

สำนักหอสมุดกลาง ' พระจอมเกล้าลาดกระบัง

การวิเคราะห์ประโยคภาษาไทยจากย่อหน้าเพื่อการแปลภาษาด้วย
คอมพิวเตอร์

THAI SYNTACTICAL ANALYSIS SYSTEM BY METHOD OF
SPLITTING SENTENCES FROM PARAGRAPH FOR MACHINE
TRANSLATION



นายสังกรศรัณย์ ล่องชุมผล
MR. SUNGKORNSARUN LONGCHUPOLE

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2538

ISBN 974- 621-463-2

ลิขสิทธิ์ของบัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เก็บเลขหมู่.....ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ เลขทะเบียน.....24783.....งเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
วัน, เดือน, ปี.....18 ส.ค. 2539

THAI SYNTACTICAL ANALYSIS SYSTEM BY METHOD OF
SPLITTING SENTENCES FROM PARAGRAPH FOR MACHINE
TRANSLATION



A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE
MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING
GRADUATE SCHOOL
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

1995

ISBN 974 - 621- 463 -2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับฉำรงใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดะทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีกรำนำไปใช้

หัวข้อวิทยานิพนธ์

การวิเคราะห์ประโยคภาษาไทยจากย่อหน้าเพื่อการแปล
ภาษาด้วยคอมพิวเตอร์

นักศึกษา

นายคังกรศรัณย์ ล่องชุมผล

อาจารย์ผู้ควบคุมวิทยานิพนธ์

รศ.ดร.รัตติกว วรากุลศิริพันธ์ุ

อาจารย์ผู้ควบคุมวิทยานิพนธ์ร่วม

ผศ.สุชุมลย์ นิลรัตน์

ระดับการศึกษา

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

ภาควิชา

อิเล็กทรอนิกส์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง

พ.ศ.

2538

บทคัดย่อ

ในระบบการแปลภาษาจากภาษาต้นฉบับไปเป็นภาษาเป้าหมายด้วยคอมพิวเตอร์นั้น ในยุคแรกจะประกอบไปด้วยการวิเคราะห์ประโยคต้นแบบที่จะแปล แล้วทำความเข้าใจในโครงสร้างและความหมายของประโยค หลังจากนั้นจะทำการค้นหาคำศัพท์จาก พจนานุกรมซึ่งเก็บคู่คำศัพท์ของภาษาต้นฉบับและภาษาเป้าหมายที่มีความหมายตรงกัน เมื่อได้คำของภาษาเป้าหมายครบถ้วนแล้วก็จะนำคำของภาษาเป้าหมายมาเรียงกัน เป็นประโยคตามหลักไวยากรณ์ของภาษาเป้าหมายนั้น ต่อมาได้มีการพัฒนาการแปลภาษาให้สามารถแปลไปเป็นภาษาเป้าหมายได้หลายภาษา โดยทำการแปลภาษาต้นฉบับเป็นภาษากลางก่อน จากนั้นก็นำประโยคที่เป็นภาษากลางไปสังเคราะห์เป็นภาษาเป้าหมายที่ต้องการอีกครั้งหนึ่ง

การวิเคราะห์ข้อความภาษาไทยเพื่อการแปลเป็นภาษาเป้าหมายนั้น จะทำการวิเคราะห์ครั้งละประโยค แล้วจะทำการแปลเป็นภาษาเป้าหมาย เนื่องจากภาษาไทยมีลักษณะเป็นการใช้คำมาเขียนติดกัน โดยไม่มีเครื่องหมายหรือช่องว่างบอกการจบหน่วยคำในประโยค ยิ่งไปกว่านั้นก็มีการนำเอาประโยคภาษาไทยหลายประโยคมาผสมกันเป็นย่อหน้า แล้วละคำที่ซ้ำกันไว้เพื่อความสวยงามของย่อหน้านั้น เช่นการละประธานที่ซ้ำกันของประโยคภาษาไทยที่นำมาเชื่อมกันโดยละประธานในประโยคภาษาไทยที่อยู่หลังคำเชื่อมเป็นต้น ปัญหาจากการที่ข้อความภาษาไทยมีประโยคหลายประโยคมาเชื่อมกันเป็นย่อหน้าและบางประโยคมีการละคำไว้ ทำให้ระบบการแปลภาษาด้วยคอมพิวเตอร์ ไม่สามารถวิเคราะห์ข้อความด้วยหลักโครงสร้างของประโยคเดียวได้

วิทยานิพนธ์นี้จะเสนอ ระบบวิเคราะห์ประโยคภาษาไทยจากย่อหน้าเพื่อการแปลภาษาด้วยคอมพิวเตอร์ โดยใช้หลักการแยกย่อหน้าออกเป็นคำก่อน แล้วหาคำกริยาเพื่อเป็นหลักในการวิเคราะห์หาจำนวนประโยคภาษาไทยที่มีในย่อหน้านั้น หลังจากนั้นหาคำสันธานที่เชื่อมประโยคภาษาไทยที่อยู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในย่อหน้า แล้วตรวจสอบประโยคภาษาไทยเหล่านั้นว่าแต่ละประโยคถูกต้องตามหลักไวยากรณ์หรือไม่ ถ้าถูกต้องก็จะแยกออกมาเป็นประโยค ผลจากการแยกประโยคภาษาไทยออกจากย่อหน้า จะทำให้สามารถวิเคราะห์ความหมายของประโยคภาษาไทยในย่อหน้าด้วยคอมพิวเตอร์ได้ความหมายครบสมบูรณ์



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Thai Syntactical Analysis System by Method of Splitting Sentences from Paragraph for Machine Translation
Student	Mr.Sungkornsarun Longchupole
Thesis Advisor	Assoc.Prof.Dr.Ruttikorn Varakulsiripunth
Thesis Co-advisor	Asst.Prof.Sukhuman Nilrat
Level of Study	Master of Engineering in Electrical Engineering
Department	Electronics, King Mongkut's Institute of Technology Ladkrabang
Year	1995

Abstract

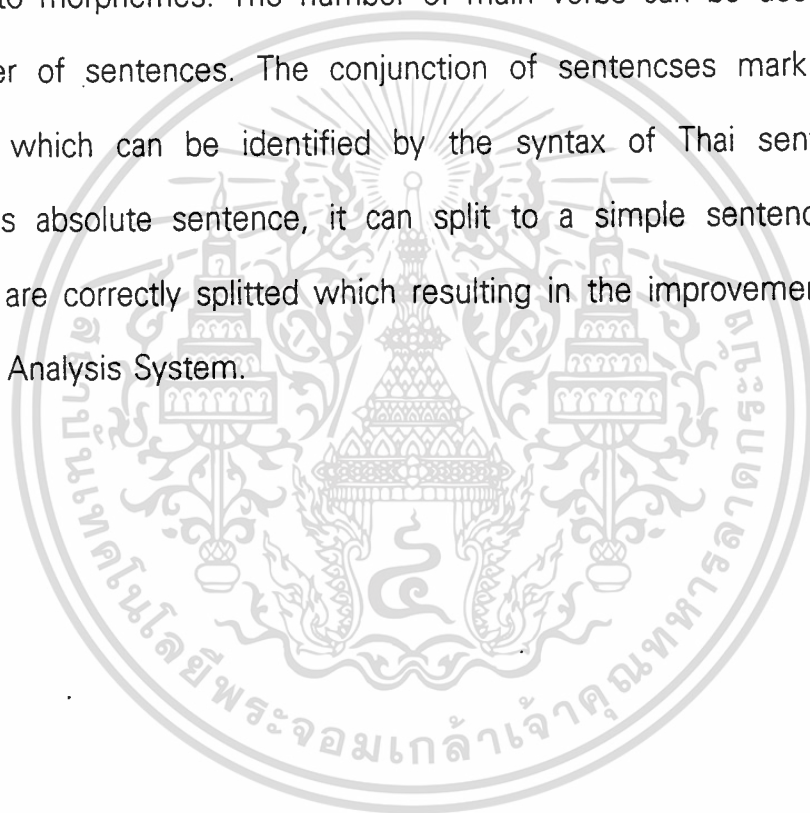
In Machine Translation system, the common method used to translate source language sentences to target language sentences is the analysis of the source sentences and finds out their meaning. Then searching for words with the same meaning in target language to generate the target sentence in structure of such target language. The continually development of Machine Translation leads to the capability of translation of Thai language to many other languages. One of the efficient method is by translate Thai sentence to interlingua sentence and finally uses interlingua sentence to translate to any target language sentence.

An analysis methodology on Thai syntax, translate Thai language to target language, is doing on each sentence one at a time. Since Thai language is agglutinative language, there is neither sign nor space which can separate

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

morphemes in the sentence. Moreover, many Thai sentences are found written one after another in a paragraph and many repeated words are neglected. For example, the neglect of redundancy subject of the sentence in a complex sentence. This cause the problems of the analysis of Thai language which cannot analyse a complex sentence by the method of the simple sentence.

This thesis proposes The Thai Syntactical Analysis System by Method of Splitting Thai Sentences from Paragraph. The method begins with splitting a paragraph to morphemes. The number of main verbs can be used to estimate the number of sentences. The conjunction of sentences mark the possible sentences which can be identified by the syntax of Thai sentence. If any sentence is absolute sentence, it can split to a simple sentence. Finally, all sentences are correctly splitted which resulting in the improvement of the Thai Syntactical Analysis System.



กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงด้วยดี เนื่องจากบุคคล และหน่วยงานต่อไปนี้ได้ช่วยสนับสนุน ด้วยดีตลอดมา ผู้เขียนขอขอบพระคุณทุกท่านดังต่อไปนี้

คุณพ่อ คุณแม่ ผู้ซึ่งคอยเป็นกำลังใจ และห่วงใยตลอดมา

รศ.ดร.รัตติกร วรากุลศิริพันธ์ และ ผศ.สุขุมลย์ นิลรัตน์ ผู้ซึ่งสละเวลาอันมีค่า คอยให้คำปรึกษาและแนะนำแนวทางที่เป็นประโยชน์เสมอมา

รศ.ประภิจ ดั่งติส้านนท์ ผู้ให้แนวทางในการทำงานให้ถึงที่สุด

Center of the International Cooperation for Computerization (CICC) JAPAN, ศูนย์วิจัยและพัฒนาปัญญาประดิษฐ์ สถาบันเทคโนโลยีพระจอมเกล้าธนบุรี, ห้องปฏิบัติการวิจัยภาษาและวิทยาการความรู้ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ กระทรวงวิทยาศาสตร์ ที่สนับสนุนข้อมูลวิจัย

มูลนิธิเพื่อการศึกษาคอมพิวเตอร์และการสื่อสาร และศูนย์เทคโนโลยีคอมพิวเตอร์และอิเล็กทรอนิกส์แห่งชาติ ที่สนับสนุนทุนวิจัย

สำนักวิจัยและบริการคอมพิวเตอร์ และ คณะวิศวกรรมศาสตร์ ที่สนับสนุนอุปกรณ์และสถานที่ทำการวิจัย

รวมทั้งทุกท่านที่ไม่มีโอกาสกล่าวนามมาในที่นี้

คังกรศรัณย์ ล่องซุนผล

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	III
กิตติกรรมประกาศ.....	V
สารบัญ.....	VI
สารบัญตาราง.....	VIII
ตาราง.....	VIII
สารบัญรูป.....	IX
รูป.....	IX
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของงานวิจัย.....	1
1.2 จุดประสงค์และขอบเขตของงานวิจัย.....	5
1.3 โครงร่างของวิทยานิพนธ์.....	6
บทที่ 2 ฐานความรู้เกี่ยวกับไวยากรณ์ภาษาไทยที่ใช้ในการวิเคราะห์.....	7
2.1 บทนำ.....	7
2.2 การแบ่งประเภทของคำภาษาไทย.....	7
2.3 โครงสร้างของประโยคภาษาไทย.....	15
2.4 โครงสร้างของย่อหน้าภาษาไทย.....	21
บทที่ 3 การจัดเก็บฐานข้อมูลไวยากรณ์ภาษาไทยในส่วนของพจนานุกรมคำศัพท์.....	27
3.1 บทนำ.....	27
3.2 ส่วนข้อมูลทางโครงสร้างและความหมายของคำศัพท์.....	28
3.3 โครงสร้างข้อมูลของฐานความรู้พจนานุกรม.....	46
3.4 การใช้งานฐานความรู้พจนานุกรม.....	55
บทที่ 4 กระบวนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า.....	60
4.1 บทนำ.....	60
4.2 ส่วนการวิเคราะห์หน่วยคำ.....	60
4.3 ส่วนการวิเคราะห์โครงสร้างของประโยคภาษาไทย.....	73
4.4 ส่วนการวิเคราะห์แยกประโยคจากย่อหน้า.....	90

บทที่ 5 ซอฟต์แวร์ของระบบ.....	94
5.1 บทนำ.....	94
5.2 ซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับหน่วยค่า	94
5.3 ซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทย.....	97
5.4 ซอฟต์แวร์ของระบบส่วนการวิเคราะห์แยกประโยคจากย่อหน้า.....	100
5.5 การใช้งานซอฟต์แวร์ระบบ.....	102
บทที่ 6 บทสรุป.....	112
6.1 บทนำ.....	112
6.2 ผลการทดสอบ.....	114
6.3 การนำผลงานวิจัยไปประยุกต์ใช้กับงานอื่นๆ.....	116
6.4 แนวทางในการพัฒนาต่อของระบบ.....	116
เอกสารอ้างอิง.....	117
ภาคผนวก ก ข้อมูลคำศัพท์.....	119
ภาคผนวก ข ข้อมูลผลการทดลอง.....	126
ประวัติผู้เขียน.....	155

สารบัญตาราง

	หน้า
ตารางที่ 3-1 แสดงโครงสร้างของระบบฐานความรู้พจนานุกรมคำศัพท์.....	28
ตารางที่ 3-2 แสดงการกำหนดค่าดัชนีตัวที่ 1 ของกลุ่มคำศัพท์ภาษาไทย.....	50
ตารางที่ 3-3 แสดงให้เห็นถึงโครงสร้างของการแบ่งคำศัพท์ชั้นที่ 2 ภายในกลุ่ม คำศัพท์ชั้นที่ 1.....	51
ตารางที่ 4-1 แสดงผลลัพธ์ที่ได้จากตัวอย่างการวิเคราะห์หน่วยคำ.....	70
ตารางที่ 4-2 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์หน่วยคำ.....	85
ตารางที่ 4-3 แสดงข้อมูลในระดับโครงสร้างประโยคของหน่วยคำ.....	85
ตารางที่ 4-4 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์ประโยคภาษาไทยจากย่อหน้า เซตที่ 1.....	88
ตารางที่ 4-5 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์ประโยคภาษาไทยจากย่อหน้า เซตที่ 2.....	89
ตารางที่ 4-6 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์ประโยคภาษาไทยจากย่อหน้า เซตที่ 3.....	89
ตารางที่ 4-7 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์ประโยคภาษาไทยจากย่อหน้า เซตที่ 4.....	89
ตารางที่ 6-1 แสดงกฎไวยากรณ์ใช้วิเคราะห์ประโยคภาษาไทย.....	113
ตารางที่ 6-2 แสดงผลการทดสอบระบบวิเคราะห์แยกประโยคภาษาไทย จากย่อหน้า.....	114

สารบัญรูป

	หน้า
รูปที่ 1.1 แสดงโครงสร้างของการแปลภาษาด้วยคอมพิวเตอร์.....	2
รูปที่ 1.2 แสดงการแปลภาษาโดยใช้สื่อภาษากลาง.....	3
รูปที่ 1.3 แสดงโครงสร้างของระบบวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า.....	5
รูปที่ 2.1 แสดงกฎไวยากรณ์ของนามวลีภาษาไทย.....	20
รูปที่ 2.2 แสดงกฎไวยากรณ์ของกริยาวลีภาษาไทย.....	21
รูปที่ 2.3 แสดงกฎไวยากรณ์ของกริยาวลีภาษาไทย (ต่อ).....	22
รูปที่ 2.4 แสดงกฎไวยากรณ์ของบุรพพทวลีภาษาไทย.....	22
รูปที่ 2.5 แสดงโครงสร้างของประโยคความเดียวในภาษาไทย.....	23
รูปที่ 2.6 แสดงโครงสร้างประโยคภาษาไทยด้วยไวยากรณ์การก.....	25
รูปที่ 2.7 แสดงหน่วยคำหลักและหน่วยพึ่งพาในโครงสร้างประโยคภาษาไทย.....	26
รูปที่ 3.1 แสดงโครงสร้างของระบบพจนานุกรมคำศัพท์และส่วนต่อเชื่อมกับระบบอื่น.....	28
รูปที่ 3.2 แสดงข้อมูลความหมายคำกับข้อมูลระดับโครงสร้างของคำและข้อมูลระดับ ความหมายของคำ.....	29
รูปที่ 3.3 แสดงรูปคำกริยากลุ่มที่ 1.....	30
รูปที่ 3.4 แสดงรูปคำกริยากลุ่มที่ 2.....	31
รูปที่ 3.5 แสดงรูปคำกริยากลุ่มที่ 3.....	31
รูปที่ 3.6 แสดงรูปคำกริยากลุ่มที่ 4.....	32
รูปที่ 3.7 แสดงรูปคำกริยากลุ่มที่ 5.....	32
รูปที่ 3.8 แสดงรูปคำกริยากลุ่มที่ 6.....	33
รูปที่ 3.9 แสดงรูปคำกริยากลุ่มที่ 7.....	33
รูปที่ 3.10 แสดงรูปคำกริยากลุ่มที่ 8.....	34
รูปที่ 3.11 แสดงรูปประโยคภาษาที่ได้ถูกวิเคราะห์แล้วโดยแสดงออกมาในรูปของ หน่วยคำและความสัมพันธ์ระหว่างหน่วยคำ.....	45
รูปที่ 3.12 แสดงโครงสร้างดัชนีของฐานความรู้พจนานุกรมคำศัพท์ภาษาไทย.....	47
รูปที่ 3.13 แสดงโครงสร้างดัชนีของพจนานุกรมคำศัพท์ภาษาไทย (ในรูปแบบของ แผนภูมิต้นไม้).....	48
รูปที่ 3.14 แสดงความสัมพันธ์ของการแบ่งกลุ่มคำศัพท์ภาษาไทยชั้นที่ 1.....	49

รูปที่ 5.9 แสดงการใช้งานซอฟต์แวร์ระบบส่วนของการวิเคราะห์ระดับหน่วยคำ.....	105
รูปที่ 5.10 แสดงผลลัพธ์ที่ได้จากซอฟต์แวร์ระบบส่วนของการวิเคราะห์หน่วยคำ.....	105
รูปที่ 5.11 แสดงการเปิดหน้าต่างแสดงผลส่วนการวิเคราะห์โครงสร้าง ประโยคภาษาไทย.....	106
รูปที่ 5.12 แสดงการใช้งานซอฟต์แวร์ระบบส่วนการวิเคราะห์โครงสร้าง ประโยคภาษาไทย.....	107
รูปที่ 5.13 แสดงผลลัพธ์ที่ได้จากซอฟต์แวร์ระบบส่วนการวิเคราะห์โครงสร้าง ประโยคภาษาไทย.....	107
รูปที่ 5.14 แสดงการเปิดหน้าต่างแสดงผลของซอฟต์แวร์ระบบส่วนวิเคราะห์ แยกประโยคภาษาไทยจากย่อหน้า.....	108
รูปที่ 5.15 แสดงการใช้งานซอฟต์แวร์ระบบของส่วนวิเคราะห์แยกประโยค ภาษาไทยจากย่อหน้า.....	109
รูปที่ 5.16 แสดงผลลัพธ์ที่ได้จากซอฟต์แวร์ระบบของส่วนแยกประโยค ภาษาไทยจากย่อหน้า.....	109
รูปที่ 5.17 แสดงข้อผิดพลาดที่เกิดจากการป้อนข้อมูลอินพุตผิดพลาด.....	110
รูปที่ 5.18 แสดงการออกจากระบบแยกประโยคภาษาไทยจากย่อหน้า.....	111

รูปที่ 3.15 แสดงความสัมพันธ์ของการแบ่งกลุ่มคำศัพท์ภาษาไทยชั้นที่ 1 (ในรูปแบบของแผนภูมิต้นไม้).....	50
รูปที่ 3.16 แสดงความสัมพันธ์ของการแบ่งกลุ่มคำศัพท์ชั้นที่ 2 ภายในคำศัพท์ชั้นที่ 1.....	52
รูปที่ 3.17 แสดงความสัมพันธ์ของการแบ่งกลุ่มคำศัพท์ชั้นที่ 2 ภายในคำศัพท์ ชั้นที่ 1 (ในรูปแบบของแผนภูมิต้นไม้).....	53
รูปที่ 3.18 แสดงความสัมพันธ์ของการแบ่งคำศัพท์ทั้ง 3 ชั้น.....	54
รูปที่ 3.19 แสดงผังงานของโปรแกรมการตรวจสอบข้อมูลกับฐานข้อมูล พจนานุกรมคำศัพท์.....	58
รูปที่ 3.20 แสดงข้อมูลที่ถูกส่งไปยังระบบวิเคราะห์โครงสร้างประโยคภาษาไทย.....	59
รูปที่ 4.1 แสดงการทำงานต่อเนื่องกันของระบบวิเคราะห์แยกประโยค ภาษาไทยจากย่อหน้า.....	61
รูปที่ 4.2 แสดงผลลัพธ์ที่เป็นไปได้ของการวิเคราะห์แยกหน่วยคำ.....	63
รูปที่ 4.3 แสดงกระบวนการวิเคราะห์หน่วยคำ.....	65
รูปที่ 4.4 แสดงขั้นตอนของระบบวิเคราะห์แบ่งคำ.....	71
รูปที่ 4.5 แสดงตัวอย่างข้อมูลในการวิเคราะห์หน่วยคำ.....	72
รูปที่ 4.6 แสดงการเริ่มต้นค้นหาหน่วยคำหลักของประโยคภาษาไทย.....	73
รูปที่ 4.7 แสดงการค้นหาหน่วยคำหลักของประโยคภาษาไทย.....	74
รูปที่ 4.8 แสดงการใช้ไวยากรณ์โครงสร้างวลีตรวจสอบโครงสร้างของประโยค ภาษาไทยในย่อหน้า.....	75
รูปที่ 4.9 แสดงการตรวจสอบหาความสัมพันธ์ระหว่างหน่วยคำหลักกับ หน่วยคำข้างเคียง.....	76
รูปที่ 5.1 แสดงผังงานของการจัดเตรียมข้อมูล.....	95
รูปที่ 5.2 แสดงไฟล์ชาร์ตของระบบวิเคราะห์แบ่งคำ.....	96
รูปที่ 5.3 แสดงผังงานของซอฟต์แวร์ของระบบส่วนการวิเคราะห์หาหน่วยคำหลัก.....	98
รูปที่ 5.4 แสดงการตรวจสอบโครงสร้างประโยคด้วยไวยากรณ์.....	99
รูปที่ 5.5 แสดงผังงานของส่วนวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า.....	101
รูปที่ 5.6 แสดงส่วนติดต่อกับผู้ใช้ของระบบแยกประโยคภาษาไทยจากย่อหน้า.....	103
รูปที่ 5.7 แสดงการป้อนข้อมูลอินพุทของข้อความย่อหน้าต้นแบบ.....	104
รูปที่ 5.8 แสดงการเปิดหน้าต่างแสดงผลของส่วนการวิเคราะห์หน่วยคำ.....	104

บทที่ 1

บทนำ

1.1 ความเป็นมาของงานวิจัย

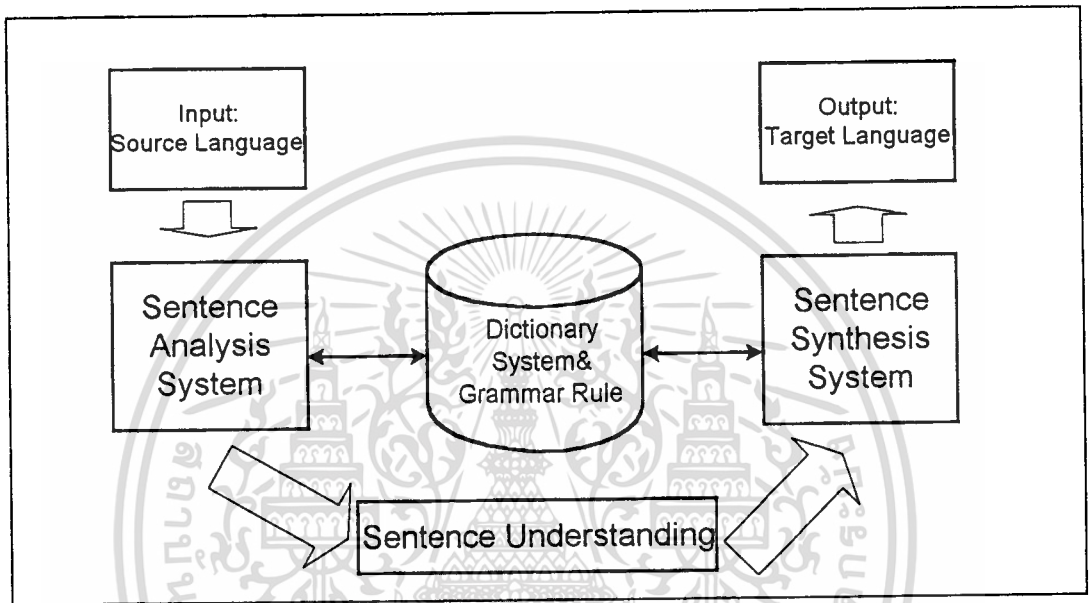
ภาษาเป็นสื่อกลางที่มนุษย์ใช้ในการสื่อสารระหว่างกัน เพื่อถ่ายทอดความคิดและความต้องการของตัวเองให้ผู้อื่นเข้าใจ ด้วยความที่มนุษย์อยู่กระจายไปทั่วโลก ภาษาที่ใช้แต่ละแห่งจึงมีความแตกต่างกันอย่างมากมายทั้งภาษาพูดที่ใช้ในการสนทนาและภาษาเขียนที่ใช้เพื่อเก็บบันทึกเป็นอักษร เมื่อเทคโนโลยีถูกพัฒนามากขึ้นมนุษย์ก็มีความสามารถที่จะสร้างพาหนะให้เดินทางไปได้ทั่วโลก ทำให้เกิดกิจกรรมต่างๆมากมายเช่น การแลกเปลี่ยนความรู้ด้านวัฒนธรรม วิชาการ การเดินทางท่องเที่ยว การค้าขายแลกเปลี่ยนสินค้าระหว่างภูมิภาคต่างๆ เป็นผลให้การสื่อสารภาษาที่มีความจำเป็นสูงขึ้นไปอย่างหลีกเลี่ยงไม่ได้ แต่ด้วยเหตุที่ทั่วโลกมีภาษาที่แตกต่างกันอยู่อย่างมากมาย การติดต่อสื่อสารจากชนชาติหนึ่งไปยังชนชาติอื่นที่ใช้ภาษาต่างกันจะต้องมีการแปลภาษาจากภาษาหนึ่งไปยังอีกภาษาหนึ่ง จึงจะสามารถติดต่อกันได้อย่างเข้าใจ การแปลภาษาจึงเป็นกุญแจสำคัญอันหนึ่งในการดำเนินชีวิตประจำวันของมนุษย์

การแปลภาษาด้วยระบบซอฟต์แวร์คอมพิวเตอร์ ได้ถูกทุ่มเทวิจัยมาเป็นเวลานาน โดยมีวัตถุประสงค์หลักเพื่อช่วยในงานแปลภาษาที่มีปริมาณมาก และต้องการทดแทนการแปลภาษาด้วยมนุษย์ ซึ่งมีค่าใช้จ่ายสูงและมีความจำกัดในด้านขีดความสามารถที่จะทำงานที่มีปริมาณมากๆ หลักของการแปลภาษาด้วยคอมพิวเตอร์จะใช้ทฤษฎีทางการประมวลผลภาษาธรรมชาติ (Natural Language Processing) ร่วมกับวิทยาการทางซอฟต์แวร์คอมพิวเตอร์ (Computer Software Technology) และภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistics) ซึ่งเป็นการจำลองการทำงานด้านการแปลภาษาของมนุษย์ คือ การรับข้อความของภาษาต้นแบบ (Source Language) นำมาทำการวิเคราะห์ (Analysis) โครงสร้างของประโยคต้นแบบ โดยใช้หลักการทางภาษาศาสตร์ เก็บไว้ในลักษณะของความเข้าใจเนื้อหาของประโยค (Sentence Understanding) จากนั้นก็ถ่ายทอดกลับออกมาเป็นภาษาเป้าหมาย (Target Language) โดยใช้การสังเคราะห์ (Synthesis) ประโยคที่มีเนื้อความในด้านความหมายเดียวกับประโยคต้นแบบ ด้วยหลักโครงสร้างประโยคของภาษาเป้าหมาย ออกมาเป็นผลลัพธ์การแปลภาษา ดังแสดงให้เห็นในรูปที่ 1.1

การแปลภาษาด้วยคอมพิวเตอร์จะทำให้งานด้านการแปลภาษาสามารถทำได้ในปริมาณมากๆ และด้วยคุณสมบัติของคอมพิวเตอร์ที่ทำงานได้เป็นระยะเวลาอันยาวนานกว่ามนุษย์ จึงสามารถทำงานได้อย่าง

เอกลีกรีนเป็นเอกลีกรีนที่สงวนไว้สำหรับกิจการเพื่อการศึกษาเท่านั้น เมื่ออนุญาตให้เข้าไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

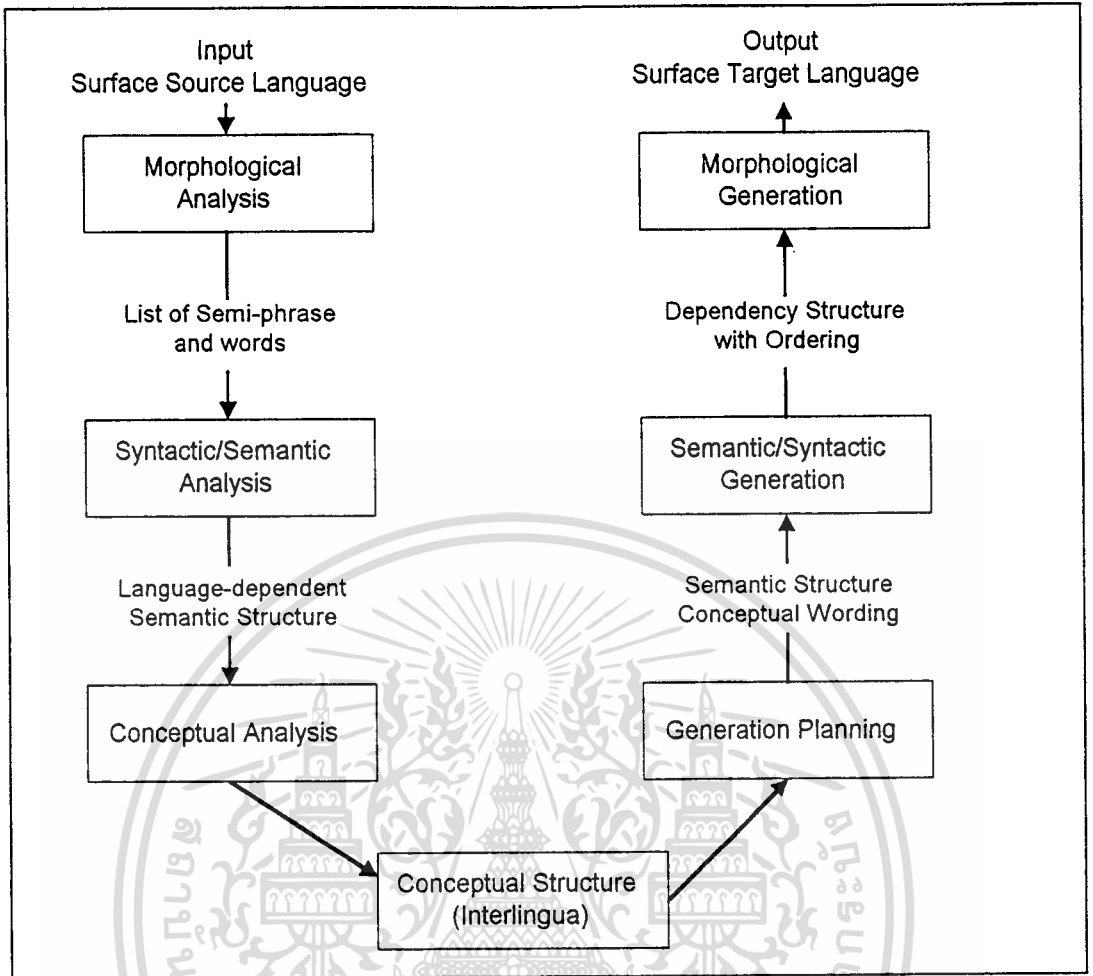
ต่อเนื่องทำให้งานเสร็จเร็วขึ้น นอกจากนี้พัฒนาการทางด้านคอมพิวเตอร์ที่ทำให้เครื่องคอมพิวเตอร์มีประสิทธิภาพสูงขึ้นแต่มีขนาดเล็กลง ก็ทำให้การใช้งานสะดวกขึ้น ดังจะเห็นได้จากพจนานุกรมอิเล็กทรอนิกส์ ซึ่งเป็นส่วนหนึ่งของงานด้านการแปลภาษาถูกสร้างให้มีขนาดเล็กจนสามารถพกพาไปด้วยได้ ดังนั้นในอนาคตจึงมีความเป็นไปได้ที่เครื่องคอมพิวเตอร์สำหรับการแปลภาษาจะสามารถพกพาไปใช้ได้ในทุกสถานที่



รูปที่ 1.1 แสดงโครงสร้างของการแปลภาษาด้วยคอมพิวเตอร์

ในประเทศไทยได้มีการเริ่มศึกษาวิจัยด้านการแปลภาษาด้วยคอมพิวเตอร์ขึ้นมาตั้งแต่ปีพ.ศ. 2524 โดยเป็นโครงการวิจัยร่วมของทบวงมหาวิทยาลัย กับมหาวิทยาลัย Grenoble ของประเทศฝรั่งเศส และมหาวิทยาลัย Sains ของประเทศมาเลเซีย โดยมีวัตถุประสงค์ที่จะทำการแปลภาษาจากภาษาอังกฤษเป็นภาษาไทย^[1] จนกระทั่งปัจจุบันนี้ การวิจัยด้านการแปลภาษาด้วยคอมพิวเตอร์ของประเทศไทยได้ประสบผลสำเร็จขั้นหนึ่งแล้ว นั่นคือการร่วมมือทำการวิจัยระหว่างกระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงานของประเทศไทยกับ Ministry of International Trade and Industry ของประเทศญี่ปุ่น ร่วมกับอีก 3 ประเทศในเอเชียคือ จีน อินโดนีเซีย และมาเลเซีย สร้างเครื่องแปลภาษาซึ่งสามารถแปลภาษาระหว่างกันได้ 5 ภาษาในระบบเดียวกัน โดยใช้สื่อภาษากลาง (Interlingua) เป็นตัวกลางในการถ่ายทอดข้อมูลโครงสร้างทางไวยากรณ์ และความหมายของประโยคภาษาต้นแบบไปสู่การสร้างประโยคภาษาเป้าหมาย ดังแสดงในรูปที่ 1.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 1.2 แสดงการแปลภาษาโดยใช้สื่อภาษากลาง

หลักการทํางานของระบบแปลภาษาโดยใช้สื่อภาษากลางจะแบ่งการทํางานออกเป็นสองส่วนหลัก^[2] คือ

ส่วนที่หนึ่ง การวิเคราะห์ประโยคต้นแบบไปเป็นประโยคภาษากลาง

ส่วนที่สอง การสังเคราะห์ประโยคภาษากลางไปเป็นประโยคเป้าหมาย

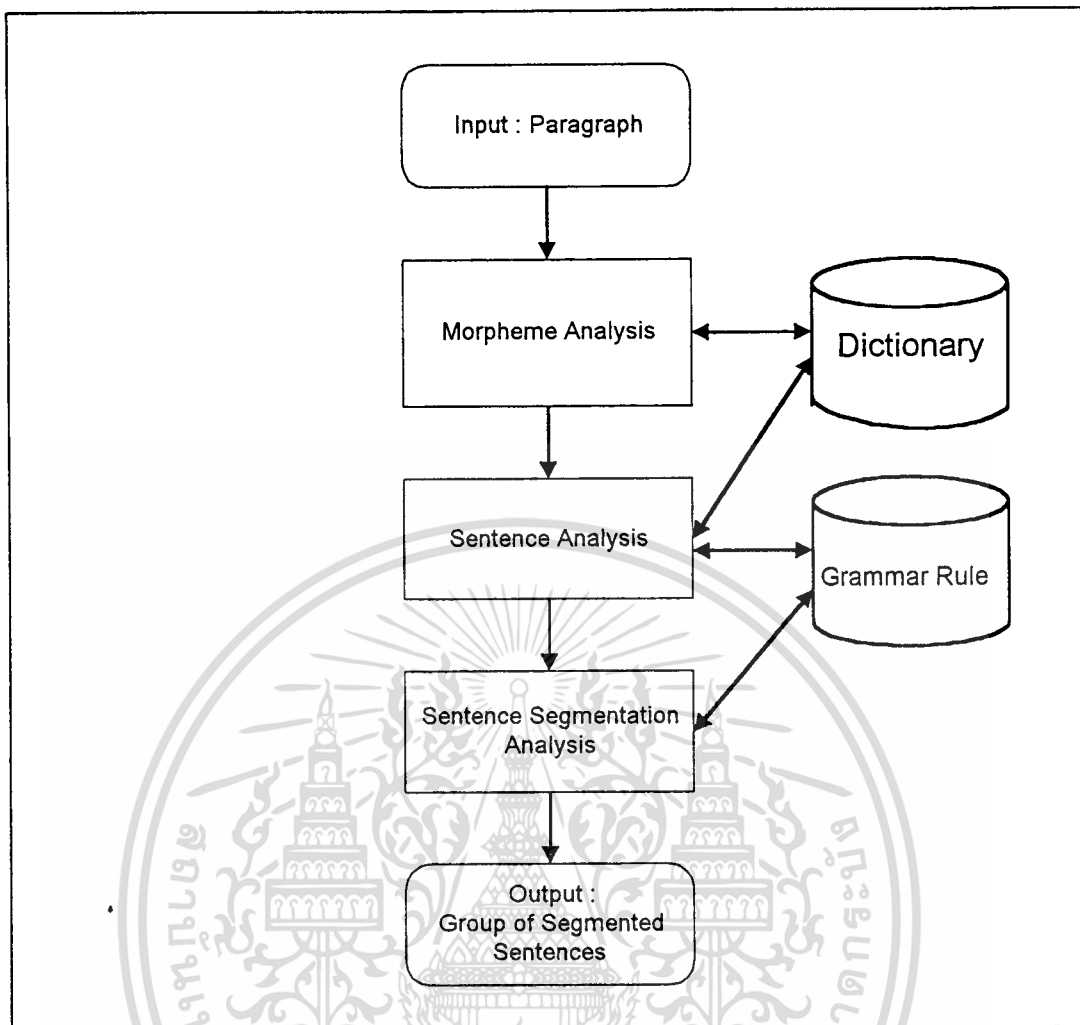
โดยการทำงานของส่วนแรกเป็นการวิเคราะห์ประโยคต้นแบบไปเป็นประโยคภาษากลางเริ่มจากการรับข้อความภาษาต้นแบบ ผ่านระบบวิเคราะห์หน่วยคำ (Morphological Analysis System) ที่จะทำการตรวจสอบหาความหมายของคำจากพจนานุกรม แล้วส่งผลลัพธ์ไปให้ระบบวิเคราะห์โครงสร้างทางวากยสัมพันธ์-อรรถศาสตร์ของประโยค (Syntactic/Semantic Analysis System) ที่จะทำการวิเคราะห์โครงสร้างทางไวยากรณ์และความหมายของประโยค ซึ่งผลลัพธ์ที่ได้จะถูกส่งต่อไปให้ระบบวิเคราะห์ความหมายลึก เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Conceptual Analysis System) ที่จะทำการวิเคราะห์ประโยคให้อยู่ในรูปของโครงสร้างทางความหมายลึก ซึ่งจะเป็นประโยคภาษากลาง

การทำงานส่วนที่สองเป็นส่วนการสังเคราะห์ภาษาเป้าหมาย ซึ่งจะนำประโยคภาษากลางที่ได้จากส่วนการวิเคราะห์ภาษาต้นแบบส่งไปยังระบบเตรียมข้อมูลสำหรับการสร้างประโยค (Generation Planning System) โดยจะเตรียมข้อมูลโครงสร้างทางความหมายของประโยคเป้าหมาย แล้วส่งผลลัพธ์ให้ระบบสร้างโครงสร้างสำหรับประโยค (Semantic and Syntactic Generation System) ที่จะจัดเรียงหน่วยคำตามโครงสร้างภาษาเป้าหมาย จากนั้นผลลัพธ์จะถูกส่งไปยังระบบกำหนดคำ (Morphological Generation System) ซึ่งจะสร้างประโยคพื้นผิวของภาษาเป้าหมาย (Surface Target Language) ถือเป็นผลลัพธ์สุดท้ายของการแปลภาษาด้วยคอมพิวเตอร์

จากการทำงานของระบบแปลภาษาที่กล่าวมาทั้งหมดนี้ ถือได้ว่าสามารถทำงานด้านแปลภาษาได้ดีในระดับหนึ่งเท่านั้น เนื่องจากข้อจำกัดของระบบที่ทำการแปลประโยคครั้งละหนึ่งประโยคเท่านั้น สำหรับการแปลภาษาไทยยังมีปัญหาที่ต้องทำการวิจัยต่อไปคือภาษาไทยเป็นภาษาที่มีลักษณะการเขียนประโยคติดกัน ต่อเนื่องไปจนหมดข้อความ นอกจากนี้ไม่มีสัญลักษณ์แบ่งคำแล้วยังไม่มีสัญลักษณ์การแบ่งประโยค ดังนั้นในการแปลข้อความภาษาไทยทีละหนึ่งหน่วยประโยค ด้วยระบบดังกล่าวนี้สามารถปฏิบัติได้ แต่ถ้าจะแปลเป็นลักษณะทีละหนึ่งหน่วยย่อนำนั้นจะต้องมีขบวนการแยกหน่วยประโยค (Sentence Segmentation) ที่อาศัยฐานความรู้ทางด้านไวยากรณ์และพจนานุกรมภาษาไทย ทั้งนี้เป็นเพราะว่าภาษาไทยไม่มีเครื่องหมายบอกการจบประโยคดังเช่นภาษาอื่นตัวอย่างเช่น เครื่องหมาย "." (Full Stop sign) ในภาษาอังกฤษ เป็นต้น

ดังนั้นวิทยานิพนธ์ฉบับนี้ขอเสนอผลงานวิจัยขั้นพื้นฐาน เพื่อเป็นแนวทางในการแก้ไขปัญหาที่กล่าวมาข้างต้น โดยจุดประสงค์หลักเพื่อวิเคราะห์แยกข้อความภาษาไทยในย่อหน้าออกมาเป็นประโยคอิสระ ด้วยหลักการวิเคราะห์หน่วยคำ การวิเคราะห์โครงสร้างประโยค และการวิเคราะห์แยกประโยคจากย่อหน้า ดังแสดงในรูปที่ 1.3



รูปที่ 1.3 แสดงโครงสร้างของระบบวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

1.2 จุดประสงค์และขอบเขตของงานวิจัย

งานวิจัยนี้ได้เสนอขอบเขตการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้าโดยได้แบ่งออกเป็นขั้นตอนหลักได้ 3 ขั้นตอนคือ

ขั้นตอนที่ 1 : ส่วนการวิเคราะห์หน่วยคำภาษาไทย

ขั้นตอนที่ 2 : ส่วนการวิเคราะห์โครงสร้างประโยคภาษาไทย

ขั้นตอนที่ 3 : ส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทั้งสามขั้นตอนจะทำงานร่วมกับฐานข้อมูลพจนานุกรมคำศัพท์ภาษาไทย ซึ่งจะมีข้อมูลของคำศัพท์ภาษาไทย และข้อมูลของกฎไวยากรณ์สำหรับกรวิเคราะห์โครงสร้างประโยคภาษาไทย โดยที่ส่วนการวิเคราะห์หน่วยคำภาษาไทยจะทำการแยกหน่วยคำภาษาไทยในย่อหน้าด้วยวิธีการ "แยกคำที่เป็นไปได้ (Possible Word Matching)" จากนั้นผลลัพธ์ที่ได้จะถูกส่งไปยังส่วนการวิเคราะห์โครงสร้างประโยคภาษาไทย โดยใช้หลักการของกฎไวยากรณ์โครงสร้างวลี (Phrase Structure Grammar) กฎไวยากรณ์การกร (Case Grammar) และกฎไวยากรณ์พึ่งพา (Dependency Grammar) ประกอบกัน เพื่อหาโครงสร้างทางวากยสัมพันธ์-อรรถศาสตร์ของประโยคภาษาไทยภายในย่อหน้า โดยงานวิจัยนี้จะทำการวิเคราะห์ข้อความภาษาไทยที่เป็นภาษาเขียน และโครงสร้างประโยคภาษาไทยที่ใช้ในงานวิจัยนี้เป็นประโยคความเดียวซึ่งหมายถึงประโยคที่มีหน่วยกริยาอยู่หน่วยเดียวในประโยค จากนั้นจะส่งไปยังส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า ซึ่งจะใช้กฎไวยากรณ์ของโครงสร้างประโยคภาษาไทยทำการวิเคราะห์ออกมาเป็นกลุ่มประโยคอิสระซึ่งเป็นผลลัพธ์ที่ต้องการของงานวิจัย

1.3 โครงร่างของวิทยานิพนธ์

วิทยานิพนธ์นี้ประกอบด้วยบทต่างๆดังนี้

บทที่ 1 ว่าด้วยเรื่องบทนำ

บทที่ 2 ว่าด้วยเรื่องฐานความรู้ทางไวยากรณ์ของภาษาไทย

บทที่ 3 ว่าด้วยเรื่องกระบวนการจัดเก็บฐานข้อมูลไวยากรณ์ในส่วนของพจนานุกรมคำศัพท์

บทที่ 4 ว่าด้วยเรื่องกระบวนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

บทที่ 5 ว่าด้วยเรื่องซอฟต์แวร์ของระบบ

บทที่ 6 บทสรุป

บทที่ 2

ฐานความรู้เกี่ยวกับไวยากรณ์ภาษาไทยที่ใช้ในการวิเคราะห์

2.1 บทนำ

ฐานความรู้ด้านไวยากรณ์ที่ใช้ในงานวิจัยนี้รวบรวมมาจากงานวิจัยด้านการวิเคราะห์โครงสร้างของภาษาไทยโดยเน้นในด้านการแปลภาษาด้วยระบบซอฟต์แวร์คอมพิวเตอร์ร่วมกับทฤษฎีทางด้านภาษาศาสตร์คอมพิวเตอร์และการประมวลผลภาษาธรรมชาติ ดังนั้นการจัดแบ่งหมวดประเภทของคำและโครงสร้างทางไวยากรณ์ภาษาจึงถือหลักตามทฤษฎีของการประมวลผลด้านภาษาธรรมชาติ

2.2 การแบ่งประเภทของคำภาษาไทย

ในการแบ่งประเภทคำ จะยึดถือหลักการแบ่งคำซึ่งพัฒนาขึ้นโดยโครงการการแปลภาษาหลายภาษาด้วยระบบซอฟต์แวร์คอมพิวเตอร์ (Multi-lingua Machine Translation Project)^{[3][4]} โดยวิเคราะห์คำเพื่อแบ่งประเภทตามหน้าที่ของคำในประโยค ที่เน้นการวิเคราะห์ภาษาไทยตามหลักการประมวลผลภาษาธรรมชาติ จึงมีบางส่วนแตกต่างไปจากหลักไวยากรณ์ทางภาษาศาสตร์ของภาษาไทยไปบ้างเล็กน้อย

การแบ่งระดับคำจะแบ่งออกเป็น กลุ่มย่อยดังต่อไปนี้

2.2.1 กลุ่มคำนาม (Noun ใช้สัญลักษณ์ N) ได้แก่คำที่ใช้เรียก คน สัตว์ สิ่งของ แบ่งเป็นคำย่อยได้ 9 ประเภทคือ วิสามัญนาม คำนามที่แสดงจำนวนนับ คำนามที่บอกลำดับที่ก่อน-หลัง คำนามที่ทำหน้าที่ได้เช่นเดียวกับคำกริยาวิเศษณ์ คำนามที่แสดงการขยายคำอื่นๆได้เช่นเดียวกับคำวิเศษณ์ คำนามที่เป็นเครื่องหมายหรือสัญลักษณ์ คำนามที่ใช้บอกยศตำแหน่งหรือบรรดาศักดิ์ สามัญนาม สมุหนาม ดังนี้

- 1) วิสามัญนาม (Proper Noun ใช้สัญลักษณ์ NPRP) เป็นคำนามที่เฉพาะ เป็นชื่อเฉพาะของ คน หรือของสิ่งใดสิ่งหนึ่ง แสดงความสัมพันธ์ในแง่การที่เป็นชื่อเฉพาะ(Proper Noun) และเรียกว่า CASE NAME และไม่ได้มีหน้าที่เป็นคำขยาย(Attribute) เช่น กรุงเทพฯ มานะ อนุสาวรีย์ชัยสมรภูมิ
- 2) คำนามที่แสดงจำนวนนับ (Cardinal Number ใช้สัญลักษณ์ NCNM) เช่น หนึ่ง สอง ร้อย ครั้ง สามารถทำหน้าที่เป็นประธาน หรือ กรรมของประโยคได้ เช่น สามมากกว่าสอง
- 3) คำนามที่บอกลำดับที่ก่อน - หลัง (Ordinal Number ใช้สัญลักษณ์ NONM) เช่น ที่หนึ่ง ที่เท่าไร ที่สุดท้าย
- 4) คำนามที่ทำหน้าที่ได้เช่นเดียวกับคำกริยาวิเศษณ์ (Adverb Like Noun ใช้สัญลักษณ์ NALN) มักจะเป็นคำนามที่ใช้แสดงเวลา เช่น วันอาทิตย์ วันนี้ ปีที่แล้ว และเป็นคำนามที่มักปรากฏอยู่หลังคำกริยา และตำแหน่งต้น หรือท้ายในประโยค เช่นเดียวกับคำกริยาวิเศษณ์(adverb) เช่น
- ร้านปิดทุกวันอาทิตย์
 - เมื่อปีที่แล้วมีประชุมใหญ่ที่กรุงเทพฯ
- 5) คำนามที่แสดงการขยายคำอื่น ๆ ได้เช่นเดียวกับคำวิเศษณ์(Adjective Like Noun ใช้สัญลักษณ์ NAJN) ได้แก่คำว่า ส่วนใหญ่ ส่วนมาก ซึ่งจะพบได้ในประโยคที่กล่าวแสดงความหมายรวมตัวอย่าง เช่น ส่วนใหญ่เรียนจบชั้นอุดมศึกษา
- 6) คำนามที่เป็นเครื่องหมายหรือสัญลักษณ์ (Label Noun ใช้สัญลักษณ์ NLBL) หมายถึง คำนามที่เป็นเครื่องหมายหรือสัญลักษณ์ และทำหน้าที่ได้เหมือนคำนามทั่วไป เช่น
- 1,2,+,-
- 7) คำนามที่ใช้บอกยศ ตำแหน่ง หรือบรรดาศักดิ์ (Title Noun ใช้สัญลักษณ์ NTTL) เป็นคำนามที่ใช้นำหน้าชื่อเพื่อแสดงถึงยศ ตำแหน่ง หรือบรรดาศักดิ์ เช่น คำว่า ร้อยโท ดร.
- 8) สามัญนาม(Common Noun ใช้สัญลักษณ์ NCMN) เป็นคำนามโดยทั่วไปที่ใช้เรียก คน สัตว์ สิ่งของ เช่น หนังสือ อาหาร หอ. รวมทั้งคำนามที่เป็นประเภทย่อยของคำนามนั้นซึ่งไม่ได้แสดงการสัมพันธ์เป็น NAME แต่การแสดงการขยาย attribute เท่านั้น เช่น ดอกไม้ป่า รถยนต์นั่ง รถเก๋ง และคำนามที่มาจากกริยาเติมคำอุปสรรค(Prefix) เช่น การกระทำ ความรัก ความสุข
- 9) สมุหนาม (Collective Noun ใช้สัญลักษณ์ NCLT) เป็นคำนามที่ใช้บอกจำนวนกลุ่มของคำนาม ได้แก่คำว่า กลุ่ม ผุ่ง เป็นต้น

2.2.2 กลุ่มคำสรรพนาม (Pronoun ใช้สัญลักษณ์ PRON) ได้แก่คำที่ใช้แทนคำนามและอาจแทนด้วยคำบ่งชี้ที่มีความหมายรวมถึงคำนามที่ละไว้ด้วย โดยแบ่งเป็นชนิดย่อยได้ 6 ประเภท คือ บุรุษสรรพนาม สรรพนามบ่งชี้ที่ใช้แทนสิ่งที่ผู้พูดชี้เฉพาะเจาะจงถึง สรรพนามชี้เฉพาะที่ใช้แทนคนหรือสิ่งต่างๆไป สรรพนามไม่ชี้เฉพาะที่ใช้แทนคนหรือสิ่งต่างๆไป สรรพนามที่แสดงการถาม ประพันธสรรพนามที่ใช้แทนนามที่อยู่ข้างหน้า ดังนี้

- 1) บุรุษสรรพนามซึ่งเป็นคำสรรพนามที่แทนตัวผู้พูด ผู้ที่พูดด้วย หรือผู้ที่กล่าวถึง (Personal Pronoun ใช้สัญลักษณ์ PPRS) เช่น กระผม ข้าพเจ้า เขา พวกเขา เป็นต้น
- 2) สรรพนามบ่งชี้ที่ใช้แทนสิ่งที่ผู้พูดชี้เฉพาะเจาะจงถึง (Demonstrative Pronoun ใช้สัญลักษณ์ PDMN) เช่น นี่คือ สิ่งที่คุณต้องการ โน่นคือ บ้านของเขา
- 3) สรรพนามชี้เฉพาะที่ใช้แทนคนหรือสิ่งต่างๆไป ไม่มีขอบเขตที่แน่นอน (Definite Pronoun ใช้สัญลักษณ์ PDEF) ได้แก่คำว่า เหล่านี้ ทั้งหมด ทั้งคู่ ทั้งสอง เช่น เราต้องการใช้ทั้งหมด
- 4) สรรพนามไม่ชี้เฉพาะที่ใช้แทนคนหรือสิ่งต่างๆไป ไม่มีขอบเขตที่แน่นอน (Indefinite Pronoun ใช้สัญลักษณ์ PIND) หมายถึง ได้แก่คำ บ้าง ทั้งหลาย ทั้งหมด เช่น บ้างชอบเดินไปทำงาน
- 5) สรรพนามที่แสดงการถาม ใช้แทนสิ่งที่ผู้พูดต้องการทราบคำตอบ (Interrogative Pronoun ใช้สัญลักษณ์ PNTR) ได้แก่คำ ใคร อะไร ที่ไหน อย่างไร เมื่อไร ทำไม เช่น เราต้องใช้อะไรในการส่งข้อมูลระหว่างคอมพิวเตอร์
- 6) ประพันธสรรพนามที่ใช้แทนนามที่อยู่ข้างหน้า และทำหน้าที่เชื่อมประโยคเข้าด้วยกัน (Relative Pronoun ใช้สัญลักษณ์ PREL) ได้แก่คำ ที่ ซึ่ง อัน ผู้ เช่น RS-232C เป็นสายที่ใช้ส่งข้อมูล

2.2.3 กลุ่มคำกริยา (Verb ใช้สัญลักษณ์ VERB) คือคำที่แสดงกริยาอาการหรือสภาพของประธานในประโยคและถือเป็นคำหลักของประโยค แยกเป็น 3 ประเภทคือ คำกริยาแสดงอาการของประธานในประโยค คำกริยาแสดงภาพ คำกริยาที่บ่งบอกคุณสมบัติของคำนาม ดังนี้

- 1) คำกริยาแสดงอาการของประธานในประโยค (Active Verb ใช้สัญลักษณ์ VACT) เช่น วิ่ง ให้ ประกาศ มอง เรียน คิด
- 2) คำกริยาแสดงภาพ ความรู้สึกของประธาน หรือแสดงผลจากการกระทำอย่างใดอย่างหนึ่ง (Static Verb ใช้สัญลักษณ์ VSTA) เช่น ราคา หนัก ลำบาก ได้ยิน เรียนรู้ เสียใจ เป็น คือ จำ

3) คำกริยาที่บ่งบอกคุณสมบัตินี้ของคำนาม (Attributive Verb ใช้สัญลักษณ์ VATT) มักมาจากคำคุณศัพท์นั่นเอง เช่น สวย ใหญ่ ผอม เย็น

2.2.4 กลุ่มคำช่วยกริยา (Auxiliary Verb ใช้สัญลักษณ์ AUX) คือคำที่เกิดร่วมกับคำกริยาเพื่อแสดงกาล มาลา วาจก ลักษณะการ รวมทั้งเจตนาของประโยค แยกเป็น 5 ประเภทคือ คำช่วยกริยาที่วางอยู่หน้า กริยาและหน้าคำปฏิเสธ"ไม่"ในประโยคปฏิเสธ คำช่วยกริยาที่วางอยู่หน้ากริยาและวางหลังคำปฏิเสธ"ไม่"ในประโยคปฏิเสธ คำช่วยกริยาที่อาจวางหน้ากริยาและวางหน้าหรือหลังคำปฏิเสธ "ไม่"ในประโยคปฏิเสธ คำช่วยกริยาที่อยู่หน้าประโยค คำช่วยกริยาที่วางหลังกริยาและวางไว้ท้ายกริยาวลี ดังนี้

- 1) คำช่วยกริยาที่วางอยู่หน้า กริยาและหน้าคำปฏิเสธ "ไม่" ในประโยคปฏิเสธ (Before Verb and Before "ไม่" ใช้สัญลักษณ์ XVBM) เช่น จะ กำลัง อาจ(จะ) เกือบ(จะ) คล้ายจะ
- 2) คำช่วยกริยาที่วางอยู่หน้า กริยาและวางหลังคำปฏิเสธ "ไม่" ในประโยคปฏิเสธ (Before Verb and After "ไม่" ใช้สัญลักษณ์ XVAM) เช่น ค่อย(จะ) น่า(จะ) ได้มัว(แต่จะ)
- 3) คำช่วยกริยาที่อาจวางหน้า กริยาและวางหน้าหรือหลังคำปฏิเสธ "ไม่" ในประโยคปฏิเสธ แต่ความหมายแตกต่างกันไป (Before Verb ; Before or After "ไม่" ใช้สัญลักษณ์ XVMM) เช่น ควร(จะ) เคย(จะ) ต้อง อยาก(จะ) ชอบ(จะ) กรุณา อย่า โปรดห้าม
- 4) คำช่วยกริยาที่อยู่หน้าประโยค (Before Verb; At the beginning of sentence ใช้สัญลักษณ์ XVBB) เช่น โปรด กรุณา ห้าม อย่า
- 5) คำช่วยกริยาที่วางหลังกริยาและวางไว้ท้ายกริยาวลี (After Verb; At the End of Verb Phrase ใช้สัญลักษณ์ XVAE) เช่น อยู่ อยู่แล้ว แล้ว ไป มา ขึ้น ลง เข้า ออก เสีย ไว้ เอา ให้ ดู

2.2.5 กลุ่มคำบ่งชี้ (Determiner ใช้สัญลักษณ์ DET) คือคำที่ใช้แสดงความเฉพาะเจาะจงของคำนาม โดยปรากฏร่วมกับคำนามและคำลักษณนาม แบ่งได้เป็นชนิดย่อยได้ 9 ประเภท ดังนี้ คำบ่งชี้ที่วางอยู่หลังคำนามหรือคำขยายนามที่ชี้เฉพาะนั้น คำบ่งชี้ที่วางอยู่หลังคำนามหรือคำขยายนามที่ชี้เฉพาะนั้น และจะมีคำลักษณนามวางอยู่ระหว่างคำนามกับคำบ่งชี้ชนิดนี้ได้ คำบ่งชี้ที่วางอยู่ระหว่างคำนามและคำลักษณนามหรืออยู่หน้าคำบอกจำนวน คำบ่งชี้ที่วางหลังคำแสดงจำนวน คำบอกจำนวนที่วางอยู่หลังคำนามหรือคำขยายนามและอาจมีคำลักษณนามอยู่ระหว่างคำนามและคำบอก

จำนวนชนิดนี้ได้ คำบอกจำนวนที่วางอยู่ระหว่างคำนามและลักษณนามหรือหน้าจำนวน(ตัวเลข) คำบอกจำนวนที่วางอยู่หลังจำนวน คำแสดงจำนวนนับ คำแสดงลำดับที่ ดังนี้

2.2.5.1 Definite Determiner (ใช้สัญลักษณ์ DED) หมายถึง คำบ่งชี้ที่ใช้ชี้เฉพาะคำนามและปรากฏอยู่ในตำแหน่ง ดังนี้

1) คำบ่งชี้ที่วางอยู่หลังคำนาม หรือคำขยาย นามที่ชี้เฉพาะนั้น (After Noun ใช้สัญลักษณ์ DDAN) แต่จะไม่มีคำลักษณนาม (Classifier) ปรากฏอยู่ระหว่างคำนามกับคำบ่งชี้ชนิดนี้เลย เช่น นี้ นั้น โน่น นู่น ตัวอย่าง รถนี้ราคาแพงมาก

2) คำบ่งชี้ที่วางอยู่หลังคำนามหรือคำขยายนามที่ชี้เฉพาะนั้น และจะมีคำลักษณนาม (Classifier) วางอยู่ระหว่างคำนามกับคำบ่งชี้ชนิดนี้ได้ (After Noun and Classifier ใช้สัญลักษณ์ DDAC) เช่น นี้ นั้น นู่น ตัวอย่าง รถคันนี้ราคาแพงมาก

3) คำบ่งชี้ที่วางอยู่ระหว่างคำนามและคำลักษณนาม หรืออยู่หน้าคำบอกจำนวน (Between Noun and Classifier ใช้สัญลักษณ์ DDBQ) เช่น ทั้ง ตั้ง สัก อีก เพียง ตัวอย่าง เครื่องทิ้งชุด, เราต้องการเครื่องอีก 3 เครื่อง

4) คำบ่งชี้ที่วางหลังคำแสดงจำนวน (After Quantitative Noun ใช้สัญลักษณ์ DDAQ) เช่น เท่านั้น พอดี ถ้วน ตัวอย่าง เรามีวิศวกรอยู่เพียง 3 คน เท่านั้น

2.2.5.2 Indefinite determiner (ใช้สัญลักษณ์ DEIN) หมายถึง คำบ่งชี้ที่ใช้แสดงจำนวนและปรากฏในตำแหน่งดังนี้

1) คำบอกจำนวนที่วางอยู่หลังคำนามหรือคำขยายนามและอาจมีคำลักษณนามอยู่ระหว่างคำนาม และคำบอกจำนวนชนิดนี้ได้ (After Classifier ใช้สัญลักษณ์ DIAC) เช่น โหนด ใด อื่น อื่นๆ ต่างๆ

ตัวอย่าง ที่มีเครื่องชนิดต่างๆให้เลือก

2) คำบอกจำนวนที่วางอยู่ระหว่างคำนาม และลักษณนาม หรือหน้าจำนวน(ตัวเลข) (Before Quantitative Noun ใช้สัญลักษณ์ DIBQ) เช่น บาง ราว เกือบ ประมาณ เกือบๆ ราวๆ

ตัวอย่าง เครื่องจะใช้เวลาประมาณ 3 นาที

3) คำบอกจำนวนที่วางอยู่หลังจำนวน (After Quantitative Noun ใช้สัญลักษณ์ DIAQ) เช่น เศษ เศษๆ กว่า กว่าๆ

เอกสารนี้ตัวอย่าง เครื่องจะใช้เวลาประมาณ 3 นาทีเศษๆ เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) คำแสดงจำนวนนับ (Cardinal Number ใช้สัญลักษณ์ DCNM) เช่น หนึ่ง ร้อย

ตัวอย่าง หนังสือสิบเล่มนี้

5) คำแสดงลำดับที่ (Ordinal Number ใช้สัญลักษณ์ DONM) เช่น ที่หนึ่ง ที่สิบ

ตัวอย่าง รถคันที่สิบ

2.2.6 กลุ่มคำกริยาวิเศษณ์ (ADVERB ใช้สัญลักษณ์ ADV) คือ คำขยายกริยาเพื่อบอกว่าเหตุการณ์หรือสิ่งที่เกิดขึ้นนั้น เกิดเมื่อไหร่ ที่ไหน หรืออย่างไร ในภาษาไทยกริยาวิเศษณ์จะไม่ปรากฏหน้าคำกริยาที่ถูกขยายเลย มีชนิดย่อยได้ 4 ประเภท คือ คำกริยาวิเศษณ์ทั่วๆไป คำกริยาวิเศษณ์ที่เป็นคำซ้ำ คำกริยาวิเศษณ์ที่มีคำอุปสรรค คำกริยาวิเศษณ์ที่ทำหน้าที่ขยายทั้งประโยค ดังนี้

1) คำกริยาวิเศษณ์ทั่วๆไป วางอยู่หลังคำกริยาที่ถูกขยาย (Adverb with Normal Form ใช้สัญลักษณ์ ADVN) เช่น จำเจ เร็ว เก่ง สม่่าเสมอ พอ เกิน ตรง

2) คำกริยาวิเศษณ์ที่เป็นคำซ้ำ(มีรูปซ้ำตัวเอง) (Adverb with Iterative Form ใช้สัญลักษณ์ ADVI) เช่น เร็วๆ เสมอๆ มั่วๆ ตัวอย่าง เขาชอบมาที่นี้เสมอๆ

3) คำกริยาวิเศษณ์ที่มีคำอุปสรรค(อย่าง โดย) ประกอบ (Adverb with Prefixed Form ใช้สัญลักษณ์ ADVP) เช่น อย่างรวดเร็ว โดยเร็ว ตัวอย่าง ฉันจะไปที่นั่นโดยเร็วที่สุด

4) คำกริยาวิเศษณ์ที่ทำหน้าที่ขยายทั้งประโยค (Sentential Adverb ใช้สัญลักษณ์ ADVS) เช่น ปกติ ธรรมดา ทั่วไป ทั่วๆไป โดยธรรมดา ตามปกติ ตัวอย่าง โดยทั่วไปจะมีฝนตกชุกในภาคใต้ของประเทศไทย

2.2.7 กลุ่มคำลักษณนาม (Classifier ใช้สัญลักษณ์ CLAS) คือคำนามที่ทำหน้าที่ประกอบนามอื่นเพื่อแสดงรูปลักษณะ ขนาด หรือประมาณของนามนั้นให้ชัดเจนยิ่งขึ้น แบ่งเป็นชนิดย่อยได้ 5 ประเภท คือ คำลักษณนามที่ใช้แสดงจำนวนหน่วยนับ คำลักษณนามที่ใช้แสดงจำนวนกลุ่ม คำลักษณนามที่ใช้แสดงจำนวนหน่วยวัด คำลักษณนามที่ใช้แสดงจำนวนความถี่ คำลักษณนามที่มาจากคำกริยา ดังนี้

1) คำลักษณนามที่ใช้แสดงจำนวนหน่วยนับ (Unit Classifier ใช้สัญลักษณ์ CNIT) ได้แก่ เล่ม ตัว

2) คำลักษณนามที่ใช้แสดงจำนวนกลุ่ม (Collective Classifier ใช้สัญลักษณ์ CLTV) ได้แก่ กอง หมู่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) คำลักษณะนามที่ใช้แสดงจำนวนหน่วยวัด (Measurement Classifier ใช้สัญลักษณ์ CMTR) ได้แก่ กิโลกรัม แก้ว

4) คำลักษณะนามที่ใช้แสดงจำนวนความถี่ (Frequency Classifier ใช้สัญลักษณ์ CFQC) ได้แก่ ครั้ง

5) คำลักษณะนามที่มาจากคำกริยา (Verbal Classifier ใช้สัญลักษณ์ CVBL) ได้แก่ มัด ห่อ

2.2.8 กลุ่มคำสันธาน (Conjunction ใช้สัญลักษณ์ CONJ) คือคำที่ใช้เชื่อมวลี หรือประโยคเข้าด้วยกัน เพื่อแสดงความสัมพันธ์กันในแง่ใดแง่หนึ่ง แบ่งเป็นชนิดย่อยได้ 3 ประเภทคือ คำเชื่อมที่ใช้เชื่อมประโยคความรวมหรือเชื่อมประโยคความเดียวสองประโยคเข้าด้วยกันเพื่อแสดงความสัมพันธ์ในลักษณะใดลักษณะหนึ่ง คำเชื่อมที่ใช้เชื่อมประโยคที่แสดงความเปรียบเทียบ คำเชื่อมที่ใช้เชื่อมประโยคความซ้อน ดังนี้

1) คำเชื่อมที่ใช้เชื่อมประโยคความรวม(Compound Sentence) หรือเชื่อมประโยคความเดียว (Simple Sentence) 2 ประโยคเข้าด้วยกัน เพื่อแสดงความสัมพันธ์ในลักษณะใดลักษณะหนึ่ง (Coordinating Conjunction ใช้สัญลักษณ์ CCRG) เช่น และเชื่อมความคล้ายตามกัน) แต่(เชื่อมความขัดแย้งกัน หรือเชื่อมความให้เลือกเอาอย่างใดอย่างหนึ่ง)

2) คำเชื่อมที่ใช้เชื่อมประโยคที่แสดงความเปรียบเทียบ (Comparative Conjunction ใช้สัญลักษณ์ CCMP) เช่น กว่า ที่สุด

3) คำเชื่อมที่ใช้เชื่อมประโยคความซ้อน(Complex Sentence) เป็นการเชื่อมประโยคกับ Clause เข้าด้วยกัน หรือ เชื่อมประโยคหลักกับประโยคย่อยเข้าด้วยกัน เพื่อขยายส่วนใดส่วนหนึ่งของประโยคหลักให้ชัดเจนขึ้น (Subordinating Conjunction ใช้สัญลักษณ์ CSBR) ตัวอย่าง เขาซื้อขนมที่เธอชอบ (ขยายกรรมให้ชัดเจน)

เด็กที่เธอเลี้ยงไว้เป็นเด็กดี(ขยายประธาน)

2.2.9 กลุ่มคำบุรพบท (Preposition ใช้สัญลักษณ์ PREP) คือคำที่นำหน้าคำนามหรือคำสรรพนาม เพื่อบอกตำแหน่งสถานที่ และความสัมพันธ์ระหว่างคำกริยาและคำนามหรือคำสรรพนามที่ตามมา แบ่งเป็นคำชนิดย่อยได้ 2 ประเภทคือ คำบุรพบทที่เป็นคำกริยาได้ด้วย คำบุรพบทอื่นๆ

1. คำบุรพบทที่เป็นคำกริยาได้ด้วย (Verbal Preposition ใช้สัญลักษณ์ RVBP) ตัวอย่าง เขาไปจากบ้าน(เป็นบุรพบท)

เขาจากเธอ (เป็นคำกริยา)

2) คำบุรพบทอื่นๆ (Preposition except "RVBP" ใช้สัญลักษณ์ RPRE) ตัวอย่าง ลืออยู่ในตู้

2.2.10 กลุ่มคำอุทาน (Interjection ใช้สัญลักษณ์ INT) ได้แก่คำที่ใช้แสดงความรู้สึกต่างๆ เช่น ดีใจ เสียใจ ตกใจ ประหลาดใจ มีเพียง 1 ประเภท คือคำอุทาน ดังนี้

1) คำอุทาน (Interjection ใช้สัญลักษณ์ INT) ได้แก่คำว่า อี้ เออ

2.2.11 กลุ่มคำที่ใช้เติมเข้าไปข้างหน้าคำอื่น (Prefix ใช้สัญลักษณ์ PIXP) ได้แก่คำที่ใช้เติมข้างหน้าคำอื่นเพื่อเปลี่ยนหน้าที่ของคำ มีเพียง 1 ประเภท คือ คำที่ใช้เติมเข้าไปข้างหน้าคำอื่น ดังนี้

1) คำที่ใช้เติมเข้าไปข้างหน้าคำอื่น (Prefix ใช้สัญลักษณ์ FIXP) ได้แก่คำว่า อย่าง ความเป็นต้น

2.2.12 กลุ่มคำลงท้าย (Ending) คือคำที่ปรากฏท้ายประโยคเพื่อแสดงคำถามหรือแสดงความรู้สึกแยกเป็น 2 ประเภทคือ คำลงท้ายในประโยคบอกเล่า คำลงท้ายประโยค

1) คำลงท้ายในประโยคบอกเล่า (Ending for Affirmative Sentence ใช้สัญลักษณ์ EAFF) เพื่อแสดงการตอบรับหรือแสดงความรู้สึก เช่น ครับ ค่ะ จ๊ะ

2) คำลงท้ายประโยค (Ending for Interrogative Sentence ใช้สัญลักษณ์ EATT) เพื่อแสดงการถามและทำให้ประโยคบอกเล่ากลายเป็นประโยคคำถามแบบตอบรับ ปฏิเสธ เช่น หรือ ไหม ใช่มั้ย ไม่ใช่ หรือ

2.2.13 กลุ่มคำปฏิเสธ (Negative ใช้สัญลักษณ์ NEG) ได้แก่คำที่ใช้แสดงความหมายปฏิเสธ มีเพียง 1 ประเภทคือ คำปฏิเสธ ดังนี้

1) คำปฏิเสธ (Negation ใช้สัญลักษณ์ NEG) ได้แก่คำว่า ไม่ได้ ไม่ มิ มีได้ เป็นต้น

2.3 โครงสร้างของประโยคภาษาไทย^{[5][6]}

ประโยคภาษาไทยประกอบด้วยหน่วยคำที่ถูกนำมาเรียงให้มีความสัมพันธ์กัน โดยพื้นฐานของภาษาไทยจะพบว่าคำประกอบขึ้นด้วยกลุ่มตัวอักษร และคำเหล่านี้จะผสมกันเป็นกลุ่มของวลี กลุ่มของวลีจะมีความสัมพันธ์กันเป็นโครงสร้างของประโยคภาษาไทย ดังนั้นการนำคำและวลีมาเขียนเรียงกันให้เป็นประโยคภาษาไทย จึงต้องมีไวยากรณ์เป็นหลักในการกำหนดตำแหน่งของคำและวลีที่จะนำมาเรียงกัน ข้อกำหนดนี้เรียกว่า “กฎไวยากรณ์” และขบวนการใช้กฎไวยากรณ์พิจารณาประโยคนี้เรียกว่า “การวิเคราะห์โครงสร้างประโยคภาษาไทย”

ไวยากรณ์ที่ใช้ในการวิเคราะห์โครงสร้างประโยคภาษาไทยในบทความนี้มีอยู่ 3 ไวยากรณ์ด้วยกันคือ

2.3.1 ไวยากรณ์โครงสร้างวลี (Phase Structure Grammar)

ไวยากรณ์โครงสร้างวลีได้กำหนดให้ ประโยคจะประกอบด้วยกลุ่มของวลีซึ่งรวมกลุ่มกันตามกฎไวยากรณ์ของโครงสร้างประโยค ส่วนวลีจะประกอบไปด้วยหน่วยคำหลายหน่วยมารวมกันเป็นวลีตามกฎไวยากรณ์ของวลี แล้ววลีต่างๆจะรวมตัวกันเป็นประโยค ดังนั้นถ้าพิจารณาโครงสร้างของประโยคภาษาไทยจากส่วนที่เล็กที่สุดรวมกันไปเป็นส่วนใหญ่ที่สุดเราจะสร้างเป็นกฎได้ดังนี้

2.3.1.1 กฎไวยากรณ์ของนามวลีภาษาไทย

นามวลีในประโยคภาษาไทยเป็นวลีที่มีค่านามเป็นแกนหลักของกลุ่มคำ โดยจะมีโครงสร้างเป็นค่านาม หรือคำสรรพนาม งานวิจัยนี้ได้แบ่งนามวลีออกเป็นสองโครงสร้างดังนี้ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยามให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก) โครงสร้างนามวลีภาษาไทยที่ 1 ได้แก่นามวลีที่ประกอบด้วยคำนาม

NP → N

ตัวอย่างเช่น

ชาวไทย

นักศึกษา

ข) โครงสร้างนามวลีภาษาไทยที่ 2 ได้แก่นามวลีที่ประกอบด้วยคำสรรพนาม

NP → PRON

ตัวอย่างเช่น

ข้าพเจ้า

เธอ

โดยที่ NP: Noun Phrase (นามวลี) คือกลุ่มของหน่วยคำที่ประกอบกันโดยมีคำนามเป็น

ตัวหลัก

N: Noun คือ คำที่อยู่ในกลุ่มคำนาม

PRON: Pronoun คือคำที่อยู่ในกลุ่มคำสรรพนาม

2.3.1.2 กฎไวยากรณ์ของกริยาวลีภาษาไทย

กริยาวลีในประโยคภาษาไทยเป็นวลีที่มีคำกริยาเป็นแกนหลักของกลุ่มคำ โดยจะมีโครงสร้างเป็นคำกริยาเพียงคำเดียว หรือคำกริยาประกอบกับส่วนขยาย งานวิจัยนี้ได้แบ่งกริยาวลีออกเป็น 7 โครงสร้างดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก) โครงสร้างกริยาวลีภาษาไทยที่ 1 ได้แก่กริยาวลีที่ประกอบด้วยคำกริยา คำนาม และบุรพบทวลี

VP → VERB + N + PP

ตัวอย่างเช่น

มี ความสามารถ ทางด้านคอมพิวเตอร์

ข) โครงสร้างกริยาวลีภาษาไทยที่ 2 ได้แก่กริยาวลีที่ประกอบด้วยคำกริยา คำนาม และคำนาม

VP → VERB + N + N

ตัวอย่างเช่น

ให้ ทิป พนักงาน

ค) โครงสร้างกริยาวลีภาษาไทยที่ 3 ได้แก่กริยาวลีที่ประกอบด้วยคำกริยา และคำนาม

VP → VERB + N

ตัวอย่างเช่น

ชอบ กีฬา

ง) โครงสร้างกริยาวลีภาษาไทยที่ 4 ได้แก่กริยาวลีที่ประกอบด้วยคำกริยา และคำกริยาวิเศษณ์

VP → VERB + ADV

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างเช่น

แสดง ตามปกติ

จ) โครงสร้างกริยาวลีภาษาไทยที่ 5 ได้แก่กริยาวลีที่ประกอบด้วยคำกริยา คำนาม และคำช่วยกริยา

VP → VERB + AUX

ตัวอย่างเช่น

ลูก

ขึ้น

ฉ) โครงสร้างกริยาวลีภาษาไทยที่ 6 ได้แก่กริยาวลีที่ประกอบด้วยคำกริยา คำบุรพบท และคำนาม

VP → VERB + PP

ตัวอย่างเช่น

อยู่

ที่ต่างประเทศ

ช) โครงสร้างกริยาวลีภาษาไทยที่ 7 ได้แก่กริยาวลีที่ประกอบด้วยคำกริยาเพียงคำเดียว

VP → VERB

ตัวอย่างเช่น

เดิน

โดยที่ VP: Verb Phrase(กริยาวลี) คือกลุ่มของหน่วยคำที่ประกอบกันโดยมีคำกริยาเป็น
ตัวหลัก

VERB :Verb คือคำในกลุ่มคำกริยา

N: Noun คือ คำที่อยู่ในกลุ่มคำนาม

ADV: Adverb คือคำในกลุ่มคำกริยาวิเศษณ์

AUX: Auxiliary Verb คือคำในกลุ่มคำช่วยกริยา

PP:Preposition Phrase คือบุรพบทวลี

2.3.1.3 กฎไวยากรณ์ของบุรพบทวลีภาษาไทย

บุรพบทวลีในประโยคภาษาไทยเป็นวลีที่มีคำบุรพบทเป็นแกนหลักของกลุ่มคำ โดยจะมี
โครงสร้างเป็นคำบุรพบท และคำนาม งานวิจัยนี้ได้แบ่งบุรพบทวลีออกเป็น 1 โครงสร้างดังนี้

ก) โครงสร้างบุรพบทวลีภาษาไทยที่ 1 ได้แก่บุรพบทวลีที่ประกอบด้วยคำบุรพบท และ
คำนาม

PP → PREP + N

ตัวอย่างเช่น

ทางด้าน

วรรณกรรม

โดยที่ PP: Preposition Phrase(บุรพบทวลี) คือกลุ่มของหน่วยคำที่ประกอบกันโดย
มีคำบุรพบทเป็นตัวหลัก

PREP: Preposition คือคำบุรพบท

N: Noun คือ คำที่อยู่ในกลุ่มคำนาม

นักศึกษา

N



NP

(ก)

ข้าพเจ้า

PRON



NP

(ข)

รูปที่ 2.1 แสดงกฎไวยากรณ์ของนามวลีภาษาไทย

2.3.1.3 กฎไวยากรณ์ของย่อหน้าภาษาไทย

ย่อหน้าภาษาไทยประกอบด้วย ประโยคภาษาไทยหลายประโยคนำมาเขียนเรียงกัน ซึ่งงานวิจัยนี้ได้แบ่งโครงสร้างของย่อหน้าภาษาไทยออกเป็น 1 โครงสร้างดังนี้

ก) โครงสร้างของย่อหน้าภาษาไทยที่ 1 ได้แก่ย่อหน้าที่ประกอบด้วย ประโยคภาษาไทย คำสันธาน และประโยคภาษาไทย โดยที่คำสันธานจะใช้เชื่อมระหว่างประโยคภาษาไทยสองประโยคที่เขียนเรียงต่อกัน ซึ่งคำสันธานที่ใช้เชื่อมประโยคนั้นจะมีหรือไม่มีก็ได้

$$P \rightarrow S(1) + [\text{CONJ}] + S(2) + \dots + [\text{CONJ}] + S(n-1) + [\text{CONJ}] + S(n)$$

ตัวอย่างเช่น

ศาสตร์

สมชายเป็นคนดี เขามีความสามารถทางด้านคอมพิวเตอร์ เนื่องจาก เขาเรียนวิศวกรรม

โดยที่ P:Paragraph คือย่อหน้าภาษาไทย

S(i):Sentence (i) คือประโยคภาษาไทยในย่อหน้าโดยประโยควางอยู่ในลำดับที่ i

ซึ่ง i = 1, 2, 3, ..., n-1, n

n คือจำนวนประโยคภาษาไทยที่อยู่ในย่อหน้าภาษาไทย

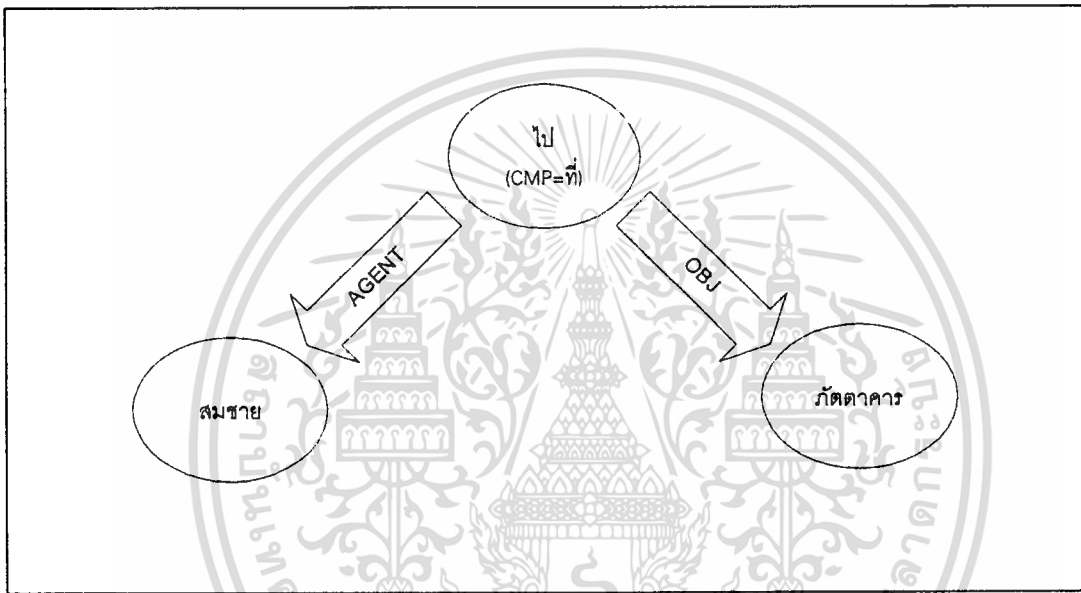
CONJ:Conjunction คือคำสันธานที่ใช้เชื่อมประโยคภาษาไทยในย่อหน้าภาษาไทย

ไทย

2.3.2 ไวยากรณ์การก (Case Grammar)

ไวยากรณ์การกได้กำหนดให้มีความสัมพันธ์ทางความหมายระหว่างคำนามกับคำกริยาในประโยคเสมอ ดังนั้นแม้ว่าการจัดวางตำแหน่งของคำนามกับคำกริยาจะสลับตำแหน่งกันก็ตามความสัมพันธ์การกก็จะยังปรากฏอยู่อย่างเดิม ดังนั้นเราจึงสามารถพิจารณาได้ว่าหน่วยคำใดทำ

หน้าที่เป็นประธาน หรือกรรมของคำกริยาหลักที่ปรากฏในประโยคภาษาไทยได้ งานวิจัยนี้ใช้กฎไวยากรณ์การกในการหาความสัมพันธ์ระหว่างหน่วยคำนามในประโยคภาษาไทยกับหน่วยคำกริยาในภาษาไทยโดยกำหนดไว้ว่า ถ้าหน่วยคำนามมีความสัมพันธ์กับหน่วยกริยาใดแล้ว หน่วยคำนามกับหน่วยกริยาที่มีความสัมพันธ์กันนั้น อยู่ในประโยคเดียวกัน ดังแสดงในรูป 2.6 หน่วยคำกริยา “ไป” ซึ่งมีความสัมพันธ์ AGENT กับหน่วยคำนาม “สมชาย” และหน่วยคำกริยา “ไป” ซึ่งมีความสัมพันธ์ OBJ กับหน่วยคำนาม “ภัตตาหาร” จะอยู่ในประโยคเดียวกัน



รูปที่ 2.6 แสดงโครงสร้างประโยคภาษาไทยด้วยไวยากรณ์การก

2.3.3 ไวยากรณ์พึ่งพา (Dependency Grammar)

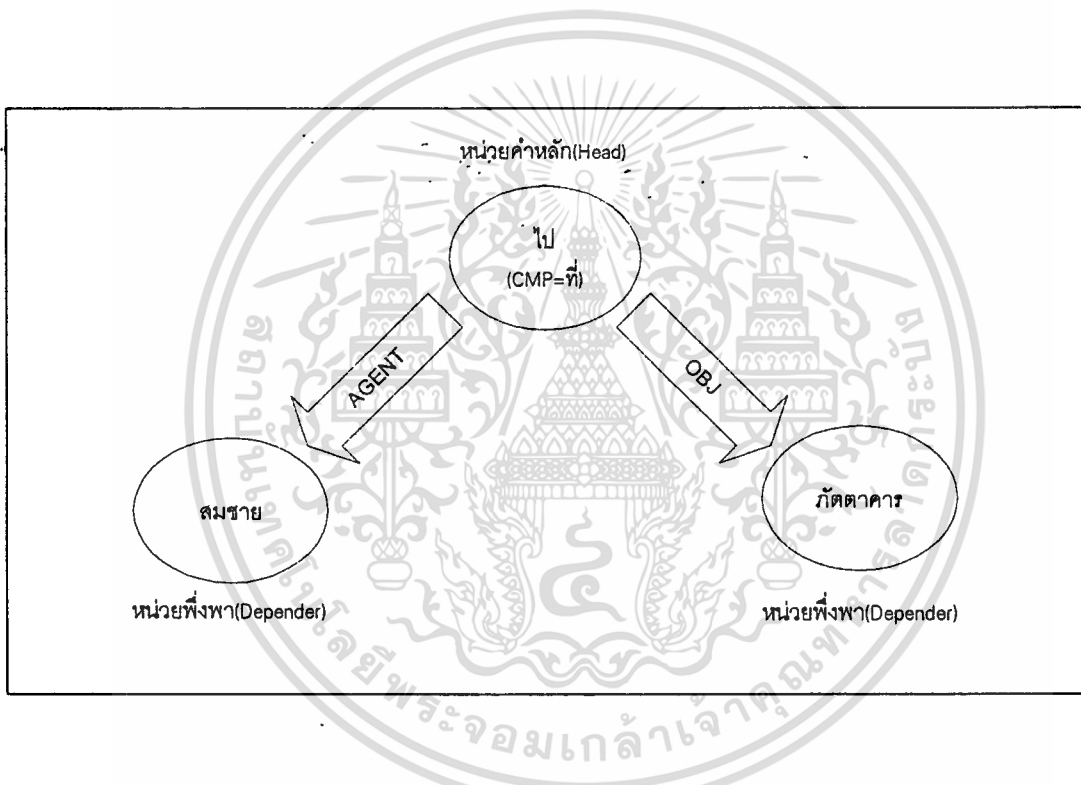
ไวยากรณ์พึ่งพาได้กำหนดไว้ว่าหน่วยคำในประโยคมีความสัมพันธ์แบบพึ่งพา โดยจะมีหน่วยคำหนึ่งทำหน้าที่เป็นหน่วยคำหลัก (Head) และหน่วยอื่นจะทำหน้าที่เป็นหน่วยพึ่งพา (Depender) ดังแสดงในรูปที่ 2.7 ซึ่งจากหลักการนี้เราจะใช้หน่วยคำหลักเป็นตัวค้นหาจุดเริ่มต้นและจุดสิ้นสุดของประโยคภาษาไทยเพื่อจะทำการแยกประโยคภาษาไทยออกจากย่อหน้าภาษาไทย โดยข้อกำหนดของการแสดงความสัมพันธ์แบบพึ่งพามีดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก) หน่วยคำในประโยคภาษาไทยต้องไม่เป็นหน่วยคำหลักของกันและกัน หมายถึงเมื่อหน่วยคำใดเป็นหน่วยคำหลักแล้วจะต้องไม่เป็นหน่วยพึ่งพาของหน่วยคำที่มีความสัมพันธ์แบบพึ่งพากับที่ตนเองเป็นหน่วยหลักอยู่ก่อนแล้ว

ข) หน่วยคำในประโยคภาษาไทยใดๆจะเป็นหน่วยพึ่งพาของหน่วยหลักมากกว่าหนึ่งหน่วยไม่ได้

ค) ทุกๆหน่วยคำในประโยคภาษาไทยย่อมมีความสัมพันธ์กับหน่วยคำอื่นๆในประโยคเสมอ



รูปที่ 2.7 แสดงหน่วยคำหลักและหน่วยพึ่งพาในโครงสร้างประโยคภาษาไทย

บทที่ 3

การจัดเก็บฐานข้อมูลไวยากรณ์ในส่วนของพจนานุกรมคำศัพท์

3.1 บทนำ

ในขบวนการวิเคราะห์โครงสร้างของภาษาไทยด้วยแนวทางของการประมวลผลภาษาธรรมชาตินั้น งานวิจัยนี้ได้แบ่งขั้นตอนของงานหลักสำคัญเป็นสองส่วน ได้แก่

ส่วนที่หนึ่งการวิเคราะห์หน่วยคำภาษาไทย

ส่วนที่สองการวิเคราะห์โครงสร้างของประโยคภาษาไทย

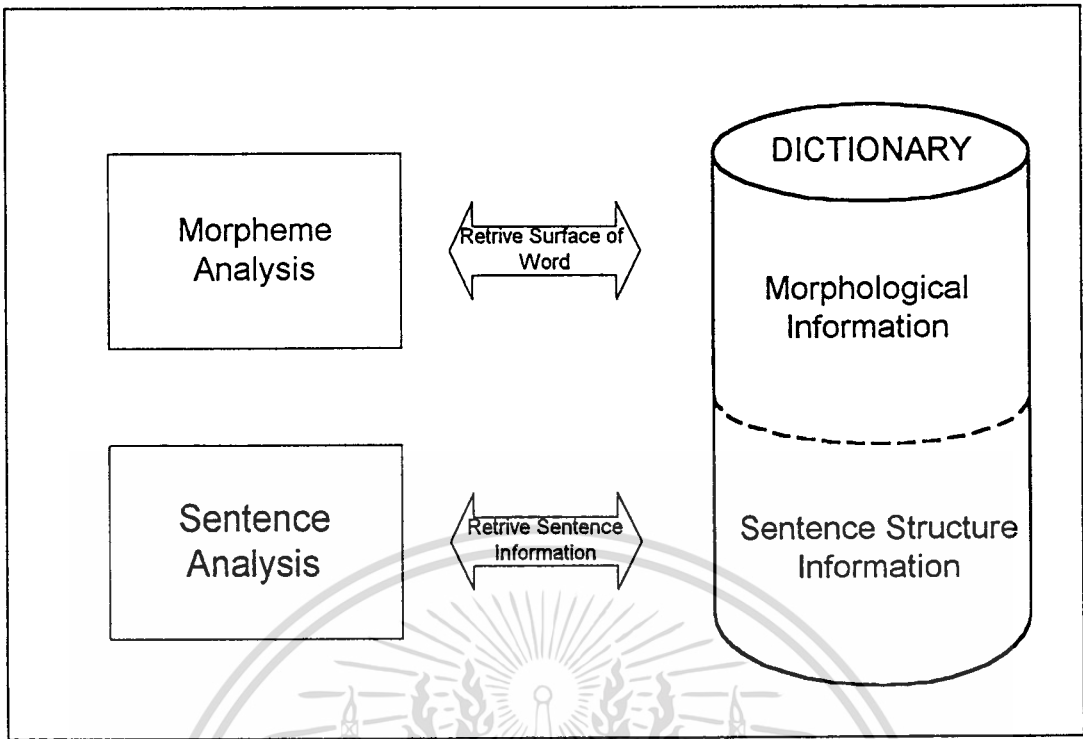
โดยให้ทั้งสองส่วนกระทำร่วมกับฐานความรู้ทางไวยากรณ์ภาษาไทยของคำภาษาไทย ที่ได้วิเคราะห์ไว้ล่วงหน้า และจัดเก็บไว้ในพจนานุกรมคำศัพท์ของระบบ เพื่อใช้ตรวจสอบความหมายทั้งหมดจากคำหรือวลีที่ต้องการวิเคราะห์

งานวิจัยนี้ได้ออกแบบระบบฐานความรู้พจนานุกรมคำศัพท์ภาษาไทยขึ้นมา โดยให้สามารถทำงานร่วมกับระบบวิเคราะห์ภาษาไทยได้ทั้งสองส่วนดังกล่าว ด้วยการวิจัยและพัฒนาระบบฐานความรู้พจนานุกรมคำศัพท์ภาษาไทยระบบเดียว เพื่อจะทำให้สะดวกในการปรับปรุงฐานความรู้ให้อ้างอิงกับข้อมูลได้ตรงกันทั้งส่วนการวิเคราะห์หน่วยคำและส่วนการวิเคราะห์โครงสร้างของประโยคภาษาไทย นั่นคือหากมีการเพิ่มหน่วยคำใดเข้ามาก็จะสามารถใช้ข้อมูลเดียวกันได้ทั้งสองส่วน รายละเอียดของข้อมูลในระบบจะได้อธิบายโดยละเอียดต่อไป

นอกจากนี้งานวิจัยนี้ยังได้พัฒนาสร้างระบบซอฟต์แวร์เพื่อการจัดเก็บข้อมูลไวยากรณ์ของคำศัพท์ในรูปของฐานข้อมูลพจนานุกรมคำศัพท์ภาษาไทย ที่มีโครงสร้างแบบฐานข้อมูลคอมพิวเตอร์ โดยออกแบบให้ง่ายต่อการค้นหาคำศัพท์ ทำให้สามารถนำไปใช้ประโยชน์ในทางอื่นได้นอกเหนือจากการนำไปใช้ในการวิเคราะห์โครงสร้างของภาษา

โครงสร้างของส่วนประกอบของข้อมูลที่จัดเก็บในฐานข้อมูลพจนานุกรมคำศัพท์ภาษาไทย ประกอบไปด้วย ดัชนีคำ (Index Word) ซึ่งเป็นรูปผิวของคำและยังเป็น key ตัวหนึ่งในการเข้าถึงข้อมูลด้วย นอกจากนี้ยังมีข้อมูลทางภาษาที่จะใช้ในการวิเคราะห์โครงสร้างของประโยคภาษาไทยได้แก่ การแบ่งกลุ่มคำ (Category and Subcategory) ข้อมูลของหน่วยคำหลักของประโยคภาษาไทย (VP) ข้อมูลโครงสร้างของประโยค และข้อมูลที่ใช้ในการแปลงความสัมพันธ์ทางวากยสัมพันธ์ไปเป็นความสัมพันธ์ทางอรรถศาสตร์ เพื่อวิเคราะห์โครงสร้างของประโยคให้อยู่ในรูปของความหมายลึก ข้อมูลการแบ่งชนิดของคำ (AKO) โครงสร้างของระบบพจนานุกรมได้ถูกแสดงไว้ในรูปที่ 3.1 และตารางที่ 3.1 จะแสดงโครงสร้างฐานข้อมูลของระบบจัดเก็บพจนานุกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 แสดงโครงสร้างของฐานข้อมูลพจนานุกรมคำศัพท์ภาษาไทยและส่วนต่อเชื่อมกับระบบอื่น

tkey	CP	cat	scat	ako	vp	tmap	cmp2	cmp3
สมชาย	SOMCHAI	N	NPRP	111				
สั่ง	ORDER	VERB	VACT	226	8	SUB=AGT,DOB=OBJ,I0B=TAR		
รับ	RECEIVE	VERB	VACT	2314	6	SUB=AGT,DOB=OBJ,CMP3=SOR		จาก

ตารางที่ 3.1 แสดงโครงสร้างของระบบฐานความรู้พจนานุกรมคำศัพท์

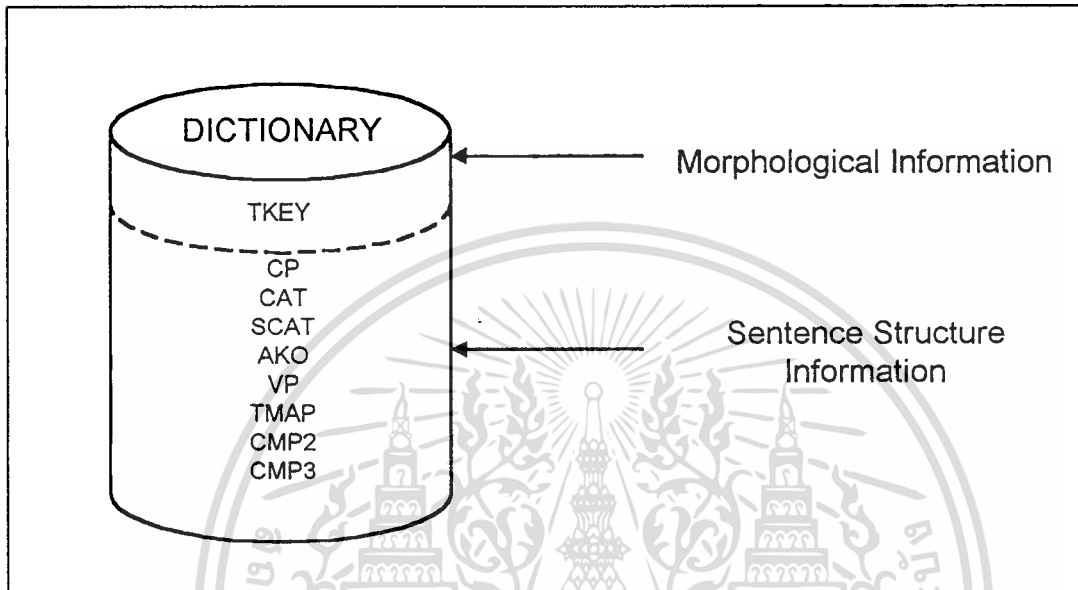
3.2 ส่วนข้อมูลทางโครงสร้างและความหมายของคำศัพท์

ในส่วนนี้จะประกอบด้วยข้อมูลสองชนิดคือ ข้อมูลระดับหน่วยคำ กับข้อมูลระดับโครงสร้างของประโยคภาษาไทยและข้อมูลระดับความหมายของคำ ดังแสดงในรูปที่ 3.2 และตัวอย่างในตาราง 3.1

3.2.1 ข้อมูลส่วนของการวิเคราะห์ในระดับหน่วยคำ (Morpheme Level Information)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลนี้เป็นส่วนสำคัญในการวิเคราะห์โครงสร้างของคำในประโยคภาษาไทย นั่นคือข้อมูลส่วนนี้จะเป็นข้อมูลที่จะใช้ในการตัดคำ (Word Segmentation) ซึ่งจะกล่าวต่อไปในบทที่ 4 โดยจะใช้รูปผิวของคำ (Surface of Words) เป็นตัวกำหนดว่าคำนี้เป็นคำที่มีอยู่จริงและเป็นส่วนหนึ่งในอัลกอริทึมที่จะใช้ตัดคำ ข้อมูลของคำจะอยู่ในส่วนของ tkey



รูปที่ 3.2 แสดงข้อมูลความหมายคำ กับข้อมูลระดับโครงสร้างของประโยคภาษาไทยและข้อมูลระดับความหมายของคำ

3.2.2 ข้อมูลส่วนของการวิเคราะห์ในระดับโครงสร้างของประโยคภาษาไทย (Syntax Level Information)

ข้อมูลระดับโครงสร้างของประโยคภาษาไทย จะถูกนำไปใช้ในส่วนของการวิเคราะห์โครงสร้างประโยคภาษาไทยโดยจะประกอบด้วย ประเภทของคำ ชนิดของคำ รูปแบบของคำกริยา และโครงสร้างของคำในระดับลึก ดังต่อไปนี้

3.2.2.1 ประเภทของคำ(Category)ใช้สัญลักษณ์ย่อ CAT)

ดังที่กล่าวมาแล้วในบทที่ 2 ว่าหน่วยคำจะถูกจัดเป็นหมวดหมู่ โดยแบ่งกลุ่มคำออกเป็น 13 กลุ่มได้แก่ กลุ่มคำนาม กลุ่มคำสรรพนาม กลุ่มคำกริยา กลุ่มคำช่วยกริยา กลุ่มคำบ่งชี้ กลุ่มคำกริยาวิเศษณ์ กลุ่มคำลักษณนาม กลุ่มคำสันธาน กลุ่มคำบุรพบท กลุ่มคำอุทาน กลุ่มคำที่ใช้เติมเข้าไปข้างหน้าคำอื่น กลุ่มคำลงท้าย กลุ่มคำปฏิเสธ โดยแต่ละกลุ่มจะถูกแบ่งเป็นกลุ่มย่อย (Subcategory ใช้สัญลักษณ์ SCAT) อีกครั้งเพื่อจะนำกลุ่มเหล่านี้ไปใช้ในการสร้างฐานความรู้กฎไวยากรณ์

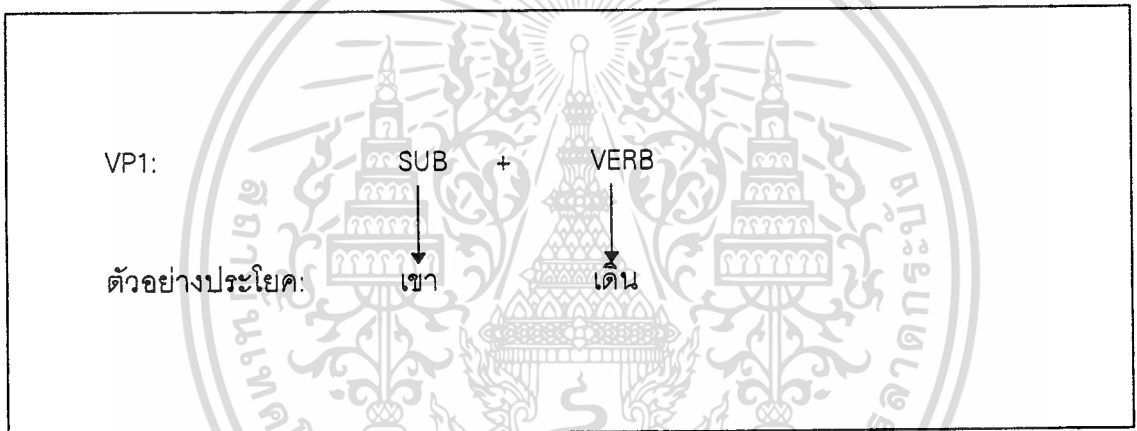
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2.2 รูปแบบของคำกริยา (Verb Pattern ใช้สัญลักษณ์ VP)

รูปแบบของคำกริยาจะเป็นข้อมูลของคำกริยาที่จะสามารถปรากฏได้ในโครงสร้างประโยครูปแบบต่างๆ ซึ่งลักษณะพิเศษนี้จะเกิดกับกลุ่มคำกริยาเท่านั้น ดังนั้นในการพิจารณาโครงสร้างของประโยค เราจึงสามารถนำข้อมูลจากคำกริยามาใช้ได้ทั้งหมด 8 ชนิดดังต่อไปนี้

รูปแบบคำกริยากลุ่มที่ 1 (VP1)

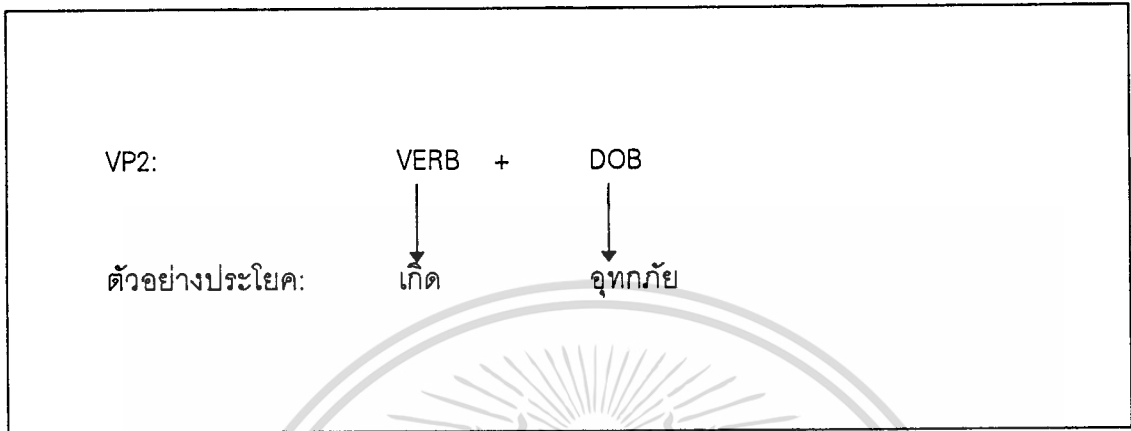
เป็นกลุ่มของรูปแบบคำกริยาที่ไม่ต้องมีกรรมมารับ ประโยคที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย ประธาน กริยา ก็มีใจความครบสมบูรณ์แล้ว ตัวอย่างของคำกริยากลุ่มนี้ได้แก่ เดิน นั่ง วิ่ง เป็นต้น



รูปที่ 3.3 แสดงรูปแบบคำกริยากลุ่มที่ 1

รูปแบบคำกริยากรุปที่ 2 (VP2)

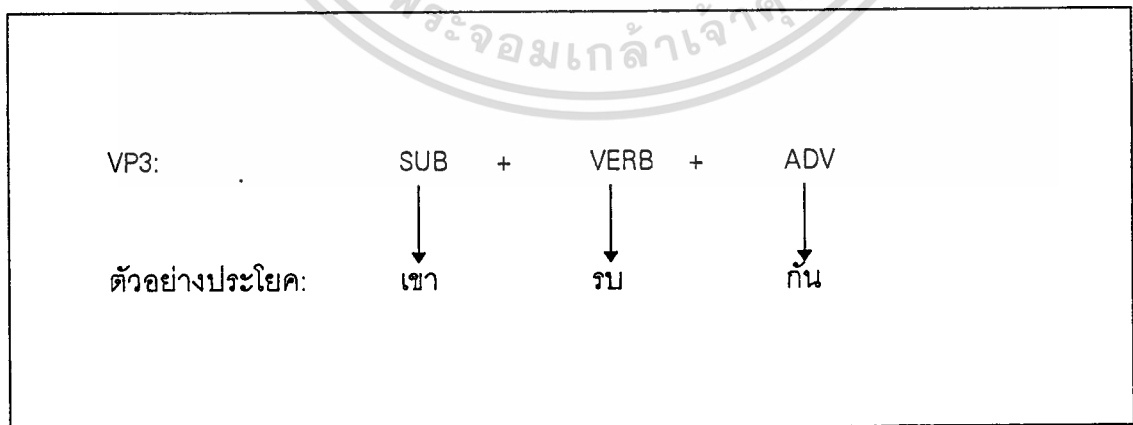
เป็นกลุ่มของรูปแบบคำกริยาที่ไม่ต้องมีประธาน ประโยคภาษาไทยที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย คำกริยา และกรรมตรง ตัวอย่างของคำในกลุ่มนี้ได้แก่ มี เกิด เป็นต้น



รูปที่ 3.4 แสดงรูปแบบคำกริยากรุปที่ 2

รูปแบบคำกริยากรุปที่ 3 (VP3)

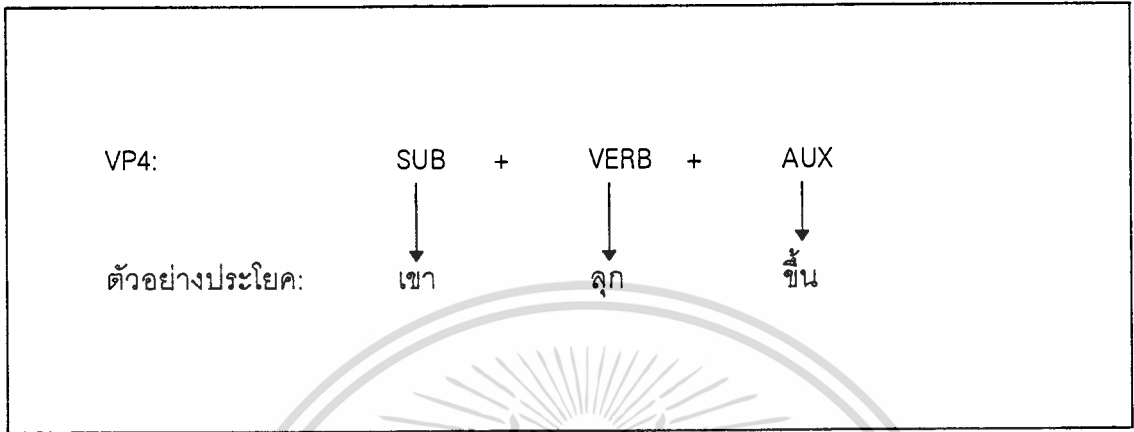
เป็นกลุ่มของรูปแบบคำกริยาที่ไม่ต้องมีกรรมมารับ แต่ต้องมีคำกริยาวิเศษณ์เพื่อให้ประโยคได้ใจความสมบูรณ์ ประโยคที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย ประธาน กริยา กริยาวิเศษณ์ ตัวอย่างของคำกริยากรุปนี้ได้แก่ รบ เป็นต้น



รูปที่ 3.5 แสดงรูปแบบคำกริยากรุปที่ 3

รูปแบบคำกริยากรุ่มที่ 4 (VP4)

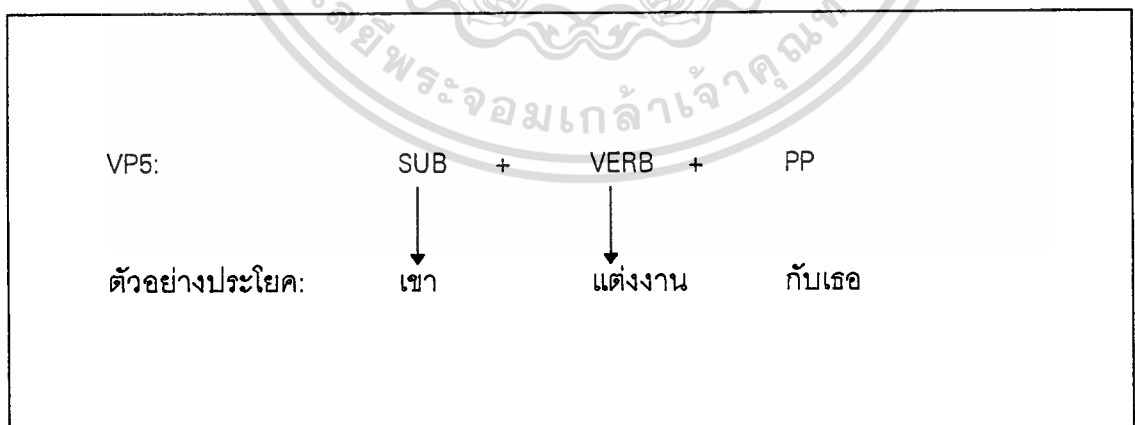
เป็นกลุ่มของรูปแบบคำกริยาที่ต้องมีคำช่วยกริยามาประกอบ ประโยคที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย ประธาน กริยา และคำช่วยกริยา ตัวอย่างของคำกริยากรุ่มนี้ได้แก่ ลูก เป็นต้น



รูปที่ 3.6 แสดงรูปแบบคำกริยากรุ่มที่ 4

รูปแบบคำกริยากรุ่มที่ 5 (VP 5)

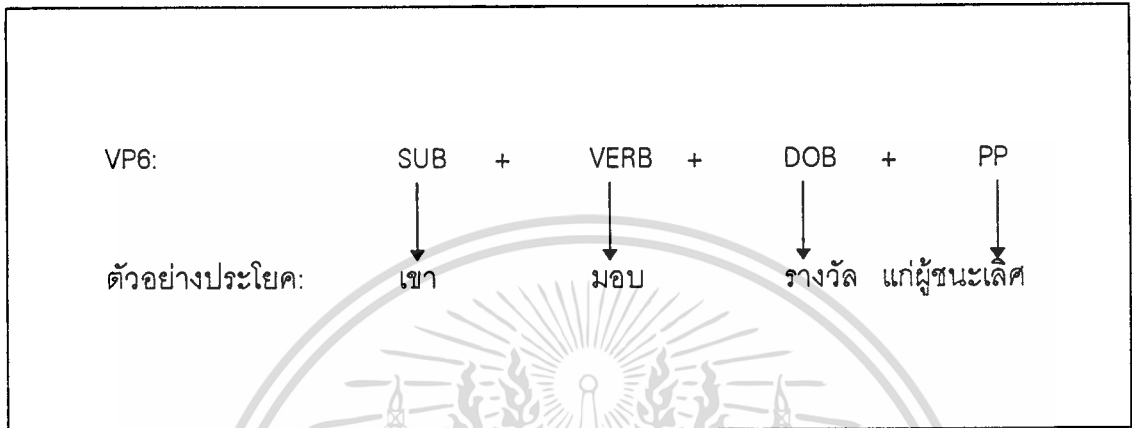
เป็นกลุ่มของรูปแบบคำกริยาที่ต้องมีบุรพทวลีมาประกอบ ประโยคที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย ประธาน กริยา และบุรพทวลี ตัวอย่างของคำกริยากรุ่มนี้ได้แก่ แต่งงาน เป็นต้น



รูปที่ 3.7 แสดงรูปแบบคำกริยากรุ่มที่ 5

รูปแบบคำกริยากลุ่มที่ 6 (VP6)

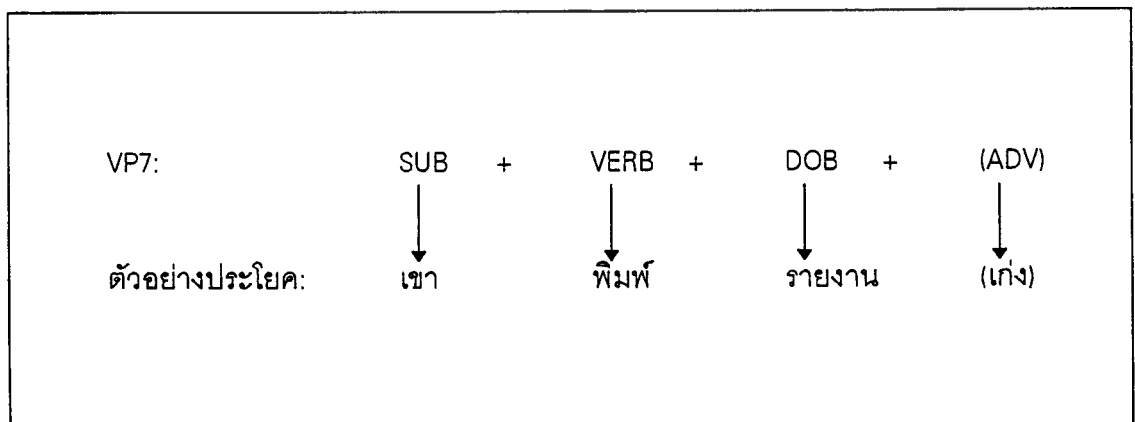
เป็นกลุ่มของรูปแบบคำกริยาที่ต้องมีกรรมตรงและบุรพทวดี ประโยคที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย ประธาน กริยา กรรมตรง และบุรพทวดี ตัวอย่างของคำกริยากลุ่มนี้ได้แก่ มอบ เป็นต้น



รูปที่ 3.8 แสดงรูปแบบคำกริยากลุ่มที่ 6

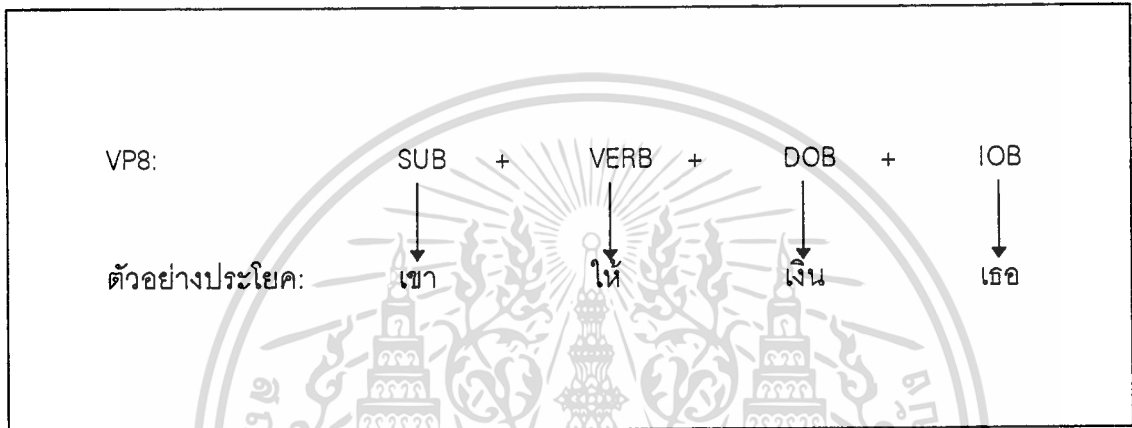
รูปแบบคำกริยากลุ่มที่ 7 (VP7)

เป็นกลุ่มของรูปแบบคำกริยาที่ต้องมีกรรมตรงมารับ ประโยคที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย ประธาน กริยา และกรรมตรง นอกจากนี้ยังอาจมีคำกริยาวิเศษณ์มาช่วยขยายความได้อีก ด้วยตัวอย่างของคำกริยากลุ่มนี้ได้แก่ พิมพ์ เป็นต้น



รูปแบบคำกริยากลุ่มที่ 8 (VP8)

เป็นกลุ่มของรูปแบบคำกริยาที่ต้องมีทั้งกรรมตรงและกรรมรอง ประโยคที่มีคำกริยาในกลุ่มนี้จะประกอบไปด้วย ประธาน กริยา กรรมตรงและกรรมรอง ตัวอย่างของคำกริยากลุ่มนี้ได้แก่ ให้ เป็นต้น



รูปที่ 3.10 แสดงรูปแบบคำกริยากลุ่มที่ 8

3.2.2.3 ชนิดของคำ (A Kind of Words ใช้สัญลักษณ์ AKO)

ชนิดของคำคือการจำแนกหมวดหมู่ของคำให้อยู่เป็นกลุ่ม โดยมีวัตถุประสงค์เพื่อจะกำหนดความหมายลึก (Deep Meaning) ของคำคำนั้น นอกจากนี้ในส่วนของทฤษฎีวิเคราะห์โครงสร้างประโยคภาษาไทย จะนำเอาชนิดของคำมาใช้เพื่อหาความสัมพันธ์ระหว่างหน่วยคำกริยา และหน่วยคำนามที่อยู่ในประโยคเดียวกัน

งานวิจัยนี้ได้แบ่งชนิดของคำภาษาไทยออกเป็นกลุ่มใหญ่ได้ 2 กลุ่มได้แก่ กลุ่มนามธรรม และกลุ่มรูปธรรม ดังนี้

ชนิดคำกลุ่มที่ 1 กลุ่มคำรูปธรรม (Concrete Words ใช้สัญลักษณ์ CCRT) แบ่งเป็น

กลุ่มย่อยที่ 1.1 สิ่งมีชีวิต (Living Thing) แบ่งเป็น

กลุ่มรหัส 1.1.1 คน (People) ตัวอย่างเช่น นายกรัฐมนตรี ครู คนไทย ช่างไม้ คนไทย

ฝรั่ง ชาวต่างชาติ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มรหัส 1.1.2 สัตว์ (Animal) ตัวอย่างเช่น นก กบ งูเหลือม คางคก ผีเสื้อ ยุง หอย
แครง ปลา เป็นต้น

กลุ่มรหัส 1.1.3 สัตว์ในวรรณคดี (Fiction Animal) ตัวอย่างเช่น ครุฑ นาค คชสีห์ ม้า
มังกร ช้างเอราวัณ เป็นต้น

กลุ่มรหัส 1.1.4 พืช (Plant) ตัวอย่างเช่น หมากรุก กุหลาบ มะลิ มะยม ประดู่ ทุเรียน
มะม่วง กล้วยา เป็นต้น

กลุ่มรหัส 1.1.5 จุลินทรีย์ (Micro Organism) ตัวอย่างเช่น เชื้อโรค จุลชีวัน จุลินทรีย์
ไวรัส เป็นต้น

กลุ่มย่อยที่ 1.2 สิ่งไม่มีชีวิต (Non-living Thing) แบ่งเป็น

กลุ่มรหัส 1.2.1 วัตถุธรรมชาติ (Natural Object) ตัวอย่างเช่น รั้งนก รั้งไหม ดาว
เคราะห์ โลก เป็นต้น

กลุ่มรหัส 1.2.2 วัตถุธาตุ (Substance) ตัวอย่างเช่น ธาตุ หน้ ไม้ แก้ว น้ เลือด
ออกซิเจน เป็นต้น

กลุ่มรหัส 1.2.3 ผืนดิน (Land) ตัวอย่างเช่น เกาะ ภูเขา ที่ราบ แหลม สมุย ชายหาด
เป็นต้น

กลุ่มรหัส 1.2.4 พื้นน้ำ (Body of Water) ตัวอย่างเช่น แม่น้ำ ทะเล ทะเลสาป คลอง คู
เป็นต้น

กลุ่มรหัส 1.2.5 สิ่งของที่มนุษย์สร้างขึ้น (Artifact) ตัวอย่างเช่น รถไฟ เรือ เกวียน ยา
ขม หมวก เป็นต้น

กลุ่มรหัส 1.2.6 ส่วนของสิ่งของที่มนุษย์สร้างขึ้น (Part of Artifact) ตัวอย่างเช่น ด้าม
ก้าน ปากขวด เป็นต้น

กลุ่มรหัส 1.2.7 อาหาร (Food) ตัวอย่างเช่น อาหาร ของกิน ทุเรียน ข้าวผัด ขนมปัง
น้ส้มสายชู เป็นต้น

กลุ่มรหัส 1.2.8 สิ่งของในวรรณคดี (Fiction Object) ตัวอย่างเช่น แก้วสารพัดนึก เป็น
ต้น

กลุ่มย่อยที่ 1.3 สถานที่ (Location) แบ่งเป็น

กลุ่มรหัส 1.3.1 พื้นที่ (Region) ตัวอย่างเช่น หมู่บ้าน อำเภอ ภาคเหนือ เขตศูนย์สูตร
นิคมอุตสาหกรรมลาดกระบัง เป็นต้น

กลุ่มรหัส 1.3.2 สถานที่ทางภูมิศาสตร์ (Geographical Area) ตัวอย่างเช่น ทุ่งหญ้า นา
ข้าว ป่าไม้ ยอดเขา เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มรหัส 1.3.3 สถานที่ในวรรณคดี (Fictional Place) ตัวอย่างเช่น สวรรค์ นรก ป่า
หิมพานต์ เขาไกรลาศ เป็นต้น

ชนิดคำกลุ่มที่ 2 กลุ่มคำนามธรรม (Abstract Words) แบ่งเป็น

กลุ่มย่อยที่ 2.1 สถานะ (State) แบ่งเป็น

กลุ่มรหัส 2.1.1 กลุ่มคำที่เกี่ยวกับเรือนร่าง (Bodily Function and Care) ตัวอย่างเช่น
คัน เป็นต้น

กลุ่มรหัส 2.1.2 การเปลี่ยนแปลง (Change) ตัวอย่างเช่น ขึ้น พัง เป็นต้น

กลุ่มรหัส 2.1.3 สัมผัส (Contact) ตัวอย่างเช่น ตี เป็นต้น

กลุ่มรหัส 2.1.4 บริโภค (Consumption) ตัวอย่างเช่น ใช้เวลา เป็นต้น

กลุ่มรหัส 2.1.5 การสร้าง (Creation) ตัวอย่างเช่น สาน ผลิต เป็นต้น

กลุ่มรหัส 2.1.6 การเคลื่อนที่ (Motion) ตัวอย่างเช่น ล่น เป็นต้น

กลุ่มรหัส 2.1.7 การรับรู้ (Perception) ตัวอย่างเช่น ได้ยิน เป็นต้น

กลุ่มย่อยที่ 2.2 การกระทำ (Action) แบ่งเป็น

กลุ่มรหัส 2.2.1 เปลี่ยนแปลง (Change) ตัวอย่างเช่น ปรับ เป็นต้น

กลุ่มรหัส 2.2.2 การสื่อสาร (Communication) ตัวอย่างเช่น พูด เป็นต้น

กลุ่มรหัส 2.2.3 แข่งขัน (Compete) ตัวอย่างเช่น แข่ง เป็นต้น

กลุ่มรหัส 2.2.4 บริโภค (Consume) ตัวอย่างเช่น ใช้เวลา เป็นต้น

กลุ่มรหัส 2.2.5 สัมผัส (Contact) ตัวอย่างเช่น ตี เป็นต้น

กลุ่มรหัส 2.2.6 กระทำด้วยความคิด (Cognitive Act) ตัวอย่างเช่น วางแผน เป็นต้น

กลุ่มรหัส 2.2.7 กระทำด้วยอารมณ์ (Emotional Act) ตัวอย่างเช่น โกรธ ร้องไห้ เป็นต้น

กลุ่มรหัส 2.2.8 เคลื่อนที่ (Move) ตัวอย่างเช่น ล่น เดิน เป็นต้น

กลุ่มรหัส 2.2.9 อยู่ (Stay) ตัวอย่างเช่น ดำรง เป็นต้น

กลุ่มรหัส 2.2.10 สร้าง (Create) ตัวอย่างเช่น ทำ เขียน เป็นต้น

กลุ่มรหัส 2.2.11 กระทำด้วยร่างกาย (Physical Act) ตัวอย่างเช่น บัสสาวะ เป็นต้น

กลุ่มรหัส 2.2.12 กระทำ (Act) ตัวอย่างเช่น ทำ ปฏิบัติ เป็นต้น

กลุ่มย่อยที่ 2.3 กิจกรรม (Activity) แบ่งเป็น

กลุ่มรหัส 2.3.1 การกระทำบนร่างกาย (Bodily Function and Care) ตัวอย่างเช่น การ
คัน เป็นต้น

กลุ่มรหัส 2.3.2 การเปลี่ยนแปลง (Change) ตัวอย่างเช่น การปรับปรุง เป็นต้น

กลุ่มรหัส 2.3.3 การสื่อสาร (Communication) ตัวอย่างเช่น การพูด เป็นต้น

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นโดยระบบอัตโนมัติของศูนย์วิจัยภาษาและการสอนภาษา
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มรหัส 2.3.4 การแข่งขัน (Competition) ตัวอย่างเช่น การแข่งขัน การประกวด เป็นต้น

กลุ่มรหัส 2.3.5 การบริโภค (Consumption) ตัวอย่างเช่น การรับประทาน การใช้ เป็นต้น

กลุ่มรหัส 2.3.6 สัมผัส (Contact) ตัวอย่างเช่น การขัดสี การตี เป็นต้น

กลุ่มรหัส 2.3.7 กระทำด้วยความคิด (Cognition) ตัวอย่างเช่น การวางแผน เป็นต้น

กลุ่มรหัส 2.3.8 การสร้าง (Creation) ตัวอย่างเช่น การสาน การทอ การเขียน เป็นต้น

กลุ่มรหัส 2.3.9 การกระทำด้วยอารมณ์ (Emotion) ตัวอย่างเช่น การโกรธ การร้องไห้ เป็นต้น

กลุ่มรหัส 2.3.10 การเคลื่อนที่ (Motion) ตัวอย่างเช่น การเดิน เป็นต้น

กลุ่มรหัส 2.3.11 การรับรู้ (Perception) ตัวอย่างเช่น การได้ยิน เป็นต้น

กลุ่มรหัส 2.3.12 การแสดงความเป็นเจ้าของ (Possession) ตัวอย่างเช่น การมี การเป็นเจ้าของ เป็นต้น

กลุ่มรหัส 2.3.13 การกระทำด้วยร่างกาย (Physical Action) ตัวอย่างเช่น การปัสสาวะ เป็นต้น

กลุ่มรหัส 2.3.14 การกระทำ (Action) ตัวอย่างเช่น การทำ การปฏิบัติ เป็นต้น

กลุ่มรหัส 2.3.15 การอยู่ (Staying) ตัวอย่างเช่น การอยู่ เป็นต้น

กลุ่มย่อยที่ 2.4 กิจกรรมทางสังคม (Social Activity) ไม่แบ่งเป็นกลุ่มย่อยและมีอยู่เพียงกลุ่มเดียว ตัวอย่างเช่น การเงิน การคลัง การธนาคาร การแต่งงาน เป็นต้น

กลุ่มย่อยที่ 2.5 ปรากฏการณ์ (Phenomenon) แบ่งเป็น

กลุ่มรหัส 2.5.1 เหตุการณ์ (Event) ตัวอย่างเช่น เหตุการณ์ อุบัติเหตุ ที่มา เศรษฐกิจ เป็นต้น

กลุ่มรหัส 2.5.2 ปรากฏการณ์ธรรมชาติ (Natural Phenomenon) ตัวอย่างเช่น ฝนตก น้ำท่วม ไฟป่า เป็นต้น

กลุ่มรหัส 2.5.3 ปรากฏการณ์ทางสังคม (Social Phenomenon) ตัวอย่างเช่น ไฟไหม้ เป็นต้น

กลุ่มรหัส 2.5.4 ปรากฏการณ์ทางสรีระ (Physiological Phenomenon) ตัวอย่างเช่น อากาศคัน ความเจ็บป่วย เป็นต้น

กลุ่มย่อยที่ 2.6 สิ่งที่เป็นนามธรรม (Abstract Thing) แบ่งเป็น

กลุ่มรหัส 2.6.1 ผลลัพธ์ของการกระทำ (Result of Action) ตัวอย่างเช่น ประกาศ
 ทฤษฎี ความรู้ เป็นต้น
 เอนไซม์เป็นเอนไซม์ที่พบในตับและไตสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มรหัส 2.6.2 สิ่งที่เป็นนามธรรมทางสังคม (Social Abstract Thing) ตัวอย่างเช่น
ชาติตระกูล เป็นต้น

กลุ่มรหัส 2.6.3 สิ่งที่เป็นนามธรรมทางประเพณี (Cultural Abstract Thing) ตัวอย่าง
เช่น มวยไทย เป็นต้น

กลุ่มรหัส 2.6.4 สิ่งประดิษฐ์ของมนุษย์ที่ไม่มีตัวตน (Immaterial Man Made Thing)
ตัวอย่างเช่น ซอฟต์แวร์ เนื้อหา เป็นต้น

กลุ่มรหัส 2.6.5 ขอบเขตของความรู้ (Field of Knowledge) ตัวอย่างเช่น เศรษฐศาสตร์
 เป็นต้น

กลุ่มรหัส 2.6.6 กีฬาและสิ่งบันเทิง (Sport and Recreation) ตัวอย่างเช่น ฟุตบอล เกม
 เป็นต้น

กลุ่มรหัส 2.6.7 สถานที่สมมุติ (Logical Place) ตัวอย่างเช่น ตลาดมืด ตลาดเสรี เป็น
 ต้น

กลุ่มย่อยที่ 2.7 ความสัมพันธ์ (Relation) แบ่งเป็น

กลุ่มรหัส 2.7.1 ที่ว่าง (Space) ตัวอย่างเช่น หน้า หลัง บน ข้างๆ เป็นต้น

กลุ่มรหัส 2.7.2 เวลา (Time) ตัวอย่างเช่น เวลา เป็นต้น

กลุ่มรหัส 2.7.3 การแสดงว่ามีอยู่ (Existence) ตัวอย่างเช่น เกิด มี ปรากฏ เป็นต้น

กลุ่มรหัส 2.7.4 กลุ่ม (Set/Group) ตัวอย่างเช่น พวก กลุ่ม คณะ เหล่า ทุกๆ เป็นต้น

กลุ่มรหัส 2.7.5 เปรียบเทียบ (Comparison) ตัวอย่างเช่น เหมือน คล้าย เท่า เป็นต้น

กลุ่มย่อยที่ 2.8 คุณสมบัติ (Attribute) แบ่งเป็น

กลุ่มรหัส 2.8.1 คุณภาพ (Quality) ตัวอย่างเช่น คุณสมบัติ คุณภาพ ลักษณะ เป็นต้น

กลุ่มรหัส 2.8.2 เงื่อนไข (Condition) ตัวอย่างเช่น สถานการณ์ เงื่อนไข ความจำเป็น
 เป็นต้น

กลุ่มรหัส 2.8.3 คุณสมบัติที่สัมผัสได้จากภายนอก (Appearance) ตัวอย่างเช่น รูปร่าง
 สี ลาย ขนาด เป็นต้น

กลุ่มรหัส 2.8.4 กลุ่ม (Set/Group) ตัวอย่างเช่น ประเภท แบบ ชนิด รุ่น เพศ เป็นต้น

กลุ่มรหัส 2.8.5 ปริมาณ (Number/Quantity) ตัวอย่างเช่น จำนวน ปริมาณ เป็นต้น

กลุ่มรหัส 2.8.6 คุณลักษณะ (Characteristic) ตัวอย่างเช่น ท่าทาง อากาาร เป็นต้น

กลุ่มย่อยที่ 2.9 มูลค่า (Value) แบ่งเป็น

กลุ่มรหัส 2.9.1 คุณภาพ (Quality) ตัวอย่างเช่น ดี ยอดเยี่ยม เลว ใจดี เป็นต้น

กลุ่มรหัส 2.9.2 เงื่อนไข (Condition) ตัวอย่างเช่น ปลอดภัย สุขสบาย ความไม่มี
 สวรรคตภาพ เป็นต้น

กลุ่มรหัส 2.9.3 คุณสมบัติที่สัมผัสได้จากภายนอก (Appearance) ตัวอย่างเช่น กลม สี เหลี่ยม แบน ลาย จุด เป็นต้น

กลุ่มรหัส 2.9.4 กลุ่ม (Set/Group) ตัวอย่างเช่น สถานงาน จำโบ้747 กระทบ เป็นต้น

กลุ่มรหัส 2.9.5 ปริมาณ (Number/Quantity) ตัวอย่างเช่น 2 นิดน้อย มาก เป็นต้น

กลุ่มรหัส 2.9.6 คุณลักษณะ (Characteristic) ตัวอย่างเช่น แน่น เป็นต้น

กลุ่มย่อยที่ 2.10 หน่วย (Unit) แบ่งเป็น

กลุ่มรหัส 2.10.1 หน่วยของวัตถุ (Unit of Object) ตัวอย่างเช่น อัน ตัว ขึ้น ไบ ลูก ฟอง ผืน บาน เครื่อง ลำ คัน ดวง เป็นต้น

กลุ่มรหัส 2.10.2 หน่วยของกลุ่มจำนวน (Unit of Collection) ตัวอย่างเช่น คู่ ชุด โหล กุลี รีม ร้อย แผง แพ เป็นต้น

กลุ่มรหัส 2.10.3 หน่วยวัด (Unit of Metrical Measurement) ตัวอย่าง เช่น ซ้อนซา ซ้อนโต๊ะ ถ้วยตวง ลิตร ตัน เมตร เดซิเบล เซลเซียส เป็นต้น

กลุ่มรหัส 2.10.4 หน่วยของความถี่ (Unit of Frequency) ตัวอย่างเช่น ครั้ง หน ที่ รอบ คราว เป็นต้น

กลุ่มรหัส 2.10.5 หน่วยของการกระทำ (Unit of Action) ตัวอย่างเช่น ห่อ กำ ม้วน มัด บัน จับ เป็นต้น

3.2.3 ข้อมูลในระดับความหมายของคำ (Semantic Level Information)

ในหลักการของภาษาศาสตร์ทางการประมวลผลภาษาธรรมชาติ (Natural Language Processing) แขนงการแปลภาษาด้วยเครื่อง (Machine Translation) นั้น ได้มีนักวิจัย หลายท่านสรุปผลการวิจัยของตนเองว่าการใช้ภาษากลาง (Intermediate Language) หรือ Interlingua ใช้สัญลักษณ์ย่อว่า IL) เป็นตัวแทนทางภาษาที่จะแสดงความหมายและโครงสร้างของหลายๆภาษาได้พร้อมกันในรูปแบบเดียว (เรียกว่า Multilingua Representation) จะเป็นวิธีการแปลหลายภาษาในระบบเดียวที่ให้ผลการแปลรวดเร็วสำหรับการแปลปริมาณมากๆ ทั้งยังประหยัดปริมาณซอฟต์แวร์และให้ผลการแปลที่ถูกต้องมากขึ้น จึงนับได้ว่าภาษากลางเป็นรูปแบบสัญลักษณ์ภาษาทางปัญญาประดิษฐ์ที่ถูกพัฒนาและกำหนดขึ้นให้เป็นตัวแทนของหลายๆภาษา ที่มีลักษณะความเป็นกลางไม่ขึ้นโดยตรงกับภาษาใด เรียกว่าเป็น "ภาษาอิสระ (Language Independent)" ดังนั้นการใช้ภาษากลางเป็นสื่อของการถ่ายทอด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลทางภาษาของประโยคจากภาษาหนึ่งไปสู่ภาษาอื่นๆ โดยเฉพาะในระบบการแปลภาษาด้วยเครื่องคอมพิวเตอร์จึงได้รับความนิยมและเป็นที่ยอมรับของนักวิจัยทั่วไป

การกำหนดส่วนประกอบหลักของภาษากลาง เพื่อแทนความหมายของภาษาต่างๆ นั้นได้มีการวิจัยและพัฒนาและนำมาใช้ในระบบแปลหลายภาษาต่างๆ ซึ่งส่วนใหญ่มีการกำหนดโครงสร้างภาษากลางโดยแบ่งเป็น 2 ส่วนได้แก่ ส่วนของหน่วยแสดงความหมาย (Unit of meaning) ของคำในภาษาดั้งเดิมเรียกว่า "Conceptual Primitive (CP)" และส่วนแสดงความสัมพันธ์ของหน่วยความหมาย เพื่อแทนความหมายของเนื้อหาในภาษาดั้งเดิม เรียกว่า ไวยากรณ์ภาษากลาง (Interlingua Grammar)

การกำหนดรูปแบบภาษากลางของงานวิจัยนี้ได้อาศัยหลักเกณฑ์ข้างต้น และได้วิจัยพัฒนาให้เหมาะสมยิ่งขึ้น เพื่อสะดวกต่อการศึกษาคำความเข้าใจและสามารถใช้เป็นตัวกลางในการถ่ายทอดข้อมูลระหว่างภาษาต่างๆ ได้ละเอียดและถูกต้องมากที่สุด ตลอดจนสะดวกต่อการนำไปประยุกต์ใช้กับภาษาต่างๆ และง่ายต่อการพัฒนาทางซอฟต์แวร์ โดยแบ่งรายละเอียดออกเป็น 3 ส่วน ได้แก่ ส่วนของหน่วยแสดงความหมายของคำ หรือ CP ส่วนแสดงความสัมพันธ์ของหน่วยคำ และส่วนแสดงรูปแบบภาษากลาง ดังจะได้อธิบายในลำดับต่อไป

3.2.3.1 หน่วยแสดงความหมายของคำหรือ CP

งานวิจัยนี้ได้กำหนด CP เพื่อเป็นตัวแทนของคำในภาษาไทย และภาษาอื่นๆ โดยพิจารณาจากความหมายลึก (Deep Meanin หรือ Concept) ของคำนั้นๆ ประกอบกับคุณสมบัติ (property) ของคำนั้นๆ และจัดหมวดหมู่คำที่มีความหมายลึกเดียวกันไว้ในกลุ่มเดียวกัน แล้วกำหนดให้มี CP ที่เป็นตัวแทนเดียวกัน ในการพิจารณาความหมายลึกเพื่อจัดกลุ่มคำดังกล่าวนี้ จะต้องอาศัยความรู้ทางวัฒนธรรมของการใช้ภาษานั้นๆ งานวิจัยนี้ได้ใช้ภาษาไทยเป็นกรณีศึกษา และความรู้ทางภาษาไทยเป็นส่วนอ้างอิงในการพิจารณากำหนดการแสดงรูปแบบของคำในรูปของ CP เพื่อแยกความหมายของคำศัพท์ที่มีรูปผิวเหมือนกันแต่ความหมายต่างกันจะใช้อักษรภาษาอังกฤษ และเขียนด้วยตัวอักษรพิมพ์ใหญ่แทนความหมายลึกของคำ

3.2.3.2 ส่วนแสดงความสัมพันธ์ของหน่วยคำ

3.2.3.21 ส่วนแสดงความสัมพันธ์ระหว่างหน่วยความหมายที่เกี่ยวกับวัตถุ

เป็นการแสดงความสัมพันธ์ระหว่างเหตุการณ์หรือสถานะกับวัตถุ (Deep Case Relation) หรือความสัมพันธ์ระหว่างหน่วยความหมายที่เป็นคำนามกับคำกริยา ความสัมพันธ์ในส่วนนี้เป็นความสัมพันธ์ทางการก (Case Relation) ซึ่งมี 6 แบบดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) ผู้กระทำ (agent ใช้สัญลักษณ์ AGT) หมายถึงความสัมพันธ์ระหว่างคำนามกับกริยาในลักษณะที่คำนามเป็นผู้ก่อการกระทำ หรือทำให้เกิดเหตุการณ์ขึ้น จะรวมทั้งสิ่งมีชีวิตและไม่มีชีวิต แต่เป็นลักษณะทำให้เกิดเหตุการณ์นั้นขึ้น ซึ่งจะมีลักษณะดังต่อไปนี้

ก. ผู้กระทำที่เป็นสิ่งมีชีวิตและมีความหมายถึงว่าเป็นคน หรือสัตว์ก็ได้ เช่น



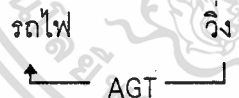
ข. ผู้กระทำ ที่เป็นองค์กร เช่น



ค. ผู้กระทำ ที่เป็นปรากฏการณ์ธรรมชาติ เช่น



ง. ผู้กระทำ ที่เป็นสิ่งไม่มีชีวิตและทำให้เกิดเหตุการณ์ขึ้น เช่น

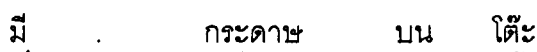


2) ผู้ถูกกระทำ (object ใช้สัญลักษณ์ OBJ) หมายถึงความสัมพันธ์ระหว่างคำนามกับคำกริยาในลักษณะที่คำนามเป็นผู้ถูกกระทำหรือผู้ที่รับสภาพหรือได้รับผลกระทบกระเทือนจากการกระทำรวมทั้งสิ่งไม่มีชีวิตที่ถูกกระทำโดยกริยา

ก. ผู้ถูกกระทำ ที่ได้รับผลจากการกระทำโดยกริยานั้นๆ เช่น



ข. ผู้ถูกกระทำ เป็นมีสถานะเป็นกรรมของประโยค เช่น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูผู้ใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) เครื่องมือ (Instrument ใช้สัญลักษณ์ INS) หมายถึงความสัมพันธ์ระหว่างคำนามกับคำกริยาในลักษณะที่คำนามเป็นเครื่องมือที่ใช้ทำกริยา หรือวิธีการที่ทำให้เกิดเหตุการณ์ขึ้นมา

ก. เครื่องมือที่มีลักษณะเป็นเครื่องมือที่ใช้ทำกริยา เช่น

เขา ขว้าง สุนัข ด้วย ก้อนหิน
 └────────── INS ─────────┘ ↑

หมายถึง ก้อนหิน เป็น INS ของกริยา ขว้าง

ข. เครื่องมือที่เป็นวิธีการให้เกิดเหตุการณ์เช่น

นักศึกษา แก้ ปัญหา ด้วย การวางแผน
 └────────── INS ─────────┘ ↑

4) จุดหมาย (Target ใช้สัญลักษณ์ TAR) หมายถึงความสัมพันธ์ระหว่างคำนามกับกริยาในลักษณะบ่งบอกถึงจุดมุ่งหมาย จุดปลายทาง จุดสิ้นสุด ของกริยา หรือเป้าหมายของกริยาที่มีได้เป็นกรรมแท้ของกริยาเช่น

นักเรียน ไป ที่ โรงเรียน
 └────────── TAR ─────────┘ ↑

หมายถึง โรงเรียน เป็น TAR ของกริยา ไป

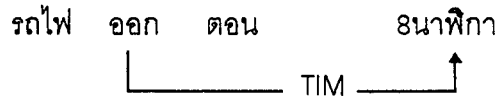
5) ต้นกำเนิด (Source ใช้สัญลักษณ์ SOR) หมายถึงความสัมพันธ์ระหว่างคำนามกับกริยาในลักษณะบอกจุดตั้งต้นของการกระทำซึ่งแสดงโดยกริยาหรือที่มาของสิ่งที่ส่งจากผู้ให้ไปยังผู้รับ หรือที่มาของเหตุการณ์ เช่น

เขา ออก จาก บ้าน
 └────────── SOR ─────────┘ ↑

หมายถึง บ้าน เป็น SOR ของกริยา ออก

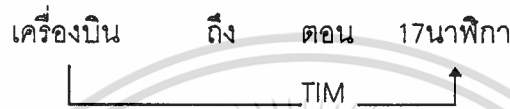
6) เวลา (Time ใช้สัญลักษณ์ TIM) หมายถึงความสัมพันธ์ระหว่างคำนามกับกริยาในลักษณะที่กล่าวถึงเวลาที่เริ่มต้น เวลาจบสิ้น ช่วงเวลาที่ใช้ในกริยานั้นๆ โดยลักษณะของเวลาจะเอกกำหนดโดย Subfeature หรือสภาพที่แสดงโดยกริยาได้แก่ ทำนั้น ไม่นุญาติให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก. เวลาที่เป็นเวลาเริ่มต้น เช่น



หมายถึง 8นาฬิกา เป็น TIM ของกริยา ออก

ข. เวลาที่เป็นเวลาจบสิ้น เช่น



ค. เวลาที่เป็นช่วงเวลา เช่น



ง. เวลาที่เป็นสภาพของการแสดงโดยกริยา เช่น

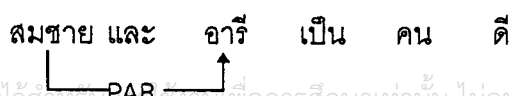


3.2.3.2.2 ส่วนแสดงความสัมพันธ์ระหว่างหน่วยความหมายภายในวัตถุ

เป็นส่วนแสดงความสัมพันธ์ระหว่างหน่วยความหมายที่ไม่ใช่คำนามกับคำกริยา ซึ่งมี 9 แบบ ดังนี้

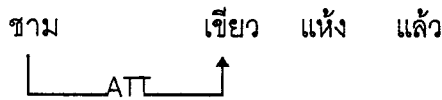
1) ตัวเชื่อมขนาน (Parallel ใช้สัญลักษณ์ PAR) หมายถึง ความสัมพันธ์ระหว่างคำนามมากกว่าหนึ่งกับคำกริยาในลักษณะที่กระทำอาการร่วมกัน หรือคำนามที่สามารถกระทำอาการมากกว่าหนึ่งอาการในเวลาเดียวกัน หรือเหตุการณ์ที่เกิดต่อเนื่องกันได้แก่

ก. PAR ของการกระทำที่ร่วมกันของคำนามมากกว่าหนึ่งเช่น

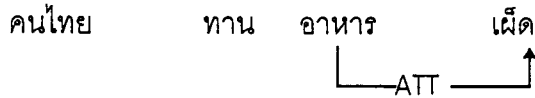


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับครูใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

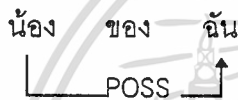
ค. สี เป็นความสัมพันธ์ที่แสดงสีของวัตถุ เช่น



ง. รสชาติ เป็นความสัมพันธ์ที่แสดงรสชาติของวัตถุ เช่น

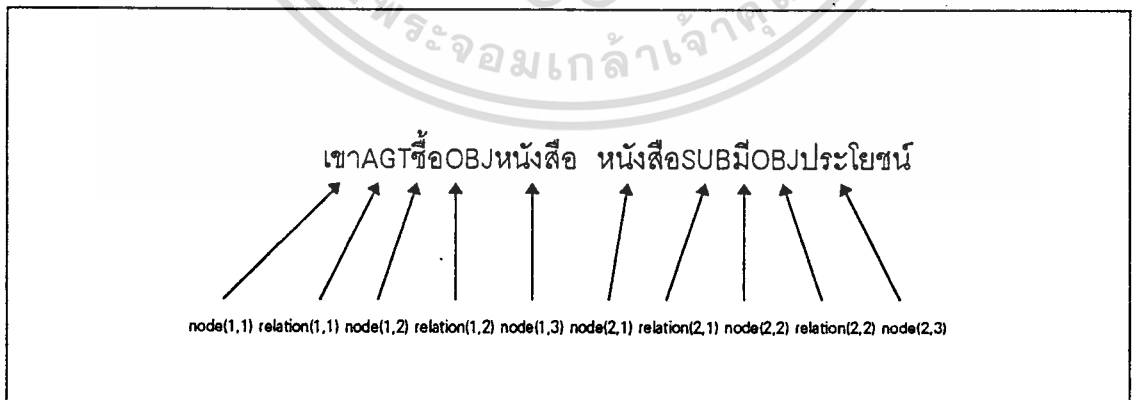


4) ผู้เป็นเจ้าของ (Possessor ใช้สัญลักษณ์ POSS) หมายถึงความสัมพันธ์ระหว่างคำนามกับคำคุณศัพท์แสดงความเป็นเจ้าของ เช่น



3.2.3.2.3 ส่วนแสดงรูปทางโครงสร้างของประโยคภาษาไทย

ในส่วนนี้จะป็นลักษณะโครงข่ายแสดงความสัมพันธ์ โดยแสดงโครงสร้างของประโยคภาษาไทยในรูปแบบของโนด (NODE) และข้อความแสดงความสัมพันธ์ (Relation) โดยโนดจะแทนหน่วยคำแต่ละหน่วยคำ และข้อความแสดงความสัมพันธ์จะแสดงความสัมพันธ์ระหว่างหน่วยคำโนดในลักษณะของคู่ความสัมพันธ์ที่ขึ้นต่อกัน (Head-Dependent Relation) ตัวอย่างเช่น



รูปที่ 3.11 แสดงรูปประโยคภาษาไทยที่ได้ถูกวิเคราะห์แล้วโดยแสดงออกมาในรูปของหน่วยคำและความสัมพันธ์ระหว่างหน่วยคำ

3.3 โครงสร้างข้อมูลของฐานข้อมูลพจนานุกรม (Data Structure of Dictionary System)

งานวิจัยนี้ได้นำเอาข้อมูลทั้งสองส่วนที่ได้กล่าวมาแล้วในหัวข้อ 3.2.1 และ 3.2.2 มาสร้างเป็นฐานข้อมูลพจนานุกรมคำศัพท์ภาษาไทย โดยใช้วิธีการค้นหาข้อมูลแบบ Fast Word Matching [7] ดังจะได้อธิบายรายละเอียดในหัวข้อต่อไป

3.3.1 การกำหนดดัชนีของกลุ่มคำศัพท์

คำศัพท์ในพจนานุกรมไทยทั้งหมดจะถูกจัดให้เป็นหมวดหมู่ และเรียงเป็นลำดับก่อนหลังในการเก็บบันทึกไว้ในฐานข้อมูลตามค่าของดัชนี 3 ตัวที่คิดจากค่า แอสกี้นัมเบอร์ของอักขระตัวที่ 1 ค่าแอสกี้นัมเบอร์ของอักขระตัวที่ 2 และจำนวนอักขระทั้งหมดของคำศัพท์นั้นๆ โดยที่กำหนดให้ค่า แอสกี้นัมเบอร์ของอักขระตัวแรกของคำศัพท์เป็นดัชนีที่แบ่งคำศัพท์ออกเป็นกลุ่มใหญ่ และให้ค่าแอสกี้นัมเบอร์ของอักขระตัวที่สองของคำศัพท์นั้นเป็นดัชนีตัวที่สอง ที่แบ่งคำศัพท์ภายในกลุ่มใหญ่ให้เป็นกลุ่มย่อย และคำศัพท์ภายในกลุ่มย่อยจะถูกจัดเรียงเป็นลำดับตามจำนวนของคำนั้นๆ โดยให้จำนวนอักขระมากอยู่ในตำแหน่งต้นๆของกลุ่มคำศัพท์นั้นๆ ตามลำดับ

ในที่นี้จะกำหนดให้ V เป็นคำศัพท์ทั้งหมดที่บรรจุอยู่ในพจนานุกรมคำศัพท์ภาษาไทยของระบบ

$I_1(j)$ เป็นดัชนีตัวแรก (First Index) ที่แสดงกลุ่มคำศัพท์ใหญ่กลุ่มที่ j ($j = 1, 2, \dots$)

$I_{1,2}(k)$ เป็นดัชนีตัวที่ 2 (Second Index) ที่แสดงกลุ่มคำศัพท์ย่อยกลุ่มที่ k ภายในกลุ่มคำศัพท์ใหญ่กลุ่มที่ j ($k=1, 2, \dots$)

$I_{1, k, 3}(N)$ เป็นดัชนีตัวที่ 3 (Third Index) ที่แสดงกลุ่มคำศัพท์ย่อยกลุ่มที่ N โดยที่คำศัพท์นั้นจะอยู่ในกลุ่มย่อยที่ k และในกลุ่มใหญ่ที่ j

เราจะแสดงความสัมพันธ์ของกลุ่มคำศัพท์ในโครงสร้างพจนานุกรมไทยในรูปของเซต (Set) ได้ดังนี้

$$V = \{ I_1(j) \{ I_{1,2}(k) \{ I_{1,k,3}(n) \} \} \} \quad \dots(1)$$

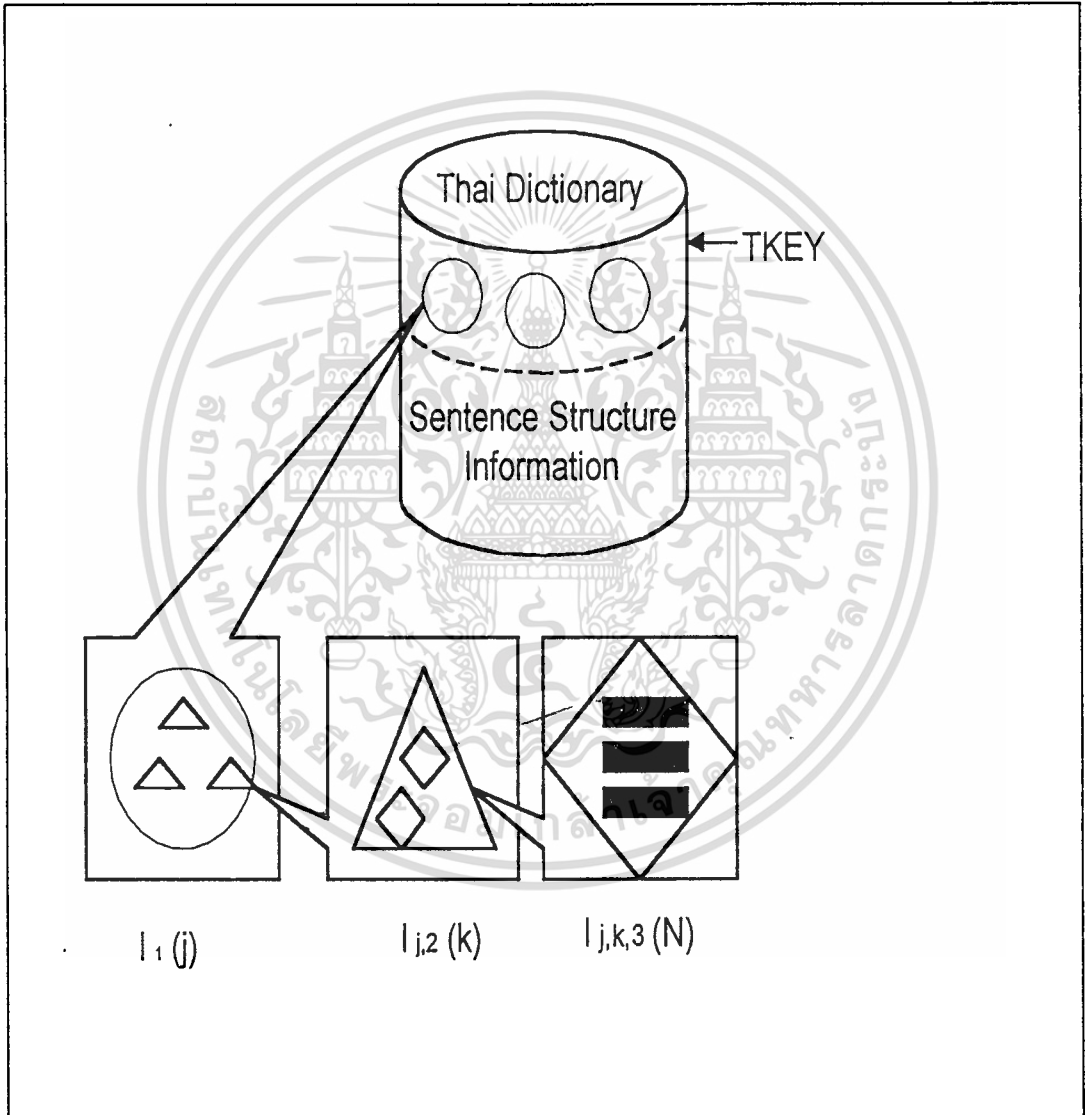
โดยที่

$$I_1(j) = \text{แอสกี้นัมเบอร์ของอักขระตัวแรกของคำศัพท์} \quad \dots(2)$$

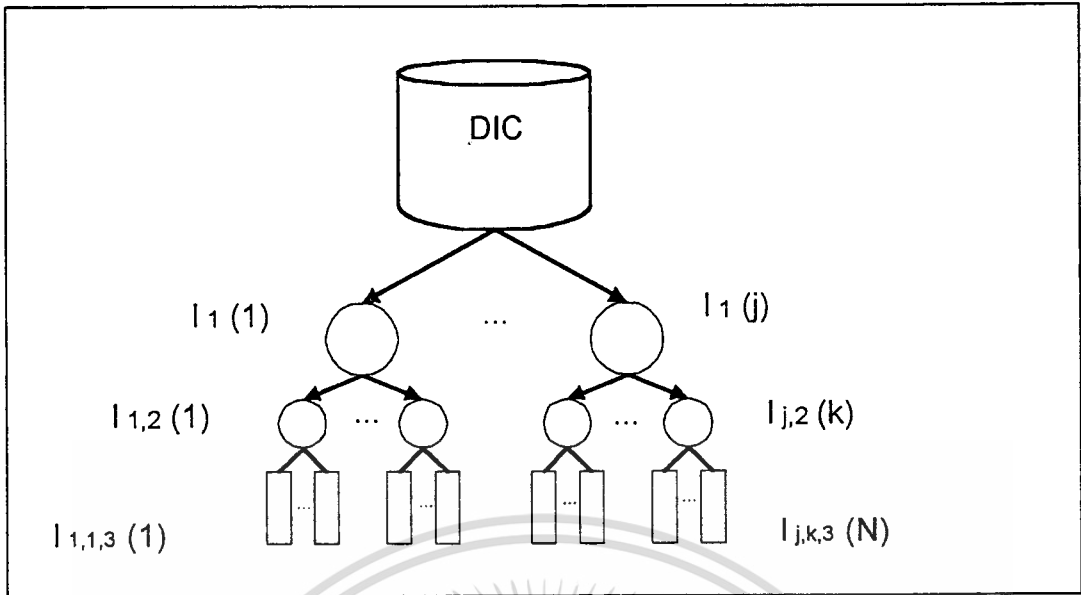
เอกสารนี้เผยแพร่โดยศูนย์บริการข้อมูลสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ไม่อนุญาตให้เผยแพร่ไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$I_{j,k,3}(N) = n \text{ เมื่อ } n = \text{จำนวนอักขระของคำศัพท์} \quad \dots(4)$$

จากนิยามของสัญลักษณ์ V , $I_1(j)$, $I_{j,2}(k)$, $I_{j,k,3}(N)$ ข้างต้นทั้งหมด เราสามารถแสดงความสัมพันธ์ของกลุ่มคำศัพท์ และการจัดเรียงคำศัพท์ทั้งหมดที่ประกอบเป็นโครงสร้างฐานข้อมูลพจนานุกรมคำศัพท์ไทยที่แสดงในรูปที่ 3.12 และรูปที่ 3.13 นั้น ได้ใช้วิธีการแสดงความสัมพันธ์แบบชาร์ต (chart) และแบบ ต้นไม้ (Tree) ตามลำดับ เพื่อให้ให้เห็นรายละเอียดของโครงสร้างมากขึ้น

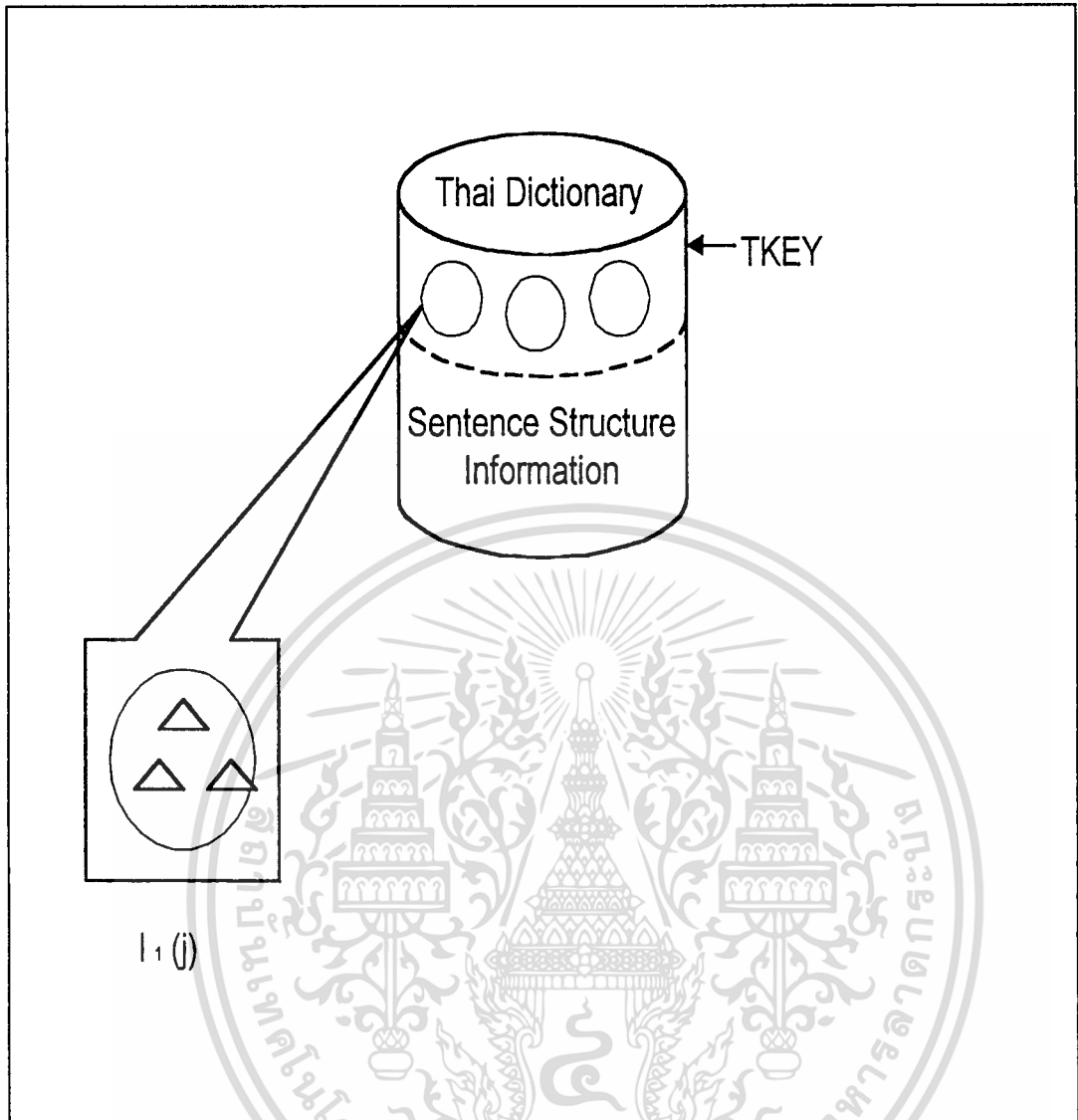


รูปที่ 3.12 แสดงโครงสร้างดัชนีของฐานข้อมูลพจนานุกรมคำศัพท์ภาษาไทย

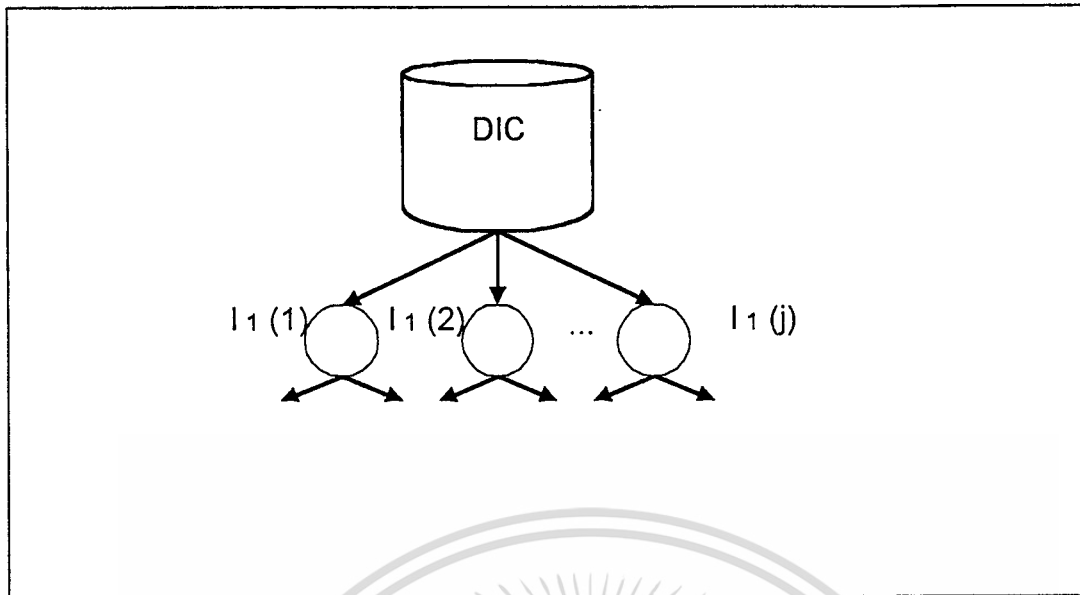


รูปที่ 3.13 แสดงโครงสร้างดัชนีของพจนานุกรมคำศัพท์ภาษาไทย(ในรูปแบบของ แผนภูมิต้นไม้)

จากการศึกษาค้นคว้าโครงสร้างของคำในภาษาไทยพบว่า อักขระที่จะเป็นตัวอักษรตัวแรกของคำศัพท์ทั้งหมดในภาษาไทยมีทั้งหมด 44 ตัวอักษร แบ่งเป็นพยัญชนะ 39 ตัวอักษร(โดยยกเว้นตัวอักษร ข ค ฤ ฦ ษ) และสระอีก 5 ตัวอักษร(แ ไ โ ใ) นั่นคือ เราจะได้ค่าของ $I_1(j)$ ทั้งหมด 44 ค่า โดยที่ j จะแสดงลำดับของกลุ่มคำศัพท์หรือ $j = 1, 2, 3, \dots, 44$ เป็นผลให้การแบ่งคำศัพท์ภาษาไทยออกเป็นกลุ่มใหญ่ๆตามอักขระตัวแรกนี้จะได้ทั้งหมด 44 กลุ่ม โดยมีดัชนี $I_1(1), I_1(2), \dots, I_1(44)$ เป็นตัวชี้บอกกลุ่มของคำศัพท์เหล่านั้นภายในโครงสร้างของพจนานุกรมคำศัพท์ภาษาไทยดังแสดงไว้ในตารางที่ 3.2 ซึ่งเป็นตารางกำหนดค่าของ $I_1(j)$ ตามลำดับของอักขระในภาษาไทยและจากสมการที่ (2) เราจะเรียกกลุ่มคำศัพท์ที่ถูกแบ่งเป็นกลุ่มใหญ่ๆ ตามค่าของแอสกินัมเบอร์ของอักขระเหล่านี้ว่า “กลุ่มคำศัพท์ชั้นที่ 1 (First Level Vocabulary)”



รูปที่ 3.14 แสดงความสัมพันธ์ของการแบ่งกลุ่มคำศัพท์ภาษาไทยชั้นที่ 1



รูปที่ 3.15 แสดงความสัมพันธ์ของการแบ่งกลุ่มคำศัพท์ภาษาไทยชั้นที่ 1 (ในรูปของแผนภูมิต้นไม้)

อักขระตัวแรกของคำศัพท์	แอสกินัมเบอร์ตัวแรกของคำศัพท์	กลุ่มดัชนีตัวที่ 1	ตัวอย่าง
ก	161	$I_1 (1)$	กลอง, กรุงเทพ
ข	162	$I_1 (2)$	ขลุ่ย, ข้าวเจ้า
ค	164	$I_1 (3)$	คนดี, ครึ่ง, ความสนใจ
ง	165	$I_1 (4)$	งานเลี้ยง, เงิน
:	:	:	:
อ	205	$I_1 (38)$	อยู่, ออกมา
ฮ	206	$I_1 (39)$	ฮกเกี้ยน, ฮังเล
เ	224	$I_1 (40)$	เก่ง, เกิน, เพราะว่า
แ	225	$I_1 (41)$	แต่, แล้ว
โ	226	$I_1 (42)$	โดยธรรมดา, โดยเร็ว
ใ	227	$I_1 (43)$	ไปที่, ไป, ได้
ใ	228	$I_1 (44)$	ใช้, ใน, ให้

ตารางที่ 3.2 แสดงการกำหนดค่าดัชนีตัวที่ 1 ของกลุ่มคำศัพท์ภาษาไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับการกำหนดค่าของ $l_{j,2}(k)$ หรือดัชนีตัวที่ 2 ตามสมการที่ (3) เพื่อแบ่งคำศัพท์ชั้นที่ 1 ให้เป็นกลุ่มย่อยตามค่าแอสกินัมเบอร์ของอักขระตัวที่สองของคำศัพท์นั้น จะต้องกำหนดค่า k ให้สัมพันธ์กันกับการปรากฏของอักขระตัวที่สองในคำศัพท์ภาษาไทย นั่นคือจำนวนกลุ่มย่อยของคำศัพท์ภายในกลุ่มคำศัพท์ชั้นที่ 1 แต่ละกลุ่มจะมีจำนวนเท่าใดนั้นขึ้นอยู่กับความเป็นไปได้ในหลักภาษาไทยของการผสมอักขระตัวที่สองกับอักขระตัวที่หนึ่งของกลุ่มคำศัพท์ชั้นที่หนึ่งดังกล่าว

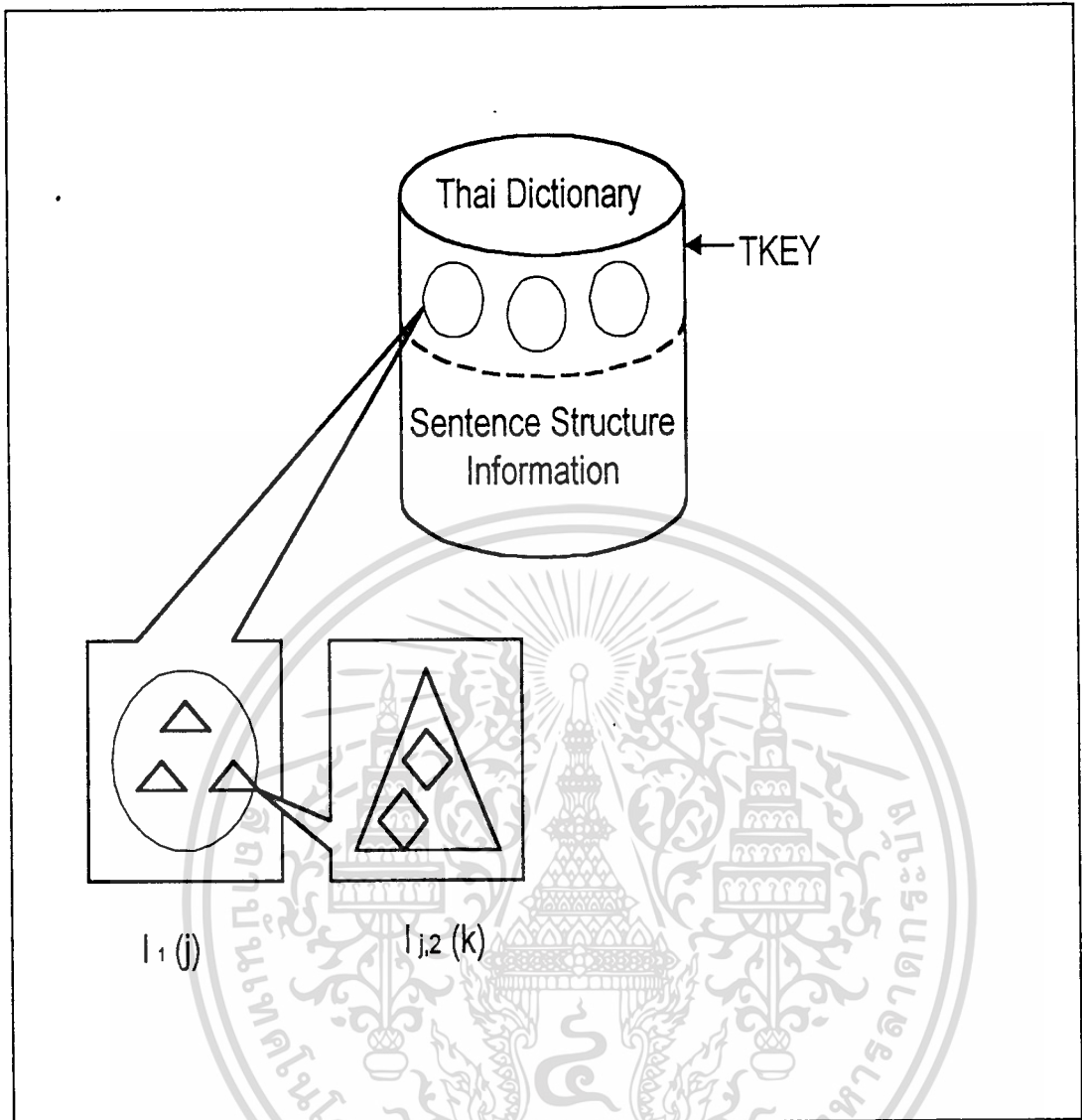
ในที่นี้กำหนดให้ k_j เป็นจำนวนอักขระตัวที่สองที่สามารถผสมกับอักขระตัวแรกของคำศัพท์ชั้นที่หนึ่งในกลุ่ม $l_{j,1}$ ดังนั้นจะได้ $k = 1, 2, \dots, k_j$ โดยที่ $k_j > 0$ เราจะแสดงความสัมพันธ์ของ $l_{j,1}$ และ $l_{j,2}(k)$ ในรูปของเซตได้ดังสมการที่ 5

$$l_{j,1} = \{ l_{j,2}(k) \} \quad \begin{array}{l} j = 1, 2, \dots, 44 \\ k = 1, 2, \dots, k_j \end{array} \quad \dots(5)$$

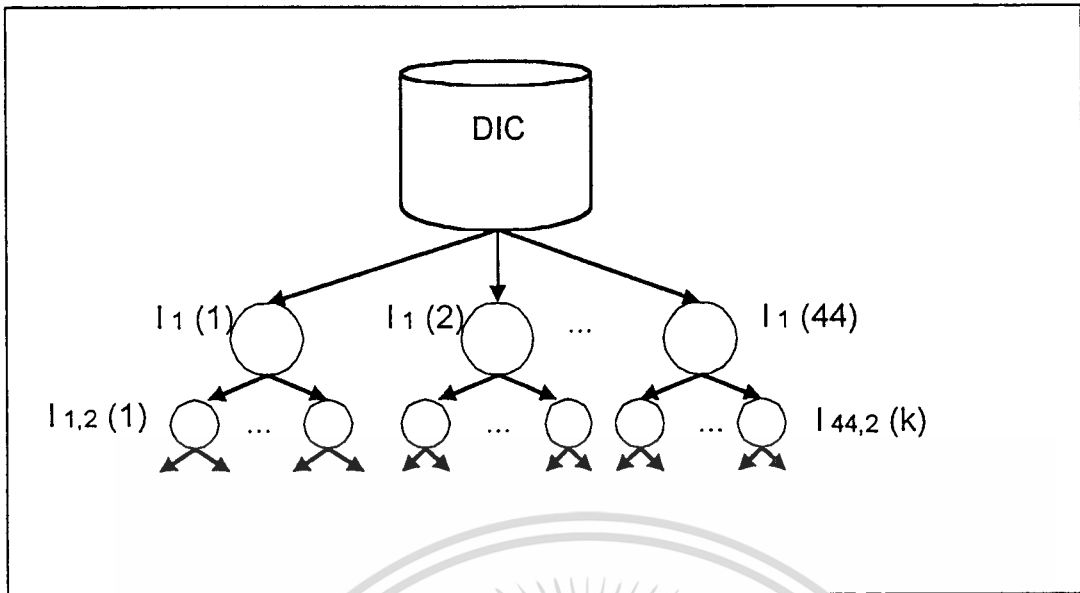
เราจะเรียกกลุ่มคำศัพท์ย่อยที่ถูกแบ่งตามค่าของแอสกินัมเบอร์ของอักขระตัวที่สองของคำศัพท์ ซึ่งถูกขีบออกด้วยดัชนี $l_{j,2}(k)$ ว่า “กลุ่มคำศัพท์ชั้นที่ 2 (Second Level Vocabulary)”

คำศัพท์ชั้นที่ 1		คำศัพท์ชั้นที่ 2			
อักขระตัวที่ 1 ของคำศัพท์	ดัชนีตัวที่ 1	อักขระตัวที่สองของคำศัพท์	แอสกินัมเบอร์ของอักขระตัวที่ 2	ดัชนีตัวที่ 2	ตัวอย่าง
ก	$l_{1,1}(1)$	ก	161	$l_{1,2}(1)$	กกข้าง
		ง	167	$l_{1,2}(2)$	กงสุล
		ช	172	$l_{1,2}(3)$	กชกร
		:	:	:	

ตารางที่ 3.3 แสดงให้เห็นถึงโครงสร้างของการแบ่งคำศัพท์ชั้นที่ 2 ภายในกลุ่มคำศัพท์ชั้นที่ 1



รูปที่ 3.16 แสดงความสัมพันธ์ของการแบ่งคำศัพท์ชั้นที่ 2 ภายในคำศัพท์ชั้นที่ 1



รูปที่ 3.17 แสดงความสัมพันธ์ของการแบ่งคำศัพท์ชั้นที่ 2 ภายในคำศัพท์ชั้นที่ 1 (ในรูปของแผนภูมิต้นไม้)

ท้ายที่สุดที่จะกล่าวถึงการกำหนดดัชนีตัวที่ 3 ได้แก่ $l_{i,k,3}(N)$ ตามสมการที่ 4 ดังที่กล่าวมาแล้วข้างต้นว่า เรากำหนดค่าดัชนีตัวที่ 3 จากจำนวนอักขระทั้งหมด นั่นคือค่า n ของคำศัพท์นั้นๆ เพื่อจัดแบ่งและเรียงคำศัพท์เป็นกลุ่มย่อยๆ ภายในกลุ่มคำศัพท์ชั้นที่สอง แต่ละกลุ่มตามค่าของ n เราเรียกการแบ่งคำศัพท์เป็นกลุ่มย่อยๆ ตามค่าของจำนวนอักขระนี้ว่า “กลุ่มคำศัพท์ชั้นที่ 3 (Third Level Vocabulary)” จากการศึกษาคุณสมบัติของคำในภาษาไทยพบว่า คำที่สั้นที่สุดในภาษาไทยจะต้องประกอบด้วยจำนวนอักขระอย่างน้อย 2 ตัว ดังนั้น จำนวนอักขระของคำใดๆจะเป็น

$$n \geq 2 \quad \dots(6)$$

ในทางตรงกันข้าม เราไม่สามารถกำหนดจำนวนอักขระสูงสุดของคำศัพท์ชั้นที่ 3 แต่ละกลุ่มได้ อย่างไรก็ตามเราทราบว่าค่า n ของคำศัพท์ใดๆ จะเริ่มจากเลขจำนวนเต็ม 2 และตามด้วย 3, 4, 5, ... ตามลำดับ ถ้ากำหนดให้ $M_{i,k}$ เป็นลำดับที่มีจำนวนอักขระสูงสุดของกลุ่มคำศัพท์ชั้นที่ 3 (ที่แสดงด้วยดัชนีตัวที่ 3 ได้แก่ $l_{i,k,3}(M_{i,k})$) ภายในกลุ่มคำศัพท์ชั้นที่ 2 กลุ่มที่ k และคำศัพท์ชั้นที่ 1 กลุ่มที่ j ตามลำดับ ดังนั้นจำนวนอักขระของคำศัพท์ใดๆจะเป็น

$$M_{i,k} \geq n \geq 2 \quad \dots(7)$$

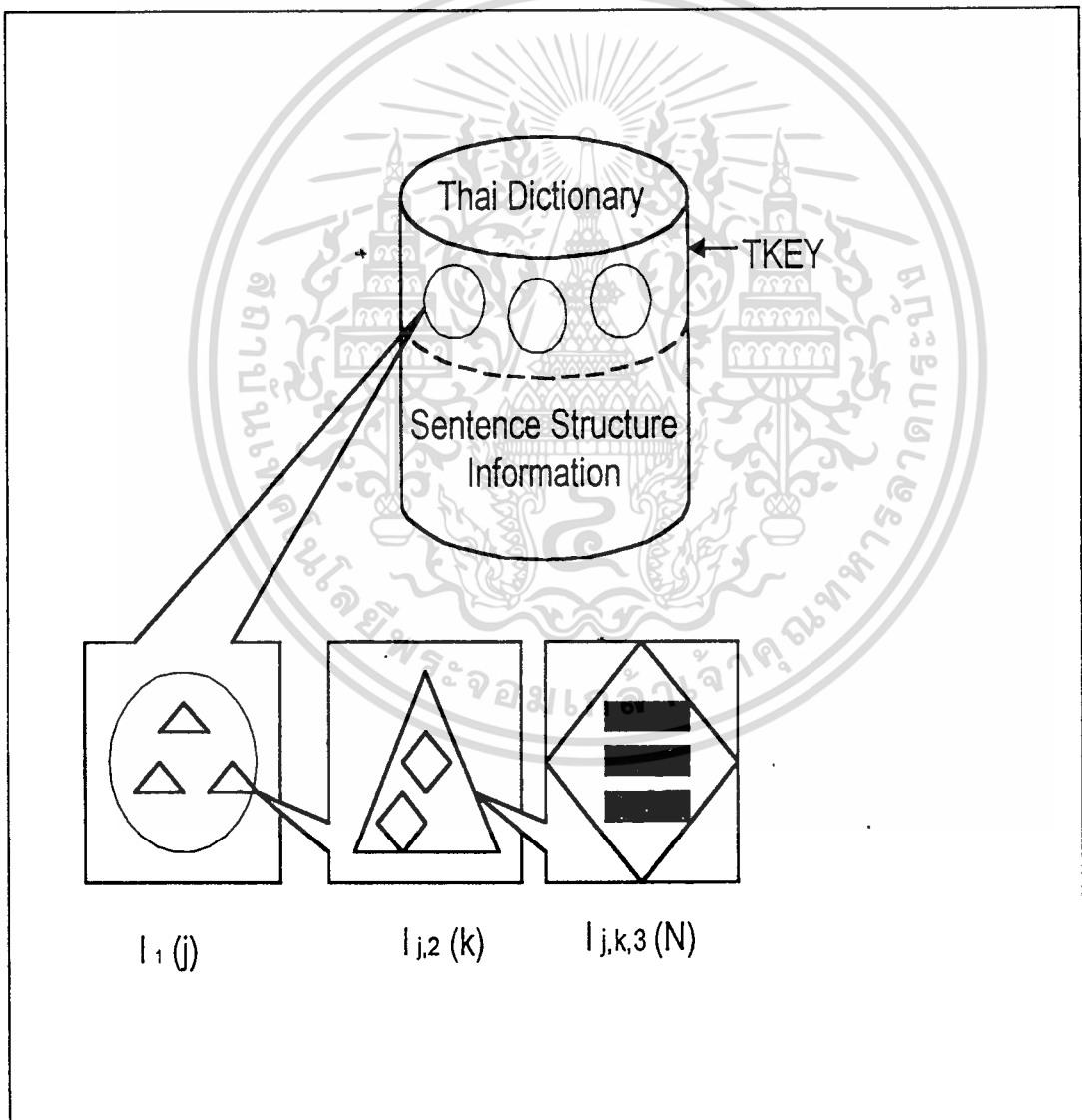
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับเมื่อ $j = 1, 2, \dots, 44$ เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$k = 1, 2, \dots, k_j$$

ในระบบ Fast Word Matching นี้ เราได้จัดเรียงคำศัพท์ชั้นที่ 3 ภายในพจนานุกรมไทยไปตามลำดับค่าของ n โดยให้ค่ามากอยู่ตำแหน่งต้นๆ ของการเก็บบันทึกลงในพจนานุกรมไทย กล่าวคือ สำหรับ j และ k ใดๆ การเรียงคำศัพท์ชั้นที่ 3 ภายในพจนานุกรมจะเป็นดังนี้

$$l_{j,k,3}(M_{j,k}) > l_{j,k,3}(M_{j,k}-1) > \dots > l_{j,k,3}(3) > l_{j,k,3}(2) \quad \dots(8)$$

หมายเหตุ เครื่องหมาย $A > B$ หมายความว่า A มีลำดับมาก่อน B



รูปที่ 3.18 แสดงความสัมพันธ์ของการจัดแบ่งคำศัพท์ทั้ง 3 ชั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 การใช้งานฐานข้อมูลพจนานุกรมคำศัพท์

จากรูปที่ 3.1 แสดงให้เห็นถึงการติดต่อกันระหว่าง ระบบฐานข้อมูลพจนานุกรม กับส่วน การวิเคราะห์หน่วยคำ และส่วนวิเคราะห์โครงสร้างประโยคภาษาไทย ซึ่งการทำงานของส่วนการ ติดต่อระหว่าง ฐานข้อมูลพจนานุกรมคำศัพท์ กับ ระบบวิเคราะห์หน่วยคำ จะมีรายละเอียดดังต่อไปนี้

3.4.1 การใช้งานฐานข้อมูลพจนานุกรมร่วมกับระบบวิเคราะห์หน่วยคำ

ข้อมูลส่วนของการวิเคราะห์ในระดับหน่วยคำ เป็นข้อมูลส่วนที่ใช้ในการวิเคราะห์แบ่งคำซึ่งใน ส่วนติดต่อนี้ระบบวิเคราะห์หน่วยคำจะใช้โปรแกรมย่อย DicChk () ในการติดต่อกับฐานข้อมูล พจนานุกรมคำศัพท์ภาษาไทย ซึ่งการทำงานของโปรแกรมย่อย DicChk () จะมีดังต่อไปนี้

สมมติให้ $C_1C_2C_3...C_n$ เป็นลำดับของอักขระของข้อความที่ต้องการค้นหาคำศัพท์ ($n=1,2,3,...$) เมื่อระบบวิเคราะห์หน่วยคำต้องการจะค้นหาคำที่อยู่ภายในข้อความต้นแบบเพื่อที่จะนำไปใช้ในการ วิเคราะห์แยกหน่วยคำ ระบบวิเคราะห์หน่วยคำจะเรียกใช้โปรแกรมย่อย DicChk() โดยให้ $C_1C_2C_3...C_n$ เป็นอินพุทของโปรแกรมย่อย ดังนั้นโปรแกรมย่อย DicChk($C_1C_2C_3...C_n$) จะทำการวิเคราะห์ $C_1C_2C_3...C_n$ โดยจะตรวจสอบว่ามีคำในพจนานุกรมที่ค่าที่ค่าเหล่านั้นแต่ละค่า มีอักขระที่ตัวเริ่มต้นด้วย $C_1C_2C_3...C_m$ ($m \leq n$) ดังนั้นโปรแกรมย่อย DicChk () จึงใช้วิธีการค้นหาคำศัพท์แบบ Fast Word Matching ในการค้นหาคำศัพท์ที่ต้องการโดยการใช้ดัชนีตัวที่ 1 คือ $I_1(j)$ ค้นหากลุ่มคำที่มีแอสกี นัมเบอร์ของอักขระตัวที่มีค่าเท่ากับแอสกี นัมเบอร์ของตัวอักษร C_1 และดัชนีตัวที่ 2 จะนำตัวอักษรตัว ต่อไปคือ C_2 มาทำการพิจารณาโดยดัชนีตัวที่ 2 คือ $I_{j,2}(k)$ จะมีค่าเท่ากับแอสกี นัมเบอร์ของตัวอักษร C_2 เป็นตัวเลือกรวมคำย่อยจากกลุ่มคำใหญ่ที่ถูกแบ่งไว้ก่อนแล้วด้วยดัชนีตัวที่ 1 เพื่อหาว่าอักขระที่ เหลือคือ $C_3C_4C_5...C_n$ มีอยู่ในกลุ่มคำศัพท์ย่อยเหล่านั้นหรือไม่ และ ดัชนีตัวที่สามคือ $I_{j,k}(N)$ จะเป็น ดัชนีที่เลือกรวมคำที่ถูกแบ่งมาก่อนหน้านี้แล้วด้วยดัชนีตัวที่ 1 และดัชนีตัวที่ 2 โดยจะนำกลุ่มของคำ ศัพท์ที่ยาวที่สุดในกลุ่มคำศัพท์ย่อยที่ถูกแบ่งด้วยดัชนีตัวที่ 1 และดัชนีตัวที่ 2 มาทำการเปรียบเทียบกับ $C_1C_2C_3...C_n$ จากนั้นก็จะนำคำศัพท์ในกลุ่มคำศัพท์ย่อยกลุ่มต่อไปที่มีขนาดสั้นกว่ามาเปรียบเทียบต่อ จนกว่าจะครบทุกกลุ่มคำศัพท์ย่อยที่ถูกแบ่งด้วยดัชนีตัวที่ 1 และดัชนีตัวที่ 2 นั้น โดยมีขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 โปรแกรมย่อย DicChk() รับกลุ่มอักขระ $C_1C_2C_3...C_n$ มาจากระบบวิเคราะห์หน่วยคำด้วย คำสั่ง DicChk($C_1C_2C_3...C_n$)

ขั้นตอนที่ 2 นำอักขระตัวที่ 1 คือ C_1 มาหาดัชนีคำศัพท์กลุ่มที่ 1 โดยที่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$I_{1,j}(j) = \text{แอสกี้นัมเบอร์ของ } C_1 \\ = A(C_1)$$

โดยที่ $A(C_n)$ เป็นแอสกี้นัมเบอร์ของ C_n ($n=1,2,3,\dots$)

ดังนั้นเราจะได้ดัชนีตัวที่ 1 ในการค้นหาคำศัพท์ที่มีอักขระตัวที่ 1 เป็น C_1

ขั้นตอนที่ 3 นำอักขระตัวที่ 2 คือ C_2 มาหาดัชนีคำศัพท์กลุ่มที่ 2 โดยที่

$$I_{j,2}(k) = \text{แอสกี้นัมเบอร์ของ } C_2 \\ = A(C_2)$$

ดังนั้นเราจะได้ดัชนีตัวที่ 1 ในการค้นหาคำศัพท์ที่มีอักขระตัวที่ 1 เป็น C_1 และดัชนีตัวที่ 2 ในการค้นหาคำศัพท์ที่มีอักขระตัวที่ 2 เป็น C_2

ขั้นตอนที่ 4 นำค่า n ซึ่งเป็นจำนวนอักขระของข้อความอินพุตมาเปรียบเทียบกับกลุ่มคำที่มีความยาวมากที่สุดในกลุ่มคำที่มีอักขระตัวที่ 1 เป็น C_1 และอักขระตัวที่ 2 เป็น C_2 นั่นคือ เริ่มที่ $i = 1, 2, 3, \dots, p$ เมื่อกลุ่มคำที่ p เป็นกลุ่มคำที่มีจำนวนอักขระที่สั้นที่สุดในกลุ่ม

$$I_{j,k,3}(i) = M_i$$

เมื่อ M_i เป็นความยาวของอักขระในกลุ่มคำศัพท์กลุ่มที่ i ที่มีอักขระตัวที่ 1 เป็น C_1 และอักขระตัวที่ 2 เป็น C_2

ถ้า $n > M_i$ เป็นจริงให้ไปทำงานต่อที่ขั้นตอนที่ 5

ถ้า $n > M_i$ เป็นเท็จให้ไปทำงานต่อที่ขั้นตอนที่ 6

ขั้นตอนที่ 5 ตัดอักขระ $C_{M_i+1}C_{M_i+2}C_{M_i+3}\dots C_n$ ออกจาก $C_1C_2C_3\dots C_n$ ดังนั้นจะเหลืออักขระเพียง $C_1C_2C_3\dots C_{n-(n-M_i)}$ และลดค่า n ออกไป $n-(n-M_i)$ เป็น $n = M_i$ จากนั้นไปทำงานต่อที่ขั้นตอนที่ 6

ขั้นตอนที่ 6 ถ้า $n < M_i$ เป็นจริงให้ไปทำงานต่อที่ขั้นตอนที่ 7

ถ้า $n < M_i$ เป็นเท็จให้ไปทำงานต่อที่ขั้นตอนที่ 8

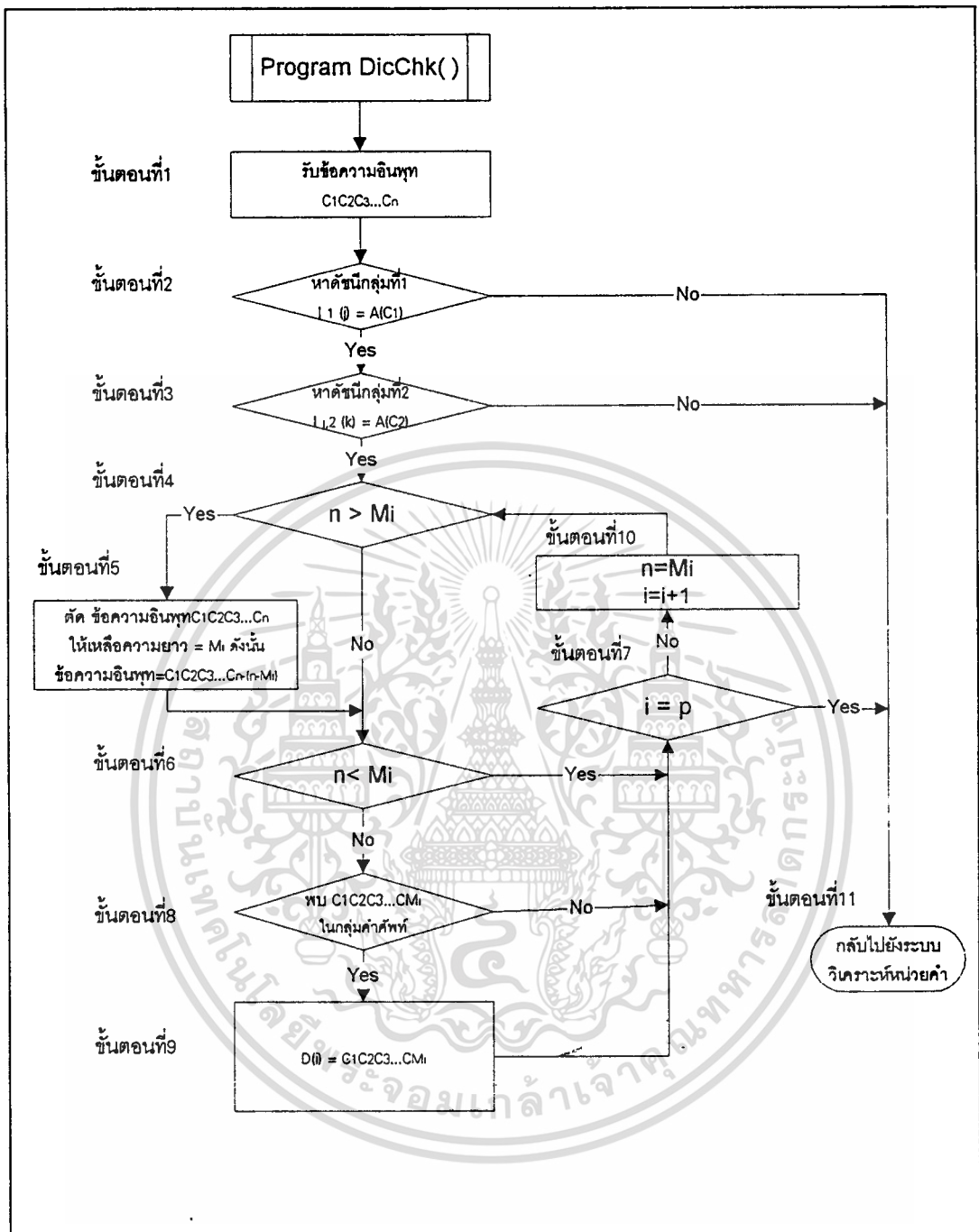
ขั้นตอนที่ 7 ถ้าเป็นกลุ่มคำศัพท์กลุ่มสุดท้าย($i=p$)ให้ไปทำงานต่อที่ขั้นตอนที่ 11
ถ้าไม่เป็นกลุ่มคำศัพท์กลุ่มสุดท้าย($i \neq p$)ให้ไปทำงานต่อที่ขั้นตอนที่ 10

ขั้นตอนที่ 8 เปรียบเทียบตัวอักษร $C_1C_2C_3...C_{n-(n-Mi)}$ กับคำศัพท์ในกลุ่มคำศัพท์ $I_{j,k,3}(i)$ ตั้งแต่ตัวแรกในกลุ่มคำศัพท์จนถึงตัวสุดท้ายในกลุ่มคำศัพท์ หากพบความเท่ากันของทุกอักษรในตำแหน่งเดียวกัน นั่นคือพบคำศัพท์ $C_1C_2C_3...C_{n-(n-Mi)}$ ในฐานข้อมูลพจนานุกรมคำศัพท์ให้ไปทำงานต่อที่ขั้นตอนที่ 9

ขั้นตอนที่ 9 เก็บคำคำศัพท์ $C_1C_2C_3...C_{n-(n-Mi)}$ ไว้ในเซตผลลัพธ์ $D(i)$ จากนั้นตัดอักษร $C_{n-(n-Mi)}$ ออก ดังนั้นจะเหลืออักษรเพียง $C_1C_2C_3...C_{n-(n-Mi-1)}$ จากนั้นไปทำงานต่อที่ขั้นตอนที่ 7

ขั้นตอนที่ 10 เพิ่มค่าดัชนีตัวที่ 3 คือ i ด้วย 1 เพื่อที่จะไปพิจารณาในกลุ่มคำศัพท์กลุ่มต่อไป

ขั้นตอนที่ 11 กลับสู่ระบบวิเคราะห์หน่วยคำ ถ้ามีเซตผลลัพธ์ $D(i)$ ให้ส่งเซตคำตอบกลับไประบบ



รูปที่ 3.19 แสดงผังงานของโปรแกรมการตรวจสอบข้อมูลกับฐานข้อมูลพจนานุกรมคำศัพท์

3.4.2 การใช้งานฐานข้อมูลพจนานุกรมร่วมกับระบบวิเคราะห์โครงสร้างประโยค

ในส่วนของการติดต่อระหว่างระบบวิเคราะห์โครงสร้างประโยคกับฐานข้อมูลพจนานุกรมนั้น
 เอกจะเป็นส่วนต่อเนื่องมาจากการค้นหาคำศัพท์ในส่วนของการวิเคราะห์หน่วยค่าที่ได้กล่าวมาแล้วในหัว
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อที่ 3.4.1 โดยเมื่อกระบวนการตรวจหาคำศัพท์ผิดพบคำที่ต้องการแล้ว ข้อมูลในส่วนของ การวิเคราะห์โครงสร้างประโยคของหน่วยคำนั้นจะถูกนำไปเก็บไว้ที่ตัวแปร ซึ่งข้อมูลของคำจะประกอบไปด้วย CAT, SCAT, AKO, VP, TMAP, CMP2 และ CMP3 ในกรณีที่คำศัพท์ผิดสามารถทำหน้าที่แตกต่างกันได้หลายหน้าที่ เช่นคำว่า เขา เป็นได้ทั้ง N และ PRON ฐานข้อมูลพจนานุกรมก็จะส่งข้อมูลทุกรายการไปยังระบบวิเคราะห์โครงสร้างประโยค เพื่อนำไปพิจารณากับหน่วยคำอื่นเพื่อหาหน้าที่ที่แท้จริง ดังแสดงในรูปที่ 3.20

หน่วยคำที่ค้นได้	ข้อมูลลำดับที่ 1 ฟิลด์ที่ 1	ข้อมูลลำดับที่ 1 ฟิลด์ที่ 2	...	ข้อมูลลำดับที่ 1 ฟิลด์ที่ n
	ข้อมูลลำดับที่ 2 ฟิลด์ที่ 1	ข้อมูลลำดับที่ 2 ฟิลด์ที่ 2	...	ข้อมูลลำดับที่ 2 ฟิลด์ที่ n

รูปที่ 3.20 แสดงข้อมูลที่ถูกส่งไปยังระบบวิเคราะห์โครงสร้างประโยคภาษาไทย

บทที่ 4

กระบวนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

4.1 บทนำ

โครงสร้างประโยคภาษาไทยมีลักษณะแตกต่างจากภาษาอื่นอย่างชัดเจนอยู่สองประการคือ

ประการที่หนึ่ง ประโยคภาษาไทยจะประกอบไปด้วยหน่วยคำหลายหน่วยคำนำมาเขียนเรียงกันตามกฎไวยากรณ์ ซึ่งการนำหน่วยคำมาเขียนเรียงติดกันนี้ ไม่มีการเว้นช่องว่างระหว่างคำ ดังจะเห็นได้จากการเขียนประโยค “โคลง เรือ” จะมีรูปเดียวกับ “โคลง เรือ” นั่นคือในการวิเคราะห์ข้อความ “โคลงเรือ” จะได้คำตอบสองแบบ ซึ่งในคำตอบทั้งหมดนี้จะมีคำตอบที่ตรงกับความเป็นจริงเพียงคำตอบเดียวเท่านั้น

ประการที่สอง การเขียนข้อความในภาษาไทยซึ่งประกอบด้วยประโยคหลายประโยคมาเรียงกัน ก็จะเป็นลักษณะการเรียงประโยคติดกันไป โดยระหว่างประโยคไม่มีเครื่องหมายหรือสัญลักษณ์ใดบ่งบอกการจบประโยค ดังนั้นในการวิเคราะห์ข้อความภาษาไทย ผู้วิเคราะห์จะต้องเป็นผู้แบ่งประโยคด้วยตัวเอง

งานวิจัยนี้จึงได้แบ่งขั้นตอนในการวิเคราะห์ออกเป็นสามส่วนใหญ่ได้ดังนี้

ส่วนที่หนึ่ง การวิเคราะห์หน่วยคำ

ส่วนที่สอง การวิเคราะห์โครงสร้างประโยค

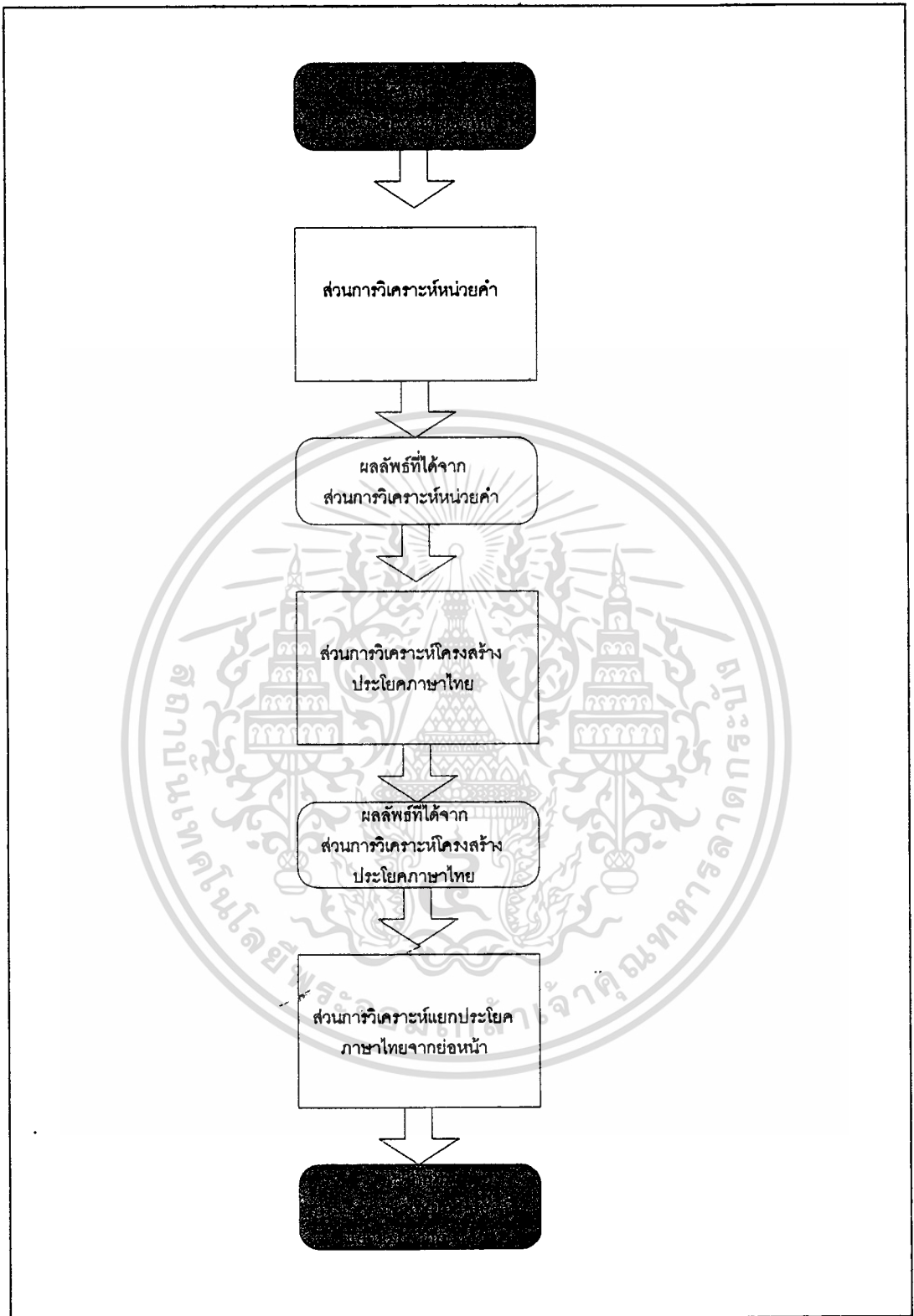
ส่วนที่สาม การวิเคราะห์แยกประโยคจากย่อหน้า

การทำงานทั้งสามส่วนจะทำงานอย่างต่อเนื่องสัมพันธ์กันตั้งแต่ต้นจนจบ ดังแสดงในรูปที่ 4.1 ซึ่งรายละเอียดจะได้กล่าวถึงในหัวข้อต่อไป

4.2 ส่วนการวิเคราะห์หน่วยคำ

การทำงานในส่วนนี้จะทำการวิเคราะห์ข้อความภาษาไทยทั้งหมด เพื่อทำการแบ่งข้อความออกเป็นหน่วยคำ ก่อนที่จะทำการวิเคราะห์ในส่วนต่อไป กระบวนการนี้เราเรียกว่า การแบ่งคำ (Word Segmentation) ซึ่งการแบ่งคำนี้ได้มีผู้คิดขึ้นมาหลายวิธีเช่น Longest Word Matching [8]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับงานวิจัยที่มีการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 แสดงการทำงานต่อเนื่องกันของระบบวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข) เครื่องหมายกำหนดการสิ้นสุดของบรรทัด 1 เครื่องหมาย คือ

ตัวอักษรแสดงการสิ้นสุดของบรรทัด (Carriage Return) ซึ่งมีรหัสแอสกี 13 และ 10 เรียงต่อกัน

4.2.2 กระบวนการแบ่งคำด้วยวิธีการ “เทียบคำที่เป็นไปได้”

วัตถุประสงค์ที่ต้องใช้วิธีการเทียบคำที่เป็นไปได้ เกิดขึ้นเนื่องมาจากการวิเคราะห์แบ่งคำจากประโยคภาษาไทย จะมีโอกาสที่จะแยกคำแล้วให้ผลลัพธ์ได้มากกว่า 1 แบบดังที่ได้กล่าวมาแล้วในหัวข้อที่ 4.1 ซึ่งแต่ละผลลัพธ์ของเซตคำตอบมีโอกาสที่จะเป็นผลลัพธ์ที่ถูกต้องได้ ดังแสดงในรูปที่ 4.2



รูปที่ 4.2 แสดงผลลัพธ์ที่เป็นไปได้ของการวิเคราะห์แยกหน่วยคำ

ซึ่งผลลัพธ์ทั้ง 4 แบบมีโอกาสที่จะเป็นการแบ่งหน่วยคำที่ถูกต้องได้ทั้งนั้น แต่การที่จะตัดสินว่าผลลัพธ์แบบใดเป็นผลลัพธ์ที่ถูกต้องตามกฎไวยากรณ์ จะต้องมีข้อมูลส่วนเพิ่มเติมเพื่อใช้ในการพิจารณา นั่นก็คือการพิจารณานำหน่วยคำโดยดูจากหน้าที่ของหน่วยคำนั้นพิจารณาเข้ากับโครงสร้างเอกสารเป็นเอกสารที่ส่งวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นับญาติให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทางไวยากรณ์ของประโยคภาษาไทย ซึ่งวิธีการที่จะพิจารณาก็คือ กฎไวยากรณ์ดังที่กล่าวไว้ในบทที่ 2 เป็นเกณฑ์ในการพิจารณาซึ่งจะกล่าวโดยละเอียดในหัวข้อต่อไป

กระบวนการวิเคราะห์หน่วยคำมีขั้นตอนหลักอยู่สองขั้นตอนดังแสดงในรูปที่ 4.3 คือ

ขั้นตอนหลักขั้นที่หนึ่ง การจัดเตรียมข้อมูล

ขั้นตอนหลักขั้นที่สอง การแบ่งคำ

โดยมีรายละเอียดดังต่อไปนี้

4.2.2.1 การจัดเตรียมข้อมูล

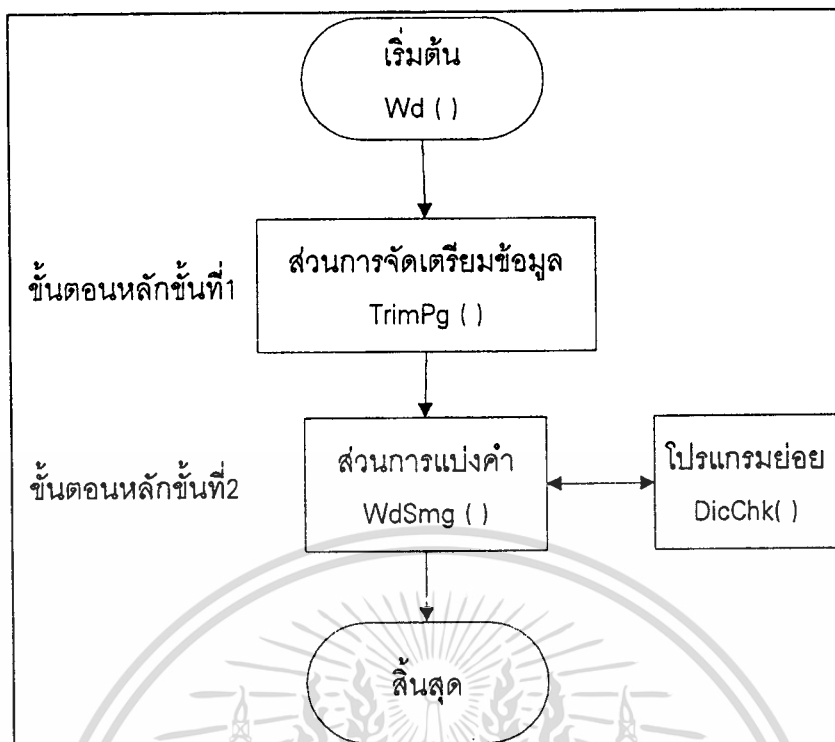
การจัดเตรียมข้อมูลถือเป็นขั้นตอนแรกสุดของการวิเคราะห์หน่วยคำ ซึ่งจะเป็นการรับข้อมูลย่อหน้าต้นแบบ และตรวจสอบความถูกต้องของการป้อนข้อมูลโดยมีขั้นตอนดังต่อไปนี้

อัลกอริทึมของส่วนการจัดเตรียมข้อมูล

ขั้นตอนที่ 1 ทำการรับข้อความย่อหน้าแบบเข้ามาสู่ระบบ

ขั้นตอนที่ 2 ทำการตรวจสอบหาข้อความที่ละตัว ตั้งแต่ตัวที่ 1 จนถึงตัวสุดท้ายของข้อความถ้าตัวอักษรใดเป็นอักษรเว้นวรรค (มีรหัสแอสกี 32) ให้ตัดออกจากข้อความ และถ้าตัวอักษรใดเป็นอักขระที่เป็นเครื่องหมายกำหนดการสิ้นสุดของบรรทัด (มีรหัสแอสกี 13 และแอสกี 10) ให้ตัดออกจากข้อความ

ขั้นตอนที่ 3 ส่งผลลัพธ์จากขั้นตอนที่ 2 ไปยังส่วนการแบ่งคำ



รูปที่ 4.3 กระบวนการวิเคราะห์หน่วยคำ

ตัวอย่างในการจัดเตรียมข้อมูล

ในการวิเคราะห์ข้อความย่อหน้าต้นแบบ

เรือโคลง เนื่องจาก

โคลงเรือ

ขั้นตอนที่ 1 รับข้อความอินพุตเข้าสู่ระบบจะได้

“เรือโคลง เนื่องจาก ◀ โคลงเรือ ▶”

ขั้นตอนที่ 2 ตัดอักขระเว้นวรรคและตัวอักขระกำหนดการสิ้นสุดของบรรทัดออกจะได้

“เรือโคลงเนื่องจากโคลงเรือ”

ขั้นตอนที่ 3 ส่งผลลัพธ์จากขั้นตอนที่ 2 ไปยังส่วนการแบ่งคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2.2 การแบ่งคำ

การแบ่งคำจะเป็นการทำงานที่ต่อเนื่องจากการตรวจสอบข้อความต้นแบบดังแสดงในรูปที่ 4.3

อัลกอริทึมของส่วนการแบ่งคำ

โปรแกรม WdSgm ()

ขั้นตอนที่ 1 รับข้อความอินพุท

ขั้นตอนที่ 2 ส่งข้อความอินพุทไปตรวจสอบกับฐานข้อมูลพจนานุกรมคำศัพท์ภาษาไทยด้วยโปรแกรม DicChk (ข้อความอินพุท)
เมื่อโปรแกรม DicChk () ส่งผลลัพธ์กลับมาเป็นคำศัพท์ที่มีอยู่ในฐานข้อมูลพจนานุกรมคำศัพท์จำนวน n ตัว ให้ไปที่ขั้นตอนที่ 3

ขั้นตอนที่ 3 ถ้า $n = 0$ ไปที่ขั้นตอนที่ 7
ถ้า $n \neq 0$ ให้ส่งคำศัพท์ตัวที่ 1 ไปที่ขั้นตอนที่ 4

ขั้นตอนที่ 4 ตัดข้อความอินพุทออกด้วยคำศัพท์ไปเก็บต่อท้ายตัวแปรที่เตรียมไว้ S โดยตัดจากอักขระตัวที่ 1 จนถึงตัวสุดท้ายของคำศัพท์ แล้วส่งข้อความอินพุทที่เหลือไปที่ขั้นตอนที่ 5

ขั้นตอนที่ 5 ถ้าอินพุทจากขั้นตอนที่ 4 ไม่มีตัวอักขระเหลือ ให้ไปที่ขั้นตอนที่ 6
ถ้าอินพุทจากขั้นตอนที่ 4 มีตัวอักขระเหลืออยู่ ให้ไปที่โปรแกรม WdSgm (ข้อความที่ได้จากขั้นตอนที่ 4) เมื่อโปรแกรมทำงานเสร็จ ให้ไปขั้นตอนที่ 7

ขั้นตอนที่ 6 นำตัวแปรที่เก็บคำศัพท์ไปเก็บไว้ที่เซตคำตอบจากการแบ่งคำ เซตที่ j ($j = 1, 2, 3, \dots$) แล้วทำการเคลียร์ข้อมูลในตัวแปร

ขั้นตอนที่ 7 กลับไปทำขั้นตอนที่ 5 ด้วยคำศัพท์ตัวต่อไปจนกว่าจะครบตัวที่ n

ตัวอย่าง ในการแบ่งคำด้วยโปรแกรม WdSgm (เรือโคลงเนื่องจากโคลงเรือ)

ขั้นตอนที่ 1 รับข้อความอินพุท

เรือโคลงเนื่องจากโคลงเรือ

ขั้นตอนที่ 2 ส่งข้อความไปที่โปรแกรม DicChk ()

DicChk (เรือโคลงเนื่องจากโคลงเรือ)

ได้ผลลัพธ์เป็นจำนวน 1 ตัว ($n = 1$)

เรือ

ขั้นตอนที่ 3 จากขั้นตอนที่ 2 $n \neq 0$ ดังนั้นส่งคำศัพท์ตัวที่ 1 คือ “เรือ” ไปขั้นตอนที่ 4

ขั้นตอนที่ 4 ตัดคำว่า “เรือ” ออกจาก “เรือโคลงเนื่องจากโคลงเรือ” จะได้ $s =$ “เรือ” เมื่อ s เป็น
ตัวแปรที่เตรียมไว้ แล้วส่ง “โคลงเนื่องจากโคลงเรือ” ไปที่ขั้นตอนที่ 5

ขั้นตอนที่ 5 เนื่องจากมีอักขระเหลืออยู่ดังนั้นส่งข้อความไปที่โปรแกรม WdSgm ()

WdSgm₂ (โคลงเนื่องจากโคลงเรือ)

เลข 2 แสดงถึงการทำงานในโปรแกรม WdSgm () ที่ถูกเรียกครั้งที่ 2

WdSgm₂ ขั้นตอนที่ 1 รับข้อความอินพุท

โคลงเนื่องจากโคลงเรือ

WdSgm₂ ขั้นตอนที่ 2 ส่งข้อความไปที่โปรแกรม DicChk ()

DicChk (โคลงเนื่องจากโคลงเรือ)

ได้ผลลัพธ์เป็นจำนวน 2 ตัว ($n = 2$)

โคลง

โค

WdSgm₂ ขั้นตอนที่ 3 จากขั้นตอนที่ 2 $n \neq 0$ ดังนั้นส่งคำศัพท์ตัวที่ 1 คือ “โคลง” ไปขั้นตอนที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

WdSgm₂ ขั้นตอนที่ 4 ตัดคำว่า “โคลง” ออกจาก “เนื่องจากโคลงเรือ” จะได้ $s_2 = \text{“โคลง”}$ เมื่อ s_2 เป็นตัวแปรที่ WdSgm₂ เตรียมไว้ แล้วส่ง “เนื่องจากโคลงเรือ” ไปที่ขั้นตอนที่ 5

WdSgm₂ ขั้นตอนที่ 5 เนื่องจากมีอักขระเหลืออยู่ดังนั้นส่งข้อความไปที่โปรแกรม WdSgm ()
 WdSgm₃ (เนื่องจากโคลงเรือ)
 เลข 3 แสดงถึงการทำงานในโปรแกรม WdSgm () ที่ถูกเรียกครั้งที่ 3

WdSgm₃ ขั้นตอนที่ 1 รับข้อความอินพุท
 เนื่องจากโคลงเรือ

WdSgm₃ ขั้นตอนที่ 2 ส่งข้อความไปที่โปรแกรม DicChk ()
 DicChk (เนื่องจากโคลงเรือ)
 ได้ผลลัพธ์เป็นจำนวน 1 ตัว ($n = 1$)
 เนื่องจาก

WdSgm₃ ขั้นตอนที่ 3 จากขั้นตอนที่ 2 $n \neq 0$ ดังนั้นส่งค่าศัพท์ตัวที่ 1 คือ “เนื่องจาก” ไปขั้นตอนที่ 4

WdSgm₃ ขั้นตอนที่ 4 ตัดคำว่า “เนื่องจาก” ออกจาก “เนื่องจากโคลงเรือ” จะได้ $s_3 = \text{“เนื่องจาก”}$ เมื่อ s_3 เป็นตัวแปรที่ WdSgm₃ เตรียมไว้ แล้วส่ง “โคลงเรือ” ไปที่ขั้นตอนที่ 5

WdSgm₃ ขั้นตอนที่ 5 เนื่องจากมีอักขระเหลืออยู่ดังนั้นส่งข้อความไปที่โปรแกรม WdSgm ()
 WdSgm₄ (โคลงเรือ)
 เลข 4 แสดงถึงการทำงานในโปรแกรม WdSgm () ที่ถูกเรียกครั้งที่ 4

WdSgm₄ ขั้นตอนที่ 1 รับข้อความอินพุท
 โคลงเรือ

WdSgm₄ ขั้นตอนที่ 2 ส่งข้อความไปที่โปรแกรม DicChk ()
 DicChk (โคลงเรือ)

โคลง

โค

WdSgm₄ ขั้นตอนที่ 3 จากขั้นตอนที่ 2 $n \neq 0$ ดังนั้นส่งคำศัพท์ตัวที่ 1 คือ “โคลง” ไปขั้นตอนที่ 4

WdSgm₄ ขั้นตอนที่ 4 ตัดคำว่า “โคลง” ออกจาก “โคลงเรือ” จะได้ $s_4 = \text{“โคลง”}$ เมื่อ s_4 เป็นตัวแปรที่ WdSgm₄ เตรียมไว้ แล้วส่ง “เรือ” ไปที่ขั้นตอนที่ 5

WdSgm₄ ขั้นตอนที่ 5 เนื่องจากมีอักขระเหลืออยู่ดังนั้นส่งข้อความไปที่โปรแกรม WdSgm ()

WdSgm₅ (โคลงเรือ)

เลข 5 แสดงถึงการทำงานในโปรแกรม WdSgm () ที่ถูกเรียกครั้งที่ 5

WdSgm₅ ขั้นตอนที่ 1 รับข้อความอินพุต

เรือ

WdSgm₅ ขั้นตอนที่ 2 ส่งข้อความไปที่โปรแกรม DicChk ()

DicChk (เรือ)

ได้ผลลัพธ์เป็นจำนวน 1 ตัว ($n = 1$)

เรือ

WdSgm₅ ขั้นตอนที่ 3 จากขั้นตอนที่ 2 $n \neq 0$ ดังนั้นส่งคำศัพท์ตัวที่ 1 คือ “เรือ” ไปขั้นตอนที่ 4

WdSgm₅ ขั้นตอนที่ 4 ตัดคำว่า “เรือ” ออกจาก “เรือ” จะได้ $s_5 = \text{“เรือ”}$ เมื่อ s_5 เป็นตัวแปรที่

WdSgm₅ เตรียมไว้ แล้วส่ง “” ไปที่ขั้นตอนที่ 5

WdSgm₅ ขั้นตอนที่ 5 เนื่องจากไม่มีอักขระเหลืออยู่ดังนั้น ไปที่ขั้นตอนที่ 6

WdSgm₅ ขั้นตอนที่ 6 นำตัวแปรที่เก็บคำศัพท์ไปเก็บไว้ที่เซตคำตอบจากการแบ่งคำเซตที่ 1

นั่นคือ ค

เรือ โคลง เนื่องจาก โคลง เรือ

จากนั้นทำการเคลียร์ S_6

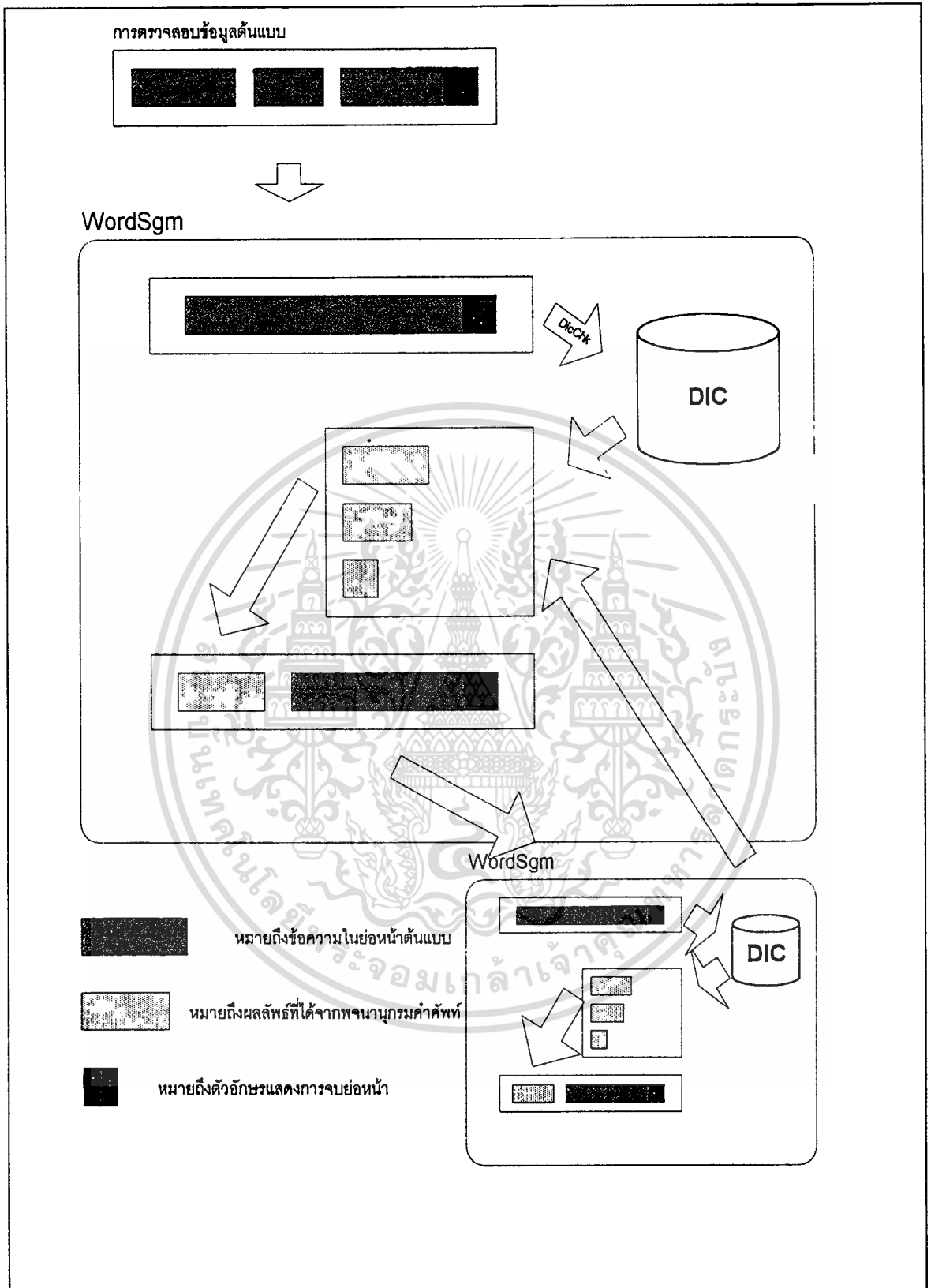
$WdSgm_6$ ขั้นตอนที่ 7 เนื่องจากมีค่าศัพท์เดียว $n=1$ จึงไม่ต้องกลับไปทำตัวต่อไป ดังนั้นไปที่ขั้นตอนที่ 8

$WdSgm_6$ ขั้นตอนที่ 8 จบการทำงานดังนั้นจึงกลับไป $WdSgm_4$ ขั้นตอนที่ 5

เมื่อทำงานมาถึงขั้นตอนนี้จะได้ชุดคำตอบที่ 1 และ เมื่อกระทำตามขั้นตอนต่อไปจนครบทุกขั้นตอน จะได้ผลลัพธ์ดังตารางที่ 4.1

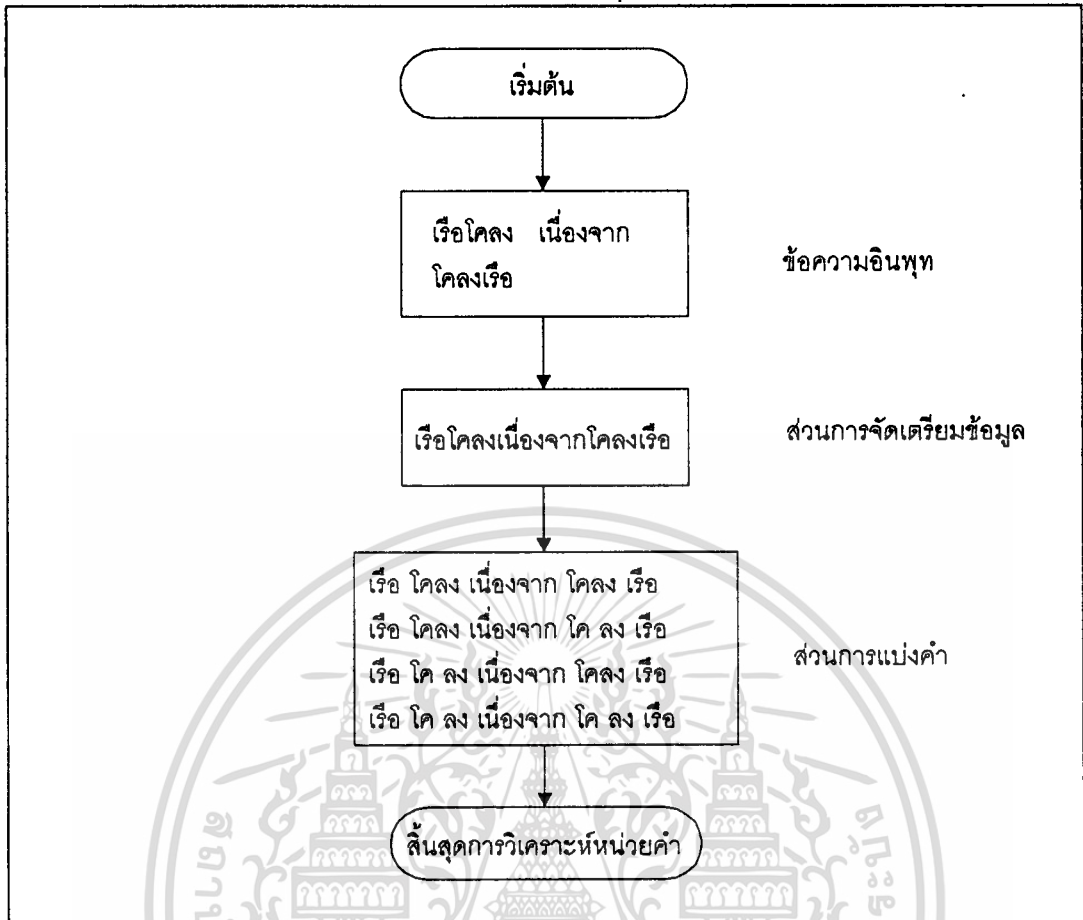
ชุดคำตอบที่ หน่วยคำ	1	2	3	4	5	6	7
1	เร็ว	โคลง	เนื่องจาก	โคลง	เร็ว		
2	เร็ว	โคลง	เนื่องจาก	โค	ลง	เร็ว	
3	เร็ว	โค	ลง	เนื่องจาก	โคลง	เร็ว	
4	เร็ว	โค	ลง	เนื่องจาก	โค	ลง	เร็ว

ตารางที่ 4.1 แสดงผลลัพธ์ที่ได้จากตัวอย่างการวิเคราะห์หน่วยคำ



รูปที่ 4.4 แสดงขั้นตอนของระบบวิเคราะห์แบ่งคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



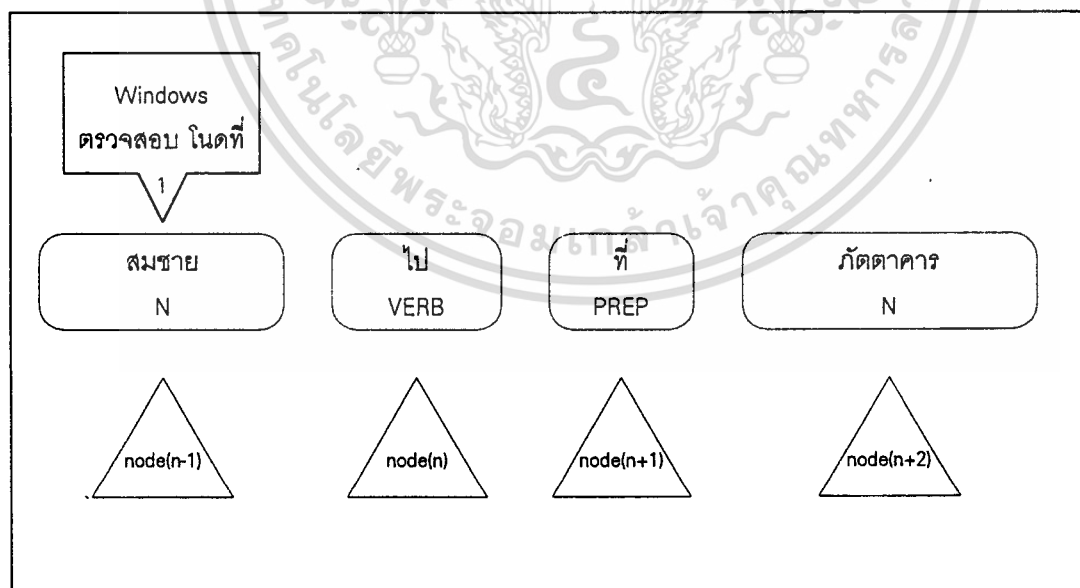
รูปที่ 4.5 แสดงตัวอย่างข้อมูลในการวิเคราะห์หน่วยค่า

4.3 ส่วนการวิเคราะห์โครงสร้างของประโยคภาษาไทย (Sentence Analysis System)

การวิเคราะห์โครงสร้างของประโยคภาษาไทยจะทำกับข้อมูลที่ผ่านมาจากการวิเคราะห์หน่วยคำมาแล้ว โดยใช้กฎไวยากรณ์ที่กล่าวไว้ในบทที่ 2 เป็นหลักในการพิจารณา ในส่วนของการวิเคราะห์โครงสร้างประโยคภาษาไทยจะกระทำทั้งการวิเคราะห์ด้านวากยสัมพันธ์และอรรถศาสตร์โดยใช้ กฎไวยากรณ์พึงพา กฎไวยากรณ์โครงสร้างวลี และกฎไวยากรณ์การกตั้งต่อไปนี้

4.3.1 การใช้ไวยากรณ์พึงพาหาหน่วยหลักของกลุ่มประโยคภาษาไทย

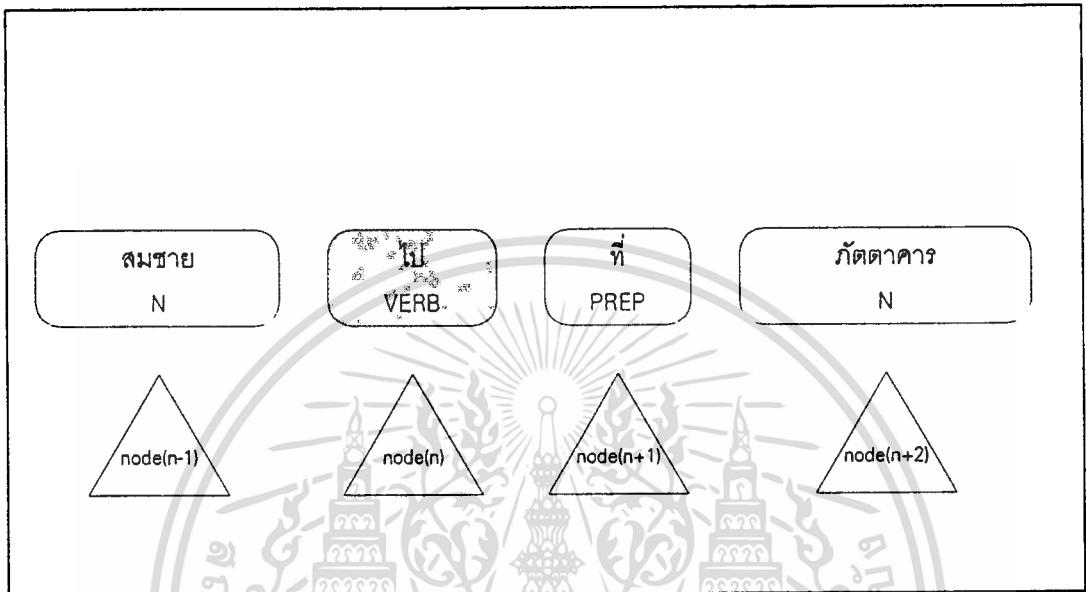
การหาหน่วยหลักของประโยคภาษาไทยจะเป็นจุดสำคัญ ที่จะใช้ในการเริ่มพิจารณาประโยคภาษาไทย เนื่องจากหน่วยคำหลักจะเป็นตัวบอกเหตุการณ์ที่เกิดขึ้นในประโยคภาษาไทย



รูปที่ 4.6 แสดงการเริ่มต้นค้นหาหน่วยคำหลักของประโยคภาษาไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

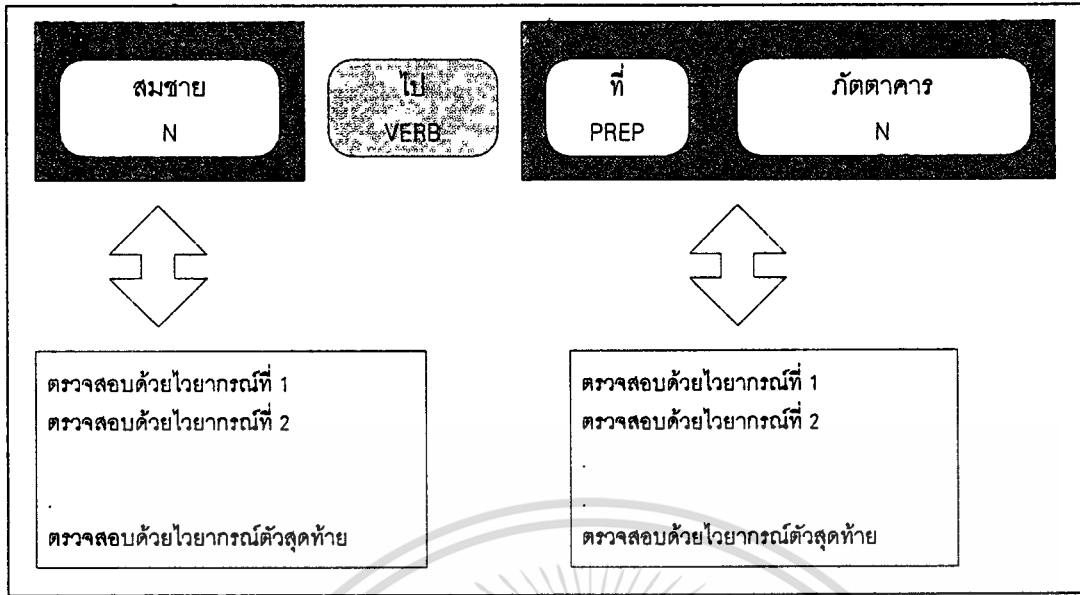
ดังนั้นในการพิจารณาหาโครงสร้างของประโยคภาษาไทยจะต้องเริ่มจากหน่วยคำหลักของประโยคภาษาไทยก่อนคือคำกริยานั้นเอง ในงานวิจัยนี้จะใช้การค้นหาคำกริยาซึ่งจะเป็นหน่วยคำหลักของประโยคภาษาไทยก่อนที่จะการวิเคราะห์ในส่วนต่อไป



รูปที่ 4.7 แสดงการหาหน่วยคำหลักของประโยคภาษาไทย

4.3.2 การใช้ไวยากรณ์โครงสร้างวลีหองค์ประกอบของโครงสร้างวลี

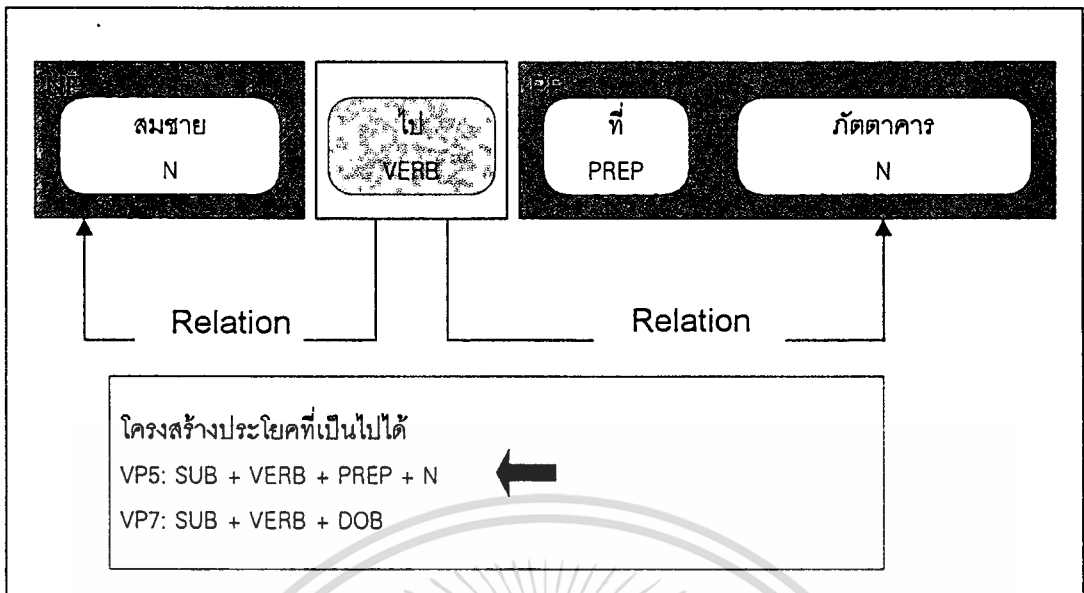
เมื่อหาหน่วยคำหลักได้แล้วก็ใช้ไวยากรณ์โครงสร้างวลีตรวจสอบด้านซ้ายมือของหน่วยคำหลักคือ โหนดที่ $n - 1$ เมื่อ n เป็นลำดับที่ที่หน่วยคำหลักอยู่ และใช้ไวยากรณ์โครงสร้างวลีตรวจสอบด้านขวามือของหน่วยคำหลักคือ โหนดที่ $n + 1$ ไปจนถึง โหนดที่ $n + 3$



รูปที่ 4.8 แสดงการใช้ไวยากรณ์โครงสร้างวลีตรวจสอบโครงสร้างของประโยคภาษาไทยในย่อหน้า

4.3.3 การใช้ไวยากรณ์การหาความสัมพันธ์ของหน่วยคำในประโยคภาษาไทย

หลังจากตรวจสอบหน่วยคำของประโยคภาษาไทยด้วยไวยากรณ์โครงสร้างวลีแล้วเราจะตรวจสอบหาความสัมพันธ์ระหว่างหน่วยคำหลักกับหน่วยคำพืงพาโดยใช้รูปแบบของคำกริยาเป็นตัวพิจารณาเปรียบเทียบหน่วยคำดังแสดงในรูปที่ 4.9



รูปที่ 4.9 แสดงการตรวจสอบความสัมพันธ์ระหว่างหน่วยคำหลักกับหน่วยคำข้างเคียง

4.3.4 กระบวนการวิเคราะห์โครงสร้างประโยคภาษาไทย

4.3.4.1 อัลกอริธึมของการตรวจสอบ

กำหนดให้

- $B_{i,k,l}$ หมายถึง หน้าหลักแบบที่ l ของหน่วยคำที่ k ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j
- j คือลำดับของเซตผลลัพธ์ที่ได้จากการแบ่งคำ
- k คือลำดับของหน่วยคำในเซตผลลัพธ์ที่ได้จากการแบ่งคำ
- l คือลำดับของหน้าหลักที่ค่านั้นมีอยู่
- S_j คือจำนวนหน่วยคำทั้งหมดในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j
- $q_{i,k}$ คือจำนวน VP ที่เป็นไปได้ทั้งหมดของหน่วยคำที่ k ในเซตผลลัพธ์ที่ j
- $r_{i,k}$ คือจำนวนหน้าหลักที่เป็นไปได้ทั้งหมดของหน่วยคำที่ k ในเซตผลลัพธ์ที่ j
- t คือจำนวนเซตผลลัพธ์ที่ได้จากการแบ่งคำ
- $Vp_{i,k}(m)$ คือรูปกริยาแบบที่ m ของหน่วยคำ k ในเซตผลลัพธ์ j

การหาหน่วยคำหลัก

โปรแกรม StnSgm ()

- ขั้นตอนที่ 1 เริ่มทำการวิเคราะห์หน่วยคำในเซตผลลัพธ์ที่ได้จากการแบ่งคำ เช็ทที่ 1 ($j = 1$) แล้วไปทำขั้นตอนที่ 2
- ขั้นตอนที่ 2 เริ่มทำการวิเคราะห์หน่วยคำที่ 1 ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเช็ทที่ j ($k = 1$), ($C=1$) แล้วไปทำขั้นตอนที่ 3
- ขั้นตอนที่ 3 เริ่มทำการวิเคราะห์หน้าที่หลักแบบที่ 1 ของหน่วยคำที่ k ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j ($l = 1$) แล้วไปทำงานขั้นตอนที่ 4
- ขั้นตอนที่ 4 ตรวจสอบหน้าที่หลักแบบที่ l ของหน่วยคำที่ k ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j ว่าทำหน้าที่เป็นคำกริยาได้หรือไม่
- ถ้า $B_{l,k} = \text{VERB}$ ให้ไปทำขั้นตอนที่ 5
- ถ้า $B_{l,k} \neq \text{VERB}$ ให้ไปทำขั้นตอนที่ 11
- ขั้นตอนที่ 5 เริ่มทำการวิเคราะห์รูปกริยา (Verb Pattern) ของหน่วยคำที่ k ในเช็ทคำตอบที่ได้จากการแบ่งคำเช็ทที่ j แบบที่ 1 ($m = 1$) แล้วไปทำขั้นตอนที่ 6 $VP_{j,k}(m)$
- ขั้นตอนที่ 6 ทำการตรวจสอบด้วยไวยากรณ์ของหน่วยคำที่ k ในเช็ทคำตอบที่ได้จากการแบ่งคำเช็ทที่ j โดยส่วนการวิเคราะห์หน่วยคำด้วยกฎไวยากรณ์ โดยใช้คำสั่ง $Pchk_{(j,k,m)}$ เมื่อได้รับผลลัพธ์แล้วไปทำขั้นตอนที่ 7
- ขั้นตอนที่ 7 ถ้าผลลัพธ์ของขั้นตอนที่ 6 พบว่าหน่วยคำที่ k ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j เป็นหน่วยคำหลักได้ ให้ไปทำขั้นตอนที่ 8
- ถ้าผลลัพธ์ของขั้นตอนที่ 6 พบว่าหน่วยคำที่ k ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j ไม่เป็นหน่วยคำหลักไม่ได้ ให้ไปทำขั้นตอนที่ 9
- ขั้นตอนที่ 8 เพิ่มค่า C ด้วย 1 เพื่อกำหนดลำดับประโยคที่จะวิเคราะห์ลำดับถัดไป
- ขั้นตอนที่ 9 ทำการตรวจสอบว่ามีรูปกริยาของหน่วยคำที่ k ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j แบบอื่นหรือไม่
- ถ้าไม่มีแบบอื่น ($m = q$) ให้ไปทำขั้นตอนที่ 11
- ถ้ายังมีแบบอื่น ($m \neq q$) ให้ไปทำขั้นตอนที่ 10
- ขั้นตอนที่ 10 เพิ่มค่าของ m ด้วย 1 แล้วไปทำขั้นตอนที่ 6
- ขั้นตอนที่ 11 ทำการตรวจว่ามีหน้าที่หลักแบบอื่นอีกหรือไม่

ถ้ามีหน้าที่หลักแบบอื่น $l \neq r_{j,k}$ ให้ไปทำขั้นตอนที่ 12

ขั้นตอนที่ 12 เพิ่มค่าของ l ด้วย 1 แล้วไปทำขั้นตอนที่ 4

ขั้นตอนที่ 13 ทำการตรวจว่าหน่วยค่านี้นี้เป็นหน่วยค่าสุดท้ายในเช็กราคาตอบที่ได้จากการแบ่งค่าเช็กราคาที่ j

ถ้า $k = s_j$ ให้ไปทำขั้นตอนที่ 15

ถ้า $k \neq s_j$ ให้ไปทำขั้นตอนที่ 14

ขั้นตอนที่ 14 เพิ่มค่าของ k ด้วย 1 แล้วไปทำขั้นตอนที่ 3

ขั้นตอนที่ 15 ทำการตรวจว่าเช็กราคาตอบนี้เป็นเช็กราคาตอบที่ได้จากการแบ่งค่าเช็กราคาสุดท้ายหรือไม่

ถ้า $j = t$ เป็นจริงให้ไปทำขั้นตอนที่ 17

ถ้า $j \neq t$ เป็นเท็จให้ไปทำขั้นตอนที่ 16

ขั้นตอนที่ 16 เพิ่มค่าของ j ด้วย 1 แล้วไปทำขั้นตอนที่ 2

ขั้นตอนที่ 17 จบการทำงาน

ส่วนการวิเคราะห์หน่วยค่าด้วยกฎไวยากรณ์

$Pch_k()$

ขั้นตอนที่ 1 รับข้อมูลมาจากขั้นตอนที่ 6 ของส่วนการแบ่งหน่วยค่า j, k, l แล้วกำหนดไว้ก่อนว่าหน่วยค่าที่ k ในเซตผลลัพธ์ j เป็นค่าหลักไม่ได้

ขั้นตอนที่ 2 ตรวจสอบตำแหน่งของหน่วยค่า

ถ้า $k > 1$ ไปที่ขั้นตอนที่ 3

ถ้า $k = 1$ ไปที่ขั้นตอนที่ 5

ขั้นตอนที่ 3 ตรวจสอบหน่วยค่าที่ $k-1$ ด้วย กฎไวยากรณ์นามวลี R1 - R4

ถ้าหน่วยค่าที่ $k-1$ เป็นนามวลีได้ ไปที่ขั้นตอนที่ 4

ถ้าหน่วยค่าที่ $k-1$ เป็นนามวลีไม่ได้ ไปที่ขั้นตอนที่ 5

ขั้นตอนที่ 4 ตรวจสอบหน่วยค่าที่ $k, k+1, k+2, k+3$ ด้วยกฎไวยากรณ์กริยาวลีส่วนที่ 1

ถ้าพบว่าตรงกับกฎใดกฎหนึ่ง ให้ทำ ACTION ตามที่กฎไวยากรณ์กำหนดให้หน่วยค่าที่ k ในเซตผลลัพธ์ j เป็นค่าหลักของประโยคที่ c ไปขั้นตอนที่ 6

ขั้นตอนที่ 5 ตรวจสอบหน่วยค่าที่ $k, k+1$ ด้วยกฎไวยากรณ์กริยาวลีส่วนที่ 2

ถ้าพบว่าถูกต้องตามกฎให้ทำ ACTION ตามที่กฎไวยากรณ์กำหนด ให้นำหน่วยค่าที่ k ในเซตผลลัพธ์ j เป็นค่าหลักของประโยคไปขึ้นตอนที่ 6
 ชั้นตอนที่ 6 ส่งผลลัพธ์กลับไปยังส่วนการหาค่าหลักชั้นตอนที่ 6

กฎไวยากรณ์สำหรับวิเคราะห์โครงสร้างประโยค

<< รูปแบบของกฎไวยากรณ์ >>

Rule No. / Rule Name /

{ COND1 [& COND2 & COND3 ...] }

= { ACTION1 [, ACTION2, ACTION3, ...] }

กำหนดให้

- Rule No. : หมายถึง ลำดับของกฎ เช่น R1 คือ กฎไวยากรณ์ที่ 1 เป็นต้น
- Rule Name : หมายถึง ชื่อของกฎที่แสดงให้เห็นว่าเป็นกฎไวยากรณ์ในการวิเคราะห์เกี่ยวกับอะไร
- COND : หมายถึง เงื่อนไขการพิจารณา ถ้าเงื่อนไขเป็นจริงทั้งหมดกับค่าที่กำลังพิจารณา แสดงว่าเกิดความสำเร็จของการวิเคราะห์กับค่าที่พิจารณา เรียกว่า “ SUCCESS “ ระบบจะทำการสร้างผลลัพธ์ให้กับขบวนการวิเคราะห์ด้วยการปฏิบัติตามคำสั่ง ACTION
- ACTION : หมายถึง การประมวลผลที่จะให้ระบบวิเคราะห์ปฏิบัติในกรณีที่มี COND ทั้งหมดเป็นจริง และระบบต้องปฏิบัติตามทุก ACTION ที่อยู่หลังเครื่องหมาย “=”

<< นิยามของฟังก์ชันและสัญลักษณ์ >>

- CatChk (P,Q,Val1) : ตรวจสอบว่าหน่วยค่าที่ Q ในเซตคำตอบที่ P โดยพิจารณาตามเงื่อนไขดังนี้
- 1) มีประเภทของค่าเท่ากับ Val1 ให้ผลลัพธ์เป็นจริง
 - มีประเภทของค่าไม่เท่ากับ Val1 ให้ผลลัพธ์เป็นเท็จ
- AKOChk (P,Q,Val1) : ตรวจสอบว่าหน่วยค่าที่ P ในเซตคำตอบที่ Q โดยพิจารณาตามเงื่อนไขดังนี้
- 1) มีชนิดของค่าเท่ากับ Val1 ให้ผลลัพธ์เป็นจริง

เอกสารนี้เป็นเอกสารที่มีชนิดของค่าไม่เท่ากับ Val1 ให้ผลลัพธ์เป็นเท็จ ก่อนอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- เครื่องหมาย “&” : เป็นเครื่องหมายที่เชื่อมระหว่างเงื่อนไขที่พิจารณาว่า เงื่อนไขที่
มาเชื่อมเป็นจริงทั้งคู่หรือไม่ ถ้ากำหนดให้จริงมีสัญลักษณ์ “T” และเท็จมีสัญลักษณ์
ลักษณะ “F” ดังนั้น ถ้า X&Y สามารถอธิบายตามหลักตรรกศาสตร์ได้ดังนี้
 - ได้ผลลัพธ์เป็นจริงกรณีเดียวคือ ถ้า X และ Y มีค่าเป็นจริงทั้งคู่ นอกนั้น ไม่ว่าจะค่า
ของ X หรือ Y ตัวใดตัวหนึ่งมีค่าเป็นเท็จ จะได้ผลลัพธ์เป็นเท็จ
- NOT () : เป็นเครื่องหมายที่จะกลับค่าทางตรรกศาสตร์ ซึ่งจะได้ผลลัพธ์เป็น จริง
หากค่าที่อยู่ในวงเล็บเป็นเท็จ และถ้าหากค่าที่อยู่ในวงเล็บเป็นเท็จ ก็จะได้ผลลัพธ์
เป็นจริง
- ChkNPSt (P,Q,Val1) : ตรวจสอบว่าประโยคที่ P ในเซตคำตอบที่ Q โดยพิจารณาตาม
เงื่อนไขดังนี้
 - 1) NPSt(P,Q) มีค่าเท่ากับ Val1 ให้ผลลัพธ์เป็นจริง
 - NPSt(P,Q) มีค่าไม่เท่ากับ Val1 ให้ผลลัพธ์เป็นเท็จ

กฎไวยากรณ์ตรวจสอบนามวลี

กฎย่อย R1 การตรวจสอบนามวลีที่ประกอบด้วยคำสรรพนาม

R1/NP \longrightarrow PRON (Human)/

CatChk (j,k-1,PRON) & ChkNPSt(j,C,NULL) & AKOChk (j,k-1,111)
= ASSIGN (REL1,AGT) , ASSIGN (NPSt (j,C) , 1) , ASSIGN(NdUChk(j, k-1) ,
NdUChk(j, k-1) + 1)

กฎย่อย R2 การตรวจสอบนามวลีที่ประกอบด้วยคำสรรพนาม

R2/NP \longrightarrow PRON /

CatChk (j,k-1,PRON) & ChkNPSt(j,C,NULL) & NOT (AKOChk (j,k-1,111))
= ASSIGN (REL1,SUB) , ASSIGN (NPSt (j,C) , 1) , ASSIGN(NdUChk(j, k-1) ,
NdUChk(j, k-1) + 1)

กฎย่อย R3 การตรวจสอบนามวลีที่ประกอบด้วยคำนาม

R3/NP \longrightarrow N(Human)/

CatChk (j,k-1,N) & ChkNPSt(j,C,NULL) & AKOChk (j,k-1,111)
= ASSIGN (REL1,AGT) , ASSIGN (NPSt (j,C) , 1) , ASSIGN(NdUChk(j, k-1) ,
NdUChk(j, k-1) + 1)

กฎย่อย R4 การตรวจสอบนามวลีที่ประกอบด้วยคำนาม เท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

R4/NP → N/

CatChk (j,k-1,N) & ChkNPSt(j,C,NULL) & NOT (AKOChk (j,k-1,111))
 = ASSIGN (REL1,SUB) , ASSIGN (NPSt (j,C) , 1) , ASSIGN(NdUChk(j, k-1) ,
 NdUChk(j, k-1) +1)

กฎไวยากรณ์ตรวจสอบกริยาวลีกลุ่มที่ 1

กฎย่อย R5 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็นVP6 คำนาม คำบุรพบท
 คำนาม

R5 / VP → VERB(VP6) N PREP N/
 VPChk (j,k, 6) & ChkVPSt(j,C,NULL) & CatChk (j, k+1, N)
 & CatChk (j, k+2, PREP) & CatChk (j, k+3, N)
 = ASSIGN (REL2, OBJ) ,
 ASSIGN (VPSt (j,C) , 6) ,
 ASSIGN(NdMrkSt(j, k+3), -1) ,
 ASSIGN(NdUChk(j, k), NdUChk(j, k) +1) ,
 ASSIGN(NdUChk(j, k+1), NdUChk(j, k+1) +1) ,
 ASSIGN(NdUChk(j, k+2), NdUChk(j, k+2) +1) ,
 ASSIGN(NdUChk(j, k+3), NdUChk(j, k+3) +1) ,
 ASSIGN(FstNd(j, C), k-1) ,
 ASSIGN(FstNd(j, C), k+3) ,
 ASSIGN(VbNd(j, C), k)

กฎย่อย R6 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็นVP8 คำนาม คำนาม

R6 / VP → VERB(VP8) N N/
 VPChk (j,k, 8) & ChkVPSt(j,C,NULL) & CatChk (j, k+1, N)
 & CatChk (j, k+2, N)
 = ASSIGN (REL2, OBJ) ,
 ASSIGN (REL3, OBJ) ,
 ASSIGN (VPSt (j,C) , 8) ,

เอกสารนี้เป็น ASSIGN(NdMrkSt(j, k+2), -1) ซึ่งงานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ASSIGN(NdUChk(j, k), NdUChk(j, k) +1),
 ASSIGN(NdUChk(j, k+1), NdUChk(j, k+1) +1) ,
 ASSIGN(NdUChk(j, k+2), NdUChk(j, k+2) +1) ,
 ASSIGN(FstNd(j, C), k-1) ,
 ASSIGN(FstNd(j, C), k+2) ,
 ASSIGN(VbNd(j, C), k)

กฎย่อย R7 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็น VP7 คำนาม คำกริยา-
วิเศษณ์

R7 / VP \longrightarrow VERB(VP7) N ADV/
 VPChk (j,k, 7) & ChkVPSt(j,C,NULL) & CatChk (j, k+1, N)
 & CatChk (j, k+2, ADV)
 = ASSIGN (REL2, OBJ) ,
 ASSIGN (VPSt (j,C) , 7) ,
 ASSIGN(NdMrkSt(j, k+2), -1) ,
 ASSIGN(NdUChk(j, k), NdUChk(j, k) +1) ,
 ASSIGN(NdUChk(j, k+1), NdUChk(j, k+1) +1) ,
 ASSIGN(NdUChk(j, k+2), NdUChk(j, k+2) +1) ,
 ASSIGN(FstNd(j, C), k-1) ,
 ASSIGN(LstNd(j, C), k+2) ,
 ASSIGN(VbNd(j, C), k)

กฎย่อย R8 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็น VP7 คำนาม

R8 / VP \longrightarrow VERB(VP7) N/
 VPChk (j,k, 7) & ChkVPSt(j,C,NULL) & CatChk (j, k+1, N)
 = ASSIGN (REL2, OBJ) ,
 ASSIGN (VPSt (j,C) , 7) ,
 ASSIGN(NdMrkSt(j, k+1), -1) ,
 ASSIGN(NdUChk(j, k), NdUChk(j, k) +1) ,
 ASSIGN(NdUChk(j, k+1), NdUChk(j, k+1) +1) ,

ASSIGN(LstNd(j, C), k+1),

ASSIGN(VbNd(j, C), k)

กฎย่อย R9 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็น VP5 คำบุรพท คำนาม

R9 / VP \longrightarrow VERB(VP5) PREP N/

VPChk (j,k, 5) & ChkVPSt(j,C,NULL) & CatChk (j, k+1, PREP)&

CatChk (j, k+2, N)

= ASSIGN (REL2, OBJ) ,

ASSIGN (VPSt (j,C) , 5) ,

ASSIGN(NdMrkSt(j, k+2), -1) ,

ASSIGN(NdUChk(j, k), NdUChk(j, k) +1) ,

ASSIGN(NdUChk(j, k+1), NdUChk(j, k+1) +1) ,

ASSIGN(NdUChk(j, k+2), NdUChk(j, k+2) +1) ,

ASSIGN(FstNd(j, C), k-1) ,

ASSIGN(LstNd(j, C), k+2),

ASSIGN(VbNd(j, C), k)

กฎย่อย R10 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็น VP3 คำกริยา-วิเศษณ์

R10 / VP \longrightarrow VERB(VP3) ADV/

VPChk (j,k, 3) & ChkVPSt(j,C,NULL) & CatChk (j, k+1, ADV)

= ASSIGN (REL2, NULL) ,

ASSIGN (VPSt (j,C) , 3) ,

ASSIGN(NdMrkSt(j, k+1), -1) ,

ASSIGN(NdUChk(j, k), NdUChk(j, k) +1) ,

ASSIGN(NdUChk(j, k+1), NdUChk(j, k+1) +1) ,

ASSIGN(FstNd(j, C), k-1) ,

ASSIGN(LstNd(j, C), k+1),

ASSIGN(VbNd(j, C), k)

กฎย่อย R11 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็น VP4 คำช่วยกริยา

เอกสารนี้เป็น R11 / VP \longrightarrow VERB(VP4) AUX/ ารศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$VPChk(j, k, 4) \& ChkVPSt(j, C, NULL) \& CatChk(j, k+1, AUX)$
 $= ASSIGN(REL2, NULL),$
 $ASSIGN(VPSt(j, C), 4),$
 $ASSIGN(NdMrkSt(j, k+1), -1),$
 $ASSIGN(NdUChk(j, k), NdUChk(j, k) + 1),$
 $ASSIGN(NdUChk(j, k+1), NdUChk(j, k+1) + 1),$
 $ASSIGN(FstNd(j, C), k-1),$
 $ASSIGN(LstNd(j, C), k+1),$
 $ASSIGN(VbNd(j, C), k)$

กฎย่อย R12 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็น VP1

$R12 / VP \longrightarrow VERB(VP1) /$
 $VPChk(j, k, 1) \& ChkVPSt(j, C, NULL)$
 $= ASSIGN(REL2, NULL),$
 $ASSIGN(VPSt(j, C), 1),$
 $ASSIGN(NdMrkSt(j, k+1), -1),$
 $ASSIGN(NdUChk(j, k), NdUChk(j, k) + 1),$
 $ASSIGN(FstNd(j, C), k-1),$
 $ASSIGN(LstNd(j, C), k),$
 $ASSIGN(VbNd(j, C), k)$

กฎไวยากรณ์ตรวจสอบกริยาวลีกลุ่มที่ 2

กฎย่อย R13 การตรวจสอบกริยาวลีที่ประกอบด้วย คำกริยาที่มีรูปแบบเป็น VP2

$R13 / VP \longrightarrow VERB(VP2) + N/$
 $VPChk(j, k, 2) \& ChkVPSt(j, C, NULL)$
 $= ASSIGN(REL2, OBJ),$
 $ASSIGN(VPSt(j, C), 2),$
 $ASSIGN(NdMrkSt(j, k+1), -1),$
 $ASSIGN(NdUChk(j, k), NdUChk(j, k) + 1),$

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อใช้ในการเรียนการสอนเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ASSIGN(FstNd(j, C) , k) ,

ASSIGN(LstNd(j, C) , k+1),

ASSIGN(VbNd(j, C) , k)

เซตผลลัพธ์ที่ 1 หน่วยคำ (j) \ (k)	1	2	3	4	5	6	7
1	เรือ	โคลง	เนื่องจาก	โคลง	เรือ ₁		
2	เรือ	โคลง	เนื่องจาก	โค	ลง	เรือ ₂	
3	เรือ	โค	ลง	เนื่องจาก	โคลง	เรือ ₃	
4	เรือ	โค	ลง	เนื่องจาก	โค	ลง	เรือ ₄

ตารางที่ 4.2 แสดงผลลัพธ์ที่ได้จากการวิเคราะห์หน่วยคำ

Tkey	B _{j,k,l}	Scat	VP (m)
โคลง	VERB	VACT	6
		VACT	7
		VSTA	1
โค	N	NCMN	
เนื่องจาก	CONJ	CONJ	
เรือ	N	NCMN	
ลง	VERB	NACT	6
		VACT	7

ตารางที่ 4.3 แสดงข้อมูลในระดับโครงสร้างประโยคของหน่วยคำ

ตัวอย่างที่ใช้ในการวิเคราะห์

ในการแสดงตัวอย่างในการวิเคราะห์โครงสร้างประโยค จะใช้ตัวอย่างเดียวกับตัวอย่างในการวิเคราะห์หน่วยคำ โดยจะใช้ผลลัพธ์จากการวิเคราะห์หน่วยคำดังแสดงในตารางที่ 4.2

ขั้นตอนที่ 1 เริ่มทำการวิเคราะห์ในเซต ผลลัพธ์ที่ได้จากการแบ่งคำ เซตที่ 1 (j=1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 2 เริ่มทำการวิเคราะห์หน่วยคำที่ 1 ในเซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ 1 นั่นคือ เริ่มทำการวิเคราะห์ที่คำว่า เรือ

ขั้นตอนที่ 3 เริ่มทำการวิเคราะห์ที่หน้าที่หลักแบบที่ 1 ของคำว่า เรือ

$$\text{ดังนั้น } B_{1,1,1} = N$$

ขั้นตอนที่ 4 $B_{1,1,1} \neq \text{VERB}$ ไปขั้นตอนที่ 11

ขั้นตอนที่ 11 เนื่องจากคำว่าเรือมีหน้าที่คำหลักเพียงหน้าที่เดียว ($r_{1,1} = 1$)

$$\text{ดังนั้น } l = r_{1,1} \text{ ไปขั้นตอนที่ 13}$$

ขั้นตอนที่ 13 เซตผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ 1 มีจำนวนหน่วยคำอยู่ 5 หน่วยคำ ($S_1 = 5$)

$$\text{ดังนั้น } k \neq S_1 \text{ ไปขั้นตอนที่ 14}$$

ขั้นตอนที่ 14 เพิ่มค่าของ k ด้วย 1 ($k=2$) ไปขั้นตอนที่ 3

ขั้นตอนที่ 3 เริ่มทำการวิเคราะห์หน้าที่หลักแบบที่ 1 ของคำว่าโคลง

$$B_{1,2,1} = \text{VERB}$$

ขั้นตอนที่ 4 $B_{1,2,1} = \text{VERB}$ ไปขั้นตอนที่ 5

ขั้นตอนที่ 5 เริ่มทำการวิเคราะห์รูปกริยาของคำว่า โคลงแบบที่ 1 ($m=1$)

$$VP_{1,2}(1) = 6$$

ขั้นตอนที่ 6 ตรวจสอบด้วยกฎไวยากรณ์ ด้วยโปรแกรม Pchk(1,2,1)

ขั้นตอนต่อไปนี้เป็นการทำงานในขั้นตอนของโปรแกรม Pchk(1,2,1)

Pchk ขั้นตอนที่ 1 รับข้อมูลมาจากระบบการแบ่งหน่วยคำ

$$VP_{1,2}(1) = 6$$

Pchk ขั้นตอนที่ 2 $k > 1$ ($k = 2$) ไปขั้นตอนที่ 3

Pchk ขั้นตอนที่ 3 ตรวจสอบหน่วยคำที่ $k-1$ คือคำว่าเรือด้วยกฎไวยากรณ์ R1-R4 พบว่า ตรงกับกฎ R4

เรือเป็นนามวลีได้ ไปที่ขั้นตอน 4

Pchk ขั้นตอนที่ 4 ตรวจสอบ โคลง, เนื่องจาก, โคลง, เรือ ด้วยกฎไวยากรณ์กริยาวลีส่วนที่ 1 พบว่าไม่ตรงกับกฎใด ไปขั้นตอนที่ 6

Pchk ขั้นตอนที่ 6 ส่งผลลัพธ์กลับไปยังส่วนการหาคำหลักขั้นตอนที่ 6

ขั้นตอนต่อไปนี้เป็นการทำงานของโปรแกรม Sgm()

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ขั้นตอนที่ 6 ผลลัพธ์จาก Pchk ปรากฏว่ายังไม่เป็นค่าหลัก ไปขั้นตอนที่ 7
- ขั้นตอนที่ 7 โคลงเป็นหน่วยค่าหลักไม่ได้ ไปขั้นตอนที่ 9
- ขั้นตอนที่ 9 ตรวจสอบแล้วพบว่ายังมีรูปกริยาแบบอื่น ($q=3$) $m \neq q$ ไปขั้นตอนที่ 10
- ขั้นตอนที่ 10 เพิ่มค่าของตัวยา ($m=2$) แล้วไปทำขั้นตอนที่ 6
- ขั้นตอนที่ 6 ตรวจสอบด้วยกฎไวยากรณ์ ด้วยโปรแกรม Pchk(1,2,2)

ขั้นตอนต่อไปนี้เป็นการทำงานในขั้นตอนของโปรแกรม Pchk(1,2,2)

- Pchk ขั้นตอนที่ 1 รับข้อมูลจากส่วนการแบ่งหน่วยค่า

$$VP_{1,2}(2) = 7$$
- Pchk ขั้นตอนที่ 2 $k > 1$ ($k=2$) ไปขั้นตอนที่ 3
- Pchk ขั้นตอนที่ 3 ตรวจสอบหน่วยค่าที่ $k-1$ คือค่าว่าเรือด้วยกฎไวยากรณ์ R1-R4 พบว่าตรงกับกฎ R4
 เรือเป็นนามวลีได้ ไปที่ขั้นตอนที่ 4
- Pchk ขั้นตอนที่ 4 ตรวจสอบ โคลง, เนื่องจาก, โคลง, เรือ ด้วยกฎไวยากรณ์กริยาวลีส่วนที่ 1 พบว่าไม่ตรงกับกฎใด ไปขั้นตอนที่ 6
- Pchk ขั้นตอนที่ 6 ส่งผลลัพธ์กลับไปยังส่วนการหาค่าหลักขั้นตอนที่ 6

ขั้นตอนต่อไปนี้เป็นการทำงานของโปรแกรม StnSgm ()

- ขั้นตอนที่ 6 ผลลัพธ์จาก Pchk ปรากฏว่า ยังไม่เป็นค่าหลัก ไปขั้นตอนที่ 7
- ขั้นตอนที่ 7 โคลงเป็นหน่วยค่าหลักไม่ได้ ไปขั้นตอนที่ 9
- ขั้นตอนที่ 9 ตรวจสอบแล้วพบว่ายังมีรูปกริยาแบบอื่น $m \neq q$ ไปขั้นตอนที่ 10
- ขั้นตอนที่ 10 เพิ่มค่าของ m ด้วย 1 ($m=3$) แล้วไปทำขั้นตอนที่ 6
- ขั้นตอนที่ 6 ตรวจสอบไวยากรณ์ด้วยโปรแกรม Pchk(1,2,3)

ขั้นตอนต่อไปนี้เป็นการทำงานในขั้นตอนของโปรแกรม Pchk(1,2,3)

- Pchk ขั้นตอนที่ 1 รับข้อมูลจากส่วนการแบ่งหน่วยค่า

$$VP_{1,2}(2) = 1$$
- Pchk ขั้นตอนที่ 2 $k > 1$ ($k=2$) ไปขั้นตอนที่ 3

Pchk ขั้นตอนที่ 3 ตรวจสอบหน่วยคำที่ $k-1$ คือคำว่าเร็วด้วยกฎไวยากรณ์ R1-R4 พบว่าตรงกับกฎ R4

เร็วเป็นนามวลีได้ ไปที่ขั้นตอนที่ 4

Pchk ขั้นตอนที่ 4 ตรวจสอบ โครง,เนื่องจาก,โคลง,เร็ว ด้วยกฎไวยากรณ์ส่วนที่ 1 พบว่าตรงกับกฎ R12 ดังนั้น โคลงเป็นคำหลักของประโยคที่ (1,1) ไปขั้นตอนที่ 6

Pchk ขั้นตอนที่ 6 ส่งผลลัพธ์กลับไปยังส่วนการหาคำหลักขั้นตอนที่ 6

ขั้นตอนต่อไปนี้เป็นการทำงานของโปรแกรม StnSgm()

ขั้นตอนที่ 6 ได้รับผลลัพธ์จาก Pchk ไปขั้นตอนที่ 7 โคลงเป็นหน่วยคำหลักได้ ไปขั้นตอนที่ 8

ขั้นตอนที่ 8 เพิ่มค่า C ด้วย 1 ($C=2$)

ขั้นตอนที่ 9 ทำการตรวจสอบแล้วพบว่าไม่มีรูปกริยาแบบอื่น $m=q$ ไปขั้นตอน 11

ขั้นตอนที่ 11 ทำการตรวจสอบพบว่าไม่มีหน้าที่หลักแบบอื่น $l = r_{1,2}$ ไปขั้นตอนที่ 13

เมื่อทำมาถึงขั้นตอนนี้ จะได้โครงสร้างประโยค 1 ประโยคเมื่อทำต่อไปจนครบทุกหน่วยคำ และทุกเซตผลลัพธ์จากการแบ่งคำแล้วจะได้ผลลัพธ์ดังตารางที่

ตัวอย่างในการวิเคราะห์

เซตผลลัพธ์ที่ 1

หน่วยคำที่	1	2	3	4	5
Tkey	เร็ว	โคลง	เนื่องจาก	โคลง	เร็ว
NduChk	1	1	0	0	0
NdMrkSt	0	0	0	0	0

ตารางที่ 4.4 แสดงผลลัพธ์ที่ได้จากระบบวิเคราะห์ประโยคภาษาไทยจากย่อหน้าเซตที่ 1

เซตผลลัพธ์ที่ 2

หน่วยคำที่	1	2	3	4	5	6
Tkey	เร็ว	โคลง	เนื่องจาก	โค	ลง	เร็ว
NduChk	1	1	0	1	1	1
NdMrkSt	0	-1	0	0	0	-1

ตารางที่ 4.5 แสดงผลลัพธ์ที่ได้จากระบบวิเคราะห์ประโยคภาษาไทยจากย่อหน้าเซตที่ 2

เซตผลลัพธ์ที่ 3

หน่วยคำที่	1	2	3	4	5	6
Tkey	เร็ว	โค	ลง	เนื่องจาก	โคลง	เร็ว
NduChk	0	0	0	0	0	0
NdMrkSt	0	0	0	0	0	0

ตารางที่ 4.6 แสดงผลลัพธ์ที่ได้จากระบบวิเคราะห์ประโยคภาษาไทยจากย่อหน้าเซตที่ 3

หน่วยคำที่	1	2	3	4	5	6	6
Tkey	เร็ว	โค	ลง	เนื่องจาก	โค	ลง	เร็ว
NduChk	0	0	0	0	1	1	1
NdMrkSt	0	0	0	0	0	0	-1

ตารางที่ 4.7 แสดงผลลัพธ์ที่ได้จากระบบวิเคราะห์ประโยคภาษาไทยจากย่อหน้าเซตที่ 4

4.4 ส่วนการวิเคราะห์แยกประโยคจากย่อหน้า (Paragraph Segmentation Analysis)

จากหัวข้อ 4.3 เราได้สร้างระบบวิเคราะห์โครงสร้างประโยค ซึ่งจะทำการวิเคราะห์หาความสัมพันธ์ทางวากยสัมพันธ์และความสัมพันธ์ทางอรรถศาสตร์ของหน่วยคำในประโยค อย่างไรก็ตามระบบวิเคราะห์ที่ได้กล่าวมาข้างต้นยังมีข้อจำกัดในการวิเคราะห์ความสัมพันธ์ระหว่างประโยคที่นำมาเขียนรวมกันเป็นย่อหน้า งานวิจัยนี้ได้ทำการวิเคราะห์แยกประโยคจากย่อหน้าโดยการวิเคราะห์หาหน่วยคำในเซตผลลัพธ์ที่ได้จากการแบ่งคำ หากหน่วยคำใดถูกวิเคราะห์เป็นส่วนหนึ่งของประโยคได้ จะทำการกำหนดไว้ว่าถูกวิเคราะห์แล้ว ในกรณีที่หน่วยคำใดไม่ถูกวิเคราะห์เป็นส่วนหนึ่งในประโยค เนื่องจากไม่สามารถทำหน้าที่ของคำในประโยคนั้นได้ ก็จะวิเคราะห์หาว่าหน่วยคำนั้นเป็นคำสันธานหรือไม่ ถ้าเป็นคำสันธานจะทำการกำหนดไว้ว่าถูกวิเคราะห์หาหน้าที่ของคำได้ โดยให้เป็นคำเชื่อมประโยค โดยมีการทำงานดังนี้

อัลกอริธึมของการแยกประโยคจากย่อหน้า

การแยกประโยคจากย่อหน้า

โปรแกรม OT ()

กำหนดให้

- $r_{j,k}$ จำนวนหน้าที่หลักในหน่วยคำที่ j ในเซตคำตอบ k
- $g_{j,k}$ จำนวนรูปแบบของหน่วยคำกริยาในหน่วยคำที่ j ในเซตคำตอบ k
- S_j จำนวนของหน่วยคำที่แบ่งได้ในเซตคำตอบ j
- $B_{j,k,l}$ หมายถึงหน้าที่หลักของหน่วยคำแบบที่ l ในหน่วยคำที่ k ในเซตคำตอบที่ได้จากการแบ่งคำเซตที่ j
- P_j ผลลัพธ์ของการแยกประโยคจากย่อหน้า (j เป็นลำดับที่ของเซตคำตอบ)

กระบวนการวิเคราะห์แยกประโยคจากย่อหน้า

- ขั้นตอนที่ 1 เริ่มทำการวิเคราะห์เซตของผลลัพธ์ จากการแบ่งคำเซตที่ 1 ($j=1$)
- ขั้นตอนที่ 2 เริ่มทำการวิเคราะห์หน่วยคำที่ 1 ในเซตของผลลัพธ์จากการแบ่งคำเซตที่ j กำหนดให้ $P_j = \text{TRUE}$ ($j = 1, 2, \dots$, โดยที่ j เป็นลำดับที่ในเซตผลลัพธ์ที่ได้จากการแบ่งคำ)
- ขั้นตอนที่ 3 ตรวจสอบว่า $NdUChk_{j,k} = 0$ หมายถึง หน่วยคำที่ k ในเซตของผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j ไม่อยู่ในประโยคที่วิเคราะห์ไว้ให้ไปขั้นตอนที่ 4
ถ้า $NdUChk_{j,k} \neq 0$ หมายถึง หน่วยคำที่ k ในเซตของผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j อยู่ในประโยคที่วิเคราะห์ไว้ให้ไปขั้นตอนที่ 5
- ขั้นตอนที่ 4 ตรวจสอบว่าหน่วยคำที่ k ในเซตของผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j อยู่ในกลุ่มคำหลักที่เป็น คำสันธาน หรือไม่
- ขั้นตอนที่ 5 ถ้าไม่อยู่ในกลุ่มคำสันธาน หมายถึงมีหน่วยคำที่ไม่สามารถจะจัดอยู่ในประโยคใดประโยคหนึ่งได้ ให้ไปขั้นตอนที่ 6 แต่ถ้าหน่วยคำนั้นอยู่ในกลุ่มคำสันธานให้ไปขั้นตอนที่ 7
- ขั้นตอนที่ 6 กำหนดให้เซตของผลลัพธ์ที่ได้จากการแบ่งคำเซตที่ j เป็นเซตที่ไม่สามารถแยกประโยคจากย่อหน้าได้ แล้วไปขั้นตอนที่ 7
- ขั้นตอนที่ 7 ถ้า $NdUChk_{j,k} \geq 2$ ให้ไปขั้นตอนที่ 8
ถ้า $NdUChk_{j,k} < 2$ ให้ไปขั้นตอนที่ 9
- ขั้นตอนที่ 8 ตรวจสอบหา C_q โดยที่ $LstNd(i, C_q) = FstNd(i, C_q+1) = k$ เพื่อหาคู่ประโยคที่มีหน่วยคำหน่วยเดียวกันอยู่ในประโยคซึ่งหน่วยคำที่ k จะอยู่ที่ในประโยคที่ C_q และ C_{q+1} ในเซตผลลัพธ์ของการแบ่งคำเซตที่ j
- ขั้นตอนที่ 9 ถ้าหน่วยคำกริยาของประโยคที่ C_q อยู่ในกลุ่มของ VP7 หรือ VP3 หรือ VP4 ให้ไปที่ขั้นตอนที่ 10
ถ้าไม่อยู่ในกลุ่มเหล่านั้น ให้ไปที่ขั้นตอนที่ 12
- ขั้นตอนที่ 10 ถ้า $Lstnd(j, C_q) - VbNd(j, C_q) = 1$ ให้ไปที่ขั้นตอนที่ 11
ถ้า $Lstnd(j, C_q) - VbNd(j, C_q) \neq 1$ ให้ไปที่ขั้นตอนที่ 12
- ขั้นตอนที่ 11 เลื่อนหน่วยคำท้ายประโยคที่ C_q ขึ้นมา 1 หน่วยคำ แล้วไปที่ขั้นตอนที่ 12
- ขั้นตอนที่ 12 ถ้าหน่วยคำกริยาของประโยคที่ C_q อยู่ในกลุ่มของ VP7 หรือ VP5 หรือ VP8 ให้ไปที่ขั้นตอนที่ 13
ถ้าไม่อยู่ในกลุ่มเหล่านั้นให้ไปที่ขั้นตอนที่ 15
- ขั้นตอนที่ 13 ถ้า $Lstnd(j, C_q) - VbNd(j, C_q) = 2$ ให้ไปที่ขั้นตอนที่ 14

- ถ้า $Lstnd(j, C_q) - VbNd(j, C_q) \neq 2$ ให้ไปที่ขั้นตอนที่ 16
- ขั้นตอนที่ 14 ถ้าหน่วยคำที่ k มีหน้าที่คำหลักเป็น N ไปขั้นตอนที่ 15
ถ้าหน่วยคำที่ k ไม่มีหน้าที่คำหลักเป็น N ไปขั้นตอนที่ 16
- ขั้นตอนที่ 15 เลื่อนหน่วยคำท้ายประโยคที่ C_q ขึ้นมา 1 หน่วยคำ แล้วไปที่ขั้นตอนที่ 16
- ขั้นตอนที่ 16 ถ้าหน่วยคำกริยาของประโยคที่ C_q อยู่ในกลุ่มของ VP6 ให้ไปที่ขั้นตอนที่ 17 ถ้า
ไม่อยู่ในกลุ่มเหล่านั้น ให้ไปที่ขั้นตอนที่ 20
- ขั้นตอนที่ 17 ถ้า $LstNd(j, C_q) - VbNd(j, C_q) = 3$ ให้ไปที่ขั้นตอนที่ 18
ถ้า $LstNd(j, C_q) - VbNd(j, C_q) \neq 3$ ให้ไปที่ขั้นตอนที่ 20
- ขั้นตอนที่ 18 ถ้าหน่วยคำที่ k มีหน้าที่คำหลักเป็น N ไปขั้นตอนที่ 19
ถ้าหน่วยคำที่ k ไม่มีหน้าที่คำหลักเป็น N ไปขั้นตอนที่ 20
- ขั้นตอนที่ 19 เลื่อนหน่วยคำท้ายประโยคที่ C_q ขึ้นมา 1 หน่วยคำ แล้วไปที่ขั้นตอนที่ 20
- ขั้นตอนที่ 20 ถ้า $k \neq S_j$ ไปขั้นตอนที่ 21
ถ้า $k = S_j$ ไปขั้นตอนที่ 22
- ขั้นตอนที่ 21 เพิ่มค่า k ด้วย 1 ไปขั้นตอนที่ 3
- ขั้นตอนที่ 22 ถ้า $j \neq t$ ไปขั้นตอนที่ 23
ถ้า $j = t$ ไปขั้นตอนที่ 22
- ขั้นตอนที่ 23 เพิ่มค่า j ด้วย 1 ไปขั้นตอนที่ 2
- ขั้นตอนที่ 24 จบการทำงาน

ตัวอย่างการวิเคราะห์แยกประโยคจากย่อหน้า OT()

ตัวอย่างนี้จะทำการวิเคราะห์ต่อจากตัวอย่างของการวิเคราะห์โครงสร้างประโยคซึ่งจากตาราง
ที่ () ซึ่งเป็นผลลัพธ์ที่ได้จากการวิเคราะห์โครงสร้างประโยค ซึ่งได้วิเคราะห์ไว้แล้วทั้ง 4 เซต

การพิจารณาเซตที่ 1 พบว่า หน่วยคำที่มี $NdUChk = 0$ มี 3 หน่วยคำคือ หน่วยคำที่ 3,4,5 ซึ่ง
เมื่อตรวจสอบกับฐานข้อมูลพจนานุกรมแล้วพบว่า มีเพียงหน่วยคำที่ 3 เพียงหน่วยคำเดียวที่เป็นคำ
สันธาน ดังนั้นเซตคำตอบนี้จะไม่สามารถวิเคราะห์แยกประโยคจากย่อหน้าได้

การพิจารณาเซตที่ 2 พบว่า หน่วยคำที่มี $NdUChk = 0$ มี 1 หน่วยคำคือ หน่วยคำที่ 3 ซึ่งเมื่อ
ตรวจสอบกับฐานข้อมูลพจนานุกรมคำศัพท์แล้วพบว่า หน่วยคำที่ 3 เป็นคำสันธาน ดังนั้นเมื่อทุก
หน่วยคำมี $NdUChk = 1$ แล้ว ก็จะทำให้การแยกประโยคออกมา โดยจะทำการแยกตั้งแต่หน่วยแรกไป
จนกว่าจะถึงหน่วยคำที่ $NdMrkSt = -1$ จากนั้น ถ้าพบว่าคำต่อไปเป็นคำสันธานจะแยกออกมา และ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตรวจสอบหน่วยคำต่อไป ถ้าหน่วยคำหน่วยต่อไปไม่ใช่คำสันธานจะทำการแยกตั้งแต่หน่วยคำหน่วยนั้นไปจนกว่าจะพบหน่วยคำที่ $NdMrkSt = -1$ อีก ท้ายที่สุดเราจะได้ผลลัพธ์คือ

ประโยคที่ 1 เรือ โคลง ตามกฎที่ R4, R12

ประโยคที่ 2 โค ลง เรือ ตามกฎที่ R4,R7

คำเชื่อมประโยคที่ 1 และ 2 เนื่องจาก

จะเห็นว่าหน่วยคำทุกหน่วยคำถูกใช้เป็นหน่วยคำในประโยคและถูกใช้เป็นคำเชื่อมประโยค ดังนั้นการแยกประโยคจากย่อหน้าในเซตคำตอบที่ 2 กระทำได้สมบูรณ์

การพิจารณาเซตที่ 3 พบว่าหน่วยคำที่มี $NdUChk = 0$ มีทุกหน่วยคำคือหน่วยคำที่ 1,2,3,4,5,6 ซึ่งเมื่อตรวจสอบกับฐานข้อมูลพจนานุกรมคำศัพท์พบว่า มีเพียงหน่วยคำที่ 3 เพียงหน่วยเดียวที่เป็นคำสันธาน ดังนั้นเซตคำตอบนี้จะไม่สามารถแยกประโยคจากย่อหน้าได้

การพิจารณาเซตที่ 4 พบว่าหน่วยคำที่มี $NdUChk = 0$ มีทุกหน่วยคำคือหน่วยคำที่ 1,2,3,4,5,6 ซึ่งเมื่อตรวจสอบกับฐานข้อมูลพจนานุกรมคำศัพท์พบว่า มีเพียงหน่วยคำที่ 3 เพียงหน่วยเดียวที่เป็นคำสันธาน ดังนั้นเซตคำตอบนี้จะไม่สามารถแยกประโยคจากย่อหน้าได้

บทที่ 5

ซอฟต์แวร์ของระบบ

5.1 บทนำ

ดังที่ได้กล่าวมาแล้วในบทที่ 1 ว่างานวิจัยนี้แบ่งการวิเคราะห์ประโยคภาษาไทยจากย่อหน้า เป็น 3 ขั้นตอนใหญ่ ที่ทำงานเป็นลำดับต่อเนื่องกันเพื่อให้สะดวกต่อการพัฒนาและให้ความละเอียด และความถูกต้องในการวิเคราะห์มากขึ้น ตลอดจนสะดวกต่อการแก้ไขเพิ่มเติมกฎไวยากรณ์ภายใน ขบวนการวิเคราะห์โครงสร้างประโยคภาษาไทย ซึ่งกระบวนการทั้งหมดได้ถูกนำมาสร้างเป็นซอฟต์แวร์ ซอฟต์แวร์ของระบบในขบวนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้ามีการประมวลผลตามลำดับคือ

ส่วนที่หนึ่ง ซอฟต์แวร์ระบบส่วนการวิเคราะห์ในระดับคำ

ส่วนที่สอง ซอฟต์แวร์ระบบส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทย

ส่วนที่สาม ซอฟต์แวร์ระบบส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

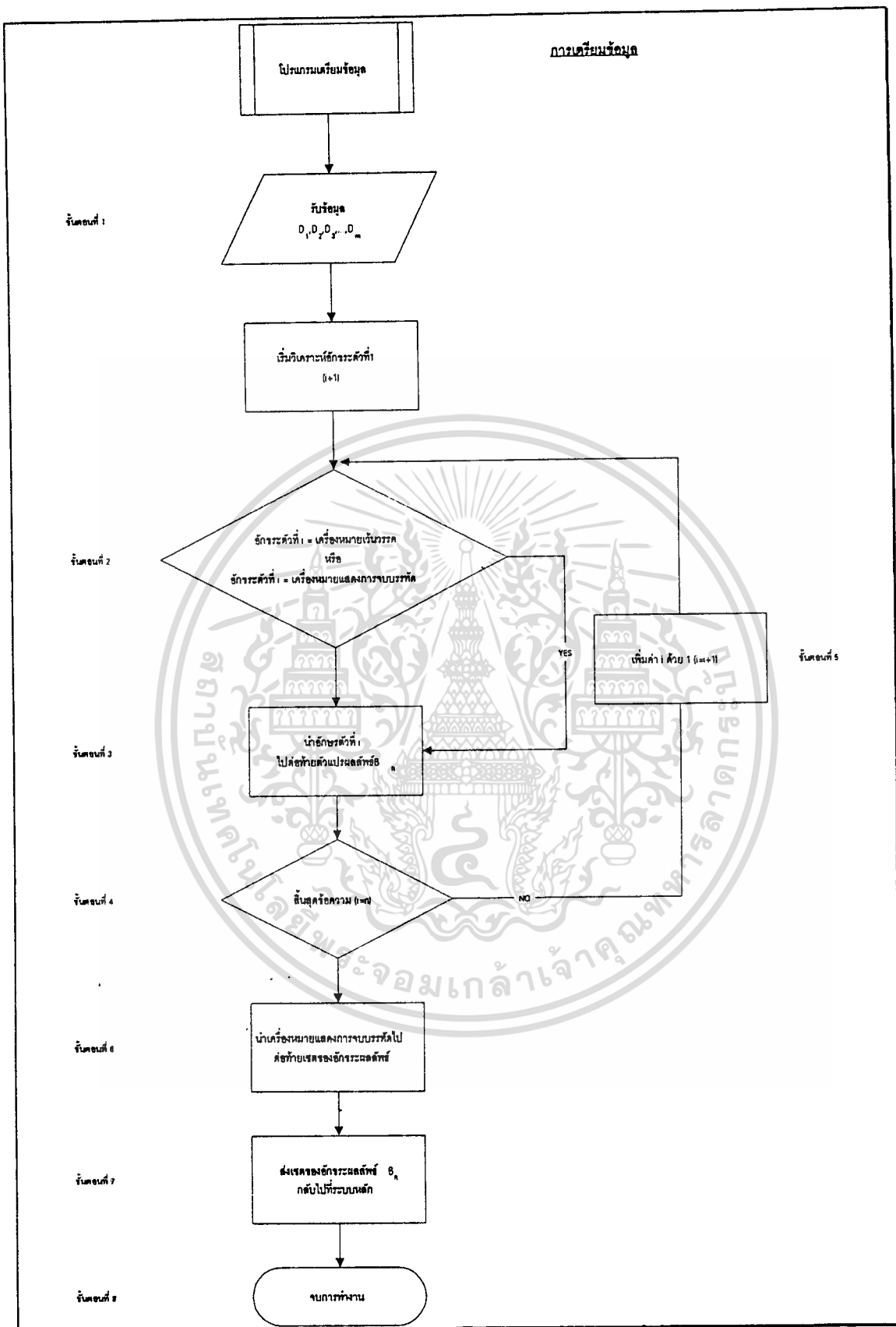
โดยมีรายละเอียดดังนี้

5.2 ซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับคำ (Morphological Analysis System)

จากอัลกอริทึมของส่วนการวิเคราะห์ในระดับคำที่กล่าวมาแล้วในข้อ 4.2 งานวิจัยนี้ได้นำมาใช้ในการสร้างซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับคำ โดยได้แบ่งการทำงานออกเป็นสองโปรแกรมดังนี้

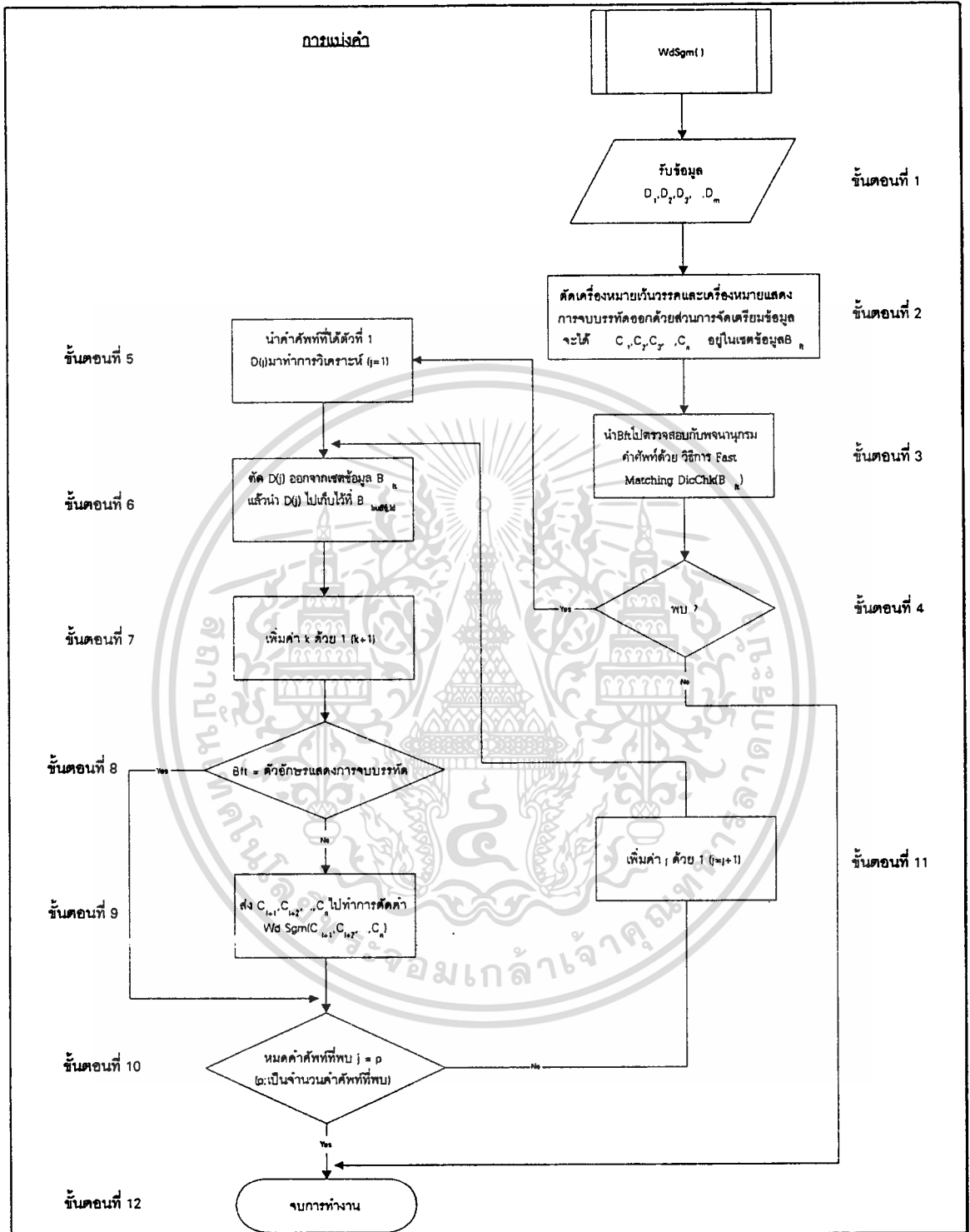
- 1) โปรแกรมการจัดเตรียมข้อมูล
- 2) โปรแกรมการแบ่งคำ

ซึ่งมีการทำงานดังแสดงไว้ในผังงานรูปที่ 5.1 และ รูปที่ 5.2



รูปที่ 5.1 แสดงผังงานของการจัดเตรียมข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานในเครือข่ายภายในเท่านั้น เมื่อผู้ผู้ดูแลระบบนำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.2 แสดงโฟลว์ชาร์ตของระบบวิเคราะห์แบ่งคำ

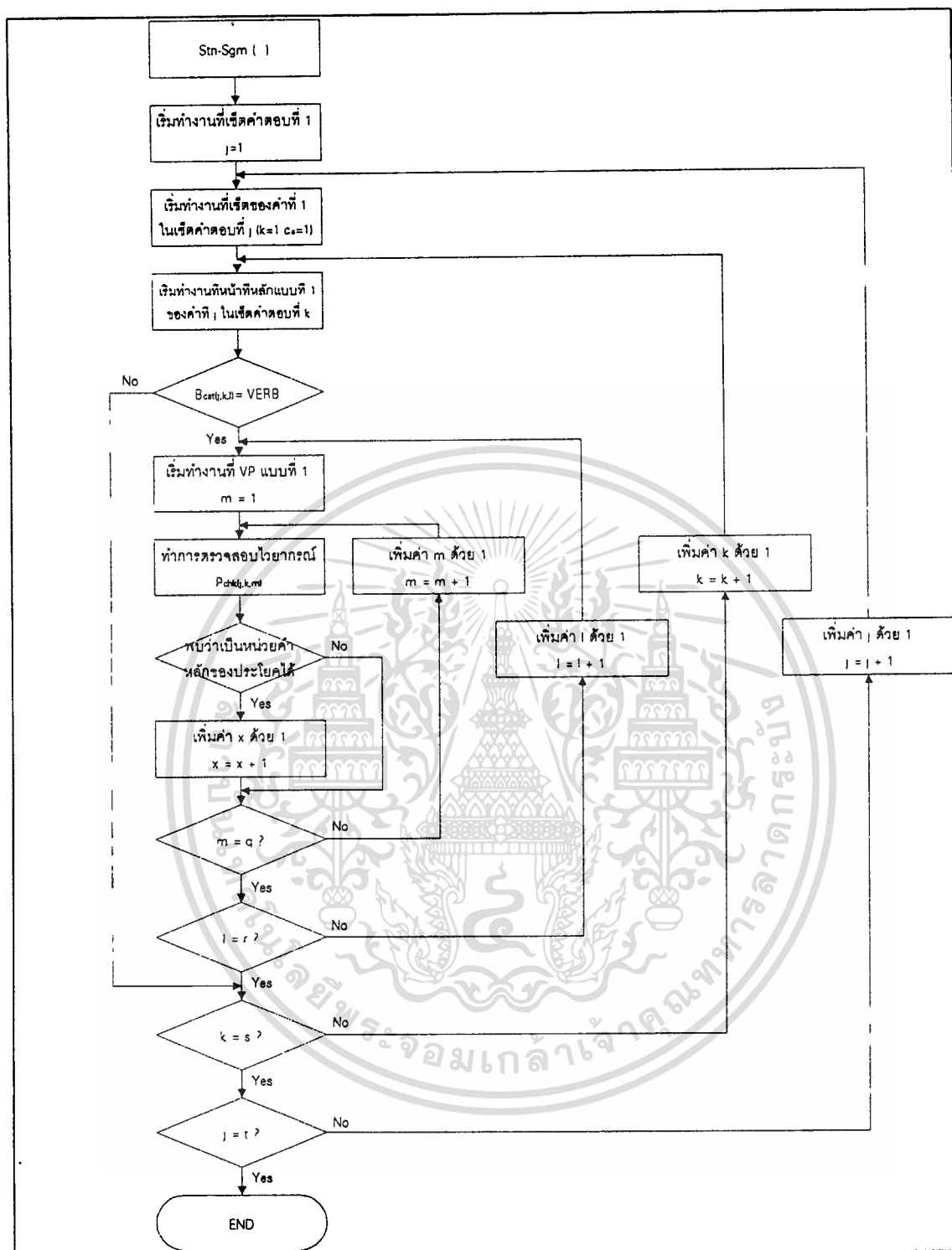
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 ซอฟต์แวร์ระบบส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทย

งานวิจัยนี้ได้สร้างซอฟต์แวร์ระบบส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทยจากอัลกอริทึมของส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทยที่ได้กล่าวมาแล้วในข้อ 4.3 โดยสร้างเป็นโปรแกรมหลักได้ 2 โปรแกรมดังนี้

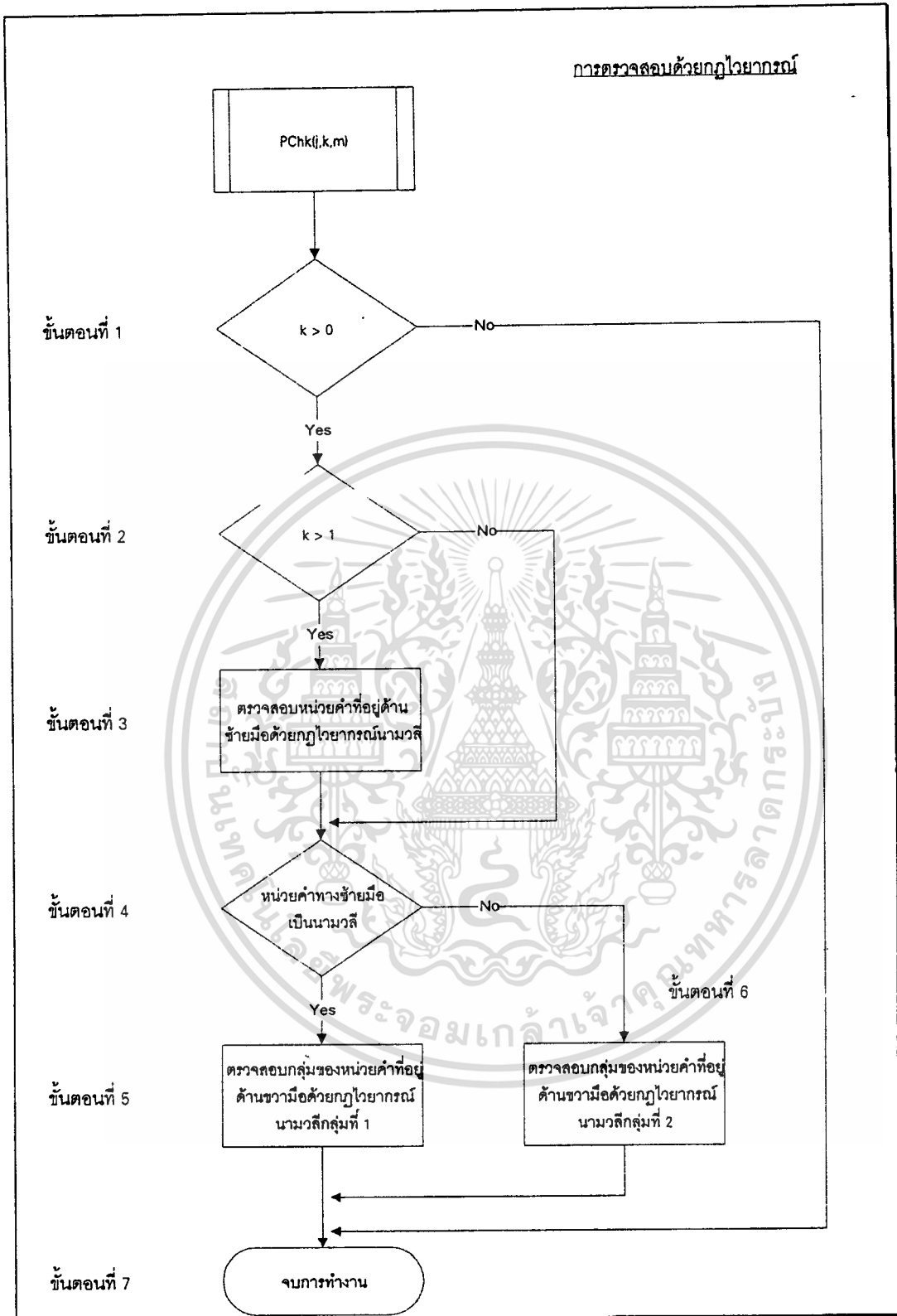
- 1) โปรแกรมหลักในการส่วนการวิเคราะห์หาหน่วยคำหลัก
- 2) โปรแกรมหลักในการตรวจสอบโครงสร้างประโยคด้วยกฎไวยากรณ์





รูปที่ 5.3 แสดงผังงานของซอฟต์แวร์ของระบบส่วนการวิเคราะห์หาหน่วยคำหลัก

การตรวจสอบด้วยกฎไวยากรณ์



รูปที่ 5.4 แสดงการตรวจสอบโครงสร้างประโยคด้วยกฎไวยากรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

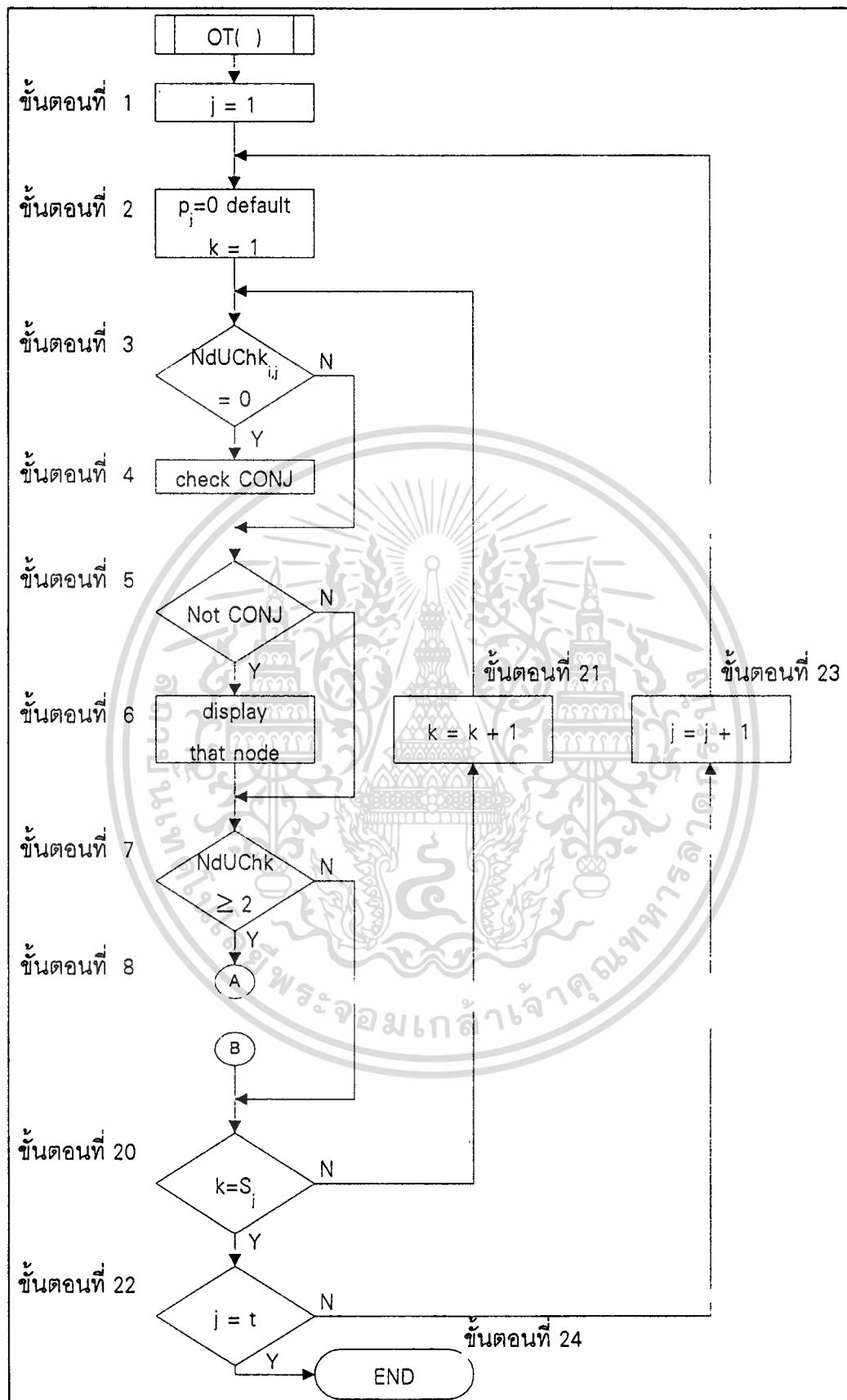
5.4 ซอฟต์แวร์ของระบบส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

งานวิจัยนี้ได้สร้างซอฟต์แวร์ของระบบส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จากอัลกอริธึมของส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้าที่ได้กล่าวมาแล้วในหัวข้อ 4.4 โดยมีโปรแกรมการทำงานหลัก 1 โปรแกรมดังนี้

- 1) โปรแกรมวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า ดังแสดงผังงานการทำงานไว้ในรูปที่

5.5





รูปที่ 5.5 แสดงผังงานของส่วนวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.5 การใช้งานซอฟต์แวร์ระบบ

การใช้งานซอฟต์แวร์ของระบบแยกประโยคภาษาไทยจากย่อหน้า จะเป็นลักษณะการติดต่อระหว่างผู้ใช้กับระบบซอฟต์แวร์ ผ่าน โปรแกรมวินโดว์ ดังแสดงไว้ในรูปที่ 5.6

การทำงานของซอฟต์แวร์ระบบ จะรับข้อมูลอินพุตจากหน้าต่าง Paragraph Input โดยมีข้อกำหนดคือต้องเป็นตัวอักษรภาษาไทย และวรรณยุกต์ไทยดังได้กำหนดไว้ในหัวข้อที่ 4.2.1

ในการสั่งให้ระบบทำงานจะใช้เมาส์เป็นตัวกำหนดการทำงานโดยทำงานร่วมกับระบบเมนูซึ่งประกอบด้วย

คำสั่งในการวิเคราะห์หน่วยคำ

คือ Word Segment.

คำสั่งในการวิเคราะห์โครงสร้างประโยค

คือ Sentence Analysis

คำสั่งในการวิเคราะห์แยกประโยคจากย่อหน้า

คือ Sentence Segment.

คำสั่งในการตรวจสอบผลลัพธ์ทั่วไป

คือ General Message

คำสั่งในการตรวจสอบผลลัพธ์ที่อยู่ในช่วงการวิเคราะห์

คือ Precessing Message

คำสั่งในการตรวจสอบผลลัพธ์ที่ผิดพลาด

คือ Error Message

รูปที่ 5.6 แสดงส่วนติดต่อกับผู้ใช้ของระบบแยกประโยคภาษาไทยจากย่อหน้า

5.5.1 การใช้งานซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับคำ

ในการใช้งานซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับคำมีการทำงาน 4 ขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1 ป้อนข้อมูลอินพุตข้อความย่อหน้าต้นแบบหนึ่งย่อหน้าตามข้อกำหนดที่กล่าวมาแล้วในข้อ 4.2.1 โดยการป้อนข้อมูลอินพุตนี้จะต้องกระทำที่หน้าต่างรับข้อมูลเข้าเท่านั้น ดังแสดงในรูปที่ 5.7

ขั้นตอนที่ 2 ทำการเปิดหน้าต่างแสดงผลโดยใช้เมาส์ชี้ไปที่ปุ่ม Word Sgmt. ดังแสดงในรูปที่ 5.8

ขั้นตอนที่ 3 ใช้คำสั่งเริ่มต้นโปรแกรมโดยใช้เมาส์ชี้ไปที่เมนู Word Segment ดังแสดงในรูปที่ 5.9

ขั้นตอนที่ 4 ตรวจสอบผลลัพธ์ที่หน้าต่าง Word Segmentation ดังแสดงในรูปที่ 5.10

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input

เร็วโคลงเนื่องจากโคลงเร็ว

General Message

Word Segmentation

Word Sgmt.
 Sentence Anal.
 Sentence Sgmt.

รูปที่ 5.7 แสดงการป้อนข้อมูลอินพุทของข้อความย่อหน้าต้นแบบ

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input

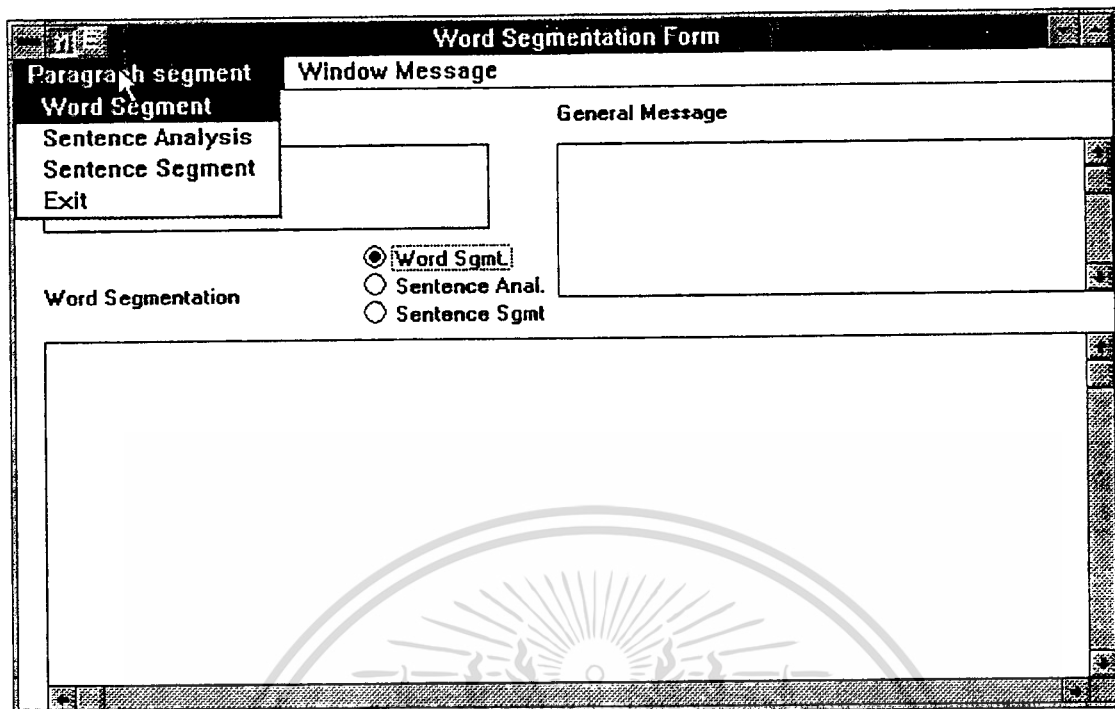
เร็วโคลงเนื่องจากโคลงเร็ว

General Message

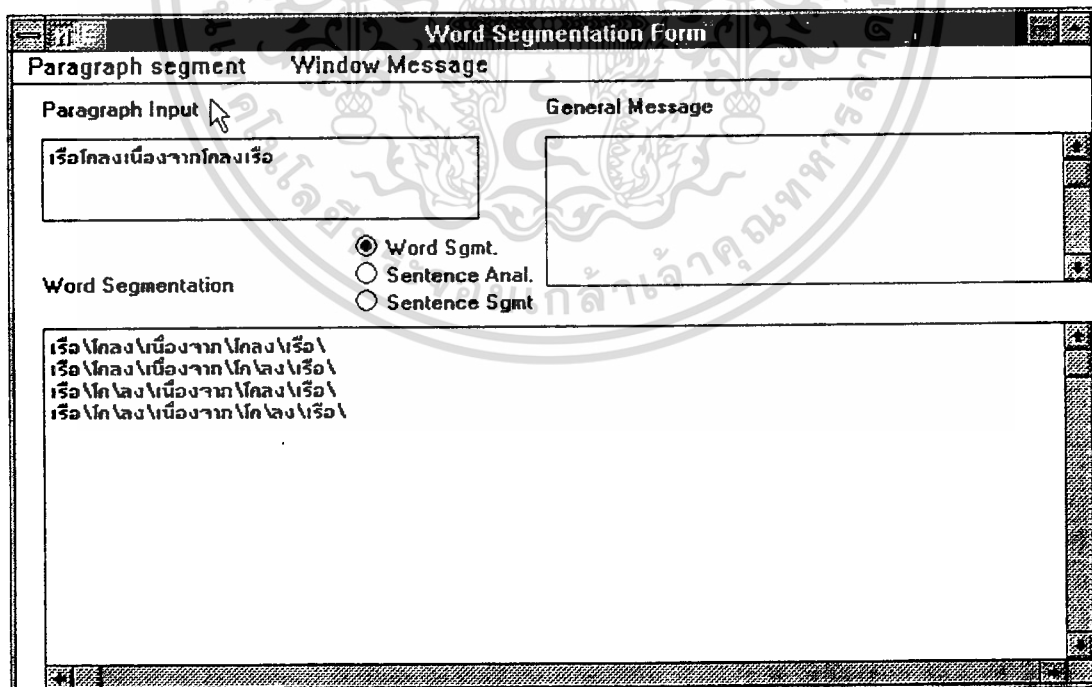
Word Segmentation

Word Sgmt.
 Sentence Anal.
 Sentence Sgmt.

เอกสารนี้เป็นเอกสารรูปที่ 5.8 แสดงการเปิดหน้าต่างแสดงผลของส่วนการวิเคราะห์หน่วยคำโยชนด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.9 แสดงการใช้งานซอฟต์แวร์ระบบส่วนของการวิเคราะห์ระดับหน่วยคำ



รูปที่ 5.10 แสดงผลลัพธ์ที่ได้จากซอฟต์แวร์ระบบส่วนของการวิเคราะห์หน่วยคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.5.2 การใช้งานซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทย

ในการใช้งานซอฟต์แวร์ของระบบส่วนการวิเคราะห์ในระดับโครงสร้างประโยคมีการทำงาน 3 ขั้นตอนดังต่อไปนี้

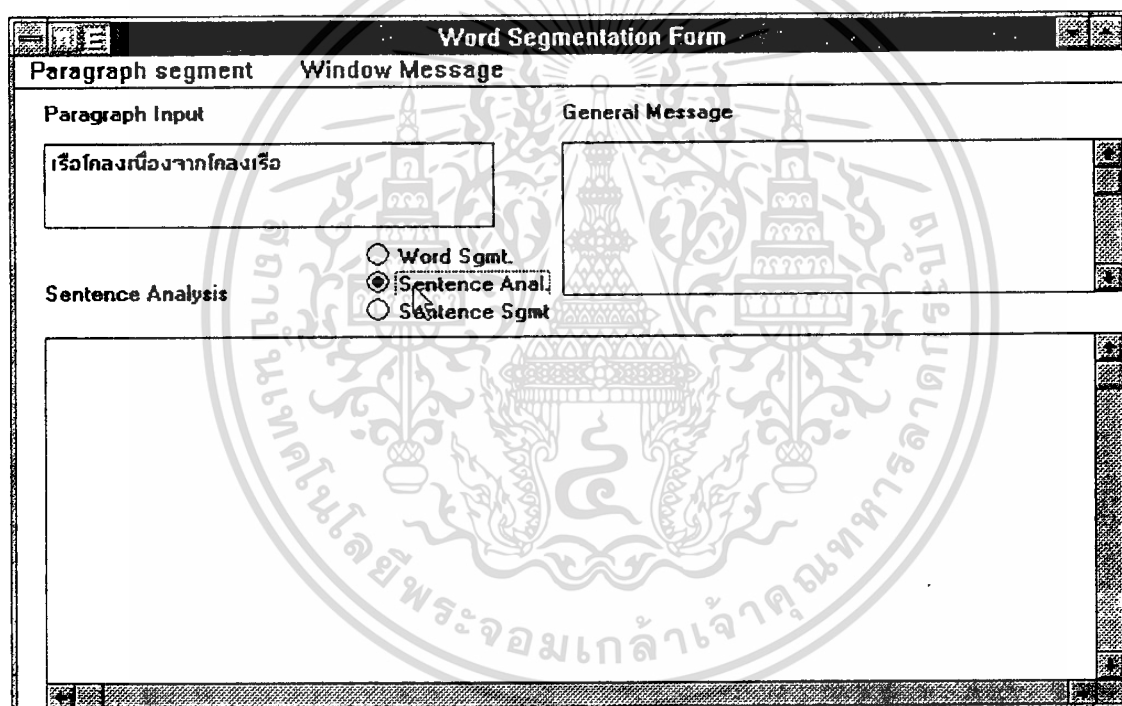
ขั้นตอนที่ 1 ทำการเปิดหน้าต่างแสดงผลโดยใช้เมาส์ชี้ไปที่ปุ่ม Sentence Anal. ดังแสดงในรูปที่

5.11

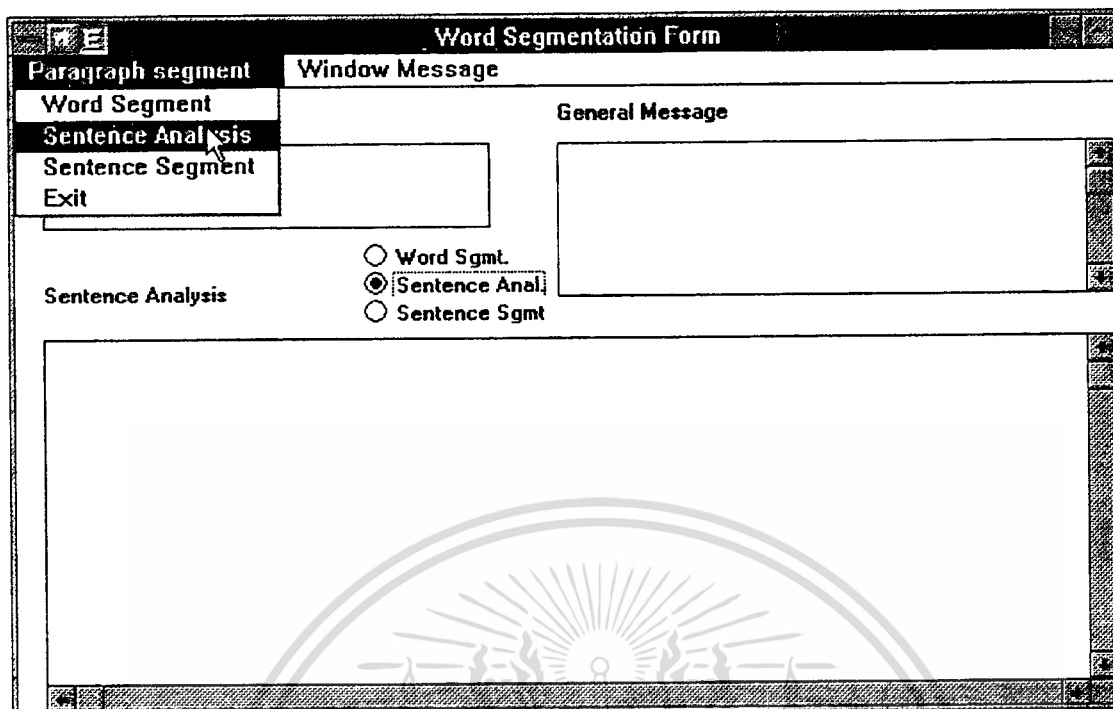
ขั้นตอนที่ 2 ใช้คำสั่งเริ่มต้นโปรแกรมโดยใช้เมาส์ชี้ไปที่เมนู Sentence Analysis ดังแสดงในรูปที่

5.12

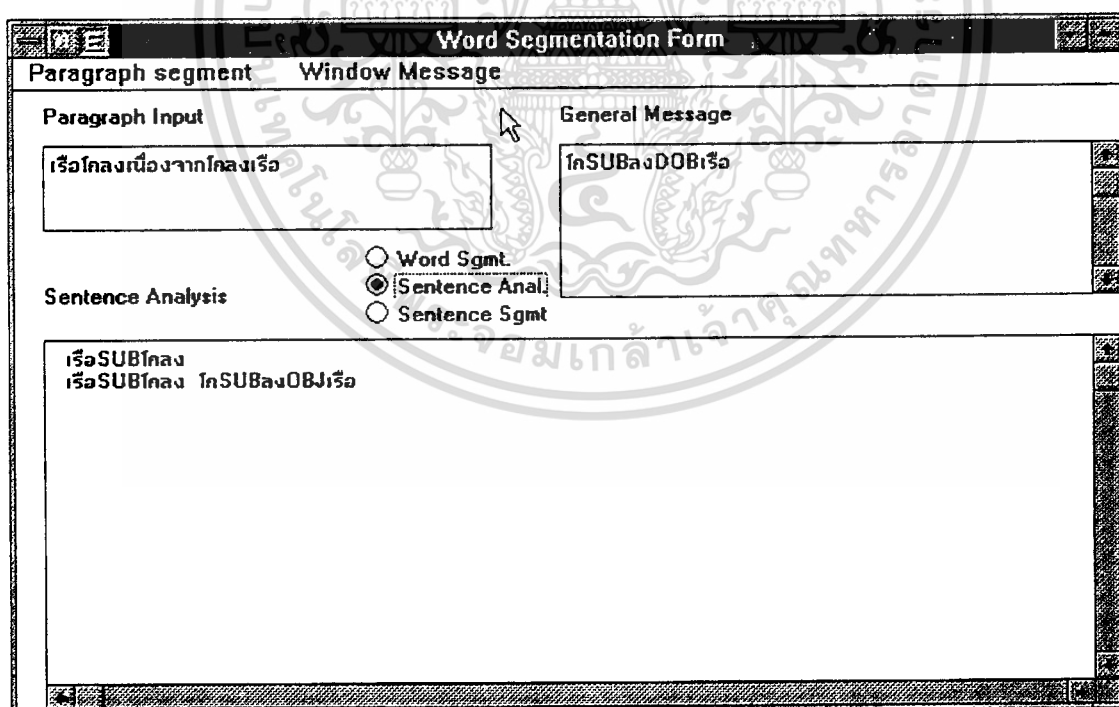
ขั้นตอนที่ 3 ตรวจสอบผลลัพธ์ที่หน้าต่าง Sentence Analysis ดังแสดงในรูปที่ 5.13



รูปที่ 5.11 แสดงการเปิดหน้าต่างแสดงผลส่วนการวิเคราะห์โครงสร้างประโยคภาษาไทย



รูปที่ 5.12 แสดงการใช้งานซอฟต์แวร์ระบบส่วนการวิเคราะห์โครงสร้างประโยคภาษาไทย



รูปที่ 5.13 แสดงผลลัพธ์ที่ได้จากซอฟต์แวร์ระบบส่วนการวิเคราะห์โครงสร้างประโยคภาษาไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

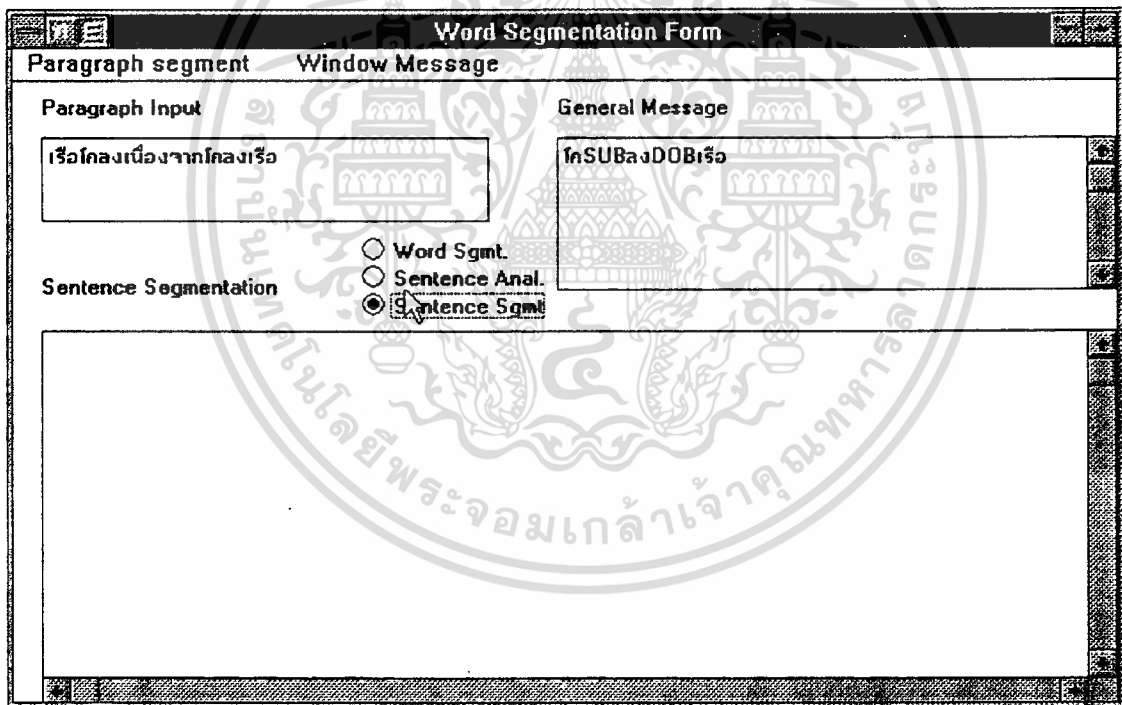
5.5.3 การใช้งานซอฟต์แวร์ของระบบส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

ในการใช้งานซอฟต์แวร์ของระบบส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้ามีการทำงาน 3 ขั้นตอนดังต่อไปนี้

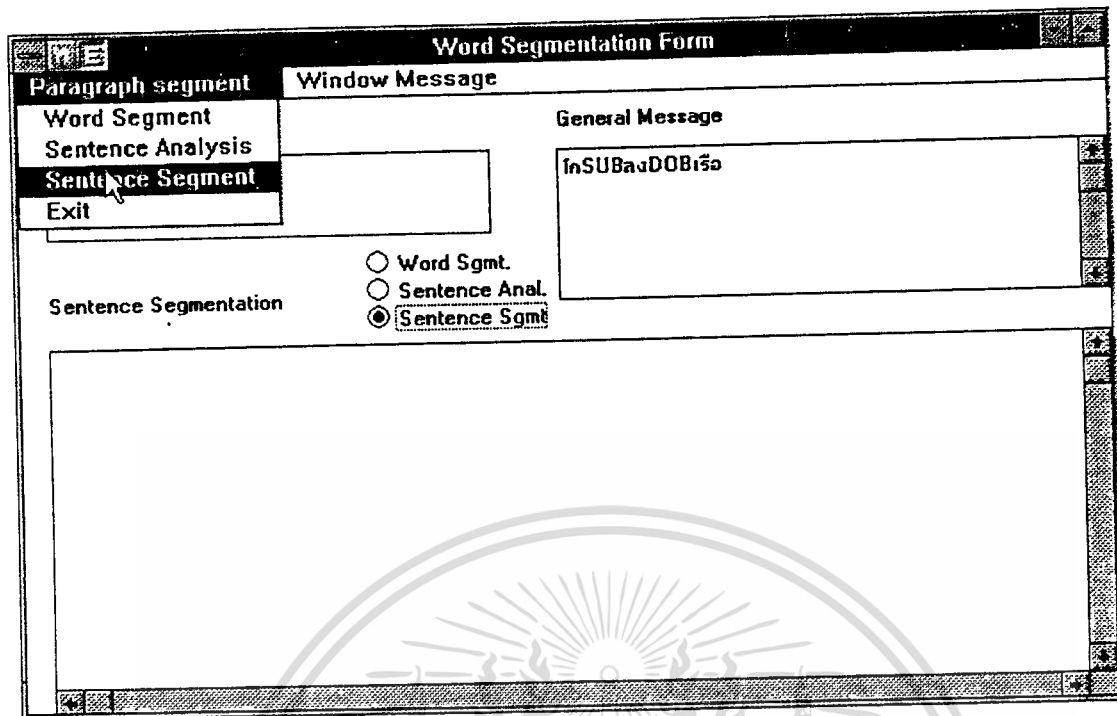
ขั้นตอนที่ 1 ทำการเปิดหน้าต่างแสดงผลโดยใช้เมาส์ชี้ไปที่ปุ่ม Sentence Sgmt ดังแสดงในรูปที่ 5.14

ขั้นตอนที่ 2 ใช้คำสั่งเริ่มต้นโปรแกรมโดยใช้เมาส์ชี้ไปที่เมนู Sentence Segment ดังแสดงในรูปที่ 5.15

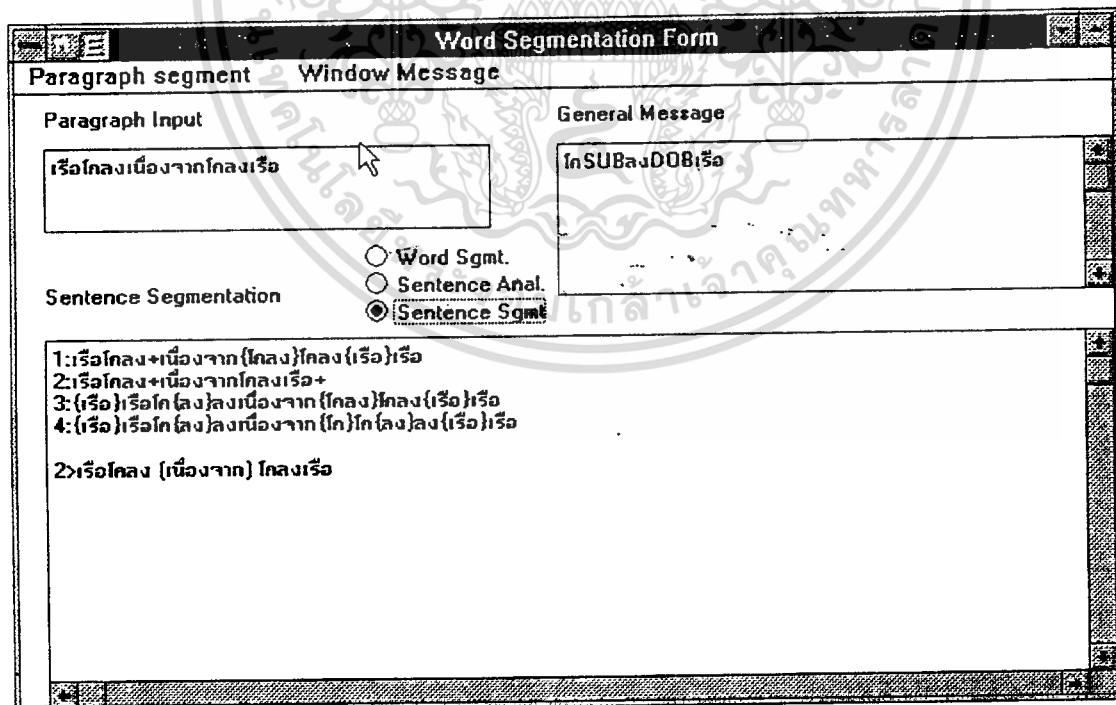
ขั้นตอนที่ 3 ตรวจสอบผลลัพธ์ที่หน้าต่าง Sentence Segmentation ดังแสดงในรูปที่ 5.16



รูปที่ 5.14 แสดงการเปิดหน้าต่างแสดงผลของซอฟต์แวร์ระบบส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า



รูปที่ 5.15 แสดงการใช้งานซอฟต์แวร์ระบบของส่วนวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า



รูปที่ 5.16 แสดงผลลัพธ์ที่ได้จากซอฟต์แวร์ระบบของส่วนแยกประโยคภาษาไทยจากย่อหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.5.5 การใช้งานซอฟต์แวร์ระบบแยกประโยคจากย่อหน้าในส่วนอื่น

นอกเหนือจากการทำงานปกติที่ได้กล่าวมาแล้วในหัวข้อที่ 5.5.1-5.5.4 ซอฟต์แวร์ของระบบแยกประโยคจากย่อหน้ายังมีหน้าต่างตรวจสอบข้อผิดพลาด ในกรณีที่มีการป้อนข้อมูลผิด โดยจะแสดงไว้ที่ หน้าต่างแสดงข้อมูลที่ผิดพลาด (Error Message) ดังแสดงในรูปที่ 5.17

The screenshot shows a window titled "Word Segmentation Form". It has two main sections: "Paragraph segment" and "Window Message".

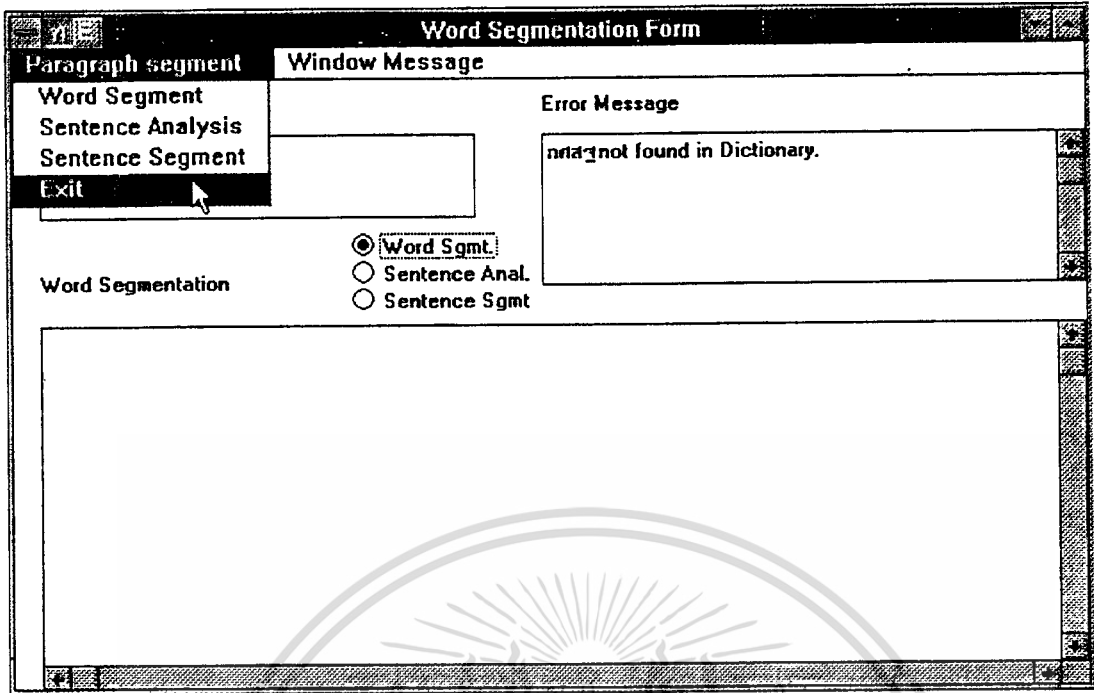
- Paragraph segment:** Contains a "Paragraph Input" field with the Thai text "กตสว" and a "Word Segmentation" section with three radio buttons: "Word Sgmt." (selected), "Sentence Anal.", and "Sentence Sgmt."
- Window Message:** Contains an "Error Message" field displaying the text "กตสวnot found in Dictionary."

The background of the window features a large, faint watermark of the Thai national emblem (Chakrasimukhachulalongkornrajavidyalaya).

รูปที่ 5.17 แสดงข้อผิดพลาดที่เกิดจากการป้อนข้อมูลผิดพลาด

5.5.6 การออกจากระบบแยกประโยคภาษาไทยจากย่อหน้า

ในการยกเลิกการใช้งานระบบแยกประโยคภาษาไทยจากย่อหน้า ผู้ใช้ต้องใช้เมาส์ชี้ไปที่คำสั่ง Exit แล้วกดปุ่มบนเมาส์ ระบบแยกประโยคภาษาไทยจากย่อหน้าจะยกเลิกการทำงานทันทีดังแสดงในรูปที่ 5.18



รูปที่ 5.18 แสดงการออกจากระบบแยกประโยคภาษาไทยจากย่อหน้า

บทที่ 6

บทสรุป

6.1 บทนำ

การวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้ามีขั้นตอนสำคัญสามส่วนที่ทำงานต่อเนื่องสัมพันธ์กัน คือ

ส่วนที่หนึ่ง ส่วนการวิเคราะห์ในระดับคำ

ส่วนที่สอง ส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทย

ส่วนที่สาม ส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

โดยส่วนการวิเคราะห์ในระดับคำจะทำการวิเคราะห์ข้อความย่อหน้าต้นแบบ ด้วยวิธีการตรวจสอบข้อความอินพุท แล้วทำการแบ่งคำออกด้วยวิธีการแบ่งคำแบบ “การแบ่งคำที่เป็นไปได้ (Possible Word Matching)”

ส่วนการวิเคราะห์ในระดับโครงสร้างของประโยคภาษาไทยจะทำการตรวจสอบความถูกต้องทางไวยากรณ์ของประโยคภาษาไทยในย่อหน้า โดยการตรวจหาคำหลักในประโยค คือหน่วยคำกริยาเป็นลำดับแรก เมื่อพบแล้วก็วิเคราะห์หน่วยทางด้านซ้ายมือ ด้วยกฎไวยากรณ์นามวลี จำนวน 4 กฎย่อย จากนั้นวิเคราะห์หน่วยคำทางขวามือด้วยกฎไวยากรณ์ กริยาวลี กลุ่มที่ 1 จำนวน 8 กฎย่อย ซึ่งเมื่อผสมกันแล้วจะได้กฎไวยากรณ์วิเคราะห์ประโยคภาษาไทยได้ 32 กฎ และเมื่อรวมกับกฎไวยากรณ์กริยาวลีกลุ่มที่ 2 อีก 1 กฎจะได้กฎไวยากรณ์สำหรับการวิเคราะห์ประโยคภาษาไทย 33 กฎ ดังแสดงในตารางที่ 6.1

ส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้าจะนำผลที่วิเคราะห์ได้จากส่วนการวิเคราะห์ในระดับโครงสร้างของประโยคภาษาไทย มาวิเคราะห์หาว่าในเซตคำตอบใดมีหน่วยคำที่ไม่อยู่ในประโยคที่ได้วิเคราะห์ไว้แล้ว หากมีก็นำหน่วยคำนั้นมาตรวจสอบหาหน้าที่หลักของคำ หากเป็นคำสันธานจะให้เป็นการเชื่อมประโยคที่อยู่ใกล้ที่สุด หลังจากนั้นจะทำการวิเคราะห์หาหน่วยคำที่อยู่ในประโยคที่อยู่ติดกัน โดยจะเป็นหน่วยคำสุดท้ายของประโยคหนึ่งและเป็นหน่วยคำที่หนึ่งของประโยคถัดไปเมื่อพบ แล้วจะทำการวิเคราะห์ย้อนกลับ (backtracking) เพื่อวิเคราะห์แยกหน่วยคำนั้นให้อยู่ในประโยคเพียงประโยคเดียว เมื่อกระทำทุกขั้นตอนแล้ววรรคและทุก “หน่วยคำ” ถูกใช้เป็นส่วนหนึ่งของประโยค หรือเป็นคำเชื่อมประโยคก็จะแยกทุกประโยคและคำเชื่อมประโยคออกจากกันได้ ผลลัพธ์สุดท้ายจะได้เซตของกลุ่มประโยคที่อยู่ในย่อหน้าโดยแต่ละประโยคจะจัดเรียงอยู่แยกกัน

ซึ่งการทำงานทั้งสามส่วนจะทำงานร่วมกับระบบพจนานุกรมคำศัพท์ ที่มีข้อมูลส่วนของการวิเคราะห์ในระดับหน่วยคำ ได้แก่ คำศัพท์ผิว และข้อมูลส่วนของการวิเคราะห์ในระดับโครงสร้างของประโยคภาษาไทย ได้แก่ ประเภทของคำ ประเภทย่อยของคำ รูปแบบของคำกริยา ชนิดของคำ

กฎไวยากรณ์ นามวลี		กฎไวยากรณ์ กริยาวลี	
กฎที่	กฎไวยากรณ์ นามวลี	กฎที่	กฎไวยากรณ์ กริยาวลีกลุ่มที่ 1
R1	NP → PRON (AKO = 111)	R5	VP → VERB (VP6) + N + PREP + N
R2	NP → PRON	R6	VP → VERB (VP8) + N + N
R3	NP → N (AKO = 111)	R7	VP → VERB (VP7) + N + ADV
R4	NP → N	R8	VP → VERB (VP7) + N
		R9	VP → VERB (VP5) + PREP + N
		R10	VP → VERB (VP3) + ADV
		R11	VP → VERB (VP4) + AUX
		R12	VP → VERB (VP1)
		กฎที่	กฎไวยากรณ์ กริยาวลีกลุ่มที่ 2
		R13	VP → VERB (VP2) + N

ตารางที่ 6.1 แสดงกฎไวยากรณ์ใช้วิเคราะห์ประโยคภาษาไทย

งานวิจัยนี้ได้นำผลการวิจัยมาสร้างเป็นซอฟต์แวร์เพื่อทำการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้าด้วยคอมพิวเตอร์ เพื่อทำการทดลอง และเพื่อเป็นต้นแบบสำหรับงานวิจัยด้านภาษาศาสตร์คอมพิวเตอร์ของภาษาไทยโดยมีขอบเขตพื้นฐานที่พัฒนาแล้วดังนี้

1) ประสิทธิภาพของส่วนการวิเคราะห์ในระดับคำ

-สามารถทำการวิเคราะห์ได้ครั้งละ 1 ย่อหน้า

-ข้อความที่จะนำมาวิเคราะห์จะต้องเป็นข้อความภาษาไทย

-หน่วยคำที่นำมาสร้างเป็นประโยคจะต้องเป็นหน่วยคำที่มีในพจนานุกรม

-มีการทำงานร่วมกับระบบพจนานุกรมคำศัพท์ที่สามารถเปลี่ยนแปลงแก้ไขเพิ่มคำ

ศัพท์ได้ ซึ่งในงานวิจัยนี้ได้วิเคราะห์หน่วยคำภาษาไทยขึ้นมาจำนวน 200 หน่วยคำ โดยรายละเอียด

ของหน่วยคำได้แสดงไว้เป็นตารางในภาคผนวกการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ประสิทธิภาพของส่วนการวิเคราะห์ในระดับโครงสร้างประโยคภาษาไทย

-สามารถวิเคราะห์ในระดับโครงสร้างประโยคเดียวในภาษาไทยได้

-สามารถวิเคราะห์โครงสร้างทางวากยสัมพันธ์ได้

-มีการทำงานร่วมกับระบบพจนานุกรมคำศัพท์ที่เก็บข้อมูลเกี่ยวกับโครงสร้างของ

ประโยค

3) ประสิทธิภาพของส่วนการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

-สามารถวิเคราะห์แยกประโยคภาษาไทยออกจากกันและตรวจสอบค่าเชื่อมประโยค

ได้

-สามารถทำการวิเคราะห์ย้อนกลับ (backtracking) เพื่อแก้ปัญหาที่ หน่วยคำที่

สามารถทำหน้าที่ในประโยคได้พร้อมกันสองประโยค

6.2 ผลการทดสอบ

ย่อหน้าที่	จำนวนประโยคในย่อหน้า	จำนวนประโยคที่วิเคราะห์ได้ถูกต้อง	จำนวนเซตคำตอบที่ได้จากการแบ่งค่า	จำนวนคำตอบที่วิเคราะห์ได้
1	3	3	8	1
2	6	5	2	2
3	2	2	1	1
4	3	3	2	1
5	2	2	1	1
6	3	3	1	1
7	3	3	2	1
8	2	2	1	1
9	4	4	2	1
10	4	3	1	0
11	2	2	1	1
รวม	34	32	22	11

ตารางที่ 6.2 แสดงผลการทดสอบระบบวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการทดลองการแยกประโยคภาษาไทยจากย่อหน้าโดยงานวิจัยนี้ได้กำหนดหน่วยคำตัว
ไว้ในฐานข้อมูลคำศัพท์จำนวน 200 หน่วยคำ และกฎไวยากรณ์วิเคราะห์โครงสร้างประโยคภาษาไทย
จำนวน 33 กฎ จากนั้นได้ให้อาสาสมัครซึ่งเป็นนักศึกษาคณะวิศวกรรมศาสตร์จำนวน 3 คนทำการ
สร้างประโยคและนำประโยคมาเรียงเป็นย่อหน้าได้จำนวน 11 ย่อหน้า ซึ่งประกอบด้วยประโยคทั้ง
หมด 34 ประโยค แล้วนำมาทดสอบกับระบบแยกประโยคภาษาไทยจากย่อหน้า จากนั้นทำการตรวจ
สอบผลลัพธ์ซึ่งทำการวิเคราะห์ด้วยมือ ซึ่งผลการทดลองมีดังตารางที่ 6.2

จากการทดลองพบว่ามีประโยคต้นแบบที่ใช้วิเคราะห์ทั้งหมด 34 ประโยค

ระบบวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้าสามารถวิเคราะห์ประโยคได้ถูกต้องจำนวน
32 ประโยค คิดเป็น 94.12 %

ระบบวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้าไม่สามารถวิเคราะห์ประโยคได้ถูกต้อง
จำนวน 2 ประโยค คิดเป็น 94.12 %

โดยประโยคที่ไม่สามารถวิเคราะห์ได้เนื่องจากกฎไวยากรณ์ไม่ครอบคลุม 1 ประโยค คิดเป็น
2.94 %

และประโยคที่ไม่สามารถวิเคราะห์ได้เนื่องจากกฎไวยากรณ์แต่ไม่ถูกความหมาย 1
ประโยค คิดเป็น 2.94 %

จำนวนย่อหน้าที่ใช้ทดลองทั้งหมด 11 ย่อหน้า

ระบบวิเคราะห์แยกประโยคจากย่อหน้าสามารถวิเคราะห์ได้ถูกต้องจำนวน 9 ย่อหน้า คิดเป็น
81.81 %

โดยข้อมูลที่ใช้ทดลองมีดังต่อไปนี้

ย่อหน้าที่ 1 เรือโคลงเพราะว่าโคลงเรือและเขาโคลงเรือ

ย่อหน้าที่ 2 ข้าพเจ้าไปกัตตาคารกับมานะ ข้าพเจ้าสั่งปลาติบมานะสั่งหูลาม ข้าพเจ้าจ่ายเงินและ
รับเงินทอน แล้วข้าพเจ้าให้ทิป

ย่อหน้าที่ 3 นักเขียนแต่งวรรณกรรมอย่างรวดเร็ว เพราะว่ามันศึกษาสนใจวรรณกรรมมาก

ย่อหน้าที่ 4 ชาวญี่ปุ่นนิยมไปที่ต่างประเทศ ชาวญี่ปุ่นชอบกรุงเทพเนื่องจากชาวไทยเป็นคนดี

ย่อหน้าที่ 5 นักศึกษาชอบคอมพิวเตอร์ เขาสั่งซอฟต์แวร์จากต่างประเทศ
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าที่ 6 ชาวไทยผลิตข้าวเจ้ามาก ต่างประเทศสั่งข้าวเจ้าจากชาวไทย ชาวไทยมีเงินมาก

ย่อหน้าที่ 7 มานะเล่นดนตรีในโรงละคร มานะใช้เปียโนตามปกติ ผู้ชมนิยมดนตรีของมานะ

ย่อหน้าที่ 8 ปลาติบเป็นของชาวญี่ปุ่นและชาวไทยมีข้าวเหนียว

ย่อหน้าที่ 9 สมชายเป็นนักศึกษา เขาเป็นคนดีเนื่องจากเขานิยมดนตรี และเนื่องจากเขามีความสามารถดี

ย่อหน้าที่ 10 วันนี้เป็นวันอาทิตย์ ซึ่งข้าพเจ้าอยู่ที่บ้าน และข้าพเจ้าใช้คอมพิวเตอร์ เนื่องจากคอมพิวเตอร์มีประโยชน์กับข้าพเจ้า

ย่อหน้าที่ 11 เกิดอุทกภัยเนื่องจากชาวไทยทำลายป่า

6.3 การนำผลงานวิจัยไปประยุกต์ใช้กับงานอื่นๆ

1. การนำไปใช้กับการแปลภาษาด้วยคอมพิวเตอร์ (Machine Translation) จะสามารถใช้งานได้
ในส่วนของการเตรียมข้อมูลสำหรับการวิเคราะห์ภาษาต้นแบบที่เป็นภาษาไทย
2. การนำไปใช้ร่วมกับโปรแกรมประมวลผลคำ (Word Processing) เพื่อทำการวิเคราะห์
เอกสารว่า เขียนถูกต้องตามหลักไวยากรณ์หรือไม่

6.4 แนวทางในการพัฒนาต่อของระบบ

1. การเพิ่มประสิทธิภาพให้ระบบ
โดยการเพิ่มกฎไวยากรณ์และ คำศัพท์ให้แก่ระบบ
2. การวิเคราะห์เรื่องราว
ปรับปรุงให้สามารถวิเคราะห์ข้อความครั้งละหลายย่อหน้า เพื่อให้สามารถวิเคราะห์เรื่องราวที่
เขียนตามหลักของการเขียนข้อความภาษาไทยแบบเรียงความได้

เอกสารอ้างอิง

- [1] นวลทิพย์ ตันติเศวตรรัตน์ "โครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์." การประชุมวิชาการครั้งที่ 1 โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์ ปีงบประมาณ 2531 ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, เล่ม 1, 2532
- [2] รัตติกกร วรากุลศิริพันธ์ "ระบบการสร้างภาษาเป้าหมายเพื่อการแปลภาษาด้วยคอมพิวเตอร์." การประชุมวิชาการครั้งที่ 1 โครงการวิจัยและพัฒนาอิเล็กทรอนิกส์และคอมพิวเตอร์ ปีงบประมาณ 2531 ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ, เล่ม 1., 2532
- [3] นวลทิพย์ ตันติเศวตรรัตน์ "โครงการพัฒนาพจนานุกรมเพื่อการแปลภาษาด้วยคอมพิวเตอร์" การแปลภาษาด้วยคอมพิวเตอร์ โครงการพัฒนาระบบเครื่องแปลภาษาสำหรับภาษาในเอเชีย ศูนย์เทคโนโลยีแห่งชาติ
- [4] Machine Translation System Laboratory CICC 1995, Technical Report of Thai Lexicon, Center of the International Cooperation for Computerization, 1995
- [5] รัตติกกร วรากุลศิริพันธ์, อรัญญา ปรีชาไว , " การวิเคราะห์โครงสร้างประโยคภาษาอังกฤษเพื่อการแปลภาษาด้วยคอมพิวเตอร์", วารสารคอมพิวเตอร์แห่งประเทศไทยในพระบรมราชูปถัมภ์, ปีที่ 19 ฉบับที่ 97/พฤษภาคม-มิถุนายน, หน้า 14-21, 2535
- [6] วันทนีย์ พันธชาติ, วรณัฐ เกิดสินธุ์ชัย, ทศนี้อยู่ เจริญพร, และ สุดา รังกุพันธ์ , "ระบบการวิเคราะห์ภาษาไทยเพื่อการแปลภาษาด้วยคอมพิวเตอร์.", วารสารฉบับพิเศษ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์ และคอมพิวเตอร์แห่งชาติ, หน้า 27-33, 2534
- [7] สมศักดิ์ จันวัน, "ระบบวิเคราะห์โครงสร้างภาษาไทยด้วยคอมพิวเตอร์", วิทยานิพนธ์มหาบัณฑิต บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, ประจำปีการศึกษา 2534
- [8] วิรัช ศรีเลิศล้ำวานิช, "การตัดคำในระบบแปลภาษา", การแปลภาษาด้วยคอมพิวเตอร์, ศูนย์เทคโนโลยีอิเล็กทรอนิกส์แห่งชาติ, หน้า50-56, 2536
- [9] กำชัย ทองหล่อ, "หลักภาษาไทย", สำนักพิมพ์บำรุงสารัน, 2530
- [10] อรัญญา ปรีชาไว, "ระบบการวิเคราะห์ประโยคภาษาอังกฤษไปสู่ภาษากลางเพื่อการแปลภาษาด้วยคอมพิวเตอร์", วิทยานิพนธ์มหาบัณฑิต บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, ประจำปีการศึกษา 2537

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [11] Wu Wang and Jianhua Chen, "Learning by Discovering Problem Solving Heuristics Through Experience" , IEEE Transactions on Knowledge and Data Engineering, Vol.3, No.4, p. 415-419, December 1991.
- [12] Marion Mast, Franz Kummert, Ute Ehrlich, Gernot A. Fink, Thomas Kuhn, Heinrich Niemann, and Gerhard Sagerer, "A Speech Understanding and Dialog System with a Homogeneous Linguistic Knowledge Base", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 16, No. 2, p. 179-194, February 1994.
- [13] Angela B. Shiflet, "Discrete Mathematics for Computer Science", West Publishing Company, 1987, 450 Pages.
- [14] วิไลวรรณ ขนิษฐานันท์, "ภาษาและภาษาศาสตร์", สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์, 2533, 195 หน้า.
- [15] Risto Miikkulainen, "Subsymbolic Natural Language Processing", A Bradfordbook, The MIT Press , 1993, 391 Pages.





เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

tkey	cat	scat	ako	vp	tmap	cm	cm
ก็	N	NPRP	111				
กงสุล	PREP	RPREP					
กรุงเทพ	N	NCMN	131				
กลอง	PRON	PPRS	111				
กลุ่ม	N	NCMN	127				
กับ	VERB	VACT	2212	6	SUB=AGT.DOB=OBJ.CMP3=TAR		ให้
การลงทุน	N	NCMN	125				
ก็ฟ้า	AUX	XVAF					
เก่ง	CONJ	CSBR					
เกิด	NEG	NEG					
เกิด	N						
เกิด	CONJ	CSBR					
เกิน	VERB	VACT		8	SUB=AGT.DOB=OBJ.IOB=BEN		
ขลุ่ย	N	NCMN	126				
ของ	ADV	ADVN	295				
ของ	VERB	VATT	295	1	SUB=ATT		
ของ	VERB	VI					
ข้าพเจ้า	N	NCMN	284				
ข้าพเจ้า	PREP	RPRE	137				
ข้าพเจ้านี้	N		283				
ขึ้น	VERB	VACT	239	5	SUB=OBJ.CMP2=PAR		กับ
ขึ้น	N	NCMN	283				
ขึ้น	N	NCMN	283				
ขึ้น	ADV	ADVN	286				
เขา	VERB	VATT	286	1	SUB=ATT		
คนดี	PRON	RPEL					
ครึ่ง	PREP	RPRE	272				
ความ	VERB	VACT	2231	7	SUB=INS.DOB=OBJ		
ความสนใจ	PREP	RVBP					
ความสามารถ	VERB	VACT	2314	6	SUB=AGT.DOB=OBJ.CMP3=BEN		แก่
คอมพิวเตอร์	N	NCMN					
โค	N	NPRP	112				
โคลง	VERB	VACT	2212	6	SUB=AGT.DOB=OBJ.CMP3=BEN		แก่
โคลง	VERB	VACT	2212	8	SUB=AGT.DOB=OBJ.IOB=TAR		
โคลง	VERB	VACT	2212	7	SUB=AGT.DOB=OBJ		
งานเลี้ยง	CONJ	.					
เงิน	N	NCMN	127				
เงินทอง	N	NPRP	111				
จังหวะ	VERB	VACT	227	7	SUB=AGT.DOB=OBJ		
จาก	N	NCMN	127				
จาก	CONJ	CSBR					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

tkey	cat	scat	ako	vp	tmap	cm	cm
จากนั้น	N	NPRP					
จ่าย	PREP	RPRE					
จ่าย	VERB	VACT	127	6	SUB=AGT,DOB=OBJ		
จ่าย	CONJ	CONJ					
จ่าย	PRON	PPRS	111				
จำใจ	N	NPRP	265				
จำนวน	N	NCMN	111				
หลวม	VERB	VACT	226	7	SUB=AGT,DOB=OBJ		
ชม	N	NCMN	234				
ชอบ	N	NCMN	111				
ชอบ	N	NCMN	131				
ช้า	N	NALN					
ช่างไม้	N	NCMN					
ขาด	PREP	RPRE					
ขาย	N	NCMN					
ชาวญี่ปุ่น	VERB	VATT		7			
ชาวไทย	VERB	VACT		1			
ให้	VERB	VACT		1			
ให้	N	NCMN					
ซอฟต์แวร์	CONJ	CSBR					
ซึ่ง	PRON	PREL					
ซึ่ง	CONJ	CONJ					
ญี่ปุ่น	VERB	VSTA		5			
ฐานะ	CLAS	CNIT					
ดนตรี	N	NLBL					
ดี	VERB	VSTA		6			
ดี	VERB	VSTA		7			
ดี	VERB	VACT		6			
โดย	N	NCMN					
โดย	N	NCMN					
โดยที่	VERB	VSTA		7			
โดยธรรมดา	PREP	RVBP					
โดยเร็ว	VERB	VACT		7			
ได้	N	NCMN	111				
ได้	PRON	PREL	111				
ตรง	N	NCMN					
ตัว	PREP	RPRE					
ต่างประเทศ	VERB	VACT		7			
ตามปกติ	VERB	VSTA		3			
แต่	N	NCMN					
แต่	VERB	VACT		4			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

tkey	cat	scat	ako	vp	tmap	cm	cm
แต่ง	VERB	VACT		3			
แต่ง	VERB	VACT		1			
ถูก	AUX	XVAE					
ถูก	VERB	VSTA		5			
ทะเล	VERB	VACT		7			
ทั้งคู่	VERB	VSTA		7			
ทั้งหมด	VERB	VSTA		1			
ทั้งสอง	PRON	PPRS	111				
ทั้งหมด	N	NCMN	111				
ทั้งหลาย	N	NCMN	111				
ทั่วไป	N	NCMN					
ทางด้าน	PREP	RPRE					
ทำลาย	N	NPRP					
ทำลาย	N	NPRP					
ที่	CONJ	CCRG					
ที่	N	NCMN	111				
ที่	N	NCMN	111				
ที่	N	NCMN					
ที่เท่าไร	N	NCMN	111				
ที่สุดท้าย	N	NCMN					
ที่หนึ่ง	VERB	VATT					
ธรรมดา	CONJ	CSBR					
เจอ	CONJ	CCRG					
นมสด	PRON	PPRS	111				
นักเรียน	N	NCMN					
นักธุรกิจ	N	NCMN					
นักศึกษา	N	NCMN					
นักอ่าน	N	NCMN					
นิยม	N	NCMN	111				
นิยม	N	NCMN					
นิยม	N	NCMN					
นี่	N	NCMN					
เนื่องจาก	ADV	ADVN					
โนน	ADV	ADVN					
โน	N	NCMN					
ในปัจจุบัน	VERB	VACT		6			
บ้าง	VERB	VACT		5			
บ้าง	VERB	VACT		7			
ปกติ	VERB	VACT		3			
ประโยชน์	VERB	VACT		1			
ปลาดี	N	NCMN					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

tkey	cat	scat	ako	vp	tmap	cm	cm
ไป	PREP	RPRE					
ไปที่แล้ว	PREP	RPRE					
เป็น	N	NCMN					
เป็น	VERB	VSTA	229	5	SUB=AGT.DOB=OBJ	ทดลอง	
เปียโน	VERB	VSTA	229	7	SUB=AGT.DOB=OBJ		
ไป	VERB	VACT	226	6	SUB=AGT.DOB=OBJ.IOB=TAR		ให้
ไป	VERB	VACT	226	8	SUB=AGT.DOB=OBJ.IOB=TAR		
ไป	VERB	VACT	226	7	SUB=AGT.DOB=OBJ		
ไปที่	VERB	VSTA	229	6	SUB=DUM1.OBJ=DUM2		ทาง
ผม	VERB	VSTA	229	7	SUB=DUM1.OBJ=DUM2		
ผลิต	VERB	VSTA	229	2	DOB=OBJ		
ผลิต	CONJ	CSBR					
ผู้	N	NPRP					
ผู้	N	NPRP	111				
ผู้ชม	N	NCNM					
ผู้ใต้	N	NCNM					
ผู้มีชื่อ	N	NCNM					
ฝูง	N	NONM					
พวกเขา	N	NONM					
พอ	N	NONM					
เพราะว่า	N	NALN					
เพลง	N	NALN					
กัฏตาการ	N	NALN					
มาก	N	NAJN					
มาก	N	NAJN					
มานะ	N	NTTL	111				
มี	N	NCLT					
มี	N	NCLT					
มี	PRON	PPRS-	111				
แม้ว่า	PRON	PPRS	111				
ไม่	PRON	PDMN					
ยากจน	PRON	PDMN					
รส	PRON	PDEF					
รสชาติ	PRON	PDEF					
ร้อย	PRON	PDEF					
ร้อยโท	PRON	PDEF					
รับ	PRON	PIND					
รับ	PRON	PIND					
รับ	PRON	PIND					
เร็ว	PRON	PREL					
เรียน	AUX	XVAE					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

tkey	cat	scat	ako	vp	tmap	cm	cm
เร็ว	ADV	ADV N					
โรงละคร	ADV	ADV N					
โรงอาหาร	ADV	ADV N					
ลง	ADV	ADV N					
ลง	ADV	ADV N					
ลุก	ADV	ADV N					
ลุก	ADV	ADV P					
ลุก	ADV	ADV P					
เล่น	ADV	ADV S					
เล่น	ADV	ADV S					
แล้ว	ADV	ADV S					
แล้ว	ADV	ADV S					
และ	ADV	ADV S					
วรรณกรรม	CONJ	CCRG					
วันนี้	VERB	VACT	228	6	SUB=AGT, DOB=TAR		กับ
วันอาทิตย์	PREP	RPRE					
วันอาทิตย์	VERB	VACT	228	5	SUB=OBJ, CMP2=TAR	ที่	
วิศวกรรม	VERB	VACT	228	7	SUB=AGT, DOB=TAR		
สนใจ	VERB	VACT	2314	8	SUB=AGT, DOB=OBJ, CMP3=SOR		จาก
สม	VERB	VACT	2314	7	SUB=AGT, DOB=OBJ, CMP3=SOR		จาก
สมชาย	VERB	VACT	2212	6	SUB=AGT, DOB=OBJ		
สม่ำเสมอ	VERB	VACT	2212	7	SUB=AGT, DOB=OBJ		
ส่วนมาก	VERB	VACT		6			
สวนสนุก	VERB	VACT		5			
สวนใหญ่	VERB	VACT		7			
สูง	VERB	VACT		6			
สูง	VERB	VACT		7			
สูง	VERB	VACT		3			
สูง	N	NCMN					
แสดง	N	NCMN					
แสดง	N	NCMN	111				
แสดง	N	NCMN					
แสดง	N	NCMN					
แสดง	N	NCMN					
หนึ่ง	N	NCMN					
หรือ	N	ADV N					
หู	VERB	VACT		6			
หูดลามู	VERB	VACT		7			
เหล่านี้	N	NCMN					
ให้	VERB	VACT		6			
ให้	VERB	VACT		7			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

tkey	cat	scat	ako	vp	tmap	cm	cm
ให้	N	NCMN					
ให้	VERB	VACT		6			
ให้	VERB	VACT		7			
อย่างรวดเร็ว	VERB	VACT		1			
อยู่	VERB	VACT		6			
อยู่แล้ว	VERB	VACT		7			
ออกมา	VERB	VACT		3			
อัน	VERB	VACT		4			
อาหาร	VERB	VSTA		2			
อิเล็กทรอนิกส์	N	NCMN					
คหภักย์	N	NCMN					





ภาคผนวก ข
ข้อมูลผลการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
เรือโคลงเพราะว่าโคลงเรือและเขาโคลงเรือ	โกSUBลงDOBเรือ
<input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt.	
Sentence Analysis	
เรือSUBโคลง เขาAGTโคลงOBJเรือ เรือSUBโคลง โกSUBลงOBJเรือ เรือSUBโคลง โกSUBลงOBJเรือ เขาAGTโคลงOBJเรือ เรือSUBโคลง โกSUBลงOBJเรือ โกSUBลงOBJเรือ	

รูปที่ ข-2 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 1

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
เรือโคลงเพราะว่าโคลงเรือและเขาโคลงเรือ	โกSUBลงDOBเรือ
<input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt.	
Sentence Segmentation	
1:เรือโคลง+เพราะว่า(โคลง)โคลง(เรือ)เรือและเขาโคลงเรือ+ 2:เรือโคลง+เพราะว่า(โคลง)โคลง(เรือ)เรือและ(เขา)เขาโคลงเรือ+ 3:เรือโคลง+เพราะว่าโคลงเรือ+และเขาโคลงเรือ+ 4:เรือโคลง+เพราะว่าโคลงเรือ+และ(เขา)เขาโคลงเรือ+ 5:(เรือ)เรือ(โก)ลง(ง)ลงเพราะว่า(โคลง)โคลง(เรือ)เรือและ(เขา)เขา(โก)โคลง(ง)ลง(เรือ)เรือ 6:(เรือ)เรือ(โก)ลง(ง)ลงเพราะว่า(โคลง)โคลง(เรือ)เรือและ(เขา)เขา(โก)โคลง(ง)ลง(เรือ)เรือ 7:(เรือ)เรือ(โก)ลง(ง)ลงเพราะว่า(โก)โคลง(ง)ลง(เรือ)เรือและ(เขา)เขา(โคลง)โคลง(เรือ)เรือ 8:(เรือ)เรือ(โก)ลง(ง)ลงเพราะว่า(โก)โคลง(ง)ลง(เรือ)เรือและ(เขา)เขา(โก)โคลง(ง)ลง(เรือ)เรือ 3>เรือโคลง (เพราะว่า) โคลงเรือ (และ) เขาโคลงเรือ	

รูปที่ ข-3 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าอินพุทที่ 2

ข้าพเจ้าไปภัตตาคารกับมานะข้าพเจ้าสั่งปลาดิบมานะสั่งหูลามข้าพเจ้าจ่ายเงินและ
ข้าพเจ้ารับเงินทอนแล้วข้าพเจ้าให้ทิป

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้ 2 คำตอบ

เซตคำตอบที่ 1 สามารถแยกประโยคได้ครบ

ประโยคที่ 1	ข้าพเจ้า ไป ภัตตาคาร กับ มานะ	ตามกฎที่ R5
ประโยคที่ 2	ข้าพเจ้า สั่ง ปลาดิบ	ตามกฎที่ R8
ประโยคที่ 3	มานะ สั่ง หูลาม	ตามกฎที่ R8
ประโยคที่ 4	ข้าพเจ้า จ่าย เงิน	ตามกฎที่ R8
ประโยคที่ 5	ข้าพเจ้า รับ เงินทอน	ตามกฎที่ R8
ประโยคที่ 6	ข้าพเจ้า ให้ ทิป	ตามกฎที่ R8

คำเชื่อมประโยคที่ 4 และ 5 และ

คำเชื่อมประโยคที่ 5 และ 6 แล้ว

เซตคำตอบที่ 2 สามารถแยกประโยคได้ครบ

ประโยคที่ 1	ข้าพเจ้า ไป ภัตตาคาร กับ มานะ	ตามกฎที่ R5
ประโยคที่ 2	ข้าพเจ้า สั่ง ปลาดิบ	ตามกฎที่ R8
ประโยคที่ 3	มานะ สั่ง หู ลาม	ตามกฎที่ R6
ประโยคที่ 4	ข้าพเจ้า จ่าย เงิน	ตามกฎที่ R8
ประโยคที่ 5	ข้าพเจ้า รับ เงินทอน	ตามกฎที่ R8
ประโยคที่ 6	ข้าพเจ้า ให้ ทิป	ตามกฎที่ R8

คำเชื่อมประโยคที่ 4 และ 5 และ

คำเชื่อมประโยคที่ 5 และ 6 แล้ว

ซึ่งเป็นเซตคำตอบที่ 1, 2 จากเซตผลลัพธ์ที่ได้จากการแบ่งคำ 2 เซต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ข-4 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 2

รูปที่ ข-5 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 2 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input

General Message

Paragraph Input: ข้าพเจ้าไปกักตุนอาหารกับมานะข้าพเจ้าสั่งปลาดี
มานะสั่งชูลานข้าพเจ้าจ่ายเงินและข้าพเจ้ารับ
เงินทอนแล้วข้าพเจ้าให้ทิป

General Message: ข้าพเจ้าSUB1ห้DOBทิป

Sentence Analysis

Word Sgmt.
 Sentence Anal.
 Sentence Sgmt

Sentence Analysis: ข้าพเจ้าAGTไปOBJกักตุนอาหารกับมานะ ข้าพเจ้าAGTสั่งOBJปลาดีOBJมานะ มานะAGTสั่งOBJชูลาน ข้าพเจ้าA
ข้าพเจ้าAGTไปOBJกักตุนอาหารกับมานะ ข้าพเจ้าAGTสั่งOBJปลาดีOBJมานะ มานะAGTสั่งOBJชูลาน OBJชูลาน ข้าพ

รูปที่ ข-6 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 2

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input

General Message

Paragraph Input: ข้าพเจ้าไปกักตุนอาหารกับมานะข้าพเจ้าสั่งปลาดี
มานะสั่งชูลานข้าพเจ้าจ่ายเงินและข้าพเจ้ารับ
เงินทอนแล้วข้าพเจ้าให้ทิป

General Message: ข้าพเจ้าSUB1ห้DOBทิป

Sentence Analysis

Word Sgmt.
 Sentence Anal.
 Sentence Sgmt

Sentence Analysis: :AGTสั่งOBJชูลาน ข้าพเจ้าAGTจ่ายOBJเงิน ข้าพเจ้าAGTรับOBJเงินทอน ข้าพเจ้าAGT1ห้OBJทิป
:AGTสั่งOBJชูลาน ข้าพเจ้าAGTจ่ายOBJเงิน ข้าพเจ้าAGTรับOBJเงินทอน ข้าพเจ้าAGT1ห้OBJทิป

รูปที่ ข-7 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 2 (ต่อ)

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
<p>ข้าพเจ้าไปทัศนศึกษากับมานะข้าพเจ้าสิ่งปลาติบ มานะสิ่งชูลลามข้าพเจ้าจ่ายเงินและข้าพเจ้ารับ เงินทอนแล้วข้าพเจ้าให้ทิป</p>	<p>ข้าพเจ้าSUBให้DOBทิป</p>
Sentence Segmentation	<input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt.
<p>1>ข้าพเจ้าไปทัศนศึกษากับมานะ+ข้าพเจ้าสิ่งปลาติบมานะสิ่งชูลลาม+ข้าพเจ้าจ่ายเงิน+และข้าพเจ้ารับเงินทอน+แล้วข้าพเจ้า 2>ข้าพเจ้าไปทัศนศึกษากับมานะ+ข้าพเจ้าสิ่งปลาติบมานะสิ่งชูลลาม+ข้าพเจ้าจ่ายเงิน+และข้าพเจ้ารับเงินทอน+แล้วข้าพเจ้า</p> <p>1>ข้าพเจ้าไปทัศนศึกษากับมานะ ข้าพเจ้าสิ่งปลาติบ มานะสิ่งชูลลาม ข้าพเจ้าจ่ายเงิน (และ) ข้าพเจ้ารับเงินทอน (แล้ว) ข้า 2>ข้าพเจ้าไปทัศนศึกษากับมานะ ข้าพเจ้าสิ่งปลาติบ มานะสิ่งชูลลาม ข้าพเจ้าจ่ายเงิน (และ) ข้าพเจ้ารับเงินทอน (แล้ว) ข้า</p>	

รูปที่ ๗-8 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 2

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
<p>ข้าพเจ้าไปทัศนศึกษากับมานะข้าพเจ้าสิ่งปลาติบ มานะสิ่งชูลลามข้าพเจ้าจ่ายเงินและข้าพเจ้ารับ เงินทอนแล้วข้าพเจ้าให้ทิป</p>	<p>ข้าพเจ้าSUBให้DOBทิป</p>
Sentence Segmentation	<input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt.
<p>านะ+ข้าพเจ้าสิ่งปลาติบมานะสิ่งชูลลาม+ข้าพเจ้าจ่ายเงิน+และข้าพเจ้ารับเงินทอน+แล้วข้าพเจ้าให้ทิป+ านะ+ข้าพเจ้าสิ่งปลาติบมานะสิ่งชูลลาม+ข้าพเจ้าจ่ายเงิน+และข้าพเจ้ารับเงินทอน+แล้วข้าพเจ้าให้ทิป+ ทกะ ข้าพเจ้าสิ่งปลาติบ มานะสิ่งชูลลาม ข้าพเจ้าจ่ายเงิน (และ) ข้าพเจ้ารับเงินทอน (แล้ว) ข้าพเจ้าให้ทิป ทกะ ข้าพเจ้าสิ่งปลาติบ มานะสิ่งชูลลาม ข้าพเจ้าจ่ายเงิน (และ) ข้าพเจ้ารับเงินทอน (แล้ว) ข้าพเจ้าให้ทิป</p>	

รูปที่ ๗-9 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 2 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าอินพุทที่ 3

นักเขียนแต่งวรรณกรรมอย่างรวดเร็วเพราะว่านักศึกษาสนใจวรรณกรรมมาก

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

ประโยคที่ 1 นักเขียน แต่ง วรรณกรรม อย่างรวดเร็ว ตามกฎที่ R3, R7

ประโยคที่ 2 นักศึกษา สนใจ วรรณกรรม มาก ตามกฎที่ R3, R7

คำเชื่อมประโยคที่ 1 และ 2 เพราะว่า

รูปที่ ข-10 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 3

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: นักเขียนแต่งวรรณกรรมอย่างรวดเร็ว เพราะว่า นักศึกษาสนใจวรรณกรรมมาก

General Message: นักศึกษาSUBสนใจDOBวรรณกรรมมาก

Sentence Analysis:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

Output: นักเขียนAGTแต่งOBJวรรณกรรมอย่างรวดเร็ว นักศึกษาAGTสนใจOBJวรรณกรรมมาก

รูปที่ ข-11 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 3

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: นักเขียนแต่งวรรณกรรมอย่างรวดเร็ว เพราะว่า นักศึกษาสนใจวรรณกรรมมาก

General Message: นักศึกษาSUBสนใจDOBวรรณกรรมมาก

Sentence Segmentation:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

Output:

- 1:นักเขียนแต่งวรรณกรรมอย่างรวดเร็ว+เพราะว่า+นักศึกษานักสนใจวรรณกรรมมาก+
- 1>นักเขียนแต่งวรรณกรรมอย่างรวดเร็ว (เพราะว่า) นักศึกษาสนใจวรรณกรรมมาก

รูปที่ ข-12 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 3

ย่อหน้าอินพุตที่ 4

ชาวญี่ปุ่นนิยมไปที่ต่างประเทศ ชาวญี่ปุ่นชอบกรุงเทพเนื่องจากชาวไทยเป็นคนดี

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

- | | |
|--|-----------------|
| ประโยคที่ 1 ชาวญี่ปุ่น นิยม ไปที่ ต่างประเทศ | ตามกฎที่ R3, R9 |
| ประโยคที่ 2 ชาวญี่ปุ่น ชอบ กรุงเทพ | ตามกฎที่ R3, R8 |
| ประโยคที่ 3 ชาวไทย เป็น คนดี | ตามกฎที่ R3, R8 |

คำเชื่อมประโยคที่ 2 และ 3 เนื่องจาก

ซึ่งเป็นเซตคำตอบที่ 1 จากเซตผลลัพธ์ที่ได้จากการแบ่งคำ 2 เซต

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: ชาวญี่ปุ่นนิยมไปที่ต่างประเทศ ชาวญี่ปุ่นชอบกรุงเทพฯ เนื่องจากชาวไทยเป็นคนดี

General Message: ชาวไทยSUBเป็นDOBคนดี

Sentence Analysis:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

ชาวญี่ปุ่นAGTนิยมOBJต่างประเทศ ชาวญี่ปุ่นAGTชอบOBJกรุงเทพฯ ชาวไทยAGTเป็นOBJคนดี

รูปที่ ข-14 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 4

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: ชาวญี่ปุ่นนิยมไปที่ต่างประเทศ ชาวญี่ปุ่นชอบกรุงเทพฯ เนื่องจากชาวไทยเป็นคนดี

General Message: ชาวไทยSUBเป็นDOBคนดี

Sentence Segmentation:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

1:ชาวญี่ปุ่นนิยมไปที่ต่างประเทศ+ชาวญี่ปุ่นชอบกรุงเทพฯ+เนื่องจากชาวไทยเป็นคนดี+
 2:ชาวญี่ปุ่น(นิยม)นิยม(ไป)ไป(ที่)ที่(ต่าง)ต่าง(ประเทศ)ต่างประเทศ(ชาว)ชาว(ญี่ปุ่น)ชาว(ญี่ปุ่น)ชอบ(ชอบ)ชอบ(กรุงเทพฯ)กรุงเทพฯ(เนื่อง)เนื่อง(จาก)

1>ชาวญี่ปุ่นนิยมไปที่ต่างประเทศ ชาวญี่ปุ่นชอบกรุงเทพฯ (เนื่องจาก) ชาวไทยเป็นคนดี

รูปที่ ข-15 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input ชาวญี่ปุ่นนิยมไปเที่ยวต่างประเทศ ชาวญี่ปุ่นชอบกรุงเทพฯเนื่องจากชาวไทยเป็นคนดี	General Message ชาวไทยSUBเป็นDOBคนดี
Sentence Segmentation <input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt	
เทพ+เนื่องจากชาวไทยเป็นคนดี+มประเทศ(ชาวญี่ปุ่น)ชาวญี่ปุ่น(ชอบ)ชอบ(กรุงเทพฯ)กรุงเทพฯเนื่องจาก(ชาวไทย)ชาวไทย(เป็น)เป็น(คนดี)คนดี ทพ (เนื่องจาก) ชาวไทยเป็นคนดี	

รูปที่ ข-16 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 4 (ต่อ)

ย่อหน้าอินพุทที่ 5

นักศึกษาชอบคอมพิวเตอร์เขาส่งซอฟต์แวร์จากต่างประเทศ

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

ประโยคที่ 1 นักศึกษา ชอบ คอมพิวเตอร์ ตามกฎที่ R3, R8

ประโยคที่ 2 เขา ส่ง ซอฟต์แวร์ จากต่างประเทศ ตามกฎที่ R1, R5

รูปที่ ข-17 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 5

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: นักศึกษาชอบคอมพิวเตอร์เขาสั่งซอฟต์แวร์จากต่างประเทศ

General Message: ซอฟต์แวร์SUBจากDOBต่างประเทศ

Sentence Analysis:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

นักศึกษาAGTชอบOBJคอมพิวเตอร์ เขาAGTสั่งOBJซอฟต์แวร์จากต่างประเทศ ซอฟต์แวร์SUBจากOBJต่างประเทศ

รูปที่ ข-18 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 5

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: นักศึกษาชอบคอมพิวเตอร์เขาสั่งซอฟต์แวร์จากต่างประเทศ

General Message: ซอฟต์แวร์SUBจากDOBต่างประเทศ

Sentence Segmentation:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

1:นักศึกษชอบคอมพิวเตอร์+เขาสั่งซอฟต์แวร์จากต่างประเทศ+

1>นักศึกษชอบคอมพิวเตอร์ เขาสั่งซอฟต์แวร์จากต่างประเทศ

I

รูปที่ ข-19 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าอินพุทที่ 6

ชาวไทยผลิตข้าวเจ้ามาก ต่างประเทศสั่งข้าวเจ้าจากชาวไทยชาวไทยมีเงินมาก

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

- ประโยคที่ 1 ชาวไทย ผลิต ข้าวเจ้า มาก ตามกฎที่ R3, R7
 ประโยคที่ 2 ต่างประเทศ สั่ง ข้าวเจ้า จาก ชาวไทย ตามกฎที่ R4, R5
 ประโยคที่ 3 ชาวไทย มี เงิน มาก ตามกฎที่ R3, R7

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input General Message

ชาวไทยผลิตข้าวเจ้ามาก ต่างประเทศสั่งข้าวเจ้า จากชาวไทยชาวไทยมีเงินมาก

Word Sgmt.
 Sentence Anal.
 Sentence Sgmt.

Word Segmentation

ชาวไทย \\\\ผลิต\\\\ข้าวเจ้า\\\\มาก \\\\ต่างประเทศ\\\\สั่ง\\\\ข้าวเจ้า\\\\จาก\\\\ชาวไทย\\\\ชาวไทย\\\\มี\\\\เงิน\\\\มาก\\\\

รูปที่ ข-20 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 6

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
ชาวไทยผลิตข้าวจำนวนมาก ต่างประเทศสั่งซื้อข้าวเจ้า จากชาวไทยชาวไทยมีเงินมาก	ชาวไทยSUBODOBเงินมาก
Sentence Analysis	<input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt
ชาวไทยAGTผลิตOBJข้าวจำนวนมาก ต่างประเทศSUBสั่งOBJข้าวเจ้าจากชาวไทย ข้าวเจ้าSUBจากOBJชาวไทย ชาวไทย	

รูปที่ ข-21 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 6

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
ชาวไทยผลิตข้าวจำนวนมาก ต่างประเทศสั่งซื้อข้าวเจ้า จากชาวไทยชาวไทยมีเงินมาก	ชาวไทยSUBODOBเงินมาก
Sentence Analysis	<input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt
ต่างประเทศSUBสั่งOBJข้าวเจ้าจากชาวไทย ข้าวเจ้าSUBจากOBJชาวไทย ชาวไทยAGTผลิตOBJเงินมาก	

รูปที่ ข-22 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 6 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
ชาวไทยผลิตข้าวจำนวนมาก ต่างประเทศสั่งข้าวเข้า จากชาวไทยชาวไทยมีเงินมาก	ชาวไทยSUBSIDOเงินมาก
Sentence Segmentation	<input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt.
1:ชาวไทยผลิตข้าวจำนวนมาก+ต่างประเทศสั่งข้าวเข้าจากชาวไทย+ชาวไทยมีเงินมาก+ 1>ชาวไทยผลิตข้าวจำนวนมาก ต่างประเทศสั่งข้าวเข้าจากชาวไทย ชาวไทยมีเงินมาก I	

รูปที่ ข-23 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 6

ย่อหน้าอินพุทที่ 7

มานะเล่นดนตรีในโรงละคร มานะใช้เปียโนตามปกติ ผู้ชมนิยมดนตรีของมานะ

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

ประโยคที่ 1	มานะ เล่น ดนตรี ใน โรงละคร	ตามกฎที่ R3, R5
ประโยคที่ 2	มานะ ใช้ เปียโน ตามปกติ	ตามกฎที่ R3, R7
ประโยคที่ 3	ผู้ชม นิยม ดนตรี ของ มานะ	ตามกฎที่ R3, R5

ซึ่งเป็นเซตคำตอบเซตที่ 1 จากเซตคำตอบที่ได้จากการแบ่งคำ 8 เซต

รูปที่ ข-24 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 7

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input	General Message
<p>มานะเล่นดนตรีในโรงละครมานะใช้เปียโนตามปกติ ผู้ชมนิยมดนตรีของมานะ</p>	<p>ผู้SUBชม</p>
<p>Sentence Analysis</p> <p> <input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt </p>	
<p>มานะAGTเล่นOBJดนตรีในโรงละคร มานะAGTใช้OBJเปียโนตามปกติ ผู้ชมAGTนิยมOBJดนตรีของมานะ มานะAGTเล่นOBJดนตรีในโรงละคร มานะAGTใช้OBJเปียโนตามปกติ ผู้AGTชม</p>	

รูปที่ ข-25 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 7

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input	General Message
<p>มานะเล่นดนตรีในโรงละครมานะใช้เปียโนตามปกติ ผู้ชมนิยมดนตรีของมานะ</p>	<p>ผู้SUBชม</p>
<p>Sentence Segmentation</p> <p> <input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt </p>	
<p>1:มานะเล่นดนตรีในโรงละคร+มานะใช้เปียโนตามปกติ+ผู้ชมนิยมดนตรีของมานะ+ 2:มานะเล่นดนตรีในโรงละคร+มานะใช้เปียโนตามปกติ+ผู้ชม+(นิยม)นิยม(ดนตรี)ดนตรี(ของ)ของ(มานะ)มานะ</p> <p>1>มานะเล่นดนตรีในโรงละคร มานะใช้เปียโนตามปกติ ผู้ชมนิยมดนตรีของมานะ</p>	

รูปที่ ข-26 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าอินพุทที่ 8

ปลาดิบเป็นของชาวญี่ปุ่นและชาวไทยมีข้าวเหนียว

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

ประโยคที่ 1 ปลาดิบ เป็น ของ ชาวญี่ปุ่น ตามกฎที่ R4, R9

ประโยคที่ 2 ชาวไทย มี ข้าวเหนียว ตามกฎที่ R3, R8

คำเชื่อมประโยคที่ 1 และ 2 และ

รูปที่ ข-27 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 8

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
ปาลาดิบเป็นของชาวญี่ปุ่นและชาวไทยมีข้าวเหนียว	ชาวไทยSUBมีDOBJข้าวเหนียว
Sentence Analysis	<input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt
ปาลาดิบSUBเป็นOBJชาวญี่ปุ่น ชาวไทยAGTมีDOBJข้าวเหนียว	

รูปที่ ข-28 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 8

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
ปาลาดิบเป็นของชาวญี่ปุ่นและชาวไทยมีข้าวเหนียว	ชาวไทยSUBมีDOBJข้าวเหนียว
Sentence Segmentation	<input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt
1:ปาลาดิบเป็นของชาวญี่ปุ่น+และชาวไทยมีข้าวเหนียว+ 1>ปาลาดิบเป็นของชาวญี่ปุ่น (และ) ชาวไทยมีข้าวเหนียว I	

รูปที่ ข-29 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าอินพุทที่ 9

สมชายเป็นนักศึกษาเขาเป็นคนดีเนื่องจากเขานิยมดนตรีและเนื่องจาก
เขามีความสามารถดี

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

ประโยคที่ 1	สมชาย เป็น นักศึกษา	ตามกฎที่ R3, R8
ประโยคที่ 2	เขา เป็น คนดี	ตามกฎที่ R1, R8
ประโยคที่ 3	เขา นิยม ดนตรี	ตามกฎที่ R1, R8
ประโยคที่ 4	เขา มี ความสามารถ ดี	ตามกฎที่ R1, R8

คำเชื่อมประโยคที่ 2 และ 3 เนื่องจาก
คำเชื่อมประโยคที่ 3 และ 4 และ
คำเชื่อมประโยคที่ 3 และ 4 เนื่องจาก

ซึ่งเป็นเซตคำตอบที่ 1 จากเซตผลลัพธ์ที่ได้จากการแบ่งคำ 2 เซต

เอกสารนี้เป็นเอกสารที่รูปที่ ๓-30 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 9
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Word Segmentation Form

Paragraph segment Window Message

<p>Paragraph Input</p> <p>สมชายเป็นนักศึกษาเขาเป็นคนดีเนื่องจากเขานิยม ดนตรีและเนื่องจากเขามีความสามารถ</p>	<p>General Message</p> <p>เขาSUBมีDOBความสามารถ</p>
<p>Sentence Analysis</p> <p><input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt</p>	
<p>สมชายAGTเป็นOBJนักศึกษา เขาAGTเป็นOBJคนดี เขาAGTนิยมOBJดนตรี เขาAGTมีDOBความสามารถ ชายSUBเป็นOBJนักศึกษา เขาAGTเป็นOBJคนดี เขาAGTนิยมOBJดนตรี เขาAGTมีDOBความสามารถ</p>	

รูปที่ ข-31 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 9

Word Segmentation Form

Paragraph segment Window Message

<p>Paragraph Input</p> <p>สมชายเป็นนักศึกษาเขาเป็นคนดีเนื่องจากเขานิยม ดนตรีและเนื่องจากเขามีความสามารถ</p>	<p>General Message</p> <p>เขาSUBมีDOBความสามารถ</p>
<p>Sentence Segmentation</p> <p><input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input checked="" type="radio"/> Sentence Sgmt</p>	
<p>1:สมชายเป็นนักศึกษา+เขาเป็นคนดี+เนื่องจากเขานิยมดนตรี+และเนื่องจากเขามีความสามารถ+ 2: (สม)สมชายเป็นนักศึกษา+เขาเป็นคนดี+เนื่องจากเขานิยมดนตรี+และเนื่องจากเขามีความสามารถ+ 1>สมชายเป็นนักศึกษา เขาเป็นคนดี (เนื่องจาก) เขานิยมดนตรี (และ) (เนื่องจาก) เขามีความสามารถ</p>	

รูปที่ ข-32 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าอินพุทที่ 10

วันนี้เป็นวันอาทิตย์ซึ่งข้าพเจ้าอยู่ที่บ้านและข้าพเจ้าใช้คอมพิวเตอร์เนื่องจาก
คอมพิวเตอร์มีประโยชน์กับข้าพเจ้า

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

ประโยคที่ 1	วันนี้ เป็น วันอาทิตย์	ตามกฎที่ R4, R8
ประโยคที่ 2	ข้าพเจ้า อยู่ ที่ บ้าน	ตามกฎที่ R1, R9
ประโยคที่ 3	ข้าพเจ้า ใช้ คอมพิวเตอร์	ตามกฎที่ R1, R8
ประโยคที่ 4	คอมพิวเตอร์ มี ประโยชน์	ตามกฎที่ R4, R8

คำเชื่อมประโยคที่ 1 และ 2 ซึ่ง
คำเชื่อมประโยคที่ 2 และ 3 และ
คำเชื่อมประโยคที่ 3 และ 4 เนื่องจาก

แต่การวิเคราะห์โครงสร้างประโยคของประโยคที่ 4 ผิดพลาด เนื่องจากวิเคราะห์ประโยคออกมาไม่ตรงกับการวิเคราะห์ด้วยมือ ดังนี้

การวิเคราะห์ด้วยระบบ ฯ

คอมพิวเตอร์ มี ประโยชน์

การวิเคราะห์ด้วยมือ

คอมพิวเตอร์ มี ประโยชน์ กับ ข้าพเจ้า

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
<p>วันนี้เป็นวันอาทิตย์ซึ่งข้าพเจ้าอยู่กับบ้านและ ข้าพเจ้าใช้คอมพิวเตอร์เนื่องจากคอมพิวเตอร์มี ประโยชน์กับข้าพเจ้า</p>	
<p>Word Segmentation</p> <p><input checked="" type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt</p>	
<p>วันนี้เป็น\วันอาทิตย์ซึ่งข้าพเจ้าอยู่กับบ้านและข้าพเจ้าใช้\คอมพิวเตอร์เนื่องจาก\คอมพิวเตอร์มี\ประโยชน์\</p>	

รูปที่ ข-33 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 10

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
<p>วันนี้เป็นวันอาทิตย์ซึ่งข้าพเจ้าอยู่กับบ้านและ ข้าพเจ้าใช้คอมพิวเตอร์เนื่องจากคอมพิวเตอร์มี ประโยชน์กับข้าพเจ้า</p>	
<p>Word Segmentation</p> <p><input checked="" type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt</p>	
<p>\และข้าพเจ้าใช้\คอมพิวเตอร์เนื่องจาก\คอมพิวเตอร์มี\ประโยชน์กับ\ข้าพเจ้า\</p>	

รูปที่ ข-34 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 10 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
<p>วันนี้เป็นวันอาทิตย์ซึ่งข้าพเจ้าอยู่ที่บ้านและ ข้าพเจ้าใช้คอมพิวเตอร์เนื่องจากคอมพิวเตอร์มี ประโยชน์กับข้าพเจ้า</p>	<p>คอมพิวเตอร์SUBมีDOBJประโยชน์</p>
Sentence Analysis	<input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt
<p>วันนี้SUBเป็นDOBJวันอาทิตย์ ข้าพเจ้าAGTอยู่OBJบ้าน ข้าพเจ้าAGTใช้OBJคอมพิวเตอร์ คอมพิวเตอร์SUBมีDOBJประโยชน์</p>	

รูปที่ ข-35 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 10

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
<p>วันนี้เป็นวันอาทิตย์ซึ่งข้าพเจ้าอยู่ที่บ้านและ ข้าพเจ้าใช้คอมพิวเตอร์เนื่องจากคอมพิวเตอร์มี ประโยชน์กับข้าพเจ้า</p>	<p>คอมพิวเตอร์SUBมีDOBJประโยชน์</p>
Sentence Analysis	<input type="radio"/> Word Sgmt. <input checked="" type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt
<p>ข้าพเจ้าAGTอยู่OBJบ้าน ข้าพเจ้าAGTใช้OBJคอมพิวเตอร์ คอมพิวเตอร์SUBมีDOBJประโยชน์</p>	

รูปที่ ข-36 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 10 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input

วันนี้เป็นวันอาทิตย์ซึ่งข้าพเจ้าอยู่ที่บ้านและ
ข้าพเจ้าใช้คอมพิวเตอร์เนื่องจากคอมพิวเตอร์มี
ประโยชน์กับข้าพเจ้า

General Message

คอมพิวเตอร์SUBDDO8ประโยชน์

Sentence Segmentation

Word Sgmt.
 Sentence Anal.
 Sentence Sgmt.

1:วันนี้เป็นวันอาทิตย์+ซึ่งข้าพเจ้าอยู่ที่บ้าน+และข้าพเจ้าใช้คอมพิวเตอร์+เนื่องจากคอมพิวเตอร์มีประโยชน์+(กับ)กับ(ข้า

รูปที่ ข-37 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 10

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input

วันนี้เป็นวันอาทิตย์ซึ่งข้าพเจ้าอยู่ที่บ้านและ
ข้าพเจ้าใช้คอมพิวเตอร์เนื่องจากคอมพิวเตอร์มี
ประโยชน์กับข้าพเจ้า

General Message

คอมพิวเตอร์SUBDDO8ประโยชน์

Sentence Segmentation

Word Sgmt.
 Sentence Anal.
 Sentence Sgmt.

+ซึ่งข้าพเจ้าอยู่ที่บ้าน +และข้าพเจ้าใช้คอมพิวเตอร์ +เนื่องจากคอมพิวเตอร์มีประโยชน์+(กับ)กับ(ข้าพเจ้า)ข้าพเจ้า

รูปที่ ข-38 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 10 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ย่อหน้าอินพุทที่ 11

เกิดอุทกภัยเนื่องจากชาวไทยทำลายป่า

ผลจากการวิเคราะห์แยกประโยคภาษาไทยจากย่อหน้า จะได้

ประโยคที่ 1 เกิด อุทกภัย ตามกฎที่ R13

ประโยคที่ 2 ชาวไทย ทำลาย ป่า ตามกฎที่ R3, R8

คำเชื่อมประโยคที่ 1 และ 2 เนื่องจาก

Word Segmentation Form	
Paragraph segment	Window Message
Paragraph Input	General Message
เกิดอุทกภัยเนื่องจากชาวไทยทำลายป่า	
Word Segmentation	<input type="radio"/> Word Sgmt. <input type="radio"/> Sentence Anal. <input type="radio"/> Sentence Sgmt.
เกิดอุทกภัยเนื่องจากชาวไทยทำลายป่า	

รูปที่ ข-39 แสดงการวิเคราะห์หน่วยคำในย่อหน้าต้นแบบที่ 11

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: เกิดอุทกภัยเนื่องจากชาวไทยทำลายป่า

General Message: ชาวไทยSUBทำลายDOBป่า

Sentence Analysis:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

เกิดOBอุทกภัย ชาวไทยAGTทำลายOBป่า

รูปที่ ข-40 แสดงการวิเคราะห์โครงสร้างประโยคภาษาไทยในย่อหน้าต้นแบบที่ 11

Word Segmentation Form

Paragraph segment Window Message

Paragraph Input: เกิดอุทกภัยเนื่องจากชาวไทยทำลายป่า

General Message: ชาวไทยSUBทำลายDOBป่า

Sentence Segmentation:

- Word Sgmt.
- Sentence Anal.
- Sentence Sgmt.

1:เกิดอุทกภัย+เนื่องจากชาวไทยทำลายป่า+

1>เกิดอุทกภัย (เนื่องจาก) ชาวไทยทำลายป่า

รูปที่ ข-41 แสดงการวิเคราะห์ประโยคภาษาไทยจากย่อหน้าในย่อหน้าต้นแบบที่ 11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน	นายคังกรศรัณย์ ล่องชุมพล
วันเดือนปีเกิด	31 มีนาคม พ.ศ. 2509
สถานที่เกิด	จังหวัดเชียงใหม่
วุฒิการศึกษาระดับปริญญาตรี	วท.บ.(ฟิสิกส์)
สถานที่สำเร็จการศึกษา	คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ พิษณุโลก
ปีที่สำเร็จการศึกษา	2530
ผลงานทางวิชาการ	<ul style="list-style-type: none"> - R. Varakulsiriputh and S.Longchupole, "Performance Analysis of Multi-Networking System", The 1990 Engineering Technology Symposium and Exhibition. - R. Varakulsiripunth and S.Longchupole, "Reliability Improvement in Loop Communication Network", The 11th Conference of Electrical Engineering. - R. Varakulsiripunth and S.Longchupole, "Thai Syntactical Analysis by Method of Splitting Sentences from Paragraph", Computer Journal, The Computer Association of THAILAND under The Royal Patronage of His Majesty The King.
ประสบการณ์การทำงาน	<ul style="list-style-type: none"> Lecturer, Computer Engineering Dept., KMITL (1994-1995) Researcher, MMT Project. NECTEC-CICC project.(1992-1994) Application Engineer, IRIS-IFEC Co.,Ltd.(1991-1992) System Engineer, IFEC Co.,Ltd. (1990-1991) Research Assistant, CRSC, KMITL. (1989-1990)