

ทฤษฎีโครงสร้างวาทะ:
จากมุมมองเชิงทฤษฎีสู่การประมวลผลภาษาไทย

Rhetorical Structure Theory:
From Theoretical Viewpoint to Thai Natural Language Processing

ศุภวัจน์ แต่รุ่งเรือง¹

Received 7 June 2021

Revised 10 August 2021

Published 11 August 2021

บทคัดย่อ

ทฤษฎีโครงสร้างวาทะเป็นทฤษฎีที่อธิบายการจัดองค์ประกอบของข้อความภายในโครงสร้างปริบท ด้วยจุดเด่นเชิงทฤษฎีที่เอื้อต่อการนำไปประยุกต์ใช้ได้หลากหลาย ทำให้ทฤษฎีนี้ได้รับการยอมรับอย่างกว้างขวางในหลายวงการ โดยเฉพาะอย่างยิ่งในวงการประมวลผลภาษาธรรมชาติ บทความชิ้นนี้มีวัตถุประสงค์เพื่อนำเสนอแนวคิดพื้นฐานและพัฒนาการของทฤษฎีดังกล่าว รวมถึงชี้ให้เห็นจุดเด่นที่เอื้อต่อการนำไปพัฒนาต่อยอดในงานหลายแขนง ยิ่งไปกว่านั้น บทความนี้จะได้กล่าวเจาะจงถึงความเกี่ยวข้องระหว่างทฤษฎีโครงสร้างวาทะกับการวิเคราะห์และประมวลผลภาษาไทย จากนั้นจึงจะชี้ให้เห็นประเด็นปัญหาที่เกี่ยวข้องของภาษาไทยเพื่อให้เป็นแนวทางการวิจัยในอนาคต

คำสำคัญ: ทฤษฎีโครงสร้างวาทะ, โครงสร้างปริบท, การประมวลผลภาษาไทย

Abstract

Rhetorical structure theory (RST) is a theory about text organization in discourse structure. Due to its theoretical strengths that allow for a wide range of applications, this theory is widely accepted in many areas, especially in area of natural language processing. This paper aims to present the basic concepts and development of this theory, as well as to highlight its strengths that lead to further development in many fields. Moreover, the relevance between Rhetorical structure theory and the analysis and processing of Thai is addressed in this paper. Then, to guide future research, problematic issues related to Thai are discussed.

Keywords: Rhetorical structure theory, Discourse structure, Thai natural language processing

¹ อาจารย์ประจำภาควิชาภาษาไทย คณะมนุษยศาสตร์ มหาวิทยาลัยเชียงใหม่

บทนำ

นับเป็นเวลากว่า 70 ปีมาแล้วที่วงการปัญญาประดิษฐ์ วิทยาการคอมพิวเตอร์ และภาษาศาสตร์ ได้พยายามผสานองค์ความรู้เข้าด้วยกันเพื่อผลักดันให้คอมพิวเตอร์สามารถวิเคราะห์และเข้าใจภาษามนุษย์ได้ภายใต้ศาสตร์ที่เรียกว่า “การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)” (Tapsai, Unger, & Meesad, 2021, p. 1) ผลจากความพยายามดังกล่าวได้ก่อให้เกิดนวัตกรรมจำนวนมากที่ช่วยอำนวยความสะดวกในการดำเนินชีวิตของมนุษย์ในปัจจุบัน ไม่ว่าจะเป็นการแปลภาษาด้วยเครื่อง การสรุปย่อข้อความในเอกสาร หรือการค้นคืนสารสนเทศจากฐานข้อมูล

การทำความเข้าใจภาษามนุษย์ของคอมพิวเตอร์นั้นมีลักษณะสอดคล้องกับการศึกษาภาษาตามแนวภาษาศาสตร์ กล่าวคือ คอมพิวเตอร์ต้องเข้าใจความสัมพันธ์ระหว่างหน่วยทางภาษาในระดับต่างๆ ที่ประกอบขึ้นเป็นข้อความตั้งแต่ระดับคำเรื่อยไปกระทั่งถึงระดับปริจเฉท ทั้งนี้ ในการประมวลผลระดับปริจเฉทนั้น อาจกล่าวได้ว่าทฤษฎีโครงสร้างวาทะ (Rhetorical Structure Theory: RST) เป็นทฤษฎีหนึ่งที่ได้รับการยอมรับอย่างแพร่หลายทั้งในเชิงทฤษฎีและการนำไปใช้งาน

ตลอดระยะเวลา นับตั้งแต่มีการเสนอทฤษฎีโครงสร้างวาทะขึ้น ทฤษฎีดังกล่าวได้ถูกนำไปประยุกต์ใช้เป็นกรอบการวิเคราะห์ข้อมูลภาษาเรื่อยมา ไม่เฉพาะสาขาภาษาศาสตร์คอมพิวเตอร์เท่านั้น แต่ยังรวมถึงสาขาปริจเฉทวิเคราะห์ ภาษาศาสตร์เชิงทฤษฎี และภาษาศาสตร์จิตวิทยาด้วย (Das & Taboada, 2018, p. 150; Taboada & Mann, 2006) ส่วนในงานประมวลผลภาษาธรรมชาติ ทฤษฎีโครงสร้างวาทะก็ถูกใช้ในการพัฒนาเครื่องมือประมวลผลข้อความหลายแขนง โดยเฉพาะอย่างยิ่งในการพัฒนาระบบถามตอบอัตโนมัติและระบบสรุปย่อข้อความ ด้วยการยอมรับอย่างแพร่หลายนี้เอง ทฤษฎีโครงสร้างวาทะจึงได้รับการตีความและพัฒนาแนวคิดเชิงทฤษฎีเรื่อยมากระทั่งถึงปัจจุบัน

ในส่วนของภาษาไทยนั้นก็ปรากฏงานวิจัยที่ประยุกต์ใช้ทฤษฎีโครงสร้างวาทะจำนวนไม่น้อย ทั้งงานที่มุ่งตีความทฤษฎีเพื่อวิเคราะห์ภาษาไทย (Intasaw & Aroonmanakun, 2013; Saetia, Taerungruang, & Chalothorn, 2020; Sinthupoun & Sornil, 2010) หรืองานที่มุ่งประยุกต์ใช้ทฤษฎีเพื่อพัฒนาเครื่องมือประมวลผลภาษาไทย (Ketui, Theeramunkong, & Onsuwan, 2015; Sukvaree, Charoensuk, Wattanamethanont, & Kultrakul, 2004; Wattanamethanont, Suvakree, & Kawtrakul, 2005) งานวิจัยเหล่านี้ได้ช่วยสะท้อนให้เห็นพัฒนาการของทฤษฎีโครงสร้างวาทะผ่านมุมมองในการวิเคราะห์ภาษาไทย ตลอดจนการประยุกต์ใช้แนวคิดเชิงทฤษฎีสู่การพัฒนาระบบประมวลผลภาษาไทยได้เป็นอย่างดี

ด้วยเหตุที่กล่าวมาข้างต้น บทความชิ้นนี้จึงมีวัตถุประสงค์เพื่อนำเสนอโน้ตค้นพื้นฐานและพัฒนาการของทฤษฎีโครงสร้างวาทะตั้งแต่ได้รับการเสนอขึ้นจนถึงปัจจุบัน อีกทั้งยังจะได้ชี้ให้เห็นถึงจุดเด่นในเชิงทฤษฎีที่เอื้อต่อการนำไปประยุกต์ในงานประมวลผลข้อความแขนงต่างๆ โดยเฉพาะอย่างยิ่ง

ในส่วนที่เกี่ยวข้องกับการประมวลผลภาษาไทย นอกจากนี้แล้ว ผู้เขียนยังจะได้นำเสนอประเด็นปัญหาและแนวทางการศึกษาวิจัยภาษาไทยเชิงทฤษฎีโครงสร้างวาทะด้วย ผลที่ได้จากการทบทวนงานที่เกี่ยวข้องกับทฤษฎีโครงสร้างวาทะในครั้งนี้ไม่เพียงแต่เป็นการประมวลองค์ความรู้ให้เป็นระบบเท่านั้น แต่ยังเป็นการเสนอมุมมองใหม่ในการวิเคราะห์ภาษาไทย อันจะสามารถนำไปประยุกต์ใช้ต่อยอดในงานประมวลผลภาษาไทยได้ในอนาคต

ผู้เขียนได้แบ่งเนื้อหาของบทความนี้ออกเป็น 6 หัวข้อด้วยกัน นอกจากหัวข้อแรกอันได้แก่บทนำนี้แล้ว ในหัวข้อที่ 2 ผู้เขียนจะนำเสนอมนทัศน์พื้นฐานและพัฒนาการของทฤษฎีโครงสร้างวาทะตั้งแต่อดีตจนถึงปัจจุบัน จากนั้นจะได้ชี้ให้เห็นจุดเด่นของตัวทฤษฎีที่เอื้อต่อการประยุกต์ใช้งานในหลากหลายแขนง ในหัวข้อที่ 3 และในหัวข้อที่ 4 จะกล่าวเจาะจงถึงความเกี่ยวข้องระหว่างทฤษฎีโครงสร้างวาทะกับการวิเคราะห์และประมวลผลภาษาไทย ส่วนในหัวข้อที่ 5 ผู้เขียนจะได้ชี้ให้เห็นประเด็นปัญหาและแนวทางการวิจัยในอนาคต จากนั้นจึงเป็นการสรุปองค์ความรู้ที่ได้กล่าวมาทั้งหมดในหัวข้อสุดท้าย

มนทัศน์พื้นฐานและพัฒนาการของทฤษฎีโครงสร้างวาทะ

ทฤษฎีโครงสร้างวาทะเป็นทฤษฎีหนึ่งในกลุ่มทฤษฎีโครงสร้างปริจเฉท (Discourse Structure Theories: DSTs) เช่นเดียวกับกับทฤษฎีโครงสร้างข้ามเอกสาร (Cross-document Theory: CDT) (Radev, 2000) หรือทฤษฎีโครงสร้างของเพนน์ (Penn's Structure Theory: PST) (Prasad et al., 2008) สมมติฐานหลักที่ทฤษฎีในกลุ่มนี้มีร่วมกันคือ ปริจเฉทมิได้เกิดการนำถ้อยคำมารวมกันอย่างไรแบบแผน แต่ประกอบขึ้นจากหน่วยปริจเฉท (discourse unit) ที่เชื่อมโยงกันอย่างมีเหตุผลและเป็นระบบเครือข่าย ด้วยเหตุนี้ กลุ่มทฤษฎีโครงสร้างปริจเฉทจึงมุ่งหาคำตอบว่าอนุพากย์ ประโยค หรือช่วงของข้อความ (text span) สามารถเชื่อมโยงกันเป็นข้อความได้อย่างไร (Hou, Zhang, & Fei, 2020, p. 1)

ทฤษฎีโครงสร้างวาทะได้รับการเสนอเป็นครั้งแรกโดยแมนน์และทอมป์สัน (Mann & Thompson, 1987, 1988) เพื่อเป็นส่วนหนึ่งของโครงการสร้างข้อความอิงคอมพิวเตอร์ (computer-based text generation) แนวคิดหลักของทฤษฎีฉบับแรกนี้คือการพิจารณาความสัมพันธ์ระหว่างหน่วยย่อยในปริจเฉท โดยยึดเจตนาของผู้ส่งสารและสารของตัวสาร และแสดงออกมาโดยใช้แผนผังต้นไม้

ในเชิงทฤษฎีนั้น แมนน์และทอมป์สันได้เสนอว่าช่วงของข้อความ (text span)² ช่วงใดๆ ที่อยู่ข้างเคียงกันย่อมมีความสัมพันธ์ (relation) ซึ่งกันและกันตามสถานะความเป็นแกนกลาง (nuclearity) สถานะความเป็นแกนกลางนี้แบ่งออกได้เป็น 2 แบบ ได้แก่ แกนกลางเดี่ยว (mononuclear) และหลายแกนกลาง (multi-nuclear) ในกรณีที่เป็นความสัมพันธ์แบบแกนกลางเดี่ยว ช่วงของข้อความหนึ่งจะมีบทบาทเป็นแกนกลาง (nucleus) และอีกช่วงของข้อความหนึ่งจะมีบทบาทเป็นบริวาร (satellite) ช่วง

² ช่วงของข้อความในทฤษฎีฉบับดั้งเดิมมักอยู่ในรูปของอนุพากย์หรือประโยค

ของข้อความที่เป็นแกนกลางจะเป็นใจความหลัก เป็นส่วนที่จำเป็นและไม่สามารถละทิ้งได้ ในขณะที่ช่วงของข้อความที่เป็นบริวารคือข้อความส่วนที่ทำหน้าที่ขยายความให้ช่วงของข้อความที่เป็นแกนกลาง เป็นส่วนที่สามารถละทิ้งได้โดยไม่กระทบใจความหลัก ส่วนกรณีที่เป็นความสัมพันธ์แบบหลายแกนกลางนั้น ช่วงของข้อความทั้งสองช่วงจะมีความสำคัญเท่าเทียมกันและต่างทำหน้าที่เป็นแกนกลาง ไม่สามารถละทิ้งส่วนใดส่วนหนึ่งได้ ช่วงของข้อความแต่ละช่วงไม่ว่าจะมีบทบาทเป็นแกนกลางหรือบริวารเมื่อประกอบเข้าด้วยกันในโครงสร้างปริจเฉทจะแสดงความสัมพันธ์อย่างใดอย่างหนึ่งต่อกันเสมอ ทั้งนี้ แมนน์และทอมป์สันเสนอให้มีความสัมพันธ์ทั้งหมด 23 แบบ เช่น สาเหตุ ภูมิหลัง หรือเงื่อนไข เป็นต้น

ในทฤษฎีฉบับของแมนน์และทอมป์สันนี้ ความสัมพันธ์ระหว่างช่วงของข้อความจะได้รับการนิยามด้วยชุดของ**ข้อบังคับ (constraint)** ของแกนกลาง (nucleus: N) และบริวาร (satellite: S) ซึ่งเกี่ยวข้องกับเป้าหมายและความเชื่อของผู้เขียน (writer: W) และผู้อ่าน (reader: R) อันจะก่อให้เกิดผลกระทบ (effect) ต่อผู้อ่าน (reader: R) เพื่อความเข้าใจที่ชัดเจนยิ่งขึ้น ขอให้พิจารณาข้อความในตัวอย่าง (1) ที่แสดงข้อบังคับของความสัมพันธ์แบบหลักฐาน (evidence)

(1) อาจารย์จักรภาพต้องอยู่ที่นี้ เพราะรถของเขาจอดอยู่หน้าคณะ

ข้อความสัมพันธ์: หลักฐาน

ข้อบังคับของ N: R อาจจะไม่เชื่อ N ในระดับที่ W พอใจ

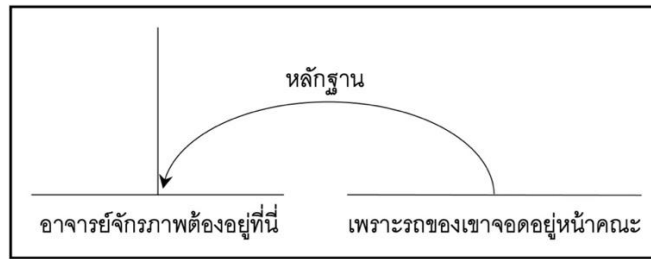
ข้อบังคับของ S: R เชื่อมั่นใน S หรือพบว่า S เชื่อถือได้

ข้อบังคับของ N+S: ความเข้าใจของ R ใน S ช่วยเพิ่มความเชื่อมั่นที่มีต่อ N

ผลกระทบ: ความเชื่อมั่นของ R ใน N เพิ่มมากขึ้น

จากตัวอย่าง (1) สามารถแบ่งช่วงของข้อความดังกล่าวได้ 2 ช่วง ได้แก่ ช่วงของข้อความ [อาจารย์จักรภาพต้องอยู่ที่นี้] ซึ่งมีบทบาทเป็นแกนกลาง (N) และช่วงของข้อความ [เพราะรถของเขาจอดอยู่หน้าคณะ] ซึ่งมีบทบาทเป็นบริวาร (S) หากพิจารณาตามข้อบังคับจะเห็นได้ว่าผู้อ่าน (R) อาจไม่เชื่อว่าอาจารย์จักรภาพอยู่ในตำแหน่งตามที่ข้อความระบุในระดับที่ผู้เขียนพอใจ แต่ในขณะเดียวกันช่วงของข้อความ [เพราะรถของเขาจอดอยู่หน้าคณะ] (S) ก็สามารถช่วยเป็นหลักฐานยืนยันความน่าเชื่อถือของช่วงของข้อความ [อาจารย์จักรภาพต้องอยู่ที่นี้] (N) ได้ตามข้อบังคับของ N+S เป็นผลให้ความเชื่อมั่นของผู้อ่าน (R) ที่มีต่อช่วงของข้อความ [อาจารย์จักรภาพต้องอยู่ที่นี้] (N) เพิ่มมากขึ้น

ในส่วนของการแทนรูปโครงสร้างปริจเฉท แมนน์และทอมป์สันเสนอให้ใช้**ผังภูมิโครงสร้าง (schema)** เพื่อแสดงการจัดองค์ประกอบของข้อความ แผนผังที่แมนน์และทอมป์สันเสนอนี้มีลักษณะเป็นแผนผังต้นไม้ที่เชื่อมโยงแต่ละช่วงข้อความเข้าด้วยกันโดยใช้เส้นแสดงความสัมพันธ์ และแต่ละเส้นจะถูกกำกับด้วยชื่อแบบของความสัมพันธ์ ดังภาพ 1

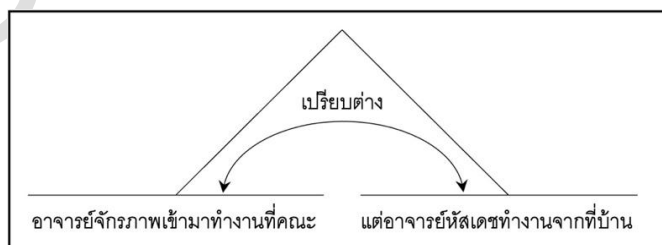


ภาพ 1 แผนผังต้นไม้อิงโครงสร้างปริเฉทของข้อความในตัวอย่าง (1)

ภาพ 1 แสดงความสัมพันธ์แบบแกนกลางเดียวระหว่างช่วงของข้อความในตัวอย่าง (1) ผ่านแผนผังต้นไม้อิงโครงสร้างปริเฉท จะเห็นได้ว่าข้อความ “อาจารย์จักรภาพต้องอยู่ที่นี้ เพราะรถของเขาจอดอยู่หน้าคณะ” สามารถแบ่งเป็นช่วงของข้อความได้ 2 ช่วง สังเกตเห็นได้จากเส้นตรงแนวนอนที่มีข้อความกำกับอยู่ ช่วงของข้อความที่เป็นแกนกลางจะมีเส้นตรงตั้งฉากกับเส้นตรงแนวนอนที่มีข้อความของช่วงนั้น ส่วนช่วงของข้อความที่เป็นบริวารจะมีเส้นโค้งที่ปลายด้านหนึ่งเป็นรูปลูกศรกำกับอยู่ โดยหัวลูกศรนี้จะชี้ไปยังเส้นแนวตั้งของช่วงของข้อความที่เป็นแกนกลางเพื่อแสดงบทบาทความเป็นบริวารของอีกช่วงของข้อความหนึ่ง และมีชื่อของความสัมพันธ์ “หลักฐาน” กำกับอยู่เหนือเส้นโค้งดังกล่าว จากแผนผังในภาพนี้จึงแปลความได้ว่าช่วงของข้อความ [เพราะรถของเขาจอดอยู่หน้าคณะ] มีบทบาทเป็นบริวาร ทำหน้าที่เป็นหลักฐานสนับสนุนช่วงของข้อความ [อาจารย์จักรภาพต้องอยู่ที่นี้] ซึ่งเป็นแกนกลาง

ส่วนความสัมพันธ์แบบหลายแกนกลางซึ่งแต่ละช่วงของข้อความล้วนมีบทบาทเป็นแกนกลางนั้น กำหนดให้ใช้เส้นทแยงฉากขึ้นไปสู่จุดแตกกิ่งเพื่อระบุบทบาทแกนกลาง ดังภาพ 2 ที่แสดงโครงสร้างปริเฉทของข้อความในตัวอย่าง (2) จากภาพ จะเห็นได้ว่าช่วงของข้อความ [อาจารย์จักรภาพเข้ามาทำงานที่คณะ] และ [แต่อาจารย์หัสเดชทำงานจากที่บ้าน] ต่างก็มีบทบาทเป็นแกนกลางที่เชื่อมโยงกันด้วยความสัมพันธ์แบบเปรียบเทียบ (contrast) อันสังเกตได้จากการกำกับด้วยเส้นโค้งที่ปลายทั้งสองด้านเป็นลูกศร

- (2) อาจารย์จักรภาพเข้ามาทำงานที่คณะแต่อาจารย์หัสเดชทำงานจากที่บ้าน

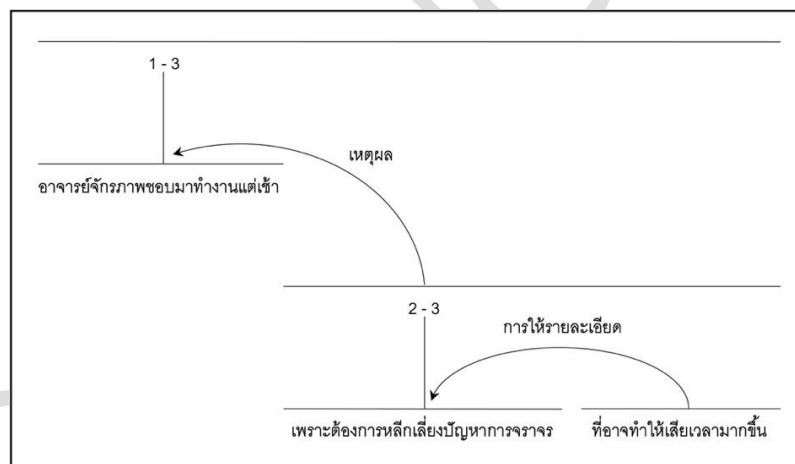


ภาพ 2 แผนผังต้นไม้อิงโครงสร้างปริเฉทของข้อความในตัวอย่าง (2)

แนวคิดเชิงทฤษฎีของทฤษฎีโครงสร้างวาทะได้รับการพัฒนาให้สามารถนำไปใช้จริงได้มากยิ่งขึ้นเมื่อคาร์ลสันและคณะ (Carlson, Marcu, & Okurowski, 2001) ได้เสนอว่า ในโครงสร้างปริเฉทมีหน่วยที่เล็กที่สุดซึ่งสามารถสื่อข้อมูลเชิงเนื้อหาและความหมายได้สมบูรณ์ เรียกว่า **หน่วยปริเฉทพื้นฐาน**

(Elementary Discourse Units: EDUs) และได้ใช้แนวคิดเรื่องหน่วยปริจเฉทพื้นฐานนี้แทนแนวคิดเรื่องช่วงของข้อความที่เสนอไว้ในทฤษฎีฉบับดั้งเดิม อย่างไรก็ตาม คาร์ลสันและคณะยังคงแนวคิดเรื่องสถานะความเป็นแกนกลางระหว่างหน่วยปริจเฉทพื้นฐานแต่ละหน่วยไว้ กล่าวคือ หน่วยปริจเฉทพื้นฐานแต่ละหน่วยในโครงสร้างปริจเฉทยังคงมีบทบาทเป็นแกนกลางหรือบริวารอยู่เช่นเดิม แต่สิ่งที่ถือว่าเป็นการพัฒนาในทฤษฎีฉบับนี้คือการกำหนดหลักการในการแบ่งขอบเขตของหน่วยปริจเฉทพื้นฐานที่ชัดเจน (Carlson & Marcu, 2001) หลักการดังกล่าวสัมพันธ์กับชุดของ **วาทสัมพันธ์** (rhetorical relation) ที่ได้เสนอขึ้นใหม่ทั้งหมด 16 กลุ่ม แบ่งเป็นประเภทของวาทสัมพันธ์ได้ 78 แบบ ยกตัวอย่างเช่น วาทสัมพันธ์แบบเปรียบเทียบ (contrast) วาทสัมพันธ์แบบลักษณะ (attribution) วาทสัมพันธ์แบบรายการ (list) วาทสัมพันธ์แบบภูมิหลัง (background) เป็นต้น และด้วยเหตุที่ทฤษฎีฉบับนี้ใช้การพิจารณาสถานะความเป็นแกนกลางควบคู่ไปกับประเภทของวาทสัมพันธ์จึงทำให้ไม่ต้องอาศัยข้อบังคับในการกำหนดประเภทของวาทสัมพันธ์อีกต่อไป ทั้งนี้ ขอให้พิจารณาการกำหนดสถานะความเป็นแกนกลางและประเภทของวาทสัมพันธ์ในข้อความในตัวอย่าง (3) เพื่อให้เข้าใจยิ่งขึ้น

- (3) อาจารย์จักรภาพชอบมาทำงานแต่เช้า เพราะต้องการหลีกเลี่ยงปัญหาการจราจรที่อาจทำให้เสียเวลามากขึ้น



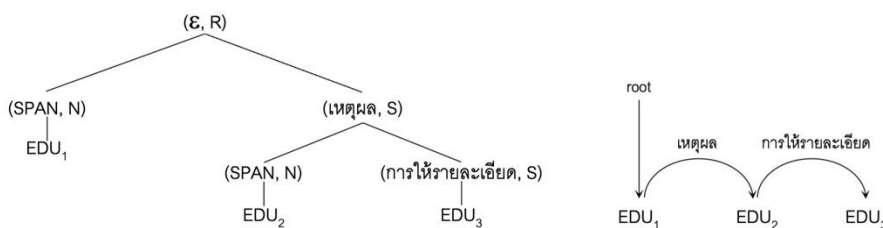
ภาพ 3 แผนผังต้นไม้โครงสร้างวาทะของข้อความในตัวอย่าง (3)

ภาพ 3 แสดงความสัมพันธ์ระหว่างหน่วยย่อยในปริจเฉทของข้อความในตัวอย่าง (3) ผ่าน **แผนผังต้นไม้โครงสร้างวาทะ** (Rhetorical Structure Tree: RS-Tree) จะเห็นได้ว่าข้อความดังกล่าวสามารถตัดแบ่งหน่วยปริจเฉทพื้นฐานได้เป็น 3 หน่วย ได้แก่ [อาจารย์จักรภาพชอบมาทำงานแต่เช้า] [เพราะต้องการหลีกเลี่ยงปัญหาการจราจร] และ [ที่อาจทำให้เสียเวลามากขึ้น] โดยระหว่างปริจเฉทพื้นฐาน [เพราะต้องการหลีกเลี่ยงปัญหาการจราจร] กับ [ที่อาจทำให้เสียเวลามากขึ้น] มีวาทสัมพันธ์แบบการให้รายละเอียด (elaboration) ซึ่งคาร์ลสันและมาร์คู (Carlson & Marcu, 2001, p. 33) ได้ระบุไว้ว่า ในวาทสัมพันธ์แบบนี้ หน่วยปริจเฉทที่เป็นบริวารจะให้ข้อมูลเพิ่มเติมเกี่ยวกับเนื้อหาของหน่วยปริจเฉทพื้นฐานที่เป็นแกนกลาง ในที่นี้จึงระบุได้ว่าหน่วยปริจเฉทพื้นฐาน [ที่อาจทำให้เสียเวลามากขึ้น] มีบทบาทเป็นบริวาร

เนื่องจากทำหน้าที่ให้ข้อมูลเพิ่มเติมเกี่ยวกับ “ปัญหาการจราจร” ในหน่วยปริศนาพื้นฐาน [เพราะต้องการหลีกเลี่ยงปัญหาการจราจร] และในลำดับขั้นที่สูงขึ้นไป หน่วยปริศนาพื้นฐาน [เพราะต้องการหลีกเลี่ยงปัญหาการจราจร] และ [ที่อาจทำให้เสียเวลามากขึ้น] ก็เป็นบริวารของหน่วยปริศนาพื้นฐาน [อาจารย์จักรภาพขอมาทำงานแต่เช้า] ตามที่คาร์ลสันและมาร์คู (Carlson & Marcu, 2001, p. 65) ระบุไว้ว่าในวาทสัมพันธ์แบบเหตุผล (reason) นั้น หน่วยปริศนาพื้นฐานที่เป็นแกนกลางจะต้องเป็นการกระทำของสิ่งมีชีวิต และหน่วยปริศนาพื้นฐานที่เป็นบริวารจะแสดงผลของการกระทำดังกล่าว จากตัวอย่างนี้จะเห็นได้ว่าทฤษฎีฉบับของคาร์ลสันและคณะนั้นได้พรรณนาหลักการในการตัดสินใจสถานะความเป็นแกนกลางของหน่วยปริศนาพื้นฐานไปพร้อมกับการระบุวาทสัมพันธ์ของหน่วยปริศนาพื้นฐานคู่หนึ่งๆ จึงไม่ต้องอาศัยข้อบังคับในการทำความเข้าใจประเภทของวาทสัมพันธ์ดังเช่นทฤษฎีฉบับของแมนน์และทอมป์สัน

ด้วยทฤษฎีฉบับที่เสนอโดยคาร์ลสันและคณะนี้มีหลักการในการกำหนดขอบเขตของหน่วยปริศนาพื้นฐานที่ชัดเจน อีกทั้งยังช่วยลดความยุ่งยากจากการตีความข้อบังคับในขั้นตอนการระบุประเภทของวาทสัมพันธ์ ทฤษฎีฉบับนี้จึงเป็นที่ยอมรับและถูกนำไปประยุกต์ใช้อย่างแพร่หลายในหลายภาษา รวมถึงภาษาไทย จนก่อให้เกิดการตีความและพัฒนาต่อยอดในเชิงทฤษฎีมาจนถึงปัจจุบัน

งานของมอเรียและคณะ (Morey, Muller, & Asher, 2018) เป็นงานอีกชิ้นหนึ่งที่พัฒนาต่อยอดแนวคิดเชิงทฤษฎีของทฤษฎีโครงสร้างวาทะเพื่อประยุกต์ใช้ในงานแจقس่วนปริศนา (discourse parsing) โดยเสนอว่างานแจقس่วนปริศนาตามแนวทฤษฎีโครงสร้างวาทะเกือบทั้งหมดได้หยิบยืมแนวคิดการแจقس่วนทางวากยสัมพันธ์เชิงส่วนประกอบ (syntactic constituency parsing) มาใช้ โดยกำกับข้อมูลและแทนรูปโครงสร้างปริศนาเป็นต้นไม้โครงสร้างวาทะแบบส่วนประกอบ (Marcu, 2000b) การกำกับข้อมูลและแทนรูปในลักษณะนี้ก่อให้เกิดปัญหาตามมาหลายประการทั้งในเชิงทฤษฎีและการใช้งาน กล่าวคือ ในเชิงทฤษฎี โครงสร้างวาทะแบบส่วนประกอบอาจก่อให้เกิดความกำกวมจากการตีความเชิงอรรถศาสตร์ได้ ส่วนในเชิงการใช้งานก็ส่งผลกระทบต่อประสิทธิภาพของระบบแจقس่วน ด้วยเหตุนี้ มอเรียและคณะจึงเสนอให้กำกับข้อมูลและแทนรูปโครงสร้างปริศนาโดยใช้ต้นไม้โครงสร้างวาทะแบบพึ่งพาอิงส่วนหลักและลำดับ (head-ordered dependency tree) แทน พร้อมทั้งได้พิสูจน์ว่าโครงสร้างแบบพึ่งพาสามารถแก้ไขปัญหาการแจقس่วนปริศนาได้ดีกว่าทั้งในเชิงทฤษฎีและการใช้งาน



ภาพ 4 ต้นไม้โครงสร้างวาทะแบบส่วนประกอบและต้นไม้โครงสร้างวาทะแบบพึ่งพาอิงส่วนหลักของข้อความในตัวอย่าง (3)

ภาพ 4 แสดงการเปรียบเทียบต้นไม้โครงสร้างวาทะแบบส่วนประกอบกับต้นไม้โครงสร้างวาทะแบบฟังก์ชันหลักและลำดับของข้อความในตัวอย่าง (3) ซึ่งแบ่งหน่วยปริจเฉทพื้นฐานในข้อความดังกล่าวได้เป็น 3 หน่วย จากภาพ แผนผังด้านซ้ายเป็นต้นไม้โครงสร้างวาทะแบบส่วนประกอบซึ่งมีการฉายภาพแบบทวิภาค (binary) ขึ้นไปเป็นจุดแตกกิ่ง แต่จุดแตกกิ่งจะกำกับด้วยค่า “SPAN” หรือประเภทวาทสัมพันธ์ และสถานะความเป็นแกนกลาง (N หรือ S) โดยมีจุดแตกกิ่งราก (root node) เป็นจุดแตกกิ่งสูงสุดในขณะที่แผนผังด้านขวาเป็นต้นไม้โครงสร้างวาทะแบบฟังก์ชันที่พิจารณาให้หน่วยปริจเฉทพื้นฐานที่มีบทบาทแกนกลางเป็นส่วนหลัก และมีเส้นโค้งปลายลูกศรชี้ไปยังส่วนฟังก์ชันอันได้แก่หน่วยปริจเฉทพื้นฐานที่มีบทบาทเป็นบริวาร

แนวคิดที่ปรากฏในงานมอเรียและคณะนี้นับว่าเป็นการพัฒนาแนวคิดเชิงทฤษฎีของทฤษฎีโครงสร้างวาทะไปอีกขั้นหนึ่ง เนื่องจากงานชิ้นนี้ได้เสนอให้พิจารณา**ความเป็นส่วนหลัก (headness)** ของหน่วยย่อยในโครงสร้างปริจเฉทที่ปรากฏในโครงสร้างแบบฟังก์ชัน อันเป็นการหยิบยืมแนวคิดด้านวากยสัมพันธ์มาใช้อธิบายข้อความในระดับปริจเฉทได้อย่างเป็นรูปธรรม กระทั่งเอื้อให้สามารถนำแนวการศึกษาทางวากยสัมพันธ์ฟังก์ชันมาประยุกต์ใช้ในการวิเคราะห์โครงสร้างปริจเฉทได้ และในแง่การใช้งานข้อค้นพบของมอเรียและคณะก็เป็นการบุกเบิกแนวทางใหม่ให้สามารถหยิบยืมแนวทางแจกส่วนทางวากยสัมพันธ์ฟังก์ชันมาใช้ในงานแจกส่วนปริจเฉทได้เช่นกัน

นอกจากนี้ยังมีงานของฮูและคณะ (Hou et al., 2020) ที่นำเสนอการตีความทฤษฎีโครงสร้างวาทะ โดยนำเสนอานิยามเชิงรูปนัยของแนวคิดที่เกี่ยวข้องกับทฤษฎีโครงสร้างวาทะเพื่อประโยชน์ในการประยุกต์ใช้ในงานแจกส่วนปริจเฉท ทั้งยังได้นำเสนอการจัดหมวดหมู่ของวิธีแจกส่วนปริจเฉทตามทฤษฎีนี้อีกด้วย งานชิ้นนี้ถือเป็นเครื่องยืนยันถึงความหนักแน่นเชิงทฤษฎีของทฤษฎีโครงสร้างวาทะได้เป็นอย่างดี ดังที่ฮูและคณะได้แสดงให้เห็นถึงการตีความทฤษฎีเพื่อใช้ต่อยอดในงานประมวลผลภาษาธรรมชาติได้อย่างกว้างขวาง

จากที่กล่าวมา จะเห็นได้ว่าตลอดระยะเวลา นับตั้งแต่ได้รับการเสนอขึ้น ทฤษฎีโครงสร้างวาทะได้รับความสนใจและมีการพัฒนาแนวคิดเชิงทฤษฎีเรื่อยมา ยังผลให้ทฤษฎีนี้มีจุดเด่นทั้งในเชิงภาษาศาสตร์และการประมวลผลภาษาธรรมชาติ กระทั่งปรากฏการประยุกต์ในงานแขนงต่างๆ อย่างแพร่หลาย ดังผู้เขียนจะได้กล่าวถึงโดยละเอียดในหัวข้อถัดไป

ทฤษฎีโครงสร้างวาทะ: จุดเด่นและการประยุกต์ใช้

จากมโนทัศน์พื้นฐานและพัฒนาการเชิงทฤษฎีของทฤษฎีโครงสร้างวาทะที่ได้กล่าวถึงในหัวข้อที่แล้ว จะเห็นได้ว่าทฤษฎีดังกล่าวนี้มีจุดเด่นในเชิงทฤษฎีที่เอื้อต่อการนำไปประยุกต์ต่อยอดได้อย่าง

กว้างขวาง จุดเด่นดังกล่าวคือการนำเสนอมุมมองในการทำความเข้าใจข้อความทั้งแง่วากยสัมพันธ์ ความหมาย และหน้าที่

ในแง่วากยสัมพันธ์นั้น ทฤษฎีโครงสร้างวาทะเสนอให้พิจารณาคูณสมบัติทางวากยสัมพันธ์ของหน่วยย่อยภายในโครงสร้างประโยคและนำมาใช้เป็นเกณฑ์ในการระบุขอบเขตของหน่วยประโยคพื้นฐาน โดยพิจารณาจากร่องรอย (clue) ทางศัพท์และวากยสัมพันธ์ และสรุปว่าในแง่วากยสัมพันธ์ หน่วยประโยคพื้นฐานอาจเป็นอนุพากย์หลัก อนุพากย์สัมพันธ์ ส่วนขยายหลังหน่วยนาม อนุพากย์ซ้อนที่แยกอนุพากย์หลักออกจากกัน รวมถึงวลีที่ขึ้นต้นด้วยตัวบ่งชี้ประโยคแข็ง (strong discourse marker) ในฐานะหน่วยประโยคพื้นฐานแบบวลี (phrasal EDU) (Carlson et al., 2001, p. 3) การพิจารณาคูณสมบัติทางวากยสัมพันธ์นี้เอื้อให้สามารถระบุขอบเขตของหน่วยประโยคพื้นฐานได้อย่างคงเส้นคงวา (consistent) สามารถนำไปประยุกต์ใช้ในงานประมวลผลภาษาธรรมชาติที่ต้องอาศัยผู้กำกับข้อมูลจำนวนมากได้ หรือกระทั่งนำไปศึกษาต่อยอดเชิงภาษาศาสตร์ในประเด็นด้านการต่อประสานระหว่างวากยสัมพันธ์กับประโยค (syntax-discourse interface) ก็ย่อมได้

ในแง่ความหมายนั้น ดังได้กล่าวไปแล้วว่าหน่วยประโยคพื้นฐานเป็นหน่วยที่ถูกนำเสนอขึ้นเพื่อใช้สื่อข้อมูลเชิงเนื้อหาและความหมายที่สมบูรณ์ในตัวเอง ทั้งนี้ ลักษณะการสื่อความหมายดังกล่าวนี้ก็เป็นผลมาจากการกำหนดเกณฑ์ทางวากยสัมพันธ์ที่ชัดเจน ในแง่นี้ หน่วยประโยคพื้นฐานจึงเป็นหน่วยที่มีเนื้อหาประพจน์ (propositional content) สมบูรณ์และสามารถใช้แทนข้อมูลที่เป็นความคิดสำคัญ (idea) เพียงความคิดเดียวได้ (ศุภวัจน์ แต่รุ่งเรือง, 2560, น. 72) จุดเด่นทางความหมายของหน่วยประโยคพื้นฐานนี้ก็เป็นอีกเหตุผลหนึ่งที่ทำให้งานประมวลผลภาษาธรรมชาติเลือกประยุกต์ใช้ทฤษฎีโครงสร้างวาทะในแขนงงานที่เน้นการประมวลผลความหมาย เช่น การวิเคราะห์ความรู้สึก หรือการสรุปย่อเอกสาร ยิ่งไปกว่านั้น ด้วยคุณสมบัติทางความหมายที่ผูกโยงกับโครงสร้างทางวากยสัมพันธ์ที่ชัดเจนนี้ ทำให้หน่วยประโยคพื้นฐานได้รับเลือกนำไปใช้เป็นหน่วยพื้นฐานในการประมวลผลแทนประโยค โดยเฉพาะอย่างยิ่ง ในภาษาที่ไม่สามารถระบุขอบเขตของประโยคได้ชัดเจน เช่น ภาษาไทย

ส่วนในแง่หน้าที่นั้น กล่าวได้ว่าเป็นการผสมผสานเอาจุดเด่นของทฤษฎีด้านวากยสัมพันธ์และความหมายมาใช้ประโยชน์ในการวิเคราะห์ข้อความในระดับประโยค กล่าวคือ ทฤษฎีโครงสร้างวาทะนำเสนอการพิจารณาหน้าที่ของหน่วยย่อยในโครงสร้างประโยคโดยใช้เกณฑ์ทางวากยสัมพันธ์เพื่อระบุขอบเขตของหน่วยประโยคพื้นฐาน และระบุหน้าที่ที่แต่ละหน่วยประโยคพื้นฐานมีต่อกันโดยพิจารณาความหมาย กลายเป็นวาทสัมพันธ์ประเภทต่างๆ ที่เชื่อมโยงหน่วยประโยคพื้นฐานเข้าไว้ด้วยกันเป็นข้อความ ทั้งนี้ หากพิจารณาในเชิงประโยควิเคราะห์ก็อาจกล่าวได้ว่า ทฤษฎีโครงสร้างวาทะสามารถอธิบายเรื่องการเกาะเกี่ยวความ (coherence) ผ่านประเภทของวาทสัมพันธ์ซึ่งระบุความสัมพันธ์ทางความหมายระหว่างหน่วยประโยคพื้นฐาน และยังสามารถอธิบายเรื่องการเชื่อมโยงความ (cohesion) ผ่านการพิจารณาดัชนี (cue) ทางศัพท์และวากยสัมพันธ์ที่ปรากฏในหน่วยประโยคพื้นฐานแต่ละหน่วย ซึ่ง

ทั้งสองส่วนนี้สามารถสะท้อนจุดเด่นของทฤษฎีโครงสร้างวาทะว่าด้วยการวิเคราะห์หน้าที่ของหน่วยต่างๆ ในบริบทได้เป็นอย่างดี

จุดเด่นเชิงทฤษฎีดังได้กล่าวมาข้างต้นได้ส่งผลให้ทฤษฎีโครงสร้างวาทะได้รับการนำไปประยุกต์ใช้งานในหลายแขนง ไม่เฉพาะงานประมวลผลภาษาธรรมชาติเท่านั้น แต่ยังรวมถึงงานที่มุ่งวิเคราะห์ข้อความเพื่อประโยชน์ด้านการเรียนการสอนการเขียนด้วย

ในด้านการนำไปประยุกต์ใช้ในงานประมวลผลภาษาธรรมชาติ ดังได้กล่าวไปแล้วว่าเมื่อแรกเริ่มนั้น ทฤษฎีโครงสร้างวาทะได้รับเสนอขึ้นเพื่อใช้ในการสร้างข้อความ (text generation) ด้วยคอมพิวเตอร์ แนวคิดหลักของงานแขนงนี้คือการวางโครงของการเกาะเกี่ยวความภายในข้อความ ด้วยเหตุนี้ ทฤษฎีโครงสร้างวาทะซึ่งอธิบายการเกาะเกี่ยวความผ่านวาทสัมพันธ์จึงมีบทบาทสำคัญในงานแขนงนี้ ดังปรากฏงานสร้างข้อความที่ประยุกต์ใช้ทฤษฎีโครงสร้างวาทะเป็นจำนวนมาก เช่น งานของวาล์สเตอร์และคณะ (Wahlster, Andre, Graf, & Rist, 1991) หรืองานของโอดอนเนลล์และคณะ (O'Donnell, Mellish, Oberlander, & Knott, 2001) เป็นต้น

การแจ่งส่วนบริบทเป็นงานแขนงหนึ่งที่น่าแนวคิดเรื่องความสัมพันธ์ระหว่างหน่วยย่อยภายในโครงสร้างบริบทที่ปรากฏในทฤษฎีโครงสร้างวาทะไปใช้อย่างแพร่หลายมากที่สุด งานแขนงนี้มีวัตถุประสงค์เพื่อให้คอมพิวเตอร์วิเคราะห์โครงสร้างบริบทของข้อความ ผลลัพธ์ที่ได้จากงานแขนงนี้จะอยู่ในรูปต้นไม้โครงสร้างบริบทซึ่งมักถูกรวบรวมเป็นคลังต้นไม้บริบท (discourse treebank) และสามารถนำไปใช้ต่อยอดในงานแขนงอื่นๆ ได้ เช่น งานสรุปย่อข้อความ (text summarization) หรืองานสกัดสารสนเทศ (information extraction) ยกตัวอย่างงานแจ่งส่วนบริบทที่ได้รับการยอมรับ เช่น งานของมาร์คู (Marcu, 2000a) และงานของบรอดและคณะ (Braud, Coavoux, & Søgaard, 2017)

การแจ่งส่วนบริบทตามแนวทฤษฎีโครงสร้างวาทะจะทำให้ทราบถึงความสำคัญของแต่ละส่วนของข้อความผ่านสถานะความเป็นแกนกลางของหน่วยบริบทพื้นฐาน อีกทั้งยังแสดงการเกาะเกี่ยวความระหว่างกันผ่านวาทสัมพันธ์ ผลลัพธ์ดังกล่าวนี้สามารถนำไปประยุกต์ใช้ต่อในงานสรุปย่อข้อความและการสกัดสารสนเทศ โดยกำหนดให้คอมพิวเตอร์เลือกเก็บหน่วยบริบทพื้นฐานที่เป็นแกนกลางไว้ เนื่องจากเป็นไปได้ว่าหน่วยบริบทพื้นฐานที่เป็นแกนกลางให้ข้อมูลที่มีความสำคัญมากกว่าหน่วยบริบทพื้นฐานที่เป็นบริวาร โดยอาศัยแนวคิดนี้ ข้อความที่ได้จากงานสรุปย่อข้อความและการสกัดสารสนเทศจึงมีเฉพาะใจความสำคัญและตรงตามความต้องการของผู้ใช้งาน ทั้งนี้ สามารถยกตัวอย่างงานในกลุ่มนี้ได้ เช่น งานของฮิระโอะและคณะ (Hirao, Yoshida, Nishino, Yasuda, & Nagata, 2013) และงานของราฮังเดลและอัครวัล (Rahangdale & Agrawal, 2014)

การวิเคราะห์ความรู้สึก (sentiment analysis) เป็นงานประมวลผลข้อความแขนงหนึ่งที่รับเอาแนวคิดและเทคนิคต่างๆ ไปประยุกต์ใช้อย่างหลากหลาย ทฤษฎีโครงสร้างวาทะก็ได้ถูกนำไปประยุกต์ใช้ในงานแขนงนี้ด้วย โดยใช้ประโยชน์จากแนวคิดเรื่องลำดับชั้นของโครงสร้างบริบทซึ่งแสดงความสัมพันธ์

ทางความหมายระหว่างข้อความแต่ละตอนของเอกสาร เป็นผลให้คอมพิวเตอร์สามารถวิเคราะห์ความรู้สึกในระดับเอกสารได้ วิธีการนี้ต่างจากแนวทางเดิมที่วิเคราะห์ความรู้สึกจากการพิจารณาความถี่ของคำศัพท์ที่แสดงความรู้สึกขั้วบวกหรือลบในข้อความ ทั้งนี้ เมื่อประยุกต์ใช้แนวคิดเรื่องลำดับชั้นของโครงสร้างปริจเฉทแล้ว ข้อความแต่ละตอนย่อมมีความสำคัญไม่เท่าเทียมกันตามสถานะความเป็นแกนกลางและความลึกของลำดับชั้นโครงสร้างปริจเฉท และมีความเป็นไปได้ว่าความรู้สึกที่ปรากฏในข้อความตอนที่สำคัญกว่าย่อมสะท้อนความรู้สึกโดยรวมในระดับเอกสาร ยกตัวอย่างงานที่อาศัยแนวคิดดังกล่าวนี้ เช่น งานของโวลล์และทาโบาดา (Voll & Taboada, 2007) งานของภาทิเยและคณะ (Bhatia, Ji, & Eisenstein, 2015) และงานของเคราส์และไฟเออร์ริเกล (Kraus & Feuerriegel, 2019) เป็นต้น

ส่วนในด้านการเรียนการสอนการเขียนนั้นก็ปรากฏงานวิจัยจำนวนมากที่ประยุกต์แนวคิดของทฤษฎีโครงสร้างวาทะเพื่อวิเคราะห์และทำความเข้าใจงานเขียนผ่านมุมมองด้านปริจเฉทวิเคราะห์ ยกตัวอย่างเช่นงานของอามาดีและปาร์ฮิซการ์ (Ahmadi & Parhizgar, 2017) ที่อาศัยแนวคิดเรื่องวาทสัมพันธ์ในการศึกษาข้อผิดพลาดด้านการเกาะเกี่ยวความในงานเขียนของชาวอิหร่านผู้เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศ โดยเปรียบเทียบข้อผิดพลาดด้านการเกาะเกี่ยวความที่พบในงานเขียนแบบพรรณนากับงานเขียนแบบแสดงเหตุผล งานของบราวน์และเวอร์ฮาเกน (Brown & Verhagen, 2018) ซึ่งศึกษาความแตกต่างระหว่างงานเขียนของผู้เรียนภาษาอังกฤษชาวญี่ปุ่นกับงานเขียนของผู้พูดภาษาอังกฤษเป็นภาษาแม่ โดยอาศัยการวิเคราะห์ค่าความถี่ของวาทสัมพันธ์ที่พบในงานเขียนดังกล่าว หรืองานของสกูฟาเกี (Skoufaki, 2020) ที่อาศัยหลักการสร้างแผนผังต้นไม้โครงสร้างวาทะในการบ่งชี้การหยุดการเกาะเกี่ยวความ (coherence break) ในงานเขียนของผู้เรียนภาษาอังกฤษชาวไต้หวัน

จากที่ได้กล่าวมาในหัวข้อนี้ จะเห็นได้ว่าทฤษฎีโครงสร้างวาทะนั้นมีจุดเด่นที่เอื้อต่อการทำความเข้าใจข้อความทั้งในเชิงโครงสร้างและความหมาย จุดเด่นดังกล่าวยังให้เกิดการประยุกต์ใช้ทฤษฎีในงานต่างๆ หลากหลายแขนง มิใช่เฉพาะในงานสร้างข้อความอิงคอมพิวเตอร์อันเป็นจุดเริ่มต้นของการเสนอทฤษฎีนี้เท่านั้น แต่ยังขยายขอบเขตไปถึงงานประมวลผลภาษาธรรมชาติแขนงอื่นที่พัฒนาขึ้นตามความก้าวหน้าของเทคโนโลยีในแต่ละช่วงระยะเวลา นอกจากนี้ ทฤษฎีโครงสร้างวาทะยังได้รับการพิจารณาในฐานะแนวคิดหนึ่งในด้านปริจเฉทวิเคราะห์ซึ่งสามารถนำไปใช้ประโยชน์ในด้านการเรียนการสอนภาษาได้อีกด้วย การประยุกต์ใช้ที่หลากหลายนี้เป็นเครื่องยืนยันถึงความหนักแน่นและความเป็นสากลของทฤษฎีโครงสร้างวาทะได้เป็นอย่างดี

ทฤษฎีโครงสร้างวาทะกับการวิเคราะห์และประมวลผลภาษาไทย

จุดเด่นของทฤษฎีโครงสร้างวาทะที่ได้กล่าวถึงในหัวข้อที่แล้วเป็นปัจจัยสำคัญที่ผลักดันให้เกิดการศึกษาทำความเข้าใจภาษาต่างๆ ตามกรอบแนวคิดของทฤษฎีโครงสร้างวาทะอย่างแพร่หลาย (Taboada & Mann, 2006, p. 572) ทั้งนี้ ภาษาไทยก็เป็นภาษาหนึ่งที่ได้รับการวิเคราะห์ตีความและ

ประมวลผลผ่านมุมมองของทฤษฎีดังกล่าวนี้ กระทั่งเกิดเป็นงานวิจัยที่ยังประโยชน์ต่อวงการประมวลผลภาษาไทยจำนวนมาก ในหัวข้อนี้จะได้กล่าวถึงแนวทฤษฎีวิเคราะห์และประมวลผลภาษาไทยผ่านงานวิจัยเหล่านั้นเพื่อประมวลเป็นฐานความรู้ในการศึกษาวิจัยต่อไป

จากการทบทวนงานที่เกี่ยวข้องกับการใช้ทฤษฎีโครงสร้างวาทะในการวิเคราะห์และประมวลผลภาษาไทย สามารถแบ่งงานวิจัยที่เกี่ยวข้องได้เป็น 2 กลุ่ม ได้แก่ กลุ่มที่วิเคราะห์ภาษาไทยตามกรอบแนวคิดของทฤษฎีโครงสร้างวาทะเพื่อใช้ประโยชน์ในการระบุขอบเขตของหน่วยปริจเฉทพื้นฐานและแฉงส่วนปริจเฉท และกลุ่มที่ประยุกต์ใช้ทฤษฎีโครงสร้างวาทะเป็นส่วนหนึ่งของการพัฒนาระบบประมวลผลภาษาไทย

งานในกลุ่มแรกที่ว่าด้วยการวิเคราะห์ภาษาไทยตามแนวทฤษฎีโครงสร้างวาทะเพื่อระบุขอบเขตของหน่วยปริจเฉทพื้นฐานและแฉงส่วนปริจเฉทนั้นเป็นงานที่ได้รับความสนใจศึกษาและพัฒนาเรื่อยมา กระทั่งปัจจุบัน เนื่องด้วยในการประมวลข้อความภาษาไทยนั้น หน่วยปริจเฉทพื้นฐานมีบทบาทสำคัญในฐานะหน่วยทางภาษาที่ใช้ในการประมวลผลแทนประโยค

ทั้งนี้ อาจกล่าวได้ว่าการตัดประโยค (sentence segmentation) ภาษาไทยนั้นเป็นข้อถกเถียงในวงการประมวลผลภาษาไทยเรื่อยมา ด้วยประโยคในภาษาไทยไม่สามารถระบุขอบเขตที่แน่นอนได้ (Intasaw & Aroonmanakun, 2013, p. 491; Lertpiya et al., 2018, p. 3; Saetia et al., 2020, p. 270) ต่างจากภาษาอังกฤษที่สามารถใช้เครื่องหมายวรรคตอนในการระบุขอบเขตของประโยคในการตัดประโยคได้ ด้วยเหตุนี้ หากต้องการนำประโยคภาษาไทยมาใช้เป็นหน่วยในการประมวลผลก็ย่อมก่อให้เกิดความยุ่งยากในการให้นิยามและระบุขอบเขตของประโยคซึ่งอาจมีได้หลากหลายตามแนวทฤษฎีไวยากรณ์สำนักต่างๆ (ศุภวัจน์ แต่รุ่งเรือง, 2560, น. 72) แนวคิดนี้สอดคล้องกับข้อเสนอของวิโรจน์ อรุณมานะกุล (2562) ที่กล่าวว่า การตัดประโยคภาษาไทยนั้นทำได้ยาก และแม้จะสามารถแยกประโยคออกมาได้ ในการประมวลผลปริจเฉทก็ยังคงต้องพิจารณาประโยคทั้งหมด รวมถึงอนุพากย์ และความสัมพันธ์ระหว่างอนุพากย์ในประโยคซับซ้อนอยู่ จึงไม่จำเป็นต้องตัดประโยค หากต้องการตัดหน่วยที่ใหญ่กว่าคำหรือวลีเพื่อใช้ในการประมวลผลก็ควรตัดเป็นหน่วยปริจเฉทพื้นฐานเพื่อให้ได้เนื้อความแต่ละส่วนสำหรับนำไปประมวลผลต่อไป

ด้วยเหตุที่กล่าวมาข้างต้น การวิเคราะห์ภาษาไทยตามแนวทฤษฎีโครงสร้างวาทะเพื่อระบุขอบเขตของหน่วยปริจเฉทพื้นฐานจึงเป็นที่สนใจอย่างมากในวงการประมวลผลภาษาไทย ดังปรากฏให้เห็นในงานของจิรวรรณ เจริญสุข และคณะ (Charoensuk, Suvakree, & Kawtrakul, 2005) ซึ่งถือเป็นงานชิ้นแรกๆ ที่วิเคราะห์ภาษาไทยตามแนวทฤษฎีโครงสร้างวาทะเพื่อตัดหน่วยปริจเฉทพื้นฐาน งานชิ้นนี้ได้ชี้ให้เห็นปัญหาของการตัดหน่วยปริจเฉทพื้นฐานในภาษาไทย จากนั้นจึงอาศัยวิธีการเรียนรู้ของเครื่องแบบต้นไม้ตัดสินใจ (decision tree) ร่วมกับกฎที่สร้างขึ้นโดยผู้เชี่ยวชาญ พัฒนาระบบตัดหน่วยปริจเฉทพื้นฐานได้สำเร็จ งานของนงนุช เกตุย และคณะ (Ketui, Theeramunkong, & Onsuwan, 2012) ที่เริ่มต้นจากวิเคราะห์

ลักษณะของภาษาไทย จากนั้นจึงนำเสนอนิยามของหน่วยปริจเฉทพื้นฐานภาษาไทยและกฎไวยากรณ์ที่สร้างขึ้นจากชุดของไวยากรณ์ไม่พึงบริบท และพัฒนาตัวแจงส่วนแผนภูมิ (chart parser) ถึงลำดับที่ยาวที่สุด โดยประยุกต์กฎไวยากรณ์ที่สร้างขึ้นเพื่อแจงส่วนข้อความให้ได้หน่วยปริจเฉทพื้นฐานออกมาเป็นผลลัพธ์ และงานของอรรถพล คงหวาน และคณะ (Kongwan, Kamaruddin, & Ahmad, 2020) ที่อาศัยการวิเคราะห์ลักษณะทางวากยสัมพันธ์ของภาษาไทยในการตัดหน่วยปริจเฉทพื้นฐาน โดยเสนอกระบวนการตัดหน่วยปริจเฉทพื้นฐานที่ใช้การพิจารณาตัวบ่งชี้ระบุนัย รูปแบบทางวากยสัมพันธ์ และกฎที่ใช้สร้างหน่วยปริจเฉทพื้นฐาน ร่วมกัน นอกจากนี้ยังมีงานของสมนึก สิ้นธุปวน และโอม ศรีนิล (Sinthupoun & Sornil, 2010) ที่ไม่เพียงเป็นการเสนอวิธีการตัดหน่วยปริจเฉทพื้นฐานด้วยแบบจำลองฮิดเดนมาร์คอฟเท่านั้น แต่ยังสามารถวิเคราะห์โครงสร้างปริจเฉทในภาษาไทยตามแนวทฤษฎีโครงสร้างวาทะเพื่อสร้างต้นไม้โครงสร้างวาทะที่มีการแสดงวาทสัมพันธ์โดยใช้วิธีการเรียนรู้ของเครื่องแบบต้นไม้ตัดสินใจอีกด้วย

ในบรรดางานที่ว่าด้วยการตัดหน่วยปริจเฉทพื้นฐานภาษาไทยนี้ยังปรากฏงานของนลินี อินตะชาว และวิโรจน์ อรุณมานะกุล (Intasaw & Aroonmanakun, 2013) ที่เสนอหลักการตัดหน่วยปริจเฉทพื้นฐานภาษาไทย ความน่าสนใจของงานชิ้นนี้คือความพยายามที่จะประยุกต์หลักการตัดหน่วยปริจเฉทพื้นฐานภาษาอังกฤษที่เสนอในงานของคาร์ลสันและมาร์คู (Carlson & Marcu, 2001) มาใช้กับภาษาไทย งานชิ้นนี้จึงมีจุดเด่นด้านการวิเคราะห์คุณสมบัติความเป็นอนุภาคของหน่วยปริจเฉทพื้นฐานในภาษาไทยเทียบกับภาษาอังกฤษ หลักการที่เสนอในงานชิ้นนี้ยังได้รับการขยายความให้ชัดเจนยิ่งขึ้นในวิทยานิพนธ์ปริญญาโทของนลินี อินตะชาว (2556) ที่เสนอการตัดหน่วยปริจเฉทพื้นฐานด้วยวิธีการเรียนรู้ของเครื่องแบบซัพพอร์ตเวกเตอร์แมชชีน ทั้งนี้ แม้เทคนิคทางคอมพิวเตอร์ที่ใช้จะถูกปรับเปลี่ยนให้พัฒนาก้าวหน้าไป แต่หลักการตัดหน่วยปริจเฉทพื้นฐานที่เสนอในงานชิ้นนี้ยังคงได้รับการยอมรับในฐานะแนวทาง (guideline) ในการกำกับข้อมูลหน่วยปริจเฉทพื้นฐานภาษาไทยที่ปรากฏในงานหลายชิ้น เช่น งานของศุภวัจน์ แต่รุ่งเรือง และวิโรจน์ อรุณมานะกุล (2558) หรืองานของสตัยา สิงห์กุล และคณะ (Singkul, Khampingyot, Maharattamalai, Taerungruang, & Chalothorn, 2019) เป็นต้น

งานของชนาธิป แซ่เตีย และคณะ (Saetia et al., 2020) ก็เป็นงานอีกชิ้นหนึ่งที่น่าหลักการกำหนดขอบเขตของหน่วยปริจเฉทพื้นฐานในภาษาไทยที่เสนอโดยนลินีและวิโรจน์ (Intasaw & Aroonmanakun, 2013) มาใช้ และชี้ให้เห็นว่าการตัดหน่วยปริจเฉทพื้นฐานตามหลักการดังกล่าวทำให้ได้หน่วยปริจเฉทพื้นฐานบางส่วนที่ขาดความสมบูรณ์ทางโครงสร้างหรือความหมาย เนื่องจากหลักการดังกล่าวยึดการพิจารณาลักษณะทางวากยสัมพันธ์ของหน่วยปริจเฉทพื้นฐานเป็นสำคัญ งานชิ้นนี้จึงเสนอให้พิจารณาลักษณะทางวากยสัมพันธ์ร่วมกับลักษณะความหมายผ่านวาทสัมพันธ์ 4 ประเภท เพื่อผสานหน่วยปริจเฉทพื้นฐานที่ถูกตัดแล้วเข้าด้วยกันให้ได้หน่วยที่มีความสมบูรณ์ทางโครงสร้างและความหมาย พร้อมใช้ในการประมวลผลในขั้นต่อไป ยิ่งไปกว่านั้น งานชิ้นนี้ยังทดลองพัฒนาแบบจำลองเพื่อตัดหน่วยปริจเฉทพื้นฐาน

โดยให้ระบบพิจารณาประเภทของวาทสัมพันธ์ผ่านการเรียนรู้เชิงลึกของเครื่อง ผลปรากฏว่าแบบจำลองสามารถตัดหน่วยปริจเฉทพื้นฐานที่มีความสมบูรณ์ทั้งโครงสร้างและความหมายได้อย่างมีประสิทธิภาพในระดับที่น่าพอใจ

เมื่อพิจารณางานในกลุ่มแรกนี้ จะเห็นได้ถึงความพยายามที่จะวิเคราะห์ภาษาไทยทั้งในเชิงวาทสัมพันธ์ ปริจเฉท และความหมาย ผ่านการตีความแนวทฤษฎีโครงสร้างวาทะในเชิงรูปนัยที่เอื้อต่อการทำงานของระบบคอมพิวเตอร์และในเชิงภาษาศาสตร์ที่เอื้อต่อการทำความเข้าใจโครงสร้างปริจเฉทของภาษาไทย ข้อค้นพบที่ปรากฏในงานกลุ่มนี้ถือเป็นก้าวสำคัญของการวิเคราะห์และประมวลผลภาษาไทย

ส่วนงานอีกกลุ่มหนึ่งที่มุ่งประยุกต์ใช้ทฤษฎีโครงสร้างวาทะในการพัฒนาระบบประมวลผลภาษาไทยนั้น พบว่างานส่วนมากเป็นไปเพื่อการพัฒนาระบบสรุปย่อข้อความ ดังปรากฏในงานของธนา สุขวารี และคณะ (Sukvaree et al., 2004) ที่ได้เสนอการสรุปย่อข้อความในวงการเกษตรโดยใช้แนวคิดของทฤษฎีโครงสร้างวาทะร่วมกับโครงสร้างความรู้ในออนโทโลยีการเกษตรที่ว่าด้วยโรคพืช งานของเมธี วัฒนเมธานนท์ และคณะ (Wattanamethanont et al., 2005) ที่เสนอวิธีการระบุความสัมพันธ์ในปริจเฉทโดยใช้แบบจำลองนาอิมฟ์เป็นตัวแทนประเภทจากลักษณะที่ประกอบไปด้วยตัวบ่งชี้ความสัมพันธ์ในปริจเฉท วลีสำคัญ และการปรากฏร่วมของคำ งานของธนา สุขวารี และคณะ (Sukvaree, Kawtrakul, & Caelen, 2007) ที่เสนอการสรุปย่อข้อความจากการเกาะเกี่ยวความในข้อความ โดยยุบรวมวาทสัมพันธ์ตามแนวทฤษฎีโครงสร้างวาทะให้เหลือ 2 ประเภท ได้แก่ การเชื่อมความ และการซ้อนความ จากนั้นจึงสร้างโครงสร้างต้นไม้การเกาะเกี่ยวเพื่อใช้ในการสรุปย่อเอกสาร และงานของนงนุช เกตุย และคณะ (Ketui et al., 2015) ที่นำเสนอแบบจำลองในการสรุปย่อเอกสารหลายชิ้น (multi-document) โดยอิงจากหน่วยปริจเฉทพื้นฐานภาษาไทย โดยแบบจำลองที่เสนอในงานชิ้นนี้มีกระบวนการทำงาน 3 ชั้น คือขั้นการตัดหน่วยข้อความที่ใช้ในการสรุปย่อข้อความ ขั้นการสร้างกราฟของหน่วยข้อความเพื่อใช้ระบุความสัมพันธ์ของหน่วยข้อความต่างๆ สำหรับเลือกมาใช้ในการสรุป และขั้นการเลือกหน่วยข้อความและการสร้างการสรุป

นอกจากงานที่เกี่ยวข้องกับการพัฒนาระบบสรุปย่อข้อความแล้ว ยังปรากฏงานอีกกลุ่มหนึ่งที่ประยุกต์ใช้ทฤษฎีโครงสร้างวาทะเป็นส่วนหนึ่งของการวิจัย ได้แก่ งานของศุภวัจน์ แต่รุ่งเรือง และวิโรจน์ อรุณมานะกุล (2558) ที่ศึกษากลไกทางภาษาที่ใช้ในการลักลอกงานวิชาการ (plagiarism) ภาษาไทยโดยใช้ทฤษฎีโครงสร้างวาทะเป็นกรอบการวิเคราะห์ งานของศุภวัจน์ แต่รุ่งเรือง (2560) ที่ทดลองเปรียบเทียบประสิทธิภาพของระบบตรวจหาการลักลอกทางวิชาการเมื่อใช้หน่วยรับเข้า (input) ที่แตกต่างกันระหว่างหน่วยปริจเฉทพื้นฐานกับย่อหน้า และงานของสิงห์กุล และคณะ (Singkul et al., 2019) ที่ทดลองแจกแจงส่วนข้อความที่ปรากฏในสื่อสังคมออนไลน์ตามแนวการแจกแจงส่วนทางวาทสัมพันธ์แบบฟิงฟา โดยใช้หน่วยปริจเฉทพื้นฐานเป็นหน่วยรับเข้าแทนประโยค

จากงานที่ยกมานี้ จะเห็นได้ว่าในภาษาไทย ทฤษฎีโครงสร้างวาทะได้รับการประยุกต์ใช้ในแขนงงานสรุปย่อข้อความมากที่สุด ทั้งนี้ เนื่องมาจากงานแขนงดังกล่าวต้องอาศัยการพิจารณาความสัมพันธ์ภายในโครงสร้างปริเฉทของข้อความ ลักษณะดังกล่าวนี้สอดคล้องโดยตรงกับแนวคิดของทฤษฎีโครงสร้างวาทะ อย่างไรก็ดี เป็นที่น่าสังเกตว่าในระยะหลัง งานแขนงอื่นๆ ก็เริ่มรับเอาแนวคิดของทฤษฎีโครงสร้างวาทะไปประยุกต์ใช้เป็นกรอบการศึกษาวิเคราะห์และพัฒนาเครื่องมือประมวลผลภาษาไทยมากขึ้นแล้ว แนวโน้มดังกล่าวนี้ นับว่าเป็นนิมิตรหมายอันดีของวงการประมวลผลภาษาไทยและวงการภาษาศาสตร์ในการพัฒนาองค์ความรู้และวิทยาการให้ก้าวหน้าต่อไป

ก้าวต่อไปของทฤษฎีโครงสร้างวาทะกับภาษาไทย

จากงานวิจัยที่ได้ยกมาในหัวข้อที่แล้ว หากพิจารณาเปรียบเทียบความก้าวหน้าของการประยุกต์ใช้ทฤษฎีโครงสร้างวาทะในการวิเคราะห์และประมวลผลภาษาไทยกับภาษาอื่นๆ โดยเฉพาะภาษาอังกฤษ ก็อาจกล่าวได้ว่าวงวิชาการภาษาไทยที่เกี่ยวข้องกับศาสตร์แขนงนี้ได้ก้าวออกมาจากจุดเริ่มต้นได้ระยะหนึ่งแล้ว อย่างไรก็ดี เมื่อพิจารณาจากแง่มุมเชิงทฤษฎีและประโยชน์ที่คาดว่าจะได้รับจากการประยุกต์ใช้ทฤษฎีโครงสร้างวาทะแล้ว ผู้เขียนเห็นว่าวงวิชาการภาษาไทยยังสามารถก้าวต่อไปได้อีกไกลโดยอาศัยแนวคิดจากทฤษฎีดังกล่าวนี้ ดังนั้นในหัวข้อนี้ ผู้เขียนจะได้ชี้ให้เห็นถึงประเด็นปัญหาที่เป็นช่องว่างทางการวิจัยว่าด้วยทฤษฎีโครงสร้างวาทะกับภาษาไทย เพื่อประโยชน์ในการพัฒนาแนวทางวิจัยด้านภาษาศาสตร์และการประมวลผลภาษาไทยในอนาคต ทั้งนี้ จะได้แบ่งกล่าวถึงประเด็นปัญหาเป็น 3 แง่มุม ได้แก่ การประมวลผลภาษาไทย การวิเคราะห์ทางภาษาศาสตร์ภาษาไทย และการเรียนการสอนการเขียนภาษาไทย รายละเอียดมีดังต่อไปนี้

ในแง่การประมวลผลภาษาไทย จะสังเกตได้ว่าแนวโน้มปัจจุบันของงานประมวลผลภาษาไทย มุ่งเน้นการประมวลผลภาษาไทยที่ปรากฏในสื่อสังคมออนไลน์ต่างๆ เพื่อใช้ประโยชน์ในเชิงพาณิชย์เป็นสำคัญ ทั้งนี้ กล่าวได้ว่าลักษณะภาษาที่ปรากฏในสื่อออนไลน์นั้นแตกต่างจากภาษาเขียนที่ปรากฏใช้ในเอกสารทั่วไปค่อนข้างมาก เป็นเหตุให้การประมวลผลภาษาไม่ได้เป็นไปโดยง่าย (Singkul et al., 2019, p. 244) โดยเฉพาะอย่างยิ่ง ในกรณีของทฤษฎีโครงสร้างวาทะที่พัฒนารอบการวิเคราะห์มาจากภาษาเขียนที่มีลักษณะเป็นทางการด้วยแล้ว การประยุกต์ใช้ทฤษฎีดังกล่าวในการประมวลผลภาษาในสื่อสังคมออนไลน์ย่อมก่อให้เกิดปัญหาบางประการตามมา

การตัดหน่วยปริเฉทพื้นฐานภาษาไทยที่ปรากฏในสื่อสังคมออนไลน์เป็นปัญหาพื้นฐานประการหนึ่งที่ต้องประสบ กล่าวคือ ระบบมักตัดหน่วยปริเฉทพื้นฐานผิดพลาด เมื่อต้องประมวลผลภาษาไทยในสื่อสังคมออนไลน์ ทั้งนี้ จากการสังเกต ผู้เขียนพบว่าภาษาไทยที่ปรากฏในสื่อสังคมออนไลน์มีลักษณะบางประการที่ส่งผลกระทบต่อการตัดหน่วยปริเฉทพื้นฐาน

- (4) [เคยเจอแบบ]_{EDU1}[ตอนเช้าต้องไปส่ง]_{EDU2}
- (5) [แบบว่าตอนเช้าต้องไปส่ง]_{EDU1}
- (6) [คือแบบจะไปทำงานสายแล้วไง]_{EDU1}
- (7) [นางเดินไปแบบเงิบๆ]_{EDU1}
- (8) [นางเดินไป]_{EDU1}[แบบไม่หันกลับมามอง]_{EDU2}
- (9) [นางเป็นคน]_{EDU1}[แบบไม่สนโลก]_{EDU2}

ตัวอย่าง (4)-(9) แสดงการตัดขอบเขตของหน่วยปริจเฉทพื้นฐานที่แตกต่างกันในข้อความที่มีคำว่า “แบบ” ปรากฏอยู่ จากตัวอย่างจะเห็นได้ว่าหน้าที่ของคำว่า “แบบ” ในข้อความข้างต้นแตกต่างกัน ส่งผลต่อความเป็นอนุพากย์ในข้อความ กล่าวคือ ในตัวอย่าง (4) สามารถวิเคราะห์ให้ “แบบ” เป็นหน่วยนาม ข้อความดังกล่าวจึงประกอบด้วยอนุพากย์ 2 อนุพากย์ หรือตัดเป็นหน่วยปริจเฉทพื้นฐานได้ว่า 2 หน่วย ในขณะที่ตัวอย่าง (5) อาจวิเคราะห์ให้ “แบบว่า” เป็นตัวบ่งชี้ปริจเฉท ข้อความดังกล่าวจึงถือเป็นอนุพากย์เดียว เช่นเดียวกับกับตัวอย่าง (6) ที่วิเคราะห์ให้ “คือแบบ” เป็นตัวบ่งชี้ปริจเฉทได้เช่นกัน ส่วนตัวอย่าง (7) “แบบ” ทำหน้าที่เป็นตัวบ่งชี้คำคุณศัพท์ “เงิบๆ” เช่นเดียวกับคำว่า “อย่าง” ในประโยค “เขากินอย่างเร็ว” ในกรณีนี้จึงวิเคราะห์ให้ข้อความดังกล่าวมีอนุพากย์เดียว ข้อความในตัวอย่าง (8) สามารถตัดหน่วยปริจเฉทพื้นฐานได้ 2 หน่วยตามจำนวนของอนุพากย์ โดย “แบบ” ในตัวอย่างนี้ทำหน้าที่เป็นคำเชื่อมอนุพากย์วิเศษณ์ และข้อความในตัวอย่าง (9) “แบบ” ทำหน้าที่เป็นคำเชื่อมอนุพากย์สัมพัทธ์ เช่นเดียวกับคำว่า “ที่” ในประโยค “ลุงที่กวาดพื้นเป็นภารโรง” จึงวิเคราะห์ได้ว่าข้อความดังกล่าวประกอบด้วยอนุพากย์ 2 อนุพากย์ จากตัวอย่าง จะเห็นได้ว่าการตัดหน่วยปริจเฉทพื้นฐานจำเป็นต้องอาศัยความเข้าใจจากการพิจารณาชนิดของคำว่า “แบบ” ที่ปรากฏในข้อความให้ถูกต้อง อย่างไรก็ตาม ยังคงมีข้อถกเถียงว่าคำว่า “แบบ” ที่ปรากฏในตัวอย่างที่ (4) และ (6)-(9) นี้เป็นคำชนิดใด ทั้งนี้ อาจวิเคราะห์ให้เป็นคำอนุภาค (particle) หรือวิเคราะห์ให้เป็นตัวบ่งชี้ปริจเฉทก็ย่อมได้ จากกรณีตัวอย่างนี้จึงเป็นที่น่าสังเกตว่าการศึกษาวเคราะห์หน้าที่ของหน่วยพื้นฐานในภาษาไทยยังคงมีความจำเป็นอยู่ ด้วยจะทำให้ทราบชนิดของคำในภาษาไทยอย่างชัดเจน และส่งผลให้พิจารณาตัดสินใจตัดหน่วยปริจเฉทพื้นฐานได้ถูกต้องแม่นยำในที่สุด

- (10) [แบงค์ชาติคาด]_{EDU1}[ว่า]ในไตรมาส 1/2565 อัตราเงินเฟ้อทั่วไป
ของไทยจะมีแนวโน้มต่ำลง]_{EDU2}
- (11) [ผม(คิด)]_{EDU1}[ว่าเขาไม่ไปหรอก]_{EDU2}

ตัวอย่าง (10) และ (11) แสดงการละหน่วยในข้อความซึ่งเป็นปัญหาอีกประการหนึ่งของการตัดหน่วยปริจเฉทพื้นฐานภาษาไทยที่ปรากฏในสื่อออนไลน์ ด้วยระบบตัดหน่วยปริจเฉทพื้นฐานที่พัฒนาขึ้นใน

ปัจจุบันเรียนรู้ขอบเขตของหน่วยปริจเฉทพื้นฐานจากการพิจารณาการปรากฏของตัวบ่งชี้ในทางวากยสัมพันธ์เป็นหลัก การละหน่วยในข้อความเช่นที่ปรากฏในตัวอย่าง (10) ที่ละคำเชื่อมอนุพากย์เติมเต็ม “ว่า” หรือตัวอย่าง (11) ที่ละกริยาหลักของอนุพากย์ “คิด” จึงส่งผลกระทบต่อความถูกต้องแม่นยำของระบบตัดหน่วยปริจเฉทพื้นฐานได้

ตัวอย่างลักษณะภาษาที่ปรากฏใช้ในสื่อสังคมออนไลน์ที่ยกมาข้างต้นนี้เป็นเพียงส่วนหนึ่งของลักษณะที่พบเท่านั้น ในการประมวลผลจริงยังอาจมีลักษณะภาษาอีกหลายประการที่ส่งผลต่อประสิทธิภาพการทำงานของเครื่องมือประมวลผลภาษา ในแง่นี้ การศึกษาวิเคราะห์ภาษาไทยในสื่อสังคมออนไลน์ตามกรอบแนวทฤษฎีโครงสร้างวาทะจึงจำเป็นอย่างยิ่งต่อวงการประมวลผลภาษาไทยในอนาคต

นอกจากการศึกษาภาษาไทยในสื่อออนไลน์แล้ว การจัดการกับอนุพากย์ฝังก็เป็นอีกเรื่องที่ต้องทำความเข้าใจเพื่อประสิทธิภาพในการตัดหน่วยปริจเฉทพื้นฐานภาษาไทย แม้จะปรากฏความพยายามในการแก้ปัญหาในงานของชานาธิป แซ่เตีย และคณะ (Saetia et al., 2020) แล้ว แต่ในด้านประสิทธิภาพการทำงานและเทคนิคทางคอมพิวเตอร์ก็ยังคงต้องการการพัฒนาให้ดียิ่งขึ้นต่อไป

ทั้งนี้ ในระยะยาว หากสามารถพัฒนาระบบแฉงส่วนปริจเฉทภาษาไทยตามทฤษฎีโครงสร้างวาทะได้อย่างมีประสิทธิภาพ ก็จะเป็นก้าวสำคัญของการประมวลผลภาษาไทย ด้วยผลลัพธ์ที่ได้จากการแฉงส่วนปริจเฉทนั้นสามารถนำไปต่อยอดได้ในงานประมวลผลภาษาธรรมชาติแขนงอื่นๆ ไม่ว่าจะเป็นการวิเคราะห์ความรู้สึก การสกัดสนเทศ การสรุปย่อข้อความ หรือการแปลภาษาด้วยเครื่อง

ในแง่การวิเคราะห์ทางภาษาศาสตร์นั้น จากการทบทวนงานที่เกี่ยวข้องยังไม่พบการทดลองวิเคราะห์ประเภทของวาทสัมพันธ์โดยประยุกต์กับภาษาไทย ทั้งนี้ การศึกษาวิจัยในหัวข้อนี้อาจนำไปสู่การเสนอชุดของวาทสัมพันธ์ชุดใหม่ที่สอดคล้องกับลักษณะการเชื่อมโยงในปริจเฉทภาษาไทย ในอีกแง่หนึ่งก็ถือเป็นการพัฒนาตัวทฤษฎีโครงสร้างวาทะเองซึ่งอาจนำไปสู่ข้อค้นพบใหม่ที่น่าสนใจ นอกจากนี้ ผลจากการศึกษาในหัวนี้ยังนำไปพัฒนาเป็นแนวทางการกำกับข้อมูลปริจเฉทภาษาไทยตามแนวทฤษฎีโครงสร้างวาทะได้

นอกจากนี้ ดังได้กล่าวไปในตอนต้นว่า ทฤษฎีโครงสร้างวาทะมีจุดเด่นประการหนึ่งคือการให้ความสำคัญคุณสมบัตินทางวากยสัมพันธ์ของข้อความภายในโครงสร้างปริจเฉท จุดเด่นประการนี้เอื้อต่อการศึกษาคำต่อประสานระหว่างวากยสัมพันธ์กับปริจเฉท โดยเฉพาะอย่างยิ่ง ในปัจจุบันที่มีความสนใจการวิเคราะห์ข้อความตามแนววากยสัมพันธ์แบบพึ่งพาในภาษาไทย แนวทางการวิจัยด้านนี้ก็อาจปรับประยุกต์แนวคิดเรื่องการแฉงส่วนทางวากยสัมพันธ์แบบพึ่งพาร่วมกับการแฉงส่วนปริจเฉทแบบพึ่งพาตามที่มีมอเรียและคณะ (Morey et al., 2018) เสนอได้ ผลที่ได้จากการศึกษาวิจัยหัวข้อนี้ นอกจากจะขยายมุมมองด้านปฏิสัมพันธ์ระหว่างระดับวากยสัมพันธ์และระดับปริจเฉทแล้ว ยังอาจนำไปสู่การพัฒนาเครื่องมือแฉงส่วนทางวากยสัมพันธ์และปริจเฉทแบบเบ็ดเสร็จในระบบเดียวได้อีกด้วย

ส่วนในแง่การเรียนการสอนการเขียนภาษาไทย อาจกล่าวได้ว่าทฤษฎีโครงสร้างวาทะมีลักษณะหลายประการที่เอื้อต่อการนำไปประยุกต์ใช้ในการเรียนการสอนการเขียน กล่าวคือ แนวคิดในทฤษฎีนี้มุ่งให้ความสำคัญกับการจัดองค์ประกอบภายในข้อความ โดยนำเสนอความคิดแต่ละความคิดออกมาในรูปแบบของหน่วยปริจเฉทพื้นฐานที่มีความสำคัญลดหลั่นกันไป อีกทั้งยังเสนอวิธีการพิจารณาการเกาะเกี่ยวความในข้อความผ่านชุดของวาทสัมพันธ์ ลักษณะดังกล่าวนี้สอดคล้องกับแนวการสอนการเขียนภาษาไทยที่มุ่งให้ผู้เรียนจัดลำดับและนำเสนอความคิดออกเป็นรูปภาษาที่มีความสอดคล้องสัมพันธ์กัน ดังนั้น หากสามารถประยุกต์ใช้แนวคิดของทฤษฎีโครงสร้างวาทะในการเรียนการสอนการเขียนภาษาไทยได้ ก็เชื่อได้ว่าการเรียนการสอนการเขียนภาษาไทยจะเป็นอย่างมีประสิทธิภาพมากยิ่งขึ้น ทั้งนี้ ในแง่การวิจัยอาจทำได้ตั้งแต่การพิจารณากลไกทางภาษาที่ใช้แสดงวาทสัมพันธ์ประเภทต่างๆ เพื่อนำมาใช้ในการสอนการเชื่อมโยงความคิดในการเขียน หรืออาจศึกษาเกี่ยวกับการจัดเรียงข้อความในการเขียนตามสถานะความเป็นแกนกลางเพื่อใช้สอนเรื่องการจัดลำดับความคิดในการเขียน

จากที่กล่าวมาในหัวข้อนี้ จะเห็นได้ว่ายังปรากฏช่องว่างทางการวิจัยว่าด้วยทฤษฎีโครงสร้างวาทะกับภาษาไทยที่สามารถเติมเต็มได้อีกมาก ทั้งในด้านการประมวลผลภาษาไทย การวิเคราะห์ภาษาไทยในเชิงทฤษฎีภาษาศาสตร์ และการเรียนการสอนทักษะการเขียนภาษาไทย ทั้งนี้ ผู้เขียนเชื่อว่าการนำทฤษฎีโครงสร้างวาทะมาใช้เป็นกรอบในการศึกษาทำความเข้าใจปรากฏการณ์ในภาษาไทยไม่เพียงแต่จะช่วยขยายขอบเขตขององค์ความรู้และเปิดมุมมองใหม่ให้แก่วงวิชาการภาษาไทยเท่านั้น แต่ผลที่ได้จากการศึกษาตามกรอบทฤษฎีนี้ยังสามารถนำไปพัฒนาต่อยอดให้ได้ประโยชน์อย่างเป็นรูปธรรมอีกด้วย

สรุป

ทฤษฎีโครงสร้างวาทะเป็นทฤษฎีหนึ่งในกลุ่มทฤษฎีโครงสร้างปริจเฉทที่ได้รับการยอมรับอย่างแพร่หลายทั้งในเชิงทฤษฎีและการนำไปประยุกต์ใช้งานในหลายแขนง แนวคิดหลักของทฤษฎีนี้คือการพิจารณาคความสัมพันธ์ระหว่างหน่วยย่อยภายในโครงสร้างปริจเฉท โดยยึดความสำคัญของสารที่สื่อผ่านรูปภาษาในหน่วยย่อย รวมถึงความสัมพันธ์ของแต่ละหน่วยย่อยมีต่อกัน ทำให้ได้โครงสร้างปริจเฉทที่มีความสัมพันธ์กันภายในเป็นแบบลำดับชั้น และด้วยการยอมรับอย่างแพร่หลาย ตลอดระยะเวลา นับตั้งแต่มีการนำเสนอ ทฤษฎีโครงสร้างวาทะได้ถูกตีความและพัฒนาแนวคิดเชิงทฤษฎีเรื่อยมาจนกระทั่งปัจจุบัน

เมื่อพิจารณาในมุมมองเชิงทฤษฎีแล้ว ทฤษฎีโครงสร้างวาทะมีจุดเด่นที่เชื่อมโยงคุณสมบัติของภาษาไว้หลายแง่มุม ทั้งในแง่วาทสัมพันธ์ แง่ความหมาย และแง่หน้าที่ จุดเด่นเหล่านี้เอื้อต่อการวิเคราะห์และทำความเข้าใจภาษาในระดับปริจเฉทได้เป็นอย่างดี จนเป็นผลให้เกิดการรับเอาแนวคิดเชิงทฤษฎีของทฤษฎีโครงสร้างวาทะไปพัฒนาต่อยอดในหลายด้านและหลายภาษา ไม่เฉพาะด้านการประมวลผลภาษาธรรมชาติเท่านั้น แต่ยังรวมถึงด้านปริจเฉทวิเคราะห์และด้านการเรียนการสอนภาษาด้วย

ในส่วนที่เกี่ยวข้องกับภาษาไทย ทฤษฎีโครงสร้างวาทะมีบทบาทสำคัญยิ่งในฐานะทฤษฎีที่นำเสนอแนวคิดเรื่องหน่วยปริจเฉทพื้นฐาน ซึ่งสามารถนำมาประยุกต์ใช้เป็นหน่วยรับเข้าในการประมวลผลภาษาไทยแทนประโยคซึ่งเป็นหน่วยทางภาษาที่ก่อให้เกิดข้อถกเถียงในวงการประมวลผลภาษาไทยเรื่อยมา ผลจากการรับเอาแนวคิดเรื่องการใช้หน่วยปริจเฉทพื้นฐานเป็นหน่วยรับเข้าในการประมวลผลภาษาไทยนี้ได้ผลักดันให้เกิดงานวิจัยที่มุ่งทำความเข้าใจและพัฒนาวิธีการตัดหน่วยปริจเฉทภาษาไทยตามมาเป็นจำนวนมาก ยิ่งไปกว่านั้น ยังปรากฏงานวิจัยหลายชิ้นที่นำแนวทฤษฎีโครงสร้างวาทะไปใช้ต่อยอดเพื่อพัฒนาระบบประมวลผลภาษาไทยในแขนงงานต่างๆ โดยเฉพาะอย่างยิ่งในแขนงงานสรุปย่อข้อความ

อย่างไรก็ดี เมื่อพิจารณาเปรียบเทียบความก้าวหน้าของการประยุกต์ใช้ทฤษฎีโครงสร้างวาทะในการวิเคราะห์และประมวลผลภาษาไทยกับภาษาอื่นๆ แล้ว จะพบว่ามีความประเด็นปัญหาที่เป็นช่องว่างทางการวิจัยซึ่งสามารถเติมเต็มได้อีกมาก ทั้งในแง่การวิเคราะห์ภาษาไทยตามแนวทฤษฎีโครงสร้างวาทะเพื่อรองรับการประมวลผลภาษาไทยในอนาคต หรือการตีความทฤษฎีโครงสร้างวาทะในฐานะทฤษฎีภาษาศาสตร์เพื่อทำความเข้าใจความสัมพันธ์ในระดับวากยสัมพันธ์และปริจเฉทในภาษาไทย หรือแม้กระทั่งการวิจัยที่ประยุกต์เอาแนวทฤษฎีโครงสร้างวาทะไปใช้ในการเรียนการสอนทักษะการใช้ภาษาไทย ทั้งนี้ การศึกษาภาษาไทยตามแนวทฤษฎีโครงสร้างวาทะไม่เพียงแต่จะเป็นการขยายขอบเขตความรู้และเปิดมุมมองใหม่ให้แก่วงวิชาการภาษาไทยเท่านั้น แต่ยังสามารถนำผลที่ได้จากการศึกษาไปพัฒนาต่อยอดให้เกิดประโยชน์อย่างเป็นรูปธรรมได้อีกด้วย

เอกสารอ้างอิง

- นลินี อินตะชาว. (2556). *การแยกอนุพากย์ภาษาไทยด้วยการใช้แบบจำลองซัพพอร์ตเวกเตอร์แมชชีน*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ), จุฬาลงกรณ์มหาวิทยาลัย, กทม.
- วิโรจน์ อรุณมานะกุล. (2562). *การตัดประโยคภาษาไทย: (1) ประโยคคืออะไร*. ค้นจาก <https://awirote.medium.com/การตัดประโยคคืออะไร-1-ประโยคคืออะไร-f009994fcc40>
- ศุภวัจนน์ แต่รุ่งเรือง. (2560). *การตรวจเทียบภายนอกหากลักลอกในงานวิชาการโดยใช้แบบจำลองซัพพอร์ตเวกเตอร์แมชชีนและการวัดค่าความละม้ายของข้อความ*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ), จุฬาลงกรณ์มหาวิทยาลัย, กทม.
- ศุภวัจนน์ แต่รุ่งเรือง และวิโรจน์ อรุณมานะกุล. (2558). กลวิธีลักลอกงานวิชาการภาษาไทย: การวิเคราะห์ทางภาษาศาสตร์. *ภาษาและภาษาศาสตร์*, 34(1), 38-65.
- Ahmadi, A., & Parhizgar, S. (2017). Coherence Errors in Iranian EFL Learners' Writing: A Rhetorical Structure Theory Approach. *Journal of Language Horizons*, 1(1), 9-37. doi: 10.22051/lghor.2017.8588.1011

- Bhatia, P., Ji, Y., & Eisenstein, J. (2015). Better Document-level Sentiment Analysis from RST Discourse Parsing. In L. Màrquez, C. Callison-Burch, & J. Su (Eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2212-2218). Lisbon, Portugal: Association for Computational Linguistics.
- Braud, C., Coavoux, M., & Søgaard, A. (2017). Cross-lingual RST Discourse Parsing. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 292-304). Valencia, Spain: Association for Computational Linguistics.
- Brown, J. D., & Verhagen, A. (2018). Using Rhetorical Structure Theory for Contrastive Purposes: A Pilot Study. *OnCUE Journal*, 11(1), 3-24.
- Carlson, L., & Marcu, D. (2001). *Discourse tagging reference manual*. CA: University of Southern California Information Sciences Institute.
- Carlson, L., Marcu, D., & Okurowski, M. E. (2001). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In *SIGDIAL '01: Proceedings of the Second SIGdial Workshop on Discourse and Dialogue* (Vol. 16, pp. 1-10). Aalborg, Denmark: Association for Computational Linguistics.
- Charoensuk, J., Suvakree, T., & Kawtrakul, A. (2005). Thai Elementary Discourse Unit Segmentation by Discourse Segmentation Cues and Syntactic Information. In *Proceedings of SNLP 2005: the 6th International Symposium on Natural Language Processing* (pp. 85-90). Chiang Rai, Thailand: Chulalongkorn University.
- Das, D., & Taboada, M. (2018). RST Signalling Corpus: a corpus of signals of coherence relations. *Language Resources and Evaluation*, 52(1), 149-184. doi: 10.1007/s10579-017-9383-x
- Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., & Nagata, M. (2013). Single-Document Summarization as a Tree Knapsack Problem. In D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, & S. Bethard (Eds.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1515-1520). Seattle, Washington, USA: Association for Computational Linguistics.
- Hou, S., Zhang, S., & Fei, C. (2020). Rhetorical structure theory: A comprehensive review of theory, parsing methods and applications. *Expert Systems with Applications*, 157, 113421. doi: 10.1016/j.eswa.2020.113421

- Intasaw, N., & Aroonmanakun, W. (2013). Basic principles for segmenting Thai EDUs. In H.-I. Lai & K. Chui (Eds.), *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)* (pp. 491-498). Taipei, Taiwan: Department of English, National Chengchi University.
- Ketui, N., Theeramunkong, T., & Onsuwan, C. (2012). A rule-based method for Thai Elementary Discourse Unit Segmentation (TED-Seg). In V. C. S. Lee & K.-L. Ong (Eds.), *KICSS '12: Proceedings of the 2012 Seventh International Conference on Knowledge, Information and Creativity Support Systems* (pp. 195-202). Melbourne, Australia: IEEE Computer Society.
- Ketui, N., Theeramunkong, T., & Onsuwan, C. (2015). An EDU-Based Approach for Thai Multi-Document Summarization and Its Application. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 14(1), Article 4. doi: 10.1145/2641567
- Kongwan, A., Kamaruddin, S. S. B., & Ahmad, F. B. K. (2020). Thai EDU Segmentation Using Clue Markers and Syntactic Information from Shallow Parser. *Journal of Theoretical and Applied Information Technology*, 98(18), 3853-3869.
- Kraus, M., & Feuerriegel, S. (2019). Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. *Expert Systems with Applications*, 118, 65-79. doi: 10.1016/j.eswa.2018.10.002
- Lertpiya, A., Chaiwachirasak, T., Maharattanamalai, N., Lapjaturapit, T., Chalothorn, T., Tirasaroj, N., & Chuangsuwanich, E. (2018). A Preliminary Study on Fundamental Thai NLP Tasks for User-generated Web Content. In K. Hashimoto, T. Supnithi, N. Hnoohom, M. Ketcham, P. Songmuang, R. Kongkachandra, K. Wongpatikaseree, N. Chumuang, W. Yimyam, & A. Jitpattanakul (Eds.), *Proceedings of the 13th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2018)* (pp. 18-25). Pattaya, Thailand: Artificial Intelligence Association of Thailand (AIAT).
- Mann, W. C., & Thompson, S. A. (1987). *Rhetorical Structure Theory: A Theory of Text Organization*. CA: Information Science Institute, University of Southern California.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3), 243-281.
- Marcu, D. (2000a). The rhetorical parsing of unrestricted texts: a surface-based approach. *Computational linguistics*, 26(3), 395-448.

- Marcu, D. (2000b). *The Theory and Practice of Discourse Parsing and Summarization*. MA: MIT Press.
- Morey, M., Muller, P., & Asher, N. (2018). A Dependency Perspective on RST Discourse Parsing and Evaluation. *Computational linguistics*, 44(2), 197-235. doi: 10.1162/coli_a_00314
- O'Donnell, M., Mellish, C., Oberlander, J., & Knott, A. (2001). ILEX: an architecture for a dynamic hypertext generation system. *Natural Language Engineering*, 7(3), 225-250. doi: 10.1017/S1351324901002698
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., & Webber, B. (2008). The Penn Discourse TreeBank 2.0. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & D. Tapias (Eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp. 2961-2968). Marrakech, Morocco: European Language Resources Association (ELRA).
- Radev, D. R. (2000). A common theory of information fusion from multiple text sources step one: cross-document structure. In L. Dybkjær, K. Hasida, & D. Traum (Eds.), *Proceedings of the 1st SIGdial workshop on Discourse and dialogue - Volume 10* (pp. 74-83). Hong Kong: Association for Computational Linguistics.
- Rahangdale, A., & Agrawal, A. J. (2014). Information Extraction Using Discourse Analysis from Newswires. *International Journal of Information Technology Convergence and Services*, 4(3), 21-30. doi: 10.5121/ijitcs.2014.4302
- Saetia, C., Taerungruang, S., & Chalothorn, T. (2020). Combining Thai EDUs: Principle and Implementation. In L. M. Nguyen, C. M. Luong, & S. Song (Eds.), *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation* (pp. 270-278). Hanoi, Vietnam: Association for Computational Linguistics.
- Singkul, S., Khampingyot, B., Maharattamalai, N., Taerungruang, S., & Chalothorn, T. (2019). Parsing Thai Social Data: A New Challenge for Thai NLP. In H. Kiyota, T. Supnithi, M. Ketcham, N. Chumuang, P. Songmuang, Supakrit Sukjarern, R. Kongkachandra, J. Donjuntai, S. Lipirodjanapong, A. Rongtong, & J. Kaewjunda (Eds.), *Proceeding of the 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP 2019)* (pp. 244-250). Chiang Mai, Thailand: Artificial Intelligence Association of Thailand (AIAT).

- Sinthupoun, S., & Sornil, O. (2010). Thai rhetorical structure analysis. *International Journal of Computer Science and Information Security (IJCSIS)*, 7(1), 95-105.
- Skoufaki, S. (2020). Rhetorical Structure Theory and coherence break identification. *Text & Talk*, 40(1), 99-124. doi: 10.1515/text-2019-2050
- Sukvaree, T., Charoensuk, J., Wattanamethanont, M., & Kultrakul, A. (2004). *RST based Text Summarization with Ontology Driven in Agriculture Domain*. Bangkok: Department of Computer Engineering, Kasetsart University.
- Sukvaree, T., Kawtrakul, A., & Caelen, J. (2007). Thai Text Coherence Structuring with Coordinating and Subordinating Relations for Text Summarization. In B. Kokinov, D. C. Richardson, T. R. Roth-Berghofer, & L. Vieu (Eds.), *Modeling and Using Context: Proceeding of the 6th International and Interdisciplinary Conference (CONTEXT 2007)* (pp. 453-466). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Taboada, M., & Mann, W. C. (2006). Applications of Rhetorical Structure Theory. *Discourse Studies*, 8(4), 567-588.
- Tapsai, C., Unger, H., & Meesad, P. (2021). *Thai Natural Language Processing: Word Segmentation, Semantic Analysis, and Application*. Cham, Switzerland: Springer.
- Voll, K., & Taboada, M. (2007). Not All Words Are Created Equal: Extracting Semantic Orientation as a Function of Adjective Relevance. In M. A. Orgun & J. Thornton (Eds.), *AI 2007: Advances in Artificial Intelligence* (pp. 337-346). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Wahlster, W., Andre, E., Graf, W., & Rist, T. (1991). Designing Illustrated Texts: How Language Production Is Influenced by Graphics Generation. In J. Kunze & D. Reimann (Eds.), *Fifth Conference of the European Chapter of the Association for Computational Linguistics* (pp. 8-14). Berlin: Association for Computational Linguistics.
- Wattanamethanont, M., Suvakree, T., & Kawtrakul, A. (2005). Thai Discourse Relations Recognition By Using naive Bayes Classifier. In *Proceedings of SNLP 2005: the 6th International Symposium on Natural Language Processing* (pp. 291-296). Chiang Rai: Chulalongkorn University.