UNSUPERVISED DEPENDENCY CORPUS ANNOTATION FOR MYANMAR LANGUAGE

[‡]Hnin Thu Zar Aye, [‡]Win Pa Pa, [†]Ye Kyaw Thu*

[‡] Natural Language Processing Lab, University of Computer Studies, Yangon, Myanmar
[†]Okayama Prefectural University (OpU), Okayama, Japan
[*]Language and Speech Science Research Lab., Waseda University, Tokyo, Japan
{hninthuzaraye, winpapa}@ucsy.edu.mm, ye@c.oka-pu.ac.jp

ABSTRACT

Dependency parsing can provide the connection of linguistic unit (words) by a directed links. This paper presents annotating a general domain corpus by using unsupervised approach by applying Universal part-of-speech (U-POS) to build Treebank for unsupervised dependency parsing of Myanmar Language. Up to now it is still hard task to obtain complete syntactic structures for Myanmar Language. Dependency structures of words in Myanmar sentences are also presented of general words and phrases orders and the relations of basic sentence structures. To annotate by using U-POS, UDPipe is used. Moreover, the preliminary results of annotated trees and parsing experiment are presented. Parsing experiments are evaluated by UDPipe in terms of unlabeled and labeled attachment scores: (UAS) and (LAS), which are 93.20%, and 91.21% in test experiment respectively.

Index Terms— annotation, dependency, unsupervised

1. INTRODUCTION

Parsing natural sentences to analyze the sentence structure has been one of the primary research fields in Natural Language Processing (NLP). Constituent, phrase, structure is more widely used to represent sentence structure such as Penn Treebank. However, it can be cumbersome for free word-orders languages. As a result, dependency structure is started to focus for parsing. Dependency structure is simpler than a phrase one to represent syntactic and semantic information of language. In a dependency structure, each node represents a word-token. Dependency parsing parses an input sentence which is word-tokens sequence with part-of-speech (POS) tags to produce a rooted dependency tree in which nodes represent words and edges show syntactic relations between the words. The statistical approaches were achieved better quality compared to the human annotations rather than rule-based approaches.

Ten years ago, two CoNLL shared tasks [6] and [10] were important milestones in dependency parsing and provided 20 treebanks of different languages with the same format. These tasks are applied as the standards to measure the quality of dependency parsers. Dependencies structures have been successfully applied to NLP applications such as machine translation in recent years [3]. Moreover, there has been a big progress in the field of unsupervised dependency parsing in order to parse from raw sentence without existing annotated treebanks as the motivation. In the recent years, many researchers are working on a project called Universal Dependencies (UD) which is a collection of treebanks for many languages, where the morphological and dependency annotation styles are unified across the languages [5]. This becomes the way for low resourced languages to obtain syntactic structures even though quality of unsupervised parsers is not good as supervised ones.

For Myanmar language, there is no resource for sentence dependency information. It is still difficult to add dependency structures of Myanmar words by human annotation and a time and cost consumed task. According to this background conditions, as our motivation, we annotated my-POS[1] corpus which contains ten thousand sentences from Wikipedia for various domain area such as economic, history, news, politic to build first Myanmar dependency corpus by unsupervised approach [7].

The paper is organized as follows: Section 2 briefly presents about Myanmar language nature and status of resource. Section 3 describes mapping U-POS for Myanmar corpus language POS tags. Section 4 presents the words dependency structures which are used in unsupervised annotation by UD frame work. Section 5 presents about Myanmar dependency parsing. Section 6 discusses the experiments. Section 7 concludes with the discussion of the future work.

2. MYANMAR LANGUAGE

Myanmar is a head-final language and morphological rich and agglutinative. Basic sentence order is Subject-Object-Verb (SOV). Myanmar noun phrases can be put in any position in sentence because noun phrases are usually ended with different types of nominal markers, postpositional markers (PPMs) to identify their roles in sentences such as subject, objects or other statement phrases. However, verbs are always put at the end of the sentences and frequently suffixed with particles and PPM. One or more noun phrases can be a

[1] https://github.com/ye-kyaw-thu/myPOS

U-POS	General POS	POS of myPOS	Description (Example [Translation])
NOUN	N	*n*	Noun (ကျောင်း [school])
		tn	Text number (တစ် [one])
		fw	Foreign word (Hardware)
PROPN	N	*n*	Proper Noun (ရန်ကုန် [Yangon])
	ABB	*abb*	Abbreviation (ဂျီဒီပီ [GDP])
PRON	PRON	pron	Pronoun (သူ [he/she])
NUM	NUM	num	Number (၁ [1] , ၂၀၀[200])
ADJ	ADJ	adj	Adjective (နီ [red])
ADV	ADV	adv	Adverb (ပြန်လည်[again])
VERB	V	v	Verb (ကူညီ [help])
CCONJ	CONJ	conj	Coordinate Conjunction (နှင့် [and])
SCONJ	CONJ	conj	Subordinate Conjunction (သောအခါ [when])
PART	PART	*part*	Particle(များ[plural marker], ဝို့ [specified marker])
		part_neg	Negative Particle (မ[not])
ADP	PPM	ppm	Post-position Marker (သည်, ကို, အား[nominal markers], ၌, မှာ [in/at/]) called as "Adoposition" in Universal POS tags
PUNCT	PUNC	punc	Punctuation (။ [.], ၊ [,])
INTJ	INT	int	Interjection (အိုး [Oh])
SYM	SB	sb	Symbol (% , =)

Table 1. Mapping Universal POS and Language POS

syntactic role. Therefore, POS information is important to identify dependencies between words. Myanmar is still low-resource language to extract syntactic structures.

3. MAPPING UNIVERSAL POS TAGS

The UD[2] is a project to develop cross-linguistically consistent treebank annotation for many languages, with the goal of multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective. Related U-POS tags of corpus language POS tags are needed for unsupervised induction of dependency syntactic structure. In order to add U-POS tags, firstly original corpus POS tags are transformed to language general POS tags which are assumed as language POS tags in CoNLL-U format[3] because POS tags of myPOS are designed to tag easily some noun types (proper name, number digit and letter, foreign words) and negative particle "မ" which is affixed or suffixed with other particles to represent negative verb or adjective. Descriptions of U-POS, general POS, and myPOS tags are listed in Table 1.

4. DEPENDENCY STRUCTURES

All phrases depend on the main sentence root which might be verb or adjective. A source word, depends on another word, head in the dependency structure of a phrase according to the dependency relation between words. If sentences have clauses, roots of clause depend on the root of sentence. Zero or more clauses can be occurred in a sentence. In this paper, dependency structures of word in phrases which cover the basic structures of sentence, noun, adjective, conjunction, adverb, and verb are presented. Detail dependency head structures will be described.

Noun can be composed with one or more words in Myanmar language and is usually suffixed with PPMs or particles (PART) to define their roles to the verb or specify the meaning of those nouns. Common occurred noun phrases will be presented as below.

Some Myanmar nouns are composed of one or more words which might be noun, adjective, or verb to represent a syntactical role and called **compound nouns**. Compound noun's general form is as *N N…N.* In the example of Figure 1, "နေ" (sun) and "ချိန်" (time) modifies "ဝင်" (go down).

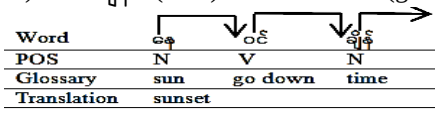


Figure 1. Example of Compound Noun

Proper noun is the unique name of common noun. Most Myanmar proper nouns are usually used with common noun. Proper noun phrases type is as the example of Figure 2.

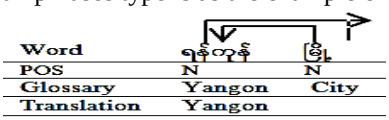
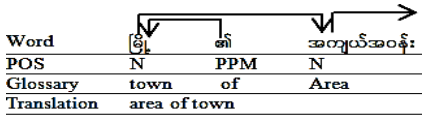


Figure 2. Example of Proper Noun

Possessive noun phrases are written with PPM, "၏" (successive of noun), and the general formula is *N ၏ N.* The

[2] http://universaldependencies.org/introduction.html
[3] http://universaldependencies.org/format.html

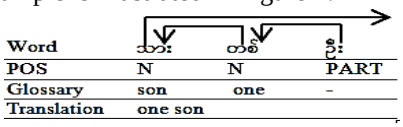
left *N* of "*၏*" is the owner of the right *N*, substance of left. The PPM-"၏" shows the possessive condition of that noun. The example is shown in Figure 3.



Word	မြို့	၏	အကျယ်အဝန်း
POS	N	PPM	N
Glossary	town	of	Area
Translation	area of town		

Figure 3. Example of Possessive Noun

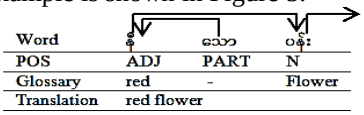
Myanmar **numeral nouns** are composed of main noun and number of counted amount and counting word which might be common noun of main or particle (PART) according to their types and shapes. The general order of numeral phrase is *N NUM/N N/PART* in which leftmost N can contain one or more noun, NUM/N is counting amount of leftmost and can be represented with digit number (NUM): ၁ (1) or digit letter(N): တစ် (one), and N or PART, numerical particle specifies the counting amount based on counting particles of noun types (e.g for normal person-" ဦး", for king-"ပါး") and shape of that noun (e.g for ball- "လုံး"), or common noun of counted noun (e.g for school- "ကျောင်း"). The varieties of numerical particles are usually used to specify counting amount. Example is illustrated in Figure 4.



Word	သား	တစ်	ဦး
POS	N	N	PART
Glossary	son	one	-
Translation	one son		

Figure 4. Example of Numeral Noun

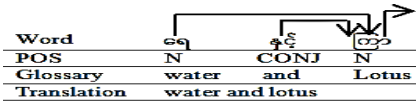
Like other languages, **adjective** modifies nouns. Myanmar *adjectives* can be before or after noun. The particle "သော" is commonly suffixed to the adjectives or root verbs which specify actions or conditions of that noun. Therefore, adjectives or root verbs depend on noun when they act as adjectives. Example is shown in Figure 5.



Word	နီ	သော	ပန်း
POS	ADJ	PART	N
Glossary	red	-	Flower
Translation	red flower		

Figure 5. Example of Adjective

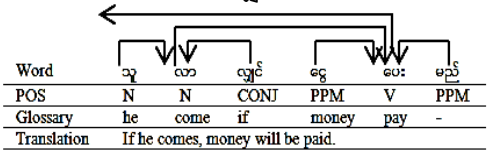
As the other languages, **conjunctions** are used to join phrases or sentences to join related phrases for coordinate meaning or sentences as subordinate case. **Coordinate** conjunctions are used to join the words or phrases or sentences. Sample dependency structure of coordinating with conjunction, "နှင့်", is shown in Figure 6.



Word	ရေ	နှင့်	ကြာ
POS	N	CONJ	N
Glossary	water	and	Lotus
Translation	water and lotus		

Figure 6. Example of Coordinate Case

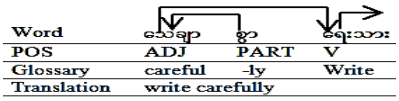
Subordinate conjunctions of Myanmar language are usually followed the verbs or adjectives of dependent sentences to join the main independent sentence to represent the full meaning. The sample dependency structure of subordinated case with conjunction, "လျှင်" is shown in Figure 7.



Word	သူ	လာ	လျှင်	ငွေ	ပေး	မည်
POS	N	N	CONJ	PPM	V	PPM
Glossary	he	come	if	money	pay	-
Translation	If he comes, money will be paid.					

Figure 7. Example of Subordinate Case

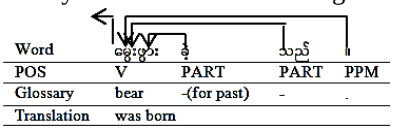
Adverb modifies verb action. Like noun, one or more words (adverb or adjective) can be used as adverb phrase. Adverb can be composed of root verb or adjective with suffixed particle, "စွာ". Dependency scheme of adverb is shown in Figure 8.



Word	ဆင်ခြင်	စွာ	ရေးသား
POS	ADJ	PART	V
Glossary	careful	-ly	Write
Translation	write carefully		

Figure 8. Example of Adverb

Verbs are usually ended with PPMs to form action type and tense and also suffixed zero or more particles to represent real status and tense and aspect of actions. Example verb phrase dependency structure is shown in Figure 9.



Word	မွေးဖွား	ခဲ့	သည်	
POS	V	PART	PART	PPM
Glossary	bear	-(for past)	-	.
Translation	was born			

Figure 9. Example of Verb

5. RELATED WORK

Myanmar sentence structure is different from other ASEAN languages of Laos, Thailand, Cambodia and Vietnam, Malaysia whose sentence structure is SVO. Among many SOV order languages, Myanmar, Japanese, Chinese and Korean have many similar styles in suffixing post markers and has been discussed in [1]. According to the similar syntactic structures, dependency-based head finalization for Myanmar language has been proposed based on dependency syntactic structure of these languages by applying the Statistical Machine Translation (SMT) in [1]. And a parsing approach of Myanmar has also been proposed by applying SMT by using Japanese as Pivot in [2]. Moreover, Japanese and Myanmar have very similar structures not only in grammar and word structure but also in syntactic annotation examples of nominal markers, particles, compound words, coordination, adjectives, and verbs as defined in UD in [11] and [13]. It can be seen that a postpositional phrase that is assigned with subject, direct or indirect object by case particles for nominal, dative, and accusative: が (ga), を (o), に (ni) in Japan-

Figure 10. Example sentence in Japanese (Left) and Myanmar (Right) with Dependencies links

ese sentence is similar to suffixed PPM in Myanmar, က (ka), ကို (ko), အား (ah), in Figure 10 in which second line is U-POS tags and third is glossaries of each word of phrases. A parallel corpus of Myanmar for Asian Language Treebank (ALT) project has been annotated for word segmentation, POS tags and syntax trees by manually [14]. But syntaxes of translated sentences from English are needed to be checked with referenced standard structures. That is why we firstly used myPOS corpus in this experiment to build direct dependency corpus model by applying indirect shared Japanese model in order to get more reliable dependency links for next sentences such as sentences from ALT corpus.

6. EXPERIMENT

The purpose of annotated corpus is for dependency parsing. This section presents annotation procedure by unsupervised approach and parsing experiments on annotated corpus.

6.1. Unsupervised Dependency Parsing

Unsupervised parsers infer the dependency structures based on language and tagset properties of dependency trees. A pure unsupervised approach should use only a raw corpus. But different unsupervised approaches have different motivations for different degree of (un) supervision based on the fact that allows to use different kinds of data and amount of knowledge about them. Moreover, current unsupervised dependency parsers were done with gold-standard supervised POS tags to build the most probable dependency tree [5]. UDPipe project tool[4] can be easily trained for new languages and requires neither additional resources such as morphosyntactic dictionaries, nor feature engineering and language-specific knowledge by the shared models of treebanks with UD of languages [9]. Therefore, we considered unsupervised approach with supervised language POS and U-POS to apply UDPipe for annotation and parsing experiment since there is no direct dependency syntax resources for Myanmar as mentioned above.

6.2. UDPipe

UDPipe can perform tokenization, morphological analysis, POS tagging, lemmatization and dependency parsing with UD. UDPipe employs *Parsito* parser that is a transition-based, non-projective dependency parser. Transition-based dependency parsing computes the dependency tree for a sentence by starting in an initial configuration and performing a sequence of transitions reaching some terminal configurations. The most popular projective stack-based arc standard system by Nivre [4] denoted as *stack* is employed. This system defines three types of transitions: *left_arc(l):*s_1 ->s_2 and *right_arc (l):*s_2 ->s_1 , which add a dependency arc with label *l* by removing second and first element of top of the stack respectively, and *shift*, which adds the next input word. Example derivation of a target dependency parse tree is shown in Figure 11. Moreover, it was proved that training a neural network (NN) classifier predicting transitions in a transition based parser with search-based oracle allowing *swap* operation [8] to more accurate fully non-projective parsing can achieve high accuracy.

Figure 12 describes the architecture of neural network classifier. For usual word embeddings, each word as a d-dimensional vector is $e_i^w \in \mathbb{R}^d$ and the full embedding matrix is $E^w \in \mathrm{R}^{d \times N_w}$ where N_w is the dictionary size. Meanwhile, POS tags and arc labels are mapped to a d-dimensional vector space, where $e_i^t, e_j^l \in \mathbb{R}^d$ are the representations of i^{th} POS tag and j^{th} arc label. Correspondingly, the POS and label embedding matrices are $E^w \in \mathrm{R}^{d \times N_t}$ and $E^l \in \mathrm{R}^{d \times N_l}$ where N_t *and* N_l are the number of distinct POS tags and arc labels. The model chooses a set of elements on the stack/ buffer positions for each type of information which might be word, POS or label, which might be useful for predictions denoted as the sets: S_w, S_t, S_l respectively [4]. Example configuration is illustrated in Figure 11 and S^t = {lc_1 (s_2).t, s_2.t, rc_1(s_2).t, s_1.t} to extract N, PPM, ADV, V, ..., PUNCT in order.

A standard neural network was built with one hidden layer, where the corresponded embedding of our chosen elements from S_w, S_t , S_l will be added to the input layer. Denoting n_w , n_t , n_l as the number of chosen elements of each type, x^w = [e_{w1}^w, e_{w2}^w ,$e_{wn_w}^w$] is added to the input layer, where S^w= { w1, … , wn_w }. Similarly, the POS tag features $x^l$and arc label features $x^l$are added to the input layer. The input layer is mapped to a hidden layer with d_h nodes through a cube activation function:

$$h = (W_1^w x^w + W_1^t x^t + W_1^l x^l + b_1)^3 \qquad (1)$$

In (1), $W_1^w \in \mathbb{R}^{d_h \times (d_{n_w})}$, $W_1^w \in \mathbb{R}^{d_h \times (d_{n_w})}$, $W_1^w \in \mathbb{R}^{d_h \times (d_{n_w})}$, and $b_1 \in \mathbb{R}^{d_h}$ is the bias. A softmax layer is finally added on the top of the hidden layer for modeling multi-class probabilities p = softmax (W_2h), where $W_2 \in \in \mathbb{R}^{|T| \times d_h}$ [4].

[4] http://ufal.mff.cuni.cz/udpipe

6.3. Experiment Setting

UDPipe and shared Japanese dependency model are used to obtain two syntactic annotated structures, dependency head and relation of CoNLL-U format by unsupervised parsing with UD. Firstly myPOS corpus is parsed with Japanese model. Parsed corpus was divided into training and test set. Training data was re-annotated on dependency heads as the reference structures and a model was built to parse by Myanmar model. Then Myanmar model is also evaluated by parsing on train and test data. The accuracies are measured by UAS and LAS and listed with data statistic in Table 2.

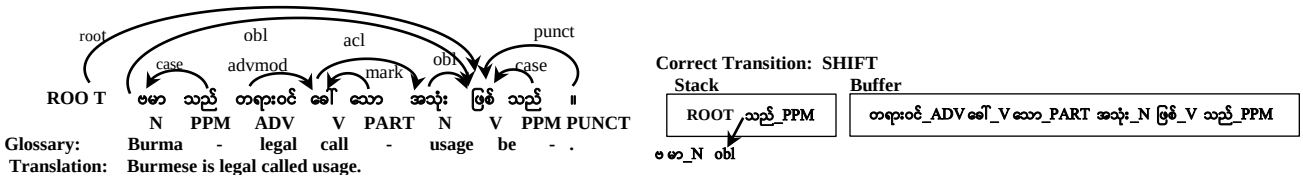
Data set	Sentences	Word Tokens	Accuracy	
			UAS	LAS
Train	10,000	217,636	93.88%	92.57%
Test	287	6,504	93.20%	91.21%

Table 2. Statistics on data and experiment results

6.4. Results and Discussions

The evaluation scores from Table 2 are over 90% and unsupervised dependency relations are acceptable. Based on the manual checked by referenced schemes, PART, PPM, and V have correct rates with 35%, 28%, and 16% respectively among other tags in training data since the grammar order and suffixed styles are similar between Japanese and Myanmar. But CONJ, N, and ADJ have very few correct rates with 5%, 3% and 2% because of the fact that Myanmar words or phrases are frequently combined as a syntactic role and that causes wrong heads in dependency trees although simple words or phrases or clauses can provide correct heads by Japanese model. However, being able to provide correct heads of roots of sentence and frequently occurred suffixed words by Japanese model are very worthy for Myanmar dependency structures. Besides unsupervised annotated result trees are also shown in Figure 13 by comparing with re-annotated referenced tree on unsupervised one. Adding manually dependency structures to the segmented and POS tagged sentences is complicated. In the annotated tree, words, "၏","မှာ","ခန့်", and "သည်", with relation (*case*) can attach to their relative nou-ns, ("မြို့", "အကျယ်အဝန်း", "မိုင်") and verb ("ရှိ"). The nodes, "မြို့", and "စတုရန်း" could not attach to their sub root heads of phrases, "အကျယ်အဝန်း" and "၁.၂၀-၃", main statement nouns of that sentence as circled nodes of referenced tree in Figure 13. It is important to provide correct links between words and main root of sentence. Defining correct links needs supervised control because suffixed nouns can be put in any position in sentence except the main root verb and Myanmar language is rich in not only vocabularies but also sentence structures. The circled nodes' heads are not like the referenced tree as we discussed above.

However, UDPipe can provide automatically not only many correct links between nodes and main root nodes but



Configuration			
Transition	**Stack = s[…3,2,1]**	**Buffer = b[1,2,3,…..]**	**Configuration: A**
	[ROOT]	[ဗမာ သည် တရားဝင် ခေါ် သော အသုံး ဖြစ် သည် ။]	∅
SHIFT	[ROOT ဗမာ]	[သည် တရားဝင် ခေါ် သော အသုံး ဖြစ် သည် ။]	
SHIFT	[ROOT ဗမာ သည်]	[တရားဝင် ခေါ် သော အသုံး ဖြစ် သည် ။]	
LEFT-ARC(obl)	[ROOT သည်]	[တရားဝင် ခေါ် သော အသုံး ဖြစ် သည် ။]	A ∪ obl (သည်, ဗမာ)
SHIFT	[ROOT သည် တရားဝင်]	[ခေါ် သော အသုံး ဖြစ် သည် ။]	
SHIFT	[ROOT သည် တရားဝင် ခေါ်]	[သော အသုံး ဖြစ် သည် ။]	
RIGHT-ARC(advmod)	[ROOT သည် ခေါ်]	[သော အသုံး ဖြစ် သည် ။]	A ∪ advmod (ခေါ် , တရားဝင်)
…….	…..	…..	…….
RIGHT-ARC(root)	[ROOT]	[]	A ∪ root (ROOT, ဖြစ်)

Figure 11. Example of transition-based dependency parsing. Top: dependency tree (left) and an intermediate configuration (right). Bottom. Transition sequence of the arc-standard system

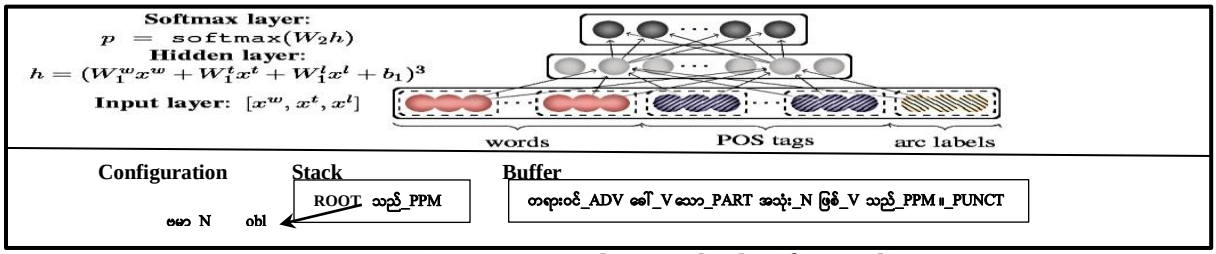


Figure 12. Neural Network Classifier Architecture

also their dependency relations with relatively high evaluation scores by a faster way rather than human annotation.

7. CONCLUSION AND FUTURE WORK

This paper has presented dependency structures and annotation from segmented and POS tagged corpus. Furthermore, the experiments and results of annotated data have been discussed for dependency parsing. As a conclusion, it can be said that UDPipe can support dependency annotation procedure of Myanmar.

Some results among annotated results of this experiment still need to be checked to be as the referenced dependency structures. Therefore, future work is to check all these results to be more reliable dependency model. After this, we will add more dependency information from new sentences such as ALT sentences or others to be more efficient for dependency parsing and Treebank.

Myanmar Word	မြို့ ၏ အကျယ်အဝန်း မှာ ၁.၂၀၃ စတုရန်း မိုင် ခန့် ရှိသည် ။
English glossary	**Town of area - 1.203 square mile about has - .**
Translation	**Area of town has about 1.203 square miles.**

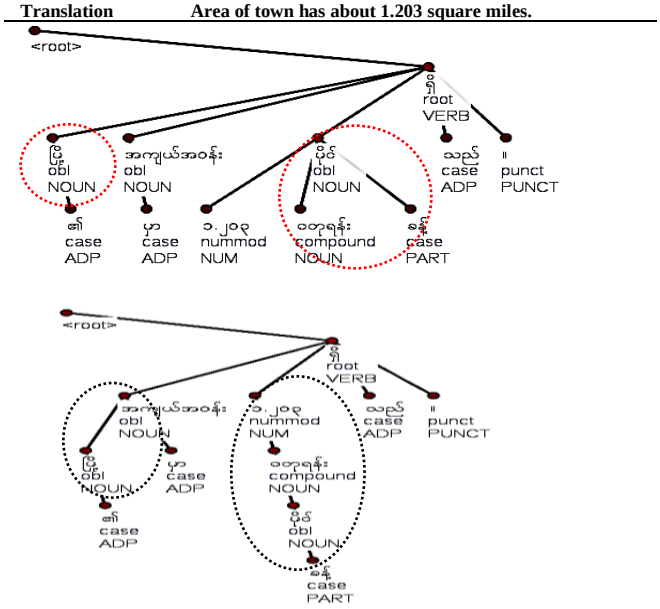


Figure 13. Top: Sentence. Middle: Unsupervised annotated tree. Bottom : Referenced dependency tree

8. REFERENCES

[1] C. Ding, et al, "Empirical Dependency-Based Head Finalization for Statistical Chinese-, English-, and French-to-Myanmar (Burmese) Machine Translation", Proceedings of the 11th International Workshop on Spoken Language Translation, Lake Tahoe, CA, USA, December 4-5, 2014, pp.184-191.

[2] C. Ding, et al, "Parsing Myanmar (Burmese) by Using Japanese as a Pivot", Proceedings of 14th International Conference on Computer Applications, Yangon, Myanmar, February, 2016, pp.158-162 .

[3] Ding, Yuan, et al, "Machine translation using probabilistic synchronous dependency insertion grammars", Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, Ann Arbor, MI, 2005, pp. 541-548.

[4] D. Chen, et al, "A fast and accurate dependency parser using neural networks", Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, ACL, Doha, Qatar, October 2014, pp. 740–750.

[5] D. Mareˇcek, "Twelve Years of Unsupervised Dependency Parsing", ITAT 2016 Proceedings, CEUR Workshop Proceedings Vol. 1649, pp. 56–62.

[6] J. Nivre, et al," The CoNLL 2007 shared task on dependency parsing", Proceedings of the CoNLL Shared Task Session of EMNLP-CoN-LL 2007, ACL, 2007, pp. 915–932.

[7] Khin War War Htike, et al, "Comparison of Six POS Tagging Methods on 10K Sentences Myanmar Language (Burmese) POS Tagged Corpus", 18th International Conference on Intelligent Text Processing and Computational Linguistics, Budapest, Hungary, April 17-23, 2017.

[8] M. Straka, et al,, "Parsing universal dependency treebanks using neural networks and search-based oracle", Proceedings of 14th International Workshop on Treebanks and Linguistic Theories, Warszawa, Poland, January, 2015.

[9] M. Straka, et al, "UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing", Proceedings of the 10th International Conference on Language Resources and Evaluation, Portorož, Slovenia, May, 2016.

[10] S. Buchholz , et al, "CoNLL-X shared task on multilingual dependency parsing", Proceedings of the 10th Conference on Computational Natural Language Learning, ACL, 2006, pp.149–164.

[11] S. Mori, et al, "A Japanese word dependency corpus", Proceedings of the 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland, May 26-31, 2014, pp. 753–758.

[12] S. Petrov, et al, "A universal part-of-speech tagset", Proceedings of the Internaional Conference on Language Resources and Evaluation, Istanbul, May 21-27, 2012, pp. 2089–2096.

[13] T. Tanaka, et al, "Universal Dependencies for Japanese", Proceedings of the 10th International Conference on Language Resources and Evaluation, Portorož, Slovenia, May, 2016.

[14] Ye Kyaw Thu, Win Pa Pa, et al, "Introducing the Asian Language Treebank (ALT)", 10th edition of the Language Resources and Evaluation Conference, Portorož, Slovenia, May 23-28, 2016.