Thai Sentence Paraphrasing from the Lexical Resource

Krittaporn Phucharasupa and Ponrudee Netisopakul
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand

phucharasupa@hotmail.com, ponrudee@it.kmitl.ac.th

Abstract

Paraphrase generation in any language has gained much attention and importance in the study of Natural Language Processing. Therefore, the focus of this paper is on Thai language paraphrase generation for the sentence level. Six sentence paraphrasing techniques for Thai are proposed and illustratively explained. In addition, the *Thai–sentence Paraphrase Generation (TPG) system* is designed using a lexical resource based system subsequently entitled the *Thai Lexical Conceptual Structure with Thai Lexicalized Tree Adjoining Grammar (TLCS–TLTAG) Resource*.

1   Introduction

For any language, putting the same content in different ways can indicate the richness of the language culture. Since the language is one of the major communication tools in every society, the ability to paraphrase what we want to say or write can also imply the society's civilization.

Paraphrasing techniques for the sentence level and others in several languages have been examined and suggested during the past several years (Stede, 1996; Dras, 1999; Barzilay and Lee, 2003; Pang et al., 2003; Qiu et al., 2006; Ellsworth and Janin, 2007; Zhao et al., 2009; Madnani and Dorr, 2010). These paraphrasing techniques were enormously used in several areas of Natural Language Processing such as Question Answering (Duboue and Carroll, 2006), Machine Translation (Shimohata, 2004; Barreiro, 2008), Summary Evaluation (Zhou et al., 2006) and Textual Entailment Recognition (Marsi et al., 2007; Malakasiotis, 2011).

In Thai language, its writing structure contains no space between words and no full stops between sentences. This could be potential problems in doing research pertaining to Thai computational paraphrasing. Nevertheless, the construction and patterns of Thai sentences have been partially investigated by a number of renown Thai linguists (Vongsantivanit, 1983; Kanchanacheeva, 1996; Thonglor, 2007; Songsilp, 2008; Settanyakan, 2011). Some researchers classified Thai verbs, identified their arguments, as well as recognized their corresponding thematic roles (Wongsiri, 1981; Sungkhavon, 1984; Panthumetha, 2010).

To be able to work on Thai sentence paraphrasing, previous research regarding constructing and paraphrasing sentences in other languages was essential and therefore surveyed (Shimohata, 2004; Barreiro, 2008; Dorr, 1994; Kozlowski et al., 2003; Fujita, 2005). It was subsequently adjusted by (Phucharasupa and Netisopakul, 2011) to fit Thai language more appropriately. Thai sentence paraphrase patterns were categorized into fourteen groups, some of which will be explained and used as examples in this research.

To achieve the goal of automatic paraphrase generation, two critical considerations must be addressed. One is that an appropriate semantic structure of the original sentence must be designed so that it facilitates the automatic system to easily generate paraphrases. The other is that the algorithm must be able to generate syntactically correct paraphrases of the original sentence and these paraphrases must faithfully preserve its original meaning.

381

The focal method for semantic representation of this research is the *Lexical Conceptual Structure (LCS)* associated with each lexical item (Fujita, 2005; Jackendoff, 1990; Dorr and Palmer, 1995) whereas the method of interest for syntactic structure representation is the *Lexicalized Tree Adjoining Grammar (LTAG)* (Joshi, 1999; Palmer and Rosenzweig, 1999) that captures the realization of the lexical item. In addition, the LTAG operations, namely, substitution and adjoining, ensure that the resulting sentence is well-formed. The above two representations, i.e., LTAG and LCS have been utilized to facilitate multilingual generation (Dorr and Palmer, 1995; Netisopakul, 1997).

In this paper, six paraphrasing techniques for generating Thai sentence paraphrases are proposed based collaboratively on LCS and LTAG. This paper is organized as follows. In the next section, the process of *the Thai–sentence Paraphrase Generation (TPG) system* is described in details. In Section 3, how each of the six paraphrasing techniques works is illustratively explained. Then, in Section 4, combinations of the proposed Thai sentence paraphrasing techniques used in some of the fourteen Thai sentence paraphrase patterns are identified along with one particular combination explicitly illustrated in details. In the last section, a conclusion and suggestions of this research are provided.

2   Processes of TPG System

Thai sentence paraphrase generation in the designed TPG system is driven by the semantic input or the *Composed LCS (CLCS)*, that is, the meaning of complex phrases composed from several *Root LCSs (RLCSs)* corresponding to individual words (Dorr, 2001). This TPG system contains three primary processes, namely, the *CLCS Decomposition*, the *Thai LTAG (TLTAG) Selection*, and the *Surface Realization* as illustrated in Figure 1.

In the very first process of the TPG system or the *CLCS Decomposition* process, one CLCS is semantically broken into many elementary LCSs corresponding to each individual word. Each elementary LCS is then normalized into its semantic base form according to the *Thai Lexical Conceptual Structure with Thai Lexicalized Tree Adjoining Grammar (TLCS–TLTAG) Resource*.

In the second process called the *TLTAG Selection*, each semantic base formed LCS is mapped with TLCS part in the TLCS–TLTAG Resource so as to pull out the corresponding TLTAG tree which defines the syntactic structure of the elementary word.

The last process entitled the *Surface Realization* combines all TLTAG trees using the LTAG operations. This process produces syntactically well-formed sentences, each of which can be read off of the leaf nodes of a combined TLTAG tree.
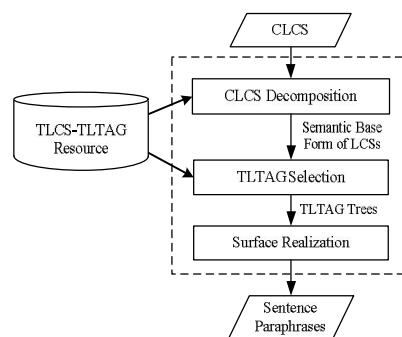

Figure 1: The Architecture of TPG System

The TLCS–TLTAG Resource is designed to assist the TPG system in generating the paraphrases because it encapsulates information necessary for the paraphrase generation process. The information in the TLCS–TLTAG Resource contains the following:

- General information of each Thai word such as the part of speech, the word sub-category, the synonyms, the antonyms, and the definition. For example, the word "เปล่ง ประกาย/*shine*" has "intransitive verb" as its part-of-speech, "Immotion Action" (Sungkhavon, 1984) as its sub-category, "ส่อง ประกาย/*glitter*" as its synonym, "หมอง/*cloud*" as its antonym, and "สะท้อนแสง/*reflect light*" as its definition.

- *Thai LCS* or TLCS semantics corresponding to individual words useful for the *CLCS Decomposition* process and the *TLTAG Selection* process.

- Syntactic structures in the TLTAG portion projected from Thai lexicon items based on the LTAG theory (Joshi, 1999).

The *Surface Realization* of TLCSs can be processed by mapping semantic arguments to the substitution nodes in TLTAGs. Considering an

example drawn from the TLCS–TLTAG Resource shown in Figure 2, a TLCS consists of a category *event*, a predicate *ACT* and its arguments *agent-agt, factitive-fac,* and *cause-cau* of the verb "ทำ/*do*". Each argument is mapped to each corresponding substitution node in the TLTAG of the verb "ทำ/*do*" as illustrated in Figure 2(a) for an affirmative sentence and in Figure 2(b) for a productive causative sentence.

(a) An Affirmative Sentence
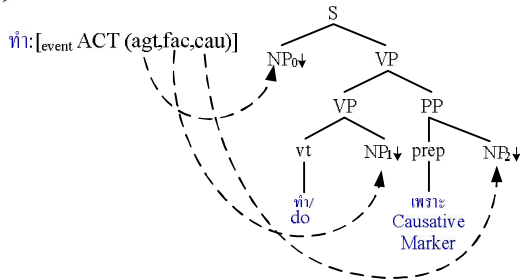
Ex. (a) นักเรียน/*students* ทำ/*do* การบ้าน/*homework* เพราะ/*because of-CausativeMarker* ครู/*teacher*

Students do their homework because of the teacher.

(b) A Productive Causative Sentence
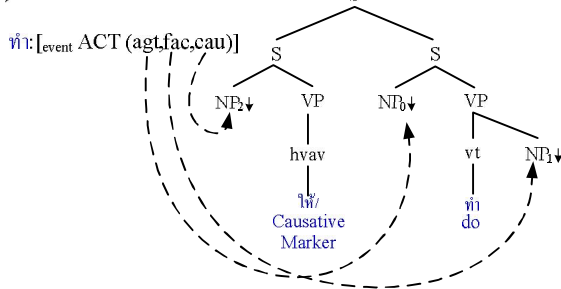
Ex. (b) ครู/*teacher* ให้/*order-CausativeMarker* นักเรียน/*students* ทำ/*do* การบ้าน/*homework*

The teacher orders the students to do their homework.

Figure 2: A Semantic (TLCS) of "ทำ/*do* something activated by a cause" Represented in Different Constructions (TLTAGs)

During the automatic paraphrasing, an initial sentence represented by a CLCS is decomposed into many TLCSs in the *CLCS Decomposition* process. Next, each decomposed TLCS is looked up in the TCLS–TLTAG Resource during the *TLTAG Selection* process to find a mapped TLCS in order to obtain its associated TLTAG. Note that the number of the obtained TLTAGs can be more than one depending on the numbers of the mapped TLCSs. The *TLTAG Selection* process may

therefore result in several surface structures indicated by each TLTAG paired with the mapped TLCS as shown in Figure 2(a) and Figure 2(b). The *Surface Realization* process links the TLCS arguments to the TLTAG empty substitution nodes according to the hierarchical order of the arguments in the thematic roles.

The *TLTAG Selection* and the *Surface Realization* processes are performed based on the fourteen Thai sentence paraphrase patterns previously suggested by (Phucharasupa and Netisopakul, 2011) using the six Thai sentence paraphrasing techniques proposed in this paper and described elaborately in the following sections.

3 Thai Sentence Paraphrasing Techniques

In Phucharasupa and Netisopakul (2011), besides exploring Thai sentence paraphrase patterns from Thai linguistic phenomena, previous research related to the analysis of language constructions and paraphrases was also reviewed. The paraphrase patterns were classified based on Thai verb classes proposed by Sungkhavon (1984). During the classification, it was noticed that one paraphrasing technique was used in several paraphrase patterns and in turn, several paraphrasing techniques could be used in one Thai paraphrase pattern.

Hence, this analysis of paraphrase patterns and techniques gives a total of six Thai sentence paraphrasing techniques to be proposed here. Later in this section, these techniques along with their operating procedures and examples will be described. Out of these six, three techniques including the *Replacement Technique*, the *Movement Technique*, and the *Left-Out/Insert Technique* involve changing individual words or phrases, all by itself. The second group of the proposed paraphrasing techniques includes the *Switching Technique* and the *Promotion/Demotion Technique*. These techniques involve making a change of the words, phrases, or clauses *in pairs*. Finally, the remaining paraphrasing technique called the *Nominalization Technique* changes the structure of the original sentence or phrase.

Throughout this section and the next, the initial sentence to be paraphrased for demonstration purposes of the six paraphrasing techniques is given in the following S_i.

(*S_i*) เขา/*he-Agent* และ/*and-ParallelMarker* เธอ/*she-Dative*

ท่องเที่ยว/*travel-MotionAction* อย่าง/*AdverbMarker*

สนุกสนาน/*joyfully-Quality* ใน/*in-PositionMarker*

กรุงเทพฯ/*Krung Thep-Locative*

He and she travel joyfully in Krung Thep.

In addition, the meaning of S_i is represented in the CLCS form shown in Figure 3 to be used as an input for starting the paraphrase generation processes.
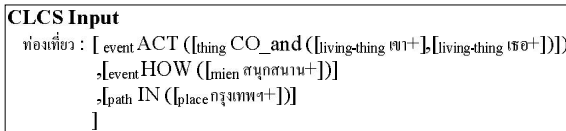
CLCS Input

ท่องเที่ยว : [$_{\text{event}}$ACT ([$_{\text{thing}}$CO_and ([$_{\text{living-thing}}$ เขา+]₋[$_{\text{living-thing}}$ เธอ+])])]

₋[$_{\text{event}}$HOW ([$_{\text{mien}}$ สนุกสนาน+])]

₋[$_{\text{path}}$ IN ([$_{\text{place}}$กรุงเทพฯ+])]

]

Figure 3: The CLCS Form for S_i

CLCS Decomposition

1) Decomposition   2) Generalization

ท่องเที่ยว: ท่องเที่ยว:

[$_{\text{event}}$ACT ([$_{\text{thing}}$CO_and ([$_{\text{living-thing}}$ เขา+]₋[$_{\text{living-thing}}$ เธอ+])])] → [$_{\text{event}}$ACT (agt)]

[$_{\text{thing}}$CO_and ([$_{\text{living-thing}}$ เขา+]₋[$_{\text{living-thing}}$ เธอ+])] → [$_{\text{thing}}$CO_and (agt,par)]

[$_{\text{living-thing}}$ เขา+]   →   [$_{\text{living-thing}}$ เขา+]

[$_{\text{living-thing}}$ เธอ+]   →   [$_{\text{living-thing}}$ เธอ+]

[$_{\text{event}}$HOW ([$_{\text{mien}}$ สนุกสนาน+])]   → [$_{\text{event}}$HOW (man)]

[$_{\text{mien}}$ สนุกสนาน+]   →   [$_{\text{mien}}$ สนุกสนาน+]

[$_{\text{path}}$ IN ([$_{\text{place}}$กรุงเทพฯ+])]   → [$_{\text{path}}$ IN (loc)]

[$_{\text{place}}$กรุงเทพฯ+]   →   [$_{\text{place}}$กรุงเทพฯ+]
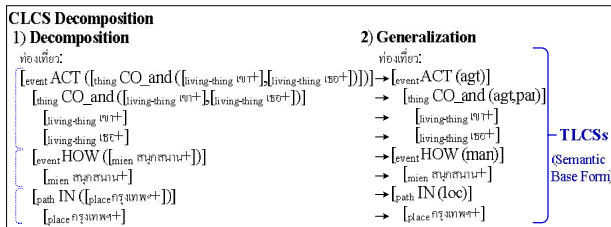
⎫ TLCSs (Semantic Base Form)

Figure 4: The Decomposition of the CLCS for S_i

When the TPG system is triggered, the CLCS input is decomposed into many TLCSs in the *CLCS Decomposition* process. Then TLCSs are normalized into semantic base forms, which will be hereafter called the "TLCSs input", as illustrated in Figure 4. Afterwards, the *TLTAG Selection* and the *Surface Realization* processes will be activated on the TLCSs input for all fourteen Thai sentence paraphrase patterns under the restriction of each pattern using the following six Thai sentence paraphrasing techniques to be described in more details now.

3.1 The Replacement Technique

This *Replacement Technique* makes use of the variety of words having similar meanings. One existing word or phrase in a sentence can then be *replaced* by a new word or phrase with the similar meaning in the same syntactic category without changing its position and its thematic role. Figure 5 also show two types of elementary TLTAG trees, according to the LTAG theory (Joshi, 1999), which correspond to TLCSs of the initial sentence S_i.

An example of using this *Replacement Technique* will be illustrated in the context of one paraphrase patterns, namely, the *Lexical*

Replacement by Its Synonym pattern. Typically, the *TLTAG Selection* process selects all elementary TLTAG trees from the TLCS–TLTAG Resource in which their TLCSs precisely agree with the TLCS input. However, in this case, the *Lexical Replacement* pattern forces the process to specifically choose the trees not just only whose TLCSs are identical to the TLCS input but also whose syntactic structures are the same as that of the TLCS input tree.
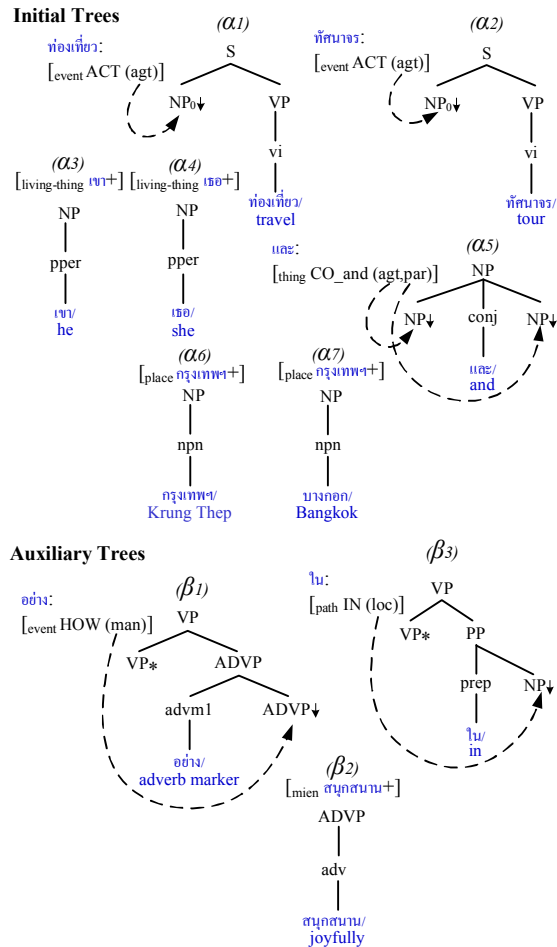
TLTAG Selection

Figure 5: The Elementary TLTAG Trees Corresponding to TLCSs for S_i

Let *α1* in Figure 5 be TLCS and TLTAG of an original decomposed word "ท่องเที่ยว/*travel*" retrieved from the TLCS–TLTAG Resource and let *α2* in Figure 5 be TLCS and TLTAG of another word "ทัศนาจร/*tour*" retrieved again from the TLCS–TLTAG Resource. Since *α2* has both the same TLCS and TLTAG as *α1*, *α2* is thus selected as *a*

384

synonym of *α1*. Similar process can be applied to another original decomposed word "กรุงเทพฯ/*Krung Thep*" and results in the TLCS and TLTAG *α6* whereas *α7* is TLCS and TLTAG for the synonym "บางกอก/*Bangkok*" of this decomposed word.

In the next step, these elementary trees are realized into well-formed surfaces by LTAG's operations shown in Figure 6.

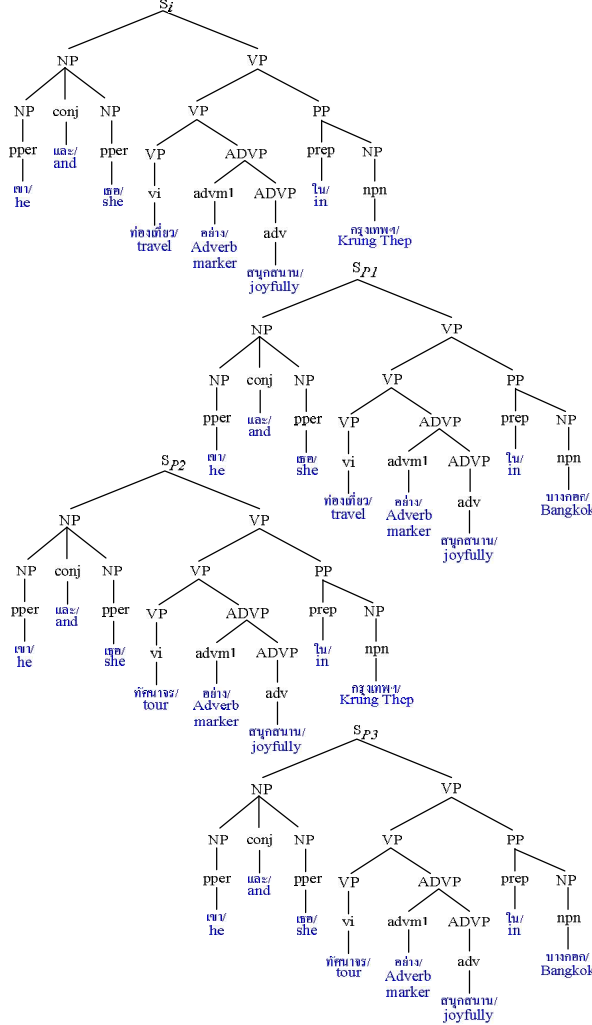Figure 6: The TLTAG Derived Trees for Both S_i and Its Paraphrases (S_{p1}-S_{p3}) Obtained from the Replacement Technique

Each sentence paraphrase can be read off of the leaf nodes of its associated TLTAG derived tree as follows.

(S_{p1}) เขา/*he-Agent*　และ/*and-ParallelMarker*　เธอ/*she-Dative*
　　　ท่องเที่ยว/*travel-MotionAction*　อย่าง/*AdverbMarker*

สนุกสนาน/*joyfully-Quality*　ใน/*in-PositionMarker*
บางกอก/*Bangkok-Locative*
He and she travel joyfully in Bangkok.

(S_{p2}) เขา/*he-Agent*　และ/*and-ParallelMarker*　เธอ/*she-Dative*
　　ทัศนาจร/*tour-MotionAction*　อย่าง/*AdverbMarker*
　　สนุกสนาน/*joyfully-Quality*　ใน/*in-PositionMarker*
　　กรุงเทพฯ/*Krung Thep-Locative*
He and she travel joyfully in Krung Thep.

(S_{p3}) เขา/*he-Agent*　และ/*and-ParallelMarker*　เธอ/*she-Dative*
　　ทัศนาจร/*tour-MotionAction*　อย่าง/*AdverbMarker*
　　สนุกสนาน/*joyfully-Quality*　ใน/*in-PositionMarker*
　　บางกอก/*Bangkok-Locative*
He and she travel joyfully in Bangkok.

3.2　The Movement Technique

In a Thai sentence, the *Movement Technique* is usually used for emphasizing on one constituent over the rest by moving the emphasized constituent to the front of the sentence. Its syntactic category and thematic role remain unchanged (Thonglor, 2007; Songsilp, 2008). For example, this *Movement Technique* is used in the *Direct Object Promotion* pattern of the fourteen Thai sentence paraphrase patterns by moving the direct object to the front of the sentence.

In another example, moving around the negative marker in a sentence can reduce or sometimes increase the negative sense of the sentence and thus make it more or sometimes less polite than the initial sentence as demonstrated in the *Moving Negation Separated from Adjective/Adverb* pattern.

The *Movement Technique* in the *Preposition Phrase Promotion* pattern will be explained here. For the given initial sentence S_i, the *TLTAG Selection* process selects TLTAG elementary trees corresponding to TLCS inputs. These elementary trees are realized into surface strings which contain the preposition modifier of the main verb. Subsequently, the *Moving Technique* will move the entire preposition branch around these three locations, namely, the front of the sentence, right after the main verb, or the back of the sentence, depending on its *promotion/demotion* switch.

In Figure 7, each sentence paraphrase can be read off of the leaf nodes of its associated TLTAG derived tree as follows.

385

(S_{p4}) ใน/*in-PositionMarker* กรุงเทพฯ/*Krung Thep-Locative*
เขา/*he-Agent* และ/*and-ParallelMarker* เธอ/*you-Dative*
ท่องเที่ยว/*travel-MotionAction* อย่าง/*AdverbMarker*
สนุกสนาน/*joyfully-Quality*

In Krung Thep, he and she are joyfully traveled.

(S_{p5}) เขา/*he-Agent* และ/*and-ParallelMarker* เธอ/*you-Dative*
ท่องเที่ยว/*travel-MotionAction* ใน/*in-PositionMarker*
กรุงเทพฯ/*Krung Thep-Locative* อย่าง/*AdverbMarker*
สนุกสนาน/*joyfully-Quality*

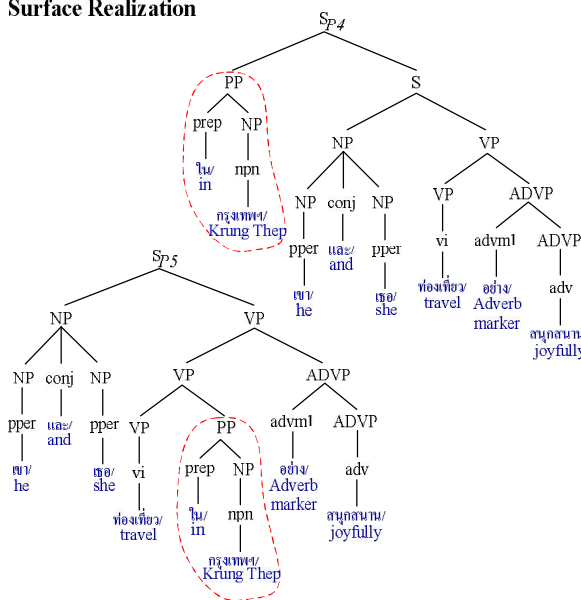He and she travel in Krung Thep that they are joyfully.

Figure 7: The TLTAG Derived Trees for Both S_{p4} and S_{p5} Obtained from the Movement Technique

3.3 The Removal/Insertion Technique

This technique comprises of two independent operations. One involves removing a word from the sentence in order to make it more compact and probably more appealing. The other operation of this technique involves inserting a word into the sentence in order to make it clearer or more sophisticated. These operations are both in fact employed in the *Quantifier Removal/Insertion* pattern but only the insertion operation will be explicitly demonstrated here.

The *Insertion Technique* first investigates the TLCS input. For the case that the initial sentence has more than one agent doing the same action such as "เขา/*he* and เธอ/*she*" of S_i taking the same

action "ท่องเที่ยว/*travel*", the quantifier "ล้วน/*all*" can be inserted after the agents and before the action/modifier to emphasize that *every* single component really performs the same action or share the same property at the same time. By inserting this type of word, the meaning of the sentence is stressed more strongly. Caused by the *Insertion Technique*, an additional tree for the quantifier "ล้วน/*all*" is selected by the *TLTAG Selection* process and then realized as part of the sentence paraphrase during the *Surface Realization* process as shown in Figure 8.
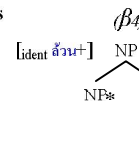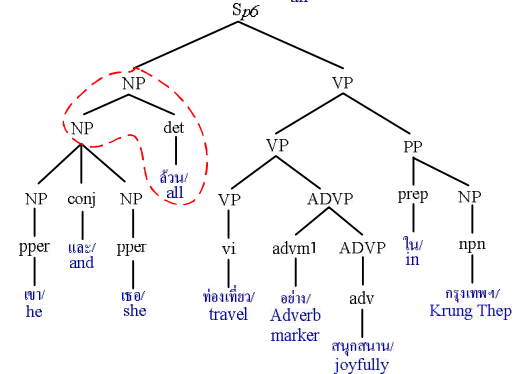

Figure 8: The Elementary Tree for "ล้วน/*all*" and the TLTAG Derived Tree for S_{p6} Obtained from the Insertion Technique

The sentence paraphrase is read off of the leaf nodes of the S_{p6} tree, as follows.

(S_{p6}) เขา/*he-Agent* และ/*and-ParallelMarker* เธอ/*she-Dative*
ล้วน/*all-Amount* ท่องเที่ยว/*travel-MotionAction*
อย่าง/*AdverbMarker* สนุกสนาน/*joyfully-Quality* ใน/*in-PositionMarker* กรุงเทพฯ/*Krung Thep-Locative*

All he and she are joyfully traveled in Krung Thep.

As for the next three paraphrasing techniques, the ideas behind each technique along with its operating procedure will be briefly discussed in this section. However, the examples of these techniques will be collaboratively demonstrated in Section 4 to show how these techniques can be used in combination to generate more complex sentence paraphrases.

3.4 The Switching Technique

This technique switches the thematic roles of the agent and the participant in the *Reciprocity* verb class (Sungkhavon, 1984). The verbs in this class must be followed by the preposition "กับ/*with-ParticipantMarker*" to indicate the togetherness of its subject and object. Every word in the Reciprocity Action verb class such as "เผชิญหน้า/*confront*", "ต่อสู้/*fight*", "สัญญา/*engage*", and "หมั้น/*engage*" etc. can switch its arguments, i.e., its thematic roles. This *Switching Technique* is exercised in the *Arguments Switching in the Reciprocity Action* paraphrase pattern as follows:

(ex_{i1}) สมศรี/*Somsri-Agent* **หมั้น**/*engages-ReciprocityAction* กับ/*with-ParallelMarker* สมชาย/*Somchai-Participant*

Somsri engages (with) Somchai.

(ex_{p1}) สมชาย/*Somchai-Agent* **หมั้น**/*engages-ReciprocityAction* กับ/*with-ParallelMarker* สมศรี/*Somsri-Participant*

Somchai engages (with) Somsri.

The switching technique can also apply to other paraphrase patterns, such as *Verb/Adverb Position Switching*, which will be demonstrated in Section 4, and *Switching Clauses in Multi-Clause sentence* as explained in the following example.

(ex_{i2}) ขโมย/*a thief-Agent* หนีไป/*flee-MotionAction* **ก่อน**/*before-TimeMarker* ตำรวจ/*a policeman-Agent* มาถึง/*arrive-MotionAction*

A thief had fled before a policeman arrived.

(ex_{p2}) ตำรวจ/*a policeman-Agent* มาถึง/*arrive-MotionAction* **หลัง**/*after-TimeMarker* ขโมย/*a thief-Agent* หนีไป/*flee-MotionAction* แล้ว/*ago-PastTense*

A policeman arrived after a thief had fled.

3.5 The Promotion/Demotion Technique

The *Promotion* mechanism usually occurs at the same time with the *Demotion* mechanism. The idea behind this technique is that as one word/phrase is promoted, another grammatically related word/phrase must be demoted. Since this technique is often used in conjunction with other techniques in generating paraphrases, the generation procedure will then be explained in Section 4.

3.6 The Nominalization Technique

The last technique to be presented changes the structure of a simple sentence/phrase but still preserves the original meaning of its initial sentence.

In Thai language, there are two prefixes for transforming a verb into an abstract noun (Thonglor, 2007). The first prefix is "การ-/*karn-*" comparable to the suffix "*-ing*" in English, to put in front of an *action* verb, e.g., "กิน/*eat*" to make a noun, e.g., "การกิน/*eating*". The second prefix is "ความ-/*kwam-*" comparable to the suffix "*-ness*" to put in front of a *mental* verb, e.g., "เสียใจ/*sad*" to make a noun, e.g., "ความเสียใจ/*sadness*". Notice that in this case, to and maintain its similar forms in both Thai and English, the Thai mental verb becomes an adjective in English.

This process can be extended to nominalize a simple sentence into a noun phrase for use in combination with the previous paraphrasing techniques for obtaining a new sentence paraphrase.

Given a simple sentence, the first step of this *Nominalization Technique* inserts the prefix "การ-/*karn-*" or "ความ-/*kwam-*" in front of the verb phrase. Then, the subject is moved to the end of the sentence and connected to the just-constructed noun phrase using the preposition marker such as "ของ/*of*" or "โดย/*by*". The new noun phrase is often used as a subject phrase or an object phrase or a modifier phrase in generating a new and more complex paraphrase as shown in the following example.

(ex_{i3}) นิวตัน/*Newton-Agent* **ค้นพบ**/*discover-TargetAction* แรงโน้มถ่วง/*gravity-Target*

Newton discovers gravity.

(ex_{p3}) [**การ**-*prefix/karn* ค้นพบ/*discover-TargetAction* แรงโน้มถ่วง/*gravity-Target*]/*AbstractNoun* **ของ**/*of-PossessorMarker* นิวตัน/*Newton-Agent*

Gravity discovering of Newton.

4 Combinations of the Proposed Thai Sentence Paraphrasing Techniques

To generate a new and probably more complex Thai sentence paraphrase, a combination of the paraphrasing techniques in Section 3 will be

employed. All possible combinations for use in the Thai sentence paraphrase patterns are depicted in Table 1. For illustration purposes, the paraphrase generation process of a combination of these particular three techniques, namely, the *Switching*, the *Promotion/Demotion* and the *Nominalization Techniques* will be applied to the *Verb/Adverb Position Switching* pattern and also fully explained now as follows.

In Thai grammar, an adverb usually acts as a modifier or sometimes an intransitive verb. This is where the *Switching Technique* comes in. However, since the syntactic functions of the verb and the adverb should also be interchanged, the adverb is grammatically promoted to a new verb while the current verb is demoted to a modifier of the new verb. Consequently, the *Promotion/ Demotion Technique* is therefore used. Last but not least, during the Demotion mechanism, the *Nominalization Technique* is also needed in transforming the current verb into an abstract noun in order to make the modifier complete.

Figure 9 illustrates an example of the above process in generating a paraphrase of the initial sentence S_i using the combination of the three mentioned techniques.

After the *TLTAG Selection* and *Surface Realization* processes yield the TLTAG Derived Trees for S_i, the *Verb/Adverb Position Switching* pattern guides the process to look for the main verb and the adverb of the sentence. The obtained main verb "ท่องเที่ยว-*vi*/*travel*" and its adverb "สนุกสนาน-*adv*/*joyfully*" are switched constituting the *Switching Technique*. Then, the adverb is promoted to a new verb "สนุกสนาน-*vi*/*enjoy*" while the old verb is demoted to a modifier for the new verb constituting the *Promotion/Demotion Technique*. During the demotion mechanism, a new elementary tree "กับ-*perp*/*with*" is acquired. This step then forces the *Nominalization Technique* to activate and form a newly transformed abstract noun "การ-*prefix*/*karn* ท่องเที่ยว-*vi*/*travel*" into a new branch of the S_{p7} TLTAG Derived Tree so that the new resulting paraphrase will be grammatically correct. Finally, the obtained sentence paraphrase can be read off of the leaf nodes of the S_{p7} tree as follows.

(S_{p7}) เขา/*he-Agent* และ/*and-ParallelMarker* เธอ/*she-Dative* **สนุก สนาน**/*enjoy-AdditionAction* **กับ**/*with-GoalMarker* [**การ-**

prefix/karn **ท่องเที่ยว**/*travel-MotionAction*]/*AbstractNoun-Complementary* ใน/*in-PositionMarker* กรุงเทพฯ/*Krung Thep-Locative*

He and she enjoy (with) traveling in Krung Thep.

In addition, other Thai sentence paraphrase patterns may use different combinations of the proposed six paraphrasing techniques to generate more complex paraphrases. For example, the *Replacement* and the *Movement Techniques* are both used in the *Negation of the Opposite Quantifier* pattern while the *Switching* and the *Promotion/Demotion Techniques* are employed in the *Preposition with Instrument-Verb Phrase Switching* pattern. Other combinations of the paraphrasing techniques used in the Thai sentence paraphrase patterns are identified and explicitly shown in Table 1.
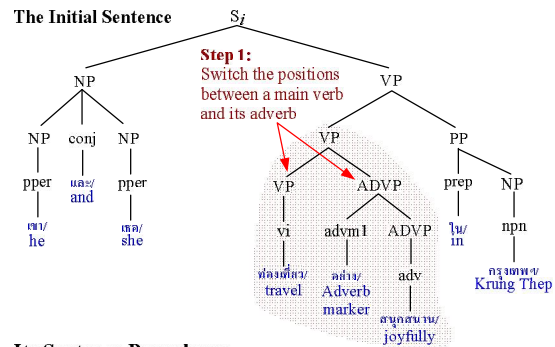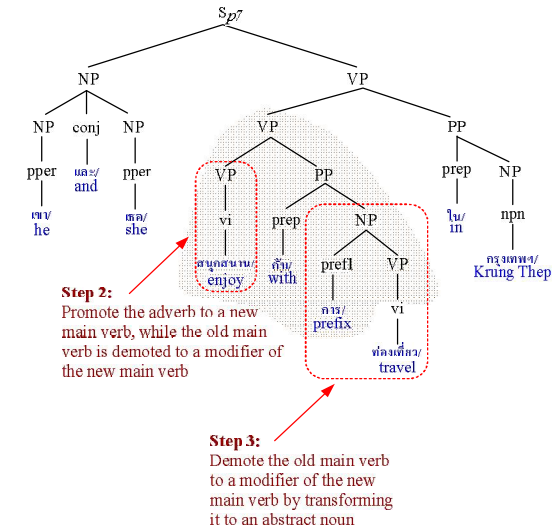

Figure 9: The TLTAG Derived Trees for Both S_i and Its Paraphrase Obtained from a Combination of the Switching, the Promotion/Demotion and the Nominalization Techniques

388

Thai sentence paraphrase patterns	Replacement	Movement	Removal/Insertion	Switching	Promotion/Demotion	Nominalization
1. Lexical Replacement						
1.1) Lexical Replacement by Its Synonym	✓					
1.2) Noun Replacement by Its Abbreviation	✓					
1.3) Common Noun Replacement by Its Definition	✓					
1.4) Grouping of Many Singular Pronouns into a Plural Pronoun	✓					
2. Preposition with Instrument-Verb Phrase Switching				✓	✓	
3. Simple Active-Passive Voices		✓		✓	✓	
4. Preposition Removal			✓			
5. Constituent Promotion/Demotion						
5.1) Direct Object Promotion/Demotion		✓				
5.2) Preposition Phrase Promotion		✓				
6. Paraphrasing in Dative Verbs						
6.1) Preposition Removal in Dative Verbs			✓			
6.2) Direct Object Promotion in Dative Verbs		✓				
6.3) Indirect Object Promotion				✓		
6.4) Passive Voice of Dative Verbs		✓			✓	
7. Arguments Switching in Reciprocity Action				✓		
8. Verb Phrase–Noun Phrase Transformation						✓
9. Verb/Adverb Position Switching				✓	✓	✓
10. Words Removal/Insertion						
10.1) Omissible Words Removal/Insertion						✓
10.2) Quantifier Removal/Insertion			✓			
11. Negation Movement						
11.1) Moving Negation Separated from Adjective/Adverb		✓				
11.2) Negation of the Opposite Quantifier	✓	✓				
12. In-Comparison Sentence Paraphrasing						
12.1) Paraphrasing in Positive Degree				✓		
12.2) Paraphrasing in Comparative Degree				✓		
12.3) Paraphrasing in Superlative Degree	✓			✓		
13. Mood Change						
13.1) Requesting ←→ Imperative Sentence	✓					
13.2) Question → Requesting or Imperative Sentence			✓			✓
14. Paraphrasing in Multi-Clause Sentences						
14.1) Switching Clauses in Multi-Clause Sentence				✓		
14.2) Collapsing A Complex Sentence into A Simple Sentence		✓	✓		✓	

Table 1: Thai Sentence Paraphrasing Techniques Identified in Thai Sentence Paraphrase Patterns

5 Conclusion

Sentence paraphrasing techniques for Thai language are discovered and proposed in this paper based mainly on the fourteen Thai sentence paraphrase patterns classified in (Phucharasupa and Netisopakul, 2011). Among these paraphrasing techniques are the *Replacement*, the *Movement*, the *Removal/Insertion*, the *Switching*, the *Promotion/Demotion* and the *Nominalization Techniques*. Some techniques involve changing only individual words or phrases and some involve changing words, phrases, or clauses in pairs. Some others may even involve changing the structure of the original sentence or phrase.

The design of *the Thai–sentence Paraphrase Generation (TPG) system* incorporating those six techniques for computationally generating paraphrases has been illustratively explained. This TPG system is based on a proposed lexical resource called the *Thai Lexical Conceptual Structure with Thai Lexicalized Tree Adjoining Grammar (TLCS–TLTAG) Resource*. This resource keeping tracks of the syntactic and the semantic structures of a lexicon simplifies Thai paraphrase generation process. The construction of this semi-automatic system is an on-going process.

References

Anabela M. Barreiro. 2008. Make It Simple with Paraphrases: Automated Paraphrasing for Authoring Aids and Machine Translation. Ph.D. Dissertation, Faculdade de Letras da Universidade do Porto, Porto, Portugal.

Aravind K. Joshi. 1999. Explorations of a Domain of Locality: Lexicalized Tree-Adjoining Grammar (LTAG). University of Utrecht (CLIN meeting).

Atsushi Fujita. 2005. Automatic Generation of Syntactically Well-formed and Semantically Appropriate Paraphrases. Ph.D. Dissertation, Nara Institute of Science and Technology, Ikoma, Nara.

Bo Pang, Kevin Knight, and Daniel Marcu. 2003. Syntax-Based Alignment of Multiple Translations: Extracting Paraphrases and Generating New Sentences. in Proceedings of HLT-NAACL, vol. 1, pp. 102–109.

Bonnie J. Dorr. 1994. Machine Translation Divergences: A Formal Description and Proposed Solution. Computational Linguistics, 20(4): 597–633.

Bonnie J. Dorr and Martha S. Palmer. 1995. Building a LCS-Based Lexicon in TAGs*. AAAI Technical Report SS-95-01, pp. 33–38.

Bonnie J. Dorr. 2001. LCS Documentation. Retrieved March 7, 2012, from University of Maryland Institute for Advanced Computer Studies, Bonnie Dorr page website: http://www.umiacs.umd.edu/~bonnie/ LCS_Database_Documentation.html

Chotikaa. Settanyakan. 2011. Translation Strategies of Focus Clausal Constructions in Academic Texts. Journal of Humanities Narasuan University, 8(1):31–54.

Erwin Marsi, Emiel Krahmer, and Wauter Bosma. 2007. Dependency-based Paraphrasing for Recognizing Textual Entailment. Proceedings of the Workshop on Textual Entailment and Paraphrasing, pp. 83–88.

Kamchai Thonglor. 2007. Principles of Thai Language. Ruamsarn (1977) Press, Bangkok, Thailand.

Krittaporn Phucharasupa and Ponrudee Netisopakul. 2011. Classification of Thai Sentence Paraphrase. International Symposium on Natural Language Processing and the Agriculture Ontology Service (SNLP-AOS 2011), pp. 197–203.

Liang Zhou, Chin-Yew Lin, Dragos S. Munteanu, and Eduard Hovy. 2006. ParaEval: Using Paraphrases to Evaluate Summaries Automatically. in Proceedings of HLT-NAACL, pp. 447–454.

Long Qiu, Min-Yen. Kan, and Tat Seng. Chua. 2006. Paraphrase Recognition via Dissimilarity Significance Classification. 2006. in Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006), pp. 18–26.

Manfred Stede. 1996. Lexical Paraphrases in Multilingual Sentence Generation, Machine Translation, vol. 11. Kluwer Academic Publishers, Netherlands, pp. 75–107.

Mark Dras. 1999. Tree Adjoining Grammar and the Reluctant Paraphrasing of Text. Ph.D. Dissertation, Department of Information and Communication Sciences, Macquarie University, Australia.

Martha Palmer and Joseph Rosenzweig. 1999. Capturing Motion Verb Generalizations in Synchronous Tree Adjoining Grammars. Kluwer Press, pp. 76–85.

Michael Ellsworth and Adam Janin. 2007. Mutaphrase: Paraphrasing with FrameNet. in Proceedings of the Workshop on Textual Entailment and Paraphrasing, pp. 143–150.

Mitsuo Shimohata. 2004. Acquiring Paraphrases from Corpora and Its Application to Machine Translation. Ph.D. Dissertation, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, Japan.

Nawawan Panthumetha. 2010. Thai Grammar, vol. 5. Faculty of Arts, Chulalongkorn University, Thailand.

Nim Kanchanacheeva. 1996. Principles of Thai Language. Thai Watana Panich Press, Bangkok, Thailand.

Nitin Madnani and Bonnie J. Dorr. 2010. Generating Phraseal and Sentential Paraphrases: A Survey of Data-Driven Methods. Computational Linguistics, 36(3):341–387.

Pablo A. Duboue and Jennifer Chu-Carroll. 2006. Answering the Question YouWish They Had Asked: The Impact of Paraphrasing for Question Answering. in Proceedings of HLT-NAACL, pp. 33–36.

Penkhae Wongsiri. 1981. Thai Intransitive Verbs: A Study and Classification in Case Grammar. M. Arts Thesis, Department of Linguistics, Graduate School, Chulalongkorn University, Thailand.

Phanu Sungkhavon. 1984. Semantic Relationships between Nous and Verb in Thai Sentences. M. Arts Thesis, Department of Thai, Graduate School, Chulalongkorn University, Thailand.

Ponrudee Netisopakul. 1997. Alternative Solution to Language Divergences: Separation of Lexical Syntax from Lexical Semantics. 9th European Summer School in Logic, Language and Information (ESSLLI'97), PhD Workshop on Natural Language Generation.

Prayoon Songsilp. 2008. Principles and Using Thai Language, vol. 1. Dhonburi Rajabhat University, Bangkok, Thailand.

Prodromos Malakasiotis. 2011. Paraphrase and Textual Entailment Recognition and Generation. Ph.D. Dissertation, Department of Informatics, Athens University of Economics and Business, Greece.

Ray S. Jackendoff. 1990. Semantic Structures. MIT Press, Cambridge, Mass.

Regina Barzilay and Lillian Lee. 2003. Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment. in Proceedings of HLT-NAACL 2003, vol. 1, pp. 16–23.

Raymond Kozlowski, Kathleen F. McCoy, and K. Vijay-Shanker. 2003. Generation of Single-Sentence Paraphrases from Predicate/Argument Structure Using Lexico-Grammatical Resources. PARAPHRASE '03 Proceedings of the Second International Workshop on Paraphrasing, vol. 16.

Shiqi Zhao, Xiang Lan, Ting Liu, and Sheng Li. 2009. Application-Driven Statistical Paraphrase Generation. in Proceedings of the 47th AnnualMeeting of the Association for Computational Linguistics (ACL) – the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (IJCNLP), pp. 834–842.

Vipa Vongsantivanit. 1983. Causative Verbs in Thai. M. Arts Thesis, Department of Linguistics, Graduate School, Chulalongkorn University, Thailand.