I.J. Information Technology and Computer Science, **2014, 10, 28-35**
Published Online September 2014 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2014.10.04

Development of Myanmar-English Bilingual WordNet like Lexicon

Soe Lai Phyue
University of Computer Studies, Mandalay, Myanmar
Email: soelaiphyue@gmail.com

***Abstract*—** A bilingual concept lexicon is of significance for Information Extraction (IE), Machine Translation (MT), Word Sense Disambiguation (WSD) and the like. Myanmar-English Bilingual WordNet like Lexicon (MEBWL) is developed to fulfill the requirements of Language Acquisition (LA). However, it is reasonably difficult to build such a lexicon is quite challenging in time and cost consuming. To overcome this challenging, this paper integrates linguistic resources, including Myanmar-English dictionary, English-Myanmar dictionary and WordNet to construct a Myanmar-English WordNet like lexicon by acquiring the lexical and conceptual knowledge from WordNet and Myanmar<->English Machine Readable Dictionaries (MRDs). The system includes three phases which include the MRD extraction phase, the link analyzing phase and the WordNet construction phase. The first phase converts the data from multiple resources with different format into a common format and joins and aligns the scattered data for smoothly access and group the data according their part of speech (POS). The link analyzing phase analyzes, classifies and generates candidates of translation links. In the constructing phase, MEBWL is constructed from the verified translation link and WordNet. Beside then, to support the inflected word of Myanmar to English words, morphological processor is designed.

***Index Terms*—** Information Extraction (IE), Machine Translation (MT), Word Sense Disambiguation (WSD), Myanmar-English Bilingual WordNet like Lexicon (MEBWL), Machine Readable Dictionaries (MRDs), Part of Speech (POS).

I. Introduction

As the processing of content information has nowadays become the center of NLP, a bilingual concept MRD is of increasingly great significance for IE, MT, WSD and the like. Since a lexical resource is the central repository of data for all language processing applications, it contains information for human consumption as well as computer programs. Each application can also be processed at levels of detail ranging from a rough approximation triggered by keywords to a deep understanding that applies all the resources of syntax, semantics and pragmatics.

Myanmar language, also known as Burmese, is the official language of the Union of Myanmar. Myanmar language is a member of the Tibeto-Burman languages, which is a subfamily of the Sino-Tibetan family of languages. Myanmar still lacks support on computers and not many NLP tools and applications are available for this language.

The language resources like lexicon is a bridge between a language and the knowledge base expressed in that language. A significant increase in the use of lexical databases has led WordNet [1] to become one of the most widely used lexical information source for NLP applications. And it is for sure that the computational linguists would find such a lexicon indispensable and useful as semantic information when facing ambiguities in languages in their applications [2].

Manual construction of WordNet is the most reliable technique for obtaining structured lexicons but it is costly and highly time-consuming [3]. Therefore, when building a Myanmar-English bilingual concept lexicon, we must take the issue of compatibility with WordNet into account. In other words, for each English concept in WordNet, there should exist a corresponding Myanmar concept in the bilingual lexicon and vice versa.

The previous work, Myanmar WordNet and lexico conceptual knowledge resources [4, 5] (Monolingual lexical database) has been developed using the semi automatic ways using the WordNet and Myanmar<-> English Machine Readable Dictionary (MRDs). Myanmar-English bilingual WordNet like lexicon is also developed using this approach. The bilingual lexicon is useful for machine translation from Myanmar to English system.

This paper is organized as follows. The related work of building WordNet like lexicon is described in section II. In section III, the nature of existing resources of the Myanmar-English MRDs and English WordNet are presented. Section IV sketches the architecture of the proposed system. Synset alignment and system evaluation are described in section V and section VI. Section VII express the design of morphological processor to support the bilingual lexicon. Then evaluation result of MEBWL is discussed in section VIII. Finally, section IX provides some concluding remarks and future directions of research.

II. Related Works

The demands on the lexicon also vary with the type of application in Natural Language Processing. Several researches were proposed to extend the design idea of WordNet to other languages.

Yang Liu et.al [6] emphasized the inheritance and transformation of the existent monolingual lexicon to develop the Chinese-English bilingual WordNet like lexcion. They extracted all the common knowledge in

WordNet as the semantic basis for further use and developed a visualized developing tool for the lexicographers to interactively operate on to express the bilingual semantics. The bilingual lexicon had thus gradually come into being in this natural process. Their approach benefited a lot by employing it to build Chinese Concept Dictionary (CCD). By now, we have fulfilled more than 32,000 Chinese-English concept pairs in noun.

Hsin-Hsi Chen and et.al [7] developed the Chinese-English WordNet and than the result is employed in Chinese-English information retrieval. They integrated five linguistic resources, including Cilin, a Chinese-English dictionary, ASBC corpus, SemCor, and WordNet, to construct a Chinese-English WordNet. A semantic tag (e.g., a Cilin sense tag and a WordNet synset) is characterized by the words surrounding it in a semantics-tagged corpus (e.g., a sense-tagged ASBC corpus and SemCor) to derive two sets of semantic vectors (i.e Cilin sense and WordNet synset vectors). And then they selected a set of English translations that are semantically coherent. Because the size of the sense vector is very large, we only consider 200 Chinese words with larger weights to reduce the complexity. Word cooccurrence model is employed and find the most similar synset from each English translation to build the bilingual WordNet.

EuroWordNet [8, 9] aimed to build a multilingual database consisting of wordNets in several European languages (English, Dutch, Italian, and Spanish). Each language-specific wordNet were structured along the same lines as WordNet (Miller et al, 1990), i.e. synonyms are grouped in synsets. The size of the database was around 25,000 comparable synsets in each language, corresponding with more or less 50,000 word meanings. Each of the European WordNets is a network of relations between word meanings in a specific language. The semantic relations are therefore considered as language-internal relations. In addition to the language-internal relations, each synset is linked to the closest synset in the Princeton WordNet1.5.

MultiWordNet [10] was built with new synsets in correspondence with the PWN synsets. The semantic relations of English synsets has been imported to Italian language, assuming that if there are two synsets in PWN and a relation holding between them, the same relation holds between the corresponding synsets in the Italian language. The resources used in construction of MultiWordNet are Italian-English dictionaries and English WordNet. To help the lexicographer in constructing MultiWordNet, two automatic procedures are introduced: Assign-procedure and Italian to English section of the Collins dictionary and it procedures as output a set of candidates. Each candidate is described as <PWN synset, confidence score> pair, where confidence score measures the degree of confidence in the link between the Italian word sense and the PWN synset. Only candidates with a confidence score greater than a certain threshold are proposed to the lexicographer.

Such multilingual database is useful for cross-language IR, for transferring of information from one resource to another or for simply comparing the different wordNets. Myanmar English bilingual WordNet is developed using the semiautomatic approach. The method uses as skeleton English WordNet and extracts its knowledge from a variety of lexical sources (taxonomies, monolingual and bilingual dictionaries). Our approach makes extensive uses of English WordNet in several steps of the building process. The system has been applied to build MEBWL.

III. Overview of the System

Myanmar <-> English MRDs and the WordNet are main reusable resources of the system to produce the Myanmar English WordNet like lexicon. The system consists of three steps.

The first step is data extraction from the existing resources and the next is the link analyzing according to the class and structure synset link criteria between Myanmar English translation link. The final step is the attachment of glossary between words. Fig. 1 shows the architecture of our design. The linguistic resources are introduced below at first.

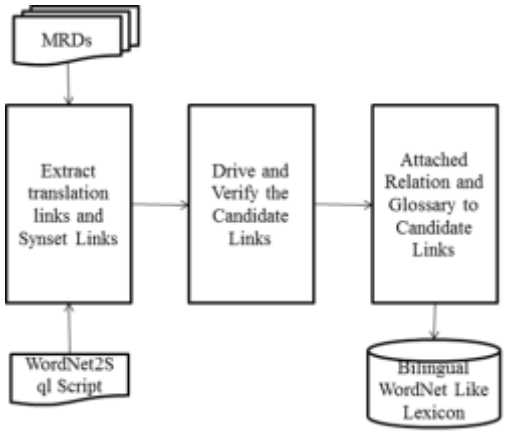

Fig. 1. Overview of the Development of Myanmar-English WordNet Like Lexicon

A. Machine Readable Dictionaries

The MRDs is the electronic version of standard dictionaries which may contain other lexicographic information that does not appear in the printed version. MRDs can be monolingual, bilingual and multilingual. The amount of semantic information useful for NLP which has been automatically extracted from dictionary definitions is severely limited. MRDs do not explicitly represent lexical relations, and gaps easily filled in by a human user cannot be automatically identified or resolved. Conventional knowledge of word usage is not even implicitly represented, as MRDs are a secondary source of word knowledge and concentrate on conveying established senses. For these reasons MRDs are not adequate as a source of lexical knowledge for computational systems. MRDs can be monolingual, bilingual, and multilingual. Any number of MRDs can be taken into account. However, the system focuses on the General Purpose Language (GPL) Myanmar <–> English Dictionaries and English WordNet used as the reusable resource.

B. WordNet Script

WordNet is a large lexical database of English, developed under the direction of George A. Miller (Emeritus). Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. Additionally, a synset contains a brief definition ("gloss") and, in most cases, one or more short sentences illustrating the use of the synset members. Word forms with several distinct meanings are represented in as many distinct synsets. Thus, each form-meaning pair in WordNet is unique. The resulting network of meaningfully related words and concepts can be navigated. WordNet is also free and public online. WordNet's structure makes it a useful tool for computational linguistics and NLP. In fact, it has become the de facto standard for several NLP tasks such as WSD. It is manually developed by a team of linguists and its design is inspired by psycholinguistic theories of human lexical memory. WordNet is also classified into 4 structures according the Part of Speech such as nouns, verbs, adjectives, and adverbs. Wordnet2sql was developed for process in a relational database with data from WordNet version 3.0. It could be easily extending or modifying the data structures.

IV. Extracting the Data

It performs several processes to extract and transform data from two main machine readable dictionaries: WordNet and WxpyDic. It consists of five steps: font conversion, format conversion, merging, cleaning and grouping. Among them, merging only concern for bilingual dictionary and others have two procedures, the first procedure involves with bilingual dictionary and the last involves with monolingual dictionary. Data from Monolingual (WordNet) and Bilingual MRDs (Myanmar<->English MRDs) are extracted for link analyzing phase.

A. Font Conversion

The data are converted from various encoding font to a common font. Our system used the Myanmar3 Unicode as a common font.

B. Format Conversion

All of the data are needed to convert desired data format. The conversion of data format transforms data from different format into the same common format, relational database, for more convenient in accessing and retrieving data. The existing resources encoding in diffident formats, WordNet is SQL script and Bilingual dictionaries are flat file. The text file of MRDs is needed to convert to get desired data format, due to Myanmar 3 font, file is converted through the java engine because of the Myanmar3. WordNet2Sql file are also convert to desire data format. WordNet2Sql source are used as MySql database.

C. Merging

The Myanmar->English and English->Myanmar MRDs are combined as a one relation for further processes. The data are also rearranges for more convenient access and retrieval. The bilingual dictionaries contain heterogeneous data: English-Myanmar and Myanmar-English dictionaries. They are in separate data source which is inconvenience in access and retrieval. To avoid the redundant and erroneous data when access and retrieve data from both, heterogenous data sources are combined into one..

D. Cleaning

All of the merged data and WordNet are needed to handle noisy, erroneous, missing or irrelevant data that usually come with the existing resources. By merging two bilingual dictionaries, it has duplicated records, missing and noisy records. And WordNet has also missing fields. Therefore for the correctness and accuracy of the result, the data cleansing is performed to eliminate such kind of data.

E. Grouping

In this step, the words in MRDs are need to be group according to their Part Of Speech (POS). Because WordNet is specialized for Noun, Verb, Adverb and Adjective and bilingual lexicon represented all POS.

F. Result of Data Extraction Phase

As aforementioned, the purpose of this module is to prepare resources (both monolingual and bilingual dictionary) for next phase (link analyzing phase). In this phase, the data which resources of monolingual (WordNet) and bilingual MRDs are converted to font and format, merged, cleaned and also grouped. The result of each process is as shown in Table 1 and this data are used in link analyzing phase to create the candidate of translation link.

Table 1. Result of Data Statistic in Data Extraction Phase

Dictionary	Original	Cleaning	Merging	Grouping						
				Noun	Verb	Adj	Adv	Pron	Prep	Conj
E-M Dic	29625	23855	_	_	_	_	_	_	_	_
M-E Dic	30722	29049	_	_	_	_	_	_	_	_
Merge	_	_	55277	31243	12706	6819	3369	170	108	185
WordNet2Sql	148730	148730	148730	_	_	_	_	_	_	_

V. Derive and Verify the Candidate Link

All data prepared by the MRD Extractor are used in Link Analyzer to generate the candidate link of Myanmar Word with English concepts by automatically constructing the link by using the criteria that lexicographer use in create the dictionary and sampling the candidate links for manually verify, then using Logistic regression (statistical method) to construct and choose the candidate links to generate Myanmar WordNet and bilingual computational lexicon. Therefore, this phase comprises 6 components includes construction candidate link, sampling, verifying, construction model, evaluation model and deploying the model.

A. Derive the Candidate Link

To construct the candidate link of Myanmar word and English concepts, 13 criteria are utilized for disambiguate the link. The criteria are categorized into class and structure of WordNet synset. Class criteria group classifies the translation between Myanmar and English words using two criteria: Monosemic and Polysemic criteria. Each can be divided into 4 sub-criteria: 1:1, 1:N, M:1, N:M.

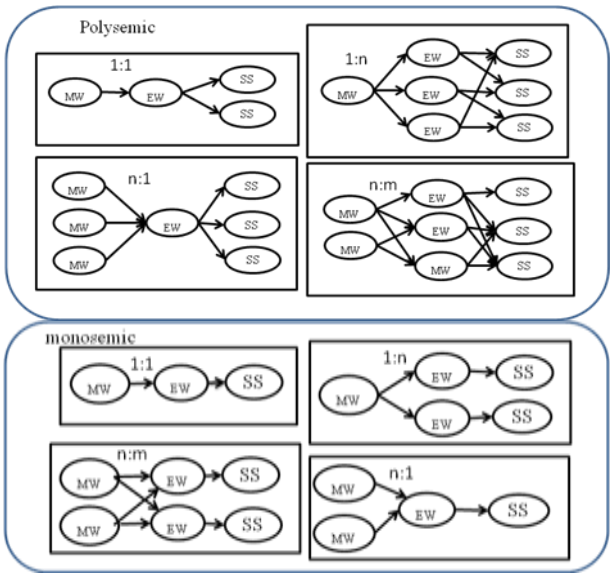

Fig. 2. Class Criterion Group

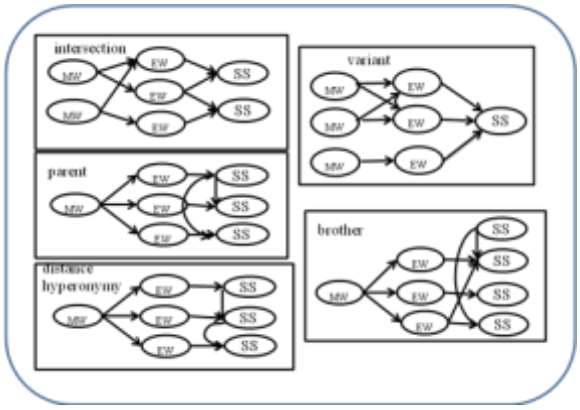

Fig. 3. Structural Criterion Group

In a structure criteria group, it takes profit of WordNet structure. It consists of 5 criteria: variant, intersection, parent, brother and distant hyponymy criteria.

The result of this process is sets of translation link of each criteria and matrix of existence of translation link in each criteria.

B. Sampling

The candidate links are sample by their criteria for verification. Candidate set of translation link is sample by using the random stratified sampling technique. In this system, the candidate links are already classified into several stratums as criteria. Then the candidate links are randomly sampled with the number of the 400 translation links in each criterion. In this research, the stratified sampling technique is applied with 95% confidence level.

C. Verify the Link

Each candidate link in the sample has to be verified by manually. The translation links are automatically evaluated by using existing resources (bilingual dictionary). The specification of evaluator is the expert in both Myanmar and English language, the specification of evaluator is compromised to non-expert but using standard bilingual dictionary in evaluation. The result of verification is represented as a summary existence matrix.

The summary existence matrix is used to construct a model for predicting the correctness of the remaining translation links and identifying the correlation and the significance of the multiple criteria. The correctness of each criteria is shown in Fig. 4.

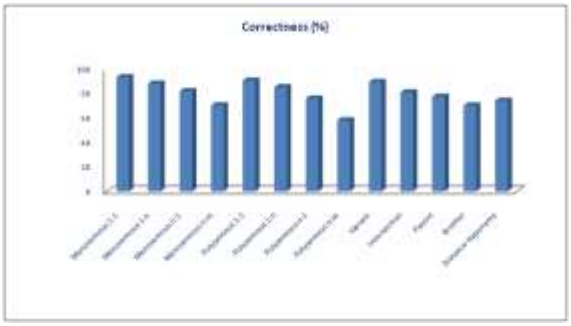

Fig. 4. Correctness of Each Criteria

D. Constructing the Model

The probability that a link would be correct can be estimated by P(OK) = NOK/NTOT. In this research, we use the logistic regression model to predict the correctness of the remaining links. In general, the logistic regression is used to predict a discrete outcome from a set of binary variables. The linear logistic regression model can be defined as Equation (1)

$$\log((P(NOK/NTOT)) = \beta_0 + \beta_1 C_{01} + \ldots + \beta_{13} C_{13} \quad (1)$$

where NOK is the number of correct evaluation for the set of solutions of every group of methods, NTOT accumulates the total number of evaluations, is a boolean variable representing the existence of link in the ith criterion, and is unknown parameter which required the least square criterion.

E. Evaluating the model

A statistical approach is used to evaluate the model. By using Pearson goodness-of-fit, determines the significance of model. A P-value of each criterion describes the significant of that criterion in the model by explaining the probability of a link of being correct. For a P-value lower than 0.05, the criterion is significant in the model. The evaluated model and each P-value is as shown in Table 2.

F. Deploying the Model

The model is deployed to construct the MEBWL. By applying the model to all remaining translation links, the semantic relations between Myanmar words and synsets are constructed according to each criterion as shown in Table 3.

Table 2. Coefficient and P-value of each Criteria of Logistic Regression Model

Criteria	Coefficient	P-Value
C01	1.54140	0.000
C02	1.02516	0.005
C03	0.54520	0.013
C04	0.20719	0.047
C05	0.25404	0.047
C06	0.15901	0.015
C07	-0.06980	0.042
C08	0.22585	0.050
C09	0.12671	0.019
C10	0.00175	0.009
C11	0.18000	0.023
C12	-0.26200	0.047
C13	0.30289	0.009

Table 3. Deploying the Logistic Regression Model

Myanmar	English	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	E(P(OK))
ကနုကမာ	oyster shell	1	0	0	0	0	0	0	0	0	0	0	0	0	0.926
ကလေးလက်တွန်းလှည်း	pram	1	0	0	0	0	0	0	0	0	0	0	0	1	0.933
ကာဘိုင်သေ နတ်	carbine	1	0	0	0	0	0	0	0	0	0	0	0	0	0.926
ကားဂိတ်	bus station	1	0	0	0	0	0	0	0	0	0	0	0	0	0.926
ကုလားအုတ်	camel	1	0	0	0	0	0	0	0	0	0	0	0	0	0.926

VI. Relation and Glossary Attachment

Construction phase of MEBWL uses the evaluated Myanmar-English translation links and the existing English WordNet to attach Myanmar words to English WordNet. This component comprises in two processes: logical inference and glossary attachment.

A. Logical Inference

English WordNet is used as a skeleton of Myanmar English bilingual lexicon, translates each English word into Myanmar and attaches English synset to each Myanmar word. Moreover, for lexicon, the internal relations (e.g. Hypernym, Hyponym, Meronym, and Holonym, anonym and entailment) are infers by applying transitivity relation.

B. Glossary Attachment

Glossary information to bilingual WordNet like lexicon is attached in this process.

VII. Morphological Processor for MEBWL

Analysis of agglutinative morphology involves many difficulties which might go unnoticed even by researchers in NLP. The development of a morphological analyzer is likely to benefit from a unifying principle which ensures that its coverage of the grammar is comprehensive in some sense. Morphological Analysis as a vocabulary acquisition strategy has both its advocates and antagonists. Morphological, or Structural, Analysis is the process of breaking down morphologically complex words into their constituent morphemes as their language nature. For this reason, we need to study the detail nature of Myanmar morphology. To generate the relevant English word as the Myanmar word, we used the finite state automaton.

A. Morphological Analyzer for Myanmar Language

Morphological Analysis as a vocabulary acquisition strategy has both its advocates and antagonists. Since most Myanmar words consist of a stem, which mainly specifies the lemma, and a set of affixes that mainly specify the morpho-syntactic features, it is appropriate to concatenate the prefix or suffix elements. We take a somewhat profit of nature of Myanmar language to defining and computing word relations to its application in a morphological processor for Myanmar morphological analyzer. The main advantage of this process is the extreme simplicity both of its tagging process and of their interpretation. Therefore, the analyzer of Morphocon consists of two processes. The first is to define the stem by segmenting with prefix and suffix and the second is recognized the inflected form of stem by their prefix and suffix. To segment and recognize

the Myanmar word, Myanmar morphological rule for inflectional form is studied.

Analysis of Myanmar Nouns: The noun can have a suffix indicating plurality. Nouns in Myanmar are pluralized by suffixing the particle တွေ if the word ends in a glottal stop) in colloquial Myanmar or များ in formal Myanmar. The particle တို့, which indicates a group of persons or things, is also suffixed to the modified noun. An example is below:

မြစ် - river
မြစ်တွေ - rivers (colloquial)
မြစ်များ - rivers (formal)
မြစ်တို့– rivers

Analysis of Myanmar Verbs: In the affirmative, the order of elements is V [one or more roots, possibly compounded] (+auxiliary verb) + aspect particle + modal ending. The roots of Myanmar verbs are almost always suffixed with at least one particle which conveys such information as tense, intention, politeness, mood, etc. Preceding the aspectual element there may be inserted the element pa indicating explicitness, which serves to mark the utterance as 'polite', though not when it is used with an imperative: သည်

Many of these particles also have formal/literary and colloquial equivalents. In fact, the only time in which no particle is attached to a verb is in imperative commands. However, Myanmar verbs are not conjugated in the same way as most European languages; the root of the Myanmar verb always remains unchanged and does not have to agree with the subject in person, number or gender.

Alone, the statement စား is imperative. The suffix တယ် can be viewed as a particle marking the present tense and/or a factual statement:

စားတယ် - I eat

The suffix ခဲ့ denotes that the action took place in the past. However, this particle is not always necessary to indicate the past tense such that it can convey the same information without it. But to emphasize that the action happened before another event that is also currently being discussed, the particle becomes imperative. Note that the suffix တယ် in this case denotes a factual statement rather than the

ငါစားခဲ့တယ် - I ate

The particle နေ is used to denote an action in progression. It is equivalent to the English '-ing'"

ငါစားနေတယ် - I am eating

This particle ပြီး, which is used when an action that had been expected to be performed by the subject is now finally being performed, has no equivalent in English. So in the above example, if someone had been expecting you to eat and you have finally started eating, the particle ပြီး is used as follows:

ငါစားပြီ - I am (now) eating

The particle မယ် (literary form: မည်) is used to indicate the future tense or an action which is yet to be performed:

ငါစားမယ် - I will eat

The particle တော့ is used when the action is about to be performed immediately when used in conjunction with မယ်. Therefore it could be termed as the "immediate future tense particle".

ငါစားတော့မယ် - I will eat (straight-away)

Analysis of Myanmar Adjectives: Myanmar does not have adjectives per se. Rather, it has verbs that carry the meaning "to be X", where X is an English adjective. These verbs can modify a noun by means of the grammatical particle တဲ့ in colloquial Myanmar (literary form: သော), which is suffixed as follows:

Colloquial: ချောတဲ့
Formal: ချောသော

Gloss: "beautiful" + adjective particle + "Noun"

Comparatives are usually ordered: X + ထက်ပို + adjective, where X is the object being compared to. Superlatives are indicated with the prefix အ + adjective + ဆုံး.

B. Morphological Generator of Equivalent English Word

For morphological generator, the word need to define their case and translate as the equivalent English meaning as "teeth". The implementation is based on the concept of validation grammars. The morphological processing is controlled by a finite automaton. The detail study of generating process for each POS is described in following.

Generate Relevant Myanmar to English Nouns: The noun in Myanmar word can have a suffix indicating plurality. It can be pluralized by suffixing the particle "တွေ" in colloquial Myanmar or "များ" in formal Myanmar. The particle "တွေ" which indicates a group of persons or things, is also suffixed to the modified noun. To generate plural or singular forms of English word, we use English grammar rules in Table 4.

Table 4. Change Form of Singular Noun to Plural Noun

Type of Noun	Rule for Forming the Plural	Examples
ends in s, x, ch, o or sh	Add 'es' to the end	
Word ends in z	Add 'zes' to the end	buzz/buzzes
End in 'y' preceded by a consonant	Change the final 'y' to 'ies'	ally/allies
End in 'y' preceded by a vowel	Add 's' to the end	
Ends in 'f' or 'fe' (but not 'ff' or 'ffe')	Change the 'f' or 'fe' to 'ves'	calf/calves, elf/elves
Irregular Noun		child/children, die/dice

Singular words which end is s, z, sh, ch or x, we add es to become plural words. Singular words which end is consonant with “y” changes the “y” to “i” and add es. All other singular words add “s”. But some nouns have irregular form e.g; man (plural men). We cannot handle this irregular noun.

Generate Relevant Myanmar to English Verbs: The In the affirmative, the order of elements is V [one or more roots, possibly compounded] (+auxiliary verb) + aspect particle + modal ending. The most commonly used verb particles and their usage are shown below with an example verb root “စား”. Alone, the statement “စား” is imperative. The suffix “တယ်” (literary form: သည်) can be viewed as a particle marking the present tense and/or a factual statement. The suffix “ခဲ့” denotes that the action took place in the past. Note that the suffix “သည်” in this case denotes a factual statement rather than the present tense. We also generate verb tense by using verb stem word and suffixes particles. Stem of verb add “ed” to become past tense. We use English grammar rule to change verb tense but some verb has irregular form e.g; past tense of “read” is also “read”. We handle irregular verb by using irregular verb list defined by Oxford Dictionary. The particle “နေ” is used to denote an action in progression. It is equivalent to the English '-ing'". This particle “ပြီ” which is used when an action that had been expected to be performed by the subject is now finally being performed, has no equivalent in English. So in the above example, if someone had been expecting you to eat and you have finally started eating, the particle “ပြီ” is used. The particle “မည်” “မယ်” “တော့မည်” ”တော့မယ်” are used to indicate the future tense or an action which is yet to be performed.

Generate Relevant Myanmar to English Adjectives: In Myanmar word, adjective is defined by the word with suffix “သော”,”သည့်”,”မည့်” Beside then, it has verbs that carry the meaning "to be X", where X is an English adjective. Comparatives are usually ordered: X + “ထက်ပို” “ပို” “ပို၍” + adjective, where X is the object being compared to. For this case , we add the X word in adding more as prefix or suffix as X-er. Superlatives are indicated with the prefix အ + adjective + ဆုံး.

VIII. Coverage of Bilingual Lexicon

As the experiments, the coverage of original MRDs and MEBWL is quite different. Some of the Myanmar word has not direct translation word which may form with several words as a phrase. As that word the origin of WordNet has not included and we omit this word in building the MEBWL.

Table 5. Dictionary Statistic of the Resources and Coverage of Bilingual Computational Lexicon

POS	Coverage of Bilingual Dictionary	Maximum Coverage of Bilingual Lexicon	Coverage of Bilingual (%)
Noun	31243	25375	81.22
Verb	12706	10737	84.5
Adjective	6859	4245	61.89
Adverb	3369	2938	87.21

The dictionary statistics and set of link derivation for Myanmar English bilingual lexicon according to their part of speech is as shown in Table 5. The quality of lexicon depends on number of synset link and the link which covered the meaning for each word. According to the result the set of Myanmar word and their relation yielded a satisfactory result, it also useful for computational lexicon.

IX. Conclusion

Myanmar English WordNet like lexicon is new state for Myanmar lexical resources. The semantic relations between English words in WordNet and the translation relations between English and Myanmar words in English<-> Myanmar machine readable dictionary are considered. Applying the class and structure methods to

noun and verb and only class method to adjective and adverb yielded a satisfactory result. Finally, a set of the verified candidate link is used to automatically construct the Bilingual lexicon.

REFERENCES

[1] C. Fellbaum, WordNet: An electronic lexical database. MIT Press, Cambridge, Massachusetts, 1998.

[2] A. Valitutti, C.o Strapparava, O. Stock, Developing Affective Lexical Resources, Psychnology Journal, 2004, Pp.61-83.

[3] X. Farreres, G. Rigau, and H.Rodr ǵuez, Using WordNet for Building WordNets, in Proceedings of COLING/ACL Workshop on the Usage of WordNet in Natural Language Processing Systems, 1998, p. 65-72.

[4] S. L. Phyue, "Construction of Myanmar WordNet Lexical Database", in IEEE student Conference on Research and Development (IEEE SCOReD 2011), Malaysia, December 2011.

[5] S. L. Phyue and A. Thida, "Development of the Lexico-Conceptual Knowledge Resource for Myanmar NLP Applications", Scientific journal of Computer Engineering (SJCE).

[6] Y. Liu, S.Yu, J.Yu, Building a Bilingual WordNet Like Lexicon: the New Approach and Algorithms, in Proceedings of the 19th international conference on Computational linguistics - Volume 2, Pp 1-5

[7] H. H. Chen, C.C. Lin and W.C. Lin, Construction of a Chinese- English WordNet and its application to CLIR, In Proceedings of the fifth international workshop on Information retrieval with Asian languages, New York, USA, 2000, Pp. 189-196.

[8] A.Artale,B.Magnini and C.Strapparava, Lexical discrimination with the Italian version of WordNet. In Proceedings of ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources, 1997.

[9] J. Atserias, S. Climent, X. Farreres, G. Rigau and H. Rodr ǵuez, Combining Multiple Methods for the automatic Construction of Multilingual WordNets, In proceedings of International Conference on Recent Advances in Natural Language Processing (RANLP'97), Tzigov Chark, 1997, Pp.143-149

[10] E. Pianta, L. Bentivogli and C. Girardi, MultiWordNet: Developing and Aligned Multilingual Database, In Proceedings of the First International Conference on Global WordNet, Mysore, India, January 21-25, 2002, Pp. 293-302.

Authors' Profiles

S. L. Phyue has received her B.C.Sc (Bechelor of Computer Science) and M.C.Sc (Master of Computer Science) degrees from University of Computer Studies, Yangon in 2004 and 2008, respectively. Currently, she is a candidate for the degree of Ph.D of Information Technology in University of Computer Studies, Mandalay in Myanmar. She research interests is Natural Language Processing, Information Retrieval and semantic technology.

How to cite this paper: Soe Lai Phyue,"Development of Myanmar-English Bilingual WordNet like Lexicon", International Journal of Information Technology and Computer Science(IJITCS), vol.6, no.10, pp.28-35, 2014. DOI: 10.5815/ijitcs.2014.10.04