

การใช้คำกริยาเป็นแกนในการสร้างกรอบมโนทัศน์ ภาษาไทยจากคลังข้อมูลด้านการเกษตร เพื่อการประมวลผลภาษาด้วยคอมพิวเตอร์

มุกชดา สุขธาราจารย์*

พุทธชาติ โปธิบาล**

นัทธชนัน นาถประพาน***

บทคัดย่อ

บทความนี้นำเสนอแนวทางการใช้คำกริยาเป็นแกนในการสร้างกรอบมโนทัศน์ภาษาไทยจากคลังข้อมูลด้านการเกษตรเพื่อการประมวลผลภาษาด้วยคอมพิวเตอร์ โดยอาศัยการทบทวนของฟิลล์มอร์(Fillmore, 1974) และลาร์สัน (Larson, 1984) ในทฤษฎีความหมายเชิงโครงสร้าง (frame semantics) เพื่อแสดงความสัมพันธ์ของคำที่ปรากฏร่วมกันในประโยค ข้อมูลที่นำมาใช้ประกอบด้วยจำนวนประโยคภาษาไทยที่มีการตัดคำอัตโนมัติ กำกับข้อมูลสารสนเทศแบบอัตโนมัติ และกำกับโครงสร้างระดับวลีพร้อมด้วยบทบาททางความหมายของคำด้วย

* นิสิตระดับดุษฎีบัณฑิต ภาควิชาภาษาศาสตร์ คณะมนุษยศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ติดต่อได้ที่: mukda.suktarachan@gmail.com

** ผู้ช่วยศาสตราจารย์ประจำภาควิชาภาษาศาสตร์ คณะมนุษยศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ติดต่อได้ที่: fhumpcpb@ku.ac.th

*** อาจารย์ประจำภาควิชาภาษาศาสตร์ คณะมนุษยศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ ติดต่อได้ที่: fhumnncn@ku.ac.th

มีจำนวน 5,784 ประโยค โดยประโยคทั้งหมดคัดเลือกมาจากประโยคที่มีคำกริยาปรากฏในความถี่สูงสุด 106 คำ จากนั้นจึงสกัดกรอบการพหุความสัมพันธ์ทางความหมายออกมาเป็นกรอบมโนทัศน์โดยใช้โปรแกรมดึงและจัดระเบียบข้อมูลอัตโนมัติ ผลการศึกษาสามารถสกัดกรอบมโนทัศน์ออกมาได้ทั้งหมด 962 กรอบ พร้อมกับเขียนข้อกำหนดการเกิดร่วมกันของคำตามโครงสร้างไวยากรณ์และความหมายให้กับประโยคทั้งหมดเพื่อประยุกต์ใช้ในการประมวลผลด้วยคอมพิวเตอร์ได้อย่างสะดวกง่ายดาย

คำสำคัญ: กรอบมโนทัศน์; ความหมายเชิงโครงสร้าง; เฟรมเน็ต; การก;
ความสัมพันธ์ทางความหมาย

Verb Centric Based Thai Concept Frame Construction for Application in Agricultural Domain Language Processing

Mukda Suktarachan^{*}

Puttachart Potibal^{**}

Natchanan Natpratan^{***}

Abstract

This paper presents a construction of Thai concept frames applied in language processing for agricultural domain based on the verb centric approach. Fillmore's (1974) and Larson's (1984) frame elements were selected to represent necessary scenarios in the domain. The source data comprised 5,784 Thai sentences which were preprocessed with automatic word segmentation, automatic POS tagging, manual shallow parsing, and manual semantic role labeling. The 106 most frequently used verbs were chosen for observation. Frame elements and syntactic-semantic relations for each of them were then constructed from the source data using an automatic capturing tool. As a result, 962 case frames

^{*} Ph.D.'s student, Department of Linguistics, Faculty of Humanities, Kasetsart University. e-mail: mukda.suktarachan@gmail.com

^{**} Assistant Professor, Department of Linguistics, Faculty of Humanities, Kasetsart University. e-mail: fhumpcpb@ku.ac.th

^{***} Lecturer, Department of Linguistics, Faculty of Humanities, Kasetsart University. e-mail: fhumncn@ku.ac.th

were constructed, while 5,784 annotated sentences were formulated as syntactic-semantic constraints for easy application to language processing.

Keywords: concept frame; Frame Semantics; FrameNet; case;
semantic relation

1. บทนำ

กรอบมโนทัศน์ (concept frame) เปรียบเสมือนฉากสถานการณ์ (scenario) ที่สามารถแสดงให้เห็นโครงสร้างของเหตุการณ์ทั้งหมดโดยการประกอบส่วนที่มีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่งเข้าด้วยกัน ซึ่งฉากสถานการณ์แต่ละฉากย่อมประกอบด้วยส่วนประกอบที่แตกต่างกัน ในทางภาษาศาสตร์ เราสามารถสร้างกรอบมโนทัศน์จากความสัมพันธ์ทางความหมายที่ปรากฏร่วมกันในประโยคได้เช่นกัน

นักภาษาศาสตร์หลายคนได้พัฒนาทฤษฎีเกี่ยวกับการสร้างกรอบความหมายหรือกรอบมโนทัศน์โดยอาศัยความสัมพันธ์ทางความหมายของคำที่ปรากฏร่วมกันในประโยค อาทิ ชาร์ล เจ. ฟิลล์มอร์ (Fillmore, 1968) เจฟฟรีย์ กรูเบอร์ (Gruber, 1965) เรย์ แจ็คเอนด์อร์ฟ (Jackendoff, 1985) โนม ชอมสกี (Chomsky, 1981) วอลเลซ เชฟ (Chafe, 1970) ทั้งนี้การสร้างกรอบความสัมพันธ์ทางความหมายของคำเป็นสิ่งที่ต้องการใช้งานสำหรับนักวิจัยในสาขาการประมวลผลภาษาธรรมชาติ (Natural Language Processing) เป็นอย่างมาก โดยเฉพาะอย่างยิ่งเมื่อมหาวิทยาลัยแคลิฟอร์เนีย เบิร์กลีย์ได้พัฒนาโครงการ Berkeley FrameNet¹ (Fillmore 1976, 1977, 1982, 1985, Fillmore and Baker 2001, 2010) ขึ้น จากนั้นนักวิจัยในหลายประเทศหลายภาษาก็ได้พัฒนา FrameNet ลักษณะเดียวกับ FrameNet ของ Berkeley โดยวิธีการนำแนวคิดในการจัดทำ FrameNet ไปใช้ หรือบางภาษาใช้วิธีแปลจากต้นฉบับของ Berkeley FrameNet งานวิจัยที่เผยแพร่ออนไลน์ในปัจจุบันมีหลายภาษา อาทิ Spanish FrameNet (Subirats and Miriam, 2003), Japanese FrameNet (Ohara et al., 2003), FrameNet Brasil (Torrent et al., 2014), Kicktionary (Schmidt, 2009), SALSA project (Burchardt et al., 2009) เป็นต้น ส่วนเฟรมเน็ตภาษาไทย ธนนท์ หลีน้อย และคณะ (Leenoi et al., 2010, pp. 277-280) ได้พัฒนา Thai FrameNet ขึ้นโดยใช้วิธีแปลจาก FrameNet ของ Berkeley และมีการเพิ่มข้อมูล

¹ <http://framenet.icsi.berkeley.edu>

ประโยคภาษาไทยมาประกอบและกำกับข้อมูลทางภาษาเพื่อให้สอดคล้องกับเฟรมภาษาไทย

อย่างไรก็ตาม การนำเฟรมเน็ตไปใช้ประโยชน์ในทางประมวลผลภาษาธรรมชาติหรือการประมวลผลภาษาด้วยคอมพิวเตอร์จำเป็นต้องคัดเลือกข้อมูลให้เหมาะสมกับการใช้งาน เนื่องจากปัจจุบันเทคโนโลยีมีความก้าวหน้ามากขึ้นเรื่อยๆ สังคมของคนเราเป็นสังคมแห่งข้อมูลขนาดใหญ่ (big data) ที่ประกอบด้วยองค์ความรู้และข้อมูลชนิดต่างๆ ที่หลากหลาย ทำให้มีการคิดค้นและพัฒนาเทคโนโลยีขึ้นมาจัดการกับข้อมูลข่าวสารเพื่อนำมาใช้ประโยชน์และตอบโจทย์ในด้านต่างๆ ทั้งในภาครัฐและเอกชน อย่างไรก็ตาม การนำความรู้มาใช้ผ่านเทคโนโลยีนั้นไม่ใช่เรื่องง่าย เนื่องจากปัจจัยทางด้านภาษา ซึ่งเป็นปัจจัยที่มีความละเอียดอ่อนและซับซ้อน ในตัวภาษาเองมีข้อมูลและสารสนเทศหลายระดับ อีกทั้งภาษามีพัฒนาการมีขอบเขต มีเครือข่าย และมีการเปลี่ยนแปลงของข้อมูลอยู่เสมอ ปัจจัยดังกล่าวก่อให้เกิดปัญหาในการประมวลผลภาษาธรรมชาติหรือการประมวลผลภาษาเชิงคำนวณซึ่งเป็นความพยายามของมนุษย์ในการพัฒนาระบบที่ทำให้คอมพิวเตอร์สามารถสื่อสารได้อย่างชาญฉลาดเช่นเดียวกับมนุษย์ ในการประมวลผลภาษาธรรมชาตินั้น ข้อมูลที่นำเข้าสู่ระบบการประมวลผลต้องมีความชัดเจน ไม่กำกวม ซึ่งเป็นเรื่องยากในการป้อนความรู้ทางภาษาให้ครอบคลุมทุกปัญหาทางภาษาให้กับคอมพิวเตอร์ ส่งผลให้การนำความรู้และข้อมูลสารสนเทศที่สกัดได้ด้วยระบบเทคโนโลยีต่างๆ ไม่ตรงตามความต้องการของผู้ใช้ ดังนั้น ในงานวิจัยนี้จึงศึกษาวิเคราะห์ข้อมูลเฉพาะด้านการเกษตรเท่านั้น ทั้งนี้เนื่องจากมหาวิทยาลัยเกษตรศาสตร์เป็นแหล่งข้อมูลด้านการเกษตรที่มีค้ำยั้ง ทั้งข้อมูลเก่าและใหม่ อีกทั้งเป็นแหล่งรวบรวมข้อมูลและให้บริการสารสนเทศด้านการเกษตรทั้งในระดับประเทศและระดับนานาชาติ การนำความรู้ด้านการเกษตรมาใช้ให้เกิดประโยชน์ต่อประชาชนและประเทศชาติจึงเป็นสิ่งที่ควรทำ เพื่อเสริมสร้างอัตลักษณ์ของมหาวิทยาลัยที่ต้องการให้ความรู้ทางด้านการเกษตรเป็นศาสตร์แห่งแผ่นดิน

1.1 ทบทวนเอกสาร

FrameNet (Fillmore et al., 2003) เป็นฐานข้อมูลคำศัพท์ ประกอบด้วย คำศัพท์ภาษาอังกฤษ กว่า 11,600 คำ มีการกำกับข้อมูลแบบครบถ้วนกว่า 6,800 ชุด และประกอบด้วยเฟรม 960 เฟรม โดยมีตัวอย่างประกอบ 150,000 ประโยค สร้างขึ้นโดยอาศัยหลักการหรือทฤษฎีความหมายเชิงโครงสร้าง (Frame Semantics) (Fillmore, 1982, 1985) ซึ่งเป็นทฤษฎีที่มองว่า กรอบ (frame) เป็นระบบทางภาษาที่สร้างขึ้นจากการรวมเอาคำศัพท์ต่างๆ ประกอบเข้ากับกฎเกณฑ์ทางไวยากรณ์ แล้วเชื่อมโยงบทบาททางความหมายของคำ (thematic roles) ที่ปรากฏในฉากสถานการณ์ (scene) กับโครงสร้างไวยากรณ์ เพื่ออธิบายว่ามนุษย์สามารถเข้าใจความหมายของคำได้อย่างไรจากความสัมพันธ์ที่สร้างขึ้นเป็นฉากสถานการณ์ ฟิลล์มอร์เรียกกรอบความหมายเชิงโครงสร้างนี้ว่าไวยากรณ์กรก (Case Grammar) (Fillmore, 1968)

บทบาททางความหมายของคำหรือการกที่ใช้ในโครงการ FrameNet นั้น เป็นการกระตบลิค ซึ่งกรอบแต่ละกรอบจะประกอบด้วยการกใดบ้างนั้นขึ้นอยู่กับ มโนทัศน์ (concept) ของกรอบ ดังตัวอย่างการกำกับข้อมูลการกของกรอบ Commerce buy ดังต่อไปนี้

1. Four years ago [_{Buyer}I] *BOUGHT*^{Target} [_{Goods}an old Harmony Sovereign acoustic guitar] [_{Money}for £20] [_{Seller}from an absolute prat].
2. [_{Buyer}Heston], who had director, script and cast approval, *BOUGHT*^{Target} [_{Recipient}Peckinpah] [_{Goods}extra filming time] [_{Means}by returning his salary to the studio], convinced they wouldn't accept it.
3. On other occasions, borrowing may be the only way [_{Buyer}you] will ever be able to afford to *BUY*^{Target} [_{Goods}something expensive] like a house.

จากตัวอย่างข้างต้นจะสังเกตได้ว่า ประธานของประโยคมีบทบาทเป็นผู้ซื้อ ซึ่งถือเป็นการกระตบลิค การกำกับการกระตบลิคนี้มีข้อเสียคือการกระตบลิคมีความซ้ำซ้อนกับการกดั้งเดิม ซึ่งจากตัวอย่างข้างต้น เราสามารถกำหนดให้บทบาทผู้ซื้อ

(Buyer) เป็นบทบาทผู้กระทำ (Agent) บทบาทสินค้า (Goods) เป็นบทบาทผู้ถูกกระทำ (Object) บทบาทผู้ขาย (Seller) เป็นบทบาทของสิ่งที่เป็นจุดเริ่มต้นการกระทำ (source) และบทบาทผู้รับ (Recipient) เป็นบทบาทของสิ่งที่ได้รับผลจากการกระทำ (Benefactive) ก็ได้

โครงสร้างเฟรม

โครงสร้างเฟรมใน FrameNet ของมหาวิทยาลัยแคลิฟอร์เนีย เบิร์กลีย์ ประกอบด้วย concept หรือโน้ตหลักซึ่งเป็นตัวแทนความหมายของกลุ่มคำที่มีความหมายรวมกลุ่มเดียวกัน ซึ่งภายในกรอบมโนทัศน์นี้จะประกอบด้วยสมาชิกที่เป็นรายการคำ (Lexical Unit หรือ LU) และมีการเชื่อมโยงกรอบมโนทัศน์ให้เป็นโครงข่ายระหว่างกรอบต่างๆ โครงสร้างเฟรมประกอบด้วยส่วนต่างๆ 5 ส่วน ดังนี้

Frame Name: คือชื่อของเฟรม

Definition: เป็นส่วนที่เป็นคำอธิบายความหมายของเฟรม

Frame elements (Fes): เป็นบทบาททางความหมายของคำ (thematic roles)

Frame-frame Relations: เป็นความสัมพันธ์ระหว่างเฟรมที่แตกต่างกัน

Lexical Units: คือรายการคำที่อยู่ใน Frame Name เดียวกัน

ส่วนที่สำคัญในโครงสร้างเฟรมใน FrameNet คือส่วนประกอบของเฟรม (frame elements) ซึ่งใน FrameNet มีการกำหนดส่วนประกอบของเฟรมเป็นส่วนที่เป็นแกน (core element) กับส่วนที่ไม่ได้เป็นแกน (non-core element) โดยในส่วนประกอบที่เป็นแกนของกรอบ (core frame elements) จะเป็นส่วนประกอบหลักของเฟรม มักเป็นส่วนที่เป็นประธานและกรรม ดังตัวอย่างประโยคต่อไปนี้

“She was born about AD 460.”

จากตัวอย่างเป็นกรอบของ “Being_born” ซึ่ง “She” จะถูกกำกับให้เป็นนามวลี เป็นส่วนประกอบของเฟรม “Child” โดยเป็นส่วนประกอบหลักหรือเป็นแกนของกรอบ (core frame elements) และเป็นประธานของประโยค สำหรับส่วนที่ไม่ได้เป็นแกนของกรอบ (non-core frame elements) เป็นส่วนที่อยู่ชายขอบ (peripheral) ไม่ได้เป็นส่วนหลักของประโยค แต่จะเสริมความให้สมบูรณ์ขึ้น มักจะเป็นส่วนที่

เกี่ยวข้องกับเวลา สถานที่ รวมทั้งส่วนอื่นๆ ที่เกี่ยวข้อง จากตัวอย่างที่กล่าวมาแล้วข้างต้น ส่วนที่พูดถึงเวลา คือ “about AD 460” ซึ่งเป็นนามวลีที่บอกกรอบเวลา ถือเป็นส่วนที่อยู่ซ้ายขอบ เป็นส่วนขยาย จึงเป็นส่วนที่ไม่ได้เป็นแกนของกรอบ ใน FrameNet จะมีการกำหนดส่วนประกอบของกรอบแต่ละกรอบไม่เท่ากัน ขึ้นอยู่กับข้อมูลที่ปรากฏในเอกสารที่กำลังศึกษา

1.2 งานวิจัยที่เกี่ยวข้อง

WordNet เป็นพจนานุกรมออนไลน์ที่อิงหลักภาษาศาสตร์เชิงจิตวิทยา โดยมีวัตถุประสงค์เพื่อให้สามารถค้นหาและนำไปใช้งานได้ง่าย สะดวกและรวดเร็ว WordNet มีลักษณะคล้ายอรรถาภิธาน (Thesaurus) เพราะมีการจัด concept ของคำเป็นแบบลาดหลั่น การค้นหาคำสามารถบอกถึงคำที่ต้องการหาและคำที่เกี่ยวข้องทั้งหมดออกมาได้ อีกทั้งมีคำอธิบายความหมายและตัวอย่างการใช้คำนั้นๆ แบบพจนานุกรมมาตรฐานทั่วไป ทำให้สะดวกกว่าการค้นหาแบบเรียงตัวอักษรอย่างในการค้นหาจากพจนานุกรมทั่วไป เพราะการค้นหาแบบเรียงตัวอักษรนั้น คำที่มีความหมายเกี่ยวข้องกัน บางครั้งเป็นคำที่มีความหมายเหมือนกัน (Synonym) ก็จะปรากฏอยู่หลายที่ ทั้งที่คำนั้นๆ มีปรากฏอยู่ในความหมายของอีกคำหนึ่งแล้ว

WordNet มีจำนวนคำทั้งหมด 95,600 คำ (51,500 simple words และ 44,100 collocations) รวม 70,100 ความหมาย

กรอบคำ (lexical matrix) ของ WordNet สร้างขึ้นเพื่อใช้ในการพิจารณาความสัมพันธ์ระหว่างรูปคำ (Form) และความหมายของคำ (meaning) ทั้งนี้เพื่อนำไปใช้ในการจัดกลุ่มของคำ (synset) ตามความหมาย ดังตัวอย่างต่อไปนี้

ตารางที่ 1

แสดงการจัดกลุ่มความหมายของคำ

word meanings	word forms				
	F1	F2	F3	...	Fn
M1	E1,1	E1,2			
M2		E2,2			
M3			E3,3		
...					
Mm					Em,n

จากตารางข้างต้นมีความหมายว่าคำใน F1 and F2 เป็นคำพ้องความหมาย (synonyms) และ F2 เป็นคำหลายความหมาย (polysemous)

ความสัมพันธ์ทางความหมายของคำ (Semantic Relations) ใน WordNet มีดังนี้

1. Synonymy คือ คำที่มีความหมายเหมือนกัน
2. Antonymy คือ คำที่มีความหมายตรงข้ามกัน
3. Hyponymy คือ คำที่มีความสัมพันธ์กันแบบเป็นคำลูกกลุ่ม (subordinate) หรือมีความสัมพันธ์แบบค่านั้นเป็นสมาชิกของคำที่มีความหมายกว้างกว่า (a kind of) นั้นเอง
4. Meronymy คือ คำที่มีความสัมพันธ์กันแบบเป็นส่วนประกอบของกันและกัน (a part of)

2. วิธีดำเนินการวิจัย

2.1 แหล่งข้อมูล

งานวิจัยนี้รวบรวมข้อมูลทางด้านการเกษตรจากเว็บไซต์ต่างๆ เพื่อนำมาใช้เป็นคลังข้อมูล โดยการเขียนสคริปต์ดึงข้อมูลที่เกี่ยวข้องกับคำศัพท์ที่ใช้เป็นคำค้นเริ่มต้นทางด้านการเกษตร เช่น คำว่า ข้าว ปุ๋ย โรคข้าว เป็นต้น เพื่อรวบรวมข้อมูลจากแหล่งข้อมูล 2 ประเภท ได้แก่ ข้อมูลเชิงวิชาการและข่าวทางด้านการเกษตรซึ่งเผยแพร่ออนไลน์ โดยข้อมูลเชิงวิชาการเป็นข้อมูลที่รวบรวมจากเว็บไซต์ขององค์กรภาครัฐที่เผยแพร่ความรู้ทางด้านการเกษตร เช่น เว็บไซต์ของโครงการสารานุกรมไทยสำหรับเยาวชน โดยพระราชประสงค์ในพระบาทสมเด็จพระเจ้าอยู่หัว เว็บไซต์กรมการข้าว เว็บไซต์สำนักงานพัฒนาการวิจัยการเกษตร (องค์การมหาชน) เว็บไซต์รักบ้านเกิด เว็บไซต์กระทรวงเกษตรและสหกรณ์ วิกิพีเดีย สารานุกรมเสรี เป็นต้น ซึ่งสามารถรวบรวมได้ทั้งหมด 3,390 ไฟล์ และรวบรวมข้อมูลข่าวเกี่ยวกับการเกษตรจากหนังสือพิมพ์ออนไลน์ 682 ไฟล์ รวมข้อมูลทั้งหมด 4,072 ไฟล์ จากนั้นจึงนำมาคัดกรองข้อมูล โดยการคัดเลือกข้อมูลที่เกี่ยวข้องกับการเกษตรเท่านั้น ทั้งนี้เนื่องจากการรวบรวมข้อมูลโดยใช้ชุดคำค้นเพื่อรวบรวมข้อมูลอัตโนมัติด้วยคอมพิวเตอร์นั้น ทำให้ข้อมูลที่ได้มีข้อมูลที่ไม่เกี่ยวข้องกับการเกษตรด้วย เพราะคำค้นที่ป้อนอัตโนมัตินั้นปรากฏอยู่ในข้อความหรือในบทความบนเว็บไซต์ ระบบจึงรวบรวมมาทั้งหมด เช่น ปรากฏในข่าวกีฬาที่มีการกล่าวถึงสถานการณ์เกี่ยวกับโรคข้าวหรือเกี่ยวกับข้าว เป็นต้น ดังนั้นจึงต้องคัดกรองข้อมูลที่ไม่เกี่ยวข้องออก โดยสามารถคัดกรองข้อมูลที่เป็นข้อมูลทางด้านการเกษตรให้เหลือ 2,000 ไฟล์

2.2 การคัดเลือกคำและประโยค

เมื่อทำการคัดเลือกไฟล์ข้อมูลที่ต้องการได้แล้ว ต่อมาก็นำไฟล์ทั้งหมดไปตัดคำและนับความถี่ของคำที่ปรากฏ โดยมีคำทั้งหมด 19,049 คำ และเนื่องจาก

ผู้วิจัยสนใจคำกริยา เพื่อนำมาสร้างกรอบในเบื้องต้น จึงได้คัดเลือกเฉพาะคำกริยาออกมา โดยมีจำนวนคำกริยาทั้งหมด 3,217 คำ และเพื่อเป็นการเลือกกลุ่มประชากรให้มีความน่าเชื่อถือ ผู้วิจัยได้นำสูตรการคำนวณหาขนาดของกลุ่มตัวอย่างของทาร์ยามาเน (Taro Yamane) มาใช้ โดยให้ค่าความคลาดเคลื่อนเป็น .10 ดังนี้

$$n = \frac{N}{1 + Ne^2}$$

E	คือ	ความคลาดเคลื่อนของการเลือกตัวอย่าง
N	คือ	ขนาดของประชากร
n	คือ	ขนาดของกลุ่มตัวอย่าง

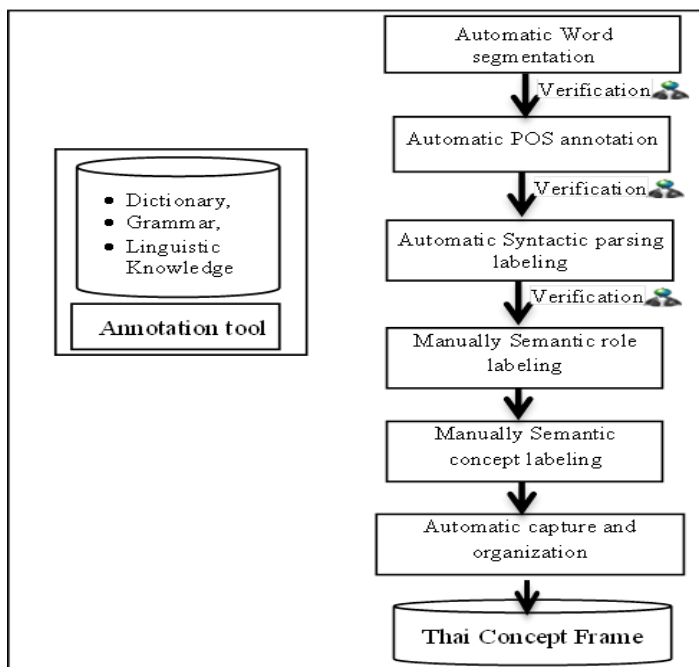
ผลลัพธ์ที่ได้ = 96.99 หมายความว่า ขนาดประชากรค่าที่ต้องใช้ คือ 97 คำ ทั้งนี้ผู้วิจัยได้ปรับจำนวนเป็น 106 คำ โดยผู้วิจัยเลือกคำที่มีความถี่สูงสุดที่ปรากฏในคลังข้อมูล 2,000 ไฟล์ ที่ได้คัดเลือกไว้แล้วตอนต้น ได้แก่ คำว่า มี ใช้ เป็น ทำ ปลุก เลี้ยง หมัก ทำให้ ให้ นำ เข้า ได้ ผสม ป้องกัน กิน เก็บ ช่วย เกิด เพาะ ผลิต กำจัด เตรียม อ่าน เริ่ม เพิ่ม พบ จัด บวก ตัด รักษา ระบาด ฟัน เจริญเติบโต อยู่ บำรุง ติด ส่ง ดูแล ต้องการ สร้าง เอา จำหน่าย แห้ง ขาย ปิด หา ผ่าน เหมาะสม ชม ชื้อ พัฒนา แบ่งปัน โต ดู ข่า สำคัญ นิยม ศึกษา เหลือ ปรับ เปิด วาง หวาน กว้าง เรียนรู้ สมบูรณ์ เต็ม สุข แดง งอก แยก เสริม รดน้ำ ขุด พัก เลือก ระบาย แซ่ ห่าง ตั้ง กล่าว แบ่ง ร่วม กัด ประสบ เปลี่ยน ปรับปรุง ออก บรรจุ ใหญ่ บริโภค ตก เท รอง รับ ปลอดภัย แน่น ต่ำ กลบ ปฏิบัติ ตรวจ ดำเนินงาน เฉลี่ย เผา ผสมพันธุ์ เพาะเลี้ยง

จากนั้นผู้วิจัยได้คัดเลือกประโยคที่มีบริบทต่างกันให้กับคำแต่ละคำ เพื่อสกัดโครงสร้างทางไวยากรณ์และความสัมพันธ์ทางความหมาย ในขั้นต้นผู้วิจัยได้คัดเลือกประโยคที่มีคำใน 106 คำดังกล่าวข้างต้นปรากฏในประโยค โดยใช้วิธีสุ่มตัวอย่างแบบเจาะจงเลือกประโยคที่มีบริบทต่างกันสำหรับคำเป้าหมาย โดยเฉลี่ย 40 ประโยคสำหรับ 1 คำเป้าหมาย แต่เมื่อนำประโยคที่คัดเลือกมาสกัดโครงสร้างกรอบพบว่าคำบางคำมีจำนวนประโยคที่สามารถสร้างเป็นกรอบได้ไม่ถึง 40 ประโยค

ผู้วิจัยจึงเลือกประโยคเพิ่มตามจำนวนที่ปรากฏจริง โดยจำนวนประโยคมีตั้งแต่ 3 ประโยค ถึง 246 ประโยค แล้วจึงนำประโยคที่คัดเลือกทั้งหมดไปกำกับข้อสนเทศในขั้นตอนถัดไป

2.3 กระบวนการสร้างกรอบมโนทัศน์

ในการสร้างกรอบมโนทัศน์ภาษาไทย (Thai concept frame) จำเป็นต้องมีการตัดประโยคเพื่อหาขอบเขตของประโยคที่ปรากฏจริงในเอกสาร และสิ่งที่สำคัญยิ่งไปกว่านั้นคือการกำกับข้อสนเทศของข้อมูล เพื่อวิเคราะห์หาความสัมพันธ์ของคำในประโยค โดยจะต้องมีการกำกับข้อสนเทศ 4 ระดับ คือระดับคำ (ชนิดของคำ) วลี การก และมโนทัศน์ แล้วจึงสกัดกรอบมโนทัศน์ออกมา โดยกระบวนการสร้างกรอบมโนทัศน์มีขั้นตอนดังภาพที่ 1 และมีรายละเอียดดังนี้



ภาพที่ 1. สถาปัตยกรรมของการสร้างกรอบมโนทัศน์

1) การตัดประโยคอัตโนมัติ (Automatic Sentence Segmentation)

การตัดขอบเขตของประโยค ซึ่งเป็นส่วนพื้นฐานที่สำคัญส่วนหนึ่ง เนื่องจากภาษาไทยไม่มีการใช้เครื่องหมายบ่งบอกว่าประโยคสิ้นสุด ณ จุดใด อีกทั้งในกระบวนการสร้างกรอบมโนทัศน์ของเฟรมจำเป็นต้องกำกับส่วนของข้อมูลตั้งแต่ระดับคำ ระดับไวยากรณ์ ระดับความหมาย ภายในประโยคหนึ่งๆ ดังนั้นขอบเขตของประโยคจึงต้องชัดเจน ในขั้นต้นงานวิจัยนี้จึงทำการตัดประโยคแบบอัตโนมัติ (พุทธชาติ โปริบาล และคณะ, 2557) แล้วตรวจสอบความถูกต้องด้วยนักภาษาศาสตร์อีกครั้งหนึ่ง ดังตัวอย่างต่อไปนี้

กรมชลประทานจัดสรรน้ำทำนาปรังในช่วงฤดูแล้งปีนี้เต็มพื้นที่
เฉพาะลุ่มเจ้าพระยา กว่า 8 ล้านไร่
พร้อมปรับแผนใหม่ทำนาปรังเร็วขึ้น

2) การตัดคำอัตโนมัติ (Automatic Word Segmentation)

การประมวลผลภาษาไทยที่ยากมากอีกส่วนหนึ่งก็คือ ขอบเขตของคำ เนื่องจากภาษาไทยเป็นภาษาที่ไม่มีการเว้นวรรคระหว่างคำอย่างภาษาอังกฤษ ทำให้การประมวลผลภาษาไทยต้องมีการตัดคำก่อน แล้วจึงต่อยอดการประมวลผลระดับอื่นต่อไป ซึ่งในงานวิจัยนี้ได้ใช้เครื่องมือตัดคำอัตโนมัติที่พัฒนาโดยหน่วยเชี่ยวชาญเฉพาะด้านการประมวลผลภาษาธรรมชาติและเทคโนโลยีสารสนเทศอัจฉริยะภาคีวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ (Sudprasert S. and Kawtrakul A., 2003) ดังตัวอย่าง

กรม ชลประทาน จัดสรร น้ำ ทำ นาปรัง ใน ช่วง ฤดู แล้ง ปี นี้ เต็ม พื้นที่
เฉพาะ ลุ่ม เจ้าพระยา กว่า 8 ล้าน ไร่

3) การกำกับชนิดของคำอัตโนมัติ (Automatic POS Tagging)

ขั้นตอนนี้เป็นการกำกับชนิดของคำ (part of speech) ได้แก่ คำนาม กริยาวิเศษณ์ คุณศัพท์ บุพบท ฯลฯ ซึ่งสามารถกำกับแบบอัตโนมัติด้วยเครื่องมือที่พัฒนาไว้แล้วโดยหน่วยเชี่ยวชาญเฉพาะด้านการประมวลผลภาษาธรรมชาติและเทคโนโลยี่

สารสนเทศอัจฉริยะ (Kawtrakul A., et al., 2004) อย่างไรก็ตาม การกำกับชนิดของคำแบบอัตโนมัติยังมีข้อผิดพลาดอยู่บ้าง จึงจำเป็นต้องอาศัยนักภาษาศาสตร์ทำการตรวจสอบความถูกต้องอีกครั้งก่อนจะดำเนินการในขั้นตอนต่อไป ดังตัวอย่าง

กรม/ncn ชลประทาน/npn จัดสรร/vt น้ำ/ncn ทำ/vt นาปรัง/ncn ใน/prep ช่วง/
ncn ฤดู/ncn แล้ง/adj ปี/cl นี้/det เต็ม/vt พื้นที่/ncn
เฉพาะ/adv ลุ่ม/ncn เจ้าพระยา/npn กว่า/adv _/blk 8/num _/blk ล้าน/num
ไร่/cl

4) การกำกับโครงสร้างวลี (Syntactic Parsing)

การกำกับโครงสร้างวลีเป็นกระบวนการกำกับประเภทของวลีว่าเป็นนามวลี กริยาวลี วิเศษณ์วลี คุณศัพท์วลี หรือบุพบทวลี เป็นต้น ซึ่งการกำกับโครงสร้างวลีนั้น ผู้วิจัยจึงเลือกใช้ทฤษฎีไวยากรณ์โครงสร้าง เพื่อกำหนดขอบเขตของคำหรือวลีที่สัมพันธ์กันในประโยค โดยมีขั้นตอนดังนี้

4.1) แยกภาคประธานและภาคแสดงออกจากกัน

4.2) วิเคราะห์โครงสร้างและกำกับชนิดของวลี เช่น เป็น นามวลี (NP) กริยาวลี (VP) วิเศษณ์วลี (ADVP) คุณศัพท์วลี (ADJP) หรือ บุพบทวลี (PP) เป็นต้น

4.3) ใช้วงเล็บครอบ [.....]/ เพื่อให้เห็นโครงสร้างภายในประโยคอย่างชัดเจน เช่น [.....]/NP [.....]/VP

ในการกำกับโครงสร้างวลี จะทำการกำกับซ้อนหน้าที่ของคำอีกชั้นหนึ่ง โดยในประโยคที่กำกับจะปรากฏทั้งหน้าที่ของคำและประเภทของวลี ดังตัวอย่าง

[กรม/ncn ชลประทาน/npn]/NP [[จัดสรร/vt น้ำ/ncn]/VP [ทำ/vt [นาปรัง/ncn]/NP
[ใน/prep [ช่วง/ncn ฤดู/ncn แล้ง/adj [ปี/cl นี้/det]/NP]/NP]/PP]/VP เต็ม/vt พื้นที่/
ncn]/VP

เมื่อกำกับโครงสร้างวลีเสร็จแล้วก็จะนำข้อมูลที่กำกับหน้าที่ของคำและโครงสร้างวลีแล้วไปกำกับบทบาททางความหมาย (semantic roles) ของคำหรือวลี ซึ่งเป็นความสัมพันธ์ทางความหมายต่อไป

5) การกำกับบทบาททางความหมายของคำ (Manually Semantic Role Labeling)

การกำกับบทบาททางความหมาย (thematic roles) ของคำหรือวลีเป็นการกำกับการ (case) ซึ่งงานวิจัยนี้จะใช้การกตาม Fillmore (1971) 9 การก ได้แก่ Agent, Instrument, Experiencer, Object, Location, Source, Goal, Time และ Benefactor และได้เพิ่มในส่วนที่ผู้วิจัยเห็นว่ากรกทั้ง 9 ของฟิลล์มอร์ยังไม่เพียงพออีก 3 การก โดยเป็นการกตามแนวคิดของลาร์สัน (Larson, 1984) ได้แก่ Accompaniment, Manner และ Measure

ไวยากรณ์การกเป็นระบบการวิเคราะห์ทางภาษาศาสตร์ที่มุ่งเน้นการเชื่อมโยงระหว่างคำกริยากับส่วนต่างๆ ที่เป็นหน่วยทางภาษา (argument) ในประโยคหรือส่วนที่เป็นประธาน กรรม เป็นต้น ในการเชื่อมโยงนี้จำเป็นต้องอิงบริบทของคำที่ปรากฏร่วมกัน ซึ่งเป็นบริบททางไวยากรณ์ โดยอาศัยบทบาททางความหมาย (thematic roles) เป็นเกณฑ์ในการกำหนดความสัมพันธ์

6) การกำกับโน้ตทัศน์ทางความหมาย (Manually Semantic Concept Labeling)

ขั้นตอนนี้เป็นการกำกับโน้ตทัศน์ทางความหมายให้กับคำหรือวลีที่สัมพันธ์กัน โดยกำกับไว้หลังการก เพื่อบ่งบอกว่าการกที่ปรากฏในประโยคนั้นเป็นการกที่เกี่ยวกับกลุ่มความหมายใด โดยกลุ่มความหมายที่ใช้ในงานวิจัยนี้เป็นกลุ่มความหมายจาก WordNet (Fellbaum, 1998) ซึ่งเป็นกลุ่มความหมายที่มีการจัดโครงสร้างความสัมพันธ์เรียบร้อยแล้ว

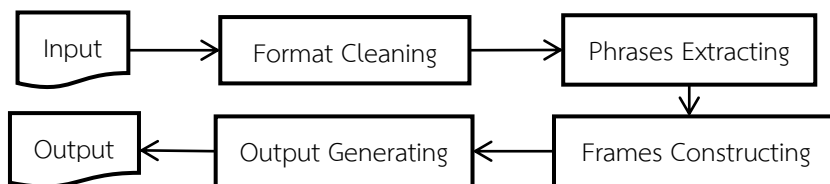
7) การดึงและจัดระเบียบข้อมูล (Automatic Capture and Organization)

ขั้นตอนนี้เป็นการดึงข้อมูลที่มีการกำกับโครงสร้างของวลี (phrase type) การก (case) และมโนทัศน์ทางความหมาย (concept) โดยใช้เครื่องมือ Automatic capture and organization (พุทธรชาติ โปธิบาล และคณะ, 2557) ดึงข้อมูล พร้อมทั้งจัดโครงสร้างเฟรมให้อยู่ในรูปแบบที่กำหนดได้แบบอัตโนมัติ โดยเครื่องมือนี้สามารถทำให้สามารถสกัดข้อสนเทศและจัดระเบียบข้อมูลได้อย่างรวดเร็วและสามารถช่วย

ตรวจสอบข้อผิดพลาดเรื่องความสอดคล้องของข้อมูล หรือความไม่ลงรอยในการกำกับข้อมูลสารสนเทศได้ด้วย

2.4 กระบวนการสร้างเฟรม

กระบวนการสร้างเฟรมจากคลังข้อมูลที่ได้รับการกำกับข้อสนเทศของคำ วลี บทบาททางความหมายและมโนทัศน์ของคำแล้ว ประกอบด้วยขั้นตอนต่าง ๆ ดังต่อไปนี้



ภาพที่ 2. แสดงผังการทำงานของระบบ

1) การปรับรูปแบบ (Format Cleaning)

ขั้นตอนนี้เป็นการปรับรูปแบบการกำกับที่ไม่ถูกต้องบางส่วนให้ถูกต้องก่อน ซึ่งรูปแบบในที่นี้หมายถึงส่วนที่เกี่ยวกับช่องว่างและเครื่องหมายบางอย่าง โดยไม่รวมถึงป้ายกำกับ (tag)

2) การสกัดวลี (Phrases Extracting)

หลังจากมีการปรับแก้รูปแบบการกำกับแล้ว ระบบจะทำการสกัดแยกวลีออกเป็นวลีเดี่ยวพร้อมป้ายกำกับ (tag) ว่าเป็นนามวลี (Noun Phrase: NP) กริยาวลี (Verb Phrase: VP) (Preposition Phrase: PP) (Adverb Phrase: ADVP) (Adjective Phrase: ADJP) โดยในขั้นตอนนี้จะได้ผลลัพธ์ออกมาเป็นรายการคู่ของวลีกับป้ายกำกับ หนึ่งวลีต่อหนึ่งป้ายกำกับ โดยเรียงจากป้ายกำกับชั้นในสุดออกไปหาชั้นนอกสุด ดังตัวอย่าง

[[กิ่ง/ncn [พันธุ์/ncn แก้วมังกร/npn]/NP]/NP [แบบ/ncn [ใน/prep [ทุ่ง/ncn
 ฟ้า/npn]/NP]/PP]/NP]/NP

จากตัวอย่างข้างต้นสามารถแยกวลีออกเป็นวลีเดี่ยวได้ดังนี้

- [กิ่ง/ncn พันธุ์/ncn แก้วมังกร/npn แบบ/ncn ใน/prep ทุ่ง/ncn ฟ้า/npn]/NP
- [กิ่ง/ncn พันธุ์/ncn แก้วมังกร/npn]/NP
- [พันธุ์/ncn แก้วมังกร/npn]/NP
- [แบบ/ncn ใน/prep ทุ่ง/ncn ฟ้า/npn]/NP
- [ใน/prep [ทุ่ง/ncn ฟ้า/npn]/NP]/PP
- [ทุ่ง/ncn ฟ้า/npn]/NP

3) การสร้างกรอบ (Frames Constructing)

ขั้นตอนนี้เป็นการสร้างกรอบหรือเฟรมจากกริยาวลีทั้งหมดในประโยค โดยเริ่มพิจารณาวลีกับป้ายกำกับจากหลังมาหน้า หากพบกริยาวลี จะทำการสร้างเฟรมของชุดคำกริยาแรกในวลีนั้น จากนั้นหากพบป้ายกำกับบทบาท (case) และความหมาย (concept) ภายในกริยาวลีนั้น จะถูกเพิ่มเข้าไปในเฟรมของชุดคำกริยานั้น หากพบกริยาวลีซ้อนอยู่ในกริยาวลี จะพักเฟรมของกริยาวลีแรกไว้แล้วเริ่มสร้างเฟรมใหม่สำหรับชุดคำกริยาใหม่ของกริยาวลีย่อยที่ซ้อนอยู่ ป้ายกำกับบทบาท และความหมายที่พบในกริยาวลีย่อยจะถูกเพิ่มเข้ากับเฟรมใหม่ แต่ไม่ถูกเพิ่มลงในเฟรมของกริยาวลีแรก กล่าวคือ ป้ายกำกับบทบาทและความหมายจะถูกเพิ่มลงในเฟรมของกริยาวลีที่มีมันปรากฏอยู่โดยตรงเท่านั้น เฟรมของกริยาวลีใหญ่และกริยาวลีย่อยจะถูกแยกออกจากกันอย่างชัดเจน

หนึ่งสำหรับป้ายกำกับบทบาทและความหมายที่อยู่ในประโยค แต่อยู่นอกเหนือกริยาวลีใดๆ จะถูกกันไว้เป็นส่วนกลางและถือว่าเป็นส่วนประกอบของเฟรมทั้งหมดในประโยคนั้น โดยจะได้รับการเพิ่มให้ทุกๆ เฟรมในประโยค หลังจากพิจารณาแล้วทั้งหมดในประโยคเรียบร้อยแล้ว สุดท้ายเขียนผลลัพธ์ลงในไฟล์ โดยเริ่มจากประโยคตามด้วยเฟรมที่ได้จากประโยคนั้นๆ หากไม่สามารถสร้างเฟรมจากประโยคนั้นได้ จะแสดงข้อผิดพลาดที่พบแทน ดังตัวอย่าง

ซึ่ง/conj [[[ทาง/prep [[สวน/ncn พรรณวดี/hpn]/NP]/Group]/PP]/AGT [จะ/
prev ปลุก/vt [[[ตะไคร้/hpn]/NP]/OBJ]/Lemongrass [[[ตาม/prep [[[หลัก/ncn
วิชา/ncn การเกษตร/ncn]/NP [ผสาน/vi [กับ/prep [ภูมิปัญญา/ncn ท้องถิ่น/ncn
]/NP]/PP]/VP]/NP]/Principle]/PP]/MAN]/VP]/S _/blk คือ/vcs

ERROR: role without concept. =>MAN

จากตัวอย่างข้างต้นแสดงว่า ประโยคนี้ไม่สามารถสร้างกรอบมโนทัศน์ได้ เนื่องจากไม่มีมโนทัศน์ของนามวลีในส่วนที่ปรากฏอยู่ในบุพพทวลี ซึ่งมีบทบาทเป็น Manner ทำให้ไม่สามารถสร้างกรอบมโนทัศน์ได้ เนื่องจากข้อมูลไม่สมบูรณ์นั่นเอง ดังนั้นระบบจึงคืนผลลัพธ์เป็นข้อผิดพลาดที่เกิดขึ้นดังตัวอย่างข้างต้น

4) การแสดงผลลัพธ์ (Output Generating)

ขั้นตอนนี้เป็นขั้นตอนรวมเฟรมของคำกริยาเดียวกันที่สกัดได้จากแต่ละประโยค เช่น ถ้าเกิดคำกริยา “ปลุก” ขึ้นในคลังข้อมูล 10 ครั้ง แต่ละครั้งจะมีการสร้างเฟรมในแต่ละประโยค ซึ่งการสร้างเฟรมจะทำได้ หากข้อมูลที่ได้รับการกำกับข้อสนเทศมีความถูกต้อง หลังจากสร้างเฟรมในแต่ละประโยคแล้ว จะมีการรวมเฟรมทั้งหมดที่สกัดได้จากไฟล์ให้เป็นเฟรมเดียวแล้วเขียนลงไฟล์ โดยเฟรมรวมนี้จะแสดงรูปคำกริยาของทุกเฟรมและบทบาทกับความหมายทั้งหมดที่ปรากฏในบทพาทนั้น

Summary :

=====

Verb : ปลุก

Role :

LOC {Area, Family}

MEA {Area, Tree, Type}

OBJ {Tree, Cassava, Plant, Rice, Rattan, Branch, Bamboo,
Longan, Rambutan, Papaya, Food, Fruit_tree, Vegetable}

TIME {Present, Rainy_season, Time}

3. ผลการวิจัย

ผลการวิจัยแบ่งเป็น 2 หัวข้อใหญ่ คือ การวิเคราะห์ความสัมพันธ์ของคำกริยากับคำบริบทในภาษาไทยและการสร้างกรอบมโนทัศน์ภาษาไทย ซึ่งในหัวข้อแรกจะเน้นผลการวิจัยเชิงปริมาณ ส่วนหัวข้อหลังจะเน้นกระบวนการสร้างกรอบมโนทัศน์และการนำไปใช้งาน เป็นผลการวิจัยเชิงคุณภาพ โดยมีรายละเอียดดังนี้

3.1 การวิเคราะห์ความสัมพันธ์ของคำกริยากับคำบริบทในภาษาไทย

จากผลการกำกับข้อสนเทศทางภาษาจำนวน 5,784 ประโยค โดยคัดเลือกประโยคที่มีคำกริยาที่มีความถี่ในการใช้มากที่สุดในกลุ่มข้อมูลด้านการเกษตร จำนวน 106 คำ และกำกับข้อสนเทศตั้งแต่ระดับคำ วลี การก จนถึงการกำกับมโนทัศน์ทางความหมายของคำหรือวลีแล้ว สามารถสกัดกรอบมโนทัศน์หรือ concept frame ออกมาได้ทั้งหมด 962 กรอบ โดยมีรายละเอียดดังนี้

ตารางที่ 2

แสดงการกำกับข้อสนเทศทางภาษาของคำกริยาที่มีความถี่ในการใช้มากที่สุดในกลุ่มข้อมูลด้านการเกษตร

ลำดับ	คำ	ความถี่ ในการปรากฏ	จำนวน			
			ประโยค	กรอบ	การก	การกสูงสุด
1	มี	7317	246	32	11	OBJ
2	ใช้	5036	212	31	11	OBJ
3	เป็น	4613	245	32	12	OBJ
4	ทำ	3465	64	18	11	OBJ
5	ปลูก	3276	85	20	9	OBJ
6	เลี้ยง	1888	55	9	7	OBJ
7	หมัก	1754	40	9	7	OBJ
8	ทำให้	1560	92	13	9	OBJ
9	ให้	1535	51	17	10	OBJ
10	นำ	1352	37	8	8	OBJ

ลำดับ	คำ	ความถี่ ในการปรากฏ	จำนวน			
			ประโยค	กรอบ	การก	การกสูงสุด
11	เข้า	1269	59	13	7	OBJ
12	ได้	1188	46	8	7	OBJ
13	ผสม	1093	53	13	9	OBJ
14	ป้องกัน	1014	50	10	7	OBJ
15	กิน	964	95	21	9	AGT, OBJ
16	เก็บ	910	98	7	5	OBJ
17	ช่วย	860	46	6	5	AGT+V2-OBJ
18	เกิด	852	50	15	9	OBJ
19	เพาะ	808	33	10	6	OBJ
20	ผลิต	805	40	16	10	OBJ
21	กำจัด	735	14	2	2	OBJ
22	เตรียม	681	27	4	5	OBJ
23	อ่าน	672	3	3	3	OBJ
24	เริ่ม	649	55	11	7	AGT
25	เพิ่ม	635	46	13	8	OBJ
26	พบ	631	38	8	5	OBJ
27	ฉีด	605	18	7	6	OBJ
28	บวก	587	4	1	2	AGT, OBJ
29	ตัด	571	60	10	7	OBJ
30	รักษา	568	41	13	9	OBJ
31	ระบาค	524	43	10	7	EXP
32	พน	510	41	9	6	OBJ, INS
33	เจริญเติบโต	507	56	7	7	EXP
34	อยู่	499	51	7	4	EXP
35	บำรุง	492	12	6	6	OBJ
36	ติด	464	43	10	6	OBJ
37	ส่ง	458	28	11	9	LOC, OBJ
38	ดูแล	446	20	6	6	OBJ
39	ต้องการ	444	22	5	5	OBJ
40	สร้าง	429	39	11	10	OBJ
41	เอา	409	67	8	7	OBJ
42	จำหน่าย	403	16	11	8	OBJ, LOC
43	แห่ง	385	55	5	5	EXP

ลำดับ	คำ	ความถี่ ในการปรากฏ	จำนวน			
			ประโยค	กรอบ	การก	การกสูงสุด
44	ขาย	383	46	19	9	OBJ
45	ปิด	360	40	4	3	OBJ
46	หา	353	52	12	8	OBJ
47	ผ่าน	333	38	18	12	OBJ
48	เหมาะสม	317	30	8	5	EXP
49	ชม	299	8	4	3	AGT
50	ซื้อ	292	69	10	8	OBJ
51	พัฒนา	283	58	7	5	OBJ
52	แบ่งปัน	279	6	3	3	OBJ
53	โต	274	56	4	4	EXP
54	ดู	261	65	7	5	OBJ
55	ฆ่า	260	58	8	5	OBJ
56	สำคัญ	255	17	2	2	EXP
57	นิยม	251	18	8	6	OBJ
58	ศึกษา	249	35	10	7	OBJ
59	เหลือ	249	199	10	6	MEA
60	ปรับ	248	27	3	3	OBJ
61	เปิด	248	20	2	2	OBJ
62	วาง	228	94	9	7	OBJ
63	หว่าน	228	22	12	8	OBJ
64	กว้าง	225	34	3	2	EXP
65	เรียนรู้	222	24	8	7	OBJ
66	สมบูรณ์	221	122	9	6	EXP
67	เติม	218	82	6	4	OBJ
68	สนุก	217	38	5	5	EXP
69	แตก	215	89	15	7	EXP
70	งอก	214	85	12	7	EXP
71	แยก	213	77	9	8	OBJ
72	เสริม	210	61	15	10	OBJ
73	รดน้ำ	208	65	8	6	OBJ, MAN
74	ขุด	201	25	5	4	OBJ
75	ปัก	201	64	5	4	EXP
76	เลือก	201	47	9	8	OBJ

ลำดับ	คำ	ความถี่ ในการปรากฏ	จำนวน			
			ประโยค	กรอบ	การก	การกสูงสุด
77	ระบาย	198	12	7	6	OBJ
78	แซ่	198	51	6	5	LOC
79	ห้วง	197	25	8	5	EXP, MEA
80	ตั้ง	196	21	8	6	OBJ
81	กล่าวว	192	23	8	5	AGT, OBJ
82	แบ่ง	190	53	12	6	OBJ
83	รวม	186	64	12	7	ACC
84	กัด	184	66	5	5	AGT, OBJ
85	ประสบ	183	62	4	5	OBJ
86	เปลี่ยน	183	17	10	8	OBJ
87	ปรับปรุง	180	59	7	6	OBJ
88	ออก	178	78	18	9	AGT, OBJ
89	บรรจุ	175	64	7	5	OBJ
90	ใหญ่	174	66	5	5	EXP
91	บริโภค	172	15	5	6	AGT, OBJ
92	ตก	170	45	12	7	EXP
93	เท	169	69	2	2	OBJ
94	รอง	164	56	5	4	OBJ
95	รับ	164	61	7	7	OBJ
96	ปลด	161	62	5	5	OBJ
97	แน่น	161	26	6	4	OBJ
98	ต่ำ	157	68	2	3	EXP
99	ปฏิบัติ	155	157	7	5	SOU, OBJ
100	ตรวจ	154	37	4	3	OBJ
101	ตรวจ	154	32	4	3	OBJ
102	ต้านทาน	152	70	4	4	OBJ
103	เฉลี่ย	152	43	3	2	EXP
104	เผา	152	61	8	7	OBJ
105	ผสมพันธุ์	139	40	10	5	AGT
106	เพาะเลี้ยง	124	14	6	7	OBJ

จากตารางข้างต้นสามารถอธิบายรายละเอียดได้ดังนี้

3.1.1 ความสัมพันธ์ระหว่างความถี่ในการปรากฏคำกริยากับการสร้างกรอบมโนทัศน์

ความถี่ของคำเป็นตัวกำหนดที่นิยมใช้มากในการประมวลผลภาษาธรรมชาติอย่างหนึ่ง โดยเฉพาะแนวทางการประมวลผลภาษาตามทฤษฎีและวิธีการทางด้านการรู้จำคำ (word recognition) ในงานวิจัยนี้ตั้งสมมุติฐานว่า คำที่มีความถี่ในการใช้สูงจะทำให้สามารถสร้าง (evoke) กรอบมโนทัศน์ได้มาก จากผลการวิจัยพบว่าความถี่ของคำกริยาเป้าหมายที่ปรากฏในคลังข้อมูลที่น่ามาใช้ในการวิจัยนี้มีความสัมพันธ์กับจำนวนกรอบมโนทัศน์ที่สามารถสร้างได้ในอันดับสูงสุดเท่านั้น โดยสังเกตได้จากตารางข้างต้นจะเห็นว่าคำว่า *มี* มีความถี่ในการปรากฏในคลังข้อมูลมากที่สุดและมีจำนวนประโยคที่ปรากฏมากที่สุด รวมถึงมีกรอบมโนทัศน์มากที่สุดด้วย แต่อย่างไรก็ตาม เมื่อสำรวจความถี่ของคำอื่นๆ พบว่าคำที่มีความถี่สูงในการปรากฏในเอกสารบางคำก็ไม่สามารถนำมาสร้างเป็นกรอบมโนทัศน์ได้เนื่องจากคำเหล่านั้นปรากฏในโครงสร้างนามวลีเป็นส่วนใหญ่ และคำบางคำมีการละส่วนสำคัญของประโยค เช่น ประธานและกรรม ทำให้ไม่สามารถสร้างกรอบการได้เช่นกัน ดังตัวอย่าง ประโยคที่ประกอบด้วยกริยาหลักคือ “อ่าน” ซึ่งปรากฏในคลังข้อมูลถึง 672 ครั้ง โดยส่วนใหญ่ปรากฏในนามวลี ส่วนที่ปรากฏเป็นกริยาแก่นของประโยคมีอยู่ 3 ครั้ง ตรงข้ามกับคำบางคำที่ความถี่ในการปรากฏไม่สูงมากนักแต่สามารถสร้างกรอบมโนทัศน์ได้จำนวนมาก

3.1.2 ความสัมพันธ์ระหว่างการกกับการสร้างกรอบมโนทัศน์

เนื่องจากงานวิจัยนี้เป็นการสร้างกรอบการกโดยพิจารณาความสัมพันธ์ของคำจากประโยคในระดับโครงสร้างผิวและประโยคที่มีลักษณะเป็นอนุพากย์เดี่ยวหรือเป็นประโยคความเดียว จำนวนการกที่ปรากฏในแต่ละประโยคนั้นพบว่ามีจำนวน 1-5 การก ซึ่งเมื่อเทียบกับการทดลองของฟิลล์มอร์จะเห็นว่า ภาษาไทยมีจำนวนการกที่ปรากฏเป็นความสัมพันธ์ของคำในประโยคมากกว่าของฟิลล์มอร์ 2 การก ดังตัวอย่างคำว่า “ทำ” ต่อไปนี้

[[[[โรค/ncn แผลง/ncn]/NP]/AGT]/Decease [ทำ/vt [[[ความ/pref1 เสียหาย/
vi]/NP]/OBJ]/Damage ให้/conj [แก่/prep [[การ/pref1 ปลูก/vt ข้าว/ncn]/BEN]/
Crop]/PP [ใน/prep [[[ปี/ncn หนึ่ง/nnum]/NP]/TIME]/Year]/PP [เป็น/vcs
[[[จำนวน/ncn]/NP]/MEA]/Quantity]/VP]/S

ทำ

AGT {Decease}

OBJ {Damage}

BEN {Crop}

TIME {Year}

MEA {Quantity}

ในแง่ของจำนวนการกทั้งหมดที่เป็นไปได้ที่จะปรากฏร่วมกับคำแต่ละคำนั้น จำนวนการกที่สามารถปรากฏร่วมกับคำหนึ่งคำโดยสรุปจากทุกประโยคที่คำกริยา เป้าหมายปรากฏพบว่ามีจำนวนการกที่เป็นไปได้ในการปรากฏต่อคำสูงสุด 12 การก โดยปรากฏร่วมกับคำว่า “เป็น” และ “ผ่าน”

ส่วนจำนวนกรอบมโนทัศน์นั้น จากผลการวิเคราะห์พบว่าจำนวนกรอบ ขึ้นอยู่กับการเรียงลำดับของการกในประโยค จำนวนกรอบมโนทัศน์มีมากหรือน้อย ไม่ได้ขึ้นอยู่กับความถี่ของคำที่ปรากฏในเอกสารแต่เพียงอย่างเดียว แต่ขึ้นอยู่กับ ตำแหน่งที่ปรากฏด้วย

คำที่มีความถี่ในการปรากฏในเอกสารมาก เช่น คำว่า “เป็น มี” เป็นต้น จะมีบริบทในการเกิดที่แตกต่างกันมาก มักจะทำให้มโนทัศน์ที่ปรากฏร่วมกับคำกริยามี ลักษณะหลากหลาย และคำนามที่ปรากฏร่วมจะมีลักษณะเป็นคำนามที่กระจายอยู่ใน วงคำศัพท์หลาย ๆ วง สามารถจัดกลุ่มความหมายของคำได้หลายความหมาย แม้จะ มีการกเดียวกัน ส่วนคำที่อยู่ในวงคำศัพท์ทางการเกษตร เช่น “ปลูก เลี้ยง หมัก ฯลฯ” ซึ่งปรากฏในเอกสารเป็นจำนวนมากเช่นกัน จำนวนการกมีไม่มากและจะปรากฏกับ มโนทัศน์ในหมวดหรือวงคำศัพท์เดียวกัน ส่งผลให้ความหมายของคำเป็น ความหมายแคบ จัดกลุ่มความหมายได้ง่าย

3.1.3 การสำคัญหรือองค์ประกอบแก่น (coreness)

ฟิลล์มอร์ได้กำหนดกรอบการก โดยมีคำกริยาเป็นแกนตามกฎโครงสร้างพื้นฐานของประโยค ซึ่งฟิลล์มอร์ให้ความสำคัญกับการกผู้ถูกกระทำว่าเป็นการกที่สำคัญที่สุด แม้ในประโยคที่มีคำกริยาบอกสภาพ (stative verb) เช่น to be red และคำกริยาแสดงกระบวนการ เช่น to become red ก็ถือว่าคำนามที่ปรากฏร่วมกับคำกริยาในประโยคเป็นการกผู้ถูกกระทำเช่นกัน (H. V. D., 1995) อย่างไรก็ตามหากพิจารณาในแนวการวิเคราะห์ภาษาของวิจินต์ ภาณุพงศ์ (2532) หน่วยกรรมกริยานั้นจะไม่มีหน่วยกรรมตามหลัง เช่น น้องร้องไห้ น้องเป็นประธาน และไม่ปรากฏกรรมในประโยคกรรมกริยา อย่างไรก็ตาม เมื่อตรวจสอบคำที่เป็น stative verbs ใน FrameNet (Online) พบว่าปัจจุบันมีการมองว่าการกที่ปรากฏร่วมกับ stative verbs เป็นการก Perceiver_agentive ดังตัวอย่างต่อไปนี้

[Perceiver_agentive He] SMELLED^{Target} [Phenomenon the burning in the air], saw the stars dimmed by smoke above him.²

สำหรับงานวิจัยนี้แม้ว่าจะยึดหลักการวิเคราะห์การกตามแนวคิดของฟิลล์มอร์ (1968, 1971) แต่ก็มี การปรับให้เข้ากับบริบทภาษาไทยตามหลักการปรากฏกรรมของวิจินต์ ภาณุพงศ์ (2532) และมองว่าฟิลล์มอร์เองก็มีการปรับเปลี่ยนแนวคิดเช่นกัน โดยผู้วิจัยวิเคราะห์ให้คำนามที่ปรากฏร่วมกับคำกริยาแสดงสภาพ เป็นได้ทั้งการกผู้กระทำ (Agent) การกผู้ประสบการณหรือการกของสิ่งที่รับรู้การกระทำ (Experiencer) และการกผู้ที่เกี่ยวข้อง (Accompaniment) เช่น

ข้าวในนาแปลงนี้]/Agent เจริญเติบโตดี
[เกษตรกร]/Experiencer ชอบข้าวหอมมะลิ
เขาเชื่อ[เธอ]/Accompaniment

ดังนั้นในงานวิจัยนี้จึงไม่ได้กำหนดว่า การกใดเป็นหน่วยแก่นของกรอบการกตายตัว ขึ้นอยู่กับการปรากฏในประโยคและบทบาทที่คำนามนั้นแสดงมากกว่า

² ที่มา: [https://framenet2.icsi.berkeley.edu/fnReports/data/lu/lu1311.xml?](https://framenet2.icsi.berkeley.edu/fnReports/data/lu/lu1311.xml?mode=annotation)

อย่างไรก็ตาม เมื่อพิจารณาจำนวนการกที่ปรากฏร่วมกับคำกริยาที่ใช้ในงานวิจัยนี้พบว่าการกรกรรม (Object) เป็นการกที่ปรากฏร่วมกับคำกริยามากที่สุด โดยสามารถเรียงลำดับความถี่ของการกที่ปรากฏร่วมกับคำกริยาซึ่งมีจำนวนมากที่สุดสำหรับคำกริยาแต่ละคำ ตามลำดับดังนี้

ตารางที่ 3

แสดงความถี่ในการปรากฏสูงสุดของการกต่อคำกริยา

การก	ความถี่ในการปรากฏสูงสุดของการก/คำกริยา
OBJ	82
EXP	18
AGT	10
LOC	3
MEA	2
ACC	1
INS	1
MAN	1
SOU	1

จากข้อมูลข้างต้นพบว่า คำที่เป็นกรรมกริยาทั้งหมดต้องการการกรกรรม แต่เนื่องจากภาษาไทยสามารถละการกรกรรมได้ จึงพบว่าการกรกรรมไม่ได้ปรากฏรูปคำที่ก่อให้เกิดกรอบการกของกรรมกริยาร่วมกับการกรกรรมผู้กระทำทั้งหมดเมื่อดูจากประโยคที่เป็นโครงสร้างผิว ดังนั้นคำกรรมกริยาบางครั้งจึงปรากฏเพียงการกผู้กระทำเพียงอย่างเดียว และบางครั้งก็ปรากฏเพียงการกรกรรม โดยไม่ปรากฏการกผู้กระทำร่วมด้วย อย่างไรก็ตามยังมีกริยาบางส่วนที่ต้องการทั้งการกผู้กระทำและการกรกรรม โดยทั้งสองส่วนมีความถี่ในการปรากฏร่วมกับคำกริยามากพอๆ กัน ได้แก่ คำว่า กิน ออก บริโภค กล่าว กัด และ บวก จากจำนวนการปรากฏร่วมกันของการกรกรรมกับกรรมกริยานี้ ทำให้เห็นว่าการกรกรรมเป็นการกที่สำคัญสำหรับกรรมกริยา เป็นองค์ประกอบที่สำคัญ อย่างไรก็ตาม หากมองในแง่ของความเป็นองค์ประกอบแก่น (coreness) ตามมุมมองของเฟรมเน็ต ในงานวิจัยนี้พบว่าในภาษาไทยไม่ปรากฏการกจำเป็นในโครงสร้างผิวที่ชัดเจน ทำให้ไม่สามารถกำหนดได้

องค์ประกอบกันซึ่งเป็นการที่ต้องปรากฏเสมอในประโยคหรือขาดไม่ได้ สำหรับคำกริยาประเภทต่าง ๆ ถึงแม้ว่าการกรกรรรมจะมีความถี่ในการปรากฏร่วมกับกรรมกริยามากที่สุด แต่ก็ไม่ได้หมายความว่า จะปรากฏในทุกประโยค เพราะภาษาไทยเป็นภาษาที่ยอมรับให้เกิดการละการกผู้กระทำหรือการกรกรรรมในประโยคได้

3.1.4 ความสัมพันธ์ระหว่างประเภทของคำกริยากับประเภทการก

คุก (Cook, 1989, pp. 194-198) ได้วิเคราะห์การปรากฏร่วมกันระหว่างประเภทของคำกริยากับประเภทของการกของฟิลล์มอร์ (Cook, 1989 cited Fillmore, 1968) และ เชฟ (Cook, 1989 cited in Chafe, 1970) ว่า กริยาประเภทต่าง ๆ เช่น กริยาแสดงสภาพ (state) กริยาแสดงกระบวนการ (process) และกริยาแสดงอาการหรือการกระทำ (action verbs) จะมีขอบเขตจำกัดในการปรากฏร่วมกัน เช่น กริยาแสดงอาการ kill จะปรากฏกับการกผู้กระทำ (Agent) และการกรกรรรม (Object) เสมอ ส่วนกริยาแสดงอาการ give ซึ่งเป็นกริยาที่มีผู้รับผลประโยชน์เข้ามาเกี่ยวข้องด้วยก็จะปรากฏร่วมกับการกผู้กระทำ (Agent) การกรกรรรม (Object) และการกผู้รับผลประโยชน์ (Benefactive) ร่วมด้วย สอดคล้องกับงานวิจัยในโครงการนี้ ซึ่งพบว่าคำกริยาแต่ละประเภทจะปรากฏร่วมกับการกเฉพาะบางการกเท่านั้น เช่น คำกริยาแสดงสภาพเป็นกลุ่มคำกริยาที่มีความสัมพันธ์กับการกผู้ประสบ (Experiencer) เช่น แดง งอก ตก ระเบิด เจริญเติบโต สมบูรณ์ เหมาะสม ใหญ่ แห้ง สุก อยู่ พัก โต ต่ำ เป็นต้น ส่วนกริยาแสดงอาการจะสัมพันธ์กับการกผู้กระทำ (Agent) และการกรกรรรม (Object) ซึ่งมีความสำคัญสำหรับทั้งอกรรมกริยาและกรรมกริยา เช่น เริ่ม ผสมพันธุ์ ชม กิน ออก บริโภค กล่าว กัด บวก เป็นต้น ส่วนกริยาแสดงการเคลื่อนที่ จะสัมพันธ์กับการกสถานที่ (Location) เช่น ส่ง จำหน่าย เป็นต้น สำหรับกริยาที่มีลักษณะเป็นกระบวนการที่ต้องการการกเครื่องมือ (Instrument) สำหรับการกอื่นๆ ก็มีความสำคัญกับกริยาแต่ละประเภทตามแต่ความจำเป็นในประโยคต่างๆ

3.2 การสร้างกรอบมโนทัศน์ภาษาไทย

การสร้างกรอบมโนทัศน์เป็นเรื่องที่มีความซับซ้อน อาศัยการวิเคราะห์ข้อมูลหลายส่วนดังนี้

3.2.1 การวิเคราะห์ความสัมพันธ์ของคำ

เพื่อให้ระบบประมวลผลภาษาธรรมชาติ (Natural Language Processing) สามารถวิเคราะห์ความหมายและสามารถตอบคำถามได้ว่า ใคร ทำอะไร ให้ใคร การวิเคราะห์ภาษาโดยการใช้ทฤษฎีด้านไวยากรณ์อย่างเดียวนั้นไม่สามารถที่จะตอบคำถามเหล่านี้ได้ทั้งหมด ดังนั้น ทฤษฎีด้านบทบาททางความหมายจึงถูกนำมาใช้ร่วมกับทฤษฎีทางไวยากรณ์ งานวิจัยนี้ใช้เครื่องมือช่วยในการสกัดข้อมูลที่มีการกำกับและตรวจสอบข้อสนเทศทางภาษาแล้วและนำมาเรียบเรียงใหม่ (Automatic Frame Capturing and Organization) (พุทธรชาติ โปธิบาล และคณะ, 2557) โดยเครื่องมือนี้จะทำให้นักวิจัยมองเห็นความสัมพันธ์ของคำหรือวลีต่าง ๆ ในประโยค เช่น ความสัมพันธ์ระหว่างประธาน กริยา และส่วนขยายของประโยค ทั้งทางด้านความหมายและด้านไวยากรณ์ โดยมีการระบุมโนทัศน์ (semantic concept) ให้กับหน่วยประกอบทุกหน่วยในประโยค เพื่อเป็นส่วนสำคัญในการระบุมโนทัศน์ของกรอบสำหรับคำกริยาแต่ละคำ ซึ่งคำกริยาเดียวกัน หากมีความหมายต่างกัน มโนทัศน์ที่ปรากฏจะแตกต่างกันไปด้วย ดังตัวอย่างต่อไปนี้

- [ควร/prev มี/vt บ่อ/ncn เฉพาะ/adj สำหรับ/conj [รักษา/vt [[โรค/ncn]/NP]/OBJ]/Disease]/VP]/S
- [เพื่อ/conj [รักษา/vt [[ความชื้น/ncn ของ/prep รอย/ncn แผล/ncn]/NP]/OBJ]/Moisture]/VP ก่อน/prep การ/pref1 นำ/vi ไป/vpost ปลุก/vt]/S
- [ถ้า/conj สงสัย/vt ว่า/conjncld _/blk เกิด/vcs จาก/prep การ/pref1 ติด/vt เชื้อ/ncn แบคทีเรีย/ncn _/blk ควร/prev ปรับปรุง/vt คุณภาพ/ncn ของ/prep น้ำ/ncn _/blk และ/conj [รักษา/vt [[ปลา/ncn]/NP]/OBJ]/fish [ด้วย/prep [[ยา/ncn ปฏิชีวนะ/ncn]/NP]/INS]/antibiotic]/PP]/VP]/S

จากตัวอย่าง คำกริยา รักษา หากมองในแง่วากยสัมพันธ์และไวยากรณ์ การก กรอบความสัมพันธ์จะแตกต่างกันเพียง 2 กรอบเท่านั้น คือ [รักษา + OBJ] และ [รักษา + OBJ + INS] แต่หากมองในแง่ความสัมพันธ์ทางความหมายด้วย จะพบว่า คำว่า รักษา นี้มีความหมายถึง 3 ความหมาย และมีกรอบโมโนทัศน์หรือกรอบความหมายต่างกันถึง 3 กรอบ คือ [รักษา + OBJ]/Disease] มีความหมายว่า ป้องกัน ซึ่งต่างกับกรอบ [รักษา + OBJ]/Moisture] ที่มีความหมายว่า สงวนไว้ และ [รักษา + OBJ]/fish + INS]/antibiotic] มีความหมายว่า เยียวยาด้วยยารักษาโรค นั้นเอง ซึ่งความหมายตรงกับพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2542³ ดังนี้

รักษ, รักษา

- (๑) ก. ระวัง เช่น รักษาสุขภาพ, ดูแล เช่น รักษาทรัพย์สินสมบัติ, ป้องกัน เช่น รักษาบ้านเมือง, สงวนไว้ เช่น รักษาความสะอาด รักษาไมตรี (ส.; ป. รกข).
- (๒) ก. เยียวยา เช่น รักษาคนไข้. (ส.; ป. รกข).

อย่างไรก็ตาม เนื่องจากภาษามีความซับซ้อน โดยเฉพาะประโยคซับซ้อน หรือประโยคความซ้อน รวมถึงการใช้คำกริยาเรียง (serial verbs) ในประโยคภาษาไทย ซึ่งจะปรากฏคำกริยามากกว่า 1 คำในประโยค ก่อให้เกิดปัญหาการสกัดองค์ประกอบของกรอบ ดังตัวอย่างต่อไปนี้

[[กรม/ncn ชลประทาน/npn]/NP]/AGT]/Organization [จัดสรร/vt [[น้ำ/ncn]/NP]/OBJ]/Water]/VP [ทำ/vt [[นาปรัง/ncn]/NP]/OBJ]/Crop [[ใน/prep [ช่วง/ncn ฤดู/ncn แล้ง/adj [ปี/cl นี้/det]/NP]/NP]/PP]/TIME]/Dry_season [เติม/prep [[พื้นที่/ncn]/NP]/LOC]/Area]/PP]/VP

จากตัวอย่างประโยคข้างต้น มีคำกริยา 2 คำ คือ จัดสรร และ ทำ และมีกรอบโมโนทัศน์ 2 กรอบ ดังนี้

³ <http://www.royin.go.th/dictionary/รักษา>

กรอบที่ 1 จัดสรร

Verb : จัดสรร

Role :

AGT {Organization}

OBJ {Water}

LOC {Area}

กรอบที่ 2 ทำ

Verb : ทำ

Role :

OBJ {Crop}

ในการแก้ปัญหา ผู้วิจัยจึงสร้างข้อกำหนดการเกิดร่วมกันของคำตามโครงสร้างไวยากรณ์และความหมาย (Syntactic-Semantic constraint) โดยการดึงคำกริยาพร้อมความสัมพันธ์ทางความหมายและมโนทัศน์ของคำทั้งประโยค จากนั้นทำการแยกและคัดเลือกกรอบมโนทัศน์ของคำกริยาที่เหมาะสมที่สุดอีกครั้งหนึ่ง โดยรายละเอียดจะกล่าวในหัวข้อเกี่ยวกับข้อกำหนดการเกิดร่วมกันของคำตามโครงสร้างไวยากรณ์และความหมาย

3.2.2 การวิเคราะห์อรรถลักษณะของคำกริยา

ส่วนประกอบของเฟรม (frame elements) ซึ่งเป็นส่วนที่แสดงถึงบทบาทของมโนทัศน์ของคำที่ปรากฏร่วมกับคำกริยาในเฟรมนั้นแบ่งเป็น 2 ส่วน คือส่วนที่เป็นแกน (core element) กับส่วนที่เป็นชายขอบ (peripheral) หรือส่วนที่ไม่ได้เป็นแกน (non-core element) ซึ่งสองส่วนนี้ขึ้นกับกฎในทางวากยสัมพันธ์ว่าเป็นส่วนสำคัญที่ต้องปรากฏร่วมกับคำกริยา หรือคำกริยาต้องการประธานและกรรมหรือไม่ แต่สำหรับโครงการวิจัยนี้ไม่ได้ให้ความสำคัญกับส่วนประกอบที่เป็นแกนหรือส่วนที่ไม่ได้เป็นแกนของประโยค เนื่องจากภาษาไทยเป็นภาษาที่สามารถละคำได้ แม้จะเป็นส่วนประกอบที่เป็นแกนคือส่วนประธานหรือกรรมของกริยาก็สามารถละได้ หากไม่ได้เป็นประโยคเริ่ม หรือแม้กระทั่งคำนามที่เป็นประธานหรือกรรมนั้นได้เคย

กล่าวถึงมาแล้วในส่วนก่อนหน้า ซึ่งอาจจะเป็นส่วนหัวเรื่องหรือส่วนสำคัญในย่อหน้า หากกล่าวถึงแล้ว ในข้อความมักจะละคำนามสำคัญนั้นและไม่กล่าวถึงอีก ซึ่งการกำหนดให้เป็นส่วนประกอบหลัก (core frame elements) หรือส่วนที่ไม่ใช่ส่วนประกอบหลัก (non-core frame elements) จึงเป็นเรื่องที่ตัดสินยากสำหรับภาษาไทยที่ใช้จริงในงานเอกสาร สิ่งสำคัญของส่วนประกอบหลักของกรอบมโนทัศน์จึงอยู่ที่การกำหนดความหมายของคำ ซึ่งจะเป็นการที่ปรากฏกับคำกริยาและมีจำนวนมโนทัศน์ที่ปรากฏพร้อมกับคำกริยานั้นมากที่สุดในประโยคนั้นเอง

การกำหนดหรือการจัดกลุ่มความหมายของคำและมโนทัศน์ของคำจะใช้ทฤษฎี Componential Analysis ในการวิเคราะห์ความหมายของคำ ดังตัวอย่างข้อมูลต่อไปนี้ซึ่งได้จากการค้นหาคำใน CU-concordance tool⁴ ของภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ซึ่งเป็นตัวอย่างที่ทำให้เห็นการแยกความหมายของคำได้ชัดเจนยิ่งขึ้น เนื่องจากคลังข้อมูลด้านการเกษตรที่นำมาใช้ในงานวิจัยนี้มีความแตกต่างทางความหมายของคำไม่มากนัก เพราะเป็นคำในแวดวงเดียวกัน

- เกษตรกรที่ปลูกกาแฟในควมามีรายได้ 3 เท่า
- เด็กชนชั้นกลางในเมืองไปปลูกข้าว ไปปลูกผัก ในแปลงทดลองของตัวเอง
- มีการไปเรียนเรื่องการผสมพันธุ์ข้าวบ้าง ปลูกปอ ปลูกมันบ้าง
- เพื่อนผมไปเรียน 7 ปี กลับมาปลูกบ้านได้หลังใหญ่อย่างกับวัง
- เขาพูดถึงแง่ดีในวิถีชีวิตแบบใหม่ มีรายได้มาก สามารถปลูกบ้านอยู่ในเมือง
- ต่อมามีการถกเถียงกันหมด ทำถนน แล้วปลูกบ้านอยู่เป็นบ้านเดี่ยวๆ อย่างฝรั่ง อยู่กับพื้น
- IOTO และ PHONO, ขจัดเซลล์ลูไลท์, การผลิตผิว, ปลูกผม, รักษาเส้นเลือดขาด, ลดน้ำหนัก
- การกำหนดนโยบายต่างๆ ของภาครัฐ เพื่อการรณรงค์ปลูกจิตสำนึกในการรักษาสิ่งแวดล้อม

⁴ <http://www.arts.chula.ac.th/~ling/ThaiConc/>

จากข้อมูลตัวอย่างข้างต้น จะเห็นว่าหากส่วนประกอบของเฟรมหรือกรอบที่สำคัญสำหรับภาษาไทยคือ ส่วนที่เป็นกรรม (object) ของคำกริยา ซึ่งเป็นข้อมูลสำคัญในการแยกความหมายของคำว่า “ปลุก” ตามมโนทัศน์หรือ concept ที่ปรากฏร่วมในขณะที่มีมโนทัศน์ของคำที่ปรากฏในตำแหน่งประธานจะเป็นมโนทัศน์ที่เป็น [+ คน] ซึ่งสามารถปรากฏกับกริยา “ปลุก” ได้ในทุกบริบท ดังนั้นจึงไม่สามารถนามโนทัศน์ในส่วนของประธานมาเป็นเครื่องตัดสินความหมายของคำว่า “ปลุก” ได้ สำหรับมโนทัศน์ของคำหรือวลีที่เป็นกรรมนั้น สามารถแบ่งเป็น 4 มโนทัศน์ใหญ่ ๆ คือ

- 1) ส่วนที่ประกอบด้วยคุณสมบัติทางความหมาย (อรรถลักษณะ) [+สิ่งมีชีวิต, +พืช] ซึ่งเป็นคุณสมบัติของ กาแฟ ข้าว ข้าวโพด มันสำปะหลัง และ ปอ
- 2) ส่วนที่เป็น [-สิ่งมีชีวิต, +รูปธรรม, +สิ่งก่อสร้าง] ซึ่งเป็นคุณสมบัติของ บ้าน
- 3) ส่วนที่เป็น [-สิ่งมีชีวิต, +รูปธรรม, +ส่วนประกอบของร่างกายมนุษย์] ซึ่งเป็นคุณสมบัติของ ผม
- 4) ส่วนที่เป็น [-สิ่งมีชีวิต, -รูปธรรม, +ความรู้สึกของมนุษย์] ซึ่งเป็นคุณสมบัติของ จิตสำนึก

คุณสมบัติทางความหมายที่ทำให้เกิดความแตกต่างกันที่จะนำไปสู่การแยกความหมายของคำ คือ คุณสมบัติที่เป็น [+สิ่งมีชีวิต] และ [-สิ่งมีชีวิต] ส่วนที่เป็น [+รูปธรรม] กับ [-รูปธรรม] ส่วนคุณสมบัติทางความหมายของคำที่เป็น [+พืช] [+สิ่งก่อสร้าง] [+ส่วนประกอบของร่างกายมนุษย์] และส่วนที่เป็น [+ความรู้สึกของมนุษย์] นั้นจะเป็นข้อจำกัดในการเกิดร่วมกันของคำนามตามคุณสมบัติทางความหมายที่กำหนดไว้ที่มีความสัมพันธ์กับคำกริยานั้นเอง คุณสมบัติทางความหมายที่แตกต่างกันนี้เองที่ทำให้สามารถแยกความหมายของคำตามมโนทัศน์ของคำว่า “ปลุก” เป็น 4 ความหมาย ความหมายแรกหมายถึง การทำให้สิ่งมีชีวิตที่เป็นพืชเจริญเติบโต ความหมายที่สองคือ การทำให้สิ่งไม่มีชีวิตที่เป็นสิ่งปลูกสร้างให้มีขึ้นหรือตงามขึ้น ความหมายที่สามคือ การทำให้สิ่งไม่มีชีวิตที่เป็นส่วนประกอบของร่างกายมนุษย์เจริญขึ้นหรือตงามขึ้น และความหมายที่สี่หมายถึง การทำให้สิ่งไม่มีชีวิตที่เป็นความรู้สึกของมนุษย์เจริญขึ้นหรือตงามขึ้น เมื่อเทียบกับความหมายตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2542 การจำแนก

ความหมายจากคลังข้อมูลจะปรากฏความหมายที่มากกว่าพจนานุกรมฉบับราชบัณฑิตยสถาน ซึ่งได้นิยามความหมายของคำว่า “ปลูก”⁵ ดังนี้

ปลูก [ปลูก] ก. เอาต้นไม้หรือเมล็ด หน่อ หัว เป็นต้น ใส่ลงในดินหรือสิ่งอื่นเพื่อให้งอกหรือให้เจริญเติบโต, ทำให้เจริญเติบโต, ทำให้งอกงาม เช่น ปลูกไมตรี; เอาสิ่งต่างๆ มาปลูกกันเข้าเพื่อทำเป็นที่อยู่หรือที่พักอาศัย โดยวิธีฝังเสาลงในดิน เช่น ปลูกบ้าน ปลูกเรือน, ปลูกสร้าง ก็ว่า, โดยปริยายหมายถึงการกระทำที่คล้ายคลึงเช่นนั้น เช่น ปลูกพลับพลายก.

สำหรับคำกริยาที่ใช้ทดลองในงานวิจัยนี้จำนวน 106 คำนั้น ผลการวิเคราะห์พบว่าคำที่มีความถี่ในการปรากฏสูง มีจำนวนมโนทัศน์ปรากฏรวมจำนวนมาก คำอาจจะไม่สามารถแยกความหมายได้มากนัก เช่น คำว่า “เป็น” ในกรอบตัวอย่างดังต่อไปนี้

Verb : เป็น

Role :

AGT {Aphis, Hurb, Part_tree, Thai_Food, Wife, Person}

ACC {Rice}

LOC {Box, Garden, Water, Globe, Region, Meeting, upland, plateau, Area, District, Hometown}

OBJ {Disease, Soil, Organic, Tunnel, Blood, Water, Lemongrass, Food, Fungus, Chairman, Committee, Vote, Carrier, Sample, Vine, Bagasse, Cause, Seasoning, Enteron, Pig, Building, Pillar, Bamboo, Animal, Danger, Aflatoxin, Horseshoe_crab, Pest, Prohibition, Case, Shell, Center, Area, Shape, Aphis, Plant, Ingredient, Seafood, Meat, Owner, Identity, Root, Problem, Species, Consumer, Concrete, Farmer, Benefit, Seaweed, Product, Season, Income, Family, Rice, Sticky, Body}

MEA {Quantity}

INS {Match, Trap}

SOU {Soil, Consumer, Province}

EXP {Mushroom, Hole, Leaf, Decease, Teenager, OccuOBJion, Thing, Animal, Boar, Main, Way, Annual_crops, Shape, Color, Food_Type, Components, People, Lemongrass, Eradication, Problem, Variety,

⁵ <http://www.royin.go.th/dictionary/ปลูก>

Danger, Favor, Concrete, Benefit, Stability, Pepper, Earthworm,
Disease, Food, Area, Cost, Seed, Merchant, Place, Plant, Vegetable,
Bran, Tray, frog, Project, Acacia, Species, Fertilizer}
TIME {Moment, Period, Morning, period}
BEN {Consumer, Farmer}
MAN {Formula, Quantity, Characteristic, Area, Easily}

จากกลุ่มโนทัศน์ เมื่อจัดกลุ่มความหมายจริงๆ แล้ว พบว่าไม่สามารถแยกความหมายได้อย่างชัดเจน เพราะมีโนทัศน์ที่ปรากฏร่วมมีความหลากหลายมากเกินไป และเมื่อเทียบกับพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 พบว่า เป็นที่เป็นการริยา มีความหมาย 2 ความหมายเท่านั้น คือ

เป็น ๑

ก. คำกริยาสำหรับแสดงความสัมพันธ์ระหว่างคำกับคำเพื่อให้เห็นว่าคำหน้าและคำหลังมีภาวะ คือ ความมี ความเป็น เกี่ยวข้องกันอย่างไร เช่น ท่านเป็นเจ้า เขาเป็นนาย.

เป็น ๒

(๑) ก. สามารถ, ได้, เช่น เต็นเป็น ราเป็น.

สำหรับคำที่อยู่ในวงศ์ศัพท์ทางการเกษตร เช่น ปลูก เลี้ยง หมัก ซึ่งปรากฏในเอกสารเป็นจำนวนมาก จะปรากฏมีโนทัศน์ในหมวดหรือวงศ์คำศัพท์เดียวกัน ทำให้จัดกลุ่มความหมายได้ง่ายและมีความหมายเจาะจง ดังตัวอย่างคำว่า “ปลูก” ในรอบต่อไป นี้ ซึ่งมีความหมายเดียว หมายถึงการทำให้สิ่งมีชีวิตที่เป็นพืชเจริญเติบโต

Verb : ปลูก

Role :

ACC {Grass}

LOC {Bed, Hole, House, Area}

OBJ {Rice, Chilli, Lemongrass, Tree, Papaya, Cassava, Plant, Rattan,
Branch, Bamboo, Longan, Rambutan, Vegetable, Fruits, Fruit_tree, Lead_tree}

SOU {Stem}

EXP {Grass, organic_vegetable}

TIME {Period, Now, Present, Season, Rainy_season, Time}

MEA {Area, Cost, Tree}

AGT {Group, Agriculturist}

MAN {Easy, Principle, Organization, Long}

3.2.3 โครงสร้างกรอบมโนทัศน์หรือโครงสร้างกรอบภาษาไทย

ข้อแตกต่างอีกประการหนึ่งระหว่างโครงการเฟรมเน็ตของมหาวิทยาลัยแคลิฟอร์เนีย เบิร์กลีย์ กับโครงการวิจัยนี้ คือส่วนที่เป็นโครงสร้างกรอบและการเชื่อมโยงกรอบมโนทัศน์ให้เป็นโครงข่ายเชิงความหมาย (Semantic network) สำหรับโครงสร้างกรอบมโนทัศน์ภาษาไทยในงานวิจัยนี้มีลักษณะที่แตกต่างจากโครงสร้างในเฟรมเน็ตของมหาวิทยาลัยแคลิฟอร์เนีย เบิร์กลีย์ คือส่วนที่เป็นความสัมพันธ์ของมโนทัศน์ ซึ่งความสัมพันธ์ของมโนทัศน์นั้นเฟรมเน็ตของมหาวิทยาลัยแคลิฟอร์เนีย เบิร์กลีย์ เชื่อมโยงระหว่างมโนทัศน์กับมโนทัศน์ (Frame-frame Relations) แต่โครงการวิจัยนี้เป็นการเชื่อมโยงระหว่างคำกริยากับมโนทัศน์แวดล้อม (Syntactic-Semantic Relation) ที่ปรากฏรวมกับการก ซึ่งได้แก่ ส่วนที่เป็นนามวลีและส่วนขยาย ส่วนประกอบต่างๆ ของกรอบมโนทัศน์ภาษาไทยในงานวิจัยนี้ประกอบด้วยส่วนหลัก 6 ส่วน ได้แก่

Frame Name: คือชื่อของเฟรม ซึ่งจะมีลักษณะเป็นมโนทัศน์ (concept)

Lexical Units: คือรายการคำที่อยู่ใน Frame Name เดียวกัน

Frame elements (FEs): เป็นบทบาททางความหมายของคำ (thematic roles)

Syntactic-Semantic Relation (SynSem): เป็นความสัมพันธ์ของมโนทัศน์ที่สามารถปรากฏในตำแหน่งของ Frame elements สำหรับมโนทัศน์ตาม Frame Name ซึ่งมโนทัศน์ทางความหมายของคำ (semantic concept) จะนำมาจาก WordNet⁶

Example Sentence: เป็นประโยคตัวอย่างที่มีการกำกับข้อสนเทศของคำ บทบาททางความหมาย (FEs) และชนิดของวลี (phrase type) แล้ว

ข้อกำหนดการเกิดร่วมกันของคำ (Syntactic-Semantic constraint): เป็นข้อกำหนดการเกิดร่วมกันของคำตามโครงสร้างไวยากรณ์และความหมาย

ส่วนประกอบของเฟรม (Frame elements (FEs)) จะยึดหลักบทบาททางความหมาย (thematic roles) หรือการกเป็นเกณฑ์ในการกำหนดส่วนประกอบต่างๆ

⁶ <https://wordnet.princeton.edu/>

ส่วนเรื่องของการกนั้น จำนวนการที่ใช้ในงานวิจัยนี้จะต่างจากการที่ใช้ในโครงการเฟรมเน็ต เนื่องจากงานวิจัยนี้เน้นความสัมพันธ์ของคำในประโยคระดับโครงสร้างผิวที่อิงบทบาททางความหมาย (thematic roles) โดยการกที่ใช้มีจำนวนไม่มาก เป็นการกดั้งเดิมของฟิลล์มอร์ ที่สร้างขึ้นในปี 1971 (Fillmore, 1971) และการกของลาร์สัน (Larson, 1984) บางส่วน พร้อมทั้งเชื่อมโยงมโนทัศน์ (semantic concept) จากพจนานุกรมที่มีความสัมพันธ์ทางความหมายของคำแบบลดหลั่นให้กับบทบาทการกอีกชั้นหนึ่ง การกดั้งเดิมของ Fillmore มีจำนวน 9 การก ได้แก่

- 1) Agent เป็นบทบาทของผู้ที่ทำให้เกิดการกระทำที่ระบุในกริยาขึ้น โดยทั่วไปมักเป็นสิ่งมีชีวิต แต่อาจใช้กับสิ่งไม่มีชีวิตเช่นหุ่นยนต์ หรือสถาบันของมนุษย์ เช่นชาติได้ด้วย
- 2) Instrument เป็นบทบาทของวัตถุหรือบางสิ่งที่เกี่ยวข้องกับการกระทำหรือสถานะที่ระบุในกริยา
- 3) Experiencer เป็นบทบาทของสิ่งที่รับรู้การกระทำหรือสถานะที่ระบุในกริยาโดยที่ไม่ได้เป็นผู้ทำให้เกิดการกระทำหรือสถานะดังกล่าวขึ้น มักเป็นสิ่งมีชีวิต
- 4) Object เป็นบทบาทสำหรับนามใดๆ ซึ่งบทบาทในการกระทำหรือในสถานะที่ระบุในกริยาขึ้นกับการตีความคำกริยา ใช้กับสิ่งที่ได้รับผลจากการกระทำหรือสถานะที่ระบุในกริยา
- 5) Location เป็นบทบาทซึ่งระบุสถานที่หรือตำแหน่งของสถานะหรือการกระทำที่ระบุในกริยา
- 6) Source เป็นบทบาทของสิ่งที่เป็นจุดเริ่มต้นการกระทำที่ระบุในกริยา
- 7) Goal เป็นบทบาทของสิ่งที่เป็นเป้าหมายของการกระทำที่ระบุในกริยา
- 8) Time เป็นบทบาทของสิ่งแสดงเวลาที่เกิดการกระทำหรือสถานะที่ระบุในกริยา
- 9) Benefactive เป็นบทบาทของสิ่งที่รับผลจากการกระทำ โดยที่ไม่ได้เป็นเป้าหมายของการกระทำที่ระบุในกริยาโดยตรง

นอกจากนี้ได้นำการกของลาร์สัน (Larson, 1984) มาใช้ 3 การก คือ

- 1) Accompaniment เป็นบทบาทของผู้ที่เกี่ยวข้องกับ Agent, Causer หรือ Affected ในเหตุการณ์อย่างใกล้ชิด ถือว่าเป็น Agent, Causer หรือ Affected ในระดับรองลงมาได้ เช่น เป็นผู้ทำเหตุการณ์ร่วมกับ Agent เป็นต้น

- 2) Manner เป็นบทบาททางคุณภาพของเหตุการณ์ แสดงให้เห็นถึงแนวทางการกระทำในกิริยาถูกกระทำว่าทำอย่างไร มีลักษณะอย่างไร
- 3) Measure เป็นบทบาททางปริมาณของเหตุการณ์ แสดงให้เห็นถึงสิ่งที่เกี่ยวข้องกับเหตุการณ์ที่สามารถวัดได้ เช่น ขนาด ปริมาณ ความถี่ เป็นต้น

ส่วนความสัมพันธ์ระหว่างความหมายกับวากยสัมพันธ์ (Syntactic-Semantic Relation (SynSem)) นั้น โครงการวิจัยนี้ใช้ความสัมพันธ์ระหว่างการกับมโนทัศน์ (concept) เป็นส่วนแสดงความสัมพันธ์ โดยมโนทัศน์ที่ใช้นำมาจากพจนานุกรมที่มีความสัมพันธ์ทางความหมายของคำแบบลดหลั่น (is-a relation) ซึ่งเป็นความหมายในลักษณะที่เป็นความคิดรวบยอดซึ่งเรียกว่า มโนทัศน์ (concept) (ราชบัณฑิตยสถาน, 2553) ชื่อพจนานุกรม WordNet ของมหาวิทยาลัยพรินซ์ตัน (Princeton University) ประเทศสหรัฐอเมริกา พัฒนาโดยศาสตราจารย์จอร์จ เอ มิลเลอร์ เป็นฐานความรู้มโนทัศน์กำกับให้กับบทบาทการกรก ดังตัวอย่างกรอบ “เพาะ” ต่อไปนี้

เพาะ

Verb : เพาะ

Role :

AGT {Person}

BEN {Chicken}

INS {Excrement, Plastic}

LOC {Food}

MEA {Method, Weight, Day, Quantity}

OBJ {Germ, Seed, Tree, Benefit, Protein, Worm, Mushroom, Tadpole, Strain, Sprout}

TIME {Season, Period, Summer, Winter}

Example sentences:

Sentence ID: 346 => เพาะ (A, B) : A - Object {Germ}; B - Location {Food};

[[เพาะ/vt [[[เชื้อ/ncn]/NP]/OBJ]/Germ [[[ใน/prep อาหาร/ncn]/PP]/LOC]/Food]/NP]/VP]/S

Sentence ID: 353 => เพาะ (A, B) : A - Object {Worm}; B - Time {Period};

[[สามารถ/prev เพาะ/vt [[[หนอน/ncn]/NP]/OBJ]/Worm ได้/vpost [[[ตลอด/prep ปี/ncn]/PP]/TIME]/Period]/VP]/S

จากตัวอย่างกรอบหรือเฟรมของ “เพาะ” ประกอบด้วยส่วนต่างๆ ดังนี้

- **Frame Name** คือ เพาะ ซึ่งจะปรากฏในบรรทัดแรกของโครงสร้างกรอบ
- **Lexical Units** คือ คำว่า เพาะ มีชนิดของคำเป็นคำกริยา
- **Frame elements (FEs)** กรอบมโนทัศน์ “เพาะ” มีส่วนประกอบที่เป็นบทบาททางความหมาย (thematic roles) ได้ 7 การก ได้แก่ Agent (AGT), Object (OBJ), Benefactor (BEN), Instrument (INS), Location (LOC), Measure (MEA) และ Time (TIME)
- **Syntactic-Semantic Relation (SynSem)** มโนทัศน์ที่สามารถปรากฏในตำแหน่งของ Frame elements สามารถอธิบายได้ว่า ในตำแหน่ง Agent (AGT) บทบาทของผู้ที่ทำให้เกิดการกระทำต้องเป็น คน {Person} ส่วนกรรมหรือ object (OBJ) กลุ่มคำที่สามารถเป็นได้ คือ เชื้อเห็ด {Germ}, เมล็ด {Seed}, ต้นไม้ {Tree}, โปรตีน {Protein}, หนอน {Worm}, เห็ด {Mushroom}, ลูกอ๊อด {Tadpole}, สายพันธุ์ {Strain}, หน่อ {Sprout} ส่วน Benefactor (BEN) หรือสิ่งที่รับผลจากการกระทำก็คือ สัตว์ประเภทไก่ {Chicken} มีเครื่องมือหรือ Instrument (INS) เป็นปุ๋ยคอก {Excrement} หรือสามารถใช้อุปกรณ์ประเภทพลาสติก {Plastic} เป็นเครื่องมือได้ นอกจากนี้สถานที่หรือตำแหน่ง (Location) ที่เพาะ คือ ในอาหาร ปริมาณ (Measure) ที่สามารถวัดได้ เช่น วิธีการ {Method}, น้ำหนัก {Weight}, จำนวนวัน {Day} และปริมาณ {Quantity} และสุดท้ายกรอบมโนทัศน์ของ “เพาะ” สามารถปรากฏร่วมกับระยะเวลา (Time) ที่เกิดการกระทำนั้น
- **Example sentences** เป็นประโยคตัวอย่างที่มีการกำกับข้อสนเทศของคำในทุกระดับแล้ว พร้อมทั้งแสดงข้อกำหนดการเกิดร่วมกันของคำตาม เพื่อให้นักวิจัยทางด้านการประมวลผลภาษาธรรมชาติสามารถนำข้อมูลไปใช้ได้อย่างสะดวกและง่าย ข้อกำหนดการเกิดร่วมกันของคำตามโครงสร้างไวยากรณ์และความหมายของ “เพาะ” เรียงตามลำดับดังนี้

Sentence ID: 346 => เพาะ (A, B) : A - Object {Germ}; B - Location {Food};

Sentence ID: 353 => เพาะ (A, B) : A - Object {Worm}; B - Time {Period};

3.2.4 ข้อกำหนดการเกิดร่วมกันของคำตามโครงสร้างไวยากรณ์

และความหมาย (Syntactic-Semantic constraint)

ในการนำความสัมพันธ์ของคำไปใช้กับการประมวลผลภาษาธรรมชาติ เช่น ระบบสกัดข้อมูลอัตโนมัติ ระบบถามตอบอัตโนมัติ ระบบแปลภาษาอัตโนมัติ ฯลฯ จำเป็นต้องกำกับความหมายให้กับคำที่สัมพันธ์กับคำกริยาทุกคำ ความหมายของคำเหล่านั้นต้องมีการจัดหมวดหมู่ตามคุณสมบัติทางความหมายแล้ว ซึ่งเป็นเรื่องที่ต้องใช้เวลาในการสร้างหมวดหมู่คำโดยแยกส่วนประกอบทางความหมายของคำให้ทุกคำ ดังนั้นเพื่อความสะดวกจึงจำเป็นต้องอาศัยความหมายของคำหรือโน้ตส์ที่มีการจัดกลุ่มไว้แล้วใน WordNet ของ Princeton University⁷ มาใช้สร้างเป็นข้อกำหนดการเกิดของคำตามโครงสร้างไวยากรณ์และความหมาย (Syntactic-Semantic constraint) ซึ่งจะมีลักษณะที่เป็นรูปธรรมและสามารถนำไปใช้ในระบบประมวลผลภาษาธรรมชาติได้ โดยลักษณะของข้อกำหนดการเกิดของคำตามโครงสร้างไวยากรณ์และความหมาย มีลักษณะดังนี้

ประโยคตัวอย่าง 1: เด็กชนชั้นกลางในเมืองไปปลูกข้าว ไปปลูกผัก ในแปลงทดลองของตัวเอง

[[เด็กชนชั้นกลางในเมือง]/AGT]/Human ไป ปลูก[[ข้าว]/OBJ]/Plant ไป ปลูก[[ผัก]/OBJ]/plant ใน [[แปลงทดลองของตัวเอง]/LOC]/Field

ข้อกำหนด (Constraint) ของ “ปลูก” สามารถเขียนได้ดังนี้

```
ปลูก (X, Y, Z)
ปลูก vt.
X__AGENT; Y__OBJ ; Z__ใน LOC
AGT → Human
OBJ → Plant
LOC → Field
```

⁷ <https://wordnet.princeton.edu/>

ประโยคตัวอย่าง 2: เพื่อนผมไปเรียน 7 ปี กลับมาปลูกบ้านได้หลังใหญ่อย่างกับวัง

[[เพื่อนผม]/AGT]/Human ไปเรียน 7 ปี กลับมา ปลูก [[บ้าน]/OBJ]/
Building [[ได้หลังใหญ่อย่างกับวัง]/MAN]/Size

ข้อกำหนด (Constraint) ของ “ปลูก” สามารถเขียนได้ดังนี้

ปลูก (X, Y, Z)
ปลูก vt.
X__AGENT; Y__OBJ ; Z__MAN
AGT → Human
OBJ → Building
MAN → Size

ประโยคตัวอย่าง 3: IOTO และ PHONO, ขจัดเซลล์ลูไลท์, การผลิตผิว, ปลูกผม, รักษา
เส้นเลือดอุดตัน, ลดน้ำหนัก,

ปลูก [[ผม]/OBJ]/Part of Human

ข้อกำหนด (Constraint) ของ “ปลูก” สามารถเขียนได้ดังนี้

ปลูก (Y)
ปลูก vt.
Y__OBJ
OBJ → Part of Human

ประโยคตัวอย่าง 4: การกำหนดนโยบายต่างๆ ของภาครัฐ เพื่อการอนุรักษ์ปลูกจิตสำนึก
ในการรักษาสิ่งแวดล้อม

ปลูก [[จิตสำนึก]/OBJ]/Conscious [[ใน การรักษาสิ่งแวดล้อม]/GOA
]/Conservation

ข้อกำหนด (Constraint) ของ “ปลูก” สามารถเขียนได้ดังนี้

ปลูก (Y, Z)
ปลูก vt.
Y__OBJ ; Z__ใน GOA
OBJ → Conscious

จากผลการวิจัยที่กล่าวมาข้างต้น ลักษณะของกรอบมโนทัศน์ในระดับ concept นี้เป็นกรอบที่รวมการเกิดรวมของคำที่อยู่ใน concept เดียวกันไว้ เพื่อแสดงความเป็นไปได้ในบริบทต่างๆ ของกรอบมโนทัศน์ในระดับข้อความ หากต้องการดูรายละเอียดปลีกย่อยของกรอบคำในแต่ละ Lexical Unit เพื่อดูความสัมพันธ์ของคำ และกรอบมโนทัศน์ในระดับประโยคและระดับข้อความ สามารถดูได้จากเว็บไซต์ <http://tcf.human.ku.ac.th/>

4. สรุป

งานวิจัยเรื่องการสร้างกรอบมโนทัศน์ภาษาไทยจากคลังข้อมูลด้านการเกษตร เพื่อการประมวลผลภาษาด้วยคอมพิวเตอร์เป็นการสร้างกรอบมโนทัศน์ในระดับประโยคโดยใช้คำกริยาเป็นแกน รวบรวมความสัมพันธ์ของคำต่าง ๆ ที่อยู่รอบๆ คำกริยานั้น แล้วจึงสรุปรวมเป็นกรอบมโนทัศน์ของคำ ทำให้ได้ความสัมพันธ์ของคำในทุกบริบทที่เป็นไปได้ ซึ่งเป็นประโยชน์สำหรับโปรแกรมประยุกต์ต่างๆ โดยเฉพาะระบบถาม-ตอบ ระบบแปลภาษา ระบบสกัดข้อความแบบอัตโนมัติ การแปลภาษาโดยอิงบริบท เป็นต้น

รายการอ้างอิง

- พุทธชาติ โปธิบาล และคณะ. (2557). *โครงการวิจัยเรื่องการสร้างกรอบมโนทัศน์ภาษาไทยจากคลังข้อมูลด้านการเกษตรเพื่อเป็นฐานความรู้ในการสกัดข้อความอัตโนมัติสำหรับระบบถาม-ตอบภาษาไทย*. กรุงเทพฯ: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ.
- ราชบัณฑิตยสถาน. (2553). *พจนานุกรมศัพท์ภาษาศาสตร์ (ภาษาศาสตร์ประยุกต์)*. กรุงเทพฯ: รุ่งศิลป์การพิมพ์ (1977).

- วจินตน์ ภาณุพงศ์. (2532). *โครงสร้างของภาษาไทย: ระบบไวยากรณ์* (พิมพ์ครั้งที่ 10). กรุงเทพฯ: มหาวิทยาลัยรามคำแหง.
- Burchardt, A. et al. (2009). FrameNet for the semantic analysis of German: Annotation, representation and automation (preprint). In Hans C. Boas (Ed.), *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Guyter.
- Chafe, W. (1970). *Meaning and the Structure of Language*. Chicago: The University of Chicago Press.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Mouton.
- Cook, Walter A., S.J. (1989). *Case Grammar Theory*. Washington, DC: Georgetown University Press.
- Eva Lavric, et al. (2008). The Kicktionary: Combining corpus linguistics and lexical semantics for a multilingual football dictionary. In *The Linguistics of Football* (Language in Performance 38) (pp. 11-23). Tübingen: Gunter Narr.
- Fellbaum C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.
- Fillmore, C. (1968). The Case for Case. In Emmon Bach and R.T. Harms (Eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston.
- Fillmore, C. (1971). Some Problems for Case Grammar. In R.J. (ed.), *O'Brien* (pp. 35-56).
- Fillmore C. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*. Conference on the Origin and Development of Language and Speech, 280, 20-32.
- Fillmore C. (1977). The need for a frame semantics in linguistics. In H. Karlgren Scriptor (Ed.), *Statistical Methods in Linguistics*.
- Fillmore C. (1982). Frame semantics. In *Linguistics in the Morning Calm* (pp. 111-137). Seoul, South Korea: Hanshin Publishing.
- Fillmore C. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6, 222-254.
- Fillmore, C. and C. Baker. (2001). Frame Semantics for Text Understanding. *Proceedings of WordNet and Other Lexical Resources Workshop*, Pittsburgh. NAACL.

- Fillmore, C. and C. Baker. (2010). A frames approach to semantic analysis. In Heine, B. and Narrog, H., (eds.), *Oxford Handbook of Linguistic Analysis*: (pp. 13-341).
- Fillmore CJ, Petruck MRL, Ruppenhofer J, Wright A. *Framenet in action: The case of attaching*. *International Journal of Lexicography*, 16, 297-332.
- Gruber, J. (1965). *Studies in lexical relations*. Massachusetts Institute of Technology.
- H. V. D. Parunak. (1995). Case Grammar: A Linguistic Tool for Engineering Agent-Based Systems. *ITI Technical Memorandum*. ps, Industrial Technology Institute, Ann Arbor. Retrieved August 3, 2016 from <http://www.iti.org/~van/casegram>,
- Jackendoff, R. (1985). The Role of Thematic Relations in Linguistic Theory. Symposium on *Thematic Roles*. Brandeis University.
- Kawtrakul A., et al. (2004). MINECoP: An Integrated Visualization Tool for Corpus Mining, *IJCNLP'2004*, HAINAN Island, China.
- Larson, M. (1984). *Meaning-based translation: A guide to cross-language equivalence*. Lanham, MD: University Press of America.
- Leenoi, D., S. Jumpathong and T. Supnithi. (2010). Building Thai FrameNet through a Combination Approach. *IALP 2010*, 277-280.
- Ohara, Kyoko Hirose et al. (2003). The Japanese FrameNet Project: A Preliminary Report. In *Proceedings of Pacific Association for Computational Linguistics (PACLING'03)* (pp. 249-254). Halifax, Canada. August, 2003.
- Schmidt, T. (2009). *Kicktionary*. The multilingual electronic dictionary of football league. Retrieved from <http://www.kicktionary.de/index.html>.
- Subirats, Carlos; Petruck, Miriam. (2003). Surprise: Spanish FrameNet. International Congress of Linguists. Workshop on Frame Semantics, Prague (Czech Republic), July 2003.
- Sudprasert S. and Kawtrakul A. (2003). "Thai Word Segmentation based on Global and Local Unsupervised Learning", NCSEC'2003, Chonburi, Thailand.
- Torrent et al. (2014). Copa 2014 FrameNet Brasil: a frame-based trilingual electronic dictionary for the Football World Cup. *COLING 2014*, (pp. 10-14). Dublin, Ireland, August 23-29.