labeling results for all trained and untrained domains.


Figure 4 | Evaluation results of our LSP between trained and untrained domain.

Figure 5 | Overall performance evaluation between our approach and other approach.

CONFLICTS OF INTEREST

There is no conflict of interest.

REFERENCES

[1] M.A. Nomani, D.M. Sharma, Towards building semantic role labeler for Indian languages, Tenth edition of the Language Resources and Evaluation Conference (LREC 2016), Portorož, Slovenia, 2016.

[2] T-H. Pham, X-K. Pham, P. Le-Hong, Building a semantic role labelling system for Vietnamese, 2015 Tenth International Conference on Digital Information Management (ICDIM), IEEE, Jeju, South Korea, 2015, pp. 77–84.

[3] D. Gildea, D. Jurafsky, Automatic labeling of semantic roles, Comput. linguist. 28 (2002), 245–288.

[4] R. Johansson, K.F. Heppin, D. Kokkinakis, Semantic role labeling with the Swedish FrameNet, Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 3697–3700.

[5] J. Rupenhofer, M. Ellsworth, M.R.L. Petruck, C.R. Johnson, C.F. Baker, J. Scheffczyk, FrameNet II: Extended theory and practice, International Computer Science Institute, Berkeley, CA, USA, 2016.

[6] R. Li, S. Li, Z. Zhang, The semantic computing model of sentence similarity based on Chinese FrameNet, 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, IEEE, Milan, Italy, 2009, pp. 255–258.

[7] C.J. Fillmore, C.F. Baker, H. Sato, The FrameNet database and software tools, Proceedings of the Third International Conference on Language Resources and Evaluation (LREC), European Language Resources Association (ELRA), Valletta, Malta, 2002, pp. 1157–1160.

[8] M. Kshirsagar, S. Thomson, N. Schneider, J. Carbonell, N.A. Smith, C. Dyer, Frame-semantic role labeling with heterogeneous annotations, Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Beijing, China, 2015, pp. 218–224.

[9] C. Liu, H.T. Ng, Learning predictive structures for semantic role labeling of NomBank, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 208–215.

[10] K.H. Ohara, S. Fujii, H. Saito, S. Ishizaki, T. Ohori, R. Suzuki, The Japanese FrameNet project: A Preliminary Report, Proceedings of the Pacific Association for Computational Linguistics, Halifax, 2003, pp. 249–254.

[11] C. Subirats-Rüggeberg, M.R.L. Petruck, Surprise: Spanish FrameNet, Proceedings of Workshop on Frame Semantics XVII International Congress of Linguists, Vol. 17, CD-Rom Publication, Prague, Czech Republic, 2003, p. 188.

[12] A. Lenci, M. Johnson, G. Lapesa, Building an Italian FrameNet through semi-automatic corpus analysis, Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, 2010.

[13] S. Koeva, Lexicon and grammar in Bulgarian FrameNet, Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, 2010.

[14] D. Leenoi, S. Jumpathong, P. Porkaew, T. Supnithi, Thai FrameNet construction and tools, Int. J. Asian Lang. Proc. 21 (2011), 71–82.

[15] A. Burchardt, K. Erk, A. Frank, A. Kowalski, S. Padó, M. Pinkal, The SALSA corpus: A German corpus resource for lexical semantics, Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), Genoa, Italy, 2006, pp. 969–974.

[16] M. Ning, S. Yanqiu, L. Chunxia, Chinese semantic role labeling based on genetic algorithm, 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, IEEE, Macau, China, 2012.

[17] K. Wang, Automatic semantic role labeling for chinese, 2010 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, IEEE, Toronto, ON, Canada, 2010.

[18] M. Paul, S. Jamal, An improved SRL based plagiarism detection technique using sentence ranking, Procedia Comput. Sci. 46 (2015), 223–230.

[19] M.T. Naing, A. Thida, A Sematic role labeling approach in Myanmar Text, International Conference on Genetic and Evolutionary Computing, Springer, Cham, 2015, pp. 423–429.

[20] N. Sethi, P. Agrawal, V. Madaan, S.K. Singh, A novel approach to paraphrase Hindi sentences using natural language processing, Indian J. Sci. Technol. 9 (2016), 1–6.

[21] D. Trandabăţ, Natural language processing using semantic frames, PhD dissection, PhD Thesis, "Alexandru Ioan Cuza" University, Faculty of Computer Science, Iaşi, 2010.

[22] R.S. Swier, S. Stevenson, Unsupervised semantic role labelling, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics (ACL), 2004, pp. 95–102.