Thai to Khmer Rule-Based Machine Translation Using Reordering Word to Phrase

Sukchatri Prasomsuk and Puthy Mol

***Abstract*—In this paper, an effective machine translation system from Thai to Khmer language on a website is proposed. To create a web application for a high performance Thai-Khmer machine translation (ThKh-MT), the principles and methods of translation involve with lexical base. Word reordering is applied by considering the previous word, the next word and subject-verb agreement. The word adjustment is also required to attain acceptable outputs. Additional steps related to structure patterns are added in a combination with the classical methods to deal with translation issues. PHP is implemented to build the application with MySQL as a tool to create lexical databases. For testing, 5,100 phrases and sentences are selected to evaluate the system. The result shows 89.25 percent of accuracy and 0.84 for F-Measure which infers to a higher efficiency than that of Google and other systems.**

***Index Terms*—Thai Khmer translation, machine translation (MT), rule based, pattern-based.**

I. Introduction

Association of Southeast Asian Nations (ASEAN) consists of ten countries with various cultures and languages. Thailand and Cambodia are included in ASEAN, and the eastern border of Thailand is adjacent to Cambodia. Therefore, efficient communication is significant for international relations between these two countries. Cambodian natives have Khmer as a national language while formal language in Thailand is Thai. The linguistic differences of Thai and Khmer in both writing and speaking contribute to a translation barrier. For instance, since Thai language has been adapted partly from Pali, Sanskrit and Old Khmer, Thai vocabulary is relatively diverse. Thai language also contains complex orthography and relational markers. Furthermore, standard written Thai is complicated due to various combinations of syllabic alphabets, which consists of 44 basic consonants, 21 vowel symbols and 4 tone diacritics, applied under the rule that all diacritics appear in front of, above or below the consonants. Furthermore, Thai syntax has a noun classifier system as well as conforms to a basic sentence structure called subject–verb–object (SVO) with a horizontal and vertical writing direction from left to right and from top to bottom, respectively. Similarly, Khmer contains 33 consonants, 23 dependent vowels and 15 independent vowels; however, no tone is presented. Due to the linguistic differences, current Thai-Khmer translation systems have scarcely achieved complete and accurate outputs. Moreover, the existent systems have rarely been created and developed. There is also a shortage of intellectuals who are competent in both languages and able to convey knowledge for creating a system of translation. As a result, the improvement of the Thai-Khmer translation system has been disrupted. Document translation between Thai and Khmer which requires high accuracy has consequently encountered difficulties. To solve the issues, machine translation (MT) from Thai to Khmer language requires development.

The proposed system in this paper implements translation techniques including rule-based algorithm with verification of sentence patterns to improve translation quality. The overview operation of the translation system is to input a Thai language text in a web application and then convert it into a desired output in Khmer. A lexical analyzer is first applied in the process to divide Thai sentences or phrases into individual syllabic words so that the separated words are analyzed and processed in the following steps resulting in Khmer sentences.

Manuscript received August 15, 2016; revised December 1, 2016.

The authors are with School of Information and Communication Technology, University of Phayao, 56000, Thailand, (e-mail: sukchatri.pr@up.ac.th, skchatri@hotmail.com, puthymol@gmail.com).

II. Related and Previous Works

There have been many attempts to research on machine translation between Thai and other languages. English-Thai machine translation was developed in 1998 with regard to the sentence-based technique which combines the rule-based and the example-based method to establish a system for English to Thai sentence translation [1]. However, the research result of performance evaluation and comparison was not indicated. In 2012, a technique called generalized patterns is presented to improve machine translation from Japanese to Thai language [2]. The method was compared to the others implemented in Google and Bing translators by executing 3,107 Japanese sentences in testing. F-Measure score was applied to assess performance of the translator.

Machine translation between Khmer and other language has also been researched. One of the studies selected Moses DoMY CE, which is statistical machine translation (SMT), as a tool to create an online system for English - Khmer translation based on Python, XML and HTML language in 2013 [3]. There is also research in 2014 on developing a French-Khmer dictionary called 'MotàMot' [4]. In 2015, an automatic machine translation was created to provide translation between Khmer and other 20 languages by using three statistical methods: the phrase-based approach, the hierarchical phrase-based approach and the operation sequence model (OSM) as well as selecting BLEU and RIBES to evaluate translation quality [5].

There is, furthermore, research specifically on Thai-Khmer machine translation. For example, Thai - Khmer machine translation on a website has been developed based

DOI: 10.7763/IJCTE.2017.V9.1142

on Java (JSP) and SQL (Appserv) with 4,000 words from a Thai - Khmer dictionary as a database [6]. In testing, 212 sample sentences have been processed, and the result has shown 72.16% of accuracy which is higher than that of Google translator. In 2014, the rule-based machine translation (RBMT) combined with statistical methods was recognized to be widely applied in automated translation [7]. The technique has shown the potential to improve translation between Thai and Khmer. Even though such classical technique is applied, the research has rarely a result with high performance.

III. Background of Thai to Khmer Translation

Sentences in Thai and Khmer language are similarly formed; on the other hand, ordering and semantic structure are different. With regard to the existent methods, the newly presented one for the proposed system is expected to balance between advantages and disadvantages of the classical techniques and be straightforward for implementation. In this paper, a process to translate Thai to Khmer language is composed of six main steps including

1) Input process: reading Thai text into the system from a website screen,
2) Word segmentation: applying LexTo[1] and the longest matching approach to divide Thai sentences into words,
3) Word search: retrieving data from the database of Thai-Khmer dictionary to find a matched-meaning word in Khmer for each Thai word,
4) Boundary check: considering a boundary of each Thai word such as conjunction, verb, adjective and surrounding nouns to inspect parts of speech,
5) Pattern verification: examining Thai sentence patterns by using the rule-based algorithm, and
6) Khmer word rearrangement: reordering Khmer words in phrases or sentences.

To build a Thai-Khmer dictionary for testing in this paper, approximately 37,052 Thai words from the Royal Institute Dictionary (RID, 1999) are translated according to the existent Thai-Khmer dictionary [8]. In the process of examining word boundaries, patterns and various conditions of grammar rules are taken into account to solve translation mistakes.

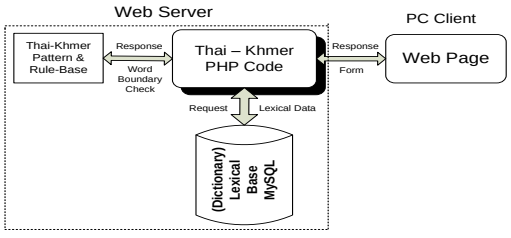


Fig. 1. Architecture of Thai-Khmer Translation on Web.

The classification of machine translation architecture which is regularly implemented on client-web server for online translation is the direct model shown in Fig. 1. The direct machine translation architecture transforms a source language sentence (Thai) into a target language sentence (Khmer). Besides, the proposed system applies the indirect architecture which is demonstrated as a diagram in Fig. 2. A sample screen of the program is also provided Fig. 3.

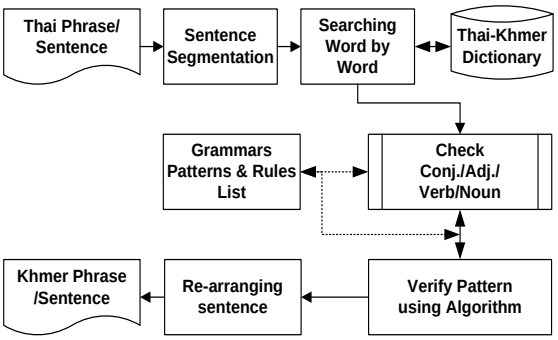


Fig. 2. Proposed system.

Fig. 3. Screen of application
(Available on http://www.ict.up.ac.th/skchatri/translate.thai-khmer/).

IV. Methodology and Proposed Algorithm of Reordering

In general, Thai and Khmer sentences are sorted verbatim. Regarding to the verbatim characteristics of these two languages, the classical algorithm of word reordering could appear to be a proper tool to cope with phrase and sentence arrangement. On the other hand, the reordering method is unable to suit all cases of input phrases and sentences since the reordering could cause translation mistakes. The analysis to deal with the issue has consequently become essential. The verification process is included in the proposed system to investigate errors; in addition, simple approaches to examine the previous word, the next word, and noun and verb positions are used to attain accurate outputs. Implementing pattern-based machine translation also alleviates the translation issue although the method is not novel. The technique based on patterns is also applied to assist translation due to its reputation of promoting translation performance. In this paper, patterns for the method are designed according to Thai and Khmer grammatical structures. The process to deal with the translation issue is arranged into four steps as follows.

1) Morphological analysis
2) Concept of pattern matching
3) Search for proposed patterns
4) Word rearrangement and translation

In the first step, LexTo software is applied to separate each word in a sentence from the others so that the morphology of

[1] LexTo is a Thai word segmentation program which was developed in 2006 by Human Language Technology Laboratory, National Electronics and Computer Technology Center (NECTEC), Thailand.

an input sentence is analyzed. Next, positions of noun (n.), verb (v.), adverb (adv.), adjective (adj.), conjunction (con.) and interjection (int) are considered to acquire a concept of pattern matching with regard to SVO sentence structure. Sample sentences are demonstrated as follows.

A simple sentence:

Example: I eat rice.

Thai sentence		Khmer sentence
S + V + O		S + V + O
ฉัน + กิน + ข้าว	=	ខ្ញុំ + ញុំ + បាយ
cʰăn + kin+ kʰâaw		kɲom + ɲam + bay

A sample sentence with color term or color perception [9]:

Example: This car is red.

Thai sentence		Khmer sentence
S + (V) + O		S + (V) + O
รถยนต์ + คันนี้ + สีแดง	=	រថយន្ត + នេះ + ពណ៌ក្រហម
rót ˈyon + kʰan-níi+ sĭi ˈdɛɛŋ		rʊət yʊən + nih + poa krahɑam

A sample sentence: The counter unit without subject (S):

Example: There are five eggs.

Thai sentence		Khmer sentence
S + V + O + unit(s)		S + V + O + unit(s)
+ มี + ไข่ + 5 + ฟอง	=	+ មាន + ស៊ុត + ប្រាំ + គ្រាប់
+ mii + kʰày + hâa + fɔɔŋ		+ mien+sut+pram+kroap

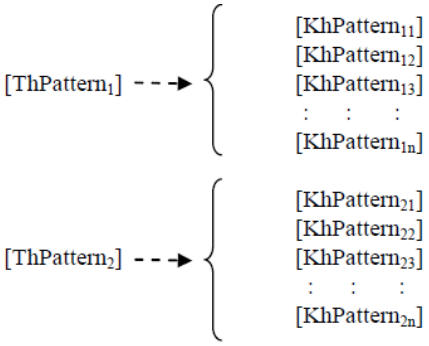
In the third step is to search for the proposed patterns in the input Thai sentence so that the output words are appropriately used in the Khmer sentence. Mapping between grammatical structures of Thai and Khmer language is undertaken. Khmer sentence patterns are then converted into proper forms through mapping algorithms.

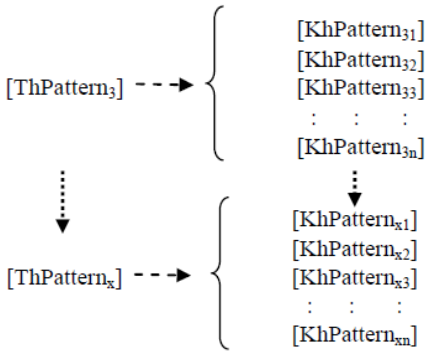
To explain the mapping process, let Thwd{x} be a Thai word and Khwd{x} be a Khmer word where x is an index of a word in the sentence. Then the pattern mapping is defined as follows.

Thai Sentence Pattern X: [ThPattern$_x$]
Thwd{1} + Thwd{2} + Thwd{3} + … + Thwd{ x_n}

is mapped to
Khmer Sentence Pattern X: [KhPattern$_x$]
Khwd{1} + Khwd{2} + Khwd{3} + … + Khwd{ x_n}

It has been found that one Thai sentence pattern is possibly mapped to more than one Khmer sentence patterns as shown below.





After mapping sentence patterns, all proposed patterns are retrieved from the corpus. For testing in this paper, only frequently used patterns are considered. The proposed patterns are searched, and the search consequently causes changes in the output construction. Sample patterns and the related algorithms are exemplified as follows.

Note: ✖ = incorrect ordering words, ✓ = Correct

A. Sample 1: [ThPattern$_1$] - - → [KhPattern$_{11}$]

Thwd{1} + Thwd{2} + Thwd{3} + … + Thwd{ x_n }

↓ ↓ ↓

Khwd{1}+Khwd{2}+(non, Khwd{3-1} or Khwd{3-2}) + … + Khwd{ x_n } [KhPattern$_{11}$]

Algorithm1.1:
IF((Thwd{2}=”number || noun || date || adv || verb || adj”) || (Thwd{4}=”number || noun || adj || verb || date”) || (Thwd{2}=”noun || adj” && Thwd1!=”verb”)) THEN
Thwd{3} is replaced by Khwd{3-2}
ELSE Thwd{3} is translated as Khwd{3-1}

The algorithm is applied if a Thai phrase contains words (Thwd{3}) such as คน, หน้า, ด้วย, กัน, นั่ง, ให้น่า, ทาง, ได้เปรียบ, จะ, แค่, ดู, ใหม่, เหมือน, ถึง, ต่อ, แก่, ใน, หยุด, ก้าว, มากกว่า, เมื่อ, ตอน, มาก, กับ, etc.

Example1.1.1: To verify a Thai word “คน” (kʰon) = person, people, human, man

If the word “คน” (kʰon) is in a position following any other words in the sentence, the Khmer word “មនុស្ស” (mɔnuh) is replaced by “ទាក់” (neak). The example sentences are provided below.

Thai: “ผมมีพี่น้องสามคน” (I have three brothers.)
(pʰŏm-mii-pʰîi-nɔ́ɔŋ-săam-kʰon), and

Khmer: ខ្ញុំមានបងប្អូនបីមនុស្ស ✖
(kɲom - mien -bɑɑŋ- pʔoon - bəy -mɔnuh)
ខ្ញុំមានបងប្អូនបីទាក់ ✓
(kɲom - mien -bɑɑŋ- pʔoon - bəy - neak)

Example1.1.2: To verify a Thai word “นั่ง” (nâŋ) = sit

If the word “นั่ง” (nâŋ) is in a position before a noun (*n.*), the Khmer word “អង្គុយ” (ʔɑŋkuy) is replaced by “នៃ” (cih). The example sentences are demonstrated as follows.

Thai: “เขาจะนั่งรถไฟไปทำงาน” (He will go to work by train.)
(kʰăw - jà - nâŋ - rót ˈfay - pay - tʰam ˈŋaan), and

Khmer: គាត់ទិញអង្ករយកទៅធ្វើការ ×

(koat - niŋ - ʔɑŋkuy - rʊət pləəŋ - tiv - tvəə-kaa)

គាត់ទិញដំរះយកទៅធ្វើការ ✓

(koat - niŋ - cih - rʊət pləəŋ - tiv - tvəə-kaa)

Algorithm1.2:

IF((Thwd{1}="ยัง || หา || ค้นหา || ไม่ || หัวใจ || โมง") || (Thwd{2}="ยัง || หา || ค้นหา || ไม่ || หัวใจ || โมง") || ((Thwd{1}="กำลัง" || Thwd{2} ="กำลัง") && Thwd{4}!="noun")) THEN Thwd{3} is translated as Khwd{3-2} ELSE Thwd3 is translated as Khwd{3-1}

The algorithm is used in a case that a Thai phrase contains words (Thwd{3}) such as อยู่, ไม่, เจอ, สบาย, เต้น, เย็น, etc.

Example1.2.1: To verify a Thai word "อยู่" (yùu) = is, am, are, was, were, be

If the word "อยู่" (yùu) follows another Thai word "กำลัง" (kamˈlaŋ) = …ing , the Khmer word "នៅ" (niv) is removed from the sentence. The example sentences are shown below.

Thai: "กำลังเขียนจดหมายอยู่ครับ" (writing a letter sir.)

(kamˈlaŋ-kʰĭan-jòt ˈmăay-yùu-kʰráp)

Khmer: កំពុងសរសេរសំបុត្រនៅបាទ ×

(kɑmpuŋ-sɑɑ seɛ-sɑmbot-niv-baat)

កំពុងសរសេរសំបុត្របាទ ✓

(kɑmpuŋ-sɑɑ seɛ-sɑmbot-baat)

Algorithm1.3:

IF(((Thwd{1}="verb") || (Thwd{2}="verb")) || ((Thwd{1}="verb || noun || adv") || (Thwd{2}="verb || noun || adv") && Thwd{4} !="noun || verb")) THEN Thwd{3} is translated as Khwd{3-2}
ELSE Thwd{3} is translated as Khwd{3-1}

The algorithm is implemented when a Thai phrase contains words (Thwd{3}) such as ทำไม, ถูก, จน, เอง, etc.

Example1.3.1: To verify a Thai word "ทำไม" (tham ˈmay) = why, for what

If Thwd{3} = "ทำไม" (tham ˈmay) is in a position after a verb (v.), Khwd{3-1} = "ហេតុអ្វី" (haet ʔvəy) is replaced by Khwd{3-2} = "ធ្វើអ្វី" (tvəə ʔvəy). The example sentences are explained as follows.

Thai: "คุณมาที่นี่ทำไม" (Why did you come here?)

(kʰun-maa-tʰîi ˈnîi-tham ˈmay)

Khmer: អ្នកមកទីនេះហេតុអ្វី ×

(neak-mɔɔk-tii nih-haet ʔvəy)

អ្នកមកទីនេះធ្វើអ្វី ✓

(neak-mɔɔk-tii nih-tvəə ʔvəy)

B. *Sample 2:* [ThPattern$_1$] - - → [KhPattern$_{12}$]

Thwd{1} + Thwd{2} + Thwd{3} + … + Thwd{ x_n }

Khwd{1}+Khwd{2}+Khwd{4}+Khwd{3}+Khwd{5} + … + Khwd{ x_n }

Algorithm2.1:

IF(Thwd{3}=="number || noun") THEN Khwd{3} and Khwd{4} is swapped position
ELSE Khwd{3} and Khwd{4} is not swapped

The algorithm is for the case that a Thai phrase contains words (Thwd{4}) such as โมง and ละ.

Example2.1: To verify a Thai word "โมง" (mooŋ) = o'clock

If the word "โมง" (mooŋ) follows a number (of time indications), the Khmer word "ព្រាំបួនម៉ោង" (pram buən-maoŋ) is replaced by drow eht "ម៉ោងព្រាំបួន" or swap the position with that of the word "ម៉ោង-ព្រាំបួន". In this case, the example sentences are provided below.

Thai: "พรุ่งนี้เริ่มต้นเก้าโมงเช้า" (Tomorrow, start at nine o'clock.)

(pʰrûŋ ˈníi-rə̂əmˈtôn-kâaw- mooŋ -cʰáaw)

Khmer: ថ្ងៃស្អែកចាប់ផ្តើមព្រាំបួនម៉ោងព្រឹក ×

(tŋay sʔaek-cap pdaəm-pram buən-maoŋ-prik)

ថ្ងៃស្អែកចាប់ផ្តើមម៉ោងព្រាំបួនព្រឹក ✓

(tŋay sʔaek-cap pdaəm-maoŋ-pram buən-prik)

C. *Sample 3:* [ThPattern$_1$] - - → [KhPattern$_{13}$]

Thwd{1} + Thwd{2} + Thwd{3} + … + Thwd{ x_n }

Khwd{1}+Khwd{3}+Khwd{4}+Khwd{2}+Khwd{5} + … + Khwd{ x_n } [KhPattern$_{13}$]

Algorithm3.1:

IF(Thwd{3}=="num") THEN Khwd{2}, Khwd{3} and Khwd{4} should be reorder as Khwd{3}+Khwd{4} +Khwd{2} ELSE do not reorder phrase

The algorithm is applied for a Thai phrase consisting of a word (Thwd{2}) อีก.

Example3.1: To verify a Thai word "อีก" (iik) = more

If the Thai word Thwd{2} = "อีก" (iik) is in front of any other words in the sentence, the Khmer phrase is required to be reordered. The example sentences are shown below.

Thai: "อีกสองเดือนฉันจะไปกัมพูชา"

(iik -sɔ̆ɔŋ-dɯan-cʰăn-jàʔ-pay-kam.pʰuuˈcʰaa)

(Next two month I will go to Cambodia.)

Khmer: ទៀតពីរខែខ្ញុំនឹងទៅកម្ពុជា ×

(tiət-pii-kʰae-kɲom-niŋ-tiv-kampuʔcie)

ពីរខែទៀតខ្ញុំនឹងទៅកម្ពុជា ✓

(pii-kʰae-tiət-kɲom-niŋ-tiv-kampuʔcie)

Reordering words and translating are in the final step to diminish the translation issue. After the pattern mapping is completed, around 37,000 words from Thai – Khmer dictionary database are retrieved to match each word which is then rearranged to be in a proper position. As a result, a Khmer sentence is attained as the output.

V. Performance Evaluation

The proposed system is assessed for translation

performance from Thai to Khmer by sentences from various sample documents as the input. In the testing process, the total phrases and sentences of 5,100 were employed: 212 sentences from Learning of Thai – Cambodian Language (Chan Rithy (ច័ន្ទ ឫទ្ធី), 2001), 170 sentences from Conversational Thai - Khmer (Seiw Som Aeoun (ស៊ីវ សំអឿន)), 150 sentences from Teaching of Conversational Thai - Cambodian (Lirm Peng (លឹម ផេង), 2003), 189 sentences from Dictionary of Idioms and Phrases for Professional Sentence Writing (Sean Cemeron Beatt (เปรมจิต บีท), 2012) and the others from a handbook, Fundamental Khmer 2 [10] and websites. Sample phrases and sentences are shown in Table I.

TABLE I: SAMPLE OF PHRASE/SENTENCES FOR TESTING

English/ Thai/ Khmer
Are you married? คุณแต่งงานหรือยัง khʰun-tɛ̀ŋ ˈŋaan-rʉ̌ʉ-yaŋ អ្នករៀបអាពាហ៍ពិពាហ៍ឬនៅ neak-riəp-ʔaapie-piʔpie-rɨɨ nɨv
What newspaper do you read? คุณอ่านหนังสือพิมพ์อะไร khʰun-àan-naŋ.sʉ̌ʉˈpʰim-aˈray អ្នកអានកាសែតអ្វី neak-ʔaan-kaasaet-ʔvəy
I'm sorry to let you waiting. ผมขอโทษที่ต้องให้คุณรอ pʰǒm-kʰɔ̌ɔˈtʰôot-tʰîi-hây-kʰun-rɔɔ ខ្ញុំសុំទោសដែលត្រូវឲ្យអ្នករងចាំ kɲom-som tooh-dael-trəv-ʔaoy-neak-rɔɔŋ cam
I will go to a party. ผมจะไปงานเลี้ยง pʰǒm -ja-pay-ŋaan-líaŋ ខ្ញុំនឹងទៅពិធីជប់លៀង kɲom-nɨŋ-tɨv-piʔtʰii-cup liəŋ

Three translation systems including Google translator [11], Chhun's translation system and the proposed system in this paper are assessed through translating the sample phrases and sentences. The translated outputs of each translation system are categorized into three groups consisting of accuracy (correct), acceptance (acceptable) and mistake (wrong).

According to 5,100 Thai sentences selected for testing, the proposed system is able to translate 4,083 words correctly (80.06%), reach the acceptable level of translation for 469 sentences (9.189%) and produce errors only in 548 sentences (10.75%). The total translation accuracy of the proposed system becomes 89.25 % which is a sum of its accuracy and acceptance value. On the other hand, Chhun's translation contributes to 3,590 correct sentences (70.38%) which is less than those of the proposed system, 658 acceptable sentences (12.9%) and 857 mistakes (16.81%). Google translation also achieve less accuracy compared to the proposed system: 1,067 correct sentences (20.9%). whereas it acquires 798 acceptable sentences (15.64%) and 3,230 mistakes (63.34%), respectively, higher than those of the proposed one.

TABLE II: F-MEASURE RESULTS FOR THAI INTO KHMER TRANSLATION

Translation Methods	Precision	Recall	F-Measure
Google	0.57	0.21	0.31
Chhun	0.84	0.70	0.76
Proposed System	0.89	0.80	0.84

Moreover, performances of all systems are compared with regard to system precision, recall and efficiency by implementing F-measure[2] as shown in Table II.

The result in Table II reveals that the proposed system attains the highest score in all evaluations: the precision is 0.89, the recall is 0.80 and the efficiency (F-Measure) is 0.84.

VI. CONCLUSION

The methodology in this paper is presented for creating Thai to Khmer machine translation system by using syntactic and semantic analysis to transform and structure patterns as well as implementing the rule-based translation. The presented processes can also simplify compound sentences into simple ones based on predefined sentence structures. The previous word, the next word and the subject-verb agreement are also considered. In addition, switching with more suitable words, reordering words and adjusting output sentences are also performed with regard to Thai and Khmer grammar. As a result, the proposed system is apparently able to improve the quality of source texts and translated outputs as well as assist Thai-Khmer language learners. Nevertheless, a larger amount of sample sentences in the corpus than that which is currently applied in the proposed system is necessary to achieve higher performance in Thai-Khmer translation. Furthermore, the larger dictionary database as well as the higher diversity of sample sources would be added to the process. Other methods or tools would also be considered to develop Thai-Khmer translation in future research.

ACKNOWLEDGMENT

This research was funded by School of Information and Communication Technology, University of Phayao. We would like to show our appreciation to Cambodian students who have assisted the research by correcting Khmer words, phrases and sentences. We are also immensely grateful to other researchers for their supports in this research project.

REFERENCES

[1] K. Chancharoen, N. Tannin, and B. Sirinaovakul, "Sentence-based machine translation for English-Thai," in *Proc. The 1998 IEEE Asia-Pacific Conference on Circuits and Systems*, Chiangmai Plaza Hotel, Chiangmai, Thailand, 24-27 Nov., 1998.

[2] P. Sae-Tang and A. Prayote, "Japanese-Thai machine translation with generalized patterns," in *Proc. 7th International Conference on Computing and Convergence Technology (ICCCT)*, Seoul, Korea, 3-5 Dec., 2012, pp. 987-992.

[3] S. Jabin, S. Samak, and K. Sokphyrum, "How to translate from English to Khmer using Moses," *International Journal of Engineering Inventions*, vol. 3, issue 2, pp. 71-81, Sep. 2013.

[4] M. Mangeot, "Mot àMot project: Conversion of a French-Khmer published Dictionary for building a multilingual lexical system," in *Proc. Languages Resources and Evaluation Conference*, pp. 1024-1031.

[5] Y. K. Thu, V. Chea, A. Finch, M. Utiyama, and E. Sumita, "A large-scale study of statistical machine translation methods for Khhmer language," in *Proc. The 29th Pacific Asia Conference on Language, Information and Computation*, Shanghai, China, October 30-November 1, 2015, pp. 259–269.

[6] C. Chhun, "Thai- Khmer simple sentence translation on Web," in *Proc. the 1st RMUTL Chiangrai National Conference, (RCCON)*, Thailand, 2015.

[7] G. Malézieux, A. Bosc, and V. Berment, "Rule-based machine translation (RBMT) with statistical knowledge," in *Proc. 25th International Conference on Computational Linguistics*, Dublin, Ireland, August 23-29, 2014, pp. 50–54.

[2]The F-measure was derived by van Rijsbergen (1979), It is a measure of a test's accuracy

[8] Y. Vanavong and R. Saravut, *Thai-Khmer Dictionary*, Nokothom Book Shop, Cambodia, pp. 10-123, 2012.
[9] S. Wattanapongsakul, "Color in Khmer : Perceprion and grammatical construction," Thesis, Silpakorn Univ., Thailand, ISBN: 974-464-290-4, 2005, pp. 44-52.
[10] P. S. Loy, *Fundamental Khmer 2*, Ramkhamhaeng University, 2014, pp. 1-50.
[11] Google web site, Google Translater. (2016). [Online]. Available: http://translate.google.com/

Sukchatri Prasomsuk was born in Lopburi, Thailand, in November 1963, and achieved the D.Ling. degree in computational linguistics in 2011 from CERLOM (Centre d'études et de recherches sur les littératures et les oralités du monde, EAD 4124) at INALCO (Institut national des langues et civilisations orientales), Paris, France and graduated in M.Eng. in information technology at RMIT (Royal Melbourne Institute of Technology), Melbourne, Australia. His first degree is in B.Sc. in mathematics at Ramkhamhaeng University, Thailand in 1985.

He is currently working as a lecturer and the head of Information Technology Subject in the School of Information and Communication Technology, University of Phayao, Thailand. Prior to that, he worked as a computer academician in the Department of Science Service, Ministry of Science and Technology, Thailand. After that since 2002, he has been as a lecture in Mae Fah Luang University, Chiangrai, Thailand. In 2006, he has joined with Franco-Thai Research Project in Computational Linguistics research, responsibility in Automatic Language Translation until finished the project in 2010. To be a member of LaLIC (Langage Logique Informatique, Cognition), 2006-2010, Université Paris-Sorbonne. His research interests are in robotics, networking, natural language processing, and at the present focus on Asian languages translation, including any tool for creating a system.

Dr. Prasomsuk gained a scholarship award of the Royal Thai government to study of master degree in 1993. Since 2005 received scholarships from France government and Naresuan university, Thailand in 2007 for doctoral research in France.

Puthy Mol was born in Takeo, Cambodia, on September 3rd 1994. He currently a bachelor degree student in information technology, faculty of Information and Communication Technology, University of Phayao, Phayao, Thailand.

Mr. Mol has received an undergraduate scholarship award of Her Royal Highness Princess Maha Chakri Sirindhorn in 2013.