



แนวทางสถิติสำหรับการรู้จำค่านามประสมและนามวลีที่มีโครงสร้างคล้าย  
ประโยคในภาษาไทยโดยอัตโนมัติ โดยใช้แบบจำลอง

คอนดิชันนอลแรนดอมฟิลด์

โดย

นายภูริเดช โพธิ์พิพัฒน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2557

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

แนวทางสถิติสำหรับการรู้จำคำนามประสมและนามวลีที่มีโครงสร้างคล้าย  
ประโยคในภาษาไทยโดยอัตโนมัติ โดยใช้แบบจำลอง  
คอนดิชันนอลแรนดอมฟิลด์

โดย

นายภูริเดช โพธิ์พิพัฒน์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์  
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์  
ปีการศึกษา 2557  
ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์



A STATISTICAL APPROACH TO AUTOMATIC RECOGNITION OF  
SENTENCE-LIKE COMPOUND NOUNS AND NOUN PHRASES IN THAI  
USING CONDITIONAL RANDOM FIELD MODELS

BY

MR. PHURIDECH PHOIPHAT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF SCIENCE  
COMPUTER SCIENCE  
FACULTY OF SCIENCE AND TECHNOLOGY  
THAMMASAT UNIVERSITY  
ACADEMIC YEAR 2014  
COPYRIGHT OF THAMMASAT UNIVERSITY

มหาวิทยาลัยธรรมศาสตร์  
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นายภูริเดช โพธิ์พิพัฒน์

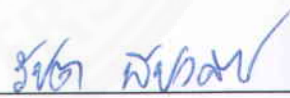
เรื่อง

แนวทางสถิติสำหรับการรู้จำคำนามประสมและนามวลีที่มีโครงสร้างคล้ายประโยคในภาษาไทยโดย  
อัตโนมัติ โดยใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิทยาศาสตรมหาบัณฑิต

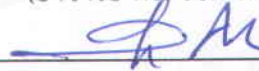
เมื่อ วันที่ 14 สิงหาคม พ.ศ. 2558

ประธานกรรมการสอบวิทยานิพนธ์



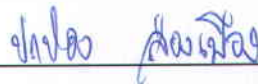
(อาจารย์ ดร. รัชต พิษวณิชย์)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก



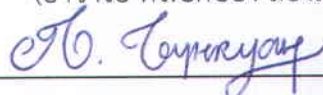
(ผู้ช่วยศาสตราจารย์ ดร. รัชฎา คงคะจันทร์)

กรรมการสอบวิทยานิพนธ์



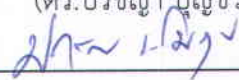
(อาจารย์ ดร. ปกป้อง ส่องเมือง)

กรรมการสอบวิทยานิพนธ์



(ดร. ปรีชญ่า บุญขวัญ)

คณบดี



(รองศาสตราจารย์ ปกรณ์ เสริมสุข)

หัวข้อวิทยานิพนธ์	แนวทางสถิติสำหรับการรู้จำคำนามประสมและนามวลีที่มีโครงสร้างคล้ายประโยคในภาษาไทยโดยอัตโนมัติ โดยใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์
ชื่อผู้เขียน	นายภูริเดช โพธิ์พิพัฒน์
ชื่อปริญญา	วิทยาศาสตร์มหาบัณฑิต
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร. รัชฎา คงคะจันทร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม (ถ้ามี)	-
ปีการศึกษา	2557

### บทคัดย่อ

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์เพื่อพัฒนาระบบการรู้จำคำนามประสมชนิดแสดงไวยากรณ์ประโยคโดยใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์สโม่เดล (CRFs) ซึ่งศึกษาประสิทธิภาพของระบบการรู้จำคำนามประสมชนิดแสดงไวยากรณ์ประโยค ที่พิจารณาโครงสร้างประโยค และคำนามหลัก รวมทั้งสิ้น 8 รูปแบบ งานวิจัยนี้ใช้คลังข้อมูล BEST เฉพาะหมวดข่าวภาษาไทยขนาด 1,000 คำ แบบจำลองที่ใช้คือ CRFSharp จากผลการทดสอบระบบ พบว่าประสิทธิภาพของระบบที่การพิจารณาโดยรวมภาพรวมทั้งหมด กฎโครงสร้างประโยค “นาม-กริยา” (ขอบเขต 2 คำ) มีความถูกต้อง (F-measure) มากที่สุด คือร้อยละ 93.06 รองลงมาคือ กฎโครงสร้างประโยค ขอบเขต 4 คำ คือร้อยละ 89.19 ตามมาด้วยการพิจารณากฎในรูปแบบอื่นๆ มีค่าประสิทธิภาพที่ใกล้เคียงกันดังนี้ คำนามหลักเกี่ยวกับเครื่องมือ คำนามหลักเกี่ยวกับอาชีพ คำนามหลักเกี่ยวกับการแพทย์ “นาม-กริยา-นาม”(ขอบเขต 3 คำ) และคำนามหลักเกี่ยวกับสถานที่ มีค่าร้อยละความถูกต้องเท่ากับ 81.10, 80.41, 80.08, 78.18 และ 76.58 ตามลำดับ สำหรับคำนามหลักเกี่ยวกับวัน เวลา มีค่าประสิทธิภาพของระบบน้อยที่สุด เท่ากับร้อยละ 68.01

**คำสำคัญ:** คำนามประสมชนิดแสดงไวยากรณ์ประโยค, คอนดิชันนอลแรนดอมฟิลด์

Thesis Title	A Statistical Approach to Automatic Recognition of Sentence-Like Compound Nouns and Noun Phrases in Thai using Conditional Random Field Models
Author	Mr. Phuridech Phopiphat
Degree	The degree of master of science
Major Field/Faculty/University	Computer science Faculty of science and technology Thammasat University
Thesis Advisor	Asst. Prof. Dr. Rachada Kongkachandra
Thesis Co-Advisor (If any)	-
Academic Years	2014

### ABSTRACT

The main purpose of this study is to develop synthetic nominal compounds recognition system using Conditional Random Fields Models (CRFs). This study uses the news corpus of 1,000 words and CRFs model applied in this research is CRF sharp. There were 8 patterns of structure and main noun word that were considered in this model. The results show that the highest recognition rates (F-measure) for pattern of 2 boundary words (noun-verb) was 93.06%. The second score of F-measure for 4 boundary words was 89.19%. Followed by, main noun word about tool, main noun word about career, main noun word about medical, 3 boundary word and main noun word about place were 81.10%, 80.41%, 80.08%, 78.18% and 76.58% respectively. The lowest score of F-measure for main noun word about time was 68.01%.

**Keywords:** Conditional Random Fields Models, synthetic nominal compounds

## กิตติกรรมประกาศ

ในการทำโครงการวิจัยนี้บรรลุวัตถุประสงค์สำเร็จลุล่วงได้ด้วยความกรุณาและความช่วยเหลือเป็นอย่างดีจากอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.รัชฎา คงคะจันทร์ คณะกรรมการ อาจารย์ ดร. รัชต พิษวณิชย์ อาจารย์ ดร.ปกป้อง ส่องเมือง และดร.ปรัชญา บุญขวัญ ที่กรุณาเสียสละเวลาให้คำแนะนำ ตรวจสอบแก้ไขข้อบกพร่อง เพื่อให้งานวิจัยมีความถูกต้องครบถ้วน

นอกจากนี้ผู้วิจัยขอขอบพระคุณ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ที่อนุเคราะห์ฐานข้อมูล คลังข้อความภาษาไทย (BEST) ซึ่งเป็นส่วนประกอบหลักที่จำเป็นอย่างยิ่งสำหรับงานวิจัยนี้

นายภูริเดช โพธิ์พิพัฒน์

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย	(1)
บทคัดย่อภาษาอังกฤษ	(2)
กิตติกรรมประกาศ	(3)
สารบัญตาราง	(7)
สารบัญภาพ	(8)
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการทำวิจัย	2
1.3 ขอบเขตของการทำวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
1.5 นิยามศัพท์เฉพาะ	3
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง	4
2.1 คำนิยามของคำประสม	4



2.1.1 โครงสร้างของค่านามประสมชนิดแสดง ไวยากรณ์ประโยค	5
2.1.2. ความสัมพันธ์ทางไวยากรณ์ของค่านาม ประสมชนิดแสดงไวยากรณ์ประโยค	5
2.1.2.1 ความสัมพันธ์ทางไวยากรณ์ของค่านาม ประสมแบบหน่วยหลักนาม	6
2.1.2.2 ความสัมพันธ์ทางไวยากรณ์ของค่านาม ประสมแบบหน่วยหลักกริยา	8
2.1.3 อรรถสัมพันธ์ของค่านามประสมชนิดแสดง ไวยากรณ์ประโยค	9
2.1.3.1 อรรถสัมพันธ์ของค่านามประสมแบบ ประธาน-กริยา-(กรรม)	10
2.1.3.2 อรรถสัมพันธ์ของค่านามประสมแบบ กรรม-กริยา	12
2.1.3.3 อรรถสัมพันธ์ของค่านามประสมแบบ หน่วยเสริม-กริยา	12
2.1.3.4 อรรถสัมพันธ์ของค่านามประสมแบบ กริยา-ประธาน	13
2.2 แนวทางในการพิจารณาความแตกต่างระหว่าง ค่านามประสมและวลี/ประโยค	14
2.2.1 เกณฑ์ทางอรรถศาสตร์	14
2.2.2 เกณฑ์ทางวិอรรถศาสตร์	15
2.2.3 เกณฑ์ทางวากยสัมพันธ์	15
2.2.4 เกณฑ์ทางข้อจำกัดของคำสมาชิก	16
2.2.5 เกณฑ์ทางคุณสมบัติหน่วยคำ	16

2.3 แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์	17
บทที่ 3 วิธีการวิจัย	19
3.1 ขอบเขตการดำเนินงานวิจัย	19
3.1.1 ข้อมูลที่ใช้ในการทำวิจัย	19
3.1.2 เครื่องมือที่ใช้ในงานวิจัย	20
3.2 กระบวนการวิจัย	21
3.2.1 กระบวนการก่อนประมวลผล	21
3.2.2 กระบวนการประมวลผล	22
3.3 ขั้นตอนการทดลอง	24
3.4 การประเมินผลการทดลอง	25
บทที่ 4 ผลการวิจัยและอภิปรายผล	26
4.1 การตั้งค่าการทดลอง	26
4.2 ผลการทดลอง	28
4.3 อภิปรายผลการทดลอง	30
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	31
5.1 สรุปผลงานวิจัย	31
5.2 ข้อเสนอแนะ	31
รายการอ้างอิง	33
ประวัติผู้เขียน	34

## สารบัญตาราง

ตารางที่	หน้า
3.1 ตัวอย่างคลังข้อมูล BEST	18
3.2 เครื่องมือที่ใช้ในการดำเนินงานวิจัย	19
3.3 ตัวอย่างบทความ	20
3.4 ตัวอย่างบทความที่มีการแจกแจงชนิดของคำ	21
3.5 ตัวอย่างคำนามหลักในคำประสมที่มีโครงสร้าง	
คล้ายประโยค	22
3.6 ตัวอย่างไฟล์สำหรับฝึกฝนที่ระบุประเภทของ	
คำนามหลัก	22
4.1 ตัวอย่างบทความข่าวสำหรับการทดลอง	25
4.2 ตัวอย่างบทความใช้กฎในการรู้จำตามคำนามหลัก	26
4.3 ตัวอย่างผลการทดลองการระบุคำนามประสมชนิด	
แสดงไวยากรณ์ประโยคอาชีพ	27
4.4 ตัวอย่างผลการทดลองระบุคำนามหลัก	
เกี่ยวกับการแพทย์	27
4.5 ตัวอย่างผลการทดลองระบุคำนามหลัก	
เกี่ยวกับเครื่องมือ	27
4.6 ตัวอย่างผลการทดลองระบุคำนามหลัก	28
เกี่ยวกับสถานที่	
4.7 ตัวอย่างผลการทดลองระบุคำนามหลัก	
เกี่ยวกับวัน เวลา	28
4.8 ตารางแสดงผลการทดลอง 8 แบบ	28

## สารบัญภาพ

ภาพที่	หน้า
1.1 แผนภูมิแสดงโครงสร้าง วากยสัมพันธ์ และอรรถสัมพันธ์ของหน่วยสมาชิก	13
2.1 รูปกราฟแสดงการเปรียบเทียบแบบจำลอง HMMs, MEMMs และ CRFs	17
3.1 กระบวนการวิจัย	20



# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

กระบวนการประสมคำ เป็นวิธีสำคัญในการสร้างคำใหม่ที่สำคัญในภาษาไทย ซึ่งทั่วไปเมื่อกล่าวถึง คำประสม (compound word) โดย (อัญชลี 2548) ได้ศึกษาเฉพาะเรื่องการสร้าง คำนามประสม ในหนังสือ คำนามประสม ศาสตร์และศิลป์ในการสร้างคำไทย ได้นิยามการประสมคำ ไว้ว่า เป็นคำที่สร้างจากการนำเอาคำเดิมที่มีอยู่แล้วในภาษาสองคำหรือมากกว่ามาประกอบเข้าด้วยกัน เพื่อสร้างคำใหม่ที่มีความหมายเฉพาะคำ คำประสมมีความแตกต่างจากประโยค เนื่องจากคำประสมมีความหมายที่จำเพาะเสมือนเป็นคำเดี่ยว (idiosyncratic meanings) ไม่ใช่เป็นผลรวมของความหมายจากหลายคำที่นำมาเรียงกัน (straight forward/compositional meaning) ซึ่งผู้เขียนได้มีการให้ข้อสังเกตว่า ในอดีตจะยึดถือเอาคำจำกัดความดังกล่าวเป็นเกณฑ์ที่ตายตัว หากไม่อยู่ในเกณฑ์ลักษณะดังกล่าวจะไม่ถือว่าเป็นคำประสม แต่ภายหลังได้มีการศึกษาของนักภาษาศาสตร์รุ่นหลัง (Lieber 1992) และ (Sadock 1998) ที่พบว่า โครงสร้างภายในคำประสมมีความหลากหลาย ซึ่งมีลักษณะทั่วไปเหมือนโครงสร้างไวยากรณ์เช่นวลีและประโยค จากสาเหตุดังกล่าวอาจจะทำให้เกิดปัญหาความกำกวมในการตีความหมายได้มากกว่าหนึ่งอย่างได้ โดยเฉพาะอย่างยิ่งในภาษาไทย ซึ่งเป็นภาษาที่มีลักษณะเฉพาะ ที่มีการเขียนคำที่ติดกันไม่เหมือนกับในภาษาอื่นๆ วลีหรือประโยคต่างๆ จะเขียนติดต่อกัน โดยจะมีการเว้นระยะก็ต่อเมื่อมีการขึ้นข้อความใหม่

ในปัจจุบันมีความพยายามในการศึกษาวิจัยเพื่อจำแนกแยกแยะคำนามประสมและไวยากรณ์วลีหรือประโยคโดยใช้การรู้จำตัวอักษร (word recognition) ซึ่งเป็นกระบวนการวิจัยสาขาหนึ่งของการประมวลผลภาษาธรรมชาติ (Natural Language Processing ย่อว่า NLP) ที่เน้นศึกษาปัญหาในการประมวลผลและใช้งานภาษาธรรมชาติ รวมทั้งการทำความเข้าใจภาษาธรรมชาติ เพื่อให้คอมพิวเตอร์สามารถเข้าใจภาษามนุษย์ โดยในกระบวนการรู้จำตัวอักษรจะใช้โมเดลทางคณิตศาสตร์มาใช้ในการศึกษาโมเดลที่ถูกนำมาใช้มีความหลากหลายและมีข้อดี ข้อเสียที่แตกต่างกัน ดังนั้นนักวิศวกรรมคอมพิวเตอร์ นักวิทยาศาสตร์คอมพิวเตอร์ และนักภาษาศาสตร์คอมพิวเตอร์ จึงศึกษาหาแนวทางและวิธีการที่มีประสิทธิภาพมากที่สุดในการวิเคราะห์คำประสม ในงานวิจัยนี้ มีความสนใจที่จะศึกษาและพัฒนาระบบวิธีการช่วยแบ่งคำของคำประสม โดยใช้โมเดลทางสถิติคอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Fields : CRFs) ในการรู้จำชื่อ โดยมีการศึกษาในต่างประเทศของ Lafferty et al. (2001) พบว่า CRFs มีข้อดีกว่าเมื่อเปรียบเทียบกับ (Hidden Markov models:

HMMs) และ (Maximum entropy Markov models: MEMMs) นอกจากนี้งานวิจัยของ Sha and Pereira (2003) ยังพบว่า ผลที่ได้จาก CRFs ดีที่สุดในการวิเคราะห์ เมื่อเปรียบเทียบกับ CoNLL task เช่นเดียวกัน

ที่ผ่านมาในประเทศไทย มีงานวิจัยที่เกี่ยวข้องกับการประมวลผลภาษาธรรมชาติของภาษาไทย ซึ่งได้มีการนำโมเดล CRFs ไปใช้ในการวิเคราะห์หาหน่วยคำไทย เพื่อนำไปใช้ในการตัดคำ และการกำกับชนิดของคำ Kruengkrai et al. (2006) โดยเมื่อศึกษาเปรียบเทียบกับวิธีการอื่น ได้แก่ Longest Matching และ Maximum Matching พบว่า ผลที่ได้จาก CRFs ดีที่สุด และงานวิจัย Haruechaiyasak, Kongyoung, and Dailey (2008) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพของระบบการตัดคำจากแบบจำลอง 4 แบบ ได้แก่ Naive Bayes, decision tree, Support Vector Machine, และ CRFs ซึ่งผลที่ออกมาปรากฏว่า CRFs ใช้กับการตัดคำภาษาไทยได้ดีกว่าแบบจำลองอื่น

## 1.2 วัตถุประสงค์ของการทำวิจัย

ศึกษาแนวทางการรู้จำคำประสมชนิดแสดงไวยากรณ์ประโยค (synthetic nominal compounds recognition) โดยใช้โมเดลทางสถิติคอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Fields : CRFs)

## 1.3 ขอบเขตของการทำวิจัย

1. คลังข้อความภาษาไทย (BEST) ที่มีการกำหนดขอบเขตของคำโดยคัดเลือกในหัวข้อข่าวภาษาไทย
2. ค้นหกริยาในโครงสร้างของคำประสมชนิดแสดงไวยากรณ์ประโยค (JJV) ในคลังข้อความภาษาไทย (BEST)
3. ใช้การเรียนรู้ของเครื่อง (Machine Learning) เทคนิค Conditional Random Fields (CRFs) เพื่อสร้างแบบจำลองสำหรับการรู้จำคำประสมชนิดแสดงไวยากรณ์ประโยค
4. การวัดผลประเมินจากรู้จำคำประสมชนิดแสดงไวยากรณ์ประโยค จากค่า ความถูกต้อง (Precision) ค่าความระลึก (Recall) และค่าประสิทธิภาพโดยรวมของระบบ (F-measure) โดยนักภาษาศาสตร์

#### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

สามารถเข้าใจหลักการทำงานของเทคนิค คอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Fields : CRFs) ในการรู้จำคำประสมชนิดแสดงไวยากรณ์ประโยค

#### 1.5 นิยามศัพท์เฉพาะ

วากยสัมพันธ์ เป็นการศึกษาว่าด้วยกฎของความสัมพันธ์ของแบบแผนองค์ประกอบของประโยคในภาษา อันเป็นส่วนหนึ่งของหลักไวยากรณ์ของภาษา เช่น ในภาษาไทย ประธาน-กริยา-กรรม เรียงคำเป็น ฉันทัน-กิ่น-ข้าว เป็นต้น

อรรถสัมพันธ์ เป็นการศึกษาที่ว่าด้วยการพิจารณาบทบาทหรือหน้าที่ของวากยสัมพันธ์ ยกตัวอย่างเช่น ผู้กระทำ ผู้ประสบ ผู้ถูกกระทำ หรือ เครื่องมือ เป็นต้น

วិอรรถสัมพันธ์ เป็นการศึกษาความหมายจากบริบท ความหมายตามเจตนาของผู้สื่อสารความหมายของถ้อยคำตามความสัมพันธ์ระหว่างบุคคล เป็นต้น

## บทที่ 2

### วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ได้พัฒนาระบบการรู้จำคำประสม โดยใช้คอนดิชันนอลแรนดอมฟิลด์โมเดล โดยมีคำนิยาม แนวคิด ทฤษฎี และงานวิจัย ที่เกี่ยวข้อง ดังนี้

- 2.1 คำนิยามของคำประสม
- 2.2 แนวทางในการพิจารณาความแตกต่างระหว่างคำนามประสมและวลี/ประโยค
- 2.3 แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์

#### 2.1 คำนิยามของคำประสม

รศ.ดร. อัญชลี สิงห์น้อย ได้กล่าวถึงคำนามประสม ในหนังสือ คำนามประสมศาสตร์ และศิลป์ในการสร้างคำไทย ปี 2548 โดยสามารถสรุปได้ดังนี้

คำนามประสมในภาษาไทยสามารถจำแนกได้เป็นสองชนิดใหญ่ๆ คือคำนามประสมชนิดแสดงไวยากรณ์ประโยค (synthetic nominal compounds) และคำนามประสมชนิดนามล้วน (noun-noun compounds) ซึ่งมีความแตกต่างกัน ไม่ว่าจะเป็นเรื่องของโครงสร้าง วากยสัมพันธ์ อรรถสัมพันธ์ และวือรรถสัมพันธ์ ซึ่งต่างก็มีปรากฏเป็นจำนวนมากมายนับไม่ถ้วน

คำนามประสมชนิดแรกคือ คำนามประสมชนิดแสดงไวยากรณ์ประโยคในคำนามประสมชนิดนี้ คำที่มาประกอบกัน อย่างน้อยต้องเป็นคำกริยาและคำนาม ซึ่งคำกริยาและคำนามนั้นมีความสัมพันธ์ทางไวยากรณ์ต่อกันเช่นเดียวกับเมื่อเกิดเป็นหน่วยประกอบของประโยคเมื่อพิจารณา วากยสัมพันธ์ และอรรถสัมพันธ์ เช่นประธานหรือผู้กระทำกริยา เป็นกรรมหรือผู้ถูกกระทำกริยา ฯลฯ ยกตัวอย่างคำนามประสมชนิดนี้ เช่น คนขับรถ ซึ่งเป็นที่รู้จักกันดีว่าเป็นคำประสมที่หมายถึง บุคคลที่มีอาชีพขับรถ คำนามชนิดนี้มีความหมายค่อนข้างตรงไปตรงมา (transparent) คล้ายประโยค แต่การนำคำว่า คน ไปประสมกับกริยาอื่น ก็สามารถสร้างคำประสมได้อีกมากมาย จึงทำให้เกิดปัญหาในการพิจารณาว่าโครงสร้างนั้นๆ เป็นคำนามประสมหรือประโยค ถึงแม้ว่าจะได้มีการนิยามไว้แล้วก็ตามว่า คำประสมมีความแตกต่างจากประโยค เนื่องจากคำประสมมีความหมายเฉพาะเสมือนคำเดี่ยว (idiosyncratic meanings) ไม่ใช่ผลรวมของความหมายจากหลายคำมาเรียงต่อกัน (straight forward/compositional meanings)



คำนามประสมชนิดแสดงไวยากรณ์ประโยค ซึ่งอาจจำแนกเป็นกลุ่มย่อย โดยพิจารณาจากโครงสร้าง วากยสัมพันธ์ และอรรถสัมพันธ์ ดังต่อไปนี้

### 2.1.1 โครงสร้างของคำนามประสมชนิดแสดงไวยากรณ์ประโยค

โครงสร้างไวยากรณ์ใดๆ ก็ตามไม่ว่าจะเป็นคำ วลี หรือประโยคจะต้องประกอบด้วยอย่างน้อย คือหน่วยหลักและอาจจะมีหน่วยขยายด้วย คำนามประสมชนิดนี้มีโครงสร้างพื้นฐานที่ประกอบด้วยหน่วยหลักและหน่วยขยายเช่นเดียวกันกับหน่วยไวยากรณ์ในระดับนามวลี ซึ่งมีหน่วยหลักอยู่ทางซ้ายสุดและมีหน่วยขยายอยู่ถัดไปทางขวา โดยทั่วไปแล้วพบว่า หน่วยหลักของคำนามประสมชนิดนี้จะเป็นคำนามและมีหน่วยขยายเป็นคำกริยาตัวเดียวหรืออาจเป็นคำกริยาและคำนามอย่างน้อยหนึ่งตัวต่อถัดไป มีคำนามประสมเพียงจำนวนหนึ่งเท่านั้นที่มีหน่วยสมาชิกเป็นกริยาล้วน คำนามประสมชนิดนี้ปรากฏในหลายรูปแบบ ซึ่งสรุปได้เป็น 5 ประเภทต่อไปนี้

<p style="text-align: center;"><u>นาม – กริยา</u></p> <p>ไม้ – กวาด = ไม้กวาด</p> <p>หมู – สับ = หมูสับ</p>	<p style="text-align: center;"><u>นาม – กริยา – นาม</u></p> <p>เครื่อง – ซัก – ผ้า = เครื่องซักผ้า</p> <p>ส้ม – ตำ – ปู = ส้มตำปู</p>
<p style="text-align: center;"><u>นาม-กริยา-นาม-นาม</u></p> <p>ไก่-ผัด-พริก-ชิง = ไก่ผัดพริกชิง</p> <p>หมู-ทอด-กระเทียม-พริกไทย = หมูทอดกระเทียมพริกไทย</p>	<p style="text-align: center;"><u>กริยา – นาม</u></p> <p>รอง- เท้า =-รองเท้า</p> <p>พัด – ลม = พัดลม</p>
<p><u>กริยา – กริยา</u></p> <p>ต้ม – ยำ = ต้มยำ</p> <p>กั้น – สาด =-กั้นสาด</p>	

### 2.1.2. ความสัมพันธ์ทางไวยากรณ์ของคำนามประสมชนิดแสดงไวยากรณ์ประโยค

ความสัมพันธ์ทางไวยากรณ์ในประโยค เป็นความสัมพันธ์ที่หน่วยกริยามีต่อหน่วยนามต่างๆ ในลักษณะที่หน่วยนามนั้นๆ เป็นประธาน (subject) กรรม (objective) และหน่วยเสริม

(oblique) ของหน่วยกริยา คำนามประสมชนิดแสดงไวยากรณ์ประโยคก็แสดงความสัมพันธ์ทางไวยากรณ์ของหน่วยสมาชิกที่เป็นหน่วยหลักและหน่วยขยายในลักษณะที่คล้ายคลึงกันและมีหลายรูปแบบ ซึ่งจะกล่าวถึงสองโครงสร้างหลัก คือ คำนามประสมที่มีหน่วยหลักเป็นคำนาม และคำนามประสมที่มีหน่วยหลักเป็นคำกริยา ดังต่อไปนี้

### 2.1.2.1 ความสัมพันธ์ทางไวยากรณ์ของคำนามประสมแบบหน่วยหลักนาม

โครงสร้างคำนามประสมประเภทที่มีหน่วยหลักเป็นคำนาม คำนามนั้นสามารถทำหน้าที่เป็นได้ทั้งประธาน กรรม และหน่วยเสริม จึงทำให้สามารถจำแนกรูปแบบความสัมพันธ์ทางไวยากรณ์ในคำนามประสมชนิดนี้ออกได้เป็นอีก 4 รูปแบบ ประธาน-กริยา-(กรรม) ประธาน-กริยา-หน่วยเสริม กรรม-กริยา และหน่วยเสริม-กริยา-(กรรม) ดังแสดงตัวอย่างดังต่อไปนี้

<p style="text-align: center;"><u>ประธาน-กริยา-(กรรม)</u></p> <p>เรือ-บิน        -----    เรือบิน</p> <p>คน-ขับ-รถ    -----    คนขับรถ</p>	<p style="text-align: center;"><u>ประธาน-กริยา-หน่วยเสริม</u></p> <p>เรือ-เดิน-สมุทร    -----    เรือเดินสมุทร</p> <p>คน-เดิน-เท้า        -----    คนเดินเท้า</p>
<p style="text-align: center;"><u>กรรม-กริยา</u></p> <p>บ้าน-เช่า        -----    บ้านเช่า</p> <p>ผัก-ดอง        -----    ผักดอง</p>	<p style="text-align: center;"><u>หน่วยเสริม-กริยา-(กรรม)</u></p> <p>      ม้า-นั่ง        -----    ม้านั่ง</p> <p>      ห้อง-นอน      -----    ห้องนอน</p> <p>      ห้อง-รับ-แขก    -----    ห้องรับแขก</p>
<p style="text-align: center;"><u>หน่วยเสริม-กริยา-(กรรม)</u></p> <p>          ม้า-นั่ง        -----    ม้านั่ง</p> <p>          ห้อง-นอน      -----    ห้องนอน</p> <p>          ห้อง-รับ-แขก    -----    ห้องรับแขก</p>	

จากตัวอย่างข้างต้นได้ว่าความสัมพันธ์ทางไวยากรณ์ของคำนามประสมที่แสดงไปข้างต้นมีความสัมพันธ์เหมือนไวยากรณ์ของประโยค เพียงแต่ถ้าพิจารณาลักษณะวากยสัมพันธ์ด้าน

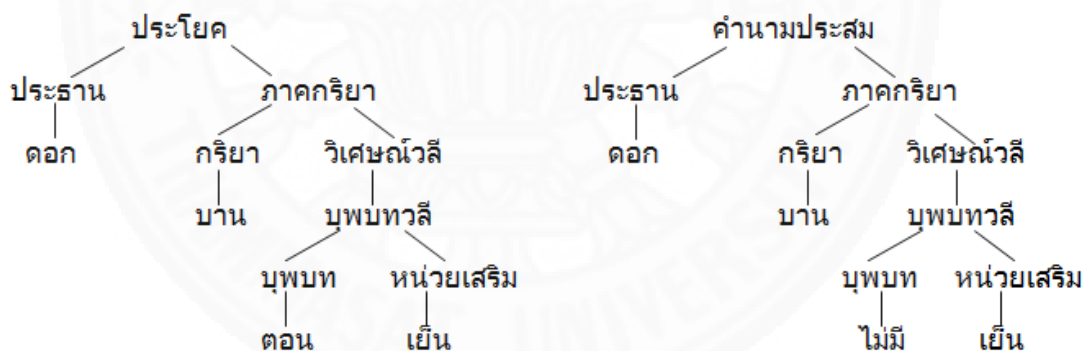
อื่นๆ เช่นลำดับการเรียงคำ (word order) และโครงสร้างภายในของหน่วยต่างๆในคำนามประสมชนิดนี้จะมีความแตกต่างไปบ้างเล็กน้อย ดังจะแสดงเปรียบเทียบให้เห็นต่อไปนี้

ประโยคและคำนามประสมแบบประธาน-กริยา-(กรรม)



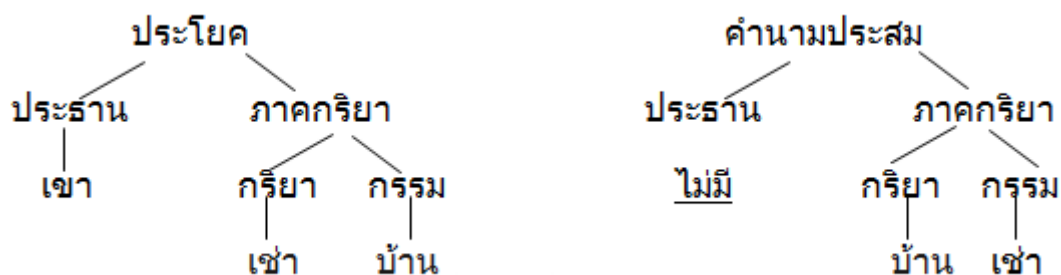
คำนามประสมแบบประธาน-กริยา-(กรรม) มีลำดับและโครงสร้างของหน่วยประธาน กริยา และกรรม เหมือนกับประโยคทุกประการ

ประโยคและคำนามประสมแบบประธาน-กริยา-หน่วยเสริม



คำนามประสมแบบประธาน-กริยา-หน่วยเสริม มีลำดับหน่วยประกอบเหมือนประโยค คือประธาน กริยา และหน่วยเสริมตามลำดับ ต่างกันเพียงแต่ในคำนามประสมไม่ปรากฏคำบุพพท

ประโยคและคำนามประสมแบบ (กรรม)-กริยา



คำนามประสมแบบกรรม-กริยา มีลำดับการเรียงคำต่างจากประโยคคือ กรรม-กริยา ในขณะที่ในประโยคเป็นกริยา-กรรม และในคำนามประสมไม่ปรากฏประธาน

ประโยคและคำนามประสมแบบหน่วยเสริม-กริยา-(กรรม)



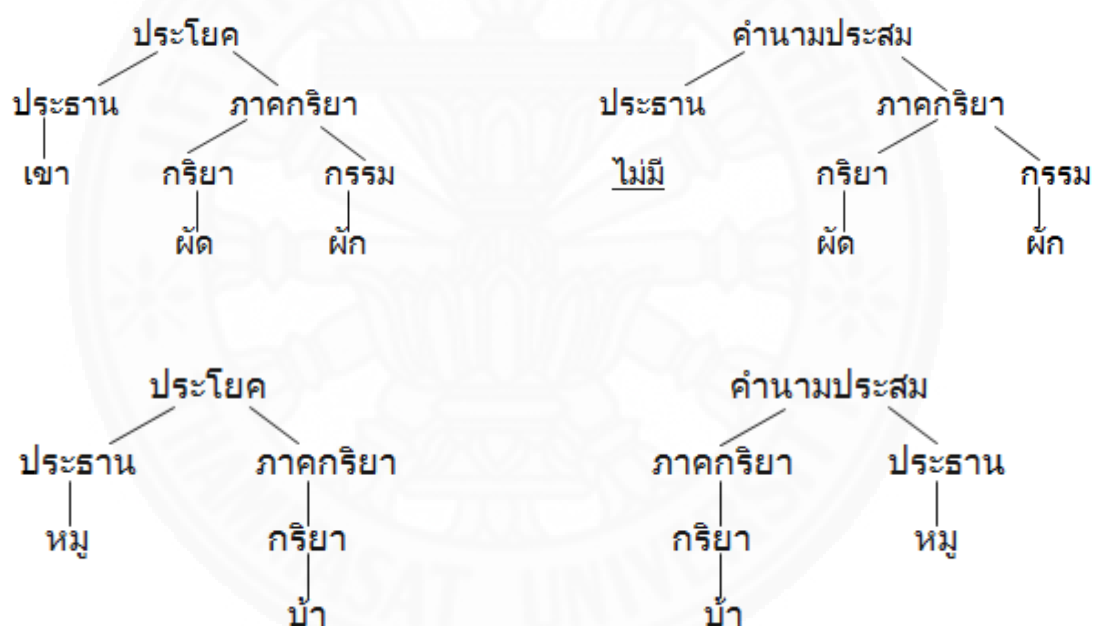
คำนามประสมแบบหน่วยเสริม-กริยา-(กรรม) มีลำดับการเรียงสลับกับประโยคและหน่วยเสริมของคำนามประสมไม่มีบุพบทนำในขณะที่หน่วยเสริมในประโยคโดยทั่วไปจะอยู่ภายใต้บุพบทวลี (มีบุพบทนำ)

### 2.1.2.2 ความสัมพันธ์ทางไวยากรณ์ของคำนามประสมแบบหน่วยหลักกริยา

คำนามประสมชนิดแสดงไวยากรณ์ประโยคที่มีหน่วยสมาชิกหลักเป็นกริยาพบว่า มีรูปแบบของความสัมพันธ์ทางไวยากรณ์ได้ 2 รูปแบบ คือ กริยา-กรรม และกริยา-ประธาน ตามตัวอย่างดังต่อไปนี้

กริยา-กรรม	กริยา-ประธาน
มัด-มัดก = มัดมัดก	มัด-มัดม = มัดมัดม
รอง-เท้า = รองเท้า	บ้า-หมู = บ้าหมู

เมื่อเปรียบเทียบกับประโยคจะเห็นได้ว่า คำนามประสมรูปแบบ กริยา-กรรม จะมีลำดับการเรียงคำของหน่วยประกอบที่เลียนแบบประโยคไร้ประธาน ในขณะที่คำนามประสมรูปแบบ กริยา-ประธาน ไม่เหมือนประโยคดังแสดงดังต่อไปนี้



### 2.1.3 อรรถสัมพันธ์ของคำนามประสมชนิดแสดงไวยากรณ์ประโยค

นอกจากคำประสมชนิดไวยากรณ์ประโยคจะมีความแตกต่างกันในเรื่องของโครงสร้างและความสัมพันธ์ทางไวยากรณ์ของหน่วยสมาชิกแล้ว ยังมีความแตกต่างในเรื่องของอรรถสัมพันธ์หรือความสัมพันธ์ทางความหมายอีกด้วย หน่วยนามในประโยคสามารถมีอรรถสัมพันธ์ต่อหน่วยกริยาได้หลายรูปแบบ เช่น เป็นผู้กระทำ ผู้ประสบ ผู้ถูกกระทำ เครื่องมือ ฯลฯ ของกริยานั้นๆ ซึ่งอาจจำแนกความแตกต่างของอรรถสัมพันธ์ระหว่างหน่วยนามกับหน่วยกริยา ดังต่อไปนี้

### 2.1.3.1 อรรถสัมพันธ์ของคำนามประสมแบบประธาน-กริยา-(กรรม)

มีหน่วยหลักประธานทำหน้าที่ทางความหมายที่แตกต่างกันได้หลายอย่าง เป็น ผู้กระทำ-กริยา-(กรรม) ผู้ประสม-กริยา-(กรรม) เครื่องมือ-กริยา-(กรรม) และผู้มีสภาพ-กริยา ดัง ตัวอย่างต่อไปนี้

#### ผู้กระทำ-กริยา-(กรรม)

คน-ขับ-รถ ----- คนขับรถ

ไก่-ชน ----- ไก่ชน

#### ผู้ประสม-กริยา-(กรรม)

คน-อ่าน ----- คนอ่าน

ผู้-ชม-รายการ ----- ผู้ชมรายการ

#### ผู้มีสภาพ-กริยา

คำ-ขาด ----- คำขาด

ปู-เค็ม ----- ปูเค็ม

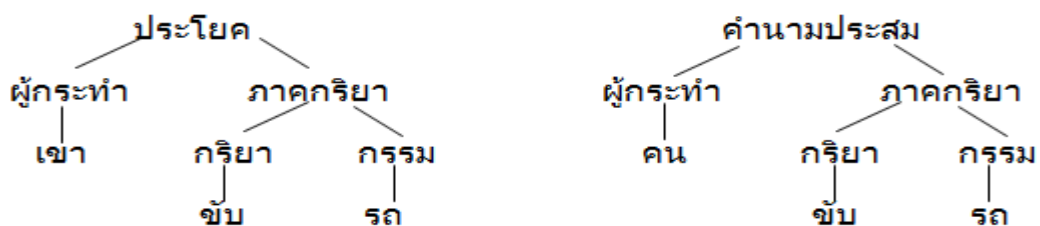
#### เครื่องมือ-กริยา-(กรรม)

ไม้-กวาด ----- ไม้กวาด

หม้อ-หุง-ข้าว ----- หม้อหุงข้าว

คำนามประสมแต่ละแบบดังกล่าวต่างมีการเกิดของหน่วยสมาชิกโดยเลียนแบบ ลำดับการเรียงคำในประโยคทั้งสิ้น ดังแสดงตามลำดับ

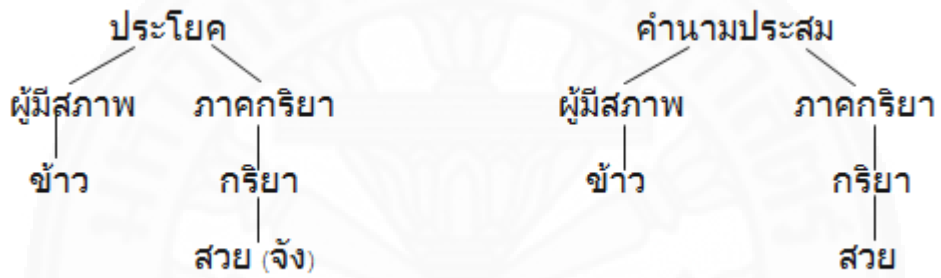
#### ผู้กระทำ-กริยา-(กรรม)



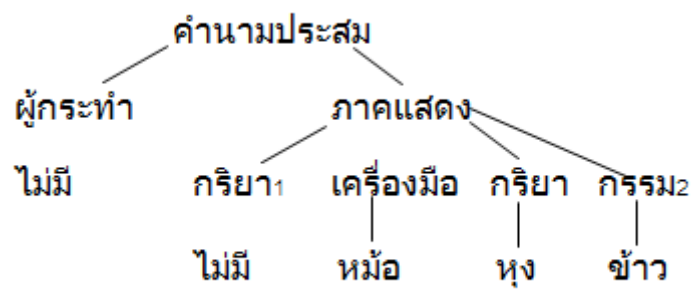
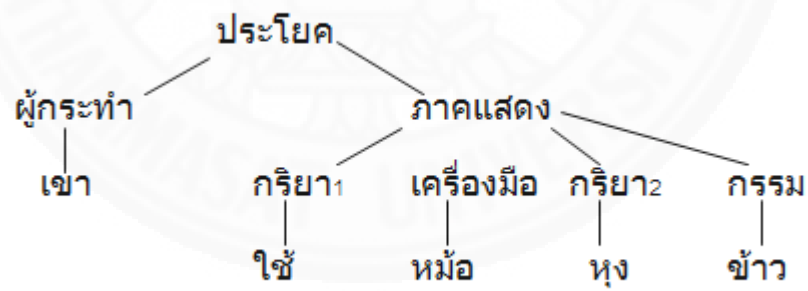
ผู้ประสบ-กริยา-(กรรม)



ผู้มีสภาพ-กริยา



เครื่องมือ-กริยา-(กรรม)



เมื่อพิจารณาถึงประเภทของหน่วยสมาชิกที่เป็นกริยาก็ย่อมพบว่าคำนามประสมที่มีอรรถสัมพันธ์แบบผู้กระทำ-กริยา-(กรรม) ผู้ประสม-กริยา-(กรรม) และเครื่องมือ-กริยา-(กรรม) มีคำกริยาประเภทเดียวกันคือคำกริยาแสดงอาการ (action verbs) คำนามประสมที่มีอรรถสัมพันธ์ในแบบผู้มีสภาพ-กริยา มีคำกริยาเป็นประเภทแสดงลักษณะ (descriptive verbs) เช่นเดียวกับที่พบในหน่วยไวยากรณ์ประโยค

### 2.1.3.2 อรรถสัมพันธ์ของคำนามประสมแบบกรรม-กริยา

โดยทั่วไปแล้ว คำนามประสมที่มีความสัมพันธ์ทางไวยากรณ์ของหน่วยสมาชิกเป็นแบบกรรม-กริยา มีหน่วยสมาชิกหลักกรรมทำหน้าที่ทางความหมายในฐานะผู้ถูกกระทำ ดังตัวอย่าง

#### ผู้ถูกกระทำ-กริยา

ลูก-อม	---- ลูกอม
กล้วย-ทอด	---- กล้วยทอด
บ้าน-เช่า	---- บ้านเช่า

### 2.1.3.3 อรรถสัมพันธ์ของคำนามประสมแบบหน่วยเสริม-กริยา

คำนามประสมชนิดแสดงไวยากรณ์ประโยคที่มีความสัมพันธ์ทางไวยากรณ์ของหน่วยสมาชิกเป็นแบบหน่วยเสริม-กริยา สามารถมีสมาชิกหลักที่เป็นหน่วยเสริมเป็นสถานที่และเวลา ซึ่งทำให้รูปแบบนี้มีสองรูปแบบย่อยที่แตกต่างกันออกไป คือ สถานที่-กริยา และเวลา-กริยา ดังแสดงตัวอย่างต่อไปนี้

#### สถานที่-กริยา-(กรรม)

ห้อง-รับ-แขก	---- ห้องรับแขก
ตลาด-นัด	---- ตลาดนัด

#### เวลา-กริยา-(กรรม)

วัน-ไหว้-ครู	----- วันไหว้ครู
วัน-หยุด	----- วันหยุด

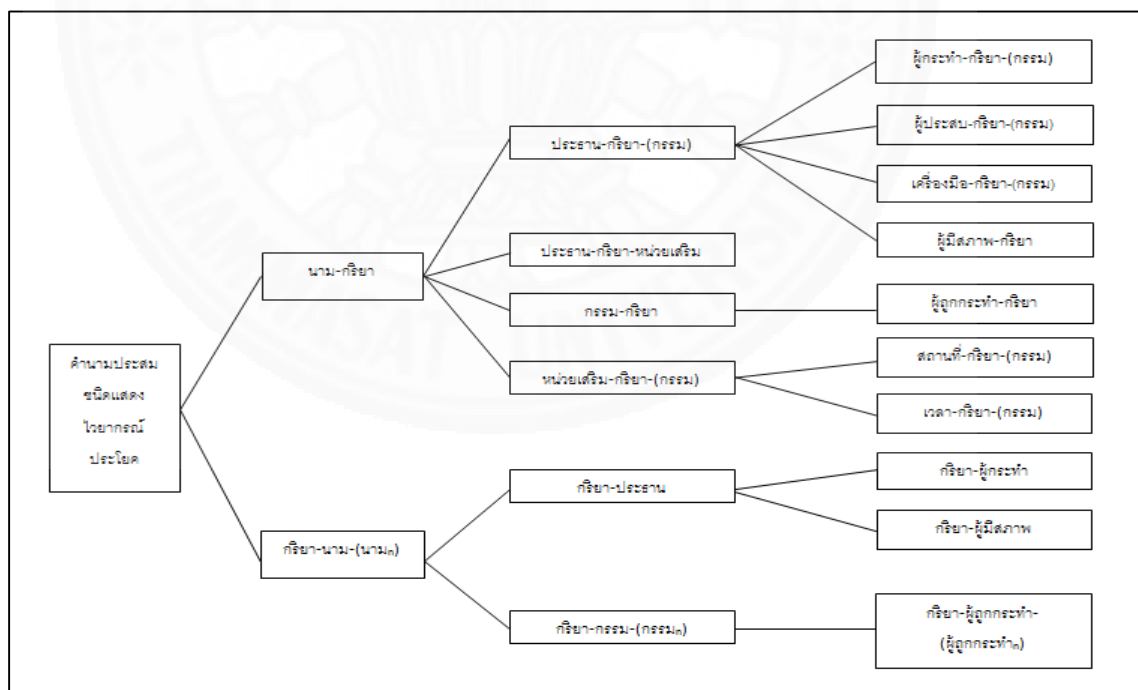


### 2.1.3.4 อรรถสัมพันธ์ของคำนามประสมแบบกริยา-ประธาน

คำนามประสมชนิดแสดงไวยากรณ์ประโยคในลักษณะกริยา-ประธานอาจจำแนกได้อีกเป็นสองกลุ่มย่อยตามความแตกต่างทางอรรถสัมพันธ์ของหน่วยสมาชิก คือ กริยา-ผู้กระทำ และกริยา-ผู้มีสภาพ ดังแสดงตัวอย่างดังต่อไปนี้

กริยา-ผู้กระทำ		กริยา-ผู้มีสภาพ	
พัด-ลม	---- พัดลม	หวาน-ใจ	----หวานใจ
กระโดด-กบ	----กระโดดกบ	ต่าง-ประเทศ	---- ต่างประเทศ
แข่ง-รถ	---- แข่งรถ		

การจำแนกคำนามประสมชนิดแสดงไวยากรณ์ประโยคออกเป็นกลุ่มย่อยๆ โดยพิจารณาจากโครงสร้าง วากยสัมพันธ์ และอรรถสัมพันธ์ของหน่วยสมาชิกดังกล่าวอาจสรุปเป็นแผนภูมิในรูปภาพที่ 1.1 ดังนี้



รูปภาพที่ 1.1 แผนภูมิแสดงโครงสร้าง วากยสัมพันธ์ และอรรถสัมพันธ์ของหน่วยสมาชิก

## 2.2 แนวทางในการพิจารณาความแตกต่างระหว่างคำนามประสมและวลี/ประโยค

ในทางปฏิบัติจริงแล้วก็อาจจะมีคำถามเกิดขึ้นว่าอะไรคือหลักเกณฑ์ที่เป็นรูปธรรมที่ชัดเจนในการจำแนกดังกล่าว บางครั้งความหมายของคำนามประสมก็ไม่ได้เปลี่ยนแปลงไปจาก ความหมายของวลีหรือประโยคอย่างชัดเจน ทำให้ไม่อาจแยกแยะหรือหาคำอธิบายได้อย่างเป็นที่น่าพอใจได้

อย่างไรก็ตาม คำนามประสมและวลี/ประโยคก็น่าที่จะต้องมีลักษณะทางไวยากรณ์บางประการที่แตกต่างกัน ดังที่ (Anderson 1985) ได้ให้เหตุผลไว้ว่า เนื่องจากคำนามประสมเป็นผลของกระบวนการสร้างคำ (word formation processes) ในขณะที่วลีหรือประโยคเป็นผลของการสร้างหน่วยทางวากยสัมพันธ์ที่ใหญ่กว่าระดับคำ (syntactic operations) ดังนั้น นอกจากบริบทรอบข้างแล้ว ลักษณะทางไวยากรณ์ที่อาจใช้ประกอบกันเป็นแนวทางในการพิจารณาว่า คำที่เกิดต่อเนื่องกันนั้นๆเป็นคำนามประสมหรือวลี/ประโยค น่าจะรวมถึงประเด็นที่เกี่ยวกับอรรถศาสตร์ (semantics) อรรถศาสตร์ (pragmatics) วากยสัมพันธ์ (syntax) ที่เกี่ยวกับลำดับการเรียงคำ (word order) และความสัมพันธ์ทางวากยสัมพันธ์ (syntactic relationship) ข้อจำกัดของการเป็นคำสมาชิก (limitation of lexicon) และ วิทยาหน่วยคำ (morphology) ดังจะกล่าวต่อไปนี้

### 2.2.1 เกณฑ์ทางอรรถศาสตร์

เป็นเกณฑ์สำคัญที่มักจะนำไปเป็นหลักในการพิจารณาคำประสม กล่าวคือ หากข้อความใดมีความหมายที่ได้จากการรวมเข้าด้วยกันของหน่วยสมาชิกอย่างตรงไปตรงมา (compositional meaning) ข้อความนั้นจะถือเป็นวลีหรือประโยค หากข้อความใดมีความผิดไปจากการรวมกันของหน่วยย่อย เกิดเป็นความหมายใหม่ที่ไม่อาจคาดเดาได้ (idiosyncratic sense) จะถือว่าเป็นคำประสม

แต่อย่างไรก็ตาม เกณฑ์ทางอรรถศาสตร์ดังกล่าวก็ไม่สามารถเป็นข้อตัดสินได้อย่างปราศจากปัญหาดังที่ (Allen 1978) ได้กล่าวไว้ว่า คำนามประสมชนิดใดก็ตามมีทั้งลักษณะที่สามารถทำนายได้จากความหมายของคำประกอบ และก็เป็นจริงอย่างที่กล่าวเมื่อพบว่ามีข้อความที่เราจัดว่าเป็นคำนามประสมในเหตุผลบางประการ แต่มีความหมายเกิดจากการรวมกันของคำสมาชิกอย่างตรงไปตรงมาในลักษณะเช่นเดียวกันกับวลีและประโยคซึ่งสามารถทำนายได้ยกตัวอย่างเช่น คนขับรถ คนทำสวน เตียนนอน ฯลฯ เราถือว่าเป็นคำประสมในขณะที่มีความหมายที่ทำนายได้เช่นเดียวกันกับวลี คนที่ขับรถ คนที่ทำสวน เตียนที่ใช้นอน

ลักษณะทางอรรถศาสตร์อีกประการหนึ่งที่มีประโยชน์ในการบอกความแตกต่างระหว่างคำประสมและวลี/ประโยค คือคำนามประสมมีอรรถสัมพันธ์ระหว่างหน่วยสมาชิกที่หลากหลาย

มากกว่าของวลีและประโยค ยกตัวอย่างเช่นในขณะที่วลีที่มีโครงสร้างแบบคำนาม คำนามสามารถมีรูปแบบอรรถสัมพันธ์เป็นไปได้เพียง 2 แบบเท่านั้นคือ สิ่งที่ถูกครอบครอง-เจ้าของ (เช่น บ้านฉัน) และ สิ่งประดิษฐ์ – วัสดุ (เช่น บ้านไม้) แต่คำนามประสมชนิดคำนาม – คำนามสามารถมีรูปแบบของอรรถสัมพันธ์ได้มากกว่าสิบแบบนอกเหนือไปจากนี้

### 2.2.2 เกณฑ์ทางวិธีวิทยา

เป็นเกณฑ์ที่สำคัญที่สุดที่ได้ในการพิจารณาตัดสินว่า คำที่เกิดต่อเนื่องกันนั้นเป็นคำนามประสมหรือวลี/ประโยค กล่าวคือ นอกจากเราอาศัยบริบททางภาษาหรือข้อความรอบข้างซึ่งเป็นข้อมูลด้านหนึ่งทางวิธีวิทยาในการพิจารณาว่า คำที่เกิดต่อเนื่องกันนั้นเป็นคำประสม วลีหรือประโยคดังที่ได้กล่าวมาแล้ว ยังมีปัจจัยทางวิธีวิทยาอื่นๆที่ใช้บ่งบอกลักษณะพื้นฐานของคำนามประสมโดยเฉพาะที่สำคัญคือ ความถาวร (permanence) หรือความเป็นปกติวิสัย (habitual nature) ของความหมายที่ได้จากการประกอบคำเข้าด้วยกัน ซึ่งสามารถนำมาเป็นเกณฑ์ในการพิจารณาแยกแยะคำประสมออกจากโครงสร้างอื่นๆ ได้ ยกตัวอย่างเช่น คำว่า คนขับรถ คนกวาดถนน ฯลฯ ใช้เรียกอาชีพของคนในสังคม ซึ่งมีความถาวรหรือเป็นสิ่งที่ทำเป็นประจำเป็นปกติวิสัย ในขณะที่วลีหรือประโยคเทียบเคียง คือ คนที่ขับรถ คนที่กวาดถนน ไม่แสดงความเป็นปกติวิสัย

### 2.2.3 เกณฑ์ทางวากยสัมพันธ์

#### 2.2.3.1 ลำดับการเรียงคำ

ลำดับการเรียงคำอาจใช้เป็นอีกแนวทางหนึ่งที่จะบอกความแตกต่างระหว่างคำประสมกับวลี/ประโยค โดยเฉพาะกรณีที่เป็นคำนามประสมชนิดแสดงวากยสัมพันธ์ประโยค ซึ่งปรากฏมีคำนามประสมจำนวนหนึ่งในชนิดนี้ที่มีการเรียงลำดับการเรียงคำแตกต่างจากประโยคตามปกติแล้วภาษาไทยมีลำดับการเรียงหน่วยหน้าที่วากยสัมพันธ์ประโยคแบบประธาน – กริยากรรม – หน่วยเสริม และค่านานประสมจำนวนหนึ่งก็จะเลียนแบบลำดับดังกล่าว เช่น ช่างตีเหล็ก ช่างปั้นหม้อ ฯลฯ แต่กรณีนั้นก็มีความประสมจำนวนหนึ่งที่มีลำดับการเรียงคำแตกต่างจากประโยค กล่าวคือ จะเห็นได้ว่าโดยปกติแล้ว กริยาของประโยคจะนำหน้ากรรมหรือหน่วยเสริม เช่น เขาเข้าบ้าน เขาล้มหมู

### 2.2.3.2 ความสัมพันธ์ของสมาชิกค่านามประสม

#### (1) สมาชิกของค่านามประสมไม่อาจเกิดแยกกันได้

ลักษณะทางวากยสัมพันธ์ที่สำคัญเป็นที่รู้จักดีประการหนึ่งในการทดสอบคำประสมคือสมาชิกของคำประสมไม่อาจแยกกันได้ ไม่ว่าจะเป็นการแยกออกจากกันหรือการแทรกคำอื่นระหว่างคำสมาชิกก็ตาม

อย่างไรก็ตามพบว่า สมาชิกของคำประสมอาจดูเหมือนว่าจะสามารถมีคำแทรกได้ดังในกรณีเช่นคำว่า คนใช้ ที่บางครั้งเป็น คนรับใช้ ซึ่งดูเหมือนว่าจะสามารถมีคำว่า รับ แทรกมาได้

#### (2) สมาชิกของค่านามประสมไม่สามารถแทนที่ด้วยคำอื่นได้

คำประสมจะไม่สามารถแยกคำประสมออกจากกันแล้วนำคำอื่นไปแทนที่ไม่ว่าตำแหน่งใดก็ตาม เช่นคำว่า พัดลม ไม่สามารถนำคำอื่นที่มีความหมายเหมือนหรือคล้ายกันไปแทนที่ส่วนประกอบใดได้ แต่อย่างไรก็ตามมีคำประสมจำนวนหนึ่ง เช่น ช่างตัดเสื้อ อาจเปลี่ยนคำได้หลากหลายเช่น ช่างเย็บเสื้อ ช่างตัดผ้า ซึ่งก็จะหมายถึงบุคคลในอาชีพดังกล่าวเหมือนกัน

#### (3) สมาชิกของค่านามประสมไม่อาจมีหน่วยขยายโดยลำพังได้

หน่วยสมาชิกของค่านามประสมไม่อาจมีหน่วยขยายได้อย่างอิสระ ถ้าหากมีคำขยายคำนั้นจะต้องขยายทั้งคำประสม ไม่เฉพาะสมาชิกตัวใดตัวหนึ่ง

#### (4) สมาชิกของค่านามประสมไม่อาจเชื่อมกับคำอื่นได้

### 2.2.4 เกณฑ์ทางข้อจำกัดของคำสมาชิก

คำประสมมีข้อจำกัดในเรื่องของความหลากหลายของคำสมาชิก คือ ไม่ใช่คำใดก็ได้ที่สามารถนำมาแทนที่สมาชิกหน่วยหนึ่งของคำประสม แต่จะจำกัดอยู่เฉพาะกับคำจำนวนหนึ่งเท่านั้น อาทิ คน ช่าง ผู้ เด็ก แม่ พ่อ ลูก หมอ ในขณะที่ประโยคสามารถมีคำอื่นใดก็ได้ในตำแหน่งนี้

### 2.2.5 เกณฑ์ทางคุณสมบัติหน่วยคำ

ลักษณะทางหน่วยคำที่คำประสมแตกต่างจากโครงสร้างที่ใหญ่กว่าเช่นวลีและประโยคคือเมื่อภาษาเปลี่ยนแปลงไป คำประสมอาจพัฒนาเปลี่ยนแปลงเป็นหน่วยที่เล็กขึ้นตัวอย่างเช่น คำเรียกผลไม้จำนวนหนึ่งที่ขึ้นต้นด้วย มะ หน่วยคำนี้มีหลักฐานว่ากร่อนมาจากคำว่า หมาก

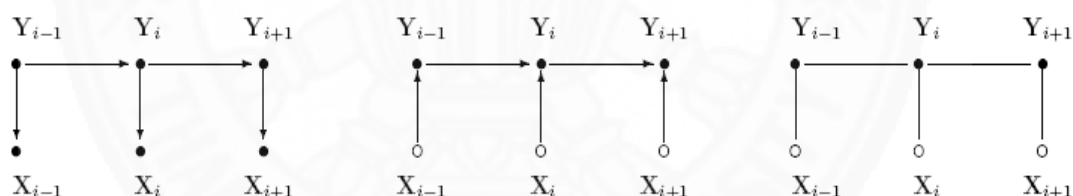
ดังที่ได้เสนอมาแล้วข้างต้น การกำหนดเกณฑ์หรือแนวทางจำนวนหนึ่งที่สามารถใช้ในการพิจารณาว่าโครงสร้างหนึ่งๆนั้นเป็นคำประสม ไม่ใช่วลีหรือประโยค เกณฑ์ต่างๆครอบคลุมทั้งในเรื่องของวากยสัมพันธ์ อรรถสัมพันธ์ วิธรรณสัมพันธ์ และกระบวนการเป็นคำ ซึ่งในหลายกรณี ไม่

สามารถที่จะพิจารณาตัดสินได้จากเกณฑ์ใดเกณฑ์เดียว จำเป็นต้องใช้กฎเกณฑ์หลายข้อประกอบกันในการพิจารณา

### 2.3 แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์

คอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Fields - CRFs) เป็นแบบจำลองในการใช้วิธีตัดคำหรือติด label ข้อความเรียงกันตามลำดับ ซึ่งสามารถนำไปประยุกต์ใช้กับการประมวลผลภาษาธรรมชาติ เช่น การรู้จำนิพจน์ระบุนาม (Named Entity Recognition) การสกัดสารสนเทศ (Information Extraction) เป็นต้น

แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ เป็นแบบจำลองที่มีลักษณะเป็น Discriminative Model โดยที่อาศัยความน่าจะเป็นแบบมีเงื่อนไขของลำดับของเลเบล (Label Sequence) ร่วมกับลำดับของข้อมูล (Observation Sequence) ทำให้ได้ความสัมพันธ์ระหว่างคุณสมบัติต่างๆ ที่เกี่ยวข้องกันของข้อมูลที่พบได้ โดยเป็นแบบจำลองที่ลดปัญหาการเกิด model bias ซึ่งแบบจำลองนี้ถูกนำเสนอโดย Lafferty et al. (2001) เมื่อนำแบบจำลอง CRFs มาวาดเป็นกราฟเปรียบเทียบกับอีกสองแบบจำลอง Hidden Markov Models (HMMs) และ Maximum Entropy Markov Models (MEMMs) จะได้ดังภาพที่ 2.1



ภาพที่ 2.1 รูปกราฟแสดงการเปรียบเทียบแบบจำลอง HMMs, MEMMs และ CRFs

จากรูปกราฟแสดงการเปรียบเทียบแบบจำลอง โดยกำหนดให้ตัวแปร  $X$  คือ ลำดับข้อมูลที่พบ (Observation Sequence) และตัวแปร  $Y$  คือ ลำดับของเลเบล (Label Sequence) ซึ่งแบบจำลอง HMMs และ MEMMs นั้นมีลักษณะเป็นกราฟระบุทิศทาง (Directed Graphical Model) โดยที่ HMMs เป็นแบบจำลองที่มีลักษณะเป็น Generative model ที่อาศัยการแจกแจงความน่าจะเป็นร่วม (Joint Probability Distribution) ทำให้ลำดับข้อมูลที่พบของ HMMs ไม่ได้มีส่วนเกี่ยวข้องกับลำดับเลเบลหรือผลลัพธ์ จึงมีปัญหว่าไม่สามารถจับความสัมพันธ์ระหว่างคุณสมบัติต่างๆ ที่เกี่ยวข้องกันในข้อมูลที่เข้ามาได้ เนื่องจากคุณสมบัติต่างๆ นั้นเป็นอิสระต่อกัน สำหรับแบบจำลอง MEMMs และ CRFs นั้นมีลักษณะเป็น Discriminative Models ลำดับข้อมูลที่พบของแบบจำลอง

MEMMs และ CRFs เป็นสิ่งที่ไม่ได้เกิดจากการประมวลผลของแบบจำลองแต่เป็นเงื่อนไขในการกำหนดสายของเลเบล มีลักษณะการแจกแจงความน่าจะเป็นแบบมีเงื่อนไข  $P(Y|X)$  นั่นคือลำดับข้อมูลที่พบ  $X$  เกิดขึ้นก่อนลำดับของเลเบล  $Y$  แบบจำลอง แต่แบบจำลอง MEMMs ยังคงพบปัญหาเนื่องจากการตัดสินใจผลในสถานะใดจะขึ้นกับสถานะปัจจุบันและลำดับข้อมูลที่พบเท่านั้น สถานะอื่นๆ ไม่มีผลต่อการคำนวณค่าความน่าจะเป็น โดยที่แบบจำลอง CRFs จึงเป็นแบบจำลองที่ลดปัญหาดังกล่าว ซึ่งจะมีลักษณะเป็นแบบจำลองกราฟไม่ระบุทิศทาง ดังนั้นจึงต่างจากแบบจำลอง MEMMs เนื่องจากแบบจำลอง CRFs หาค่าความน่าจะเป็นของเลเบลถัดไปโดยนำความน่าจะเป็นเลเบลก่อนหน้าทั้งหมดที่มีลำดับเหตุการณ์เป็นเงื่อนไขมาคำนวณค่าความน่าจะเป็นด้วย น้ำหนักของคุณสมบัติต่าง ๆ จากสถานะที่ต่างกันจึงมีการปรับสมดุลให้ค่าไม่เอนเอียงไปสถานะใดสถานะหนึ่ง

ดังนั้นแบบจำลอง CRFs จึงถูกนำมาใช้ในการศึกษาวิจัยในหลายวิจัยที่ได้จากการสืบค้นมีดังต่อไปนี้

งานวิจัยของ Fuchun Peng และคณะ ในปี 2004 นำเอาเทคนิค CRFs มาประยุกต์ใช้ในการตัดคำของภาษาจีน จากผลงานวิจัยพบว่า CRFs สามารถช่วยเพิ่มประสิทธิภาพในการตัดคำในภาษาจีน

งานวิจัยของ John Lafferty และคณะ ในปี 2001 นำเอาเทคนิค CRFs มาช่วยในการแบ่งแยกข้อมูลที่มีอยู่ต่อเนื่องกัน เปรียบเทียบประสิทธิภาพกับวิธี Maximum Entropy Markov Models (MEMMs) ผลงานวิจัยพบว่า CRFs มีข้อผิดพลาด (error) เท่ากับ 4.6 % ขณะที่ MEMMs มีข้อผิดพลาด (error) ถึง 42%

งานวิจัยของ ชูชาติ ไชยยะศักดิ์ และคณะ ในปี 2551 นำเอาเทคนิค CRFs มาเป็นวิธีตัดคำร่วมกับการตัดคำแบบใช้พจนานุกรมเพื่อที่จะแก้ไขปัญหาในส่วนของการตัดคำที่กำกวม โดยเรียกวิธีนี้ว่า LearnLexTo เปรียบเทียบกับการตัดคำแบบใช้พจนานุกรมเพียงอย่างเดียว ผลการศึกษาพบว่า LearnLexTo มีค่า F-measure metric 87.67% ซึ่งมากกว่าการตัดคำแบบใช้พจนานุกรมอย่างเดียวที่มีค่า F-measure metric 82.71%

งานวิจัยของ นัชชา ธีระสาโรช และคณะ ในปี 2553 นำเอาเทคนิค CRFs ในการศึกษาระบบการรู้จำนิพจน์ระบุนาม (Thai named entity recognition) ศึกษาเปรียบเทียบ input ที่ได้จากกร word-segmented กับ input syllable-segmented ผลการศึกษาพบว่า syllable segmented ดีกว่า word-segmented เล็กน้อย

## บทที่ 3

### วิธีการวิจัย

สำหรับในบทนี้จะกล่าวถึงขอบเขตการดำเนินงานวิจัย เครื่องมือหรือแบบจำลองที่ใช้ในการพัฒนาระบบ ขั้นตอนการทดลองและวิธีการวัดผลลัพธ์ของการทดลอง

#### 3.1 ขอบเขตการดำเนินงานวิจัย

##### 3.1.1 ข้อมูลที่ใช้ในการทำวิจัย

คลังข้อความสำหรับใช้ในงานวิจัยนี้เป็นคลังข้อความภาษาไทย (BEST) ซึ่งมีการกำกับขอบเขตของคำขนาด 5 ล้านคำ ประกอบด้วย 4 หมวด คือ บทความวิชาการ (Article), สารานุกรม (Encyclopedia), ข่าว (News), และ นวนิยาย (Novel) จัดทำขึ้นโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) โดยงานวิจัยนี้คัดเลือกใช้คลังข้อความภาษาไทย (BEST) เฉพาะในหมวดของข่าว (News) มาใช้สำหรับทำการวิจัยเท่านั้น เนื่องจากบทความของข่าวมีรูปแบบของการเขียนที่สามารถพบได้ในชีวิตประจำวัน ดังแสดงตัวอย่างข้อมูลในตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างคลังข้อมูล BEST สำหรับฝึกฝนและทดสอบระบบ

นาย/NN/TTL\_B|ประเสริฐ/NR/PER\_B| |บุญสัมพันธ์/NR/PER\_I| |ประธาน/NN/O|เจ้าหน้าที่/  
 NN/O|บริหาร/JJV/O| |และ/CNJ/O|กรรมการ/NN/O|ผู้/FXN/O|จัดการ/JJV/O|ใหญ่/JJA/O| |  
 บมจ.ปตท./NR/NAME\_B|กล่าว/VV/O|ว่า/COMP/O| |ใน/P/O| |1-2/CD/MEA\_B| |วัน/  
 CL/MEA\_B|นี้/DDEM/O| |อาจ/AUX/O|จะ/AUX/O|ประกาศ/VV/O|ลด/VV/O|ราคา/NN/O|ขาย/  
 JJV/O|ปลีก/ADV/O|น้ำมัน/NN/O|ใน/P/O|ใน/P/O|ประเทศ/NN/O|ได้/AUX/O| |หาก/CNJ/O|  
 ราคา/NN/O|น้ำมัน/NN/O|ตลาดสิงคโปร์/NR/NAME\_B|ลด/VV/O|ลง/VV/O|ฮวบ/ADV/O|ตาม/  
 VV/O|ราคา/NN/O|น้ำมัน/NN/O|ดิบ/JJA/O|สหรัฐ/NR/ORG\_B|ที่/COMP/O|ลด/VV/O|ลง/  
 VV/O|ประมาณ/PDT/O| |6/CD/MEA\_B| |ดอลลาร์/CL/MEA\_B|สหรัฐ/NR/MEA\_I|ต่อ/PDT/O|  
 บาร์เรล/CL/O| |ซึ่ง/COMP/O|เป็น/VV/O|ผล/NN/O|มา/VV/O|จาก/P/O|การ/FXN/O|หัวนเกรง/  
 VV/O|ปัญหา/NN/O|ผล/NN/O|กระทบ/JJV/O|สินเชื่อ/NN/O|ซบไฟร์ม/NN/O|รอบ/NN/O|ใหม่/  
 JJA/O|ของ/P/O|สหรัฐ/NR/ORG\_B| |และ/CNJ/O|หาก/CNJ/O|ปัญหา/NN/O|ความ/FXN/O|ดึง  
 เครียด/VV/O|ด้าน/NN/O|อื่น/DDEM/O|ๆ/PU/O| |โดย/COMP/O|เฉพาะ/ADV/O|ปัญหา/NN/O|  
 อิหร่าน/NR/ORG\_B|หมด/ADV/O|ไป/VV/O|

### 3.1.2 เครื่องมือที่ใช้ในงานวิจัย

ในส่วนนี้จะกล่าวถึงเครื่องมือต่างๆ ที่นำมาใช้ในการดำเนินงานวิจัย ซึ่งแสดงในตารางที่

### 3.2

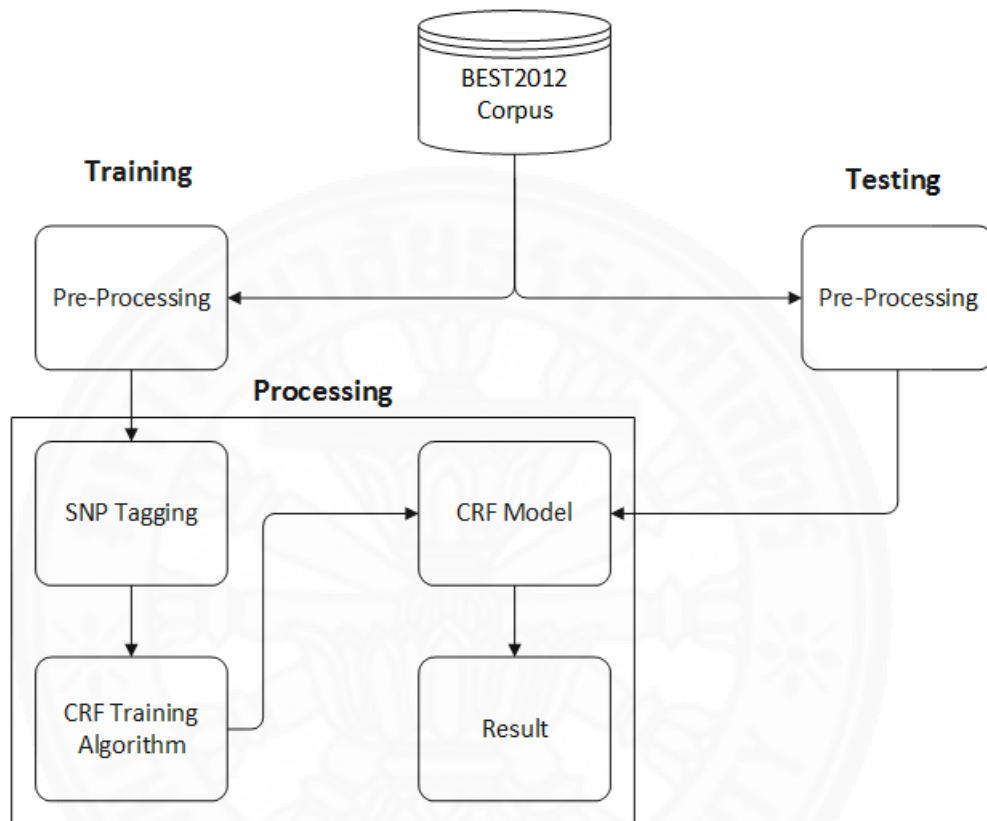
ตารางที่ 3.2 เครื่องมือที่ใช้ในการดำเนินงานวิจัย

ฮาร์ดแวร์ (Hardware)	Intel® Core™ i7-4710HQ CPU @ 2.50GHz Main Memory 8.00 GB Hard Disk 500 GB Graphic Chip GeForce GTX 850M
ซอฟต์แวร์ (Software)	ระบบปฏิบัติการ Windows 8.1 คลังข้อมูลข่าวภาษาไทย (BEST) Visual Studio 2013 CRFSharp สำหรับฝึกฝนและทดสอบค่านาม ประสมชนิดไวยากรณ์ประโยค สามารถดาวน์โหลดได้จาก <a href="https://crfsharp.codeplex.com/">https://crfsharp.codeplex.com/</a>



### 3.2 กระบวนการวิจัย

ในส่วนนี้จะอธิบายถึงกระบวนการต่างๆ ในการดำเนินงานวิจัย ดังแสดงในรูปที่ 3.1



รูป 3.1 กระบวนการวิจัย

ซึ่งกล่าวในรายละเอียดถึงกระบวนการวิจัยที่ประกอบด้วย 2 กระบวนการดังต่อไปนี้

**3.2.1 กระบวนการก่อนประมวลผล (Preprocessing)** เป็นกระบวนการในการเตรียมพร้อมข้อมูลในการฝึกฝน โดยนำคลังข้อความภาษาไทย (BEST) เฉพาะในหมวดของข่าวที่มีคำประสม (News) จำนวน 1,000 คำ มาแบ่งชุดข้อมูลสำหรับการฝึกฝนและทดสอบ โดยขั้นตอนก่อนประมวลผลนี้จะนำประโยคที่มีชนิดของคำที่เป็น JJV และนำประโยคที่ได้มาตัดคำและกำกับชนิดของคำ ดังแสดงในตาราง 3.3

ตารางที่ 3.3 ตัวอย่างบทความ “นายเอสมัครงานเป็นคนขับรถ”

นาย	เอ	สมัคร	งาน	เป็น	คน	ขับ	รถ
NN	NN	VV	NN	P	NN	JJV	NN

การกำกับคำนามประสมชนิดไวยากรณ์ประโยค เริ่มจากการค้นหาชนิดของคำที่เป็น JJV หลังจากนั้นให้เริ่มหาขอบเขตของนามวลี โดยใช้คำนามหลักในคำนามประสมชนิดไวยากรณ์ประโยคมาใช้ในการค้นหาตำแหน่งเริ่มต้นของคำ หลังจากนั้นเมื่อได้ตำแหน่งเริ่มต้นของคำแล้วให้ทำการหาตำแหน่งสิ้นสุดของคำ โดยสังเกตจากชนิดของคำที่อยู่นอกเหนือจาก NN, NR, JJV, JJA ซึ่งได้แสดงตัวอย่างในตารางที่ 3.4

ตารางที่ 3.4 ตัวอย่างบทความที่มีการแจกแจงชนิดของคำ

คำ	ชนิดของคำ	คำอธิบาย	คำกำกับ
จัด	VV	คำกริยา	S
เตรียม	VV	คำกริยา	S
อุปกรณ์	NN	คำนาม	B_SNP
กล่อง	NN	คำนาม	M_SNP
ส่ง	JJV	คำกริยา	M_SNP
ดวง	NN	คำนาม	M_SNP
อาทิตย์	NN	คำนาม	E_SNP

3.2.2 กระบวนการประมวลผล (Processing) เป็นกระบวนการรู้จำคำนามประสมชนิดไวยากรณ์ประโยคด้วยเทคนิค CRFs โดยมีการกำหนดกฎการรู้จำคำนามประสมชนิดไวยากรณ์ประโยคที่จะนำมาใช้สำหรับเทคนิคการเรียนรู้เพื่อระบุคำนามประสมชนิดไวยากรณ์ประโยคสำหรับในงานวิจัยนี้ได้แบ่งออกเป็น 2 กรณี ดังนี้

3.2.2.1. พิจารณาคำนามหลักของคำนามประสมชนิดไวยากรณ์ประโยค แบ่งออกเป็น 5 ประเภท ได้แก่ คำนามหลักเกี่ยวกับอาชีพ คำนามหลักเกี่ยวกับการแพทย์ คำนามหลักเกี่ยวกับเครื่องมือ คำนามหลักเกี่ยวกับสถานที่ และคำนามหลักเกี่ยวกับวัน เวลาหรือเหตุการณ์ต่างๆ ตามที่แสดงในตารางที่ 3.5

ตารางที่ 3.5 ตัวอย่างค่านามหลักในคำประสมที่มีโครงสร้างคล้ายประโยค

ความหมายทางอรรถศาสตร์	ตัวอย่างค่านามหลัก	ตัวอย่างคำประสม
1. ค่านามหลักเกี่ยวกับอาชีพ	คน พนักงาน เจ้าหน้าที่ อาจารย์ ครู ช่าง เด็ก	คนขับรถ อาจารย์สอนภาษาไทย พนักงานเก็บตัวโดยสาร
2. ค่านามหลักเกี่ยวกับ การแพทย์	โรค ไวรัส เชื้อ เซลล์ วัคซีน	เข็มฉีดยา วัคซีนป้องกันโรค
3. ค่านามหลักเกี่ยวกับ เครื่องมือ	เครื่อง ที่ ไม้ ชูต อุปกรณ์ เครื่องใช้	เครื่องซักผ้า ไม้แขวนเสื้อ ที่ตัด กระดาษ กรรไกรตัดเล็บ
4. ค่านามหลักเกี่ยวกับ สถานที่	โรง ร้าน ตึก อาคาร ที่ แปลง สวน ศูนย์ ย่าน ห้อง แผลง	ร้านขายของชำ โรงเพาะ ศูนย์ อนุรักษ์พันธุ์สัตว์ทะเล แปลงปลูก ผัก แปลงขายผ้า
5. ค่านามหลักเกี่ยวกับวัน เวลา หรือเหตุการณ์ต่างๆ	วัน งาน เทศกาล พิธี	วันสถาปนาโรงเรียน งานเลี้ยง ต้อนรับพนักงานใหม่

โดยงานวิจัยนี้ได้นำประเภทของค่านามหลักมาใช้กำหนดคุณสมบัติ เพื่อใช้ระบุประเภทของค่านามหลัก โดยมีตัวอย่างในตารางที่ 3.6 ดังต่อไปนี้

ตารางที่ 3.6 ตัวอย่างไฟล์สำหรับฝึกฝนที่ระบุประเภทของค่านามหลัก

คำ	ชนิดของ คำ	อาชีพ	การแพทย์	เครื่องมือ	สถานที่	วัน เวลา	คำกำกับ
แพทย์	NN	0	0	0	0	0	S
ห้วง	VV	0	0	0	0	0	S
โรค	NN	0	1	0	0	0	B_SNP

คำ	ชนิดของ คำ	อาชีพ	การแพทย์	เครื่องมือ	สถานที่	วัน เวลา	คำกำกับ
ระบด	JJV	0	0	0	0	0	E_SNP
จาก	P	0	0	0	0	0	S
ภาวะ	NN	0	0	0	0	0	S
โลก	NN	0	0	0	0	0	S
ร้อน	JJA	0	0	0	0	0	S

**3.2.2.2. พิจารณาจากโครงสร้างประโยค** ตามที่ได้มีการกล่าวถึงในบทที่ 2 หัวข้อ 2.1.1 ที่ได้มีการจำแนกไว้ 5 ประเภท แต่ในงานวิจัยนี้สนใจศึกษาใน 3 ประเภทที่ขึ้นต้นด้วยคำนาม ซึ่งสามารถพบได้บ่อย ดังต่อไปนี้

- (1) กำหนดขอบเขตสองคำ นาม-กริยา
- (2) กำหนดขอบเขตสามคำ นาม-กริยา-นาม
- (3) กำหนดขอบเขตสี่คำ นาม-กริยา-นาม-นาม

นาม-นาม-กริยา-นาม

นาม-กริยา-นาม-กริยา

นาม-นาม-นาม-กริยา

### 3.3 ขั้นตอนการทดลอง

งานวิจัยนี้ใช้บทความข่าวภาษาไทยที่มีคำนามประสมชนิดไวยากรณ์ประโยค จำนวน 1,000 คำ มาทดสอบโดยแบ่งตามกรณีดังที่ได้กล่าวในข้อ 3.2.2 และใช้เทคนิค 10-fold cross validation (Ron, 1995) มาแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน โดยใช้ข้อมูล 9 ส่วนสำหรับการฝึกฝน และ 1 ส่วนสำหรับการทดสอบ โดยเปลี่ยนชุดข้อมูลสำหรับการฝึกฝนและทดสอบไปเรื่อยๆ จนกระทั่งถึงส่วน

สำหรับการทดสอบเป็นชุดที่ 10 เพื่อสร้างแบบจำลอง ซึ่งผลลัพธ์ที่ได้จากแบบจำลองจะถูกตรวจสอบด้วยนักภาษาศาสตร์ โดยประเมินประสิทธิภาพจากค่าความระลึก ค่าความแม่นยำ และค่าประสิทธิภาพแยกจากตามกรณีเพื่อประเมินผลการทำลอง

### 3.4 การประเมินผลการทดลอง

ทำการเปรียบเทียบผลการทดลองการรู้จำคำนามประสมชนิดไวยากรณ์ประโยค โดยใช้แบบจำลอง CRFs เปรียบเทียบในแต่กฎที่ระบุไว้ในหัวข้อ 3.2.2 โดยมีสูตรในการคำนวณดังต่อไปนี้

ค่าประสิทธิภาพโดยรวมของระบบ (F-measure)

$$F\text{-Measure} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall})$$

ความแม่นยำ (Precision)

$$\text{Precision} = \text{Corr} \times 100 / \text{OutputWord}$$

ค่าความระลึก (Recall)

$$\text{Recall} = \text{Corr} \times 100 / \text{RefWord}$$

โดยกำหนดให้

Corr = จำนวนคำที่เลือกได้ถูกต้อง

OutputWord = จำนวนคำที่ระบบเลือกออกมาทั้งหมด

RefWord = จำนวนคำที่เลือกโดยนักภาษาศาสตร์

## บทที่ 4 ผลการวิจัยและอภิปรายผล

### 4.1 การตั้งค่าการทดลอง

**4.1.1 ชุดข้อมูล** ในงานวิจัยนี้ทำการทดลองโดยใช้บทความข่าวภาษาไทยจากคลังข้อมูลภาษาไทย BEST โดยมีตัวอย่างบทความข่าวสำหรับการทดลองดังแสดงในตารางที่ 4.1 ต่อไปนี้

ตารางที่ 4.1 ตัวอย่างบทความข่าวสำหรับการทดลอง

ต่อมา/CNJ/O| เวลา/NN/DTM\_B| |10.00/OD/DTM\_I| |น./CL/DTM\_I| |วัน/NN/O| |เดี่ยว/JJA/O|  
 กัน/REFX/O| |ที่/P/O| |ศาลากลาง จ.อุดรธานี/NR/PER\_B| | นาย/NN/TTL\_B| |สุพจน์/NR/PER\_B| |  
 เลาวินัยศิริ/NR/PER\_I| | ผวจ./NN/ABB\_DES\_B| | อุดรธานี/NR/LOC\_B| | เรียก/VV/O| | ประชุม/  
 VV/O| | ด่วน/JJA/O| | เพื่อ/CNJ/O| | เตรียม/VV/O| | งาน/NN/O| | "/PU/O| | ชาว/FXG/O| | อุดร/NR/O| | |รัฐ/  
 VV/O| | รัก/VV/O| | สามัคคี/VV/O| | "/PU/O| | มี/VV/O| | ตัว/NN/O| | แทน/JJV/O| | จาก/P/O| | มณฑล  
 ทหารบก24/NR/PER\_B| |, /PU/O| | กองบิน 23/NR/PER\_B| |, /PU/O| | ตำรวจ/NN/O| | ภูธร/NN/O|  
 |, /PU/O| | ตำรวจ/NN/O| | ตระเวน/JJV/O| | ชาย/NN/O| | แดน/NN/O| |, /PU/O| | ฝ่าย/NN/O|  
 ปกครอง/JJV/O| | และ/CNJ/O| | หน่วย/NN/O| | งาน/NN/O| | อื่น/DDEM/O| |, /PU/O| | โดย/  
 COMP/O| | จะ/AUX/O| | จัด/VV/O| | ให้/COMP/O| | มี/VV/O| | การ/FXN/O| | เดิน/VV/O| | ธรรมดา/VV/O| |  
 ร่วม/VV/O| | กัน/REFX/O| | ทำ/VV/O| | ให้/COMP/O| | บ้านเมือง/NN/O| | สงบสุข/VV/O| | | โดย/  
 COMP/O| | จะ/AUX/O| | แบ่ง/VV/O| | ขบวน/NN/O| | เป็น/P/O| | |2/CD/NUM\_B| | สาย/CL/MEA\_B| |  
 สาย/CL/O| | แรก/OD/O| | รวม/VV/O| | ตัว/NN/O| | ที่/P/O| | |ร.ร. บ้านหมากแข้ง/NR/PER\_B| | |สาย/  
 NN/O| | สอง/CD/O| | ที่/P/O| | |ร.ร. คุณากร/NR/PER\_B| | |แล้ว/AUX/O| | เดิน/VV/O| | มุ่งหน้า/VV/O|  
 มา/VV/O| | ยิง/P/O| | สนามทุ่งศรีเมือง/NR/PER\_B| | เวลา/NN/DTM\_B| | |10.00/OD/DTM\_I| | |น./  
 CL/DTM\_I| | |วัน/NN/DTM\_B| | |อาทิตย์/NN/DTM\_I| | |ที่/P/DTM\_I| | |3/OD/DTM\_I| | |สิงหาคม/  
 NN/DTM\_I| | |นี้/DDEM/O| | |จาก/P/O| | |นั้น/DDEM/O| | |ได้/AUX/O| | |ประชุม/VV/O| | |ลับ/ADV/O| | |เพื่อ/  
 CNJ/O| | |เตรียม/VV/O| | |รับ/VV/O| | |สถานการณ์/NN/O| | |รุนแรง/JJA/O|

**4.1.2 วิธีการเลือกขอบเขตค่านามประสมชนิดแสดงไวยากรณ์ประโยค การฝึกฝนแบบจำลอง** ในการเลือกขอบเขตค่านามประสมชนิดไวยากรณ์ประโยค ระบบที่พัฒนาขึ้นมาสำหรับเตรียมข้อมูลและกำกับคำจะเริ่มจากค้นหาชนิดของคำที่เป็น JJV หลังจากนั้นจะเริ่มค้นหาคำขึ้นต้นของค่านามประสมชนิดไวยากรณ์ประโยค โดยใช้ค่านามหลักในค่านามประสมชนิดแสดงไวยากรณ์ประโยคมาใช้ในการค้นหา หลังจากนั้นเมื่อได้ตำแหน่งเริ่มต้นของคำแล้วให้ทำการหาตำแหน่งสิ้นสุดของคำ โดยสังเกตจากชนิดของคำที่เป็นส่วนหนึ่งในโครงสร้างของค่านามประสมชนิดแสดงไวยากรณ์ประโยคได้ คือ NN, NR, JJV, JJV และนำข้อมูลที่ได้กำกับขอบเขตของคำไปฝึกฝนแบบจำลอง CRFs

ตัวอย่างบทความใช้กฎในการรู้จำตามค่านามหลัก ค่านามหลักอาชีพ + โครงสร้างของค่านามประสมชนิดแสดงไวยากรณ์ประโยค (NN – JJV – NN – NN) แสดงในตารางที่ 4.2

ตารางที่ 4.2 ตัวอย่างบทความใช้กฎในการรู้จำตามค่านามหลัก

นาง	NN	0	0	0	0	0	0	S
สุตารัตน์	NR	0	0	0	0	0	0	S
เกยูราพันธุ์	NR	0	0	0	0	0	0	S
รัฐมนตรี	NN	1	0	0	0	0	0	B_SNP
ว่าการ	JJV	0	0	0	0	0	0	M_SNP
กระทรวง	NN	0	0	0	0	0	0	M_SNP
สาธารณสุข	NN	0	0	0	0	0	0	E_SNP
เปิดเผย	VV	0	0	0	0	0	0	S
ผล	NN	0	0	0	0	0	0	S
การ	FXN	0	0	0	0	0	0	S
ตรวจสอบ	VV	0	0	0	0	0	0	S
การ	FXN	0	0	0	0	0	0	S
ตรวจ	VV	0	0	0	0	0	0	S
แยก	VV	0	0	0	0	0	0	S
เชื้อ	NN	0	0	0	0	0	0	S
โรค	NN	0	0	0	0	0	0	S

## 4.2 ผลการทดลอง

ตัวอย่างผลการทดลองการระบุค่านามประสมชนิดแสดงไวยากรณ์ประโยคอาชีวะ ค่านามหลักเกี่ยวกับการแพทย์ ค่านามหลักเกี่ยวกับเครื่องมือ ค่านามหลักเกี่ยวกับสถานที่ ค่านามหลักเกี่ยวกับวัน เวลา แสดงในตารางที่4.3 – 4.7

ตารางที่4.3 ตัวอย่างผลการทดลองการระบุค่านามประสมชนิดแสดงไวยากรณ์ประโยคอาชีวะ

พาณิชย์[S] เดือน[S] **พ่อค้าขายเครื่องสังฆทาน[SNP]** ปิด[S] ป้าย[S] ราคา[S]  
 [S] ร.ต.อ.[S] เฉลิม[S] อยู่บำรุง[S] **รัฐมนตรีว่าการกระทรวงมหาดไทย[SNP]** กล่าว[S] ถึง[S]  
 กรณี[S] ที่[S] จะ[S] มี[S] การ[S] ปรับ[S] คณะ[S] รัฐมนตรี[S]  
 หลัง[S] การ[S] สอบสวน[S] จึง[S] ส่ง[S] ตัว[S] ให้[S] **พนักงานสอบสวน[SNP]** ดำเนิน[S] คดี[S]  
 ตาม[S] กฎหมาย[S] ต่อ[S] ไป[S]

ตารางที่4.4 ตัวอย่างผลการทดลองระบุค่านามหลักเกี่ยวกับการแพทย์

'[S] ออสซี่[S] '[S] พลิต[S] ได้[S] **ยาฆ่าหัตถ์นก[SNP]**  
 มี[S] เด็ก[S] เสีย[S] ชีวิต[S] ด้วย[S] **ไข้หัตถ์นกมรณะ[SNP]** ราย[S] วัน[S]  
 พ.ต.ท.[S] ทักษิณ[S] ออก[S] วิทย์[S] ยืนยัน[S] อีก[S] ครั้ง[S] ว่า[S] **โรคระบาด[SNP]** ใน[S] ไก่  
 [S] ไม่[S] ไข้[S] ไข้[S] หวัด[S] นก[S]

ตารางที่4.5 ตัวอย่างผลการทดลองระบุค่านามหลักเกี่ยวกับเครื่องมือ

ตำรวจ[S] ภูธร[S] จ.[S] บุรีรัมย์[S] บุ๊ก[S] ตรวจ[S] ยึด[S] **ไม้แปรรูป[SNP]** หวงห้าม[S] กว่า[S]  
 80[S] แผ่น[S]  
 นัก[S] พนัน[S] กำลัง[S] ช่วย[S] กัน[S] ทำลาย[S] หลักฐาน[S] ต่างๆ[S] โดย[S] ใส่[S] ไฟ[S] บัค  
 คารา[S] ลง[S] ใน[S] **เครื่องตัดกระดาษ[SNP]**  
 การ[S] แสดง[S] จะ[S] เน้น[S] เต็ม[S] ที่[S] เครื่อง[S] สาย[S] ตาม[S] ด้วย[S] สตรีง[S] และ[S]  
 ปิด[S] ด้วย[S] **เครื่องเป่า[SNP]**



ตารางที่ 4.6 ตัวอย่างผลการทดลองระบุค่านามหลักเกี่ยวกับสถานที่

เพื่อนบ้าน[S] ได้[S] เข้า[S] เยี่ยม[S] อาการ[S] ที่[S] ห้อง[S] ไอซียู[S] โรงพยาบาลค่ายสรรพลี ประสงค์[SNP]
ปตส.[S] -[S] ผอ.[S] ศูนย์[S] กระทรวง[S] ไอซีที[S] บุค[S] บ้านพักย่านพระประแดง[SNP] จับ[S] นัก[S] เรียน[S] หญิง[S] ม.[S] ต้น[S]
น้ำ[S] เอ่อ[S] เข้า[S] ท่วม[S] เขต[S] เทศบาล[S] เมือง[S] อุดรดิตต์[S] ซึ่ง[S] เป็น[S] ย่าน เศรษฐกิจสำคัญ[SNP] ของ[S] จังหวัด[S]

ตารางที่ 4.7 ตัวอย่างผลการทดลองระบุค่านามหลักเกี่ยวกับวัน เวลา

พ.ต.ท.[S] ทักษิณ[S] ชินวัตร[S] รัฐมนตรี[S] กล่าว[S] ถึง[S] สถานการณ์ไก่อตาย[SNP] ที่[S] เกรง [S] กัน[S] ว่า[S] จะ[S] เป็น[S] ไข้[S] หวัด[S] นก[S]
นอกจาก[S] นี้[S] นาง[S] สดศรี[S] ยัง[S] กล่าว[S] ถึง[S] เรื่อง[S] กำหนด[S] วันเลือกตั้ง[SNP] ด้วย[S]
น.ส.[S] ลัดดา[S] กล่าว[S] ว่า[S] งานมอเตอร์โชว์[SNP] มัก[S] พุด[S] กัน[S] ว่า[S] ไม่[S] ได้[S] ไป[S] ดู[S] รถ[S]

งานวิจัยนี้แบ่งการทดลองออกเป็น 8 แบบ โดยประเมินประสิทธิภาพของผลการทดลอง จาก การหาค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่า F-Measure ซึ่งมีผลการทดลอง ดังที่แสดงในตารางที่ 4.8 ต่อไปนี้

ตารางที่ 4.8 แสดงผลการทดลอง 8 แบบ

	Precision (%)	Recall (%)	F- Measure (%)
<b>กฎโครงสร้างประโยค</b>			
ขอบเขต 2 คำ (นาม กริยา)	93.75	92.38	93.06
ขอบเขต 3 คำ (นาม กริยา นาม)	78.86	77.53	78.18

	Precision (%)	Recall (%)	F- Measure (%)
ขอบเขต 4 คำ	89.66	88.74	89.19
<b>ชนิดคำนามหลัก</b>			
นามวลีเกี่ยวกับอาชีพ	81.77	79.10	80.41
นามวลีเกี่ยวกับ การแพทย์	80.93	79.25	80.08
นามวลีเกี่ยวกับ เครื่องมือ	81.54	80.67	81.10
นามวลีเกี่ยวกับสถานที่	77.18	76.00	76.58
นามวลีเกี่ยวกับวัน เวลา	69.15	66.92	68.01

#### 4.3 อภิปรายผลการทดลอง

จากผลการทดลองสรุปได้ว่า ประสิทธิภาพของระบบที่การพิจารณาโดยรวมภาพรวมทั้งหมด กฎโครงสร้างประโยค “นาม-กริยา” (ขอบเขต 2 คำ) มีความถูกต้อง (F-measure) มากที่สุด คือ 93.06 รองลงมาคือ กฎโครงสร้างประโยค ขอบเขต 4 คำ คือร้อยละ 89.19 ตามมาด้วยการพิจารณากฎในรูปแบบอื่นๆ มีค่าประสิทธิภาพที่ใกล้เคียงกันดังนี้ คำนามหลักเกี่ยวกับเครื่องมือ คำนามหลักเกี่ยวกับอาชีพ คำนามหลักเกี่ยวกับการแพทย์ “นาม-กริยา-นาม”(ขอบเขต 3 คำ) และ คำนามหลักเกี่ยวกับสถานที่ มีค่าร้อยละความถูกต้องเท่ากับ 81.10 80.41 80.08 78.18 และ 76.58 ตามลำดับ สำหรับคำนามหลักเกี่ยวกับวัน เวลา มีค่าประสิทธิภาพของระบบน้อยที่สุด เท่ากับร้อยละ 68.01 โดยพบว่าในโครงสร้างของคำประสมชนิดไวยากรณ์ประโยคในประเภทของคำนามหลักเกี่ยวกับวัน เวลานั้น จะพบชนิดของคำประเภทอื่นนอกเหนือจากคำนามและคำกริยาได้มากกว่าประเภทของคำนามหลักอื่น เช่น |งาน/NN|เผยแพร่/JJV|ความ/FXN|รู้/VV|ทาง/NN|กฎหมาย/NN|แก้/P|ประชาชน/NN| |พิธี/NN|ต้อนรับ/JJV|คณะ/NN|ผู้/FXN|แทน/VV| เป็นต้น

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงการสรุปผลการวิจัย การรู้จำคำนามประสมชนิดไวยากรณ์ประโยค จากคลังข้อมูล BEST พิจารณาเฉพาะข่าวภาษาไทย และข้อเสนอแนะหรือแนวทางการพัฒนางานวิจัยในอนาคต

#### 5.1 สรุปผลงานวิจัย

งานวิจัยนี้นำเสนอการพัฒนาระบบการรู้จำคำนามประสมชนิดแสดงไวยากรณ์ประโยค โดยใช้แบบจำลองคอนดิชันนอลแรนดอมฟลตสโมเดล (CRFs) และศึกษาเปรียบเทียบประสิทธิภาพของระบบการรู้จำคำนามประสมชนิดแสดงไวยากรณ์ประโยค โดยใช้แบบจำลองที่พิจารณาโครงสร้างประโยค และคำนามหลัก รวมทั้งสิ้น 7 รูปแบบ งานวิจัยนี้ใช้คลังข้อมูล BEST เฉพาะหมวดข่าวภาษาไทยขนาด 1,000 คำ แบบจำลองที่ใช้คือ CRFSharp จากผลการทดสอบระบบ พบว่าประสิทธิภาพของระบบที่การพิจารณาโดยรวมภาพรวมทั้งหมด กฎโครงสร้างประโยค “นาม-กริยา” (ขอบเขต 2 คำ) มีความถูกต้อง (F-measure) มากที่สุด คือร้อยละ 93.06 รองลงมาคือ กฎโครงสร้างประโยค ขอบเขต 4 คำ คือร้อยละ 89.19 ตามมาด้วยการพิจารณากฎในรูปแบบอื่นๆ มีค่าประสิทธิภาพที่ใกล้เคียงกันดังนี้ คำนามหลักเกี่ยวกับเครื่องมือ คำนามหลักเกี่ยวกับอาชีพ คำนามหลักเกี่ยวกับการแพทย์ “นาม-กริยา-นาม”(ขอบเขต 3 คำ) และคำนามหลักเกี่ยวกับสถานที่ มีค่าร้อยละความถูกต้องเท่ากับ 81.10 80.41 80.08 78.18 และ 76.58 ตามลำดับ สำหรับคำนามหลักเกี่ยวกับวัน เวลา มีค่าประสิทธิภาพของระบบน้อยที่สุด เท่ากับร้อยละ 68.01

#### 5.2 ข้อเสนอแนะ

สำหรับงานวิจัยนี้ได้นำเสนอเกี่ยวกับการนำแบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ เพื่อใช้ในการรู้จำคำประสมที่มีโครงสร้างเหมือนประโยค และได้้นำโครงสร้างประโยค และคำนามหลักประเภทต่างๆ มาใช้ในการค้นหาขอบเขตของคำ อย่างไรก็ตามผลลัพธ์ที่ได้ยังมีความผิดพลาดอยู่ จากปัญหาที่เกิดจากการหาขอบเขตของคำไม่ครบถ้วนตัวอย่างเช่น คำประสมที่มีโครงสร้างเหมือนประโยคมีเครื่องหมายหรือตัวเลข คำประสมมีคำเชื่อม และคำประสมมีคำช่องว่าง เป็นต้น

จากข้อจำกัดดังกล่าว ทำให้งานวิจัยนี้สามารถนำไปศึกษาหรือพัฒนาเพิ่มเติมเพื่อให้ผลการทดลองมีความถูกต้องมากยิ่งขึ้น ได้แก่

1. ศึกษาเพิ่มเติมในส่วนของโครงสร้างของค่านามประสมชนิดแสดงไวยากรณ์ประโยค ซึ่งผู้วิจัยพบว่าภายในโครงสร้างของค่านามประสมชนิดแสดงไวยากรณ์ประโยคยังสามารถพบชนิดของคำประเภทอื่นได้ นอกเหนือจากค่านามและคำกริยา

2. พัฒนาเพิ่มเติมในส่วนของประเภทของค่านามหลักที่พบได้ในโครงสร้างของค่านามประสมชนิดแสดงไวยากรณ์ประโยค เพื่อให้สามารถกำหนดขอบเขตของโครงสร้างประสมชนิดแสดงไวยากรณ์ประโยคได้ถูกต้องและแม่นยำมากยิ่งขึ้น



## รายการอ้างอิง

### หนังสือและบทความในหนังสือ

อัญชลี สิงห์น้อย,ค่านามประสมศาสตร์และศิลป์ในการสร้างคำไทย,(กรุงเทพมหานคร: สำนักพิมพ์แห่ง  
จุฬาลงกรณ์มหาวิทยาลัย,2548) น. 39-55 ,108-119

### บทความวารสาร

Anderson, Stephen R.1985. “Grammatical categories.” In Shopen, T. 9ed) Language  
Typology and Syntactic Description, 150-201.

Haruechaiyasak, C., Kongyoung, S., and Dailey, M. N. 2008. A Comparative Study on  
Thai Word Segmentation Approach. In Proceedings of ECTI-CON. Krabi.

Kruengkrai, C., Sornlertlamvanich, V., and Isahara, H. 2006. A Conditional Random  
Field Framework for Thai Morphological Analysis. In Proceeding of the Fifth  
International Conference on Language Resources and Evaluation. Genoa.

Lafferty, J., McCallum, A. and Pereira, F. 2001. Conditional Random Fields:  
Probabilistic Models for Segmenting and Labeling Sequence Data. Proceedings  
of the 18th International Conference on Machine Learning 2001 (ICML 2001),  
pages 282-289

Lieber, Rochelle.1992. Deconstructing Morphology.Chicago:The University of Chicago  
Press.

Sadock, Jerrold.1998. “On the autonomy of compounding morphology.” In  
Lapointe,S.J.,Brentari, D.K. and Farrell ,P.M. (eds) Morphology and Its Relation  
to Phonology and Syntax. Stamford, California: CALI Publications 161-187.

### สื่ออิเล็กทรอนิกส์

คลังข้อความ BEST สืบค้นจาก <http://thailang.nectec.or.th/downloadcenter>

## ประวัติผู้เขียน

ชื่อ	นายภูริเดช โพธิ์พัฒน์
วันเดือนปีเกิด	30 ตุลาคม พ.ศ. 2527
ตำแหน่ง	วิศวกร สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยี แห่งชาติ (สวทช.)
ทุนการศึกษา (ถ้ามี)	-
ผลงานทางวิชาการ	-
ประสบการณ์ทำงาน	พ.ศ. 2555 วิศวกร สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) พ.ศ. 2552 โปรแกรมเมอร์ บริษัท Humanica พ.ศ. 2551 Application Developer Venda Software Development