Rule-based Machine Translation from Myanmar to Kayah-Li Language

Zar Zar Linn[1], Moh Moh Aung[2], May Su Hlaing[3], Htwe Ei Hlaing[4], Poe Ei Phyu[5], Kyawt Kyawt Htay[6]

Faculty of Computer Science, University of Computer Studies (Loikaw)

[1]zzlinn@gmail.com, [2]moemoea2009@gmail.com, [3]maysuhlaingster@gmail.com, [4]hlaing.cf@gmail.com, [5]poeeiphyu571@gmail.com, [6]kyawtkyawthtay2014@gmail.com

***Abstract*— Machine Translation is one of Natural Language Processing (NLP) tasks which are modern computational technologies. The machine translation is to translate text from one natural language to another. This involves accounting for the grammatical structure of each language and using rules, and grammars to transfer the grammatical structure of the source language (SL) into the target language (TL). This paper presents Myanmar to Kayah-Li for translating well-structured Myanmar sentences into well-structured Kayah-Li sentences using a rule- based machine translation.**

***Keywords*— grammatical structure, rule-based, machine translation, natural language processing.**

I. Introduction

Machine translation (MT) is a process, sometimes referred to as Natural Language Processing which uses a bilingual data set and other language assets to build language and phrase models used to translate text. As computational activities become more mainstream and the internet opens up the wider multilingual and global community, research and development in machine translation continues to grow at a rapid rate.

A few different types of machine translation are available in the market today, the most widely use being Statistical Machine Translation (SMT), Rule-Based Machine Translation (RBMT), and Hybrid Systems, which combine RBMT and SMT.

Language is very important part of the communication. There are many different languages spoken in Myanmar among which is the Myanmar language. This paper presents the language translation from Myanmar to Kayah-Li language.

Kayah Li (Red Karen) is a Central Karenic language of the Tibeto Burman Language family. It is spoken by the Kayah Li people from Myanmar. (Li means "red") The Kayah people live predominantly in Kayah State, Shan State and along Thailand's northwestern border. In more recent years, Kayah Li people have been given the opportunity to emigrate all over the world from New Zealand and Australia to Finland, the USA and other countries. While there is no recent census data to draw official figures from, an estimate of the Kayah Li speaking population is over 100,000 people.

While the Kayah language has three dialects, western, eastern and northern, the reality is that different speech varieties can be found from village to village.One man illustrated the richness of varieties with some hyperbole, complaining that even in one household, the Kayah spoke differently from each other [6].

II. Related Works

[13] proposed 39 function tags for Myanmar Language and addressed the question of assigning function tags to Myanmar words and used a small functional annotated tagged corpus as the training data. In the task of function tagging, we used the output of morphological analyzer which tagged the function of Myanmar sentences with correct segmentation, POS (part-of-speech) tagging and chunking information. Naïve Bayesian statistics was used to disambiguate the possible function tags of each word in the sentence. The performance of

function tagging for simple and complex sentences were evaluated.

[14] tried to disambiguate for syntactic analysis system by many dependency rules and segmentation. Segmentation is made during parsing. If two adjacent morphemes had no syntactic relations, their syntactic analyzer made new segment between these two morphemes, and found out all possible partial parse trees of that segmentation and combined them into complete parse trees. Also the adjacent-rule and adverb subcategorization were used to disambiguate of syntactic analysis. The syntactic analyzer system used morphemes for the basic unit of parsing. All possible partial parse trees on each segmentation process made and tried to combine them into complete parse trees.

[8] introduced mechanism which converts multi sentences, question sentences of English to Sanskrit text to speech conversion. They stated that the model consists of array of translation rules to translate from source to target sentence, which is the frame of Rule based Machine Translation System.

[9] developed a rule-based machine translation system for English to Malayalam language pair. Their system took in the English sentence as input and parse with the help of Stanford Parser.

III. MACHINE TRANSLATION TECHNIQUES

Machine Translation refers to the use of computers to automate some of the tasks or the entire task of translating between human languages. The major machine translation techniques are Statistical Machine Translation (SMT), Example Based Machine Translation (EBMT) and Rule Based Machine Translation (RBMT). The RBMT is used in this paper.

A. Rule-based Machine Translation

The main approach of RBMT systems is based on linking the structure of the given input sentence with the structure of the demanded output sentence, necessarily preserving the unique meaning. The following example can illustrate the general frame of RBMT [15]:

ကျွန်တော် ထမင်း စားသည်

Source Language = Myanmar

Target Language = Kayah-Li

Minimally, to get a Kayah-Li translation of this Myanmar sentence one needs:

1. A dictionary that will map each Myanmar word to an appropriate Kayah-Li word.
2. Rules representing regular Myanmar sentence structure.
3. Rules representing regular Kayah-Li sentence structure.

And finally, we need rules according to which one can relate these two structures together.

Accordingly, we can state the following **stages of translation:**

1st: getting basic part-of-speech information of each source word:

ကျွန်တော်= ကတ္တား; စားသည် =ကြိယာ; ထမင်း = ကံ

2nd: getting syntactic information about the verb “to eat”:

နာမ်-စားသည်- နာမ်; စားသည် – Present Simple

Often only partial parsing is sufficient to get to the syntactic structure of the source sentence and to map it onto the structure of the target sentence.

3rd: translate Myanmar words into Kayah Li [1]:

ကျွန်တော် (နာမ်) => ဗဌ (နာမ်)

စားသည် (ကြိယာ) => ဠ်ဂဲဠ်ဠ် (ကြိယာ)

ထမင်း (ကံ) => ဧဥဌဇ (ကံ)

4th: Mapping dictionary entries into appropriate inflected forms (final **generation**) [2]:

ကျွန်တော်　　ထမင်း　စားသည်

(Myanmar)

ကျွန်တော်　　စားသည် ထမင်း

(Kayah Li)

ꤜꤢ꤭　　ꤢꤩꤏꤢꤧꤣ　　ꤘꤢꤨꤤꤣ

IV. Myanmar Language

The Myanmar language, Burmese, belongs to the Tibeto-Myanmar language group of the Sino-Tibetan family. It is also morphologically rich and agglutinative language. Myanmar words are postpositionally inflected with various grammatical features [13].

A. Grammatical Hierarchy in Myanmar

The grammatical hierarchy is a useful notion of successively included levels of grammatical construction operating within and between grammatical levels of analysis. This hierarchy is generally assumed in this study as a heuristic principle for the purposes of laying a foundational understanding of Burmese grammatical units and constructions. This hierarchy is a compositional hierarchy in which lower levels typically are filler units for the next higher level in the hierarchy. Fig.1 shows the grammatical hierarchy from the lowest level to the highest [13].

Text
Paragraph
Sentence
Clause
Phrase
Word
Morpheme

Figure 1 Grammatical Hierarchy

V. Sentences of Myanmar Language

There are two kinds of sentences according to the syntactic structure of Myanmar language. They are simple sentence (SS) and complex sentence (CS). Fig. 2 describes the syntactic structure of Myanmar language.

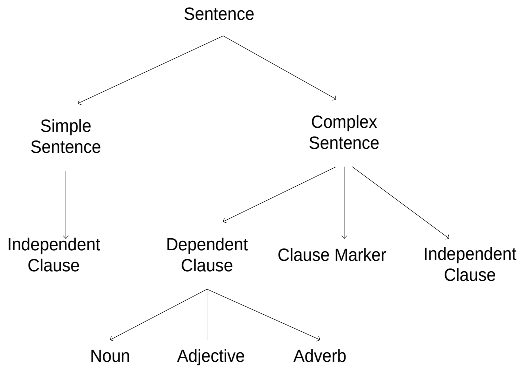

Figure. 2 Syntactic Structure

VI. Kayah-Li *Language*

In the western Kayah-Li national script there are 24 consonants, 10 digits and 9 vowels, 1 diphthong and 3 tones [6, 7].

A. Kayah-Li Alaphbet

Table I, II, III, and IV describe Kayah-Li consonants, numbers, vowels and vowels diacritics, and tone marks respectively [3, 4, 5].

TABLE I

KAYAH-LI CONSONANTS

Consonants/ꤜꤢꤨꤕꤜꤢꤨꤔꤢꤧꤗꤢꤨꤌꤢꤨ					
ꤊ [ka] က	ꤋ [kha] ခ	ꤌ [ga] ဂ	ꤍ [nga] င	ꤎ [sa] စ	ꤏ [sha] ဆ
ꤐ [zha] ဇ	ꤑ [nya] ည	ꤒ [ta] တ	ꤓ [hta] ထ	ꤔ [na] န	ꤕ [pa] ပ
ꤖ [pha] ဖ	ꤗ [ma] မ	ꤘ [da] ဒ	ꤙ [ba] ဘ	ꤚ [ra] ရ	ꤛ [ya] ယ

ꤜ	ꤝ	ꤞ	ꤟ	ꤠ	ꤡ
[la]	[wa]	[tha]	[ha]	[va]	[ca]
လ	ဝ	သ	ဟ	ဗ	ကျ

TABLE II

KAYAH-LI NUMBERS

Numbers / ꤗꤢꤒꤢꤩ									
꤀	꤁	꤂	꤃	꤄	꤅	꤆	꤇	꤈	꤉
0	1	2	3	4	5	6	7	8	9

TABLE III

KAYAH-LI VOWELS AND VOWELS DIACRITICS

Vowels/ꤢꤚꤟꤢꤩ				Vowels diacritics/ꤗꤢꤒꤢꤩ				
ꤢ	ꤣ	ꤤ	ꤥ	ꤢꤦ	ꤢꤧ	ꤢꤨ	ꤢꤩ	ꤢꤪ
[a]	[ɣ]	[i]	[o]	[ɯ]	[ɛ]	[u]	[e]	[ɔ]
အာ့	အဝ်	အီး	အို့	အု	အဲ	အူး	အေး	အော

TABLE IV

KAYAH-LI TONE MARKS

Tone Symbols/ꤜꤟꤢꤩꤜꤟꤢꤥ			
꤭	꤬	꤫	¯
high tone	low tone	mid tone	mid-high tone

B. Syllable Structure and Special Features

Kayah-Li has four syllable types: V, CV, CCV and CCCV (where V stands for vowel and C for consonant).

An important difference between Kayah and English is that there are no closed syllables in Kayah other than a glottal stop. This means that a final consonant like the ‘p’ in ‘stop’ never occurs after a vowel in Kayah. This is a challenge for the Kayah person who is learning English and will require special attention to pronouncing the final consonants in English.

The syllable types, abbrevations, sentence structure, sentence type in Kayah-Li are described in Table V, VI, VII, VIII respectively [7].

TABLE V

SYLLABLE TYPES

Syllable Type	Phonetic	Kayah Li	English Meaning
V	a˧	ꤢ꤭	he/she/it
V	o˥	ꤥ꤭	drink
CV	tə˧	ꤒꤣ꤭	one
CV	bi˥	ꤙꤤ	when
CCV	plu˥	ꤕꤜꤢꤨ	punch
CCV	kru˧	ꤊꤚꤢꤨ꤭	firewood
CCCV	klja˥	ꤊꤜꤛꤢ	road
CCCV	prja˧	ꤕꤚꤛꤢ꤭	quickly

TABLE VI

ABBREVIATIONS

Abbreviations	English	Kayah Li
adj	adjective	ꤊꤟꤢ꤫ꤒꤟꤢ꤭ꤒꤟꤢꤩ
adv	adverb	ꤊꤟꤢ꤫ꤗꤟꤢꤩ꤆ꤢ꤭
dem	demonstrative	ꤊꤟꤢ꤫꤃ꤟꤢꤩꤜꤟꤒꤣ꤫
der	derivative	ꤊꤟꤢ꤫ ꤢꤚꤟꤒꤣ꤫
n	noun	ꤊꤟꤢ꤫ꤢꤧꤗꤟꤝꤢ꤭
num	number	꤃ꤟꤢ
phr	phrase	ꤜꤢ꤭ꤢꤚꤟꤥꤓꤢꤨ
Part	particle	ꤜꤟꤢ꤭ꤊꤟꤢ꤫ꤓꤢꤨ
Prep	preposition	ꤊꤟꤢ꤫ꤢꤟꤢꤨ꤭
Pron	pronoun	ꤊꤟꤢ꤫ꤢꤧꤗꤟꤝꤢ꤭ ꤃ꤢ꤭ꤒꤟꤥ
V	verb	ꤊꤟꤢ꤫ꤗꤟꤢꤩꤒꤟꤢꤧ

TABLE VII

SENTENCE STRUCTURE

ဝါကျတည်ဆောက်ပုံ/ ꤊꤟꤥꤜꤟꤢꤩꤊꤜꤢ꤭ꤟꤢꤩ꤆ꤢꤧ

Subject	Verb	Object	
ꤔꤟꤢꤧ သင် ကတ္တား	ꤡꤟꤣ သွားသည် ကြိယာ	꤆ꤢꤒꤢꤩ ဘယ်မှာလဲ ကံ	
ꤠꤢ꤭ ကျွန်ုပ်	ꤡꤟꤣ သွားသည်	꤃ꤣ သို့	ꤢꤩꤢ꤭꤇ꤢꤧꤜꤟꤢꤧ꤮ မန္တလေး

ကတ္တား	ကြိယာ	ဝိဘတ်	ကံ
ꤠꤢ꤭ ကျွန်ုပ် ကတ္တား	ꤢꤧ꤬ စားသည် ကြိယာ	ꤘꤤ꤬၊ ထမင်း ကံ	

TABLE VIII

SENTENCE TYPES

ဝါကျအမျိုးအစား/ ꤢꤧ꤬ꤗꤢꤧ꤬ꤟꤢ ꤕꤢꤓꤟꤢ

ဝါကျအမျိုးအစား		
၁	အဆိုပြုဝါကျ/ꤢꤧ꤬ꤗꤢꤧ꤬ꤟꤢ	ꤠꤢ꤭ ꤔꤟꤢ ꤢꤧ꤬ ꤙꤢ꤬ꤓꤢ꤬ꤜꤢ ꤢꤧ꤬ꤢ
		ကျွန်ုပ်မန္တလေးသို့ သွားမည်။
၂	အငြင်းဝါကျ/ꤢꤧ꤬ꤗꤓꤢ ꤜꤢꤧ꤬ꤗꤢꤧ꤬ꤟꤢ	ꤟꤢ꤬ꤚꤢ꤬ꤒꤢ꤬ ꤠꤢ꤭ꤞꤢꤔꤟꤢ ꤢꤧ꤬ ꤞꤟꤢꤗꤜꤢꤙꤢ꤬ ꤚꤢ
		ဒီနေ့ ကျွန်တော်ကျောင်းမသွားချင်ဘူး။
၃	မေးခွန်းဝါကျ/ꤢꤧ꤬ꤗꤚꤢꤐꤢ꤬	ꤗꤢ꤬ꤗꤢꤧ꤬ꤚꤢ꤬ꤛꤢ ꤔꤢꤧ꤬ꤢꤟꤢ꤬ꤢꤧꤚꤢ
		သင်ဘာဖြစ်လို့ကျောင်းမလာတာလဲ။
၄	အမိန့်ပေးဝါကျ/ꤢꤧ꤬ꤗ ꤢꤧ꤬ꤚꤢꤧ꤬	ꤒꤢ꤬ꤟꤗꤜꤒꤢ꤬ꤞꤢ꤬ꤔꤟꤢ꤬ ꤓꤢ꤬ꤔꤟꤢꤔꤢꤧ ꤢꤧ꤬ ꤢ꤬ꤐꤜꤢ꤬ꤚꤢ꤬ꤛꤢ ꤚꤢ
		စာသင်ချိန်မှာမည်သူမျှအပြင်မထွက်ရ။
၅	ခွင့်တောင်းဝါကျ/ꤢꤧ꤬ꤗ ꤐꤓꤒꤚꤢꤧ꤬	ꤢꤟꤢꤔꤢꤧ꤬ꤟꤢ꤬ ꤠꤢ꤭ ꤢꤧ꤬ ꤢ꤬ꤐꤢ꤬
		ကျွန်ုပ်ကို အထဲဝင်ခွင့်ပြုပါ။
၆	တောင်းပန်ဝါကျ/ꤢꤧ꤬ꤗ ꤐꤓꤒꤞꤢꤢꤧ꤬	ꤠꤢ꤭ꤢꤧ꤬ꤟꤢ꤬ꤚꤢꤧ꤬ꤔꤢ ꤒꤢꤧ꤬ꤢꤧ꤬ ꤚꤢ꤬ꤐꤢ
		ကျွန်တော် နဲနဲနောက်ကျသွားတယ်။
၇	အာမေဍိတ်/ꤢꤧ꤬ꤗ ꤙꤟꤢ ꤢꤧ꤬	ꤐꤢ ! ꤗꤢꤧ꤬ꤢ ‿ !
		အို့ အမေရေ့။

C. Sentences Types

There are four sentence types. They are:

1. Declarative sentences that have a positive meaning are called affirmative sentences and a negative meaning are called negative sentences.
2. Imperative sentences are to give commands, warnings, suggestions, or advice. Imperative sentences can also be used to make a request.
3. Interrogative sentences are to ask a question to get information.
4. Exclamatory sentences are to express strong emotions [16].

D. Tenses

The following tenses are referenced from the storis in Kayah-Li [10, 11, 12].

Present Tense

ကျွန်တော် ထမင်း စားသည် (Myanmar)

ကျွန်တော် စားသည် ထမင်း (Kayah Li)

va˥ e˧ di˧

ꤠꤢ꤭ ꤢꤧ꤬ ꤘꤤ꤬

Past Tense

ကျွန်တော် ထမင်း စားခဲ့သည် (Myanmar)

ကျွန်တော် စား **ခဲ့သည်** ထမင်း(Kayah Li)

va˥ e˧ tʰʷɯᵊ˧ hɔ˥ di˧

ꤠꤢ꤭ ꤢꤧ꤬ ꤓꤟꤢ꤬ ꤢꤧ꤬ ꤘꤤ꤬

Future Tense

ကျွန်တော် ထမင်း စားမယ် (Myanmar)

ကျွန်တော် စားမယ် ထမင်း

(Kayah Li)

va˥ e˧ di˧ pa˧

ꤠꤢ꤭ ꤢꤧ꤬ ꤘꤤ꤬ ꤕꤢ꤬

Present ContinuousTense

ကျွန်တော် ထမင်း စား နေသည်

(Myanmar)

ကျွန်တော် စား နေသည် ထမင်း (Kayah Li)

va˥ e˧ pa˧ di˧ prɛ˥

ꤠꤢ꤭ ꤢꤧ꤬ ꤕꤢ꤬ ꤘꤤ꤬ ꤕꤚꤢꤧ꤭

Present Perfect Tense

ကျွန်တော် ခုလေးတင် ထမင်း စား ပြီးပြီ (Myanmar)

ကျွန်တော် စား ပြီးပြီ ထမင်း ခုလေးတင် (Kayah Li)

va˥ e˧ tʰʷɯᵊ˧ hɔ˥ di˧ kʰo˧ ɲạ˥

ꤠꤢ꤭ ꤢꤧ꤬ ꤓꤟꤢ꤬ ꤢꤧ꤬ ꤘꤤ꤬ ꤋꤪ꤬ꤑꤢ꤭

VII. Conclusion

Rule-based machine translation is described for Myanmar to Kayah-Li language. Also presents Myanmar to Kayah-Li for translating well-structured Myanmar sentences into well-structured Kayah-Li sentences. In this study, the nature of the Kayah-Li grammartical rules are similar to the nature of English grammartical rules. In the future, this paper can be upgraded for Myanmar and Kayah-Li morphological features, lexicon and bidirectional machine translation approach.

Acknowledgment

Our sincere thanks go to U Moe Zet Aung Lay, Loikaw University for linking the Kayah-Li language experts.

We are very grateful to U Khu Phya Reh, Tutor, History Department, Loikaw University for his patience, guidance and necessary instruction of Kayah-Li Language during the period of the paper. We would like to express our special thanks to U Myar Doo for his invaluable guidance and expertise for Kayah-Li Language.

References

[1] Basic Grammar to learn and study Kayah Language.
[2] Burmese-Kerenni, Lexicon.
[3] Grade 1 (ꤊꤢꤧꤛꤢ ꤁, ꤢꤊꤟꤊꤜꤛꤊꤢꤛꤢ) , Kayah Text Book, Kayah National Literacy and Culture Committee, 2016.
[4] Grade 2 (ꤊꤢꤧꤛꤢ ꤂, ꤢꤊꤟꤊꤜꤛꤊꤢꤛꤢ) , Kayah Text Book, Kayah National Literacy and Culture Committee, 2016.
[5] Grade 3 (ꤊꤢꤧꤛꤢ ꤃, ꤢꤊꤟꤊꤜꤛꤊꤢꤛꤢ) , Kayah Text Book, Kayah National Literacy and Culture Committee, 2016.
[6] Htay Reh, "English-Kayah Reverse Dictionary", ISBN 9781-1-4675-6167-9, in cooperation with Payap University/Linguistics Institute, publish for Kayah National Literacy and Culture Committee.
[7] "Kayah Literacy", National Race and Culture Department, University for the Development of the National Races of the Union, Sagaing, May, 2013.
[8] Mr.Uday C. Patkar et.al, "Transformation of Multiple English Text Sentences To Vocal Sanskrit Using Rule Based Technique", International Journal of Computers and Distributed Systems, Vol. No.2, Issue 1, December 2012.
[9] R. Harsha Wardhan, "Rule Based Machine Translation System for English to Malayalam Language", Centre for Excellence in Computational Engineering and Networking, December, 2011.
[10] Stories in Kayah-Li, Book 1 (ꤜꤊꤧ ꤁), published by Kayah National Literacy and Culture Committee, December, 2013.
[11] Stories in Kayah-Li, Book 2 (ꤜꤊꤧ ꤂), Kayah-Li Translation Committee, 2011.
[12] Stories in Kayah-Li, Book 3 (ꤜꤊꤧ ꤃), Kayah-Li Translation Committee, 2011.
[13] W. W. Thant, T. M. Htwe, and N. L. Thein, "Function Tagging for Myanmar Language", Inernational Journal of Computer Applications, Vol. 26, No. 2, July, 2011.
[14] Y. Park and H. Kwon, "Korean Syntactic Analysis using Dependency Rules and Segmentation ", Proceedings of the Seventh International Conference on Advanced Language Processing and Web Information Technology(ALPIT2008), Vol.7, pp.59-63, China, July 23-25, 2008
[15] https://en.wikipedia.org/wiki/Rule-based_machine_translation
[16] http://www.really-learn-english.com/types-of-sentences-in-english.html