Thai Poetry Translation to English with Tuning Module

Sajjaporn Waijanya, Anirach Mingkhwan
Faculty of Information Technology
King Mongkut s University of Technology
North Bangkok, Bangkok, Thailand
sajjaporn.w@gmail.com, anirach@ieee.org

ABSTRACT: *The complexities of poetry translation always challenges researchers in the field of NLP (National Language Processing) and MT (Machine Translation). This paper will focus on the Thai poetry type "Klonn-Pad" and aim to translate into English keeping terms of prosody. The results of translation between MT Forward and Dictionary Base Forward will be compared as well as between MT Backward and Dictionary Base Backward. BLEU (Bilingual Evaluation Understudy) metric will be used to compare between reference and candidates of English results. The abbreviations are as follows, T1 (G) translated by Google API, T1 (D) translated by Dictionary Base, T3 (G) translated by Google API with Tuning and T3 (D) translated by Dictionary Base. The BLEU score of T1 (G) equaled 0.344, T1 (D) equaled 0.888, T3 (G) equaled 0.674 and T3 (D) equaled 0.889. The survey results regarding native Thai people opinions to this case study, sorted by highest score first was T3 (D), T1 (D), T3 (G), and T1 (G). Based on this study, it can be concluded that the machine translators are unable to provide good enough translation for Thai poetry compared with the dictionary base. Thus, it is necessary and reasonable that Thai poetry needs to have a special dictionary with tuning algorithms to provide better results. Moreover, another reason is that poetry may not relate completely to syntax structures.*

Keywords: Poetry Translation, Poem Translation, Thai Poem, Thai Poetry

1. Introduction

Poetry Translation is a very challenging job for the translation field. Poetry translators must understand the poem as a whole in one language and then rewrite to another language. It is very important that output of poetry translation should still be poetry. The Poetry.org has defined the meaning of poetry which is Poetry (ancient Greek: ποιεω (poieo) = I create) is an art form in which human language is used for its aesthetic qualities in addition to, or instead of, its notional and semantic content[1]. Poetry has a different prosody (syntax/form) in different countries. This can refer to different linelength (number of syllables), rhyme, meter and patterns. Thai poetry consists of five major types which are Khlong, Chann, Khapp, Klonn and Raai[1]. There are many famous poets and world-class poetries such as William Shakespeare with his famous poem "*Julius Caesar and Hamlet*". A

---

[1]Types of Thai poetry

Chinese poet of the Tang Dynasty, Li Po (also known as Li Bai, Li Pai, Li T ai-po, and Li T ai-pai). Li Po wrote occasional verse and poems about his own life. His poetry is known for its clear imagery and conversational tone. Thai Poet, Phra Sunthorn Vohara, known as Sunthorn Phu, (26 June 1786–1855) is Thailand s best-known [2] royal poet. In 1986, the 200th anniversary of his birth, Sunthorn Phu was honored by UNESCO as a great world poet. His Phra Aphai Mani poems described a fantastical world and poetry type is Klonn-Pad. Although many researchers have translated poetry before, this study takes a new approach by employing Dictionary base Machine Translation (MT) to achieve meaningful translation.

Referring to a previous published paper by the author of this research "*Thai Poetry in Machine Translation*" [3], it was found that MT is unable to understand poetry patterns, and had difficulty with original words and sentences. The reason for that is the operating principle of MT itself. They use Phrase based methods while translating from the original to another language. But Thai Poetry can be written in incomplete sentences. Moreover, Thai words especially words in poetry are very complex. The reason why poets use more difficult words is a matter of felling, the beauty of these words and also the beauty of the poetry itself.

2. Related Work

This research will focus on the comparison of machine translator and dictionary bases. To validate output of language translation, "the backward translation" with BLEU (Bilingual Evaluation Understudy) will be used in this researche. Related work now follows below.

2.1 Poetry Translating as Expert Action
Francis R. Jones published a book called "Poetry Translating as Expert Action: Processes, Priorities and Networks" [4]. This book shows how experienced poetry translators translate poems and bring them to readers, and how they not only shape new poems, but also help communicate images of the source culture. It uses cognitive and sociological translation studies methods to analyses real data, most of it from two contrasting source countries, the Netherlands and Bosnia. Case studies, including think-aloud studies, analyses how translators translate poems. In interviews, translators explain why and how they translate. However this book did not include Thai poetry in the case studies.

2.2 Machine translation system for patent documents combining rule-based translation and statistical post editing applied to the NTCIR-10 Patent MT Task
Terumasa ehara from Yamanashi Eiwa College published a Machine translation system for patent documents combining rule-based translation and statistical post editing applied to the NTCIR-10 Patent MT Task [5]. This paper uses both advantages of rule-based method and statistical methods and they use an inverse translation method. The procedure of their method is described in Figure 1.
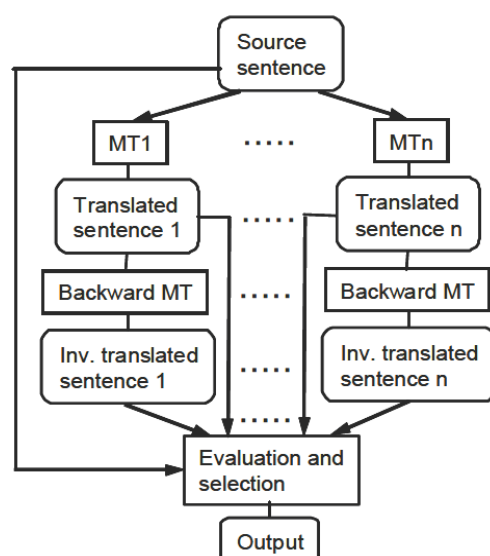

Figure 1. Translation selection method using inverse translation (Source: Proceedings of the 10th NTCIR Conference, June 18-21, 2013, Tokyo, Japan, p. 336)

These inversely translated sentences 1 to n are compared with the original source sentence and the system calculates evaluation scores for each inversely translated sentences. The system selects translated sentence i as an output, where i is the sentence number that inversely translated sentence i has the best score. They made a preliminary experiment comparing three criteria: sentence level BLUE, RIBES and IMPACT to be the evaluation method.

3.3 Backward Word Translation: Lexical vs. Conceptual Mediation or 'Concept Activation' vs. 'Word Retrieval'?
Angeliki Salamoura and John N.Williams published "BackwardWord Translation: Lexical vs. Conceptual Mediation or „Concept Activation vs. „Word Retrieval ?" [6]. This study addresses the question of whether backward translation (L2-to-L1) exploits a different mental process from forward translation (L1-to-L2). They investigate the difference between the two translation-directions in the case of „concept activation versus „word retrieval by employing three matching tasks in L2 in combination with an L2-to-L1 translation task within a group of dominant Greek-English bilinguals. In the number and word matching tasks subjects are asked to decide whether pairs of digits and words are the „ same or „ different , while in the semantic matching task whether pairs of words are semantically related or not (e.g.vegetable-carrot vs. window-tree). They use the correlation analysis of the results obtained within and between the two experiments and the questionnaires. In conclusion part, they mention it seems that semantic retrieval is a necessary component of bilingual lexical processing reflecting a conceptual level interface between the two languages.

3. Tools and Methodology

3.1 Machine Translator
A Machine Translator MT (Machine translation) is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one natural language to another. MT has two major types and consists of Rule Base Machine Translation and Statistical Machine Translation Technology. Referring to a previous published paper by the author of this research, a statistical machine translator by API from the Google Translator will be used to translate Thai poetry for the part of MT.

3.2 Standard Dictionary
The dictionary is used as a tool in this paper, 4 types of standard dictionaries as follows.

• Thai to Thai dictionary, Thai to Thai dictionary is used to search "*direct synonyms*" words from ROYIN dictionary [7].

• Thai to English dictionary, NECTEC's Lexitron-2 Dictionary [8].

• English to Thai dictionary, English-Thai: NECTEC's Lexitron-2 Dictionary.

• English-Thai: HOPE dictionary and English-Thai: Nontri Dictionary [9].

Dictionaries will be applied to the part of Translation, Backward Translation and Tuning. Moreover, collections of synonyms words from each dictionary are used to create a new dictionary.

3.3 Synonyms Dictionary for Poem
Normally, synonymous words can be searched from a dictionary but unfortunately not for poem words. Thus, it is necessary to correct the meaning of poem and keep the prosody (syntax). Moreover, feeling and beautiful words are important in every poem. It is necessary to collect and create a new database of those words as poem words corpus according to a Standard dictionary as direct synonyms cannot support all poem words.

3.4 Thai Word Segmentation
Lexto-wordseg [10] is an open source software for word segmentation. This Thai word segmentation uses the longest matching and rule based. The results from Lexto can be classified as separated unknown words, known words, the ambiguous, English words, numbers, and special characters.

3.5 BLEU (Bilingual Evaluation Understudy)
BLEU (Bilingual Evaluation Understudy) [11] is an algorithm for evaluating the quality of text which has been translated from machine translator from one natural language to another language. Quality is considered to be the correspondence between a machine's output and that of a human. BLEU uses a modified form of precision to compare a candidate translation against

multiple reference translations. Equation of BLUE is showed in equation (1) and (2).

$$BLEU = BP \cdot exp \left(\sum_{n=1}^{N} w_n \, log \, p_n \right) \qquad (1)$$

When P_n: Modified n-gram precision, Geometric mean of $p1, p2,..pn.$

BP: Brevity penalty (c = length of MT hypothesis (candidate), r = length of reference)

$$BP = \begin{cases} 1 & if \, c > r \\ e^{(1-r/c)} & if \, c \leq r \end{cases} \qquad (2)$$

In this baseline, $N = 4$ and uniform weights $wn = 1/N$ are used.

4. Our Proposed Approach

4.1 System Flow Chat
To study Thai Poetry in Machine Translation, Thai poetry Klonn-Pad[2] 2 Bot[3] (8 Wak[4]) was used as an input of processing as shown in Figure 2. And five modules have been designed for translating Thai poetry to English which keeps the prosody term.
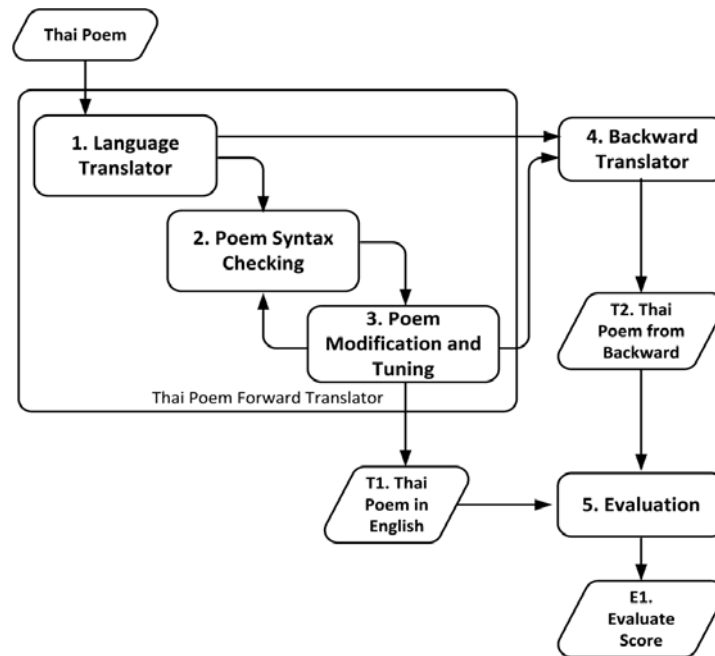

Figure 2. System flow chat of Thai to English poetry translation

4.2 Rhyme Prosody for Thai Poetry Klonn-Pad (2 Bot) and Case Study
The research objective attempts to have the poetry translation of Klonn-Pad in Thai to the English language. The rhyme and prosody term as shown in Figure 3. Thai poem name "Deūxnchāy"[5] from the book "Oh jảo dòk máai oie"[6] has been selected for the case study as shown in Figure 4 which presents mapping of Rhyme words with "underline" symbol.

---

[2] Klonn-Pad: Thai poetry in the most common eight-syllable variety in line.

[3] one stanza of poetry.

[4] a group of syllables written together in one line, the number of syllable depend upon the type

[5] Gaillardia, name of flower in Thai we call "Deūxnchāy" means moon-shine
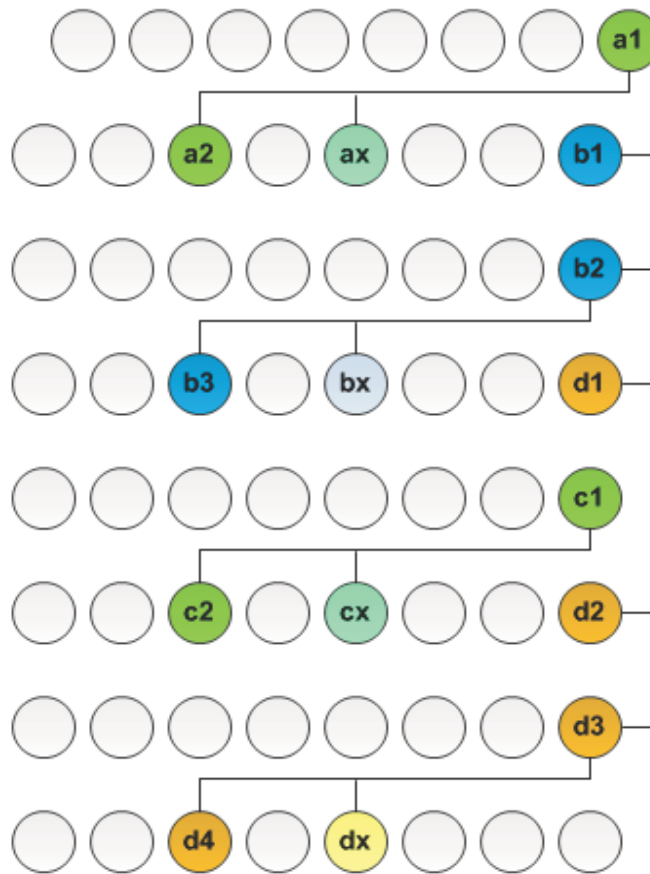
[6] Name of book, its mean Beautiful flowers

Figure 3. Structure of Klonn-Pad Rhyme and prosody (2 Bot)

Figure 4. Thai poem name "Deūxnchāy" from Book "Oh jảo dòk máai oie"

4.3 Module 1: Language Translator

Referring to the purpose of this study to compare translation results from MT and Dictionaries. Two sub-modules in the Language Translator have been defined. There are 1.1 Google API Machine Translate (MT) and 1.2 Dictionary base Translate (Dictionary). Figure 5 shows a system flow chart and presents the first step.

Output from 1.1 $MT = T1$ (G) is presented in Figure 6. This refers to the author s previous paper for steps of translation. The results from 1.2 Dictionary base $T1$ (D) are presented in Figure 7 and step details of this module are described in Figure 8.
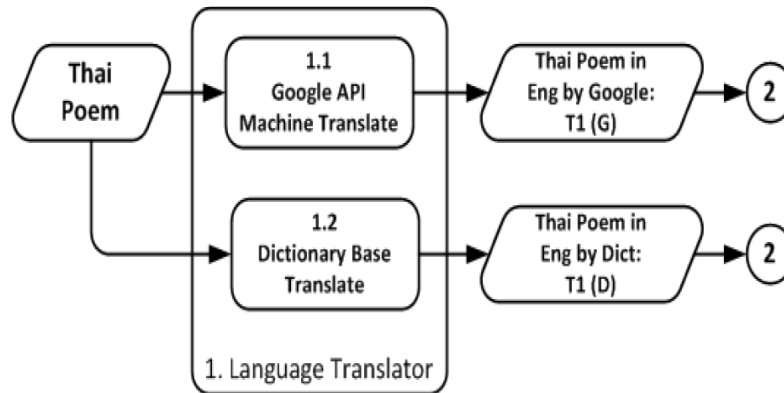
Figure 5. Module 1: Language Translator

1) From Case Study, Translation by Google API: T1 (G)

Herbaceous species are easy to grow

Your beauty is *"Kwai"* mind nostalgic.

Yellow, some red, some elephants *Ospin*.

I washed *Enidain* home all day.

"Kwai" Yes to the first two races.

Synonyms tunable only realistic there.

Particularly fertilizer *"Kwai"* more elven species.

To conjured up randomly rants flattery.

Figure 6. T1(G) Output from Google Translator

2) From Case Study, Translation by Dictionary base: T1 (D)

Be annual grow very easy

Name beautiful outstanding "Deunchay" heart yearn

Color yellow, Red really beautiful

Flower beautiful thick forest bloom for all day.

"Deunchay" yes will race moonlight shine

Only name be same melodious truthfully that

More get apply fertilizer " Deunchay" more beautiful

Too much invent describe praise.

Figure 7. T1(D) Output from Dictionary Base

System flow chart in Figure 8 describes steps of translation as follows.

• *Module 1.2.1 Separate Wak:* Input (poem) this process will check "*end of line*" which is comparable to "*end of Wak*".

• *Module 1.2.2 Word Segmentation:* In this step the Thai Segment "Lexto" is used to separate Thai words into types and applies color Symbols. Example as "โศภิน"(S̄ophin) and "เฉิดฉิน"(C̄heidc̄hin) will be displayed in red due to it has no meaning in a normal dictionary.
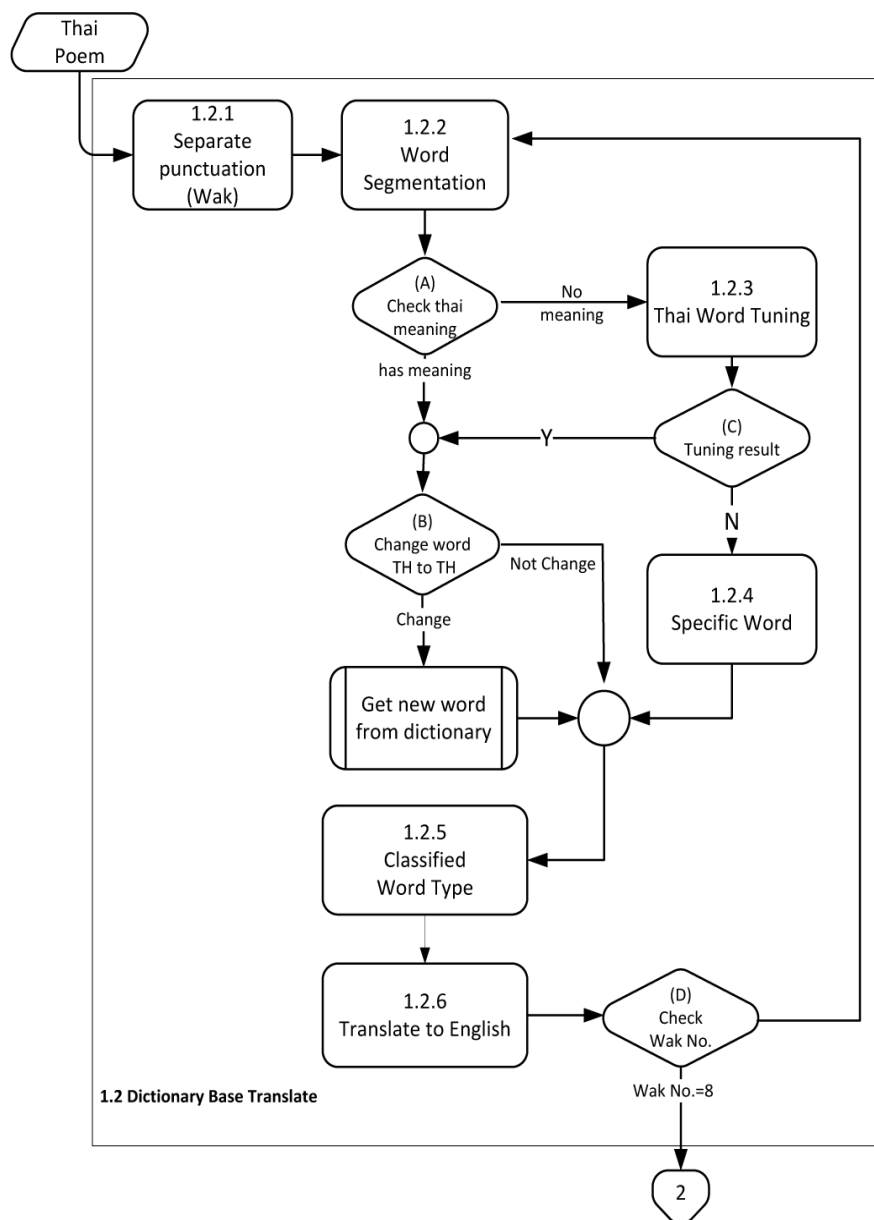
Figure 8. System Flow chart of Language Translator module by Dictionary Base

• **Module 1.2.3 Thai word tuning:** In the case that a Thai words has no direct meaning such as "โศภิน" (S̄ophin) and "เชิดชิน""(C̄heidc̄hin) from Module 1.2.2. but they are Thai words so they can be translated into Thai. After that, they will be translated to Thai synonyms by Synonyms Dictionary which is especially created for Thai poem translating.

• **Module 1.2.4 Specific word:** In the case that a Thai word has no meaning such as "เดือนฉาย" (Deụ̄xnc̄hāy). It is name of flower, it will be defined by type of word "specific word" and not translated to English. But it will be written in terms of phonetic.

• **Module 1.2.5 Classified word type:** In some cases, words may have more than one meaning. The different meanings will depend on the word s type. This step will classify the type of word before translating to a reasonable word.

• **Module 1.2.6 Translate to English:** In this step, the output from the previous step is used to select a reasonable word. If the type of word is specific, it will transform to be phonetic. For normal cases, it will be translated to a English word.

For the example of a Thai word such as "ง่าย" (Ng̀āy). In the module 1.2 condition (A), it s checked for the Thai meaning.

If the Thai meaning is found, the dictionary will be used for translation which is based on the type of word. For example as "ง่าย" (Ngāy) will be translated to "*easy*". The word mapping model is presented in Figure 9.
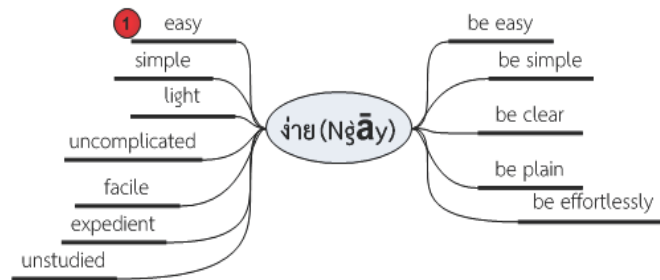

Figure 9. Mapping model of Word "H2"" (Ngy)

4.4 Module 2: Poem Syntax Checking and Module 3: Modification and Tuning

After gaining the output from Module 1 T1 (G) and T1 (D), Rhyme and prosody will be checked by Module 2: Poem Syntax Checking. Any error from Module 2 will be input for Module 3 which tunes the words and syllables from T1 (G) and T1 (D) then the output will be generated by T3 (G) and T3 (D). Step of Syntax Checking described in Figure 10 and Module 3 Tuning and modification described in Figure 11 and steps are as follows.
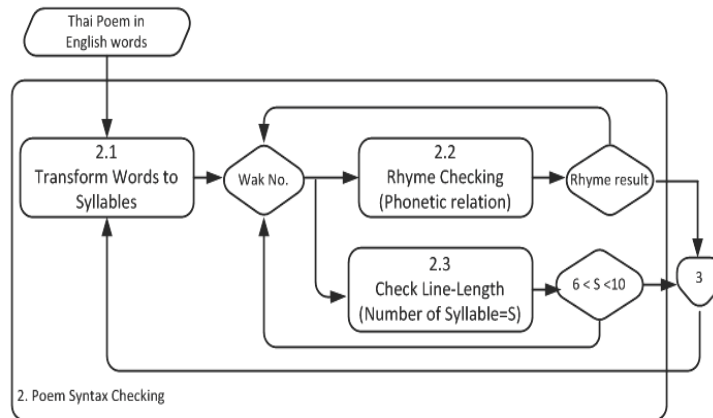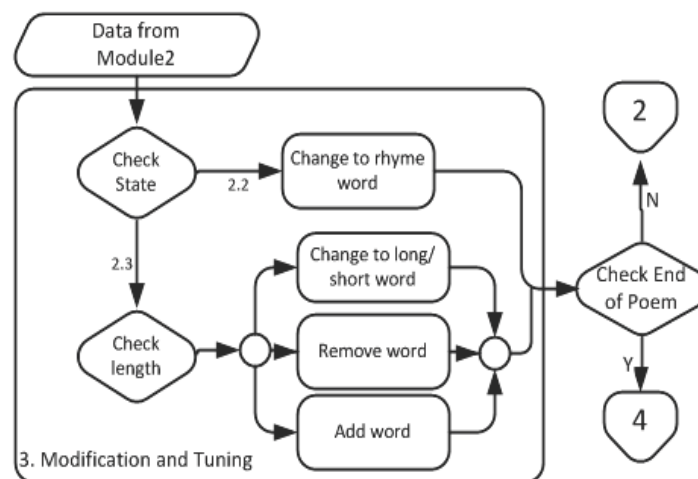

Figure 10. Module 2: Poem Syntax Checking

Figure 11. Module 3: Poem modification and Tuning

Referring to output T1 (G), Rhyme was checked by rule of Klonn-Pad prosody in Figure 3, 100% Rhyme error was found. The example is a1= grow and a2 = beauty, were changed with synonyms words in Module 3.

Another example, words "โสภิน" (Ṣ๐ꝑhⅰn) and "เฉิดฉิน" (C̄heidc̄hin), in module1:T1 (G) google MT changed them in order to $b2$ = "Ospin" and $b3$ = "Eฺ๐idฺฺin" , they are not readable, no meaning and not rhyme with $b1$ = nostalgic. Moreover, $b1$ and $b2$ are not transliterated words but they are Thai words so they can be translated into Thai. After that, they will be translated to Thai synonyms by Synonyms Dictionary which is especially created for Thai poem translating. Figure 12 shows how to gain "*aesthetic*" for $b3$.

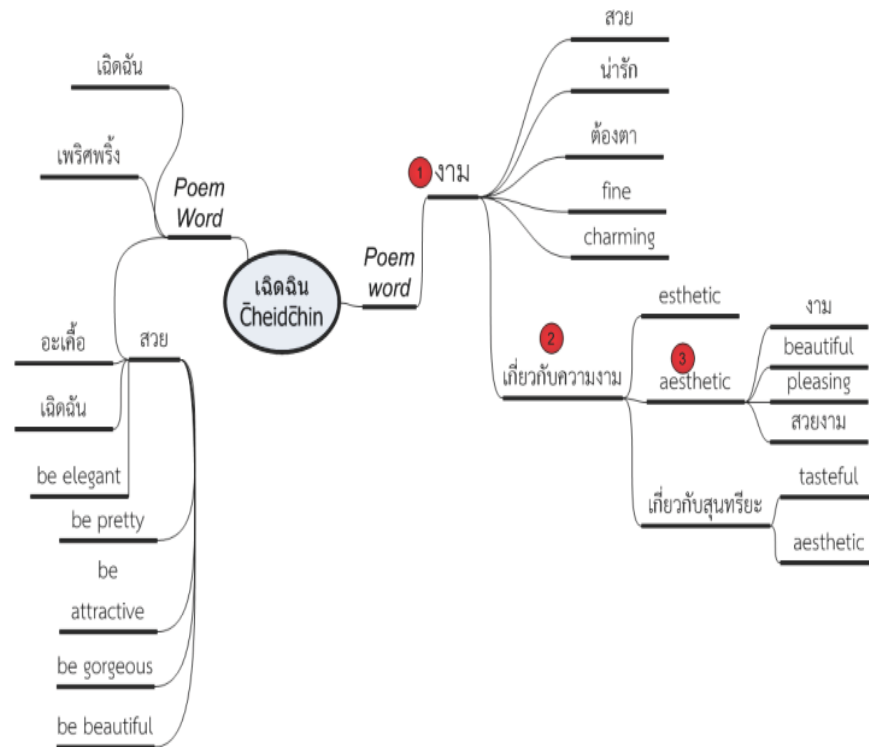

Figure 12. Mapping Poem word by Synonyms Dictionary Poem

4.5 Module 4: Backward Translator

The Backward Translator Module has been created in this paper. Output from this module will be evaluated. The backward translation is a process to ensure the quality of translation. A flow chart of the Backward Translator is shown in Figure 13.
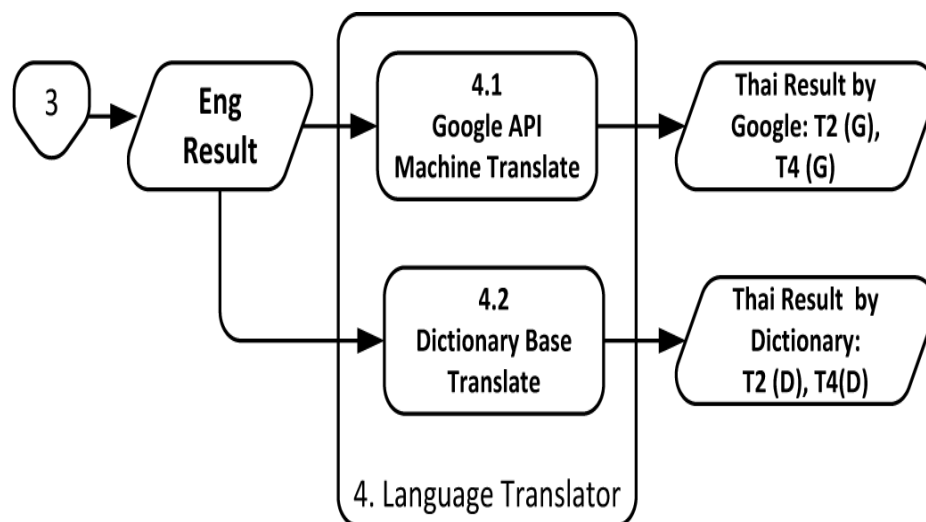

Figure 13. Module 4: Backward Translator

Step of backward translator in this paper will be start from output of Translation (Module3) there are T1 (G), T3 (G), T1 (D) and T3 (D). Module 4 will translate English to Thai and output of this module will be evaluated. The reference of evaluation is original Thai poem "Deūxnchāy".

5. Measurement And Experiment Results

5.1 Evaluate by BLEU Score

To measure and evaluate 4 English poem outputs, the original English poetries are used for references (From the owner of this Thai poetry) which are then compared to 4 outputs translated by Google and Dictionary Base without and with tuning. Example of the calculated BLEU score is shown in Table I and the final score of each English output is shown in Table II.

Reference	Candidate by T3(G)	BLEU score
Seasonal plant very easy grow	Herbaceous species are easy to growing	0.905304
Name melodious "Duenshine" alluring	Your charming is "Duenshine" mind nostalgic	0.860265
Some Yellow some Red be pleasing	Yellow, some red, some elephants magic.	0.956932
All day, blooming flower very pretty	I washed aesthetic home all day.	0.887559
"Duenshine" never compete with moonlight	"Duenshine" Yes to the first two rivalry.	0.839673
Both names sound alike pleasingly	Synonyms tunably only realistic there.	0
More fertilizer "Duenshine" bloom many	Particularly fertilizer "Duenshine" more elven species.	0.946048
Difficulty describe emblazon.	Conjured up randomly rants flattery	0
	Average BLEU Score	**0.674**

Table1. The Example Of Calculated Bleu Score

Candidate Poem	BLEU score
T1(G)	0.344
T1(D)	0.888
T3(G)	0.674
T3(D)	0.889

Table 2. Bleu Score of All Output

5.2 Evaluate by survey from Thai native

All Thai outputs from Poem Translator and Backward Translator are evaluated and given scores by a human. Rang of scores are 0 to 1 and surveyed by 10 native Thai people. Average score of each output is shown in Table III.

Conclusion and Future Work

The output results from comparisons of MT and translation by dictionary uncovered interesting results. However, it was not

Translated Poem	Score	Backward Translator	Score
T1(G)	0.3	T2(G)	0.4
T1(D)	0.5	T2(D)	0.6
T3(G)	0.6	T4(G)	0.5
T3(D)	0.8	T4(D)	0.7

Table 3. Average Score From Human Evaluation

possible to directly use a standard dictionary. The reason is that poem words and the Thai language have a complicated structure. Thus, this paper has presented a synonyms dictionary for poems. This can be used for tuning and modification which should overcome the above problems and improve quality of output in terms of meaning and Klonn-Pad Rhymes prosody also.

The BLEU score in this paper used 1 reference poem and 4 candidate poems after it was translated by MT and a dictionary with and without tuning. The evaluation was via Wakk which calculated the average of Bot. T1(D) and T3(D) gained BLEU 0.888 and 0.889. Both candidates used a dictionary. T1(D) output was achieved before tuning and T3(D) after the tuning step. T1 (G) and T3(G) gained BLEU 0.344 and 0.674. When translated by Google API MT, T1(G) was achieved before tuning and T3(G) after tuning. However, the dictionary is necessary for translation of poetry more accurately.

For backward translation, output from the Dictionary Base was found to be better than MT. This was also backed up by information collected by the survey. Average score arranged by highest first was T4 (D), T2(D), T4(G) and T2(G)

This paper used only 1 case study. This concept must be proven extensively with more poetry. The prosody and meaning of poetry are very important when translated into other languages because poetry reflects art, culture and feelings. The Synonyms Dictionary for poems should be developed further. Moreover, in the future, it should group words by sense and collect words to construct a large corpus to increase accuracy of translation.

Acknowledgment

References

[1] What is poetry, http://www.poetry.org/whatis.htm, Last Access Date : 28-03-2014

[2] P.E.N. International Thailand-Centre Under the Royal Patronage of H.M. The King. Anusorn Sunthorn Phu 200 years, Amarin printing (1986)

[3] Sajjaporn W., Anirach M., (2012). Thai Poetry in Machine Translation.: Proceedings of IC2IT2012, 64-69, Bangkok

[4] Francis R. Jones. (2011). Poetry Translating as Expert Action Arts & Disciplines, John Benjamins Publishing, Netherland

[5] Terumasa EHARA. (2013). Machine translation system for patent documents combining rule-based translation and statistical postediting applied to the NTCIR-10 PatentMT Task. *In*: Proceedings of the 10th NTCIR Conference,335-338, Tokyo, Japan

[6] Angeliki Salamoura, John N.Williams. (2010). BackwardWord Translation: Lexical vs. Conceptual Mediation or, Concept Activation vs., Word Retrieval . *In*: Research Centre for English and Applied Linguistics, University of Cambridge, 31-55

[7] ROYIN dictionary, http://rirs3.royin.go.th/dictionary.asp

[8] LEXiTRON, http://lexitron.nectec.or.th/2009_1/

[9] Nontri dictionary, ftp://ftp.cs.washington.edu/pub/thaisys/BBS

[10] Lexto-Wordseg, http://www.sansarn.com/lexto/

[11] Papineni, K., Roukos, S., Ward, T., Zhu, W.J. (2002) Bleu: a method for automatic evaluation of machine translation. *In*: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 311-318