A FRAMEWORK FOR IMPLEMENTATION OF THAI TO ENGLISH MACHINE TRANSLATION SYSTEMS

Tawee Chimsuk

A Dissertation Submitted in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Computer Science)
School of Applied Statistics
National Institute of Development Administration
2010

A FRAMEWORK FOR IMPLEMENTATION OF THAI TO ENGLISH MACHINE TRANSLATION SYSTEMS

Tawee Chimsuk

School of Applied Statistics

Associate Professor.....Major Adviser

(Surapong Auwatanamongkol, Ph.D)

The Examining Committee Approved This Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Computer Science).

Associate Professor.....Committee Chairperson

(Pipat Hiranvanichakorn, D.E)

Assistant Professor.....Committee

(Ohm Sornil, Ph.D)

Associate Professor.....Committee

(Wirote Aroonmanakun, Ph.D)

Instructor.....Dean

(Lersan Bosuwan, Ph.D.)

August 2010

ABSTRACT

Title of Dissertation	A Framework for Implementation of Thai to English Machine Translation Systems
Author	Mr. Tawee Chimsuk
Degree	Doctor of Philosophy (Computer Sciences)
Year	2010

Machine translation (MT) is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one language to another. It can help people to communicate with others with different natural languages. Machine Translation (MT) between Thai and English languages has been an ongoing and challenging research topic. Most research has been done on English to Thai machine translation, but not the other way around.

Developing a machine translation system which can correctly translate any sentence from one language to another, seems to be a formidable and challenging task due to the big differences between syntax and semantics of the two languages. This suggests that such a system should be implemented incrementally. Starting with a system that can translate simple phrases and sentences, it can then be expanded to include grammar rules and capabilities to translate new sentences which are more complex and have not previously been successfully translated.

This research proposes a new framework to develop Thai to English machine translation systems incrementally. It is based on the interlingua approach, where input sentences are translated into intermediate representations before they are converted into equivalent sentences in the other language. Lexical Functional Grammar (LFG) - or an LFG tree structure - is used as an intermediate representation in this research. The framework has been used to implement a prototype system that can translate

about 200 Thai phrases and sentences into English phrases and sentences, resulting in a satisfactory quality of the translation output.

ACKNOWLEDGEMENTS

The author would like to thank Associate Professor Dr. Surapong Auwatanamongkol, and Associate Professor Dr. Wirote Aroonmanakun, my major advisors and supervisory professors on this research work, for their kindness in providing valuable advice, suggestions and guidance throughout the time of formulating the thesis. I would also like to thank the other thesis committees, Associate Professor Dr. Pipat Hiranvanichakorn and Assistant Professor Dr. Ohm Sornil, for their kind advice and suggestions leading to the completion of the thesis.

I would like to thank the Royal Thai Navy, especially the Royal Thai Naval Academy – my institute - for providing support for my study toward a doctoral degree at the National Institute of Development Administration. Furthermore, the author would like to thank Admiral Pongcharoon Vichayapai Bunnag and Captain Youwadee Premwichai for their kindness in promoting me.

The author also wishes to express his thanks to all professors of the School of Applied Statistics, the National Institute of Development Administration (NIDA), who have taught the author and provided me with valuable advice and suggestions. I would like to thank all staff of the school especially the staff of the school computer center for their help and supports during my study at NIDA.

The author would like to thank the National Institute of Development Administration which partly funded the work of the thesis, thereby enabling its successful completion.

Finally, the author is grateful and would like to present the accomplishment of this thesis to the author's father and mother, and other members of the author's family, who consistently supported, suggested and encouraged the author in different ways until this study was successfully completed.

Tawee Chimsuk

August 2010

TABLE OF CONTENTS

LIST OF TABLES

Tables **Page**

LIST OF FIGURES

Figures **Page**

CHAPTER 1

INTRODUCTION

1.1 Statement of the Problem

Nowadays information technology has enabled people to easily communicate with others and to search for information in cyberspace. Most information in cyberspace is written in popular languages such as English, French, German, Spanish, etc. English, especially, has become the common language for information exchange and dissemination. Most Thai people cannot write English very well, therefore language becomes a barrier for Thais to present their information to the outside world. It would therefore be convenient if we had an effective machine translation system that could translate Thai texts into English texts. An effective machine translation system could help people speaking two different languages communicate more easily. Implementing such a system is not an easy task as the two languages differ in their morphology, syntax and semantics. This suggests that such a system should be implemented incrementally. We may first want to build a system that can translate simple Thai phrases and sentences into English, and then subsequently expand the system's capability to enable it to cope with new sentences that have not been successfully translated before.

1.2 Objectives of the Study

This research aims at developing a new framework for implementing Thai to English machine translation incrementally. The framework can help simplify the implementation process of the system as grammar rules, lexicon and their corresponding translation modules can be progressively added into a system so that

new sentences can be successfully translated. The framework utilizes the Interlingua approach, in which Thai sentences are parsed into interlingua which is derived from a Lexical Functional Grammar (LFG) tree structure.

The LFG tree structure consists of two main structures: Functional structure (F-structure) and Constituent structure (C-structure). The F-structure consists of abstract functional information about components in a sentence, e.g. subject, object and abstract features such as tenses and cases. The C-structure contains syntactical structure information of the sentence by means of phrase structure trees.

The framework for implementing the machine translation system is divided into four phases.

The first phrase performs Thai syntactic analysis on a given Thai sentence. GFU-LAB software is modified to perform bottom-up parsing (Left-Corner parsing) using LFG rules for Thai sentences. The output of the first phase is a nested list structure consisting of the two structures of LFG, C-structure and F-structure. If the given Thai sentence or phrase cannot be parsed successfully, the syntax of the Thai sentence or phrase is analyzed manually using IC theory and new Thai grammar rules found by the analyses will be added to the system. Alternatively, the output of that particular Thai sentence or phrase is forwarded to the second phase.

The second phase converts the list structure of Thai LFG tree - produced from the first phase - into an interlingua in a form of a general tree structure. The third phase transforms the interlingua into a corresponding English C-structure tree. The transformation process has two steps. The first step matches the child nodes of the root in F-structure that represents the sentence level of the Thai LFG-tree against a set of predefined patterns at the sentence level of an English sentence. The predefined patterns are called sentence level FtoC patterns. Once a match is found, the corresponding root and child nodes for an output English C-structure will be created. The second step is performed recursively at each lower level or the phrase levels. The step matches the current node and its children of the F-Structure against the predefined set of phase level FtoC patterns. If a match is found, the corresponding nodes in the partially constructed English C-structure tree will be created. If a match is not found and equivalent nodes of English C-structure cannot be constructed successfully, the syntactical structure of the sentence or phrase is analyzed and new

corresponding FtoC patterns along with necessary English C-node construction modules are added into the system.

The fourth phase constructs an English sentence by traversing the English C-structure tree derived from the third phase and printing out all English words of the leaf nodes in the direction from left to right. The list of the printed words forms the English phrase or sentence which has the same meaning as the given Thai phrase/sentence.

1.3 Scope and Limitations of the Study

A prototype system which can perform Thai-English translation will be built based on the framework. Thai and English sentence grammar rules and their lexicon will be constructed based on 200 sentences/phrases from Orchid Corpus of NECTEC. Since the purpose of the implantation of the prototype system is to verify whether the framework can be implemented, all sentences/phrases used to develop the system will be limited to simple phrases/sentences. The quality of the translation output from the system will be evaluated by three Thai graduate students who are proficient in English and currently studying for Masters degrees in Languages and Communications at NIDA.

CHAPTER 2

LITERATURE REVIEW

Machine translation (MT) or Automatic translation is a sub-field of computational linguistics (Arnold, Balkan, Meijer, Humphreys and Sadler, 1994: 1). It translates sentences from one language (source language) to another language (target language). In addition, it is multi-disciplinary involving many fields such as artificial intelligence, linguistics, statistics, philosophy and so on. The aim of machine translation is to make it easier for people with different languages to communicate. A machine translation system that produces a high quality, highly accurate translation with high efficiency can help people with different natural languages to communicate without the need for a human translator.

2.1 A Brief History of Machine Translation

It has been controversial in the science community as to who was the first to introduce the concept of machine translation (MT) (Arnold et al., 1994: 12-15). The first proposal to build a machine translation system was put forward in 1947 after discussions between Andrew D. Booth, a British crystallographer, and Warren Weaver of the Rockefeller Foundation. In 1949, Weaver proposed that computers may be programmed to perform machine translations. By the early 1950s, there were many research groups working on machine translation in both Europe and the USA. In 1964, the Automatic Language Processing Advisor Committee (ACPAC) presented the ACPAC Report to indicate the state of Machine translations. It reported a hopeless state of machine translation research which had failed to produce a useful result in the form of an acceptable translation. This led to the termination of Government funding for machine translation projects in the USA.

For the next ten years, research in machine translation was mostly funded by the Mormon Church to do bible translations. The funding helped lead the way in several research projects on computational linguistics and Artificial intelligence. It was not until the late 1970s that machine translation research began to produce some successful results. Many machine translation systems have been successfully implemented, for instance, English-French Machine translation SYSTRAN system, Russian-English Machine Translation system built by USAF and NASA, EUROTRA project of the CEC and so on. These machine translation systems can be considered as some of the largest research projects in Natural Language Processing.

2.2 Problems in Machine Translation

When devising a machine translation system, many problems can be encountered which can complicate the construction. The problems can be categorized as follows (Arnold et al., 1994: 105-121)

2.2.1 Ambiguity

Two types of ambiguities may arise during the translation

2.2.1.1 Lexical Ambiguity

Any word in one language can have more than one meaning which causes lexical ambiguity and incorrect translation.

For example

a. “I show a Thai boxing”.

b. “The show is starting in the morning”.

From the above two sentences the word “Show” has two parts of speech, verb and noun and thus two meanings. In (a). its part of speech is a VERB while (b) is a NOUN. Suppose we want to translate the sentences in the example to Thai language. Simply looking up the meaning from a bilingual dictionary of English-Thai language may not be able to determine the exact meaning of that word for the particular sentence. Considering the word position and its usages in the sentence can help to identify the correct meaning of the word for a particular sentence.

2.2.1.2 Structural Ambiguity

A structural ambiguity can exist in a phrase or sentence if there are more than two structural interpretations for the phrase or sentence.

For example

“Somchai saw Dang using spyglass”.

A sentence can be interpreted in two ways as follows

1) Somchai used spyglass to see Dang
2) Dang was using spyglass, and Somchai saw him.

2.2.2 Lexical and Structural Differences

At lexical level, a single word in one language can possibly be translated into several words in other language. For example, the English word: eat can have two Thai words with similar meaning, i.e. กิน or รับประทาน

At structure level, different structures can carry the same meaning. These differences create problems of choosing which of the words or structures are to be used for translating the given sentence.

2.2.3 Multiword Units; Idioms and Collocations.

2.2.3.1 Idioms

In general, idioms need to be treated as a single word or translated using special treatment.

For example,

Somchai is making much noise. cut it out!.

The idiom **cut it out** in the above sentence means stop doing something (making much noise in the sentence context).

There are two approaches to handle the idioms.

1) Represent the idiom as a single unit and translate the unit into corresponding word or structure of the target language.
2) Use special grammar rules to parse the idiom into special representation and then convert the representation into corresponding structure of the target language.

2.2.3.2 Collocations

A collocation is a compound word containing two or more words that often go together. There are various types of collocations, for instance,

Adverb + adjective

Adjective + noun

Noun + Noun

Noun + Verb

Verb + Noun

Verb + Expression with preposition

Verb + adverb

For example in the case of adjective + Noun

The doctor ordered him to take regular exercise**.**

Translation of collocations can be handled differently from idioms, e.g. a collocation can be translated into a structure with dependent parts before the structure can be translated into a corresponding structure of the target language

2.3 Machine Translation Methods.

Several machine translation methods have been proposed (Arnold et al., 1994: 59-81,173-195). They are as follows:

2.3.1 Rule-based Method

It was the first method for machine translation that utilized collection of rules (grammar rules) and dictionary for translation. The method can be subdivided into three approaches as shown in figure 2.1.

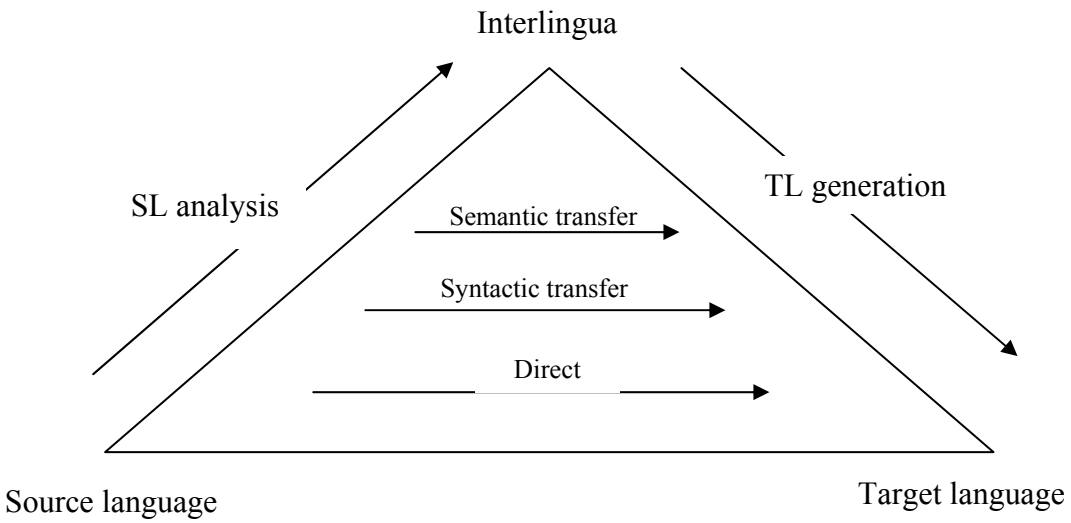


Figure 2.1 The Vauquois Triangle.

2.3.1.1 Direct Translation

This method is used in the case where source and target languages have similar grammar rules and so the translation is usually performed at word level. Little translation of structures at phrase or sentence level is performed. The translation requires a large bilingual dictionary and grammar rules to specify the word order and morphology to produce the sentence in the target language.

2.3.1.2 Transfer Translation

This method is an improvement of the direct translation method. There are three steps involved in this approach

1) It first translates the source text into abstract representation which may be language dependent

2) It then converts the abstract representation into equivalent target language representation

3) Finally it converts the target language representation into final target language text

2.3.1.3 Interlingua Translation

This method requires translation of a source language sentence into an intermediate representation or interlingua. The intermediate representation must be able to express the semantics of the translated sentence. In addition, the Interlingua must be language independent so it can be converted into several target languages. This method normally consists of two stages. The first stage translates the source sentence into an Interlingua and the second stage converts the Interlingua into a sentence of the target language. There were a number of machine translation researches using interlingua approaches. For instances, Dorr (1987) used a tree structure interlingua, Lee, Yi, Seneff and Weinstein (2001: 1-6) employed a semantic frame as interlingua, Xuan, Huiping, and Huowang (2002: 464-472) employed a framework as interlingua, and Turhan (1997: 320-323) used a case frame as interlingua.

2.3.2 Knowledge-Based Machine Translation

This approach is similar to the interlingua approach, namely the input sentence language is translated into an intermediate form which is independent of any language. The difference between the two approaches is that the intermediate form of the knowledge based machine translation is intended to represent the semantics of the input sentence rather than the syntactics of the sentence. The processes of this approach include two phases: the syntactic analysis that translates the source texts into intermediate form representing only its meaning (semantic knowledge), and the transformation of the intermediate form or semantic representation into the target language. Hence, the meaning (knowledge) is extracted from the input sentence and used during both analysis and transformation phases. This approach attempts to acquire and encode various kinds of meaning or knowledge from sentences for the purpose of disambiguation.

2.3.3 Corpus-Based Machine Translation

This approach has gained more popularity for machine translation following the developments in machine translation technology. The Corpus-based machine

translation systems can learn how to translate words, phrases and sentences from previously translated materials (corpus). The method uses either statistical techniques or analogical mapping to select the best translation from the corpus.

2.3.4 Hybrid Approaches

The hybrid approaches may combine two or more methods in performing the translation, for instance, statistical-based method combined with traditional rule-based approach.

2.4 Characteristics of Thai Language

The characteristics of Thai language have been extensively discussed in scientific publications (Banchob Bandhumedha, 2002; Amara Prasithrathsint, Yuphaphann Hoonchamlong and Saranya Savetamalya, 2003; Nittaya Kanchanawan, 1999; Navavan Bandhumedha, 2006). Important characteristics of the Thai language are as the follows

2.4.1 Properties of Thai Language

2.4.1.1 Thai word is an isolate language, having its complete, exclusive meaning. Some word may have several meanings. The position of a word in a sentence is used to identify which particular meaning it represents. The order or arrangement of words in a sentence also plays a role in determining the part of speech for the words in the sentence. To indicate gender, number, tense, or mood of a word, other words must be added to form a compound word (Banchob Bandhumedha, 2002: 1-2).

2.4.1.2 A Thai sentence can be divided into four parts

1) Subject part, this part must precede a predicate part and can be placed at any position in a sentence.

2) Predicate part, this part must follow a subject part but precede any object parts. In a Thai sentence there can be more than one predicate part.

3) Object part, this part always follow the predicate part, however some predicate parts may not need object parts.

4) Modifier part, e.g. an adjective or adverb. An adjective expands a subject part or object part and must be placed after the subject or the object part. An adverb expands a predicate part and must be placed after the predicate part.

2.4.2 Types of Thai Sentences

There are four types of Thai sentences as follows (Banchob Bandhumedha, 2002: 101-103; Nittaya Kanchanawan, 1999: 143-144)

2.4.2.1 Order Sentences

These types of sentences do not need a subject part as it implies that the listener is the subject part who is ordered to do something. If there is a subject part, it wants to emphasize who is the subject for the order. For example,

แดง/มา (Dang come)

2.4.2.2 Persuasive Sentences

They ask someone to do something so they share the same properties as the order sentence. The sentence must end with words such as น่า, น่ะ, ซิ to show the persuasion.

2.4.2.3 Interrogative Sentences

Adverbs of question such as Who, What, Where, When, How must exist in this type of sentences.

2.4.2.4 Affirmative Sentences

They are always used to state general facts or tell general stories.

2.4.3 Thai Language Processing Problems

2.4.3.1 Word Segmentation Problem

A Thai text is a string of symbols with no explicit word boundary, which is similar to Japanese, Chinese and Korean languages. To perform Thai language processing, Thai text must be segmented into strings of words. Complications in Thai word segmentation can arise as there can be more than one

word segmentation pattern for some given sentences. Several research studies have been done on Thai word segmentation. (Sojka and Antoš, 2003: 65-72) employed pattern matching to Thai texts. (Tanapong Potipiti, Virach Sornlertlamvanich and Thatsanee Charoenporn, 2000: 176-181) used statistical methods to determine the best word segmentation pattern for a Thai text. (Surapant Meknavin, Paisarn Charoenpornsawat, and Boonserm Kijsirikul, 1997) used machine learning techniques with feature-base for Thai word segmentation. (Paisarn Charoenpornsawat, Boonserm Kijsirikul, and Surapant Meknavin, 1998) used machine learning to learn feature-based for identifying Thai unknown word boundaries and Thai Proper names. (Boonserm Kijsirikul, Paisarn Charoenpornsawat and Surapant Meknavin, 1999) showed the effective result of Winnow than RIPPER to Thai Named-Entity Identification. (Wirote Aroonmanaku, 2002: 68-75) presented approach of Thai word segmentation based on syllable segmentation and syllable merging.

2.4.3.2 Sentence Segmentation Problem

There is no explicit Thai sentence marker like a full stop in English. Generally, a space can be put at the end of a Thai sentence to indicate the end of the sentence, however it does not always indicate the end of a sentence. Pradit Mittrapiyanuruk and Virach Sornlertlamvanich (2000: 23-28) used a tokenization concept, word segmentation and a part-of-speech trigram model to check the sequence of POS to determine the end of a sentence. Paisarn Charoenpornsawat and Virach Sornlertlamvanich (2001) used technique on machine learning on training corpus to build a method for segmenting a Thai paragraph into sentences.

2.4.4 Comparison of Thai and English Language Characteristics

The similarities and differences between Thai and English can be summarized as follows (Wittaya Nathong, 2003: 51-214)

2.4.4.1 Inflection Language

English languages are Inflectional languages. Morphemes of English words can be changed internally to indicate number, gender, tense and so on. In contrast, Thai language, morphemes can not be changed to indicate number, gender and so on.

1) Plural Noun Inflection

To indicate plurality of a noun in English, the suffix –s is appended at the end of the noun, while in Thai an adjective is used to express plurality of a noun.

2) Possessive Cases

Apostrophe s is employed to indicate possession in English language. Thai language uses a preposition ของ (of) to indicate possession.

3) Pronoun Inflection

English language uses I, we, you, he, she, it, they and who as subjective forms of pronoun inflection and uses me, us, him, her, them and whom as objective forms of pronoun inflection while you and it can be used for both cases. Thai language has no Pronoun inflection.

4) Verb Inflection

There are five forms of verb inflection in the English language, simple form, singular, past tense, past participle and present participle. There is no verb inflection in Thai language.

5) Comparison of Adjective and Adverb Inflection

Inflection of adjectives and adverbs in English language indicate the degrees of the adjectives or adverbs. Three degrees can be expressed, i.e. positive (normal) degree, comparative degree and superlative degree. Thai language uses the word กว่า to indicate comparative degree and the word ที่สุด to indicate superlative degree.

2.4.4.2 Derivational Suffixes

For English language derivational suffixes such as ly, ing can be appended to a word to alter their part of speech, i.e. verb to noun, verb to adjective, noun to adjective, adjective to noun and so on. For Thai language the word การ can be put in front of a verb and thus constitute a noun compound word, ความ put with an adjective to make a noun compound word and อย่าง together with an adjective to make an adverb compound word.

2.4.4.3 Basic Sentence Structure

Basic sentences are sentences comprising only three word types; noun, verb and adjective. The structures of English and Thai basic sentences can be summarized as follows in table 2.1

Table 2.1 Thai and English Basic Sentence Structures

Thai sentence structures	English sentence structure
Noun + Transitive verb + Noun แดง เตะ ฟุตบอล (Dang kicks football.)	D + N + TV + D + N The boy hits his brother.
Noun + Intransitive verb แดง วิ่ง (Dang runs.)	D + N + Int. V. The boy runs
No Structure	D + N + L-verb + Adj The boy is smart.
Noun + Linking verb + Noun แดง เป็น นักกวี (Dang is a writer.)	D + N + L-verb + D + N The boy is a sportsman.

Note: D = noun-determiners
N = noun (include pronoun)
Int.V = Intransitive verb
TV = Transitive verb
L-verb = linking verb

2.5 Research Works on Thai-English Machine Translation

Machine Translation (MT) between the Thai and English languages has been an ongoing research topic within Thai natural language processing. Most of the research was done on English-to-Thai machine translations, but not the other way around. Machine translation systems were implemented both for research and for

commercial application. Virach Sornlertlamvanich, Paisarn Charoenpornsawat, Mothika Boriboon and Lalida Boonmana (2000) implemented an English-to-Thai MT system that used the Interlingua approach extended from the CROSSROAD system, a MT system originally designed for Japanese. Three types of grammar , namely Phrase Structure grammar, Dependency grammar and Case grammar, were all used within the English-to-Thai MT system. The dictionary used in the system consisted of S-BLOCK comprising of the syntactic information for English words, C-BLOCK comprising of the semantic information for English words and the final block – G-BLOCK comprising of the syntactical information for Thai words. Prachya Boonkwan and Asanee Kawtrakul (2002) implemented an Internet-based English-to-Thai MT system translation assistant system which performed sentence structure transfer between the two languages. Since the system was semi-automatic, users could manually choose the most appropriate translation and could manually input new rules into the system to resolve some of the translation problems such as structural ambiguities. The system could be divided into four components. The first component, syntactical analysis, required users to identify a sentence boundary before the parsing could be performed. The second component performed probabilitic structural transformation between the two languages. The third component performed Thai sentence generation using linguistic knowledge of translation as well as unknown English-Thai word alignments. In other approaches, a hybrid combinding rule-based and example-based machine translation, was proposed, for instance, (Kaewchai Chancharoen, Nisanad Tannin and Booncharoen Sirinaovakul, 1998) proposed a system that used grammar rules to parse English sentences into phrase structure form and then performed mapping of the structures into a Thai sentence. This system used a corpus-based approach to calculate the probabilities of words to be selected and used for the translation. Another approach (Booncharoen Sirinaovakul and Kaewchai Chancharoen, 2002) proposed machine-aided translations based on language structure matching and transformation. Sentences were parsed into phrase structure format using a bottom-up technique. Equipped with a large bilingual dictionary for mapping words between the two languages, the system could solve ambiguities by selecting a proper Thai word for a given English word. One of the popular web-based MT

systems was established by internet giant Google company. The system could perform bi-directional multilingual translations using the corpus-based approach. It used a very large dictionary to keep several billion words in all the languages it supported. Another MT system was implemented by Asia Online Pte. The system could translate ten Asian languages and more than twenty-three European languages (about 500 different language pairs). The system adopted a statistical machine translation (SMT) platform that could learn from human feedback to improve the quality of its translation. Other linguist tools, such as sentence alignment, optical character recognition (OCR), word segmentation and so on, were also incorporated into the system.

2.6 Lexical Functional Grammar

Lexical functional grammar (LFG) was developed using a linguistic theory to study several aspects of linguistic structures and the relations between them (Falk, 2001: 1-89). LFG was developed by Joan Bresnan and Ron Kaplan for natural language processing. It views a language as being made up of multiple dimensions of structures. Each of these dimensions is represented as a distinct structure with its own rules, concepts, and forms. LFG grammar introduced several structures to represent the syntactic and semantic information of a sentence. Two LFG structures, used in this research are Functional structure and Constituent structure. Functional structure (f-structure) represents the abstract grammatical functions and features of a sentence such as subject, object, tense and case. Constituent structure (C-structure) represents the word order and phrasal grouping of a sentence in a form of phrase structure trees. LFG grammar has previously been used in natural processing, for instance, Güngördü and Oflazer (1994: 494-500) for Turkish language, (Maosong, 2001) for Chinese language and (Briffault, Chibout, Sabah and Vapillon, 1997: 1-14) for French language.

2.6.1 Constituent Structure (C-structure)

The C-structure of a sentence is represented by a phrase structure tree. Considering the following phrase structure grammar rules

$$S \rightarrow \underset{(\uparrow SUBJ) = \downarrow}{NP} \quad \underset{(\uparrow = \downarrow)}{VP}$$

$$VP \rightarrow V \quad \underset{(\uparrow OBJ) = \downarrow}{NP}$$

$$NP \rightarrow N \quad \underset{(\uparrow MODI) = \downarrow}{N}$$

$N \rightarrow N$

$N \rightarrow$ cross cut picture

$N \rightarrow$ body

$V \rightarrow$ snapshot

$N \rightarrow$ computer

A phrase structure tree for a Thai sentence “คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย”, which can be described by the above grammar rules is shown figure 2.2. The phrase structure tree can be constructed through the syntactical analysis of the sentence.

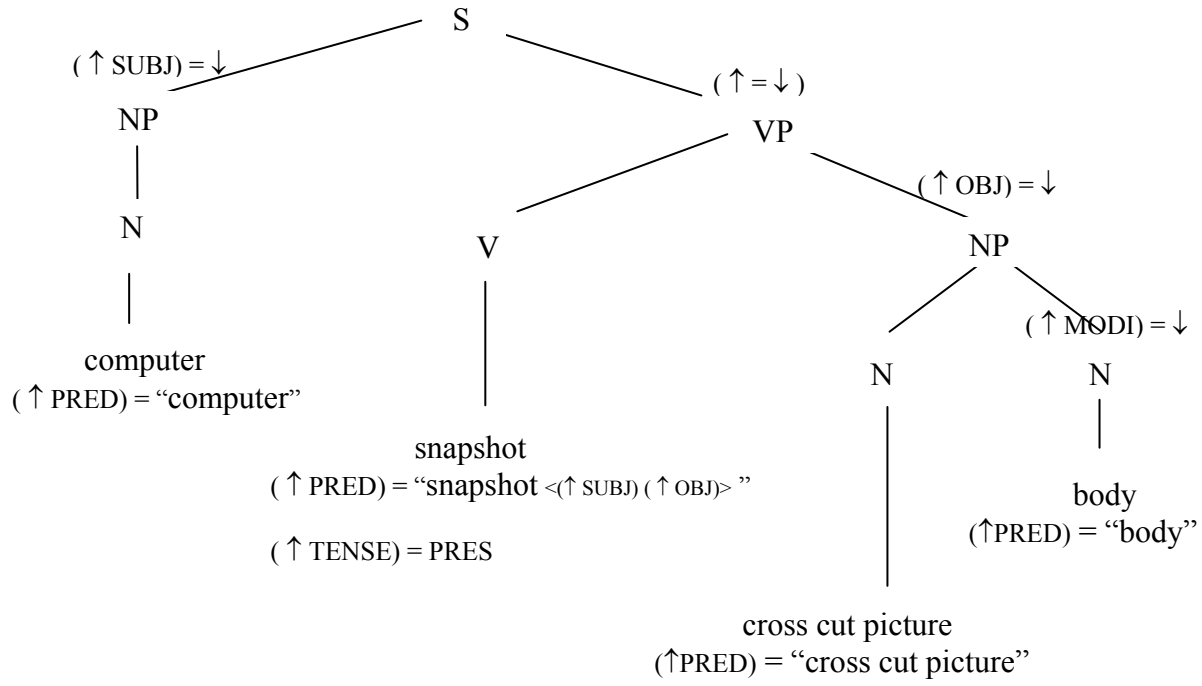


Figure 2.2 An Example of a Phrase Structure Tree (C-structure of a Thai Sentence “คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย”)

2.6.2 Functional Structure (F-structure)

Phrase structure grammar rules are annotated with functional schemata using two metavariables up-arrows (↑) and down-arrows (↓). Consider the functional annotated phrase structure grammar rules and the example of phrase structure tree derived from the grammar rules as shown above. The up and down arrow indicated at any node in the tree describes how the functional information of that node will be transferred to or shared with the f-structure of its parent node. Up arrows refer f-structure of the parent node and down arrows refer to the f-structure of the node itself.

According to the example structure tree the functional information contained in the NP node (NP’s f-structure) will be transferred to the SUBJ part of the parent’s f-structure (f-structure of S node). While the functional information in the VP node (f-structure of VP node) is all transferred to or shared with its parent node.

A PRED in the tree describes a semantic form of a lexical entry. The functional information (↑PRED) = “Computer” at the Computer node is transferred up to its parent node and assigned to be the SUBJ part of the S node. The functional information (↑PRED) = “cross cut picture” at Cross cut picture node is also transferred to be the OBJ part of the S node. As the (↑PRED) = “snapshot<(↑SUBJ) (↑OBJ)>” and (↑TENSE) = “PRES” are all transferred to the VP node and finally consolidated at the S node, the f-structure of the whole sentence at the S node are as follows,

$$\begin{pmatrix} \text{SUBJ} & [\text{PRED} \quad \text{‘Computer’}] \\ \text{OBJ} & \begin{pmatrix} \text{PRED} & \text{‘Cross cut picture’} \\ \text{MODI} & [\text{PRED} \quad \text{‘Body’}] \end{pmatrix} \\ \text{TENSE} & \text{PRES} \\ \text{PRED} & \text{‘snapshot <(↑ SUBJ)(↑ OBJ)>’} \end{pmatrix}$$

Figure 2.3 A f-Structure of Thai Sentence “คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย”

The information in the f-structure, as shown in figure 2.4, is displayed in a form of Attribute-value matrix (AVM). The left-hand side of AVM is an attribute name, and the right hand side is its corresponding value of the attribute.

2.7 Case Grammar

Case Grammar was developed by Charles J. Fillmore in 1968 (Cook, 1989: 1-13). The grammar was designed to describe the logical form of sentence in terms of

a predicate and a series of case-labeled arguments. Semantic relationships between noun and verb in a sentence are classified into six cases as follows;

1) Agentive
2) Instrument
3) Dative
4) Factitive
5) Locative
6) Objective

Phanu Sungkhavon (1984: 73-123) proposed that the semantic relationships for Thai sentences should be classified into 16 cases as follows,

1) Agent: the relationship between a noun and a verb of action where the action is performed by the noun.

For example:

แดง/วิ่ง (Dang runs)

วิ่ง is an active verb.

แดง is the Agent.

2) Dative: the relationship between a noun and a verb where the noun is the dative of the verb.

For example:

บ้าน/เก่า (Old house)

เก่า is an attribute verb.

บ้าน is the Dative of an attribute.

3) Experiencer: the relationship between a noun and a verb where the noun expresses a feeling or experiences of the verb.

For example:

แดง/เหนื่อย (Dang is tired)

เหนื่อย is a stative verb.

แดง is the Experiencer.

4) Object: the relationship between a noun and an action verb where the noun is affected or changed by the action verb.

For example:

พ่อ/ซ่อม/เก้าอี้ (Father repairs a chair)

ซ่อม is an action verb.

พ่อ is an Agent.

เก้าอี้ is an Object.

5) Goal: the relationship between a noun and a verb of action where the noun is the result of the action performed by the agent.

For example:

น้อย/ทอ/ผ้า (Noi weaves textile)

ทอ is an action verb.

น้อย is an Agent.

ผ้า is an Goal.

6) Target: the relationship of a noun and a verb where the noun is the target of the verb.

For example:

นิด/ไป/ตลาด (Nid goes to the market)

ไป is an action verb.

นิด is an Agent.

ตลาด is an target.

7) Locative: the relationship between a noun and a verb where the noun is a location of the verb.

For example:

นิด / นั่ง /บนเก้าอี้ (Nid sits on a chair)

นั่ง is an action verb.

นิด is an Agent.

เก้าอี้ is an Locative.

8) Participant: the relationship between a noun and a verb where the noun participates in the action or the feeling of the verb.

For example:

หน่อย / แต่งงาน / กับเดช (Noi is married to Dej)

แต่งงาน is an action verb.

หน่อย is an Agent.

เดช is a Participant.

9) Quantity: the relationship between a noun and a verb, revealing a condition or state.

For example:

แหวน / ราคา / 20,000 บาท (The ring costs 20,000 baht)

ราคา is an attribute verb.

แหวน is a Dative.

20,000 บาท is a Quality.

10) Complementary: the relationship in which a noun is complementary to a verb.

For example:

น้อย / เล่น / ฟุตบอล (Noi plays football)

เล่น is an action verb.

น้อย is an Agent.

ฟุตบอล is a complementary.

11) Time: the relationship between a noun and a verb where the noun tells the time of the action of the verb.

For example:

นิด / ตื่น / ตี 4 (Nid wakes up at 4.00 am)

ตื่น is an action verb.

นิด is an Agent.

ตี 4 is a Time.

12) Cause: the relationship between a noun and a verb where a noun is the cause of action or state, or an experiencer's feeling.

For example:

นิด / เรียน / หมอ เพราะแม่ (Because of her mother, Nid decided to study in a faculty of medicine)

เรียน is an action verb.

นิด is an Agent.

แม่ is a Cause.

หมอ is a Complementary.

13) Source: the relationship between a noun and a verb of action where the noun is a state before the action takes place.

For example:

พ่อ / กลับ / จากที่ทำงาน (Father came back from his office)

กลับ is an action verb.

พ่อ is an Agent.

ที่ทำงาน is a Source.

14) Instrument: the relationship between a noun and a verb of action where the noun is an instrument or raw material used to perform the action.

For example:

พ่อ / เปิด / ประตู ด้วย กุญแจ (Father opened the door with the key)

เปิด is an action verb.

พ่อ is an Agent.

ประตู is an Object.

กุญแจ is an Instrument.

15) Manner: the relationship between a noun and a verb of action where the noun tells about the manner of the action.

For example:

แดง / เรียน / หนังสือ ด้วยความตั้งใจ (Dang studies diligently)

เรียน is an Action verb.

น้อง is an Agent.

หนังสือ is a Complementary.

ความตั้งใจ is a Manner.

16) Beneficiary: the relationship between a noun and a verb of action where the noun is a beneficiary from the action.

For example:

พ่อ / ทำงาน / เพื่อลูก **(**father works to support his children)

ทำงาน is an action verb. พ่อ is an Agent.

ลูก is a Beneficiary.

Consider a sentence “Dang kicks football” represented in the logical form of the case grammar in figure 2.5

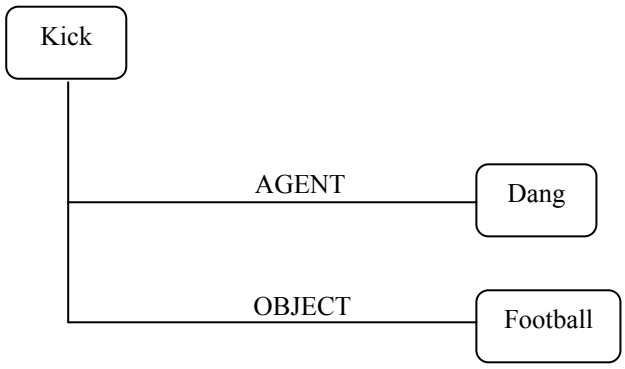


Figure 2.4 Case Grammar of "Dang Kicks Football"

According to the principle of Case Grammar, "Dang" is an agent of the verb "kick" while "football" is the object to the verb "kick".

2.8 Immediate Constituent Analysis (IC)

Immediate Constituent Theory was introduced by Leonard Bloomfield an American linguist in 1933. This theory has been used by linguists to analyze the grammar of sentences. In the analysis, a sentence is repeatedly divided into successive levels of disjoint constituents until word constitutes are reached and analysis is performed for each level of constituent(s) (Amara Prasithrathsint, Yuphaphann Hoonchamlong and Saranya Savetamalya, 2003: 151-157). A sample of analysis on the grammar of a Thai sentence using IC-analysis is shown in table 2.2

Table 2.2 A Sample of IC-analysis of a Thai Sentence

<table>
<tr><td>แต่ละ</td><td>ชุด</td><td>จะ</td><td>ทำหน้าที่</td><td>ต่างกัน</td></tr>
<tr><td colspan="3">แต่ละ ชุด</td><td colspan="2">จะทำหน้าที่ ต่างกัน</td></tr>
<tr><td colspan="5">แต่ละ ชุด จะทำหน้าที่ ต่างกัน</td></tr>
</table>

2.9 Parsing or Syntactical Analysis

Parsing (syntactical analysis) is the process of analyzing a sequence of tokens or words for a given sentence to determine whether the sentence's grammatical structure follows designated grammar rules. The parser, a program that performs parsing, also produces a data structure called a parse tree, representing the syntactical structure of the given sentence for later processing (Aho, Sethi, and Ullman, 1986: 159-267; Jurafsky et al., 2000: chap. 10).

In general, parsing can be performed in two approaches

2.9.1 Top-Down Parsing

It constructs a parse tree from the root node down to the leaves of the tree. The parser starts with the whole sentence and breaks it down into smaller parts until terminal symbols are reached. For example,

Consider the following grammar

$$S \rightarrow bCa$$
$$C \rightarrow xy \mid x$$

and the input string of bxa.

To construct a parse tree for the input string using the top-down approach, the parser first creates a single node labeled S. From the given grammar, it chooses the first rule containing left hand-side symbol of current leftmost Nonterminal S to expand the tree as in figure 2.5

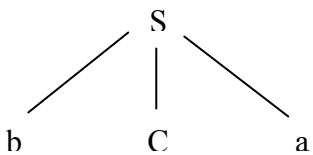


Figure 2.5 The First Expansion of Parse tree by Top-Down Parsing

As the current leftmost leaf of the tree is a terminal symbol and matching the current symbol pointed by an input pointer, so the input pointer is advanced to the next input symbol. Next, symbol of the parse tree is C so it is expanded using the first choice of C (C → xy). The output of tree after the expansion is shown in figure 2.6

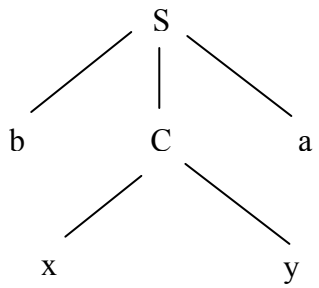


Figure 2.6 The Second Expansion of Parse Tree by Top-Down Parsing

Now the current leftmost leaf of the tree is a terminal and matches with a current input symbol, the input pointer is then advanced to the next input symbol. As the current input symbol does not match with the leftmost terminal leaf labeled y, the parser must backtrack and go back to expand C with other choice from the second choice of C, a parse tree in figure 2.7 is built and the input pointer is also moved back to the symbol X.

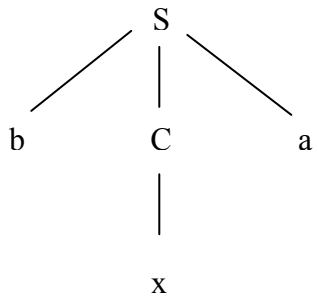


Figure 2.7 The Final Parse Tree Generated by Top-Down Parsing

As the current leftmost terminal x matches with the current input, the input pointer is then advanced to next symbol y, which then matches the leftmost terminal. The parser can halt because there are no more inputs. Parsing was successful and can now report the output.

2.9.2 Bottom-up Parsing

It starts from the words of the input and attempt to build a parse tree from leaves up to the root by building internal nodes according to given grammar rules. It succeeds if the parser can build a tree rooted with starting symbol S.

For example,

Consider the following grammar

S → xMNy
M → Mcd | c
N → a

and the input string is ***xccday***.

The input string is scanned to find a substring that matches the right side of a grammar rule and replaces the substring with the left hand-side symbol of the rule. The parser chooses the leftmost c and replaces it with M according to the rule M → c. Next, the parser finds the substring Mcd of the result string and replaces the substring Mcd with the symbol M according to M → Mcd. Next, the result string is replaced by N according to N → a. It now becomes xMNy which can be replaced by S according to the first grammar rule. Hence, the given input string can be reduced to the starting symbol S following some grammar rules (see figure 2.8). Therefore, the input string is accepted by the parser.

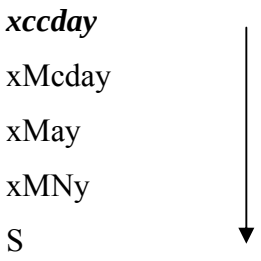


Figure 2.8 Bottom-up Parsing on the Input String xccday

2.10 The Parser for LFG

GFU-LAB is to be used in our implementation. It is a software package developed to support lexical functional grammar parsing. It was developed by (Juan C. Ruiz Anton) and written in prolog programming language. The parser of the software employs a left corner algorithm for parsing.

The Left-corner algorithm combines bottom-up and top-down methods. It avoids going the wrong way -which is prone to occur in pure top-down and pure bottom-up approaches - and thus causes unnecessary backtracking. Let us consider English grammar rules as given in figure 2.9.

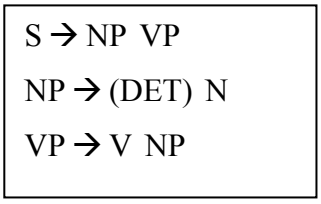


Figure 2.9 Existing English Grammar Rules

Given an English sentence “The hunter saw tigers” which is to be parsed using the left corner algorithm parts of speech for each word in the sentence, the outcome of “The hunter saw tigers” is shown in figure 2.10

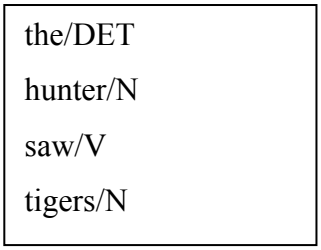


Figure 2.10 Parts of Speech of Words in the Sentence "The Hunter saw Tigers"

Initially, a symbol '*' is placed in front of the first word of the given sentence to indicate the current word position for parsing (see figure 2.11.) The following steps of the left corner parsing are to be performed: First step: to assign a current goal category and push it into stack, as shown in figure 2.12

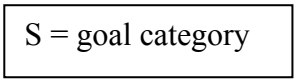


Figure 2.11 An Initial Stack that S is the Current Goal Category

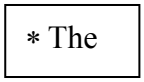


Figure 2.12 The Initial Current Word Position

Second step: to determine the part of speech of the current word that is "The", i.e. DET, then advance the '*' to next word in the input sentence and replaces "The" with DET, as shown in figure 2.13

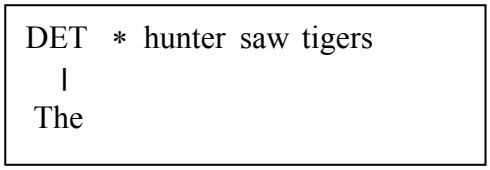


Figure 2.13 The Current Word Position After the Second Step

Third step: to find a grammar rule which the first element on the right-hand side is DET, which matches the word “The” in the sentence. The rule “NP → DET N” is found, this rule is then pushed into the stack. Place the ‘*’ after the “DET” to indicate the current goal category position and set the Nonterminal (N) to be the current goal category, as shown in figure 2.14

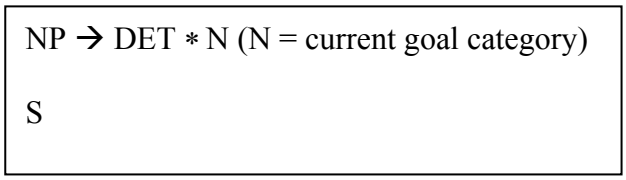


Figure 2.14 The Current State of the Stack after the Third Step

Fourth step: to determine the part of speech of the current word “hunter”, which is “N”. Since the current goal category is the same as the part of speech of the current word, the current word position ‘*’ is advanced to the next word and replaces “DET hunter” with NP, as shown in figure 2.15

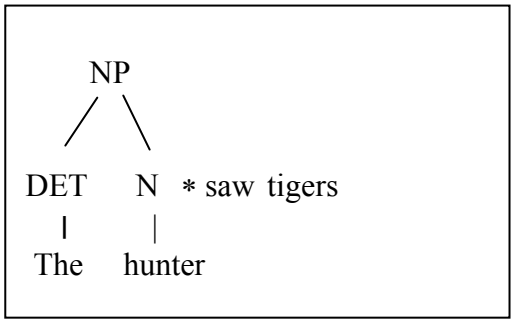


Figure 2.15 A Current Word Position after the Fourth Step

As all RHS symbols of the rule NP → DET N have been successfully matched against words in the given sentence, the rule is popped from the stack.

Fifth step: to look for rules that have NP as the first elememt of the RHS. The rule S → NP VP, V is found and pushed onto the stack. Since NP has been completed.

the '*' is then placed after NP and indicates the Nonterminal (VP) to be the current goal category, as shown in figure 2.16

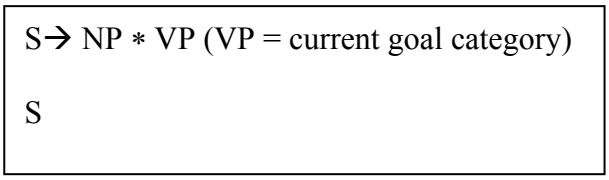


Figure 2.16 The Current State of the Stack after the Fifth Step

Sixth step: to determine the part of speech of the current word "saw" which is V. As the goal category is not the same as the part of speech of the current word, i.e. "VP ≠ V", the rule that contains V as the first symbol of its RHS is considered. Hence the rule VP → V NP is pushed onto the stack. The '*' is then placed after the "V" to indicate the current goal category which is the Nonterminal (NP), as shown in figure 2.17.

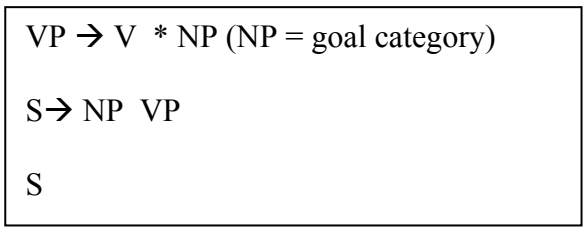


Figure 2.17 The Current State of the Stack after the Sixth Step

As the symbol V has been successfully parsed, the '*' is advanced to the next word in the input sentence, as shown in figure 2.18

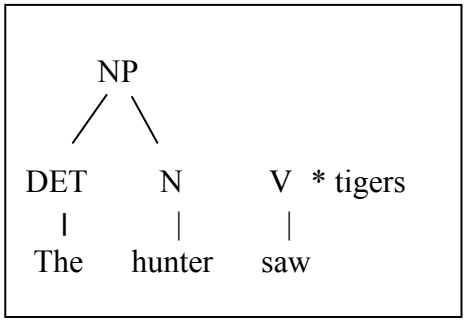


Figure 2.18 A Current Word Position at the "tiger" after the Sixth Step

Seventh step: to determine part of speech of the current word "tigers". The current goal category is "NP" while the part of speech of the current word of "tigers" is "N". A grammar rule (NP → N) with the first element of RHS is N (matched with the part of speech of current word) is pushed onto the stack as shown in figure 2.19

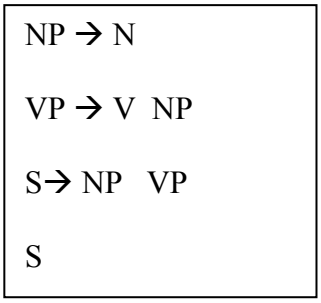


Figure 2.19 The Current State of the Stack after the Seventh Step

Eighth step: since there is no further goal to be completed in rule NP → N, the rule is popped, and so are the other rules on the stack. Thus all are popped from the stack. Since the initial goal S is also popped and there is no more word in the sentence to be parsed,the parsing process can be terminated. The final parse tree generated from the left corner parsing is shown in figure 2.20

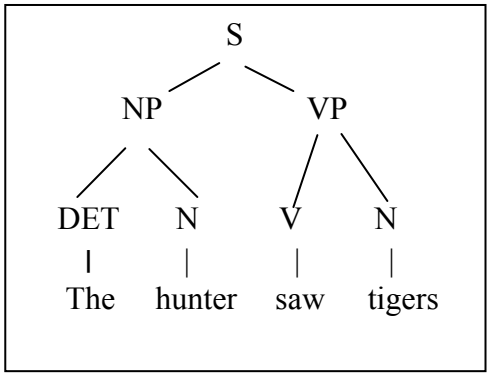


Figure 2.20 A Final Parse Tree of the Left Corner Parsing for the Sentence " The Hunter saw Tigers"

2.11 Orchid Corpus

Orchid Corpus (Open linguistic Resource Channeled toward Interdisciplinary research) was established to support research in natural language processing. It is a cooperative research between Linguistic and Knowledge Science Laboratory (LINKS) of NECTEC, Thailand and Kansai Advanced Research Center (KARC) of Japan. The Orchid Corpus contains about 2 MB of texts from proceeding of a NECTEC annual conference. Texts in the corpus are tagged using the primary part-of-speech (POS) tag set of LINKS. Texts in the corpus are divided into three levels: paragraph, sentence, and word levels. Probabilistic trigram techniques are used for word segmentation and POS tagging. The some of data from orchid corpus are shown in figure 2.21

```
%TTitle: การประชุมทางวิชาการ ครั้งที่ 1
%ETitle: [1st Annual Conference]
%TAuthor:
%EAuthor:
%TInbook: การประชุมทางวิชาการ ครั้งที่ 1, โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์, ปีงบประมาณ 2531, เล่ม 1
%EInbook: The 1st Annual Conference, Electronics and Computer Research and Development Project, Fiscal Year 1988, Book 1
%TPublisher: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน
%EPublisher: National Electronics and Computer Technology Center, Ministry of Science, Technology and Energy
%Page:
%Year: 1989
%File:
#P1
#1
การประชุมทางวิชาการ ครั้งที่ 1//
การ/FIXN
ประชุม/VACT
ทาง/NCMN
วิชาการ/NCMN
<space>/PUNC
ครั้ง/CFQC
ที่ 1/DONM
//
#2
โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์//
โครงการวิจัยและพัฒนา/NCMN
อิเล็กทรอนิกส์/NCMN
และ/JCRG
คอมพิวเตอร์/NCMN
//
#P2
#1
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ//
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ/NPRP
//
#2
กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน//
กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน/NPRP
//
```

Figure 2.21 Examples of Data from Orchid Corpus

Thai part-of-speech used as the tagset for orchid. The tagset using with orchid is categorized as follows

2.11.1 Noun

It is used to refer to people, animal, place, thing or abstraction and can be divided into 6 subcategories.

1) Proper noun (NPRP)

It is the name of a particular person, place, or thing.

2) Cardinal number noun (NCNM)

It is used to express number but is not the one used in a quantitative expression.

3) Ordinal number noun (NONM)

It is used to put thing in a numerical order.

4) Label noun (NLBL)

It is used to label an item.

5) Common noun (NCMN)

It is used to refer to a person, a place or a thing in general.

6) Title noun (NTTL)

It is a descriptive or distinctive appellation to show a person's rank, occupation, etc.

2.11.2 Pronoun

It may replace a noun or noun phrase and can be subcategorized into 4 subcategories.

1) Personal pronoun (PPRS)

It represents a person to refer to oneself, the people or things to talk about.

2) Demonstrative pronoun (PDMN)

It refers to something being either near to or distant from the speaker.

3) Interrogative pronoun (PNTR)

It is used to form a question

4) Relative pronoun (PREL)

It introduces a relative clause.

2.11.3 Verb

It occurs as part of the predicate of a sentence, or refers to an action, state or attribute and can be divided into 3 subcategories.

1) Active verb (VACT)

It indicates an action of an entity.

2) Stative verb (VSTA)

It refers to a state and is not used in progressive aspect.

3) Attributive verb (VATT)

It refers to an attribute. For example, ดี (good).

2.11.4 Auxiliary

It is used with a verb of a sentence indicating a grammatical function such as aspect, voice, mood or tense and can be divided into 5 subcategories.

1) Pre-verb auxiliary, before negator (XVBM)

It is used preceding a verb and in the position preceding the negator "ไม่" (not) in a negative sentence.

2) Pre-verb auxiliary, after negator (XVAM)

It is used preceding a verb and in the position following the negator "ไม่" (not) in a negative sentence.

3) Pre-verb auxiliary, before or after negator (XVMM)

It is used preceding a verb and in the position either preceding or following the negator "ไม่" (not) in a negative sentence.

4) Pre-verb auxiliary in imperative mood (XVBB)

It is used preceding a verb, usually in the initial position of an imperative sentence to show the grammatical function of imperative mood.

5) Post-verb auxiliary (XVAE)

It is used following a verb.

2.11.5 Determiner

It is used with a noun restricting the meaning of the noun in some way and can be divided into 9 subcategories.

1) Definite determiner: after a noun without a classifier in between (DDAN)

It is used immediately after a noun and no classifier allowed in between.

2) Definite determiner: allowing a classifier in between (DDAC)

It follows a noun or a classifier.

3) Definite determiner: between a noun and a classifier, or preceding a quantitative expression (DDBQ)

It is used between a noun and a classifier or, preceding a quantitative expression.

4) Definite determiner: following a quantitative expression (DDAQ)

It is used after a quantitative expression.

5) Indefinite determiner: following a noun and allowing a classifier in between (DIAC)

It is used after a noun with or without a classifier in between.

6) Indefinite determiner: between a noun and a classifier, or preceding a quantitative expression (DIBQ)

It is used between a noun and a classifier or, preceding a quantitative expression.

7) Indefinite determiner: following a quantitative expression (DIAQ)

It is used following a quantitative expression. For example, กว่า (over), เศษ (more).

8) Determiner: cardinal number expression (DCNM)

It is a cardinal number used in quantitative expression

9) Determiner: ordinal number expression (DONM)

It is an ordinal number used to express the order.

2.11.6 Adverb

It describes or modifies to the meaning of a verb, another adverb or a sentence and can be divided into 4 subcategories.

1) Adverb with a normal form (ADVN)

It is used in the base form and is not in the form of repetition or is not derived from a verb by adding an adverbial prefix "**โดย**" or "**อย่าง**".

2) Adverb with an iterative form (ADVI)

It is used in an iterative form by combining with "**ๆ**".

3) Adverb with an prefixed form (ADVP)

It is derived from a verb by adding a prefix such as "**โดย**" or "**อย่าง**".

4) Sentential adverb (ADVS)

It shows the speaker's attitude to or evaluation of what is said in the rest of the sentence.

2.11.7 Classifier

It is used with a noun showing the sub-class to which a noun belongs to express the unit of a noun in counting. It can be divided into 5 subcategories.

1) Unit classifier (CNIT)

It represents an individual thing, animal or person.

2) Collective classifier (CLTV)

It is used to represent a set, group, class or type of thing, animal or person.

3) Measurement classifier (CMTR)

It is used to represent the measuring unit of a noun

4) Frequency classifier (CFQC)

It is used to represent the frequency of a noun.

5) Verbal classifier (CVBL)

It derives from a verb to represent the unit of a noun.

2.11.8 Conjunction

It joins words, phrases, or clauses together. The conjunction is divided into 3 subcategories.

1) Coordinating conjunction (JCRG)

It joins linguistic units which are equivalent or of the same rank.

2) Comparative conjunction (JCMP)

It joins a standard to which someone or something is referred and an independent clause.

3) Subordinate conjunction (JSBR)

It joins an independent clause and a dependent clause

2.11.9 Preposition

It is used preceding a noun or a pronoun to link it grammatically to other word or phrase (RPRE).

2.11.10 Interjection (INT)

It indicates an emotional state or attitude such as delight, surprise, shock, and disgust, but has no referential meaning.

2.11.11 Prefix

It is added to the beginning of a word changing the meaning or function of the word and can be divided into 2 subcategories.

1) Nominal prefix (FIXN)

It is added to the beginning of a verb (or a verb phrase) changing the function of the verb (or the verb phrase) to be a noun (or a noun phrase).

2) Adverbial prefix (FIXV)

It is added to the beginning of a verb (or a verb phrase) changing the function of the verb (or the verb phrase) to be an adverb (or an adverb phrase).

2.11.12 Sentence Particle

1) Ending for affirmative sentence (EAFF)

It is added to the end of a sentence to indicate the mood of an utterance.

2) Ending for interrogative sentence particle (EITT)

It is added to the end of a sentence to indicate the mood of interrogative, and yes-no question.

2.11.13 Negator (NEG)

It is used to negate a sentence.

2.11.14 Punctuation (PUNC)

A mark or a sign (such as an opening/closing bracket, opening/closing quote, comma, colon, semicolon, dash, exclamation mark, period, and question marker).

CHAPTER 3

THE PROPOSED FRAMEWORK

3.1 The Proposed Framework for Developing Thai to English Machine Translation Systems

The proposed framework adopts the interlingua rule-based approach for the process of translating a Thai sentence into an Interlingua and translating the Interlingua into the corresponding English sentence. The reasons to choose the interlingua rule-based paradigm instead of the Statistical approach, are the follows: First, the rule-based approach does not need large bilingual corpora compared to the statistical one. The preparation of a large corpus of several million words of Thai-English translations is very time consuming. Second, we believe that the MT system must be built incrementally to gradually enhance the quality of a translation and the rule-based approach is more suitable as the system can be extended with new rules to cope with more complex sentences. Some translation problems - such as lexical ambiguity and structural ambiguity - can be handled and resolved by using rules. Third, the rule-based MT approach can provide details or explanations of how the translation steps are performed to achieve the output translation. Fourth, a Statistics-based MT system requires a process to search for the best match between sentences in a corpus and the given sentence and so requires a more powerful machine to run it compared to the rule-based one . Finally, the proposed framework adopts an interlingua approach so the interlingua can be extended to support a sentence translation from one to many languages while the statistical approach translates only one pair of languages at a time.

Our framework to develop Thai to English Machine Translation System can be divided into four phases as shown in figure 3.1

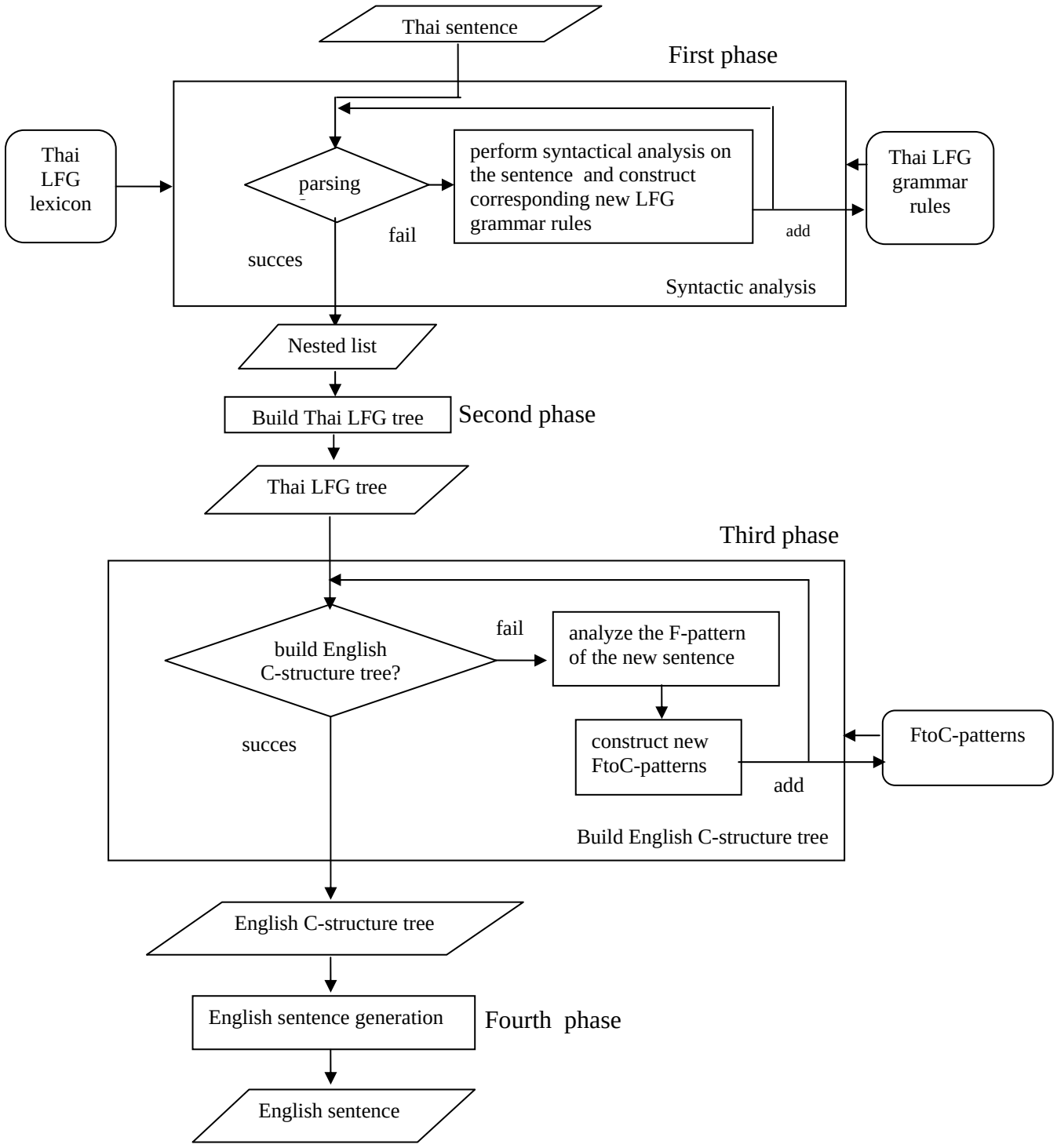


Figure 3.1 The Framework to Develop a Thai to English Machine Translation System

3.2 Phases of the Framework of Thai to English Machine Translation Systems

Each phase of the framework can be described as follows

3.2.1 First Phase: Thai Syntactical Analysis

This phase performs a syntactical analysis on a given Thai sentence using LFG parser in GFU-LAB. The parser was modified to parse Thai sentences and generate Interlingua. The syntactic analysis requires two sets of data as follows,

3.2.1.1 Thai LFG Grammar Rules

To develop Thai LFG grammar rules (Amara et al., 2003); Navavan Bandhumedha, 2006; Nittaya Kanchanawan, 1999) about 200 noun phrases, verb phrases, preposition phrases and simple sentences from Orchid Corpus were analyzed using IC (Immediate constituent) analysis.

Orchid Corpus (Open linguistic Resource Channeled toward Interdisciplinary research) was established to support research in natural language processing. It is a cooperative research between Linguistic and Knowledge Science Laboratory (LINKS) of NECTEC, Thailand and Kansai Advanced Research Center (KARC) of Japan. The Orchid Corpus contains about 2 MB of texts from proceeding of a NECTEC annual conference. Texts in the corpus are tagged using the primary part-of-speech (POS) tag set of LINKS. Texts in the corpus are divided into three levels: paragraph, sentence, and word levels. Probabilistic trigram techniques are used in word segmentation and POS tagging.

Each phrase or sentence is repeatedly divided into successive constituents until the smallest constituents - consisting of only a word or meaningful part of a word - are reached. Each step of the constituent division can then be translated into a Thai grammar rule in a context-free format. Finally, to convert the Thai grammar rules in context-free format into LFG grammar rules, functional annotations must be added, some the grammatical functions based on LFG grammar such as subject, object, complement and so on, must be identified through Thai case grammar analysis (Phanu Sungkhavorn, 1984: 27-123). There are about 218 Thai

grammar rules that were derived from data from Orchid Corpus and used in the system. These grammar rules can be categorized into two levels, namely sentence level and phrase level (e.g. noun phrase, verb phrase and prepositional phrase). The process of constructing Thai LFG grammar rules from a given Thai sentence can be shown in figure 3.2

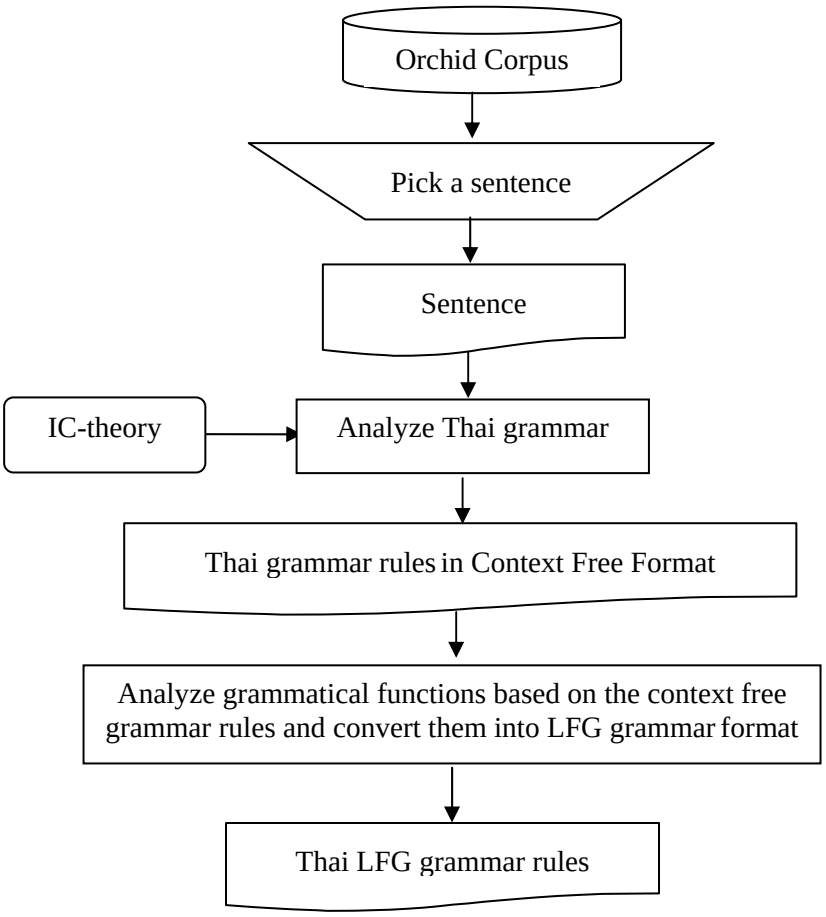


Figure 3.2 The Processes to Develop Thai Grammar Rules

For example, consider the following Thai sentence from Orchid corpus shown in figure 3.3

คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย // คอมพิวเตอร์/ NCMN ถ่าย/VACT ภาพตัดขวาง/ NCMN ร่างกาย/NCMN

Figure 3.3 A Sample Sentence from Orchid Corpus

The steps of deriving the grammar rules from the given sentence using IC-theory can be shown in figure 3.4. The output from the analysis is a parse tree shown in figure 3.5.

คอมพิวเตอร์	ถ่าย	ภาพตัดขวาง	ร่างกาย
คอมพิวเตอร์	ถ่าย	ภาพตัดขวาง ร่างกาย	
คอมพิวเตอร์	ถ่าย ภาพตัดขวาง ร่างกาย		
คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย			

Figure 3.4 The Steps of Deriving Grammar Rules from a Thai Sentence "คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย" using IC-theory

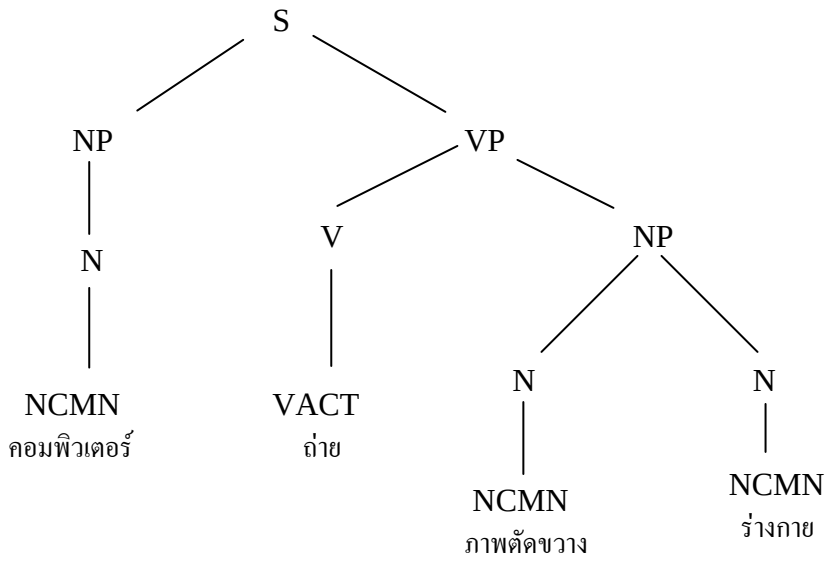


Figure 3.5 An Output Parse Tree of the Given Sentence from the Analysis using IC-Theory

The grammar rules in context-free grammar form are then extracted from the parse tree as shown in figure 3.6

S → NP VP

NP → Noun Noun

VP → Verb

NP → Noun

Noun → NCMN [คอมพิวเตอร์]

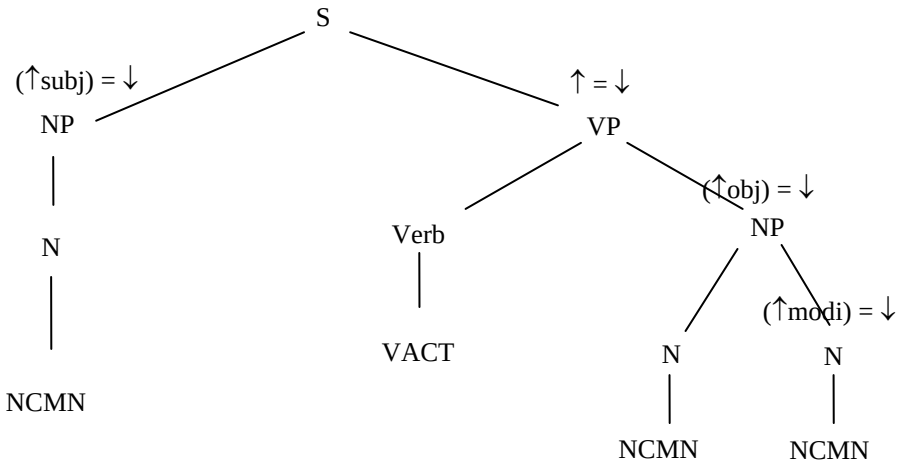
Noun → NCMN [ภาพตัดขวาง]

Noun → NCMN [ร่างกาย]

Verb → VACT [ถ่าย]

Figure 3.6 Context-free Grammar Rules Corresponding to the Given Thai Sentence

Finally, the context-free grammar rules are converted into LFG grammar format, as shown in figure 3.7. Notice that functional annotations are added into the context free grammar rules



S → NP VP
(↑subj) = ↓

NP → Noun Noun
(↑modi) = ↓

VP → Verb NP
(↑obj) = ↓

NP → Noun

Noun → NCMN [คอมพิวเตอร์]

Noun → NCMN [ภาพตัดขวาง]

Noun → NCMN [ร่างกาย]

Verb → VACT [ถ่าย]

Figure 3.7 Final Thai LFG Grammar Rules Derived from the Given Sentence

3.2.1.2 Thai LFG Lexicon

The Thai LFG lexicon contains information on 800 Thai words in LFG format. Each of the Thai words in the lexicon is tagged with its part of speech e.g. noun, verb, classifier, determiner, preposition and so on. For example, the Thai word “คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย” contains information in LFG format as follows

คอมพิวเตอร์ = NOUN

head = คอมพิวเตอร์

pred = computer
type = NCMN
ontology = computer

ถ่าย = VERB

head = ถ่าย

pred = snapshot
type = VACT
ontology = verb

ภาพตัดขวาง = NOUN

head = ภาพตัดขวาง

pred = cross cut picture
type = NCMN
ontology = sight

ร่างกาย = NOUN

head = ร่างกาย

pred = body
type = NCMN
ontology = person

The output of the first phase, after a given Thai sentence or phrase has been parsed, is a nested list structure, consisting of two structures of LFG, namely C-structure and F-structure. If the given Thai sentence or phrase can be parsed successfully, the output of that Thai sentence or phrase is forwarded to the second phase. Otherwise, the syntax of the Thai sentence or phrase is analyzed using IC theory and new Thai grammar rules founded by the analyses will be added to the system so that the sentence or phrase can be successfully parsed. Figure 3.8 (a) shows an example output of the syntactical analysis for given a Thai phrase “คณะ วิศวกรรมศาสตร์“ the output of the syntactical analysis is as figure 3.8 (a)

[[modi=[wordtype=CLASSIFIER, head=คณะ, sem=[pred=faculty, type=CLTV, num=pl, ontology=community]], wordtype=NOUN, head=วิศวกรรมศาสตร์, sem=[type=NPRP, pred=engineering, ontology=applied_science, modifier=[pred=faculty, type=CLTV, num=pl, ontology=community]]]]

Figure 3.8 (a) The Output Resulting from Syntactical Analysis for a Thai Phrase “คณะ วิศวกรรมศาสตร์“

Another output example from the syntactic analysis is shown in figure 3.8 (b). The output for a Thai sentence is “คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย“. Note that Thai words in a given sentence are transformed into equivalent English words which are stored inside the PRED nodes of the F-structure.

[[subj=[wordtype=NOUN, head= คอมพิวเตอร์, sem=[pred=computer, type=NCMN, ontology=computer, case=agentive]], obj=[sem=[case=objective, pred=cross cut picture, type=NCMN, ontology=sight, modifier=[pred=body, type=NCMN, ontology=person_body]], modi=[sem=[pred=body, type=NCMN, ontology=person_body], wordtype=NOUN, head=ร่างกาย], wordtype=NOUN, head= ภาพตัดขวาง], wordtype=VERB, head=ถ่าย , sem=[pred=snapshot, type=VACT, ontology=verb, subject=[case=agentive, pred=computer, type=NCMN, ontology=computer], object=[case=objective, pred=cross cut picture, type=NCMN,

Figure 3.8 (b) The Output from Syntactic Analysis Performed on the Sentence “คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย“

3.2.2 Second Phase: Building a Thai LFG tree

In this phase the output from the first phase- which is in the nested list format- is converted into a general tree structure, called a Thai LFG tree. The tree structure will serve as an interlingua for the translation. The LFG tree comprises nodes that hold syntactic and semantic information of the given Thai sentence. i.e. C-structure and F-structure. For example, the output of figure 3.8 (a) can be converted into a corresponding Thai LFG tree as shown in figure 3.9 (a)

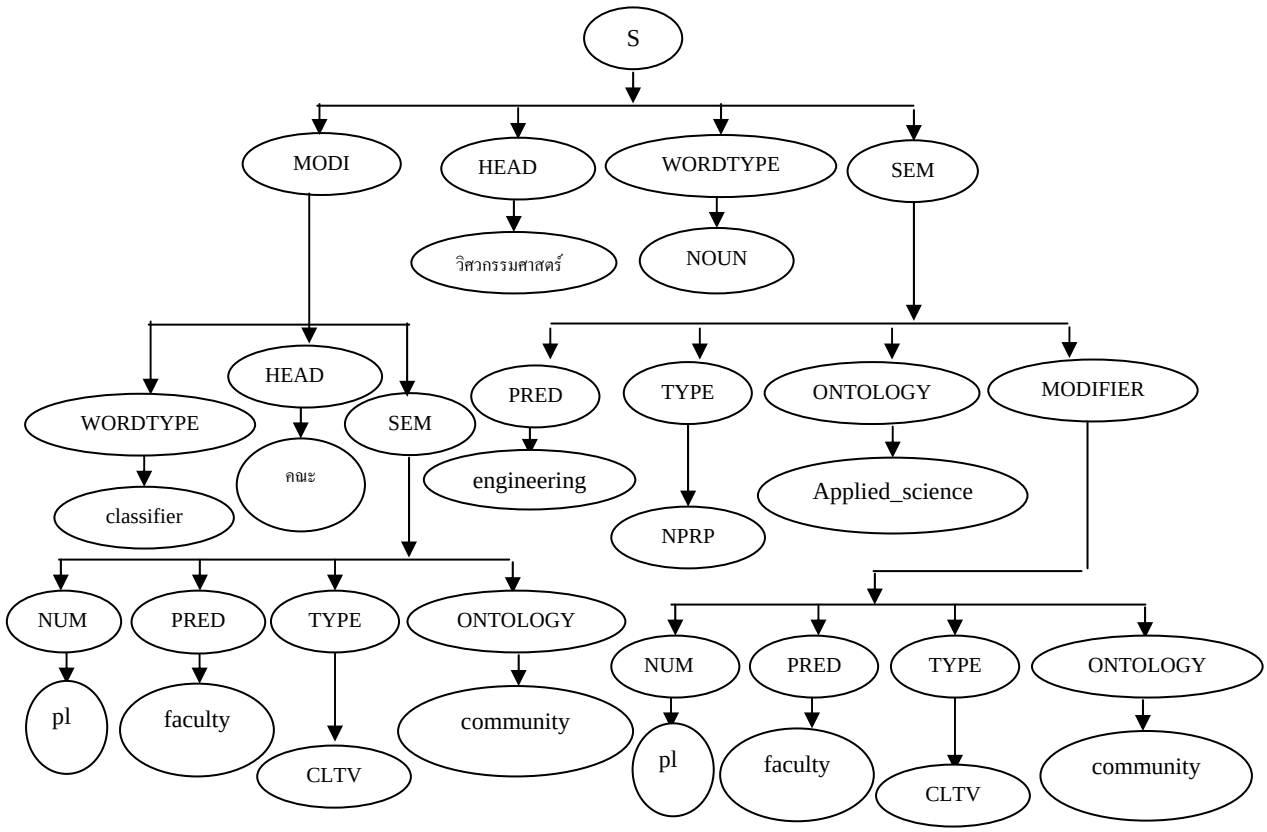


Figure 3.9 (a) The Thai LFG tree (Interlingua) of the Phrase "คณะ วิศวกรรมศาสตร์"

The output of figure 3.8 (b) can also be converted into a corresponding Thai LFG tree as shown in figure 3.9 (b)

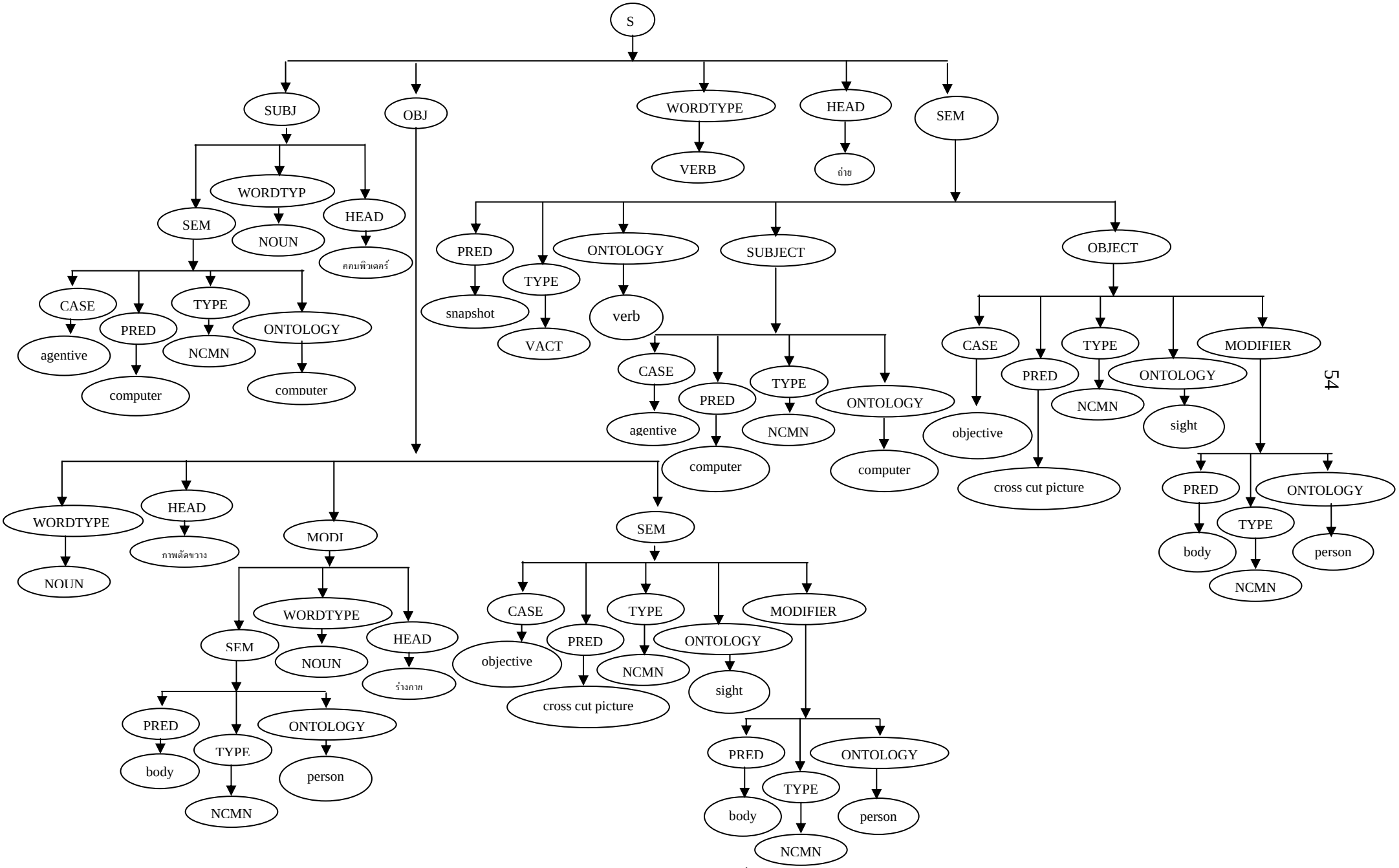


Figure 3.9 (b) Thai LFG tree (Interlingua) of the Sentence “คอมพิวเตอร์ ถ่าย ภาพตัดขวางร่างกาย”.

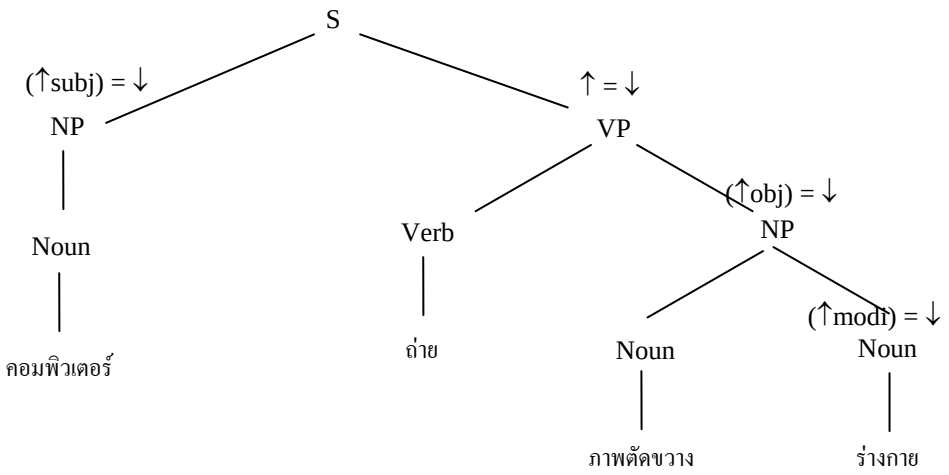


Figure 3.10 C-Structure Tree of Thai Sentence “คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย“

The Thai LFG tree shown in figure 3.9 (b) consists of a C-structure corresponding with the C-structure tree shown in figure 3.10. In particular, the root of the Thai LFG tree S has a Subj, Obj wordtype, and head Sem subtrees. The Subj subtree contains information of the subject “คอมพิวเตอร์”, which is extracted from the lexicon. The word type and head subtrees contain information on the “ถ่าย”, which is extracted from the lexicon. The Obj subtree contains information on the object which is a noun phrase consisting of “ภาพตัดขวาง” as the head of the phrase and the ”ร่างกาย” as the modifier of the phrase. Finally, the SEM subtree, representing F-structure contains the semantic information for all components of the sentences e.g. subject, object, and verb.

3.2.3 Third Phase: Building English C-Structure Tree

In this phase, an English C-structure - equivalent to the F-structure of Thai LFG tree derived from the second phase - is created through a pattern matching and transformation process. Predefined patterns, called FtoC patterns, are used for the process. Each of the FtoC patterns contains a mapping between the substructure pattern of F-structure and its corresponding substructure pattern in the English C-structure. A substructure of F-structure in the LFG tree is matched against

substructure patterns of FtoC patterns, and if a match is found, a corresponding substructure of C-structure (English) is created. The advantages of using FtoC patterns to construct C-structure (English) out of the given F-structure, are as follows. First, the FtoC patterns can be constructed through pattern analysis between the F-structure produced from Thai sentences and their corresponding English C-structures. Second, FtoC patterns are basically mapping rules which can be added to or modified easily to handle unseen sentences or phrases in the future. These mapping rules are suitable for incrementally building the machine translation system. The FtoC patterns can be divided into two sets, one for the sentence level and the other for the phrase level.

3.2.3.1 Sentence Level Patterns

This level contains functional information about subject, object and so on. The sentence level patterns are the matching between functional information of a given Thai sentence and functional information of the English sentence. The ordering of functional information of the equivalent English sentence for a given Thai sentence is based on English grammar. For example, if the Thai sentences contain Subject, Verb and Object nodes respectively, then, the corresponding pattern of the equivalent English sentence is the following: Subject, Verb and Object. The current implementation of the system employs 25 FtoC patterns for the sentence level.

3.2.3.2 Phrase Level Patterns

This level contains functional information and additional modifier information. The phrase level patterns are the matching between functional information and additional modifier of a given Thai phrase and functional information and additional modifier of the equivalent English phrase. The ordering of this information in the English phrase for a given Thai phrase is dependent on the English phrase grammar. To perform pattern matching of nodes at phrase level the type of information of nodes kept in the F-structure must be used for the matching. For example, a given Thai phrase “คณะ/CLTV วิศวกรรมศาสตร์/NPRP” type or Orchid tag sets of คณะ(CLTV) and วิศวกรรมศาสตร์(NPRP) are used during node matching.

The FtoC patterns are predefined and need to be developed through the pattern analysis process. Figure 3.10 shows the process of developing FtoC patterns through the pattern analysis (performed by a linguistic expert). The current implementation of the system employs 116 FtoC patterns for noun phrases, verb phrases and preposition phrases.

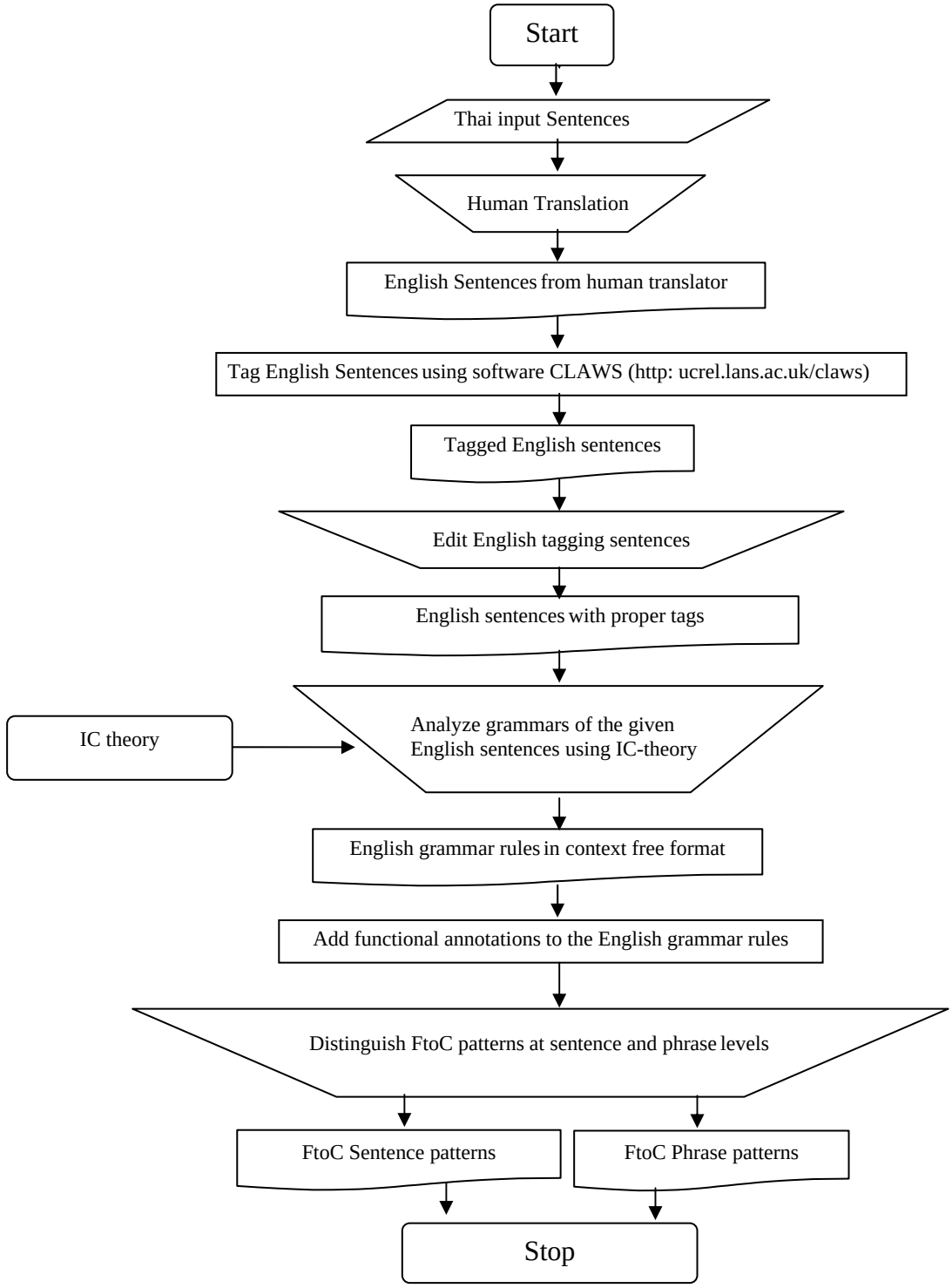


Figure 3.11 Process to Develop FtoC-Patterns

There are two steps in the third phase. The first step matches the child nodes of the root of F-structure tree against predefined FtoC patterns at the sentence level. If a match is found, the child nodes are rearranged and converted into the equivalent nodes of English C-structure at the sentence level.

The second step is performed recursively at each lower level of the root or the phrase levels. Phase-level transformation is performed for each of the lower levels. The phase-level transformation involves matching the current node and its children against the predefined set of phase level FtoC patterns. If a match is found, necessary rearrangement and transformation takes place. The transformation process is performed recursively from the root to the leaves of the F-structure tree until the whole tree structure has been processed. Each of the FtoC patterns consists of a sequence of constitutes at the right hand side of a Thai LFG grammar rule. Examples of FtoC patterns at a sentence level are shown in Table 3.1 while some examples of FtoC patterns at a phrase level are shown in Table 3.2. The symbols used in the table are as follows: V_1, V_2, and V_s represent finite verbs while V_3 denotes a participle verb. NP represents a noun phrase and PP denotes a preposition phrase. Like the grammar rules used in the first phase, the FtoC patterns are built incrementally. If an equivalent English C-structure is constructed successfully, the output English C-structure will be forwarded to the next phase. Otherwise the syntactical structure of the sentence or phrase is analyzed and new corresponding FtoC-patterns along with necessary ordering/conversion modules are added into the system so that the equivalent English C-structure can be successfully constructed.

Table 3.1 Examples of Sentence Level FtoC-Patterns

	F-structure Patterns				English C-structure Patterns			
PASSIVE VOICE	**VERB**	**SUBJECT**	**OBLIQUE PHRASE**		**SUBJECT**	**VERB**	**OBLIQUE PHRASE**	
		(NP)	(PP)		(NP)	(is + V_3) (are + V_3)	(PP)	
FUTURE	**VERB**	**SUBJECT**	**OBJECT**		**SUBJECT**	**VERB**	**OBJECT**	
	(ADJUNCT + V)	(NP)	(NP)		(NP)	(will + V_1) (shall + V_1)	(NP)	
PAST	**VERB**	**SUBJECT**	**OBJECT 1**	**OBJECT 2**	**SUBJECT**	**VERB**	**OBJECT 2**	**OBJECT 1**
	(ADJUNCT + V)	(NP)	(NP)	(NP)	(NP)	(V_2)	(NP)	(NP)
PRESENT	**VERB**	**SUBJECT**	**OBJECT**		**SUBJECT**	**VERB**	**OBJECT**	
	(V + Modifier)	(NP singular)	(NP)		(NP singular)	(Vs + Modifier)	(NP)	
	VERB	**SUBJECT**	**OBJECT**		**SUBJECT**	**VERB**	**OBJECT**	
	(V + Modifier)	(NP plural)	(NP)		(NP plural)	(V_1 + Modifier)	(NP)	

Table 3.2 Examples of Phrase Level FtoC-Patterns

	F-structure Patterns				English C-structure Patterns			
NOUN Phrase	**NOUN**	**MODIFIER**		**The**	**MODIFIER**	**of**	**NOUN**	
	(Proper Noun)	(Classifier)			(Classifier)		(Proper Noun)	
	NOUN	**MODIFIER**		**The**	**MODIFIER**	**NOUN**		
	(Common Noun)	(Ordinal Number)			(Ordinal Number)	(Common Noun)		
	NOUN	**MODIFIER**		**The**	**MODIFIER**	**of**	**NOUN**	
	(Proper Noun)	(Common Noun)			(Common Noun)		(Proper Noun)	
	NOUN	**MODIFIER1**	**MODIFIER2**	**The**	**MODIFIER2**	**MODIFIER1**	**of**	**NOUN**
	(Proper Noun)	(Common Noun)	(Common Noun)		(Common Noun)	(Common Noun)		(Proper Noun)
	NOUN	**MODIFIER**			**MODIFIER**	**NOUN**		
	(Common Noun)	(Common Noun)			(Common Noun)	(Common Noun)		
PREPOSITION Phrase	**PREPOSITION**	**OBJECT**			**PREPOSITION**	**OBJECT**		
		(Noun phrase)				(Noun phrase)		
VERB Phrase	**VERB**	**NEGATOR**	**OBJECT**		**DO + NOT**	**V_1**	**OBJECT**	
			(Noun phrase)				(Noun phrase)	

Let's consider the construction process of the English C-structure tree for a Thai sentence "คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย" (computer snapshot a body cross cut picture). As the output from the second phase - Thai LFG tree- contains both C-structure and F-structure, only the F-structure will be used for the English C-structure tree construction. The F-structure of the processed Thai sentence is shown in figure 3.12

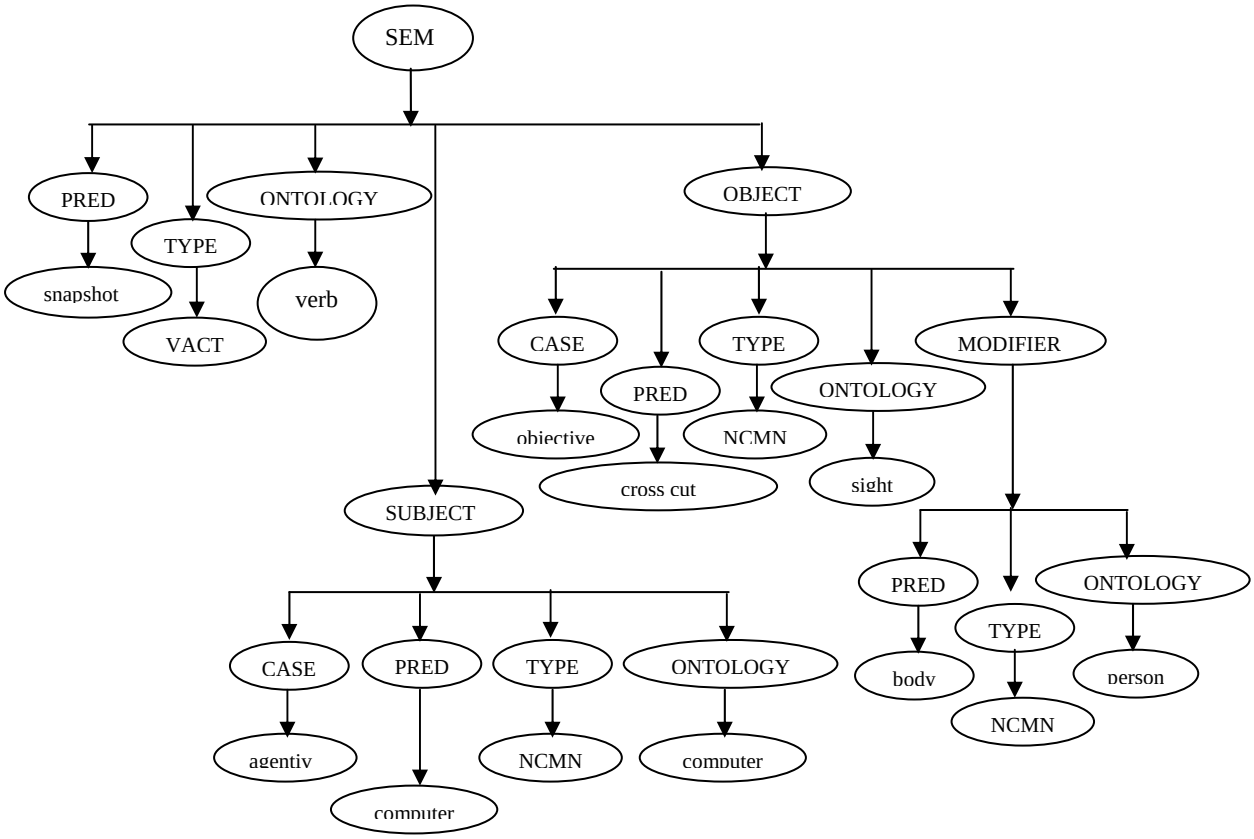


Figure 3.12 The F-Structure of the Sentence "คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย"

Next, the F-structure of the Thai LFG tree is compared against FtoC-patterns at sentence level patterns.

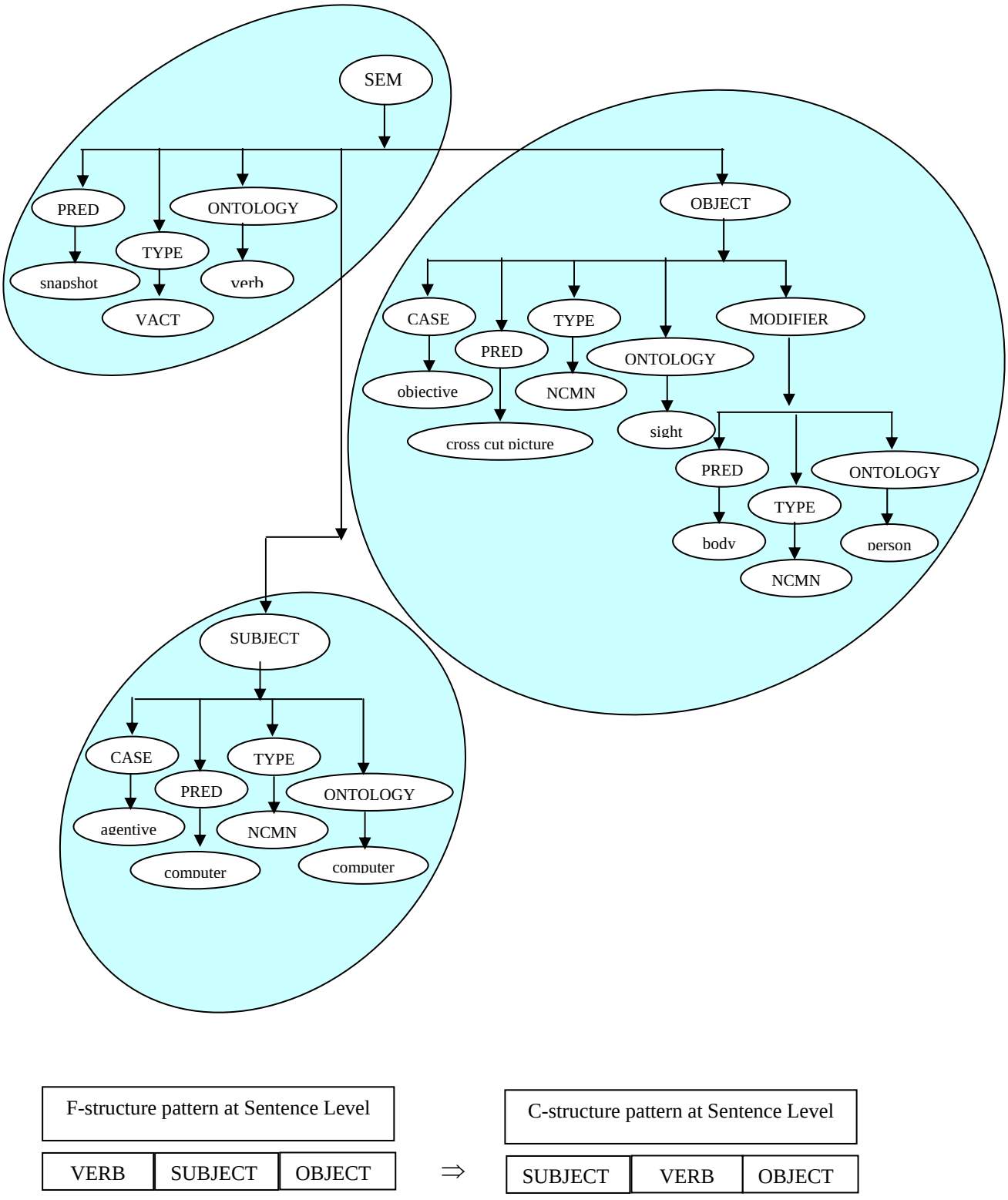


Figure 3.13 (a) A Match Found Between the F-Structure of a Thai LFG Tree and an FtoC Pattern at Sentence Level

Once a match is found as shown in figure 3.13 (a), An English C-structure tree that has aStart node is partially created according to the C-structure pattern of the matched FtoC patterns. The root of the newly created tree contains three sub-trees, namely Subject Sub-tree, Verb Sub-tree and Object sub-tree, as shown in figure 3.13(b).

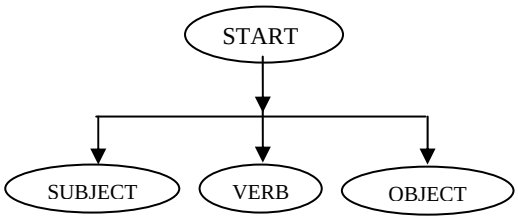


Figure 3.13 (b) The Partially Created English C-Structure Tree at the Sentence Level

The matching process continues for each subtree of the root of the F-structure (see figure 3.13 (c). As to the Subject subtree, this subtree contains only the information about the word “คอมพิวเตอร์ ” with no child subtrees. Thereforeall information about the word is transferred to the Subject node of the English C-structure tree currently under construction. As to the Verb subtree, it contains the information about the word “ ถ่าย” and has no child subtrees. So, all information about the word is transferred to the Verb node of the English C-structure tree. Finally, the Object subtree contains information about the word “ ภาพตัดขวาง” and a Modifier subchild containing information about the word “ร่างกาย”. The matching process must therefore continue at the phrase level on the Object subtree.

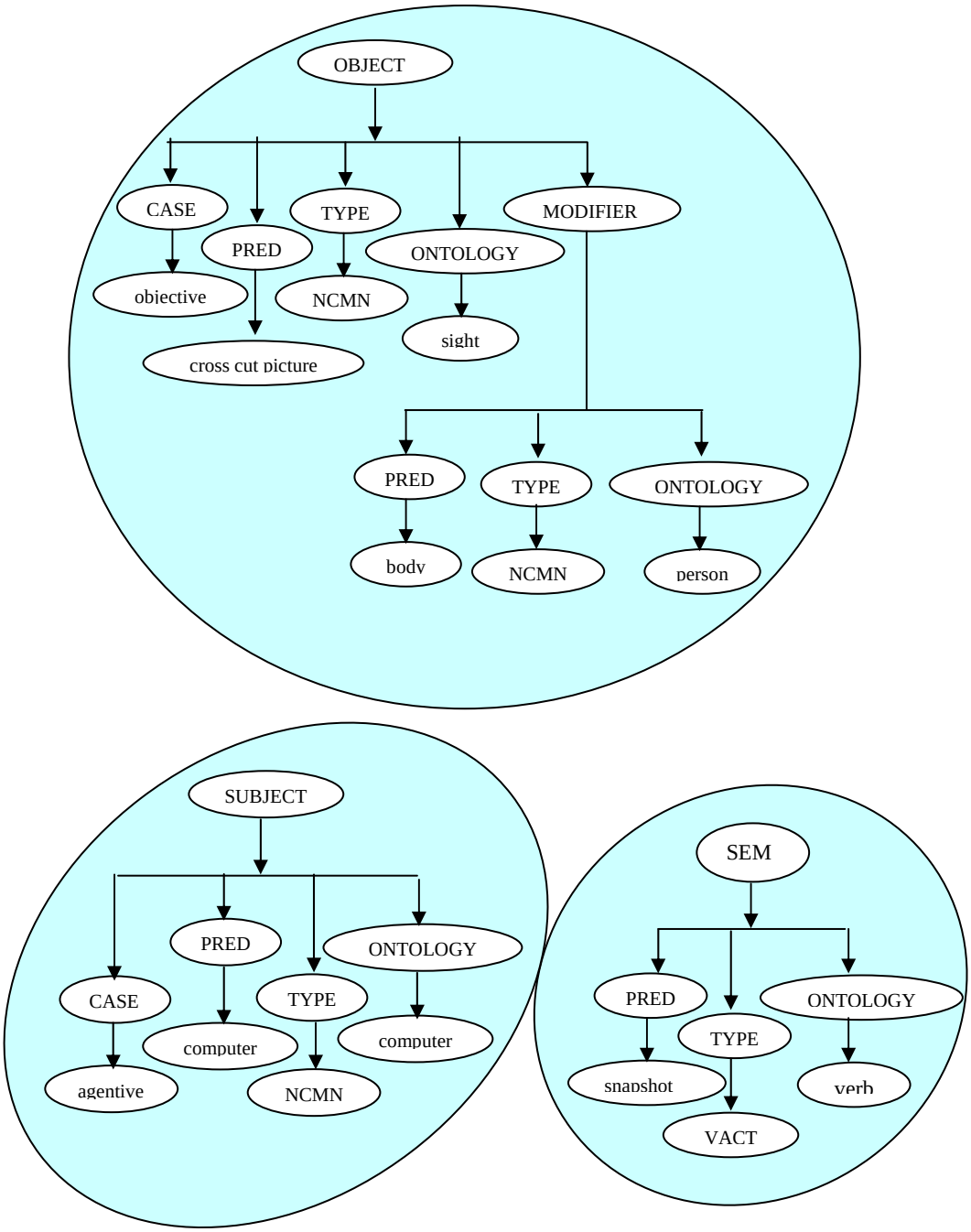
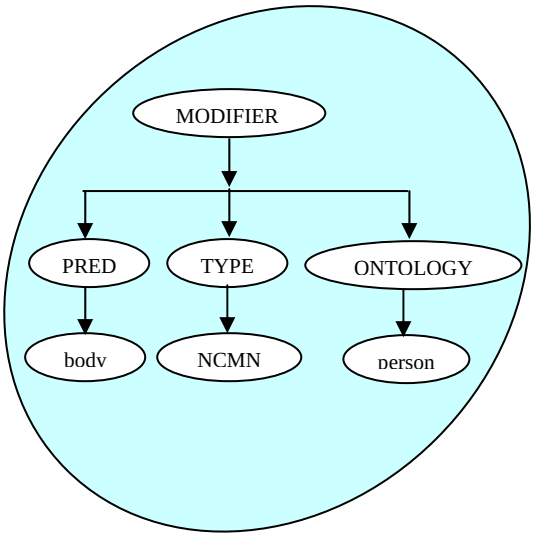
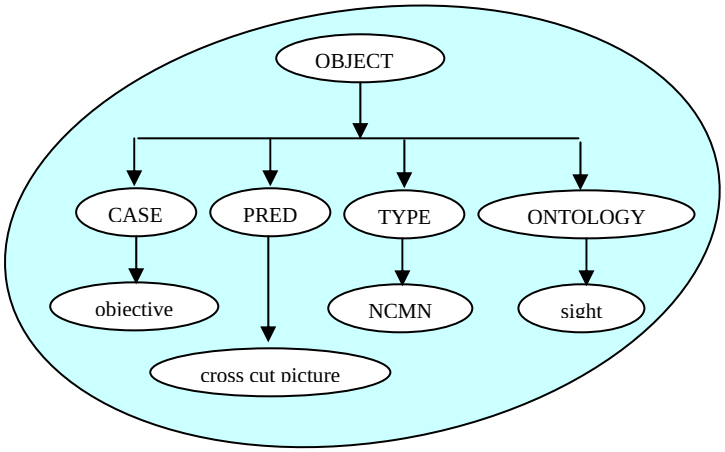


Figure 3.13 (c) The Subtrees of Root of the F-Structure for the Given Thai Sentence

Next, the root and its children of the Object sub-tree are matched against the FtoC-patterns at the phrase level. A match is found, as shown in figure 3.13 (d)





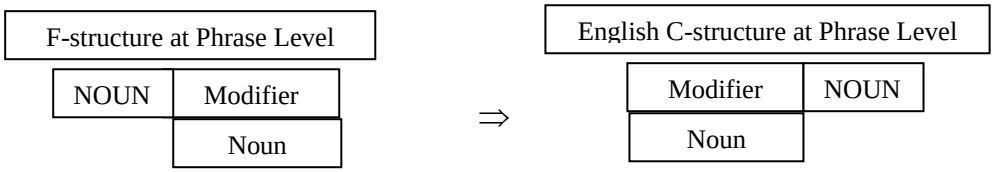


Figure 3.13 (d) A Match is Found Between the Object Subtree and a FtoC-pattern at Phrase Level

Since the Object subtree contains two children, the noun word “cross cut picture” and a Modifier. The information of the noun word and information of the modifier are transferred to the corresponding Noun and Modifier in the English C-structure. The output of the final step is the English C-structure tree shown in figure 3.13 (e)

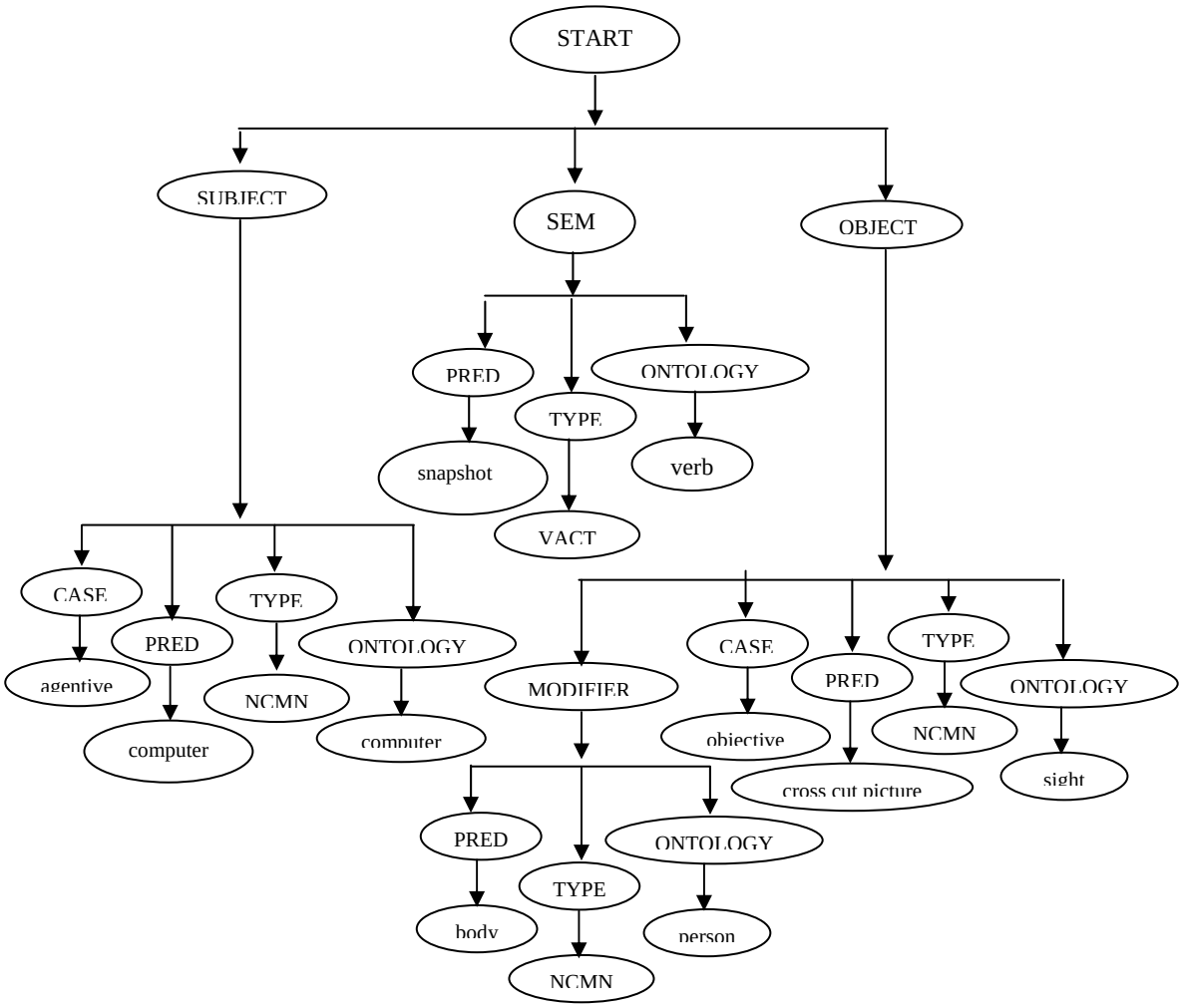


Figure 3.13 (e) The Final English C-Structure Tree

Tense, voice and article of nouns in the English output sentence must be in accordance with the meaning of the given Thai sentence. Some words to indicate tense, voice or articles must be inserted into the output sentence. Other word

transformations must be performed to indicate plurality of nouns, a compound word and an English plural noun should end with -s or -es. For instance, a Thai phrase “การ/FIXN พัฒนา/VACT”, which contains two parts of Orchid tag set - Nominal prefix and Active verb - must be replaced with only one English word “development”.

3.2.4 Fourth Phase: Generating an English Sentence

To generate an English sentence, the system traverses an English C-structure tree from left to right and prints out all English words kept in the leaf nodes. The output from this fourth phase will be the English phrases/sentences corresponding with the given Thai sentence.

CHAPTER 4

EVALUATION OF THE PROPOSED FRAMEWORK

To verify whether the proposed framework can be used effectively for building a Thai-to-English machine translation system, a system prototype is built according to the framework. Two stages of building the prototype are implemented so the prototype is constructed incrementally according to the framework.

4.1 The First Stage: Building a Preliminary System

In this stage about 100 Thai phrases and simple sentences (no serial verb construction) are selected from the Orchid Corpus, with Thai grammar rules being created to support the phrases and sentences using the IC-analysis. A Thai LFG lexicon based on the Thai phrases and simple sentences is also created. Following the framework, each of the phrases/sentences is input into the prototype system one at a time. The grammar rules and FtoC patterns were built incrementally. After the completion of the first stage, the prototype can handle the translation of phrases and simple sentences. Hence, the prototype can be used as the preliminary system for extension so the system can handle the translation of more complex sentences.
All translations resulting from the preliminary system were recorded for subsequent evaluation of the quality of the translations.

4.2 The Second Stage: Enhancing the Preliminary System

An additional 100 more complex sentences (no serial verb constructions) were selected from the corpus and then processed in the same way as in the first stage. The lexicon for those words contained in the 100 complex sentences were also added into the current Thai LFG lexicon. Most of the additional 100 sentences could not be parsed successfully because the current grammar rules in the system were applicable

to simple sentences and phrases only. Hence, grammar rules associated with each sentence that could not be parsed successfully were analyzed using IC analysis and subsequently added into the system. In addition, grammar rules and the FtoC patterns associated with each of the additional phrases/sentences must be created or modified in order to successfully translate the given Thai sentences/phrases. For the second stage, FtoC patterns were added or modified while a few FtoC phrase level patterns were only partially added or modified as they had been almost fully created in the first stage. The iterative processes were performed until all additional 100 complex sentences were successfully translated. As more sentences were input into the prototype system, its capability increases and some sentences can be translated successfully without any changes or modifications to the system.

4.3 Evaluation of Quality of the Output Translation of the Prototype System

To evaluate the quality of the translation of the prototype, three senior master students of the School of Language and Communication of National Institute of Development Administration (NIDA), Thailand, who had taken courses in translation, were presented with the 200 input phrase/sentences as well as the corresponding results of the machine translations from the prototype. The students were asked to assess and classify the quality of the translation into three levels. The first level - with a score of 3 - is given to a translation that is acceptable, which corresponds to that the output sentence could be understood and has the same meaning as the source sentence. The second level - with a score of 2 - is given to a translation that is moderately acceptable. This means that the output sentence might have small errors but could still be understood and carried the same meaning as the source sentence. The third level - with a score of 1- is given to a translation that is not acceptable or has to be rejected outright. This means that the output sentence could not be understood or carries the same meaning as the source sentence. The scores given by the three students are summarized in Table 4.1

Table 4.1 The Summary of Scores Given by the Three Students

	Acceptable (3) (%)	**Moderate(2) (%)**	**Rejected (1) (%)**
First student	22	57	21
Second student	31.5	48	20.5
Third student	32.5	37.5	30
Average percentages	28.7	47.5	23.8
Average Score		**2.05**	

Table 4.1 shows that most of the translations (about 76.2 %) produced by the prototype system are either acceptable or moderately acceptable. The remainder - about 23.8 % - are rejected. The average score given by the three students was 2.05.

In addition to the 200 tested phrases/sentences from the orchid corpus 50 additional Thai phrases/sentences were created, using existing Thai grammar rules and lexicons. They were also entered into the system. The prototype system was able to successfully translate these additional phrases/sentences. The additional phrases/sentences and their corresponding translation output are listed in appendix L. The results from the evaluation demonstrate that the prototype built according to the framework can be an effective tool for Thai-to-English machine translation, although there are still improvements required in the future, for example improvements in the Thai to English word selection, FtoC mapping as well as the transformation process.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusions

This research proposes a new framework to develop Thai to English machine translation incrementally. The framework is an iterative process for which complex grammar rules, lexicons and FtoC patterns can be added incrementally into the system. Each iteration of the process consists of four stages: syntactic analysis, building a Thai LFG tree, mapping the Thai F-structure of the Thai LFG-tree into an English C-structure tree, and, finally, English sentence generation. Left-corner parsing is the parsing technique used for syntactic analysis of the framework, as it combines the advantages of top-down and bottom-up parsing methods. The output from the parsing is a Thai F-structure tree for a given sentence. To transform a Thai F-structure tree into an equivalent English C-structure tree, the process of tree node matching and transformation is performed, starting from the root down to the leaves. FtoC patterns corresponding between Thai F-structure tree and English C-structure tree are predefined for matching and transformation. There are two pattern types, namely sentence level and phrase level patterns. The set of FtoC patterns can be modified and new FtoC patterns can be added to successfully translate unseen phrases/sentences. As the set of FtoC patterns can be easily expanded and modified, the machine translation system can be built incrementally.

5.2 Future Work

Although, the prototype system can achieve a certain level of success in terms of the quality of the translations, future development work should be pursued so that

the system can handle more complex sentences, tenses, voices and articles. Pattern matching of FtoC patterns in the third phase of the framework is limited to 1-to-1 matching between F-structure and English C-structure nodes. In the future, pattern matching should be extended to m-to-n matching. For instance, m-to-n matching will be required for the translation of a Thai sentence shown in Figure 5.1

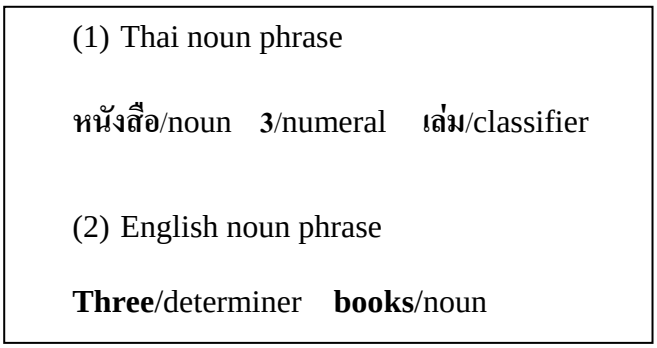


Figure 5.1 A Thai Noun Phrase that Requires m-to-n Matching of FtoC patterns for its Translation

In figure 5.1 a Thai noun phrase contains the word “หนังสือ” as a head noun, “3” as a numeral, and “เล่ม” as a classifier, while its corresponding English noun phrase contains only a noun “book” and its modifier “three”. As the English phrase contains fewer nodes than its Thai counterpart, the transformation cannot be performed through 1-to-1 mapping/transformation. Finally, the framework may need an additional process to select the most appropriate English word out of a set of Thai words with the same meaning. This process should yield better translation quality for the system.

Some parts of the proposed framework require some human assistance, for instance, the construction processes of LFG grammar rules, lexicons, FtoC patterns and pattern transformation procedures. Research on building automatic tools to learn Thai grammar rules from corpus and to learn FtoC patterns and their transformation from a corpus of translation outputs could be pursued, potentially resulting in less human involvement being required in the system development.

BIBLIOGRAPHY

Aho, Alfred V.; Sethi, Ravi and Ullman, Jeffrey D. 1986. **Compilers: Principles, Techniques and Tools.** Reading, Mass: Addison-Wesley.

Amara Prasithrathsint; Yuphaphann Hoonchamlong; Saranya Savetamalya. 2003. **Grammatical Theories**. 2nd ed. Bangkok: Chulalongkorn University. (In Thai)

Arnold, Douglas; Balkan, Lorna; Meijer, Siety; Humphreys, R. Lee and Sadler, Louisa. 1994. **Machine Translation: An Introductory Guide**. London: Blackwells-NCC.

Azar Betty Schrampfer. 1989. **Understanding and Using English Grammar.** 2nd ed. New Jersey: Prentice-Hall.

Banchob Bandhumedha. 2002. **Structure of Thai.** 13nd ed. Bangkok: Ramkhamhaeng University. (In Thai)

Blackburn, Patrick; Bos, Johan and Striegnitz, Kristina. 2550. **Learn Prolog Now Version 1.2.5, 2550.** Retrieved August 13, 2004 from http://www.coli.uni-saarland.de/~kris/learn-prolog-now/html/node1.htm

Booncharoen Sirinaovakul and Kaewchai Chancharoen. 2002. English-Thai Structure-Based Machine Translation. **Computational Intelligence**. 18 (August): 294-312.

Boonserm Kijsirikul; Paisarn Charoenpornsawat and Surapant Meknavin. 1999. Comparing Winnow and RIPPER in Thai Named-Entity Identification. In **Proceedings of the Natural Language Processing Pacific Rim Symposium 1999 (NLPRS'99)**, Bejing, China.

Briffault, Xavier; Chibout, Karim; Grard, Sabah and Vapillon, Jrame. 1997. A Linguistic Engineering Environment Using LFG (Lexical Functionnal Grammar) and CG (Conceptual Graphs). In **Proceedings of the LFG97 Conference**, University of California, San Diego, Miriam Butt and Tracy Holloway Kings, eds., CSLI Publications.

CLAWS Part of Speech Tagger for English. Retrieved August 13, 2004 from http://ucrel.lancs.ac.uk/claws/

Cook, Walter A. S. J. 1989. **Case Grammar Theory.** Washington, DC: Georgetown University Press.

Cranias, Lambros; Papageorgiou, Harris and Piperidis, Stelios. 1994. A Matching Technique in Example-Based Machine Translation. In **Proceedings of the 15th Conference on Computational Linguistics Volume 1**, Kyoto, Japan Pp. 100–104.

Deitel, H. M. and Deitel, P. J. 2005. **Java: How To Program**. 6th ed. Upper Saddle River, NJ: Pearson.

Domnern Garden and Sathienpong Wannapok. 2006. **Thai-English Dictionary**. 3rd ed. Bangkok: Matichonbook. (In Thai)

Dorna, Micheal; Frank, Anette; Genabith, Josef Van and Emele, Martin C. 1998. Syntactic and Semantic Transfer with F-structures. In **Proceedings of the 17th International Conference on Computational Linguistics Volume 1,** Montreal, Quebec, Canada. Pp. 341–347.

Dorr, Bonnie. 1987. UNITRAN: An Interlingual Approach to Machine Translation. In **Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence**, Seattle, WA.

Eisele, Andreas and Dörre, Jochen. 1986. A Lexical Functional Grammar System in PROLOG. In **Proceedings of the 11th Conference on Computational Linguistics**, Bonn, Germany. Pp. 551–553.

Emele, Martin C. and Dorna, Micheal. 1998. Ambiguity Preserving Machine Translation using Packed Representations. In **Proceedings of the 17th International Conference on Computational Linguistics Volume 1**, Montreal, Quebec, Canada. Pp. 365–371.

Falk, Yehuda N. 2001. **Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax.** California: Center for the Study of Language and Information Stanford.

Franz, Alexander; Horiguchi, Keiko; Duan, Lei; Ecker, Doris; Koontz, Eugene and Uchida, Kazami. 2000. An Integrated Architecture for Example-Based Machine Translation. In **Proceedings of the 18th Conference on**

Computational linguistics Volume 2, Saarbrücken, Germany. Pp. 1031–1035.

Frey, Werner and Reyle, Uwe. 1983. A Prolog Implementation of Lexical Functional Grammar as a Base for a Natural Language Processing System. In **Proceedings of the First Conference on European Chapter of the Association for Computational Linguistics**, Pisa, Italy. Pp. 52-57.

Furuse, Osamu and Iida, Hitoshi. 1992. Example-Based Method for Transfer-Driven Machine Translation. In **Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation.** Pp. 139-149.

Goodrich, Micheal T.; Tamassia, Roberto; Mount, David. 2004. **Data Structures and Algorithms in C++.** New York: Wileg. Pp. 533-564.

Graham, Yvette; Bryl, Anton and Genabith, Josef Van. 2009. F-Structure Transfer-Based Statistical Machine Translation. In **Proceedings of the LFG09 Conference**, Dublin City University, Miriam Butt and Tracy Holloway King, eds., CSLI Publications.

Güngördü, Zelal and Oflazer**,** Kemal. 1994. Parsing Turkish using the Lexical Functional Grammar Formalism. In **Proceedings of the 15th Conference on Computational Linguistics**, Kyoto, Japan. Pp. 494–500.

Güvenir**,** H. Altay and Cicekli, Ilyas. 1996. Learning Translation Templates from Bilingual Text. In **Proceedings of AIMSA'96,** Sozopol, Bulgaria. Pp. 3-11.

Hakkani, Dilek Zeynep; Tür, Gökhan; Oflazer, Kemal; Mitamura, Teruko and Nyberg, Eric H. 1998. An English to Turkish Interlingual MT System. In **Machine Translation and the Information Soup,** MT in a Growing Field of Language Technologies, Association for Machine Translation in the Americas (AMTA-98). Pp. 83-94.

Human language Technology Laboratory. **LEXITRON [online].** Retrieved August 13, 2004 from http://lexitron.nectec.or.th/

Janya In-ong. 2006. **Modern English Grammar in the 21st Century (Perfect Edition).** Bangkok: D.K. today. (In Thai)

Jurafsky, Daniel and Martin, James H. 2000. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.** Upper Saddle River, NJ: Prentice-Hall.

Kaewchai Chancharoen. 1999. Pattern-based Machine Translation for English-Thai. **KMITT Research and Development Journal**. 22 (September-December): 35-45.

Kaewchai Chancharoen; Nisanad Tannin and Booncharoen Sirinaovakul. 1998. Sentence-Based Machine Translation for English-Thai. In **The 1998 IEEE Asia-Pacific Conference on Circuits and Systems**, November 24-27, Chiangmai Plaza Hotel, ChiangMai, Thailand.

Kaewchai Chancharoen; Nisanad Tannin and Booncharoen Sirinaovakul. 1999. Pattern-Based Machine Translation for English-Thai. In **The 13th Pacific Asia Conference on Language, Information and Computation, Proceedings,** February 10-11, Taiwan, R.0.C.

Kanlaya Naruedomkul and Cercone, Nick. 2002. Generate and Repair Machine Translation. **Journal Computational Intelligence.** 18 (March): 254-269**.**

Kaplan, Ronald M.; Netter, Klaus; Wedekind, Jürgen and Zaenen, Annie. 1989. Translation by Structural Correspondences. In **Proceedings of the Fourth Conference on European Chapter of the Association for Computational Linguistics**, Manchester, England. Pp. 272-281.

Kim, Jung-Jae; Choi, Key-Sun and Chae, Young-Soog. 2000. Phrase-Pattern-Based Korean to English Machine Translation Using two Level Translation Pattern Selection**.** In **Proceedings of the 38th Annual Meeting on Association for Computational Linguistics Hong Kong**. Pp. 31–36.

Lee, Young-Suk; Yi, Wu Sok; Seneff, Stephanie and Weinstein, Clifford J. 2001. Interlingua-Based Broad-Coverage Korean-to-English Translation in CCLINC. In **Proceedings of the First International Conference on Human Language Technology Research**, San Diego. Pp.1-6.

Li, Tangqui; Nyberg, Eric H. and Carbonell, Jaime G. 1996. **Chinese Sentence Generation in a Knowledge-Based Machine Translation System**

(1996). Technical Report CMU-CMT-96-148. Retrieved March 13, 2007 from http://www.lti.cs.cmu.edu/Research/Kant/PDF/chinese-tr.pdf

Lonsdale, Deryle W.; Franz, Alexander M. and Leavitt, John R. R. 1994. Large-Scale Machine Translation: An Interlingua Approach. In **Proceedings of the 7th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems**, Austin, Texas, United States. Pp. 525-530.

Maosong, Sun. 2001. LFG for Chinese: Issues of Representation and Computation. In **Proceedings of the LFG01 Conference.** University of Hong Kong, Hong Kong.

Matsumoto, Yuji; Ishimoto, Hiroyuki and Utsuro, Takehito. 1993. Structural Matching of Parallel Texts. In **Proceedings of the 31st Annual Meeting on Association for Computational Linguistics**, Columbus, Ohio. Pp. 23–30.

Moore, Robert and Dowding, John. 1991. Efficient Bottom-Up Parsing. In **Proceedings of the Workshop on Speech and Natural Language**, Pacific Grove, California. Pp. 200-203.

Nagao, Makato. 1984. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In **Proceedings of the International NATO Symposium on Artificial and Human Intelligence**, Lyon, France. Pp. 173–180.

National Electronics and Computer Technology Center. n.d. **Thai Concept Classification (Technical Report).** Bangkok: Linguistics and Knowledge Science Laboratory (LINKS), NECTEC.

National Electronics and Computer Technology Center. n.d. **Interlingual Expression for Thai Language (Technical Report).** Bangkok: Linguistics and Knowledge Science Laboratory (LINKS), NECTEC.

Navavan Bandhumedha. 2006. **Thai Grammar.** 3rd ed. Bangkok: Chulalongkorn University Publishers. (In Thai)

Nittaya Kanchanawan. 1999. **Analysis of Thai Structure**. 2nd ed. Bangkok. Ramkhamhaeng University Publishers. (In Thai)

Paisarn Charoenpornsawat and Virach Sornlertlamvanich. 2001. Sentence Break Disambiguation for Thai. In **Proceedings of the 19th International Conference on Computer Processing of Oriental Languages,** Seoul, Korea.

Paisarn Charoenpornsawat; Boonserm Kijsirikul and Surapant Meknavin. 1998. Feature-Based Proper Name Identification in Thai. In **Proceeding of the National Computer Science and Engineering Conference'98 (NCSEC'98),** Bangkok, Thailand.

Paisarn Charoenpornsawat; Boonserm Kijsirikul and Surapant Meknavin. 1998. Feature-Based Thai Unknown Word Boundary Identification Using Winnow. In **Proceedings of the 1998 IEEE Asia-Pacific Conference on Circuits and Systems (APCCAS'98)**, Chieng Mai, Thailand.

Phanu Sungkhavon. 1984. **Semantic Relationships Between Nouns and Verbs in Thai Sentences.** Master's thesis, Chulalongkorn University. (In Thai)

Prachya Boonkwan and Asanee Kawtrakul. 2002. Plaesarn: Machine-Aided Translation Tool for English-to-Thai. In **COLING-02 on Machine Translation in Asia Volume 16**. Pp. 1-7.

Pradit Mittrapiyanuruk and Virach Sornlertlamvanich. 2000. The Automatic Thai Sentence Extraction. In **The Fourth Symposium on Natural Language Processing (SNLP2000),** Chiang Mai, Thailand. Pp. 23-28.

Radford, Andrew. 2004. **Minimalist Syntax: Exploring the Structure of English.** United Kingdom: Cambridge University Press.

Reyle, Uwe and Frey, Werner. 1983. A Prolog Implementation of Lexical Functional Grammar. In **Proceedings of the Eighth International Joint Conference on Artificial Intelligence-Volume 2**, Karlsruhe, West Germany. Pp. 693-695.

Riezler, Stefan; King, Tracy H.; Kaplan, Ronald M.; Crouch, Richard; Maxwell lll., John T. and Johnson, Mark. 2002. Parsing the Wall Street Journal Using a Lexical-Functional Grammar and Discriminative Estimation Techniques. In **Proceedings of the 40th Annual Meeting on the Association for Computational Linguistics (ACL),** Philadelphia, Pennsylvania July 2002. Pp. 271-278.

The Royal Institute Dictionary. 2003. **The Royal Institute Dictionary B.E. 2542**. Bangkok: Nanmeebooks. (In Thai)

Ruiz, Juan C. Anton. 2004. **GFU-LAB**. Retrieved August 13, 2004 from http://www3.uji.es/~ruiz/gfulab/

Seljan**,** Sanja**.** 2006**.** The Role of the Lexicon in Lexical-Functional Grammar – Example on Croatian. In **Proceedings of the 5th Slovenian and 1st International Conference, Language Technologies, IS-LTC 2006,** Ljubljana, Slovenia. Pp. 198-203.

Shirai, Satoshi; Bond, Francis and Takahashi, Yamato. 1997. A Hybrid Rule and Example Based Method for Machine Translation. In **The Natural Language Processing Pacific Rim Symposium (NLPRS-97)**, Phuket, Thailand. Pp. 49-54.

Sojka, Petr and Antoš, David. 2003. Context Sensitive Pattern Based Segmentation: A Thai Challenge. In **Proceedings of EACL 2003 Workshop Computational Linguistics for South Asian Languages - Expanding Synergies with Europe.** Budapest: Association for Computational Linguistics. Pp. 65-72.

Surapant Meknavin; Paisarn Charoenpornsawat and Boonserm Kijsirikul. 1997. Feature-Based Thai Word Segmentation. In **Proceedings of the Natural Language Processing Pacific Rim Symposium 1997 (NLPRS'97),** Phuket, Thailand.

SWI-Prolog 5.6 Program. n.d. Retrieved August 13, 2004 from http://www.swi-prolog.org/

Takeda, Koichi. 1996. Pattern-Based Context-Free Grammars for Machine Translation. In **Proceedings of the 34th Annual Meeting on Association for Computational Linguistics,** Santa Cruz, California. Pp. 144–151.

Takeda, Koichi. 1996. Pattern-Based Machine Translation. In **Proceedings of the 16th Conference on Computational Linguistics Volume 2,** Copenhagen, Denmark. Pp. 1155–1158.

Takenobu, Tokunaga; Makoto, Iwayama; Hozuyoi, Tanaka and Tadashi, Kamiwaki. 1988. LangLAB: A Natural Language Analysis System**.** In **Proceedings**

of the 12th Conference on Computational Linguistics Volume 2, Budapest, Hungry. Pp. 655–660.

Tambouratzis, G.; Sofianopoulos, S.; Spilioti, V.; Vassiliou, M.; Yannoutsou, O. and Markantonatou, S. 2006. Pattern Matching-Based System for Machine Translation (MT). In **Advances in Artificial Intelligence,** (Vol. 3955). G. Antoniou, G. Potamias, C. Spyropoulos and D. Plexousakis, eds. Germany: Springer Berlin / Heidelberg. Pp. 345-355.

Tanapong Potipiti; Virach Sornlertlamvanich and Thatsanee Charoenporn. 2000. Automatic Corpus-based Thai Word Extraction. In **The Fourth Symposium on Natural Language Processing (SNLP2000),** Chiang Mai, Thailand. Pp. 176-181.

Turhan, C. K. 1997. An English to Turkish Machine Translation System Using Structural Mapping. In **Proceedings of the Fifth Conference on Applied Natural Language Processing,** Washington, DC. Pp. 320-323.

Virach Sornlertlamvanish and Monthika Boriboon. n.d. **Thai Concept Classification (Technical Report).** Retrieved August 13, 2006 from http:// www.links.nectec.or.th/pub/tech/akotech.ps.g

Virach Sornlertlamvanish and Wantanee Phantachat. n.d. **Interlingual Expression for Thai Language (Technical Report).** Retrieved August 13, 2006 from http://www.links.nectec.or.th/pub/tech/iltr.ps.gz

Virach Sornlertlamvanish; Naoto Takahashi and Hitoshi Isahara. 2004. **Orchid Corpus.** Retrieved August 13, 2004 from http://www.hlt.nectec.or.th/orchid/corpus/corpus.html.

Virach Sornlertlamvanich; Paisarn Charoenpornsawat; Mothika Boriboon and Lalida Boonmana. 2000. ParSit: English-Thai Machine Translation Services on Internet. In **12th Annual Conference, ECIT and New Economy, National Electronics and Computer Technology Center,** Bangkok, June 2000. (In Thai).

Wirote Aroonmanakul. Collocation and Thai Word Segmentation. 2002. In **Proceedings of the Fifth Symposium on Natural Language Processing & The Fifth Oriental COCOSDA Workshop.** Pathumthani: Sirindhorn International Institute of Technology. Pp. 68-75.

Wittaya Nathong. 2003. **Contrastive Analysis of English and Thai**. 13th ed. Bangkok: Ramkhamhaeng University. (In Thai).

Xavier, Briffault; Karim, Chibout; Gérard, Sabah and Jérôme, Vapillon. 1997. A Linguistic Engineering Environment using LFG (Lexical Functional Grammar) and CG (Conceptual Graphs). In **Proceedings of the LFG97 Conference**, University of California, San Diego. Pp. 1-14.

Xuan, Qi**;** Huiping, ZHOU and Huowang**,** CHEN**.** 2002**.** An Interlingua-Based Chinese-English MT System. **Journal of Computer Science and Technology**. 17 (July): 464–472.

Yasukawa, Hideki. 1984. LFG System in Prolog. In **Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting on Association for Computational Linguistics.** Stanford, California. Pp. 358–361.

APPENDICES

APPENDIX A

Thai Part-of-Speech Tagset used in Orchid Corpus

No.	POS	Description	Example
1	NPRP	Proper noun	วินโดวส์ 95, โคโรน่า, โค้ก, พระอาทิตย์
2	NCNM	Cardinal number	หนึ่ง, สอง, สาม, 1, 2, 3
3	NONM	Ordinal number	ที่หนึ่ง, ที่สอง, ที่สาม, ที่1, ที่2, ที่3
4	NLBL	Label noun	1, 2, 3, 4, ก, ข, a, b
5	NCMN	Common noun	หนังสือ, อาหาร, อาคาร, คน
6	NTTL	Title noun	ดร., พลเอก
7	PPRS	Personal pronoun	คุณ, เขา, ฉัน
8	PDMN	Demonstrative pronoun	นี่, นั่น, ที่นั่น, ที่นี่
9	PNTR	Interrogative pronoun	ใคร, อะไร, อย่างไร
10	PREL	Relative pronoun	ที่, ซึ่ง, อัน, ผู้
11	VACT	Active verb	ทำงาน, ร้องเพลง, กิน
12	VSTA	Stative verb	เห็น, รู้, คือ
13	VATT	Attributive verb	อ้วน, ดี, สวย

No.	POS	Description	Example
14	XVBM	Pre-verb auxiliary, before negator “ไม่”	เกิด, เกือบ, กำลัง
15	XVAM	Pre-verb auxiliary, after negator “ไม่”	ค่อย, น่า, ได้
16	XVMM	Pre-verb, before or after negator “ไม่”	ควร, เคย, ต้อง
17	XVBB	Pre-verb auxiliary, in imperative mood	กรุณา, จง, เชิญ, อย่า, ห้าม
18	XVAE	Post-verb auxiliary	ไป, มา, ขึ้น
19	DDAN	Definite determiner, after noun without classifier in between	นี่, นั่น, โน่น, ทั้งหมด
20	DDAC	Definite determiner, allowing classifier in between	นี้, นั้น, โน้น, นู้น
21	DDBQ	Definite determiner, between noun and classifier or preceding quantitative expression	ทั้ง, อีก, เพียง
22	DDAQ	Definite determiner, following quantitative expression	พอดี, ถ้วน
23	DIAC	Indefinite determiner, following noun; allowing classifier in between	ไหน, อื่น, ต่างๆ
24	DIBQ	Indefinite determiner, between noun and classifier or preceding quantitative expression	บาง, ประมาณ, เกือบ

No.	POS	Description	Example
25	DIAQ	Indefinite determiner, following quantitative expression	กว่า, เศษ
26	DCNM	Determiner, cardinal number expression	หนึ่งคน, เสือ 2 ตัว
27	DONM	Determiner, ordinal number expression	ที่หนึ่ง, ที่สอง, ที่สุดท้าย
28	ADVN	Adverb with normal form	เก่ง, เร็ว, ช้า, สม่ำเสมอ
29	ADVI	Adverb with iterative form	เร็วๆ, เสมอๆ, ช้าๆ
30	ADVP	Adverb with prefixed form	โดยเร็ว
31	ADVS	Sentential adverb	โดยปกติ, ธรรมดา
32	CNIT	Unit classifier	ตัว, คน, เล่ม
33	CLTV	Collective classifier	คู่, กลุ่ม, ฝูง, เชิง, ทาง, ด้าน, แบบ, รุ่น
34	CMTR	Measurement classifier	กิโลกรัม, แก้ว, ชั่วโมง
35	CFQC	Frequency classifier	ครั้ง, เที่ยว
36	CVBL	Verbal classifier	ม้วน, มัด
37	JCRG	Coordinating conjunction	และ, หรือ, แต่
38	JCMP	Comparative conjunction	กว่า, เหมือนกับ, เท่ากับ
39	JSBR	Subordinating conjunction	เพราะว่า, เนื่องจาก, ที่, แม้ว่า, ถ้า
40	RPRE	Preposition	จาก, ละ, ของ, ใต้, บน

No.	POS	Description	Example
41	INT	Interjection	โอ้ย,โอ้, เออ, เอ๋, อ๋อ
42	FIXN	Nominal prefix	การทำงาน, ความสนุกสนาน
43	FIXV	Adverbial prefix	อย่างเร็ว
44	EAFF	Ending for affirmative sentencev	จ๊ะ, จ้ะ, ค่ะ, ครับ, นะ, น่า, เถอะ
45	EITT	Ending for interrogative sentence	หรือ, เหรอ, ไหม, มั้ย
46	NEG	Negator	ไม่, มิได้, ไม่ได้, มิ
47	PUNC	Punctuation	(,), ", „, ;

NOUN

A word which is used to refer to people, animal, place, thing or abstraction. The noun is subcategorized into 10 subcategories.

Proper noun (NPRP)

A noun which is the name of a particular person, place, or thing. It is easy to get confused in distinguishing proper noun and common noun in Thai language because there is no any distinguishable spelling such as using capital letter in English language.

NPRP can be identified as,

a) Name of product **วินโดวส์** 95 (Windows 95), **โคโรน่า** (Corona), **โค้ก** (Coke)

b) Abbreviation name **จส.**100, **เน็คเท็ค** (NECTEC)

c) Name of person, group of person, company, etc.

d) Geographical name, such as name of region, continent, country, province, etc.

e) Astronomical name พระอาทิตย์ (the sun), ทางช้างเผือก (Milky way)

f) Chemical name โปรตีน (protein), ออกซิเจน (oxygen)

g) Scientific name

h) Name of artificial place

i) Name of language, race, religion, etc.

Cardinal number noun (NCNM)

A noun which is used to express number but is not the one used in a quantitative expression. It can be written with a figure or with a word. For example, หนึ่ง (one), สอง (two), สาม (three), 1, 2, 3.

Ordinal number noun (NONM)

A noun which is used to put thing in a numerical order. It can be a figure or a word. For example, ที่หนึ่ง (first), ที่สอง (second), ที่สาม (third), ที่1 (1st), ที่2 (2nd), ที่3 (3rd).

Label noun (NLBL)

A noun which is used to label an item. For example, 1, 2, 3, 4, ก, ข, a, b.

Common noun (NCMN)

A noun which is used to express the name of a person, place or thing but in particular. For example, หนังสือ (book), อาหาร (food), อาคาร (building), คน (person).

Title noun (NTTL)

A noun which is a descriptive or distinctive appellation. It is used to show a person's rank, occupation, etc. For example, ดร. (Dr.), พลเอก (General).

PRONOUN

A word which may replace a noun or noun phrase. The pronoun is subcategorized into 6 subcategories.

Personal pronoun (PPRS)

A pronoun which represents a person. It is used to refer to oneself, the people one is talking to, or the people or things one is talking about. For example, คุณ (you), เขา (he), ฉัน (I).

Demonstrative pronoun (PDMN)

A pronoun which refers to something being either near to or distant from the speaker. For example, นี่ (this), นั่น (that), ที่นั่น (there), ที่นี่ (here).

Interrogative pronoun (PNTR)

A pronoun which is used to form a question. It refers to the information that someone is asking about. For example, ใคร (who), อะไร (what), อย่างไร (how).

Relative pronoun (PREL)

Relative pronoun (PREL):- A pronoun which introduces a relative clause. For example, ที่, ซึ่ง, อัน(which, that), ผู้ (who).

VERB

A word which occurs as part of the predicate of a sentence, or which refers to an action, state or attribute. The verb is subcategorized into 3 subcategories.

Active verb (VACT)

A verb which indicates an action of an entity. For example, ทำงาน (work), ร้องเพลง (sing), กิน (eat).

Stative verb (VSTA)

A verb which refers to a state. The stative verb is not used in progressive aspect. For example, เห็น (see), รู้ (know), คือ (be).

Attributive verb (VATT)

A verb which refers to an attribute. For example, อ้วน (fat), ดี (good), สวย (beautiful).

AUXILIARY

A word which is used with a verb of a sentence indicating a grammatical function such as aspect, voice, mood or tense. An auxiliary has a rather fixed position relating to the verb of the sentence. The auxiliary is subcategorized into 5 subcategories.

Pre-verb auxiliary, before negator (XVBM)

An auxiliary which is used preceding a verb and in the position preceding the negator "ไม่" (not) in a negative sentence. For example, ฝนเกิดไม่ตก, การบ้านเกือบเสร็จแล้ว, คนไข้กำลังหลับ.

Pre-verb auxiliary, after negator (XVAM)

An auxiliary which is used preceding a verb and in the position following the negator "ไม่" (not) in a negative sentence. For example, เขาไม่ค่อยมาที่นี่, บ้านนี้น่าอยู่, เราได้เห็นฝีมือเขาแล้ว.

Pre-verb auxiliary, before or after negator (XVMM)

An auxiliary which is used preceding a verb and in the position either preceding or following the negator "ไม่" (not) in a negative sentence. For example, เธอ (ไม่) ควรไปพบเขา or เราควร (ไม่) พูดเลยวันนี้, เล็ก (ไม่) เคยเอาใจใส่เรา or เล็กเคย (ไม่) เอาใจใส่เรา, เรา (ไม่) ต้องบอกเขาก่อน or เราต้อง (ไม่) บอกเขาก่อน.

Pre-verb auxiliary in imperative mood (XVBB)

An auxiliary which is used preceding a verb, usually in the initial position of an imperative sentence. It is used to show the grammatical function of imperative mood. For example, กรุณา, จง, เชิญ (please), อย่า, ห้าม (don't).

Post-verb auxiliary (XVAE)

An auxiliary which is used following a verb. For example, แก้วแตกไปหลายใบ or เด็กกินไปเล่นไป, ฉันเห็นมากับตา or เขาพักที่นี่มาหลายวันแล้ว, ยกมือขึ้น.

DETERMINER

A word which is used with a noun restricting the meaning of the noun in some way. The determiner is subcategorized into 9 subcategories.

Definite determiner: after a noun without a classifier in between (DDAN)

A definite determiner which is used immediately after a noun and no classifier allowed in between. For example, นี่ (this), นั่น, โน่น (that), ทั้งหมด (all).

Definite determiner: allowing a classifier in between (DDAC)

A definite determiner which follows a noun or a classifier. For example, นี้ (this), นั้น (that), โน้น, นู้น (over there).

Definite determiner: between a noun and a classifier, or preceding a quantitative expression (DDBQ)

A definite determiner which is used between a noun and a classifier or, preceding a quantitative expression. For example, ทั้ง (all), อีก (the other), เพียง (only).

Definite determiner: following a quantitative expression (DDAQ)

A definite determiner which is used after a quantitative expression. For example, พอดี (quite right), ถ้วน (exactly).

Indefinite determiner: following a noun and allowing a classifier in between (DIAC)

An indefinite determiner which is used after a noun with or without a classifier in between. For example, ไหน (which), อื่น (other), ต่างๆ (several).

Indefinite determiner: between a noun and a classifier, or preceding a quantitative expression (DIBQ)

An indefinite determiner which is used between a noun and a classifier or, preceding a quantitative expression. For example, บาง (some), ประมาณ (about), เกือบ (almost).

Indefinite determiner: following a quantitative expression (DIAQ)

An indefinite determiner which is used following a quantitative expression. For example, กว่า (over), เศษ (more).

Determiner: cardinal number expression (DCNM)

A determiner which is a cardinal number used in quantitative expression. It can be a figure or a word. For example, หนึ่ง/DCNM คน/CNIT, เสือ/NCMN 2/DCNM ตัว/CNIT.

Determiner: ordinal number expression (DONM)

A determiner which is an ordinal number used to express the order. For example, ที่หนึ่ง (first), ที่สอง (second), ที่สุดท้าย (last).

ADVERB

A word which describes or modifies to the meaning of a verb, another adverb or a sentence. The adverb is subcategorized into 4 subcategories.

Adverb with a normal form (ADVN)

An adverb which is used in the base form and is not in the form of repetition or is not derived from a verb by adding an adverbial prefix "**โดย**" or "**อย่าง**". For example, **เก่ง** (smartly), **เร็ว** (quickly), **ช้า** (slowly), **สม่ำเสมอ** (always).

Adverb with an iterative form (ADVI)

An adverb which is used in an iterative form by combining with "**ๆ**". For example, **เร็วๆ** (quickly), **เสมอๆ** (always), **ช้าๆ** (slowly).

Adverb with an prefixed form (ADVP)

An adverb which is derived from a verb by adding a prefix such as "**โดย**" or "**อย่าง**". For example, **โดยเร็ว** (quickly).

Sentential adverb (ADVS)

An adverb which shows the speaker's attitude to or evaluation of what is said in the rest of the sentence. It is usually placed at the beginning of the sentence, but some are also used in other position. For example, **โดยปกติ** (generally), **ธรรมดา** (normally).

CLASSIFIER

A word which is used with a noun showing the sub-class to which a noun belongs. It is originally used to express the unit of a noun in counting. The classifier is subcategorized into 5 subcategories.

Unit classifier (CNIT)

A classifier which represents an individual thing, animal or person. For example, **ตัว** (a classifier for animals or furniture), **คน** (a classifier for human beings), **เล่ม** (a classifier for books).

Collective classifier (CLTV)

A classifier which is used to represent a set, group, class or type of thing, animal or person. For example, **คู่** (pair), **กลุ่ม** (group), **ฝูง** (flock), **เชิง**, **ทาง**, **ด้าน**, **แบบ**, **รุ่น** (type).

Measurement classifier (CMTR)

A classifier which is used to represent the measuring unit of a noun. For example, กิโลกรัม (kilogram), แก้ว (glass), ชั่วโมง (hour).

Frequency classifier (CFQC)

A classifier which is used to represent the frequency of a noun. For example, ครั้ง, เที่ยว (time).

Verbal classifier (CVBL)

A classifier which derives from a verb to represent the unit of a noun. For example, ม้วน (a classifier for a thing resulted from rolling), มัด (a classifier for a thing resulted from tying).

CONJUNCTION

A word which joins words, phrases, or clauses together. The conjunction is subcategorized into 3 subcategories.

Coordinating conjunction (JCRG)

A conjunction which joins linguistic units which are equivalent or of the same rank. For example, และ (and), หรือ (or), แต่ (but).

Comparative conjunction (JCMP)

A conjunction which joins a standard to which someone or something is referred and an independent clause. For example, กว่า (than), เหมือนกับ, เท่ากับ (equally).

Subordinate conjunction (JSBR)

A conjunction which joins an independent clause and a dependent clause (a clause which must be used with other clause to form a complete grammatical structure). It is usually used to indicate time, condition, purpose, reason, result, concession, or place of an event. For example, เพราะว่า, เนื่องจาก (because), ที่ (that), แม้ว่า (even if), ถ้า (if).

PREPOSITION

A word which is used preceding a noun or a pronoun to link it grammatically to other word or phrase.

Preposition (RPRE)

A word which is used preceding a noun, a pronoun, or other substantive to form a phrase functioning as a modifier of a verb, or a noun and that typically expresses a spatial, temporal and so on. For example, จาก (from), ละ (each), ของ (of), ใต้ (under), บน (on).

INTERJECTION

Interjection (INT)

A word which indicates an emotional state or attitude such as delight, surprise, shock, and disgust, but has no referential meaning. For example, โอ้ย,โอ้, เออ, เอ่, อ่อ.

PREFIX

A word which is added to the beginning of a word changing the meaning or function of the word. The prefix is subcategorized into 2 subcategories.

Nominal prefix (FIXN)

A word which is added to the beginning of a verb (or a verb phrase) changing the function of the verb (or the verb phrase) to be a noun (or a noun phrase). There are 2 words in this category, "การ" and "ความ". For example, การ/FIXN ทำงาน/VACT (working), ความ/FIXN สนุกสนาน/VSTA (amusement).

Adverbial prefix (FIXV)

A word which is added to the beginning of a verb (or a verb phrase) changing the function of the verb (or the verb phrase) to be an adverb (or an adverb phrase). There is 1 word in this category, "อย่าง". For example, อย่าง/FIXV เร็ว/VATT (quickly).

SENTENCE PARTICLE

Ending for affirmative sentence (EAFF)

A word which is added to the end of a sentence to indicate the mood of an utterance, such as affirmation, interrogation, imperative, persuasion, threatening and so on. It also gives the information about seniority, familiarity, and social status which is the relation between the speaker and the listener. For example, จ๊ะ, จ้ะ, ค่ะ, ครับ, นะ, น่า, เถอะ.

Ending for interrogative sentence particle (EITT)

A word which is added to the end of a sentence to indicate the mood of interrogative, and yes-no question. It can be used either with or without an interrogative pronoun. For example, หรือ, เหรอ, ไหม, มั้ย (yes or no).

NEGATOR

Negator (NEG)

A word which is used to negate a sentence. For example, ไม่, มิได้, ไม่ได้, มิ (not).

PUNCTUATION

Punctuation (PUNC)

A mark or a sign (such as an opening/closing bracket, opening/closing quote, comma, colon, semicolon, dash, exclamation mark, period, and question marker). For example, (,), ", ,, ;.

APPENDIX B

EXAMPLES OF TESTING PHRASES/SENTENCES

%TTitle: การประชุมทางวิชาการ ครั้งที่ 1

%ETitle: [1st Annual Conference]

%TAuthor:

%EAuthor:

%TInbook: การประชุมทางวิชาการ ครั้งที่ 1, โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์, ปีงบประมาณ 2531, เล่ม 1

%EInbook: The 1st Annual Conference, Electronics and Computer Research and Development Project, Fiscal Year 1988, Book 1

%TPublisher: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, กระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน

%EPublisher: National Electronics and Computer Technology Center, Ministry of Science, Technology and Energy

%Page:

%Year: 1989

%File:

#P1

#1

การประชุมทางวิชาการ ครั้งที่ 1//

การ/FIXN

ประชุม/VACT

ทาง/NCMN

วิชาการ/NCMN

<space>/PUNC

ครั้ง/CFQC

ที่ 1/DONM

//

#2

โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์//

โครงการวิจัยและพัฒนา/NCMN

อิเล็กทรอนิกส์/NCMN

และ/JCRG

คอมพิวเตอร์/NCMN

//

#3

ปีงบประมาณ 2531//

ปีงบประมาณ/NCMN

<space>/PUNC

2531/NCNM

//

#4

เล่ม 1//

เล่ม/CNIT

<space>/PUNC

1/DONM

//

#P2

#1

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ//

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ/NPRP

//

#2

กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน//

กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน/NPRP

//

#P3

#1

วันที่ 15-16 สิงหาคม 2532//

วัน/NCMN

ที่ 15/DONM

<minus>/PUNC

16/DONM

<space>/PUNC

สิงหาคม/NCMN

<space>/PUNC

2532/NCNM

//

#2

ห้องประชุม ชั้น 4//

ห้องประชุม/NCMN

<space>/PUNC

ชั้น/CNIT

<space>/PUNC

4/NCNM

//

%TTitle: สาร ฯพณฯ รัฐมนตรีว่าการกระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน

%ETitle: [Message of H.E. the Minister of Ministry of Science, Technology and Energy]

%TAuthor: นายประจวบ ไชยสาส์น

%EAuthor: [Mr. Prajuab Chaiyasarn]

%TInbook: การประชุมทางวิชาการ ครั้งที่ 1, โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์, ปีงบประมาณ 2531, เล่ม 1

%EInbook: The 1st Annual Conference, Electronics and Computer Research and Development Project, Fiscal Year 1988, Book 1

%TPublisher: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, กระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน

%EPublisher: National Electronics and Computer Technology Center, Ministry of Science, Technology and Energy

%Page:

%Year: 1989

%File:

#P1

#1

สาร//

สาร/NCMN

//

#P2

#1

ฯพณฯ รัฐมนตรีว่าการกระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน//

ฯพณฯ/NTTL

<space>/PUNC

รัฐมนตรีว่าการ/NCMN

กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน/NPRP

//

#P3

#1

ประเทศไทยได้มีการปรับเปลี่ยนโครงสร้างในการพัฒนาเศรษฐกิจของประเทศ \\

จากประเทศเกษตรกรรมไปสู่ความเป็นประเทศอุตสาหกรรมมากยิ่งขึ้น//

ประเทศไทย/NPRP

ได้/XVAM

มี/VSTA

การ/FIXN

ปรับเปลี่ยน/VACT

โครงสร้าง/NCMN

ใน/RPRE

การ/FIXN

พัฒนา/VACT

เศรษฐกิจ/NCMN

ของ/RPRE

ประเทศ/NCMN

<space>/PUNC

จาก/RPRE

ประเทศ/NCMN

เกษตรกรรม/NCMN

ไปสู่/RPRE

ความ/FIXN

เป็น/VSTA

ประเทศอุตสาหกรรม/NCMN

มาก/ADVN

ยิ่งขึ้น/ADVN

//

#2

ในการดำเนินการเพื่อให้บรรลุวัตถุประสงค์ดังกล่าว \\

จะต้องอาศัยปัจจัยพื้นฐานหลายประการในการเป็นตัวเร่งและเป็นฐาน เช่น การพัฒนาเทคโนโลยีที่

ใช้ใน\\

การผลิตของภาคอุตสาหกรรม//

ใน/RPRE

การ/FIXN

ดำเนินการ/VACT

เพื่อให้/JSBR

บรรลุ/VSTA

วัตถุประสงค์/NCMN

ดังกล่าว/DDAC

<space>/PUNC

จะ/XVBM

ต้อง/XVMM

อาศัย/VSTA

ปัจจัยพื้นฐาน/NCMN

หลาย/DIBQ

ประการ/CNIT

ใน/RPRE

การ/FIXN

เป็น/VSTA

ตัวเร่ง/NCMN

และ/JCRG

เป็น/VSTA

ฐาน/NCMN

<space>/PUNC

เช่น/RPRE

<space>/PUNC

การ/FIXN

พัฒนา/VACT

เทคโนโลยี/NCMN

ที่/PREL

ใช้/VACT

ใน/RPRE

การ/FIXN

ผลิต/VACT

ของ/RPRE

ภาคอุตสาหกรรม/NCMN

//

#3

กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน จึงได้ให้ความสำคัญเป็นลำดับสูงในการ

พัฒนา\\

อุตสาหกรรมอิเล็กทรอนิกส์และคอมพิวเตอร์//

กระทรวงวิทยาศาสตร์ เทคโนโลยีและการพลังงาน/NPRP

<space>/PUNC

จึง/XVBM

ได้/XVAM

ให้ความสำคัญ/VACT

เป็น/VSTA

ลำดับ/NCMN

สูง/VATT

ใน/RPRE

การ/FIXN

พัฒนา/VACT

อุตสาหกรรม/NCMN

อิเล็กทรอนิกส์/NCMN

และ/JCRG

คอมพิวเตอร์/NCMN

//

#4

ซึ่งอุตสาหกรรมนี้จะมีบทบาทที่สำคัญมากในภาคอุตสาหกรรม//

ซึ่ง/JSBR

อุตสาหกรรม/NCMN

นี้/DDAC

จะ/XVBM

มี/VSTA

บทบาท/NCMN

ที่/PREL

สำคัญ/VATT

มาก/ADVN

ใน/RPRE

ภาคอุตสาหกรรม/NCMN

//

#5

โดยเป็นปัจจัยพื้นฐาน หรือส่วนประกอบที่สำคัญของการผลิตผลิตภัณฑ์อุตสาหกรรมแทบทุก

สาขา//

โดย/JSBR

เป็น/VSTA

ปัจจัยพื้นฐาน/NCMN

<space>/PUNC

หรือ/JCRG

ส่วนประกอบ/NCMN

ที่/PREL

สำคัญ/VATT

ของ/RPRE

การ/FIXN

ผลิต/VACT

ผลิตภัณฑ์อุตสาหกรรม/NCMN

แทบ/DIBQ

ทุก/DIBQ

สาขา/CNIT

//

APPENDIX C

EXAMPLE OF F-STRUCTURE AND C-STRUCTURE

#1001 ระบบ MU พัฒนา ด้วย ภาษา LISP.

SYNTACTIC REPRESENTATION:

```
+-
¦ subj +-
¦     ¦ modi +-
¦     ¦     ¦ wordtype=NOUN
¦     ¦     ¦ head=ระบบ
¦     ¦     +-
¦     ¦ wordtype=NOUN
¦     ¦ head=MU
¦     +-
¦ obl_instr +-
¦         ¦ obj1 +-
¦         ¦     ¦ wordtype=NOUN
¦         ¦     ¦ head=ภาษา
¦         ¦     +-
¦         ¦ obj2 +-
¦         ¦     ¦ wordtype=NOUN
¦         ¦     ¦ head=LISP
¦         ¦     +-
¦         ¦ wordtype=PREPOSITION
¦         ¦ head=ด้วย
¦         +-
¦ wordtype=VERB
¦ head=พัฒนา
```

+-

SEMANTIC REPRESENTATION:

```
+-
¦ head=พัฒนา
¦ pred=develop
¦ type=VACT
¦ ontology=verb
¦ subject +-
¦       ¦ case=objective
¦       ¦ head=MU
¦       ¦ pred=MU
¦       ¦ type=NPRP
¦       ¦ ontology=tool
¦       ¦ modifier +-
¦       ¦       ¦ head=ระบบ
¦       ¦       ¦ pred=system
¦       ¦       ¦ type=NCMN
¦       ¦       ¦ ontology=equipment
¦       ¦       +-
¦       +-
¦ func_instrument +-
¦               ¦ head=ด้วย
¦               ¦ pred=with
¦               ¦ type=PP
¦               ¦ ontology=prep
¦               ¦ object1 +-
¦               ¦       ¦ case=result
¦               ¦       ¦ head=ภาษา
¦               ¦       ¦ pred=language
¦               ¦       ¦ type=NCMN
¦               ¦       ¦ ontology=language
```

```
¦             ¦       +-
¦             ¦ object2 +-
¦             ¦       ¦ case=result
¦             ¦       ¦ head=LISP
¦             ¦       ¦ pred=LISP
¦             ¦       ¦ type=NPRP
¦             ¦       ¦ ontology=software
¦             ¦       +-
¦             +-
¦
+-
```

#1007 โนด จะ แทน **CP** และ **Case.**

SYNTACTIC REPRESENTATION:

```
+-
¦ subj +-
¦     ¦ wordtype=NOUN
¦     ¦ head=โนด
¦     +-
¦ adjunct +-
¦       ¦ wordtype=AUXILIARY
¦       ¦ head=จะ
¦       +-
¦ obj +-
¦     ¦ coor1 +-
¦     ¦       ¦ wordtype=NOUN
¦     ¦       ¦ head=CP
¦     ¦       +-
¦     ¦ coor2 +-
¦     ¦       ¦ wordtype=NOUN
```

```
¦   ¦     ¦ head=Case
¦   ¦     +-
¦   ¦ wordtype=CONJUNCTION
¦   ¦ head=และ
¦   +-
¦ wordtype=VERB
¦ head=แทน
+-
```

SEMANTIC REPRESENTATION:

```
+-
¦ head=แทน
¦ pred=replace
¦ type=VSTA
¦ ontology=verb
¦ subject +-
¦       ¦ case=experiencer
¦       ¦ head=โนด
¦       ¦ pred=node
¦       ¦ type=NCMN
¦       ¦ ontology=object
¦       +-
¦ adjunction +-
¦         ¦ head=จะ
¦         ¦ pred=will
¦         ¦ type=XVBM
¦         ¦ ontology=auxi
¦         +-
¦ object +-
¦       ¦ case=complement
¦       ¦ head=และ
¦       ¦ pred=and
```

```
¦       ¦ type=JCRG
¦       ¦ ontology=coor_join
¦       ¦ coordination1 +-
¦       ¦               ¦ head=CP
¦       ¦               ¦ pred=CP
¦       ¦               ¦ type=NCMN
¦       ¦               ¦ ontology=symbol
¦       ¦               +-
¦       ¦ coordination2 +-
¦       ¦               ¦ head=Case
¦       ¦               ¦ pred=Case
¦       ¦               ¦ type=NCMN
¦       ¦               ¦ ontology=symbol
¦       ¦               +-
¦       +-
+-
```

#1009 3 กำหนดค่า TCAT ให้กับ CP แต่ละ ตัว.

SYNTACTIC REPRESENTATION:

```
+-
¦ subj +-
¦      ¦ wordtype=NOUN
¦      ¦ head=3
¦      +-
¦ obj +-
¦     ¦ wordtype=NOUN
¦     ¦ head=TCAT
¦     +-
¦ obl_benefic +-
¦             ¦ obj +-
¦             ¦     ¦ modi1 +-
```

```
¦           ¦   ¦     ¦ modi +-
¦           ¦   ¦     ¦     ¦ wordtype=DETERMINER
¦           ¦   ¦     ¦     ¦ head=แต่ละ
¦           ¦   ¦     ¦     +-
¦           ¦   ¦     ¦ wordtype=CLASSIFIER
¦           ¦   ¦     ¦ head=ตัว
¦           ¦   ¦     +-
¦           ¦   ¦ wordtype=NOUN
¦           ¦   ¦ head=CP
¦           ¦   +-
¦           ¦ wordtype=PREPOSITION
¦           ¦ head=ให้กับ
¦           +-
¦ wordtype=VERB
¦ head=กำหนดค่า
+-
SEMANTIC REPRESENTATION:
+-
¦ head=กำหนดค่า
¦ pred=determine
¦ type=VACT
¦ ontology=verb
¦ subject +-
¦       ¦ case=agentive
¦       ¦ head=3
¦       ¦ pred=3
¦       ¦ type=NLBL
¦       ¦ ontology=ordinal
¦       +-
¦ object +-
¦      ¦ case=result
```

```
¦       ¦ head=TCAT
¦       ¦ pred=TCAT
¦       ¦ type=NCMN
¦       ¦ ontology=object
¦       +-
¦ func_beneficiary +-
¦               ¦ head=ให้กับ
¦               ¦ pred=for
¦               ¦ type=PP
¦               ¦ ontology=prep
¦               ¦ object +-
¦               ¦       ¦ head=CP
¦               ¦       ¦ pred=CP
¦               ¦       ¦ type=NCMN
¦               ¦       ¦ ontology=symbol
¦               ¦       ¦ modifier1 +-
¦               ¦       ¦       ¦ head=ตัว
¦               ¦       ¦       ¦ pred=unit
¦               ¦       ¦       ¦ type=CNIT
¦               ¦       ¦       ¦ num=sg
¦               ¦       ¦       ¦ ontology=equipment
¦               ¦       ¦       ¦ modifier +-
¦               ¦       ¦       ¦       ¦ head=แต่ละ
¦               ¦       ¦       ¦       ¦ pred=each
¦               ¦       ¦       ¦       ¦ type=DIBQ
¦               ¦       ¦       ¦       ¦ ontology=compare
¦               ¦       ¦       ¦       +-
¦               ¦       ¦       +-
¦               ¦       +-
¦               +-
+-
```

#10010 6 กำหนดค่า รูปแบบ กริยา ให้กับ **predicate CP.**

SYNTACTIC REPRESENTATION:

```
+-
¦ subj +-
¦     ¦ wordtype=NOUN
¦     ¦ head=6
¦     +-
¦ obj1 +-
¦     ¦ wordtype=NOUN
¦     ¦ head=รูปแบบ
¦     +-
¦ obj2 +-
¦     ¦ wordtype=NOUN
¦     ¦ head=กริยา
¦     +-
¦ obl_benefic +-
¦           ¦ obj +-
¦           ¦     ¦ wordtype=NOUN
¦           ¦     ¦ head=predicate_CP
¦           ¦     +-
¦           ¦ wordtype=PREPOSITION
¦           ¦ head=ให้กับ
¦           +-
¦ wordtype=VERB
¦ head=กำหนดค่า
+-
```

SEMANTIC REPRESENTATION:

```
+-
¦ head=กำหนดค่า
```

¦ pred=determine
¦ type=VACT
¦ ontology=verb
¦ subject +-
¦ ¦ case=agentive
¦ ¦ head=6
¦ ¦ pred=6
¦ ¦ type=NLBL
¦ ¦ ontology=ordinal
¦ +-
¦ object1 +-
¦ ¦ case=result
¦ ¦ head=รูปแบบ
¦ ¦ pred=pattern
¦ ¦ type=NCMN
¦ ¦ ontology=sight
¦ +-
¦ object2 +-
¦ ¦ case=result
¦ ¦ head=กริยา
¦ ¦ pred=verb
¦ ¦ type=NCMN
¦ ¦ ontology=word
¦ +-
¦ func_beneficiary +-
¦ ¦ head=ให้กับ
¦ ¦ pred=for
¦ ¦ type=PP
¦ ¦ ontology=prep
¦ ¦ object +-
¦ ¦ ¦ head=predicate_CP

```
¦           ¦     ¦ pred=predicate_CP
¦           ¦     ¦ type=NCMN
¦           ¦     ¦ ontology=word
¦           ¦     +-
¦           +-
¦
+-
```

APPENDIX D

EXAMPLES OF NESTED LIST OF THAI SENTENCES

#1001 ระบบ **MU** พัฒนา ด้วย ภาษา **LISP.**

[[subj=[modi=[wordtype=NOUN, head=ระบบ, sem=[head=ระบบ, pred=system, type=NCMN, ontology=equipment]], wordtype=NOUN, head=MU, sem=[head=MU, pred=MU, type=NPRP, ontology=tool, modifier=[head=ระบบ, pred=system, type=NCMN, ontology=equipment], case=objective]], obl_instr=[sem=[head=ด้วย, pred=with, type=PP, ontology=prep, object1=[case=result, head=ภาษา, pred=language, type=NCMN, ontology=language], object2=[case=result, head=LISP, pred=LISP, type=NPRP, ontology=software]], obj1=[sem=[case=result, head=ภาษา, pred=language, type=NCMN, ontology=language], wordtype=NOUN, head=ภาษา], obj2=[sem=[case=result, head=LISP, pred=LISP, type=NPRP, ontology=software], wordtype=NOUN, head=LISP], wordtype=PREPOSITION, head=ด้วย], wordtype=VERB, head=พัฒนา, sem=[head=พัฒนา, pred=develop, type=VACT, ontology=verb, subject=[case=objective, head=MU, pred=MU, type=NPRP, ontology=tool, modifier=[head=ระบบ, pred=system, type=NCMN, ontology=equipment]], func_instrument=[head=ด้วย, pred=with, type=PP, ontology=prep, object1=[case=result, head=ภาษา, pred=language, type=NCMN, ontology=language], object2=[case=result, head=LISP, pred=LISP, type=NPRP, ontology=software]]]]]

#1007 โนด จะ แทน CP และ Case.

[[subj=[wordtype=NOUN, head=โนด, sem=[head=โนด, pred=node, type=NCMN, ontology=object, case=experiencer]], adjunct=[wordtype=AUXILIARY, head=จะ, sem=[head=จะ, pred=will, type=XVBM, ontology=auxi]], obj=[sem=[case=complement, head=และ, pred=and, type=JCRG, ontology=coor_join, coordination1=[head=CP, pred=CP, type=NCMN, ontology=symbol], coordination2=[head=Case, pred=Case, type=NCMN, ontology=symbol]], coor1=[wordtype=NOUN, head=CP, sem=[head=CP, pred=CP, type=NCMN, ontology=symbol]], coor2=[sem=[head=Case, pred=Case, type=NCMN, ontology=symbol], wordtype=NOUN, head=Case], wordtype=CONJUNCTION, head=และ], wordtype=VERB, head=แทน, sem=[head=แทน, pred=replace, type=VSTA, ontology=verb, subject=[case=experiencer, head=โนด, pred=node, type=NCMN, ontology=object], adjunction=[head=จะ, pred=will, type=XVBM, ontology=auxi], object=[case=complement, head=และ, pred=and, type=JCRG, ontology=coor_join, coordination1=[head=CP, pred=CP, type=NCMN, ontology=symbol], coordination2=[head=Case, pred=Case, type=NCMN, ontology=symbol]]]]]

#1009 3 กำหนดค่า TCAT ให้กับ CP แต่ละ ตัว.

[[subj=[wordtype=NOUN, head=3, sem=[head=3, pred=3, type=NLBL, ontology=ordinal, case=agentive]], obj=[sem=[case=result, head=TCAT, pred=TCAT, type=NCMN, ontology=object], wordtype=NOUN, head=TCAT], obl_benefic=[sem=[head=ให้กับ, pred=for, type=PP, ontology=prep, object=[head=CP, pred=CP, type=NCMN, ontology=symbol, modifier1=[head=ตัว, pred=unit, type=CNIT, num=sg, ontology=equipment, modifier=[head=แต่ละ, pred=each, type=DIBQ, ontology=compare]]]], obj=[sem=[head=CP, pred=CP, type=NCMN, ontology=symbol, modifier1=[head=ตัว, pred=unit, type=CNIT, num=sg, ontology=equipment, modifier=[head=แต่ละ, pred=each, type=DIBQ, ontology=compare]]], modi1=[sem=[head=ตัว, pred=unit, type=CNIT, num=sg,

ontology=equipment, modifier=[head=แต่ละ, pred=each, type=DIBQ, ontology=compare]], modi=[wordtype=DETERMINER, head=แต่ละ, sem=[head=แต่ละ, pred=each, type=DIBQ, ontology=compare]], wordtype=CLASSIFIER, head=ตัว], wordtype=NOUN, head=CP], wordtype=PREPOSITION, head=ให้กับ], wordtype=VERB, head=กำหนดค่า, sem=[head=กำหนดค่า, pred=determine, type=VACT, ontology=verb, subject=[case=agentive, head=3, pred=3, type=NLBL, ontology=ordinal], object=[case=result, head=TCAT, pred=TCAT, type=NCMN, ontology=object], func_beneficiary=[head=ให้กับ, pred=for, type=PP, ontology=prep, object=[head=CP, pred=CP, type=NCMN, ontology=symbol, modifier1=[head=ตัว, pred=unit, type=CNIT, num=sg, ontology=equipment, modifier=[head=แต่ละ, pred=each, type=DIBQ, ontology=compare]]]]]]]

#10010 6 กำหนดค่า รูปแบบ กริยา ให้กับ predicate CP.

[[subj=[wordtype=NOUN, head=6, sem=[head=6, pred=6, type=NLBL, ontology=ordinal, case=agentive]], obj1=[sem=[case=result, head=รูปแบบ, pred=pattern, type=NCMN, ontology=sight], wordtype=NOUN, head=รูปแบบ], obj2=[sem=[case=result, head=กริยา, pred=verb, type=NCMN, ontology=word], wordtype=NOUN, head=กริยา], obl_benefic=[sem=[head=ให้กับ, pred=for, type=PP, ontology=prep, object=[head=predicate_CP, pred=predicate_CP, type=NCMN, ontology=word]], obj=[sem=[head=predicate_CP, pred=predicate_CP, type=NCMN, ontology=word], wordtype=NOUN, head=predicate_CP], wordtype=PREPOSITION, head=ให้กับ], wordtype=VERB, head=กำหนดค่า, sem=[head=กำหนดค่า, pred=determine, type=VACT, ontology=verb, subject=[case=agentive, head=6, pred=6, type=NLBL, ontology=ordinal], object1=[case=result, head=รูปแบบ, pred=pattern, type=NCMN, ontology=sight], object2=[case=result, head=กริยา, pred=verb, type=NCMN, ontology=word], func_beneficiary=[head=ให้กับ, pred=for, type=PP, ontology=prep, object=[head=predicate_CP, pred=predicate_CP, type=NCMN, ontology=word]]]]]

APPENDIX E

EXAMPLES OF THAI LEXICON

กัน = ADVERB

U/wordtype = ADVERB

U/ head = กัน

U/sem/head = กัน

U/sem/ pred = together

U/ sem/type = ADVN

U/ sem/ontology = join_separate.

ทั้งหมด = ADVERB

U/wordtype = ADVERB

U/head = ทั้งหมด

U/sem/head = ทั้งหมด

U/sem/pred = all

U/sem/type = ADVN

U/sem/ontology = quantity.

kbyte = CLASSIFIER

U/wordtype = CLASSIFIER

U/head = kbyte

U/sem/head = kbyte

U/sem/pred = kbyte

U/sem/type = CMTR

U/sem/ontology = measure

ชุด = CLASSIFIER

U/wordtype = CLASSIFIER

U/head = ชุด

U/sem/head = ชุด

U/sem/pred = set

U/sem/type = CLTV

U/sem/num = pl

U/sem/ontology = set.

แต่ = CONJUNCTION

U/wordtype = CONJUNCTION

U/ head = แต่

U/sem/head = แต่

U/ sem/pred = but

U/ sem/type = JCRG

U/ sem/ontology = coor_join.

แต่ละ = DETERMINER

U/wordtype = DETERMINER

U/head = แต่ละ

U/sem/head = แต่ละ

U/sem/pred = each

U/sem/type = DIBQ

U/sem/ontology = compare.

Facility = NOUN

U/wordtype = NOUN

U/head = Facility

U/sem/head = Facility

U/sem/pred = Facility

U/sem/type = NCMN
U/sem/ontology = asset.

กรรมการ = NOUN

U/wordtype = NOUN
U/ head = กรรมการ
U/sem/head = กรรมการ
U/ sem/pred = committee
U/ sem/type = NCMN
U/ sem/ontology = team_work.

การก = NOUN

U/wordtype = NOUN
U/head = การก
U/sem/head = การก
U/sem/pred = Case
U/sem/type = NCMN
U/sem/ontology = symbol.

การเงิน = NOUN

U/wordtype = NOUN
U/ head = การเงิน
U/sem/head = การเงิน
U/ sem/pred = finance
U/ sem/type = NCMN
U/ sem/ontology = account.

การงบประมาณแผ่นดิน = NOUN

U/wordtype = NOUN

U/ head = การงบประมาณแผ่นดิน

U/sem/head = การงบประมาณแผ่นดิน

U/ sem/pred = budget

U/ sem/type = NCMN

U/ sem/ontology = account.

ข้อมูล = NOUN

U/wordtype = NOUN

U/head = ข้อมูล

U/sem/head = ข้อมูล

U/sem/pred = data

U/sem/type = NCMN

U/sem/ontology = information.

คอมพิวเตอร์ = NOUN

U/wordtype = NOUN

U/ head = คอมพิวเตอร์

U/sem/head = คอมพิวเตอร์

U/ sem/pred = computer

U/ sem/type = NCMN

U/ sem/ontology = computer.

เครือข่าย = NOUN

U/wordtype = NOUN

U/head = เครือข่าย

U/sem/head = เครือข่าย

U/sem/pred = network

U/sem/type = NCMN

U/sem/ontology = network.

ตาราง = NOUN

U/wordtype = NOUN
U/ head = ตาราง
U/ sem/head = ตาราง
U/ sem/pred = table
U/ sem/type = NCMN
U/ sem/ontology = information

ภาษา = NOUN

U/wordtype = NOUN
U/head = ภาษา
U/sem/head = ภาษา
U/sem/pred = language
U/sem/type = NCMN
U/sem/ontology = language.

APPENDIX F

THAI INPUT SENTENCES AND THEIR TRANSLATION OUTPUT

Input sentences	English Translation Sentences	Mean
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ.	National Electronics and Computer Technology Center	3
คณะ วิศวกรรมศาสตร์.	the faculty of engineering	3
รูป ที่1.	the first figure	3
รัฐมนตรีว่าการกระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน .	the minister of Science Technology and Energy	3
เล่ม 1.	the first volume	3
ในกรณีที่ ออกแบบ ผิดพลาด.	in case of wrong design	2.33
เพื่อ แสดง ตัวหนังสือ.	so that show a character	1.33
ทำให้ ใช้เวลา น้อยลง.	so less spend	1
ปีงบประมาณ 2531.	2531 BE budget year	1.67
คณะ กรรมการบริหาร ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ.	the administrative committee of National Electronics and Computer Technology Center	2.33
โครงการ เครือข่ายคอมพิวเตอร์.	a computer network project	3
หลักสิบ หลักร้อย.	ten digit hundred digit	1.33
13. ออปโตอิเล็กทรอนิกส์.	13 .optoelectronic	2.33
บุญรัตน์ อัตชู.	Boonrat Attachoo	3
ดัง รูป ที่1.	as the first figure	2.67
มี อินพุท อย่างไร.	how does an input have ?	1.67
แสดงออก ทาง มอนิเตอร์.	display on a monitor	3
แต่ละ ช่อง ประกอบ ด้วย.	each channel composes of	3
และ มี เอาท์พุท อย่างไร.	and how does an output have ?	1.67
ซึ่ง ได้รับ ตาม เป้าหมาย.	that receive according to a goal	2
ทำให้ งาน ล่าช้า ไป.	so a work is delayed	2.33
ตาราง ที่2 งบประมาณ.	the second table budget	2
ผล การ ดำเนินงาน ของ.	operation result of	2
โครงการวิจัยและพัฒนา อิเล็กทรอนิกส์ และ คอมพิวเตอร์.	electronic and computer Research Project and Development	1.67
คณะ กรรมการ นโยบาย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ	the policy committee group of National Electronics and Computer Technology Center	2
เลขาธิการ คณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ กรรมการ.	the secretary general of Office of the National Economics and Social Development Board committee	1.67

Input sentences	English Translation Sentences	Mean
คณะ วิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี.	the faculty of engineering King Mongkuts University of Technology Thonburi	2.67
ปลัด กระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน รองประธานกรรมการ.	the deputy of Science Technology and Energy vice president	1
ผู้อำนวยการ สำนักวิจัยและบริการคอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าลาดกระบัง.	the director of Computer Research and Service Center King Mongkuts Institute of Technology	2.67
คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย.	a computer takes a body cross cut picture	2.33
ไม่ ใช้ หลักหน่วย.	do not use a digit	3
7. เครื่องควบคุม มอเตอร์กระแสสลับ.	7. A.C. motor controller machine	2.33
ฯพณฯ รัฐมนตรีว่าการ กระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน .	H.E. minister of Science Technology and Energy	2.67
ดร. บุญวัฒน์ อัตชู.	Dr Boonwat Attachoo	2.67
สรุป ข้อมูล เกี่ยวกับ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ.	summarize the data about National Electronics and Computer Technology Center	2.67
จะ เสีย ค่าใช้จ่าย น้อยลง.	will lose less an expense	1
ชั้น 6 สำนักงานปลัดกระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน.	the sixth floor of Office of the Permanent Secretary of MOST	1.67
15 สิงหาคม 2532.	August 15 2532 B.E .	1.67
เพื่อ พิจารณา โครงการ และ งบประมาณ	so that consider project and budget	1.33
เพื่อ เผยแพร่ ต่อ เอกชน ต่อไป.	so that announce to a private sector later	1.33
ดัง แสดง ใน รูป ที่ 4.1.	as display in the fourth point one figure	2
ดัง แสดง ใน รูป ที่ 5.1.	as display in the fifth point one figure	2
แล้ว จึง ส่ง ให้ ผู้ประกอบการ .	then send to an entrepreneur	2.67
เครื่อง ดังกล่าว ติดตั้ง ที่ โรงพยาบาลสยาม.	that machine is installed at Siam hospital	3
ราคา ประมาณ 20ล้าน บาท.	about price twenty million baht	1.33
ตาราง ที่1 จำนวน บุคลากร.	the first table personnel number	2.33
รูปแบบ การ ทำงาน ของ เครื่อง .	working layout of an instrument	2
ประธานสภา มนตรีสภา สมาคมวิทยาศาสตร์และเทคโนโลยีแห่งประเทศไทย กรรมการ.	the counselor of council chief of council of The Council of Scientific and Technological Associations of Thailand committee	2.33
ห้องประชุม ชั้น 4.	the fourth floor meeting room	2.67
5. อุปกรณ์ตรวจรู้ และ อุปกรณ์ขยายสัญญาณ.	5. sensor and signal amplifier device	2.33
รศ. ดร. สมเกียรติ ศุภเดช.	Associate Professor Dr Somkriate Supadej	2
สำหรับ ต้นแบบ ดังกล่าว ได้แก่.	for that prototype is	2.67
จะ อ่าน ผล ได้ ยาก.	will difficultly read a result	1.33
แต่ละ ส่วน เชื่อมต่อ กัน ได้ อย่างไร.	how did each part connect together ?	2
และ จะ แปลง สัญญาณ ใน ช่วงสะบัดกลับ.	and will convert the signal in a shaking back stage	2.67
สำหรับ ใช้ ใน เครื่องใช้ไฟฟ้า และ อิเล็กทรอนิกส์.	for use in electric equipment and electronic	2.67
ดัง แสดง ใน ตาราง 4.1 .	as display in the fourth point one table	2
เมื่อ ผ่าน วงจรขยาย ภาค แรก แล้ว.	when passed the first section amplifier circuit	2

Input sentences	English Translation Sentences	Mean
เพราะ สร้าง ภาพ ด้วย วิธี เดียวกัน.	because the picture is created with same method	2.33
รองปลัด กระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน กรรมการ และ เลขานุการ.	the Deputy Undersecretary of Science Technology and Energy committee and secretary	2.33
โครงการ คอมพิวเตอร์ อิเล็กทรอนิกส์ ทาง การแพทย์.	a medicine course electronic computer project	1
โครงการ การ พัฒนา คอมพิวเตอร์ ซอฟต์แวร์.	a software computer development project	2.67
ไม่ สามารถ แสดงสัญญาณ ได้ ชัดเจน.	do not clearly display signal	1.67
12. ระบบควบคุม การ ผสม อาหารสัตว์.	12. animal feed mixer controller system	1
16. สารนิเทศ ห้องสมุด ของ มหาวิทยาลัยเกษตรศาสตร์.	16. library information of Kasetsart University	1.67
นาย สมภพ อมาตยกุล กรรมการ.	Mr Sompob Amartayakul committee	2.67
จึง ใช้เวลา 24960 ไมโครวินาที.	take 24960 microsecond	2.33
แต่ละ Block มี หน้าที่ อะไร.	what does each block have function ?	1.33
แต่ละ Block เชื่อมต่อ กัน อย่างไร.	how does each block connect together ?	2
การ ประชุม ทาง วิชาการ ครั้ง ที่1.	the first time academic course conference	1.67
อย่างไรก็ตาม โครงการวิจัย นี้ ยัง ไม่ เสร็จสิ้น สมบูรณ์.	however this research project has not finished completely	1.67
วิธี นี้ ใน สมัยก่อน ใช้ กัน มาก.	this method in the past uses much together	1
ภาควิชา อิเล็กทรอนิกส์ คณะ วิศวกรรมศาสตร์ สถาบัน เทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง.	the engineering faculty electronic department of King Mongkuts Institute of Technology	1.33
1 คอมพิวเตอร์ 32 บิท.	1 computer 32 bit	2
3 วงจรรวม เชิง พาณิชย์ สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี.	3 commercial base Integrated Circuit of King Mongkuts University of Technology Thonburi	1.67
จะ มี ลักษณะ ดัง รูป 1.	will have the appearance as the first figure	2.33
10 โครงสร้าง ข้อมูล ของ ข้อมูล ส่วนตัว .	10 data structure of personal data	1.67
เวลา ใน การ สแกน รวดเร็ว.	the time in scanning quickly	1.33
2 โครงสร้าง อุปกรณ์ ใน การ วิจัย.	2 device structure in research	2.67
โดย ใช้ อุปกรณ์ รังสีเอ็กซ์ และ ระบบคอมพิวเตอร์.	by using X ray and computer system device	2.67
ดัง รูป ที่5.	as the fifth figure	2.67
รูป ที่4 ผัง การ ทำงาน ระบบ สินค้าคงคลัง.	the fourth figure inventory system working chart	2.33
และ ปรับ กระแส ได้ สูงสุด 25 mA.	and electricity was adjusted maximum 25 mA	2
แต่ละ ชุด จะ ทำหน้าที่ ต่างกัน .	each set will perform differently	3
จอมอนิเตอร์ แสดง อักษร ขนาด 12 นิ้ว.	a monitor displays 12 inch character size	2.67
หน่วยความจำ หลัก 2 เม็กกะไบท์.	main memory 2 megabyte	2
พร้อมกับ แสดง เส้นตัดขวาง ของ ภาพตัดขวาง ด้วย.	together with display the cross cut line of a cross cut picture too	1.67
รูป ที่1 คอมพิวเตอร์ ถ่าย ภาพตัดขวาง ร่างกาย รุ่น ที่1.	the first figure a computer photographs body cross cut picture the first version -	2
รูป ที่6 ระบบ ไมโครคอมพิวเตอร์ ชุด แรก.	the sixth figure the first set microcomputer system	2.33
รูป ที่7 ไมโครคอมพิวเตอร์ ชุด ที่สอง.	the seventh figure the second set microcomputer	2.33
รูป ที่8 อุปกรณ์ รังสีเอ็กซ์ ชุด ใหม่.	the eighth figure the new set X ray device	2
ผลงานวิจัย 3 การ พัฒนา ระบบอินพุทเอาท์พุท และ ระบบ	a research 3 IO system and supporting and translation system	2

Input sentences	English Translation Sentences	Mean
สนับสนุนการแปลภาษา.	development	
ผลงานวิจัย 5 ระบบเตรียมข้อมูลคอมพิวเตอร์อัตโนมัติ ภาษาไทย และ ภาษาอังกฤษ.	a research 5 Thai language and English language automatic computer data preparing system	2
บทคัดย่อ การ เสนอ ผลงาน.	a result of work presentation abstract	1.67
ไม่ ผิด ระเบียบ ด้าน การเงิน การงบประมาณแผ่นดิน.	do not against a budget finance area regulation	2
หน่วยงาน ต่างๆ สามารถ นำไปใช้ ได้ .	other institutes apply	1.33
โปรแกรมคอมพิวเตอร์ ก็ จะ ต้อง ใหญ่ ขึ้น เท่านั้น.	a computer program expand only	1.67
แต่ เป็น ภาษาสากล.	but become an International language	1.67
โดย ใช้ ระบบ VENUS ของ ประเทศญี่ปุ่น.	by using the VENUS system of Japan	2.67
3.2 ภาค กำเนิด สัญญาณ .	3.2 part make signal - 32 version of signal generating	1.33
ระบบ MU พัฒนา ด้วย ภาษา LISP.	system of MU is developed with language of LISP	2
โนด จะ แทน CP และ Case.	a node will replace CP and Case	2.67
3. กำหนดค่า TCAT ให้กับ CP แต่ละ ตัว	3. determines a TCAT for each CP unit	2.67
6..กำหนดค่า รูปแบบ กริยา ให้กับ predicate CP.	6. determines a verb pattern for a predicate CP	2.67
ประโยค นี้ จะ มี รูปแบบ ภาษากลาง .	this sentence will have an Interlingua pattern	1.33
ส่วน Case นั้น จะ ถูก แทนที่ ด้วย คำบุพบท และ คำสันธาน .	that Case part will be replaced with Preposition and Conjunction	2.67
สุดท้าย คณะ ผู้วิจัย ได้ กำหนด โครงสร้าง อิเล็กทรอนิกส์ ดิกชั่นนารี.	the last researcher group assigned an Electronic dictionary structure	2
แต่ละ บรรทัด ปิดท้าย ด้วย null .	each line is ended with null	2.33
รูป ที่5 แสดง Language Dependent Semantic Structure .	the fifth figure illustrate a Dependent Semantic Structure Language	2
1. หา syntactic role ของ แต่ละ node .	first find the syntactic role of each node	2
6. ศึกษา โครงสร้าง ของ พจนานุกรม.	6. study the structure of dictionary	2.33
บทความ ผลงานวิจัย นี้ เสนอ ระบบเตรียมข้อมูลคอมพิวเตอร์ อัตโนมัติภาษาไทยและภาษาอังกฤษ .	this research article proposes an Automatic Recognition of Thai English Characters System	1.33
ลักษณะ ลายเส้น ของ ตัวอักษร จะ ถูก แทน ด้วย รหัส 1.	the lined figure appearance of alphabet will be replaced with first code	2
พัฒนา ซอฟต์แวร์ ของ Multi font Printed Thai Character recognition .	first develop the software of Multi font Printed Thai Character recognition	2
2 . สร้าง Card ซอฟต์แวร์ Multi font Printed Thai Character recognition.	2. build a Multi font Printed Thai Character recognition Card software	2
4. สร้าง Card ซอฟต์แวร์ ของ Multi font Printed English Character recognition .	4. build the software Card of Multi font Printed English Character recognition	2.33
ข้อมูล ชุด นี้ พิมพ์ คอมพิวเตอร์.	this set data is printed computer	1.33
2. คนไทย กิน ซุป ด้วย ช้อน .	2. Thai people eat the soup with spoon	2.67
บาง การก อาจ แสดง ด้วย คำ ทาง ไวยากรณ์ หรือ คำเชื่อม บาง คำ	some Cases display word or conjunction	1.33
1. สำรวจ สถานภาพ ระบบ ฐานข้อมูล ห้องสมุด หน่วยงาน เครือข่าย	first explore an organization library database system status	1.33
ฐานข้อมูล มี โครงสร้าง ฐานข้อมูล ต่างๆ กัน .	a database have other database structure together	1
รูป ที่ 3 แสดง ลวดลาย ของ NMOS และ PMOS ทรานซิสเตอร์.	the third figure display the pattern of transistor NMOS and PMOS	2
รูป ที่ 11 แสดง คุณสมบัติ ของ ทรานซิสเตอร์ แบบ nMOS.	the eleventh figure display the attribute of nMOS type transistor	2

Input sentences	English Translation Sentences	Mean
เครื่อง จะ พิมพ์ ใบเสร็จรับเงิน ให้ .	an instrument will print receipt	2
8042 จะ ควบคุม สัญญาณ เหล่านี้ .	8042 will control these signal	2.67
Extension slot มี ทั้งหมด จำนวน 8 slot.	an extension slot have all amount 8 slot	2
System bus ถูก สร้าง มา จาก CPU 80386SX.	a system bus is created from 80386SX CPU	2.67
System address bus มี ขนาด 20 บิท เท่านั้น.	a system address bus have only 20 bit size	2
ขนาด ของ data bus จะ เป็น ขนาด 8 บิท.	Size of data bus will be 8 bit size	2
bus ของ DRAM จะ อยู่ ทาง local bus .	bus of DRAM will be on a local bus	2.33
ขนาด ของ ROM จะ เป็น ชนิด 16 บิท .	size of ROM will be 16 bit type	2.67
สัญญาณเลือก ROM จะ ถูก สร้าง มา โดย 82335 .	a ROM signal will be built by 82335	2.67
แบบ 8 บิท จะ เป็น slot ขนาด 62 ขา .	8 bit type will be 62 leg size slot	2
LA17 LA23 เป็น สัญญาณ แอดเดรส ของ ระบบ.	a LA17 LA23 is the address signal of a system	3
เครื่องต้นแบบ เครื่อง ที่1 มี บล๊อคไดอะแกรม .	the first machine prototype have a block diagram	2
2 POWER DIVIDER คือ ตัวแบ่งกำลังสัญญาณ .	2 POWER DIVIDER is signal divider	2.67
อีก ทาง หนึ่ง ป้อน กลับมา เข้า พรีเสกลเลอร์ .	another one way feed into a prescaler	2
ระบบเก็บข้อมูล อัตโนมัติ แบบ เอนกประสงค์.	the automatic data storage system multipurpose type	2
Data Logger แต่ละ ตัว จะ มี จุดวัด 16 จุด.	each Data Logger unit will have checkpoint 16 point	1.67
เครื่อง จะ บันทึก ข้อมูล ท่าน ได้ ใน หน่วยความจำ.	an instrument will record your data in memory	2
2.1 Data Logger จะ มี ข้อกำหนด ดังนี้ .	2.1 Data Logger will have principle as follows	2.33
3. หน่วยเวลา เป็น หน่วยเวลา มาตรฐาน ให้กับ ซีพียู .	3. Clock Calender is the standard Clock Calender for a CPU	1.67
5.2 ซอฟท์แวร์ ควบคุม เครื่อง Data Logger .	5. 2 software control a Data Logger instrument	3
ยกระดับ มาตรฐานการครองชีพ ของ ประชาชน.	4 increase the living standard of citizen	2
รูป ที่ 3.1 แสดง บล็อคไดอะแกรม ของ ตัวควบคุม แบบ ลำดับ.	the third point one figure display the block diagram of order type controller	1.67
แบตเตอรี เป็น ตัวจ่ายไฟเลี้ยง ให้กับ หน่วยความจำ ชนิด RAM.	7 battery is the power service for RAM type memory	2
วงจรอินพุท จะ ต้อง มี คุณสมบัติ ดังนี้.	an input circuit will have attribute as follows	2.33
3 ต้อง สามารถ แยก ระบบ ออก จาก เอาท์พุท .	3 separate the system from output	2
เครื่อง จะ แสดง ERROR .	an instrument will display ERROR	2.67
เครื่อง จึง จะ ทำงาน ต่อไป ได้.	an instrument will work later	2.33
โปรแกรม ขั้นบันได จะ ถูก อ่าน ออกมา ทีละ ขั้น.	staircase program will be read each step	2
SERVICE มี หน้าที่ เช่นเดียวกับ โหมด โปรแกรม .	a SERVICE have the function as program mode	1.67
4.2 คำสั่ง เบื้องต้น ของ PC .	basic 4 2 instruction of PC	2
โครงสร้าง ของ ระบบ จะ เป็น แบบ โมดูล .	structure of system will be a module type	2.67
รูป ที่2 แสดง รูปลักษณะ ของ แผงควบคุม .	the second figure display the characteristic of control panel	2.33
แผนผัง โครงสร้าง ภายใน ของ ไอซี PWM แสดง ดัง รูป ที่3 .	an internal structure diagram of PWM IC show as the third figure	2
ส่วน รูปคลื่นสัญญาณ PWM จะ เป็น ดัง รูป ที่5 .	PWM signal wave form part will be as the fifth figure	2.33
คุณสมบัติ ต่างๆ ดังกล่าว ข้างต้น สามารถ ใช้ งาน ได้ผลดี .	other property above that perform work effectively	1.67
รูป ที่9 แสดง ภาพ วงจรควบคุม ของ เครื่องต้นแบบ .	the ninth figure show the controller circuit picture of model	1.67

Input sentences	English Translation Sentences	Mean
ลักษณะ สมบัติ กระแส แรงดัน แสดง ตาม รูป ที่8 .	pressure electricity property character show according to the eighth figure	1.67
1. เตาเผา พร้อม ชุดควบคุมอุณหภูมิ ประกอบ ด้วย หม้อแปลง และ วารีแอต .	1 kiln together temperature controller compose with transformer and variate	1.33
จำนวน ผู้ใช้ มี มากขึ้น ตามลำดับ .	user number have increasingly respectively	1
ปัญหา ด้าน เนื้อหา สารนิเทศเฉพาะทาง มี ศัพท์เทคนิค มาก .	specific information content area problem have much technical word	1.67
1 ศึกษา สถานภาพ ต่างๆ ของ ระบบ เครือข่ายคอมพิวเตอร์ .	first study the other status of computer network system	2.67
โน๊ต ดนตรี นั้น จะ แบ่ง ออกเป็น ช่วงเสียง .	that music notation will be divided to octave interval	1.67
วงจรนับ ลง จะ รับ ข้อมูล ตัวโน๊ต จาก ภาคเก็บค่าข้อมูลตัวโน๊ต ของเพลง .	a counter circuit will receive note data from song note storage section	2.33
2 ภาคสร้างความถี่มาตรฐาน ยกเว้น ตัวกำเนิดความถี่มูลฐาน เช่น cystal .	2 standard frequency part except the fundamental frequency origin such as a cystal	2.33
เอาท์พุต ของ comparator ผ่าน เข้า Full Adder .	output of comparator flow Full Adder	1.67
รูป ที่ 10 แสดง จำนวน ขา ของ วงจรวีแอลเอสไอ .	the tenth figure show the pin amount of VLSI circuit	2
แผนภาพ ของ ชิพ Gate array เป็น ดัง รูป ที่ 2 .	the Gate array diagram of chip is as the second figure	2.33
ทั้งสอง ส่วน นี้ จะ ป้อนเข้า คอมพิวเตอร์.	this both part will feed computer	1.67
ไมโครคอมพิวเตอร์ ขนาด 16 บิท มี หน่วยความจำหลัก 640 kbyte .	16 bit size microcomputer have 640 kbyte main memory	2
โครงการ นี้ คณะผู้วิจัย ได้ ดำเนินการ เสร็จสิ้น แล้ว.	this research team project performed completely already	1.33
คณะผู้วิจัย ขอขอบคุณ มา ณ ที่ นี้ .	a research team thanks at this here	1.33
ปัจจุบัน ได้ นำเข้า เครื่อง ดังกล่าว ของ บริษัทฮิตาชิ .	a present imported that machine of Hitachi companyb	1.33
1. Electromygraphy ใช้ electrode แบบ เข็ม .	1. Electromygraphy use a pin electrode type	2.33
8. รูปคลื่น ของ สัญญาณ กระตุ้น square wave .	8. wave form of signal activate a square wave	2.33
ตัวอย่าง รูป ที่5 คือ MOVE.	the fifth figure example is a MOVE	2.33
โนด จะ แทน CP และ Case .	a node will replace CP and Case	2.67
ตอนกลาง จะ เป็น บรรทัด ของ ตัวอักษร ทั่วไป .	a center will be the line of general alphabet	1.67
โครงการ นี้ จะ สนับสนุน ระบบ สารนิเทศ ใน ระดับ ประเทศ .	this project will support the information system in national level	2.67
6. สมาคมส่งเสริมเทคโนโลยี มี ศูนย์ข้อมูล ทาง เทคโนโลยี .	6. Technological Promotion Association have technology area data center	1.67
ศัพท์สัมพันธ์ ROOT มี คำศัพท์ ทั้งสิ้น ประมาณ 18800 คำ.	ROOT relationship vocabulary have all vocabulary about 18800 word	1.67
แต่ละ สาขา จะ มี คำศัพท์ ประมาณ 1200 คำ .	each branch will have vocabulary about 1200 word	1.67
คณะอนุกรรมการ ควร มี องค์ประกอบ ดังนี้ .	a subcommittee group have infrastructure as follows	2
มาตรฐาน เหล่านี้ จะ มี ชุด อักษร ใน กลุ่ม Go ร่วมกัน .	a standard will have a character set	1.33
มาตรฐาน นี้ เป็น ฉบับร่าง ครั้ง ที่ 2 .	this standard is the second time outline	2
ภาษา C ยัง ทำงาน ได้ เช่นเดียวกับ ภาษาระดับต่ำ อีกด้วย .	a C language worked as low level language also	1.67
วิธี นี้ จะ มี ตัวแปร ประเภท Facility .	this method will have Facility type varaible	2
ตาราง ที่ 1 แสดง ผล ของ คำสั่ง H PINCHK .	the first table display the result of H PINCHK instruction	2
รูป ที่ 5 บล็อกไดอะแกรม แสดง โครงสร้าง ของ EPM5064 .	the fifth block diagram figure display the structure of EPM5064	2
6 Display เป็น seven_segments 6 ตัว .	the sixth Display is 6 unit seven segments	1.67

Input sentences	English Translation Sentences	Mean
รูป ที่ 10 แสดง วงจรรวม ส่วน ที่ 1 .	the tenth figure illustrate the first part Integrated Circuit	2
ผู้ใช้ จะ ควบคุม เครื่อง ได้ .	an user will control instrument	2
หน่วยคำ ต่างๆ เหล่านี้ สามารถ สลับ ตำแหน่ง กัน ได้.	these other word units change position together	1.33
vp จะ ต้อง เป็น เอกพจน์ ด้วย เป็นต้น .	a vp will be singular too etc	1.67
5.1 พัฒนา PATR PARSER ด้าน โปรแกรม ภาษา Prolog .	5.1 develop a Prolog language program area PATR PARSER	2.33
โปรแกรม TSP เขียน ขึ้น ด้วย ภาษา LPA Prolog 2.5 .	TSP program is written with language the second point two LPA	1.33
Total		2.05

APPENDIX G

SENTENCE LEVEL FtoC-PATTERNS

SENTENCE LEVEL (THAI SENTENCE)

Type	THAI SENTENCE PATTERNS						ENGLISH SENTENCE PATTERNS					
pres	VP verb	SUBJ NP	OBJ NP	OBLIQUE PP			SUBJ NP	VP verb	OBJ NP	OBLIQUE PP		
	VP verb	SUBJ NP	OBJ 1 NP	OBJ 2 NP			SUBJ NP	VP verb	OBJ 2 NP	OBJ 1 NP		
	VP verb	SUBJ NP	OBJ NP	MODIFIER adverb			SUBJ NP	VP verb	OBJ NP	MODIFIER adverb		
	VP verb	SUBJ NP	MODIFIER 1 adverb	MODIFIER 2 adverb			SUBJ NP	VP verb	MODIFIER 1 adverb	MODIFIER 2 adverb		
	VP verb	SUBJ NP	OBLIQUE PP	MODIFIER adverb			SUBJ NP	VP verb	OBLIQUE PP	MODIFIER adverb		
	VP verb	SUBJ NP	OBJ 1 NP	OBJ 2 NP	OBLIQUE PP		SUBJ NP	VP verb	OBJ 2 NP	OBJ 1 NP	OBLIQUE PP	
	VP verb	SUBJ NP	OBJ NP				SUBJ NP	VP verb	OBJ NP			
	VP verb	SUBJ NP	OBJ 1 NP	OBJ 2 NP	OBJ 3 NP		SUBJ NP	VP verb	OBJ 3 NP	OBJ 2 NP	OBJ 1 NP	
	VP verb	SUBJ NP	OBLIQUE PP				SUBJ NP	VP verb	OBLIQUE PP			
	VP verb	SUBJ NP	OBJ 1 NP	OBJ 2 NP	OBJ 3 NP	OBJ4 NP	SUBJ NP	VP verb	OBJ 4 NP	OBJ 3 NP	OBJ 2 NP	OBJ 1 NP

Type	THAI SENTENCE PATTERNS			ENGLISH SENTENCE PATTERNS		
future + passive	VP adjunct + verb	SUBJ NP	OBLIQUE PP	SUBJ NP	VP will, shall + be + v_3	OBLIQUE PP
	VP adjunct1 + adjunct 2 + verb	SUBJ NP	OBLIQUE PP	SUBJ NP	VP will, shall + be + v_3	OBLIQUE PP
	VP adjunct1 + adjunct 2 + verb + adjunct 3	SUBJ NP	OBLIQUE PP	SUBJ NP	VP will, shall + be + v_3	OBLIQUE PP

Type	THAI SENTENCE PATTERNS						ENGLISH SENTENCE PATTERNS					
future	VP adjunct + verb	SUBJ NP	OBJ NP				SUBJ NP	VP will, shall + v	OBJ NP			
	VP adjunct + verb	SUBJ NP	OBJ NP	ADJUNCTION auxiliary			SUBJ NP	VP will, shall + v	OBJ NP			
	VP adjunct + verb	SUBJ NP	OBJ 1 NP	OBJ 2 NP			SUBJ NP	VP will, shall + v	OBJ 2 NP	OBJ 1 NP		
	VP adjunct + verb	SUBJ NP	OBJ NP	OBLIQUE PP			SUBJ NP	VP will, shall + v	OBJ NP	OBLIQUE PP		
	VP adjunct + verb	SUBJ NP	OBJ NP	MODIFIER adverb			SUBJ NP	VP will, shall + v	OBJ NP	MODIFIER adverb		
	VP adjunct+ verb	SUBJ NP	OBJ NP	ADJUNCTION auxiliary	OBLIQUE PP		SUBJ NP	VP will, shall + v	OBJ NP	OBLIQUE PP		
	VP adjunct + verb	SUBJ NP	OBJ 1 NP	OBJ 2 NP	OBLIQUE PP		SUBJ NP	VP will, shall + v	OBJ 2 NP	OBJ 1 NP	OBLIQUE PP	
	VP adjunct + verb	SUBJ NP	OBJ 1 NP	MODIFIER adverb	OBJ 2 NP		SUBJ NP	VP will, shall + v	OBJ 1 NP	MODIFIER adverb	OBJ 2 NP	
	VP adjunct + verb	SUBJ NP	OBJ 1 NP	OBJ 2 NP	OBLIQUE PP	MODIFIER adverb	SUBJ NP	VP will, shall + v	OBJ 2 NP	OBJ 1 NP	OBLIQUE PP	MODIFIER adverb

Present Perfect

THAI SENTENCE PATTERNS			ENGLISH SENTENCE PATTERNS		
VP	SUBJ	MODIFIER	SUBJ	VP	MODIFIER
Adjunct + V	NP	Adverb	NP	has,have + V_3	Adverb

Present Passive

THAI SENTENCE PATTERNS			ENGLISH SENTENCE PATTERNS		
VP	SUBJ	OBLIQUE	SUBJ	VP	OBLIQUE
ADJUNCT + V	NP	PP	NP	Is, am, are + V_3	PP

Past

THAI SENTENCE PATTERNS				ENGLISH SENTENCE PATTERNS			
VP	SUBJ	OBJ 1	OBJ 2	SUBJ	VP	OBJ 2	OBJ 1
ADJUNCT + V	NP	NP	NP	NP	V_2	NP	NP

APPENDIX H

NOUN PHRASE LEVEL FtoC PATTERNS

PHRASE LEVELS - NP
Thai Patterns

NOUN

NOUN	MODIFIER
Common Noun	Ordinal Number

NOUN	MODIFIER
Common Noun	NP(Class + Deter)

NOUN	MODIFIER
Proper Noun	Common Noun

NOUN	MODIFIER
Common Noun	Cardinal Number

NOUN	MODIFIER
Common Noun	Common Noun

NOUN	MODIFIER
Common Noun	CoorCon

NOUN	MODIFIER
Common Noun	Label Noun

NOUN	MODIFIER
Common Noun	Determiner

PHRASE LEVELS- NP
English Patterns

NOUN

The	MODIFIER	NOUN
	Ordinal Number	Common Noun

MODIFIER	NOUN
NP(Class + Deter)	Common Noun

The	MODIFIER	of	NOUN
	Common Noun		Proper Noun

MODIFIER	NOUN
Cardinal Number	Common Noun

MODIFIER	NOUN
Common Noun	Common Noun

NOUN	MODIFIER
Common Noun	CoorCon

MODIFIER	NOUN
Label Noun	Common Noun

MODIFIER	NOUN
Determiner	Common Noun

PHRASE LEVELS - NP
Thai Patterns

NOUN	MODIFIER
Common Noun	Ordinal Noun

NOUN	MODIFIER
Common Noun	NP

NOUN	ADJUNCTION
Common Noun	PP

NOUN	MODIFIER
Ordinal Number	Common Noun

NOUN	MODIFIER
Ordinal Noun	Classifier

NOUN	MODIFIER
Label Noun	Common Noun

NOUN	MODIFIER
Label Noun	CoorCon

NOUN	MODIFIER
Title Noun	NP

NOUN	MODIFIER
Title Noun	Title Noun

NOUN	MODIFIER
Proper Noun	Proper Noun

PHRASE LEVELS- NP
English Patterns

MODIFIER	NOUN
Ordinal Noun	Common Noun

MODIFIER	NOUN
NP	Common Noun

NOUN	ADJUNCTION
Common Noun	PP

NOUN	MODIFIER
Ordinal Number	Common Noun

NOUN	MODIFIER
Ordinal Noun	Classifier

NOUN	MODIFIER
Label Noun	Common Noun

NOUN	MODIFIER
Label Noun	CoorCon

NOUN	MODIFIER
Title Noun	NP

NOUN	MODIFIER
Title Noun	Title Noun

NOUN	MODIFIER
Proper Noun	Proper Noun

PHRASE LEVELS - NP	
Thai Patterns	
MODIFIER	NOUN
Classifier	Proper Noun

CLASSIFIER	MODIFIER
	Ordinal Number

NOUN	MODIFIER
Ordinal Noun	Classifier

CLASSIFIER	MODIFIER
	Determiner

CLASSIFIER	MODIFIER
	Determiner

NOUN	MODIFIER
Common Noun	Determiner

NOUN	MODIFIER
	NP

NOUN	MODIFIER
	Pronoun

NOUN	MODIFIER
Proper Noun	NP

NOUN	MODIFIER
Proper Noun	NP

NP	ADJUNCTION
	PP

PHRASE LEVELS- NP			
English Patterns			
The	MODIFIER	of	NOUN
	Classifier		Proper Noun

The	MODIFIER	CLASSIFIER
	Ordinal Number	

The	NOUN	MODIFIER
	Ordinal Noun	Classifier

MODIFIER	CLASSIFIER
Determiner	

MODIFIER	CLASSIFIER
Determiner	

MODIFIER	NOUN
Determiner	Common Noun

MODIFIER	NOUN
NP	

MODIFIER	NOUN
Pronoun	

The	MODIFIER	of	NOUN
	NP		Proper Noun

The	MODIFIER	of	NOUN
	NP		Proper Noun

NP	ADJUNCTION
	PP

PHRASE LEVELS - NP
Thai Patterns

Nominal Prefix	MODIFIER
	noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Nominal Prefix	NP

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Ordinal Number	Common Noun

NOUN	MODIFIER	ADJUNCTION
Common Noun	Nominal Prefix	PP

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Common Noun	Common Noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Common Noun	Determiner

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	Determiner

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Classifier	Common Noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	Common Noun

NOUN	MODIFIER	ADJUNCTION
Common Noun	Nominal Prefix	PP

PHRASE LEVELS- NP
English Patterns

Nominal Prefix	MODIFIER
	noun

MODIFIER 2	Nominal Prefix	NOUN
NP		Common Noun

MODIFIER 1	NOUN	MODIFIER 2
Ordinal Number	Common Noun	Common Noun

MODIFIER	NOUN	ADJUNCTION
Nominal Prefix	Common Noun	PP

MODIFIER 2	MODIFIER 1	NOUN
Common Noun	Common Noun	Common Noun

MODIFIER 2	MODIFIER 1	NOUN
Determiner	Common Noun	Common Noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	Determiner

MODIFIER 2	MODIFIER 1	NOUN
Common Noun	Classifier	Common Noun

MODIFIER 1	NOUN	MODIFIER 2
Determiner	Common Noun	Common Noun

Nominal Prefix	NOUN	ADJUNCTION
	Common Noun	PP

PHRASE LEVELS - NP Thai Patterns		
NOUN	MODIFIER	ADJUNCTION
Common Noun	Common Noun	PP

PHRASE LEVELS- NP English Patterns		
MODIFIER	NOUN	ADJUNCTION
Common Noun	Common Noun	PP

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	Classifier

MODIFIER 1	MODIFIER 2	NOUN
Determiner	Classifier	Common Noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Classifier	Determiner

MODIFIER 2	NOUN	MODIFIER 1
Determiner	Common Noun	Classifier

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	NP	NP

MODIFIER 2	MODIFIER 1	NOUN
NP	NP	Common Noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	NP

MODIFIER 1	NOUN	MODIFIER 2
Determiner	Common Noun	NP

NOUN	MODIFIER	ADJUNCTION
Label Noun	Nominal Prefix	PP

NOUN	Nominal Prefix	ADJUNCTION
Label Noun		PP

NOUN	MODIFIER 1	MODIFIER 2
Label Noun	Common Noun	Common Noun

NOUN	MODIFIER 2	MODIFIER 1
Label Noun	Common Noun	Common Noun

NOUN	MODIFIER 1	MODIFIER 2
Label Noun	NP	NP

NOUN	MODIFIER 1	MODIFIER 2
Label Noun	NP	NP

NOUN	MODIFIER 1	MODIFIER 2
Proper Noun	Title Noun	Common Noun

MODIFIER 1	MODIFIER 2	of	NOUN
Title Noun	Common Noun		Proper Noun

NOUN	MODIFIER 1	MODIFIER 2
Proper Noun	Title Noun	Proper Noun

MODIFIER 1	NOUN	MODIFIER 2
Title Noun	Proper Noun	Proper Noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	Ordinal Noun

NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	Ordinal Noun

PHRASE LEVELS - NP		
Thai Patterns		
NOUN	MODIFIER 1	MODIFIER 2
Common Noun	Determiner	Common Noun

CLASSIFIER	MODIFIER 1	MODIFIER 2
	Determiner	Determiner

CLASSIFIER	MODIFIER 1	MODIFIER 2
	Determiner	Determiner

Nominal Prefix	MODIFIER 1	MODIFIER 2
	NP	NP

NOUN	MODIFIER 1	MODIFIER 2
Proper Noun	NP	Proper Noun

NOUN	MODIFIER 1	MODIFIER 2
Proper Noun	NP	Proper Noun

NOUN	MODIFIER 1	MODIFIER 2
Proper Noun	NP	NP

NOUN	MODIFIER 1	MODIFIER 2
Proper Noun	NP	NP

NP	MODIFIER 1	MODIFIER 2
	NP	NP

NOUN	MODIFIER 1	MODIFIER 2	MODIFIER 3
Common Noun	Determiner	Determiner	Classifier

NOUN	MODIFIER 1	MODIFIER 2	MODIFIER 3
Proper Noun	Common Noun	Common Noun	Common Noun

PHRASE LEVELS- NP		
English Patterns		
MODIFIER 1	MODIFIER 2	NOUN
Determiner	Common Noun	Common Noun

MODIFIER 1	MODIFIER 2	CLASSIFIER
Determiner	Determiner	

MODIFIER 2	MODIFIER 1	CLASSIFIER
Determiner	Determiner	

MODIFIER 2	MODIFIER 1	Nominal Prefix
NP	NP	

MODIFIER 1	NOUN	MODIFIER 2
NP	Proper Noun	Proper Noun

The	MODIFIER 1	of	NOUN	MODIFIER 2
	NP		Proper Noun	Proper Noun

The	MODIFIER 1	of	NOUN	MODIFIER 2
	NP		Proper Noun	NP

The	MODIFIER 2	MODIFIER 1	of	NOUN
	NP	NP		Proper Noun

NP	MODIFIER 2	MODIFIER 1
	NP	NP

NOUN	MODIFIER 1	MODIFIER 2	MODIFIER 3
Common Noun	Determiner	Determiner	Classifier

The	MODIFIER 2	MODIFIER 1	of	NOUN	MODIFIER 3
	Common Noun	Common Noun		Proper Noun	Common Noun

PHRASE LEVELS - NP
Thai Patterns

NOUN	Nominal Prefix	MODIFIER 2	MODIFIER 3
Common Noun		NP	NP

NOUN	MODIFIER 1	MODIFIER 2	MODIFIER 3
Common Noun	Common Noun	Common Noun	Common Noun

NOUN	MODIFIER 1	MODIFIER 2	MODIFIER 3
Proper Noun	Label Noun	NP	NP

NOUN	MODIFIER 1	MODIFIER 2	ADJUNCTION
Label Noun	NP	NP	PP

NOUN	MODIFIER 1	Nominal Prefix	MODIFIER 2
Label Noun	NP		NP

NOUN	MODIFIER 1	MODIFIER 2	MODIFIER 3
Proper Noun	Title Noun	Proper Noun	NP

NP	MODIFIER 1	MODIFIER 2	MODIFIER 3
	NP	NP	NP

NOUN	MODIFIER 1	MODIFIER 2	ADJUNCTION
Ordinal Noun	NP	NP	PP

NOUN	MODIFIER 1	MODIFIER 2	MODIFIER 3	MODIFIER 4	MODIFIER 5
Common Noun	Determiner	Common Noun	Nominal Prefix	Common Noun	Common Noun

PHRASE LEVELS- NP
English Patterns

MODIFIER 3	MODIFIER 2	Nominal Prefix	NOUN
NP	NP		Common Noun

MODIFIER 3	MODIFIER 2	MODIFIER 1	NOUN
Common Noun	Common Noun	Common Noun	Common Noun

MODIFIER 1	MODIFIER 3	MODIFIER 2	of	NOUN
Label Noun	NP	NP		Proper Noun

NOUN	MODIFIER 2	MODIFIER 1	ADJUNCTION
Label Noun	NP	NP	PP

NOUN	MODIFIER 2	Nominal Prefix	MODIFIER 1
Label Noun	NP		NP

MODIFIER 1	NOUN	MODIFIER 2	MODIFIER 3
Title Noun	Proper Noun	Proper Noun	NP

NP	MODIFIER 3	MODIFIER 2	MODIFIER 1
	NP	NP	NP

NOUN	MODIFIER 2	MODIFIER 1	ADJUNCTION
Ordinal Noun	NP	NP	PP

The	MODIFIER 1	NOUN	MODIFIER 5	MODIFIER 4	MODIFIER 3
	Determiner	Common Noun	Common Noun	Common Noun	Nominal Prefix

APPENDIX I

SUBORDINATION CONJUNCTION

Subordinating Conjunction (Subor)

Thai Patterns

VERB PHRASE	SubOr	MODIFIER
VP		Adverb

VERB PHRASE	SubOr	OBJECT
VP		NP

VERB PHRASE	SubOr	OBLIQUE PHRASE
VP		(PP)

VERB PHRASE	SubOr	SUBJECT
VP		NP

VERB PHRASE	SubOr	SUBJECT
VP		NP

VERB PHRASE	SubOr	OBLIQUE PHRASE	MODIFIER
VP		(PP)	Adverb

VERB PHRASE	SubOr	OBJECT 1	OBLIQUE PHRASE
VP		NP	(PP)

Subordinating Conjunction (Subor)

English Patterns

SubOr	MODIFIER	VERB PHRASE
	Adverb	VP

SubOr	VERB PHRASE	OBJECT
	VP	NP

SubOr	VERB PHRASE	OBLIQUE PHRASE
	VP	(PP)

SubOr	SUBJECT	VERB PHRASE
	NP	VP

SubOr	SUBJECT	VERB PHRASE
	NP	VP

SubOr	VERB PHRASE	OBLIQUE PHRASE	MODIFIER
	VP	(PP)	Adverb

SubOr	OBJECT 1	VERB PHRASE	OBLIQUE PHRASE
	NP	VP	(PP)

Subordinating Conjunction (Subor)

Thai Patterns

VERB PHRASE	SubOr	OBJECT 1	OBJECT 2
VP		NP	NP

VERB PHRASE	SubOr	OBJECT 1	OBLIQUE PHRASE	MODIFIER
VP		NP	(PP)	Adverb

VERB PHRASE	SubOr	OBJECT 1	OBJECT 2	OBLIQUE PHRASE
VP		NP	NP	(PP)

VERB PHRASE	SubOr	OBJECT 1	OBJECT 2	ADJUNCTION
VP		NP	NP	Auxiliary

Coordinating Conjunction (CoorCon)

CoorCon	CoorCon1	CoorCon2
	NP	NP

VERB PHRASE	CoorCon	OBJECT
VP		NP

VERB PHRASE	CoorCon	OBJECT 1	OBJECT 2
VP		NP	Pronoun

VERB PHRASE	CoorCon	OBJECT	OBLIQUE PHRASE
VP		NP	(PP)

Subordinating Conjunction (Subor)

English Patterns

SubOr	VERB PHRASE	OBJECT 2	OBJECT 1
	VP	NP	NP

SubOr	VERB PHRASE	OBJECT 1	OBLIQUE PHRASE	MODIFIER
	VP	NP	(PP)	Adverb

SubOr	VERB PHRASE	OBJECT 2	OBJECT 1	OBLIQUE PHRASE
	VP	NP	NP	(PP)

SubOr	VERB PHRASE	OBJECT 2	OBJECT 1
	VP	NP	NP

Coordinating Conjunction (CoorCon)

CoorCon1	CoorCon	CoorCon2
NP		NP

CoorCon	VERB PHRASE	OBJECT
	VP	NP

CoorCon	OBJECT 2	Auxiliary Verb	OBJECT 1	VERB PHRASE
	Pronoun		NP	VP

CoorCon	VERB PHRASE	OBJECT	OBLIQUE PHRASE
	VP	NP	(PP)

Subordinating Conjunction (Subor)

Thai Patterns					
VERB PHRASE	CoorCon	OBJECT 1	ADJUNCTION	MODIFIER	OBJECT 2
VP		NP	AUXILIARY	Adverb	NP

Subordinating Conjunction (Subor)

English Patterns				
CoorCon	OBJECT 1	VERB PHRASE	MODIFIER	OBJECT 2
	NP	VP	Adverb	NP

APPENDIX J
Verb Phrase level FtoC patterns

Thai Patterns - VP			English Patterns - VP	
VP	ADJUNCTION		VP	(Indicate Tense)
verb	Auxiliary		verb	
VP	OBLIQUE		VP	OBLIQUE
verb	(PP)		verb	(PP)
VP	ADJUNCTION		VP	(Indicate Tense)
verb	Auxiliary		verb	
VP	MODIFIER		VP	MODIFIER
verb	Adverb		verb	Adverb
VP	MODIFIER	ADJUNCTION	VP	MODIFIER
verb	Adverb	Auxiliary	(verb)	Adverb
VP	ADJUNCTION 1	ADJUNCTION 2	VP	
verb	Auxiliary	Auxiliary	verb	
VP	ADJUNCTION 1	ADJUNCTION 2	VP	
verb	Auxiliary	Auxiliary	verb	
VP	ADJUNCTION 1	ADJUNCTION 2	VP	
verb	Auxiliary	Auxiliary	verb	

Thai Patterns - VP

VP	ADJUNCTION	MODIFIER
verb	Auxiliary	Adverb

VP	OBJECT 1	OBJECT 2
verb	Noun	Pronoun

VP	NEGATOR	OBJECT
verb		Noun

VP	OBJECT	OBLI PH
verb	Noun	(PP)

VP	OBJECT	MODIFIER
(verb)	Noun	Adverb

VP	ADJUNCTION	OBJECT
verb	Auxiliary	NP

VP	ADJUNCTION	NEGATOR	MODIFIER
verb	Auxiliary		Adverb

VP	ADJUNCTION	OBJECT	MODIFIER
verb	Auxiliary	Noun	Adverb

VP	ADJUNCTION	OBJECT	OBLIQUE PHR
verb	Auxiliary	Noun	(PP)

English Patterns - VP

VP	MODIFIER
verb	Adverb

OBJECT 2	Auxiliary Verb	OBJECT 1	VP
Pronoun		Noun	verb

Auxiliary Verb	NEGATOR	VP	OBJECT
		verb	Noun

VP	OBJECT	OBLI PH
verb	Noun	(PP)

VP	MODIFIER	OBJECT
verb	Adverb	Noun

VP	OBJECT
verb	NP

NEGATOR	VP	MODIFIER
	verb	Adverb

VP	OBJECT	MODIFIER
verb	Noun	Adverb

VP	OBJECT	OBLI PH
verb	Noun	(PP)

Thai Patterns - VP

VP	ADJUNCTION 1	OBJECT	ADJUNCTION 2
verb	Auxiliary	Noun	Auxiliary

VP	ADJUNCTION 1	ADJUNCTION 2	ADJUNCTION 3
verb	Auxiliary	Auxiliary	Auxiliary

VP	ADJUNCTION	MODIFIER 1	MODIFIER 2
verb	Auxiliary	Adverb	Adverb

VP	ADJUNCTION 1	OBJECT	ADJUNCTION 2	MODIFIER
verb	Auxiliary	Noun	Auxiliary	Adverb

VP	ADJUNCTION 1	OBJECT	MODIFIER	ADJUNCTION 2
verb	Auxiliary	Noun	Adverb	Auxiliary

VP	ADJUNCTION 1	ADJUNCTION 2	OBJECT	MODIFIER
verb	Auxiliary	Auxiliary	Noun	Adverb

VP	ADJUNCTION 1	ADJUNCTION 2	OBJECT	ADJUNCTION 3
verb	Auxiliary	Auxiliary	Noun	Auxiliary

English Patterns - VP

VP	OBJECT
verb	Noun

VP
verb

VP	MODIFIER 1	MODIFIER 2
verb	Adverb	Adverb

VP	MODIFIER	OBJECT
verb	Adverb	Noun

VP	OBJECT	MODIFIER
verb	Noun	Adverb

VP	OBJECT	MODIFIER
verb	Noun	Adverb

VP	OBJECT
verb	Noun

Thai Patterns - VP

VP	ADJUNCTION 1	ADJUNCTION 2	MODIFIER	ADJUNCTION 3
verb	Auxiliary	Auxiliary	Adverb	Auxiliary

VP	NEGATOR	OBJECT 1	OBJECT 2	OBJECT 3	OBJECT 4
verb		Noun	Noun	Noun	Noun

VP	NEGATOR	ADJUNCTION 1	OBJECT	ADJUNCTION 2	MODIFIER
verb		Auxiliary	Noun	Auxiliary	Adverb

VP	ADJUNCTION 1	ADJUNCTION 2	OBJECT 1	MODIFIER	OBJECT 2
verb	Auxiliary	Auxiliary	Noun	Adverb	Pronoun

English Patterns - VP

VP	MODIFIER
verb	Adverb

NEGATOR	VP	OBJECT 4	OBJECT 3	OBJECT 2	OBJECT 1
	verb	Noun	Noun	Noun	Noun

Auxiliary Verb	NEGATOR	MODIFIER	VP	OBJECT
		Adverb	verb	Noun

VP	OBJECT 1	MODIFIER	OBJECT 2
verb	Noun	Adverb	Pronoun

APPENDIX K

PREPOSITION PHRASE LEVEL FtoC PATTERNS

Thai Patterns

PP	OBJECT
	NP

PP	OBJECT 1	OBJECT 2
	NP	NP

English Patterns

PP	OBJECT
	NP

PP	OBJECT 2	OBJECT 1
	NP	NP

APPENDIX L

ADDITIONAL TESTING SENTENCES

Thai Sentences	English Translation Sentences
เพื่อ แสดง โครงการ.	so that show a project
ในกรณีที่ แสดง ภาพตัดขวาง.	in case of show a cross cut picture
แสดงออก ทาง ออปโตอิเล็กทรอนิกส์.	display on an optoelectronic
ทำให้ คอมพิวเตอร์ ล่าช้า ไป.	so a computer is delayed
ซึ่ง ได้รับ ทาง ร่างกาย.	that receive on a body
โครงการวิจัยและพัฒนา อุปกรณ์ขยายสัญญาณ และ มอเตอร์กระแสสลับ.	signal amplifier device and AC motor Research Project and Development
เพื่อ พิจารณา เครือข่ายคอมพิวเตอร์ และ เครื่องควบคุม.	so that consider computer network and controller machine
จะ เสีย งบประมาณ ดังนี้.	will lose as follows a budget
สรุป งาน ตาม นโยบาย.	summarize the job according to a policy
แล้ว จึง ส่ง ทาง คอมพิวเตอร์.	then send on a computer
รูป ที่1 ติดตั้ง ที่ ประเทศญี่ปุ่น.	the first figure is installed at Japan
และ จะ แปลง นโยบาย ตาม กรรมการ.	and will convert the policy according to a committee
แต่ละ Case มี รูปแบบ อย่างไร.	how does each Case have pattern ?
1. หน่วยความจำ 8 kbyte.	1. memory 8 kbyte
ออปโตอิเล็กทรอนิกส์ แสดง เอาท์พุท จำนวน 12 ชั้น.	an optoelectronic displays 12 floor output number
8042 จะ แทน electrode และ DRAM.	8042 will replace electrode and DRAM
1. กำหนดค่า มาตรฐาน รังสีเอ็กซ์ ตาม ระบบควบคุม.	first determines an X ray standard according to a controller system
รหัส นี้ จะ มี ลักษณะ ลำดับ.	this code will have an order character
ส่วน รูปแบบ นั้น จะ ถูก แทนที่ ตาม วงจรรวม และ อุปกรณ์ขยายสัญญาณ.	that layout part will be replaced according to IC and signal amplifier device
แต่ละ จุด ปิดท้าย ด้วย node.	each point is ended with node
รูป ที่ 10 แสดง ตำแหน่ง ตัวโน๊ต.	the tenth figure show a note position
1. หา บล็อกไดอะแกรม จาก แต่ละ โหมด.	first find the block diagram from each mode
6 ศึกษา โครงสร้าง จาก ฐานข้อมูล.	6 study the structure from database
ลักษณะ โครงสร้าง ของ อุปกรณ์ตรวจรู้ จะ ถูก แทน	the structure appearance of sensor will be replaced as first signal

Thai Sentences	English Translation Sentences
เช่นเดียวกับ สัญญาณ 1	
2. สร้าง ระบบ ฐานข้อมูล อิเล็กทรอนิกส์ดิกชั่นนารี.	2. build an Electronic dictionary system database
คำศัพท์ ประเภท นี้ พิมพ์ ภาษาสากล.	this vocabulary type is printed International language
ฉบับร่าง จะ พิมพ์ ลวดลาย ให้.	an outline will print pattern
ไมโครคอมพิวเตอร์ ถูก สร้าง มา จาก คณะ ผู้วิจัย.	a microcomputer is created from researcher committee
ชั้น ของ electrode จะ อยู่ ใน เกท.	layer of electrode will be in a gate
ลักษณะ ของ คำสั่ง จะ เป็น เทคโนโลยี 640 kbyte.	appearance of instruction will be 640 kbyte technology
ผลงานวิจัย จะ ถูก สร้าง มา จาก ผู้วิจัย.	a research will be built from researcher
เครือข่ายคอมพิวเตอร์ เป็น ผัง โครงสร้าง ของ ไมโครคอมพิวเตอร์ .	a computer network is the structure chart of a microcomputer
แรงดัน อัตโนมัติ ภายใน ทรานซิสเตอร์.	the automatic pressure transistor internal
ระบบสนับสนุนการแปลภาษา จึง จะ ทำงาน ต่างกัน ได้.	a supporting and translation system will work differently
คณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ จะ แสดง ใบเสร็จรับเงิน.	an Office of the National Economics and Social Development Board will display receipt
หน่วยความจำหลัก มี องค์ประกอบ ของ สัญญาณ วงจรนับ.	a main memory have the infrastructure of counter circuit signal
เวลา ของ ซีพียู จะ เป็น ชนิด มาตรฐาน.	time of CPU will be a standard type
คณะ กรรมการ นโยบาย จะ เป็น เช่นเดียวกับ กลุ่ม แรก.	policy committee group will be as the first group
ขั้น ที่สอง แสดง คำสั่ง ตัวตั้งเวลา ของ ตัวจ่ายไฟเลี้ยง.	the second step show the setting time instruction of power service
1. ค่าใช้จ่าย และ งบประมาณ ประกอบ ด้วย ตำแหน่ง และ บุคลากร.	1 expense and budget comprise with position and personnel
ข้อมูล คำบุพบท มี ต่างกัน ดังนี้.	Preposition data have differently as follows
ตัวควบคุม บรรทัด นั้น จะ แบ่ง จาก ประโยค ภาษาสากล.	that line controller will be divided from International language sentence
ลักษณะ ของ กระแส ผ่าน เข้า เครื่องใช้ไฟฟ้า.	appearance of current flow electric equipment
กลุ่ม แรก แสดง ชุด ข้อมูล ของ ตัวแบ่งกำลังสัญญาณ.	the first group show the data set of signal divider
คำเชื่อม นี้ ข้อมูล ได้ ดำเนินการ ต่อไป ตามลำดับ.	this data conjunction performed next respectively
การแพทย์ ได้ นำเข้า ชิพ ดังกล่าว จาก ประเทศญี่ปุ่น.	a medicine imported mention chip from Japan
1. ประชาชน ใช้ เตาเผา แบบ ตัวตั้งเวลา.	1 citizen use a setting time kiln type
ภาพ จะ แทน ตาราง และ ข้อมูล.	a picture will replace table and data
กรรมการบริหาร จะ ต้อง มี ข้อมูล ชัดเจน.	an administrative will have data clearly
รูป 1 แสดง ระบบ ของ คำสั่ง และ คุณสมบัติ เครื่องใช้ไฟฟ้า.	the first figure display the system of electric equipment instruction and property

BIOGRAPHY

NAME	Mr. Tawee Chimsuk
ACADEMIC BACKGROUND	1994 B.S. (Computer Science, Science) Ramkhamhaeng University 1998 M.S. (Computer Science) The National Institute of Development Administration (NIDA)
PRESENT POSITION	1. 1998-2001: Programmer at Sattahip Naval base Chonburi Province 2. 2001-Present: Lecturer at Royal Thai Naval Academy Samutprakarn Province
EXPERIENCES	1. Tawee Chimsuk and Surapong Auwatanamongkol. 2009. A Thai to English Machine Translation System using Thai LFG tree structure as Interlingua**.** In Proceedings International Conference of World Academy of Science, Engineering and technology, 60. 2. Tawee Chimsuk and Surapong Auwatanamongkol. 2010. An Incremental framework for a Thai-English Machine Translation System using a LFG tree structure as an Interlingua. In International Journal on Computer Science and Engineering (IJCSE) Vol. 02, No. 02, 2010, 280-288.