

วงศ์ศัพท์ในแบบเรียนภาษาไทย: การศึกษาตามแนวทางภาษาศาสตร์คลังข้อมูล*

ธารทอง แจ่มไพบูลย์** และ วิโรจน์ อรุณมานะกุล***

บทคัดย่อ

บทความชิ้นนี้มุ่งศึกษาคำศัพท์ที่ปรากฏในแบบเรียนภาษาไทยในระดับเริ่มเรียน-มัธยมศึกษาปีที่ 6 โดยแบ่งออกเป็น 4 ช่วงชั้น เพื่อศึกษาว่าคำศัพท์ที่ปรากฏในช่วงชั้นที่สูงกว่าจะมีมากกว่าและครอบคลุมคำศัพท์จากช่วงชั้นที่ต่ำกว่า รวมถึงยากกว่าในแง่จำนวนพยางค์ ความถี่ในทางความหมาย และมีความหลากหลายทางความหมายมากกว่าหรือไม่ อย่างไร โดยมีสมมติฐานว่า 1) คำศัพท์ในช่วงชั้นที่สูงกว่ามีมากกว่าและครอบคลุมคำศัพท์ในช่วงชั้นที่ต่ำกว่า 2) คำศัพท์ที่พบในช่วงชั้นที่สูงกว่ามีค่า 3 พยางค์และคำที่มากกว่าช่วงชั้นที่ต่ำกว่า 3) คำกริยาจากช่วงชั้นที่สูงกว่ามีความหมายและการใช้มากกว่าช่วงชั้นที่ต่ำกว่า จากรายการคำศัพท์ของช่วงชั้นต่าง ๆ ที่ได้จากคลังข้อมูลแบบเรียนภาษาไทยขนาด 3 ล้านคำแสดงให้เห็นว่าคำศัพท์ในช่วงชั้นที่สูงกว่าไม่ได้ครอบคลุมคำศัพท์ทั้งหมดจากช่วงชั้นที่ต่ำกว่า ส่วนเรื่องความยากง่าย พบว่า คำ 3 พยางค์รวมถึงคำที่มีลักษณะทางความหมายเป็นคำที่มีแนวโน้มเพิ่มสูงขึ้นตามช่วงชั้นที่สูงขึ้น จึงสามารถใช้เป็นปัจจัยบ่งความยากง่ายได้ ในแง่ของความหลากหลายทางความหมายของคำที่มีความถี่สูงซึ่งได้เลือกศึกษาคำกริยา *ไป-มา* ไม่พบว่ามี การเพิ่มขึ้นของความยากในมิตินี้ เนื่องจากความหมายทั้งหมดของทั้งสองคำถูกพบใช้ตั้งแต่ในระดับชั้นเริ่มเรียนแล้ว และยังพบอีกว่าช่วงชั้นประถมปลายเป็นช่วงชั้นที่มีความหลากหลายของคำศัพท์มากที่สุดจึงน่าจะมีความสำคัญต่อการขยายตัวของวงศ์ศัพท์มากกว่าช่วงชั้นอื่น ๆ

คำสำคัญ: ภาษาศาสตร์คลังข้อมูล คำศัพท์ แบบเรียนภาษาไทย

* บทความนี้เป็นส่วนหนึ่งของวิทยานิพนธ์ระดับดุษฎีบัณฑิตเรื่อง “พัฒนาการของคำศัพท์ที่มีความถี่ในการปรากฏสูงในหนังสือเรียนวิชาภาษาไทย: การศึกษาตามแนวทางภาษาศาสตร์คลังข้อมูล” ของผู้วิจัยชื่อแรก ได้รับทุนสนับสนุนจากสำนักงานกองทุนสนับสนุนการวิจัย ในโครงการปริญญาเอกกาญจนาภิเษก (คปก.)

** ดุษฎีบัณฑิต ภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท แขวงวังใหม่ เขตปทุมวัน กรุงเทพมหานคร 10330
อีเมล: tantong.c@gmail.com

*** รองศาสตราจารย์ ภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ถนนพญาไท แขวงวังใหม่ เขตปทุมวัน กรุงเทพมหานคร 10330
อีเมล: awirote@gmail.com



Vocabulary in Thai Textbooks: A Corpus-based Study*

Tantong Champaiboon** and Wirote Aroonmanakun***

Abstract

This study analyzed vocabulary in Thai textbooks used for teaching children ranging from preschool to the twelfth grade, which was categorized into four levels of education. The hypotheses of this study were 1) the vocabulary of textbooks in the higher levels subsumed vocabulary in the lower levels; 2) the number of three-syllable words and opaque words in higher levels were higher than those in the lower levels; and 3) the same verb found in the higher levels had more meanings and usages than those in the lower levels. According to the list of vocabulary taken from three-million-word Thai Textbook Corpus, it was found that the vocabulary in the higher levels did not include every word in the lower levels. In terms of word difficulty, the findings revealed that the number of three-syllable words, and the opaque words tended to increase in the higher levels. As a result, these factors could be used to indicate the difficulty of vocabulary. However, in the case of high frequency words with multiple meanings

* This article is a part of the first author's dissertation title "Development of high frequency words in Thai language textbooks: a corpus linguistics study". Financial support from the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0203/2551) is acknowledged.

** Ph.D. graduate, Department of Linguistics, Faculty of Arts, Chulalongkorn University
Phayathai Road, Wangmai, Pathumwan, Bangkok 10330, THAILAND.
E-mail: tantong.c@gmail.com

*** Associate Professor, Department of Linguistics, Faculty of Arts, Chulalongkorn University
Phayathai Road, Wangmai, Pathumwan, Bangkok 10330, THAILAND.
E-mail: awirote@gmail.com

like the verbs ‘go’ (/paj/) and ‘come’ (/maa/), it was found that there was no increment of meanings according to the levels of education since all of the meanings of both words were already used at the first level. Besides, the findings indicated that the vocabulary in level 2 (4th-6th grade) had the widest range of vocabulary, which contributed to the increasing number of vocabulary of learners more than other levels did.

Keywords: Corpus-Based Study, Vocabulary, Thai Textbook

บทนำ (Introduction)

ปัญหาเกี่ยวกับการเรียนการสอนภาษาไทยโดยเฉพาะอย่างยิ่งประเด็นที่ว่าเด็กไทยที่ศึกษามาระยะหนึ่งแล้วยังอ่านไม่ออกเขียนไม่ได้ เป็นประเด็นสำคัญประเด็นหนึ่งที่ยังมีการถกเถียงกันอยู่ในปัจจุบัน มีหน่วยงาน นักวิชาการ ผู้เชี่ยวชาญด้านภาษาหลายฝ่ายต่างหยิบยกประเด็นนี้ขึ้นมาพูดคุยและเห็นว่าปัญหาดังกล่าวอาจเกิดขึ้นจากหลายสาเหตุ ตั้งแต่หลักสูตร นโยบาย ผู้บริหาร ครูผู้สอน ผู้ปกครอง นักเรียน รวมไปถึงแบบเรียนหรือสื่อการเรียนการสอนภาษาไทย เฉลิมลาภ ทองอาจ (2557) ตั้งข้อสังเกตว่าจะเป็นไปได้หรือไม่ที่สื่อการเรียนการสอนภาษาไทย “มีลักษณะยึดยึดความรู้ปฏิกิริยาที่ไม่จำเป็น ไม่เหมาะสมกับพัฒนาการของเด็ก” ซึ่งเราไม่อาจปฏิเสธได้ว่าสื่อการเรียนการสอนเหล่านี้มีบทบาทอย่างมากต่อการเรียนรู้ของนักเรียน เนื่องจากเป็นสิ่งที่บรรจุ “ความรู้” เกี่ยวกับภาษาไทยในแง่มุมต่าง ๆ ทั้งคำศัพท์ หลักไวยากรณ์ วิธีการประสมคำ ฯลฯ โดยหลักการแล้วแง่มุมต่าง ๆ เหล่านี้ควรพัฒนาระดับความยากขึ้นตามช่วงชั้นที่สูงขึ้นด้วย

การเรียนรู้คำศัพท์เป็นเรื่องพื้นฐานในการเรียนรู้ภาษา แนวคิดเรื่องการคัดเลือกคำศัพท์ให้เหมาะสมกับการเรียนการสอนภาษาในระดับต่าง ๆ เป็นที่ยอมรับและใช้กันมานาน เนื่องจากการรู้คำศัพท์เป็นปัจจัยหนึ่งที่มีความสำคัญต่อการอ่านเพื่อความเข้าใจ (Loraine, 2008; กระทรวงศึกษาธิการ, 2532) การคัดเลือกคำศัพท์ที่เหมาะสมจะมีส่วนช่วยในการเรียนรู้อย่างมาก หากคำศัพท์เหล่านั้นสอดคล้องกับพัฒนาการการเรียนรู้ของนักเรียนแล้ว นักเรียนจะสามารถทำความเข้าใจตามได้ง่ายและเกิดพัฒนาการที่มีลำดับต่อเนื่องจากง่ายไปยาก จนนำไปสู่สัมฤทธิ์ผลในการเรียนรู้ภาษา

หลักเกณฑ์สำคัญข้อหนึ่งที่ใช้ประกอบการคัดเลือกคำศัพท์ที่เหมาะสมต่อการเรียนการสอนในระดับต่าง ๆ คือ การพิจารณาจากความถี่ของคำที่ปรากฏใช้จริง โดยรายการคำศัพท์หลายรายการคัดเลือกจากคำศัพท์ที่มีความถี่ในการปรากฏสูงจากแหล่งข้อมูลต่าง ๆ (Carroll, Davies, & Richman, 1971; Coxhead, 2000; West, 1953; Xue & Nation, 1984) โดยอาศัยพื้นฐานที่ว่าคำที่ปรากฏบ่อย ๆ เป็นองค์ประกอบสำคัญและจำเป็นในการสื่อสาร ผู้เรียนภาษาควรศึกษาคำเหล่านั้นก่อนคำอื่น ๆ ที่ปรากฏด้วยความถี่ต่ำ

จากแนวคิดดังกล่าวนำไปสู่การวิจัยระดับชาติในประเทศไทย โดยระหว่างปี พ.ศ. 2529-2531 กระทรวงศึกษาธิการ (2532) ได้ดำเนินการศึกษาคำพื้นฐานที่ใช้ในการเรียนการสอนภาษาไทย ระดับประถมศึกษา จากการสัมภาษณ์นักเรียนกลุ่มตัวอย่าง คัดเลือกคำศัพท์ที่ปรากฏด้วยความถี่สูงมาวิเคราะห์ได้บัญชีคำพื้นฐานทั้งสิ้น 4 บัญชี ได้แก่ บัญชีคำพื้นฐานหรือคำค้นตาสำหรับเด็กเล็ก บัญชีคำพื้นฐานชั้นประถมศึกษาปีที่ 1 2 และ 3 รวมทั้งสิ้น 3,258 คำ โดยมีบางคำเป็นคำรวมอยู่ในบัญชีคำของแต่ละกลุ่มอายุด้วย จากการศึกษาในครั้งนี้ กระทรวงศึกษาธิการได้ระบุข้อเสนอแนะว่าควรมีการศึกษาคำจากหนังสือเรียน หนังสืออ่านประกอบ คำที่ใช้ในวิทยุ โทรทัศน์ และหนังสือพิมพ์ เพื่อนำมาเปรียบเทียบกับคำที่นักเรียนใช้พูด และควรนำคำที่ศึกษาเหล่านั้นมาจัดเป็นประมวลคำศัพท์วิชาการในสาขาวิชาต่าง ๆ เพื่อ

นำไปใช้สร้างสื่อการเรียนการสอน นอกจากนี้ การศึกษาคำพื้นฐาน น่าจะมีการศึกษาเป็นระยะ ๆ เพื่อให้ทันกับความเจริญของสังคมด้วย อย่างไรก็ตามหลังจากนั้นก็ยังไม่ผู้ใดศึกษาเกี่ยวกับคำศัพท์พื้นฐานดังกล่าวอีกเลย

ผู้วิจัยมีความเห็นว่าการหารายการคำศัพท์ที่สามารถใช้ในการเรียนการสอนภาษาไทยได้นั้น เป็นประเด็นสำคัญที่ต้องมีการศึกษาวิจัยอย่างลึกซึ้ง และเพื่อที่จะได้ข้อมูลเบื้องต้น การสำรวจและวิเคราะห์ข้อมูลเนื้อหาภาษาที่พบในแบบเรียนที่มีอยู่นั้นเป็นสิ่งจำเป็น เพื่อให้เห็นรายการคำศัพท์ที่ได้บรรจุในหลักสูตรจริง ๆ ที่เกิดขึ้น ว่ามีการเปลี่ยนแปลงในแต่ละระดับอย่างไร การศึกษาเปรียบเทียบลักษณะนี้จะช่วยให้เห็นภาพการเปลี่ยนแปลงและการพัฒนารายการคำศัพท์ที่ใช้ได้ชัดเจนขึ้น ซึ่งเป็นพื้นฐานที่จำเป็นต่อการนำเสนอรายการคำที่เหมาะสมแก่การเรียนการสอนภาษาไทยในระดับชั้นต่าง ๆ ต่อไป งานวิจัยชิ้นนี้จึงเป็นการต่อยอดงานวิจัยด้านคำศัพท์พื้นฐานภาษาไทยที่กระทรวงศึกษาธิการได้เคยทำไว้ในอดีต โดยจะศึกษาวิเคราะห์คำศัพท์ที่ปรากฏในแบบเรียนภาษาไทย ครอบคลุมการศึกษาชั้นพื้นฐานทั้งหมด (เริ่มเรียน-มัธยมศึกษาตอนปลาย) ตั้งแต่หลักสูตร 2503-2544

คำถามสำคัญในงานวิจัยชิ้นนี้คือคำศัพท์ในแบบเรียนภาษาไทยจากช่วงชั้นต่าง ๆ มีลักษณะเป็นอย่างไร คำศัพท์จากช่วงชั้นที่สูงกว่าจะมีจำนวนมากกว่าและครอบคลุมคำศัพท์ที่ปรากฏในช่วงชั้นที่ต่ำกว่าหรือไม่ และเมื่อพิจารณาความยากง่ายในแง่จำนวนพยางค์ ความทึบใสทางความหมาย และลักษณะทางความหมายของคำศัพท์จะพบว่าช่วงชั้นที่สูงกว่านั้นยากกว่าช่วงชั้นที่อยู่ต่ำกว่าในทุกมิติหรือไม่ อย่างไร

งานวิจัยชิ้นนี้เป็นงานวิจัยเชิงปริมาณเป็นหลัก มุ่งเน้นการวิเคราะห์ข้อมูลเชิงประจักษ์ โดยมีการเก็บรวบรวมข้อมูลเป็นจำนวนมากภายใต้หลักเกณฑ์ที่ชัดเจน วิเคราะห์ข้อมูลและใช้วิธีการทางสถิติมายืนยันผล แต่การวิเคราะห์ความหมายของคำจำเป็นต้องลงรายละเอียด ในส่วนนี้จึงดำเนินตามหลักเกณฑ์ของงานวิจัยเชิงคุณภาพ เพื่อแยกแยะความหมายแบบต่าง ๆ โดยพิจารณาจากบริบทที่ละตัวอย่าง ต้องใช้ความรู้ความเข้าใจทางภาษาศาสตร์แบบเฉพาะในการวิเคราะห์ความหมาย จึงไม่สามารถตรวจสอบความน่าเชื่อถือด้วยวิธี Intercoder for Reliability ได้ ซึ่งถือเป็นข้อจำกัดในงานวิจัยชิ้นนี้

วัตถุประสงค์การวิจัย (Objectives)

1. เพื่อศึกษาพัฒนาการและการเปลี่ยนแปลงของรายการคำศัพท์ในแบบเรียนภาษาไทย ตั้งแต่ระดับเริ่มเรียน-มัธยมศึกษาตอนปลาย ว่ามีขนาดของวงศัพท์ขยายออกไปอย่างไร
2. เพื่อศึกษาความยากง่ายในแง่จำนวนพยางค์และความทึบใสทางความหมายของคำศัพท์ในช่วงชั้นต่าง ๆ
3. เพื่อศึกษาการพัฒนาและเปลี่ยนแปลงทางความหมายของคำกริยาที่ปรากฏบ่อยเป็นคู่แรกจากช่วงชั้นต่าง ๆ

สมมติฐานการวิจัย (Research Hypothesis)

1. คำศัพท์ในช่วงชั้นที่สูงกว่าจะมีจำนวนมากกว่าและครอบคลุมคำศัพท์ในช่วงชั้นที่ต่ำกว่า
2. คำศัพท์ที่พบในช่วงชั้นที่สูงกว่ามีคำ 3 พยางค์ขึ้นไปมากกว่า และมีคำที่พบมากกว่าช่วงชั้นที่ต่ำลงมา
3. คำกริยาที่ปรากฏบ่อยเป็นคู่แรกจากช่วงชั้นที่สูงกว่าปรากฏลักษณะทางความหมายและการใช้มากกว่าช่วงชั้นที่ต่ำลงมา

นิยามศัพท์เฉพาะเชิงปฏิบัติการ (Operational Definition)

1. รูปศัพท์ หมายถึง คำศัพท์แต่ละคำที่มีรูปเขียนแตกต่างกัน ตรงกับโน้ตศัพท์ในภาษาอังกฤษว่า type เช่น *ฉันชอบเดินไปกินไป* ประโยคนี้มี 5 รูปศัพท์ ได้แก่ *ฉัน ชอบ เดิน ไป กิน*
2. รูปปรากฏ หมายถึง จำนวนการปรากฏของรูปศัพท์ทุก ๆ ครั้ง ตรงกับโน้ตศัพท์ในภาษาอังกฤษว่า token เช่น *ฉันชอบเดินไปกินไป* ประโยคนี้มี 6 รูปปรากฏ ได้แก่ *ฉัน ชอบ เดิน ไป กิน ไป* (นับ ไป 2 ครั้ง)
3. คำ 3 พยางค์ หมายถึง คำที่มีจำนวนพยางค์ตั้งแต่ 3 พยางค์ขึ้นไป
4. คำประกอบ คือ คำที่สร้างขึ้นจากคำ มากกว่า 1 คำมาประกอบกันเป็นคำใหม่ รวมความถึง คำประสม คำซ้ำ คำซ้อน และคำประสม
5. ความทึบใสทางความหมาย หมายถึง ความสัมพันธ์ระหว่างความหมายของคำประกอบกับความหมายของคำย่อยภายในคำประกอบนั้น
6. คำใส (transparent word) หมายถึง คำที่ความหมายของคำประกอบเทียบเท่าหรือใกล้เคียงกับการนำเอาความหมายของคำย่อยภายในคำประกอบ มาประกอบกัน
7. คำทึบ (opaque word) หมายถึง คำที่ความหมายของคำประกอบไม่สามารถตีความได้โดยนำเอาความหมายของคำย่อยภายในคำประกอบมาประกอบกัน แต่มีความหมายอื่นที่ต่างไป

ทบทวนวรรณกรรม/แนวคิดที่เกี่ยวข้อง (Literature Review)

การทบทวนวรรณกรรมในส่วนนี้จะเน้นที่หลักและเกณฑ์ในการวิเคราะห์ข้อมูล โดยจะอธิบายเกี่ยวกับแนวคิดในการพิจารณาความง่ายของคำเป็นลำดับแรก แล้วตามมาด้วยแนวคิดเกี่ยวกับความหลากหลายทางความหมายของคำ ทั้งนี้จะใช้แนวคิดต่าง ๆ เหล่านี้เป็นแนวทางในการวิเคราะห์ข้อมูลรายการคำศัพท์ทั้งหมด

ความยากง่ายของคำ

ความยากง่ายของคำเกี่ยวข้องกับการเลือกรายการคำที่เหมาะสมสำหรับชั้นเรียนต่าง ๆ โดยหลักคำยากไม่ควรปรากฏมากนักในช่วงชั้นต้น ๆ คำถามสำคัญ คือ เราจะพิจารณาอย่างไรว่าคำใดยากหรือง่ายกว่ากัน หากมองว่าคำที่มีมิติด้านรูปกับความหมาย ในมิติรูป มีหลายงานที่มองความยากง่ายว่าขึ้นกับจำนวนพยางค์ของคำ โดยมองว่าคำที่มีจำนวนพยางค์มากจะยากกว่าคำที่มีพยางค์น้อย เช่น งานของกิตติพงษ์ ทวีวงศ์ (2538) ที่ศึกษาเปรียบเทียบระดับความยากง่ายต่อการอ่านของหนังสือพิมพ์บางกอกโพสต์และเดอะเนชั่น นอกจากจำนวนพยางค์ ก็มีผู้ที่มีมองว่าจำนวนพยางค์ขั้นต้นที่มีมากกว่าจะทำให้คำนั้นยากกว่าด้วยเช่นกัน (กระทรวงศึกษาธิการ, 2532)

ในมิติความหมาย ความยากง่ายของคำอาจเห็นได้จากการประสมคำขึ้นเป็นคำใหม่ คำประสมมีทั้งแบบที่เป็นคำทึบ คำใส ความทึบใสทางความหมายคือสิ่งซึ่งแสดงความเกี่ยวข้องสัมพันธ์กันระหว่างความหมายของคำประกอบกับความหมายของคำย่อยที่เป็นองค์ประกอบของคำประกอบนั้น ๆ หากความหมายของคำประกอบเทียบเท่าหรือใกล้เคียงกับการนำเอาความหมายของคำย่อยภายในคำมาประกอบกัน จะจัดว่าคำนั้น ๆ เป็นคำใส (transparent word) ในทางกลับกันหากความหมายของคำประกอบไม่สามารถตีความได้จากการประกอบความหมายของคำย่อยภายในคำ แต่มีความหมายอื่นก็จะจัดให้คำนั้น ๆ เป็นคำทึบ (opaque word) โดยผู้วิจัยนำความคิดนี้มาจากงานของเพียร์ศิริ วงศ์วิภาณนท์ (2531) ที่ศึกษาเกี่ยวกับคำซ้อนในภาษาไทยโดยระบุว่า ความหมายของคำซ้อนอาจไม่เกี่ยวข้องกับผู้ประกอบของคำซ้อนนั้น ๆ ก็ได้ เช่น *เสียดสี กดขี่ ย่อยยับ* เป็นต้น ซึ่งแนวคิดเรื่องความทึบใสทางความหมายนี้สามารถใช้อธิบายคำในกลุ่มคำประสมได้ด้วยเช่นกัน เช่นคำว่า *หางเสือ* ที่หมายถึงเรือถือท้ายเรือ และแม้ว่าในงานของเพียร์ศิริจะไม่ได้ระบุลงไปชัดเจนว่าความหมายในลักษณะดังกล่าวจะเกี่ยวข้องกับความยากง่ายของคำ แต่ในเชิงการใช้ภาษา หากผู้เรียนไม่สามารถเดาความหมายของคำหนึ่ง ๆ ได้เนื่องจากความหมายนั้น ๆ ไม่เกี่ยวข้องสัมพันธ์กับองค์ประกอบภายในคำเลย คำดังกล่าวก็น่าจะเข้าใจได้ยากกว่าคำที่สามารถเดาความหมายจากองค์ประกอบได้

ไม่เพียงความทึบใสทางความหมายของคำเท่านั้น ความหมายที่ถูกใช้จริงของคำก็ที่น่าสนใจเนื่องจากคำจำนวนไม่น้อยในภาษาไทยเป็นคำหลายความหมาย ความหมายพื้นฐานที่สุดซึ่งมีแนวโน้มว่าเป็นความหมายแรกที่ถูกบรรจุไว้ในพจนานุกรมน่าจะเป็นความหมายที่ง่ายที่สุด ในขณะที่ความหมายอื่น ๆ ที่ขยายจากความหมายพื้นฐานดังกล่าวน่าจะเพิ่มระดับความยากมากยิ่งขึ้นไปด้วย

นอกจากมิติรูปและความหมายของคำแล้ว ความยากง่ายยังอาจพิจารณาจากการใช้หรือความเข้าใจในการใช้คำต่าง ๆ ซึ่งก็มีงานของกระทรวงศึกษาธิการ (2532) ที่กำหนดความยากโดยการทดสอบว่านักเรียนรู้จักและเข้าใจคำต่าง ๆ มากน้อยเพียงใดผ่านการอ่านออกเสียงและสะกดคำ คำที่นักเรียนส่วนใหญ่อ่าน และ/หรือ สะกดผิดก็จะถูกจัดกลุ่มว่ายากกว่าคำที่นักเรียนอ่านและสะกดได้ถูกต้อง

ในบทความชิ้นนี้ ผู้วิจัยต้องการศึกษาเปรียบเทียบวงศัพท์ระหว่างช่วงชั้นต่าง ๆ โดยจะไม่ลงรายละเอียดไปที่การใช้จริงของนักเรียนได้ จึงเน้นศึกษาจากรายการคำศัพท์ที่ปรากฏใช้จริงในแบบเรียนระดับต่าง ๆ โดยจะวิเคราะห์เปรียบเทียบความยากง่ายในเรื่องจำนวนพยางค์ ความทึบใสทางความหมาย รวมถึงความหลากหลายทางความหมายของคำ

ความหลากหลายทางความหมายของคำ

คำศัพท์ในภาษาไทยจำนวนหนึ่งเป็นคำที่มีหลายความหมาย หลายหน้าที่ อาจเป็นได้ทั้งที่พ้องกันโดยบังเอิญ ไม่มีความเกี่ยวข้องทางความหมายกัน หรืออาจเป็นคำหลายความหมายที่เกิดจากการขยายความหมายและหน้าที่ (Semantic/Functional Extension) รวมไปถึงการผ่านกระบวนการกลายเป็นคำทางไวยากรณ์ (Grammaticalization) คำว่า *ไป* และ *มา* ก็อยู่ในกลุ่มนี้ด้วย

ผู้วิจัยพบว่า คำว่า *มา* ปรากฏอยู่ในบทแรกของหนังสือเรียนภาษาไทยชั้นประถมศึกษาปีที่ 1 หลักสูตร 2533 ฉบับกระทรวงศึกษาธิการ นั่นคือประโยค “*ตา มา รถไฟ*” จะเห็นได้ว่า *มา* ในที่นี้ ทำหน้าที่เป็นกริยาหลักและกริยาเดียวของประโยค จึงสันนิษฐานได้ว่า *ไป* และ *มา* ในฐานะคำกริยาหลัก ควรเป็นความหมายหลักพื้นฐานความหมายแรกที่เราควรเรียนรู้ก่อน ซึ่งสอดคล้องกับการนิยามความหมายของทั้งสองคำนี้ในพจนานุกรมเล่มต่าง ๆ (นพวรรณ พันธุมธธา, 2555; “พจนานุกรมฉบับมติชน,” 2547; ราชบัณฑิตยสถาน, 2556) ที่นิยามความหมายแรกของ *ไป* และ *มา* เป็นความหมายของคำกริยาหลักในประโยค นอกจากนี้ ยังสันนิษฐานต่อไปว่า *ไป* และ *มา* ในฐานะคำกริยาหลักและกริยาเดียวของประโยคน่าจะพบมากที่สุดในช่วงชั้นที่ 1 หรือประถมต้น เนื่องจากโครงสร้างประโยคจะไม่ซับซ้อนนัก

ตามที่ได้กล่าวไว้ข้างต้นว่า *ไป* และ *มา* สามารถปรากฏในตำแหน่งคำกริยานำและกริยารองเพื่อช่วยเสริมความหมายของคำกริยาหลักได้ด้วยนั้น โครงสร้างการปรากฏดังกล่าวมีลักษณะเป็นคำกริยาที่เรียงต่อกัน หรือที่นักภาษาศาสตร์เรียกว่า หน่วยสร้างกริยาเรียง (Serial verb construction) ซึ่งเกรียงไกร วัฒนาสวัสดิ์ (2544) อธิบายไว้ว่า หมายถึง หน่วยสร้างที่เกิดจากการเรียงกันของคำกริยาสองตัวหรือมากกว่า โดยที่คำกริยาทุกตัวเหล่านั้นต้องสามารถปรากฏเป็นกริยาเดียวในประโยคได้ อาจมีประธานร่วมกันหรือไม่ก็ได้ แต่ต้องมีกาล (Tense) หรือการณลักษณะ (Aspect) เดียวกัน ต้องไม่มีลักษณะใด ๆ ที่แสดงว่ากริยาตัวที่หนึ่งและสองอยู่คนละประโยครวมถึงต้องไม่ปรากฏคำเชื่อมระหว่างกริยาทั้งสองนั้น นอกจากนี้ คำกริยาที่ปรากฏอยู่ในหน่วยสร้างกริยาเรียงอาจเป็นกริยาแสดงการกระทำ (Action verb) หรือกริยาแสดงสภาพ (State verb) ก็ได้ โดยจุดกำเนิดของหน่วยสร้างกริยาเรียงนี้เกิดจากปรากฏการณ์สูญรูป (Zero anaphora) คือ มีการละค่านามในตำแหน่งประธานหรือกรรมของประโยค ทำให้เกิดการเรียงลำดับเหตุการณ์ขึ้นในภาษา (Serialization) จนเกิดเป็นกริยาเรียงดังกล่าวขึ้น

ด้วยเหตุที่มีการละค่านามในตำแหน่งต่าง ๆ ของประโยค ประโยคที่เกิดขึ้นใหม่จึงน่าจะมีความซับซ้อนมากขึ้น *ไป* และ *มา* ที่เป็นส่วนหนึ่งของกริยาเรียงจึงน่าจะมีโครงสร้างที่ซับซ้อนกว่ากริยา *ไป-มา*

ที่พบเดี่ยว ๆ ยิ่งไปกว่านั้นยังพบอีกว่า *ไป* และ *มา* ที่ปรากฏอยู่ในหน่วยสร้างกริยาเรียงนี้มีลักษณะทางความหมายที่แตกต่างกันไปขึ้นกับตำแหน่งและคำกริยาที่ปรากฏร่วมด้วย กล่าวคือ เมื่อ *ไป* และ *มา* ปรากฏหน้าคำกริยาอื่น จะแสดงความหมายในเชิงการเคลื่อนที่เพื่อ *ไป/มา* กระทำการบางอย่าง อาจเป็นการบอกทิศทาง เช่น *เขาไปซื้อของ* หรือแสดงการเปลี่ยนสภาพบางอย่าง เช่น *อย่ามาฉฉฉฉฉฉ* เป็นต้น หรือเมื่อ *ไป* และ *มา* ตามหลังคำกริยาอื่น ก็แสดงความหมายได้สองความหมาย คือ แสดงทิศทางการเคลื่อนที่ของคำกริยาที่อยู่ข้างหน้า เช่น *เขาเดินไปโรงเรียน* และแสดงการณลักษณะ เช่น *เขาวิ่งมาลักลักแล้ว* (ชาภินี มณีนาวาชัย, 2551)

ความหมายประจำคำในฐานะคำกริยาหลักของ *ไป* และ *มา* หมายถึง “การเคลื่อนที่” จะเลือกใช้คำใดขึ้นกับจุดอ้างอิงของผู้พูด ความหมายของกริยาหลักดังกล่าวค่อนข้างใกล้เคียงกับ *ไป* และ *มา* ในฐานะกริยานำและกริยารองที่ยังคงเสริมความหมายด้านทิศทาง ความหมายส่วนนี้จึงน่าจะถูกสอนในลำดับต่อมา เนื่องจากแม้ว่าความหมายจะสื่อถึงทิศทาง คล้ายคลึงและใกล้เคียงกับคำกริยาหลัก แต่ลักษณะทางโครงสร้างมีความซับซ้อนมากกว่า ส่วน *ไป* และ *มา* ที่มีความหมายซับซ้อนห่างจากการแสดงทิศทางการเคลื่อนที่ คือ แทนที่จะเป็นการเคลื่อนที่ทางกายภาพ กลับแปรไปในการเคลื่อนที่ทางเวลาซึ่งมีความเป็นนามธรรมมากกว่า รวมถึงความหมายในเชิงการเปลี่ยนสภาพ และความหมายที่แสดงการณลักษณะลักษณะดังกล่าวน่าจะซับซ้อนและเข้าใจยากขึ้นอีกขึ้น ดังนั้น ลักษณะความหมายดังกล่าวจะพบมากขึ้นในช่วงชั้นที่สูงขึ้นไปอีก ส่วนสำนวนที่มี *ไป* และ/หรือ *มา* เป็นองค์ประกอบถือเป็นกลุ่มก่อนความหมายใหม่ ผู้เรียนต้องเรียนรู้และตีความเพื่อเข้าใจความหมายของสำนวนเหล่านั้น ก็น่าจะถือได้ว่ามีระดับความยากเพิ่มขึ้นอีกชั้นหนึ่ง ดังนั้น รูปแบบโครงสร้างการปรากฏของ *ไป* และ *มา* ที่เริ่มเข้าใจได้ยากขึ้น ทั้งในฐานะคำกริยานำและกริยารอง รวมถึงส่วนหนึ่งของสำนวน น่าจะเป็นโครงสร้างที่ปรากฏน้อยหรือไม่ปรากฏเลยในช่วงชั้นต้น แต่เพิ่มจำนวนมากขึ้นในช่วงชั้นที่สูงกว่า

ระเบียบวิธีวิจัย (Research Methodology)

เพื่อจะตอบคำถามในงานวิจัยนี้ ผู้วิจัยจำเป็นต้องสร้างคลังข้อมูลแบบเรียนภาษาไทยขึ้นมาก่อนเป็นอันดับแรก แล้วจึงจะสามารถนำข้อมูลที่ได้จากคลังข้อมูลภาษามาศึกษาวิเคราะห์เป็นลำดับต่อไป ในหัวข้อนี้จะแยกออกเป็นสองส่วน คือ 1) การสร้างคลังข้อมูล และ 2) การวิเคราะห์ข้อมูล

การสร้างคลังข้อมูล

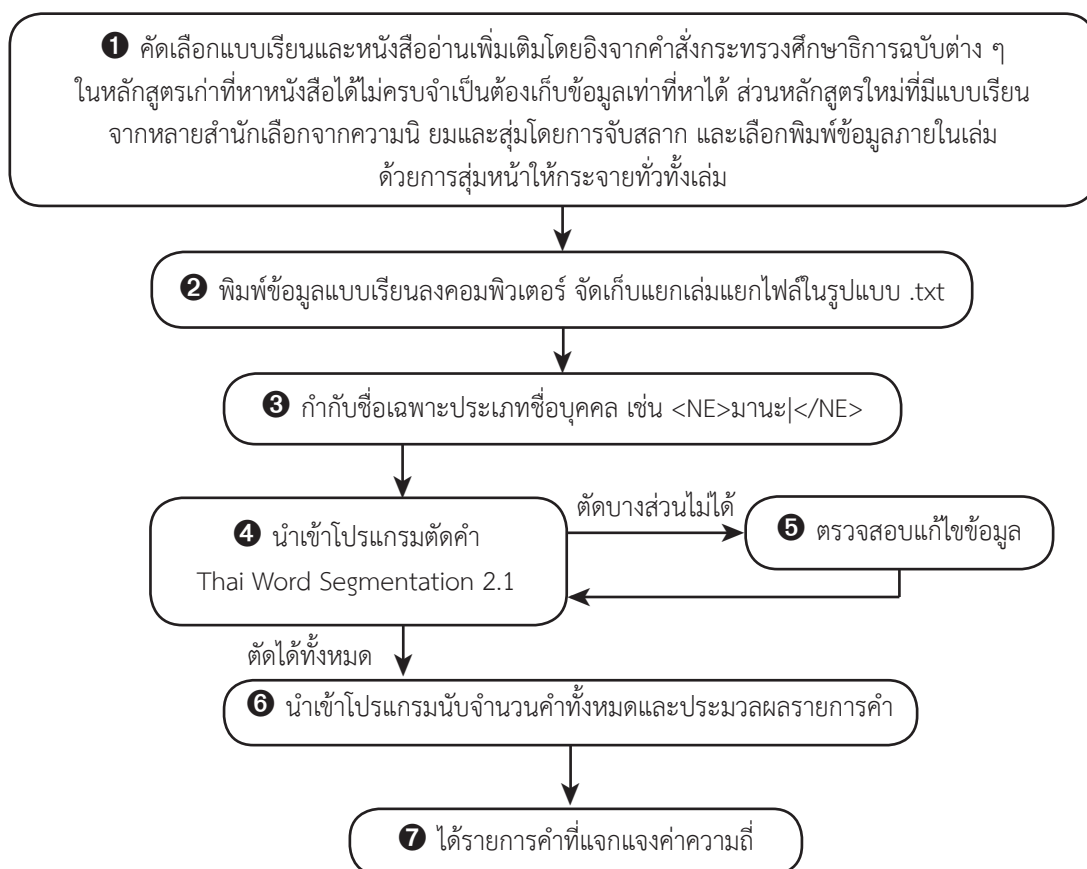
การสร้างคลังข้อมูลภาษาในงานวิจัยทางภาษาศาสตร์คลังข้อมูลนั้นมีความสำคัญและใช้เวลามาก ผู้วิจัยได้สร้างคลังข้อมูลแบบเรียนภาษาไทย (Thai Textbook Corpus: TTC) ขึ้น โดยการสุ่มแบบเรียนและหนังสืออ่านเพิ่มเติมสาระการเรียนรู้ภาษาไทย ในระดับการศึกษาขั้นพื้นฐาน ช่วงชั้นที่ 1 (ระดับเริ่มเรียน-ประถมศึกษาปีที่ 3) ช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 4-6) ช่วงชั้นที่ 3 (มัธยมศึกษาปีที่ 1-3) และช่วงชั้นที่ 4 (มัธยมศึกษาปีที่ 4-6) ตามที่ระบุไว้ในหลักสูตร ดังต่อไปนี้

1. หลักสูตร พ.ศ. 2503
2. หลักสูตรประถมศึกษาและมัธยมศึกษาตอนต้น พ.ศ. 2521
3. หลักสูตรมัธยมศึกษาตอนปลาย พ.ศ. 2524
4. หลักสูตร พ.ศ. 2521 (ฉบับปรับปรุง พ.ศ. 2533)
5. หลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2544

ทั้งนี้ หลักสูตร 2503 เป็นหลักสูตรเดียวที่มีการแบ่งช่วงชั้นแตกต่างจากหลักสูตรอื่น ๆ กล่าวคือ ช่วงชั้นที่ 1 จะครอบคลุมระดับเริ่มเรียน-ประถมศึกษาปีที่ 4 ช่วงชั้นที่ 2 ครอบคลุมชั้นประถมศึกษาปีที่ 5-7 ช่วงชั้นที่ 3 ครอบคลุมชั้นมัธยมศึกษาปีที่ 1-3 และช่วงชั้นที่ 4 ครอบคลุมชั้นมัธยมศึกษาปีที่ 4-5 นอกจากนี้หลักสูตรในข้อ 2 และ 3 ที่ระบุข้างต้นผู้วิจัยจะถือเป็นข้อมูลแบบเรียนจากยุคเดียวกันเนื่องจากเกิดขึ้นในเวลาใกล้เคียงกัน และจะเรียกรวมว่าเป็นข้อมูลจากหลักสูตร 2521 อย่างไรก็ตาม ในบทความชิ้นนี้ จะยังไม่พิจารณาปัจจัยความแตกต่างในด้านมิติช่วงเวลา แต่สนใจศึกษาพัฒนาการและการเปลี่ยนแปลงของวงศ์พินตามช่วงชั้นที่สูงขึ้นก่อน

ข้อมูลแบบเรียนภาษาไทยถูกสุ่มขึ้นจากแบบเรียนตามช่วงชั้นและหลักสูตรดังกล่าวข้างต้นรวมทั้งสิ้น 354 เล่ม คัดเลือกเฉพาะส่วนที่เป็นข้อความต่อเนื่องและไม่รวมข้อมูลแบบฝึกหัด แบบคัดลายมือ แบบเรียนท่องถิ่น/แบบเรียนเฉพาะกลุ่ม ภาษาในวรรณคดีชั้นสูง และคำประพันธ์ประเภทที่มีฉันทลักษณ์ซับซ้อน¹ คลังข้อมูลนี้เป็นตัวแทนภาษาไทยแบบทั่วไป ขั้นตอนการสร้างคลังข้อมูลอาจสรุปเป็นแผนผังได้ ดังนี้

¹ จำเป็นต้องรวมคำประพันธ์ประเภทกลอนสี่ กลอนแปด เอาไว้ด้วย เนื่องจากข้อมูลลักษณะนี้ปรากฏเป็นจำนวนมากในช่วงชั้นที่ 1



ภาพที่ 1: ขั้นตอนการสร้างคลังข้อมูล

Figure 1: The Process of Corpus Compilation

ที่มา: จากการรวบรวมข้อมูล

Source: From data collection

สาเหตุที่ต้องกำกับข้อมูลที่เป็นชื่อเฉพาะ ก็เพื่อจำแนกชื่อเฉพาะเหล่านี้ออกจากคำทั่วไปที่ต้องการศึกษา เนื่องจากความถี่ในการปรากฏของคำเป็นปัจจัยสำคัญในการศึกษา เมื่อชื่อเฉพาะบางชื่อไปพ้องกับคำทั่วไป เช่นคำว่า มานะ ในแง่ชื่อเฉพาะถือเป็นหนึ่งในตัวละครหลักของแบบเรียนหลักสูตร 2521 แต่คำว่า มานะ นี้ ก็มีความหมายในฐานะคำทั่วไป ผู้วิจัยเห็นว่าควรแยก มานะ ที่เป็นชื่อเฉพาะออก เพื่อให้ความถี่ของคำ ๆ นี้ เป็นความถี่ที่เกิดจากการปรากฏใช้แบบคำทั่วไปไม่ใช่ชื่อเฉพาะ จึงได้ดำเนินการกำกับชื่อเฉพาะที่ปรากฏในแบบเรียนให้มากที่สุดเท่าที่จะทำได้

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลที่ได้จากคลังข้อมูลที่สร้างขึ้นนั้นแบ่งออกเป็น 3 ส่วน ได้แก่ 1) การวิเคราะห์วงศัพท์ของช่วงชั้นต่าง ๆ 2) ความยากง่ายของคำศัพท์ในแง่จำนวนพยางค์และความถี่ใส่ทางความ

หมาย และ 3) ความยากง่ายของคำศัพท์ในแง่ความหมายและการใช้ แต่ละส่วนจะมีรูปแบบการวิเคราะห์ที่แตกต่างกัน รายละเอียดแต่ละส่วน เป็นดังนี้

การวิเคราะห์วงศัพท์ในแต่ละช่วงชั้น ใช้โปรแกรมคอมพิวเตอร์ที่เขียนขึ้นมาประมวลผลข้อมูลรายการคำ พิจารณาข้อมูลจาก 4 ช่วงชั้น ทั้งรูปศัพท์และรูปปรากฏ ดูการกระจายตัวของคำศัพท์ตามช่วงความถี่ต่าง ๆ เปรียบเทียบการปรากฏของคำศัพท์ในแต่ละช่วงชั้น พิจารณาจำนวนคำศัพท์ที่เพิ่มขึ้นและเกิดร่วมกัน คำนวณออกมาเป็นร้อยละ โดยผู้วิจัยต้องการศึกษาว่ามีการเพิ่มขึ้นของวงศัพท์ที่นักเรียนในแต่ละระดับจะได้พบในแบบเรียนเมื่อเรียนจบช่วงชั้นนั้น ๆ มากน้อยเพียงใด ผู้วิจัยจึงจะเปรียบเทียบรายการคำศัพท์ที่พบในแต่ละช่วงชั้น กล่าวคือ เทียบรายการศัพท์ประสมปลายกับประสมต้น เทียบรายการคำศัพท์ของมัธยมต้นกับของประถมต้นและปลายรวมกัน แล้วจึงเทียบรายการคำศัพท์ของมัธยมปลายกับของทั้งชั้นประถมต้น ประสมปลายและมัธยมต้นรวมกัน

การวิเคราะห์ความยากง่ายของคำศัพท์ในแง่จำนวนพยางค์และความทึบใสทางความหมาย เมื่อได้รายการคำศัพท์ทั้งหมดมาแล้ว พิจารณาจำนวนพยางค์และความทึบใสทางความหมายของคำศัพท์ทีละคำ มิติเรื่องจำนวนพยางค์วิเคราะห์จำแนกเป็น 1) คำ 1 พยางค์ 2) คำ 2 พยางค์ และ 3) คำที่มีตั้งแต่ 3 พยางค์ขึ้นไป ส่วนเรื่องความทึบใสทางความหมาย ผู้วิจัยได้ปรับเปลี่ยนจากเกณฑ์การวัดระดับความสัมพันธ์ทางความหมาย (degree of semantic relatedness) ของนาگیและแอนเดอร์สัน (Nagy & Anderson, 1984) จำแนกออกเป็น 4 กลุ่ม SEM 0-3 รายละเอียดของแต่ละกลุ่ม เป็นดังนี้

- 1) SEM 0 คือ คำเดี่ยว อาจมี 1 พยางค์หรือหลายพยางค์ก็ได้ เช่น *ฉัน ข้าพเจ้า นาฬิกา* เป็นต้น
- 2) SEM 1 คือ คำประกอบ ประเภทคำใส โดยเป็นคำที่ความหมายรวมมาจากองค์ประกอบมีส่วนหลักส่วนขยาย (Head-modifier) หรือคำที่องค์ประกอบตัวใดตัวหนึ่งแทนได้ (ในกรณีที่ต้ององค์ประกอบทั้งหมดมีความหมายไปในทางเดียวกัน) ได้แก่ คำประสมและคำซ้ำทั้งหมด คำซ้อนและคำประสมบางคำ เช่น *นักเรียน สบายสบาย เด็ก ๆ ร่างกาย ดอกไม้* เป็นต้น
- 3) SEM 2 คือ คำประกอบ ก้ำกึ่ง ไม่ใส ไม่ทึบ ความหมายของคำยังเกี่ยวข้องกับความหมายขององค์ประกอบบ้าง แต่ไม่ชัดเจนนัก หรือคำที่ความหมายอยู่ที่ยึดองค์ประกอบตัวใดตัวหนึ่งเท่านั้น (กรณีที่ต้ององค์ประกอบมีความหมายไม่เป็นไปในทางเดียวกัน) ได้แก่ คำซ้อนและคำประสมบางคำ เช่น *เดินทาง ช่วยเหลือ แทรกซ้อน* เป็นต้น
- 4) SEM 3 คือ คำประกอบ ประเภทคำทึบ ความหมายย้ายที่ เปลี่ยนไปโดยสิ้นเชิง ได้แก่ คำซ้อนและคำประสมบางคำ เช่น *รับรอง หลักฐาน ผลักดัน ดุดัน* เป็นต้น

เมื่อได้ผลการวิเคราะห์ทั้งสองส่วนดังกล่าวแล้ว นำผลวิเคราะห์ที่ได้เข้าโปรแกรมคอมพิวเตอร์ที่เขียนขึ้นเพื่อประมวลผลออกมาว่าคำศัพท์ในแต่ละช่วงชั้นมีจำนวนพยางค์และความทึบใสทางความหมายแตกต่างกันหรือไม่ ในส่วนนี้จะใช้สถิติไคสแควร์ (chi-square) เพื่อทดสอบความเกี่ยวข้องสัมพันธ์กันระหว่างตัวแปรช่วงชั้นกับความยากง่ายในมิติจำนวนพยางค์และความทึบใสทางความหมาย

การวิเคราะห์ความยากง่ายของคำศัพท์ในแง่ความหมายและการใช้ การวิเคราะห์ในส่วนนี้ จำเป็นต้องลงลึกในรายละเอียด จึงไม่สามารถวิเคราะห์ความหมายของทุกคำได้ ผู้วิจัยจึงเลือกคำกริยาที่ปรากฏบ่อยที่สุดคู่แรกซึ่งเป็นคู่ตรงข้ามกันมาศึกษา คือคำว่า *ไป* และ *มา* สุ่มวิเคราะห์ตัวอย่างจากแต่ละคลังข้อมูลย่อย คลังละ 100 ตัวอย่าง รวมเป็น *ไป* 400 ตัวอย่าง และ *มา* 400 ตัวอย่าง ซึ่งจากการทบทวนวรรณกรรม ลักษณะความหมายของ *ไป* และ *มา* อาจจำแนกออกเป็นกลุ่มจากง่ายไปหายากได้ ดังนี้

- 1) กริยาเดี่ยว-แสดงการเคลื่อนที่
- 2) กริยานำ/กริยารอง-แสดงการเคลื่อนที่เพื่อไปหรือมาทำอะไรบางอย่าง (เสริมความหมายเกี่ยวกับทิศทาง)
- 3) กริยานำ/กริยารอง-แสดงการเปลี่ยนสภาพ การณ์ลักษณะ และแสดงความเกินพอดี (เสริมความหมายที่ไม่เกี่ยวกับทิศทาง)
- 4) ปรากฏอยู่ในสำนวน

เมื่อได้ผลการวิเคราะห์ลักษณะทางความหมายจากตัวอย่างที่สุ่มมาทั้งหมดแล้ว นำมาแจกแจงเป็นตารางเพื่อดูสัดส่วนลักษณะความหมายที่เกิดขึ้นในแต่ละช่วงชั้น รวมถึง *ไป* และ *มา* ที่ปรากฏในช่วงชั้นที่สูงกว่ามีความหมายมากกว่าคำคู่นี้ที่ปรากฏในช่วงชั้นที่ต่ำกว่าหรือไม่

ผลการศึกษา (Results)

ในหัวข้อนี้จะแบ่งออกเป็น 3 ส่วน ซึ่งสอดคล้องไปกับการวิเคราะห์ข้อมูลที่ได้กล่าวถึงไปแล้ว คือ 1) วงศัพท์ของช่วงชั้นต่าง ๆ จากคลังข้อมูลภาษา 2) ความยากง่ายของคำศัพท์ในแง่จำนวนพยางค์และความซับซ้อนทางความหมาย และ 3) ความยากง่ายของคำศัพท์ในแง่ความหมายและการใช้

วงศัพท์ของช่วงชั้นต่าง ๆ จากคลังข้อมูลภาษา

คลังข้อมูลที่ได้มีขนาด 3,037,772 รูปปรากฏ ซึ่งเป็นคำศัพท์ 19,494 รูปศัพท์ สามารถแบ่งออกเป็นคลังข้อมูลย่อย (Subcorpora) 4 คลังตามช่วงชั้นต่าง ๆ โดยคลังข้อมูลย่อยที่ 1 คือ คลังข้อมูลของช่วงชั้นที่ 1 (TTC-Level 1: TTC-L1) คลังข้อมูลย่อยที่ 2 คือ คลังข้อมูลของช่วงชั้นที่ 2 (TTC-Level 2: TTC-L2) คลังข้อมูลย่อยที่ 3 คือ คลังข้อมูลของช่วงชั้นที่ 3 (TTC-Level 3: TTC-L3) และคลังข้อมูลย่อยที่ 4 คือ คลังข้อมูลของช่วงชั้นที่ 4 (TTC-Level 4: TTC-L4) รายละเอียดตามตารางที่ 1

ตารางที่ 1: จำนวนคำในคลังข้อมูลย่อย

Table 1: Word Count in Subcorpora

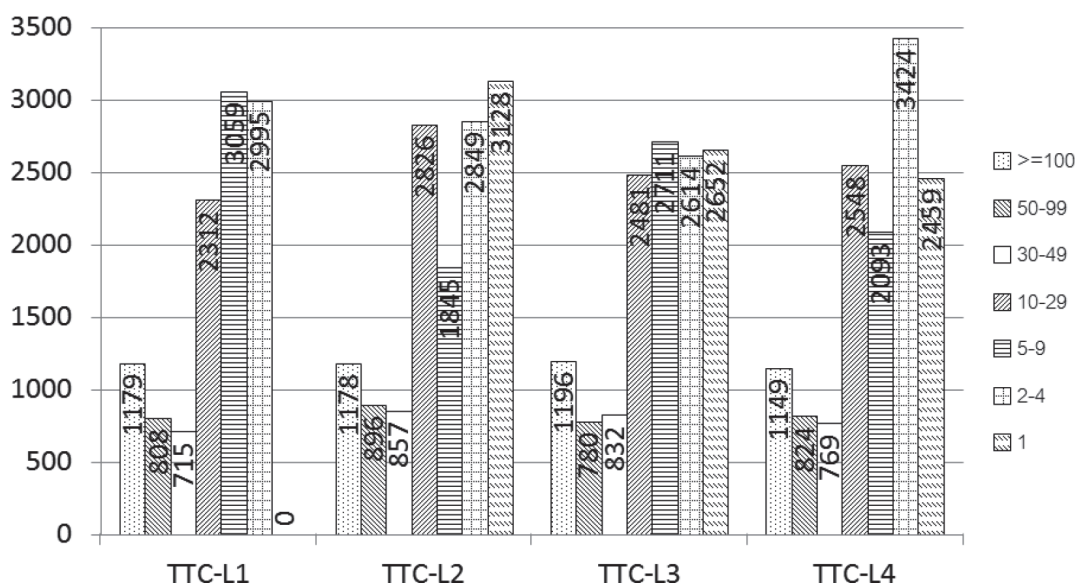
คลังข้อมูลย่อย	รูปศัพท์	รูปปรากฏ
TTC-L1	11,068	444,044
TTC-L2	13,579	704,289
TTC-L3	13,266	852,099
TTC-L4	13,266	1,037,340
รวม	19,494	3,037,772

ที่มา: จากการรวบรวมข้อมูล

Source: From data collection

แต่ละคำที่ปรากฏในรายการคำนั้นมีความถี่แตกต่างกันตั้งแต่ 1-63,126 ครั้ง คำศัพท์ราว 100 รูปศัพท์แรกมีความถี่สะสมรวมที่ร้อยละ 50 จากความถี่รวมทั้งหมดของคลังข้อมูล จึงจะเห็นได้ว่าคำที่ใช้มากจริง ๆ มีอยู่ราว 100 คำ ซึ่งหลายคำมีลักษณะเป็นคำหลายหน้าที่ หลายความหมาย เช่น *ที่ไปมา ขึ้น ลง เข้า ออก ว่า ให้* เป็นต้น จำนวนครั้งในการปรากฏของคำศัพท์แต่ละคำถือว่าต่างกันอย่างชัดเจนในที่นี้จึงจะจำแนกจำนวนคำศัพท์ที่ปรากฏในแต่ละช่วงความถี่ไว้ด้วย และเพื่อเปรียบเทียบสัดส่วนการปรากฏในคลังข้อมูลย่อยทั้งสี่ที่แต่ละคลังย่อยมีจำนวนรูปปรากฏไม่เท่ากัน จึงไม่สามารถนำความถี่ที่เกิดขึ้นจริง (raw frequency) มาเปรียบเทียบกันโดยตรงได้ จำเป็นต้องปรับฐานของจำนวนรวมรูปปรากฏให้เท่ากันก่อน โดยวิธีที่นิยมใช้คือการปรับฐานไปอยู่ที่ 1,000,000 เท่ากัน (per million words) (McEnery, Xiao, & Tono, 2006, p. 53) ซึ่งจากการปรับฐานดังกล่าวส่งผลต่อความถี่ในการปรากฏของคำศัพท์ใน TTC-L1 เนื่องจากเป็นคลังข้อมูลย่อยที่มีขนาดเล็กที่สุด รูปปรากฏที่เดิมปรากฏเพียง 1 ครั้ง จึงถูกปรับขึ้นมาเป็น 2 ทั้งหมด ใน TTC-L1 จึงไม่ปรากฏรูปศัพท์ที่มีรูปปรากฏเพียง 1 ครั้งด้วยเหตุผลดังกล่าว

ผลที่ได้จากคลังข้อมูลย่อยทั้งสี่เป็นไปในทิศทางเดียวกัน กล่าวคือ คำที่ปรากฏ 100 ครั้งขึ้นไปมีอยู่เกือบ 1,200 รูปศัพท์ หรือราวร้อยละ 10 ของจำนวนรูปศัพท์ทั้งหมดในแต่ละคลังข้อมูลย่อย เมื่อนับรวมคำที่ปรากฏมากกว่า 30 ครั้งในทั้งสี่คลังข้อมูลย่อยก็จะมีจำนวนรูปศัพท์พอ ๆ กัน คิดเป็นร้อยละ 25 ของวงศ์พททั้งหมด ในขณะที่คำซึ่งปรากฏด้วยความถี่ต่ำกว่า 30 ครั้งมีปรากฏเป็นจำนวนมาก ราวร้อยละ 75 ของวงศ์พทในแต่ละคลังย่อย แต่คำที่ปรากฏบ่อยในข้อมูลนี้เมื่อคิดเป็นขนาดของรูปปรากฏก็จะเป็นปริมาณส่วนใหญ่ของคลังข้อมูล โดยเมื่อนับจำนวนความถี่รวมของคำที่ปรากฏมากกว่า 30 ครั้งจะคิดเป็นประมาณร้อยละ 93 ทั้งสี่คลังข้อมูลย่อยซึ่งก็เป็นลักษณะปกติของการใช้ภาษาโดยทั่วไป รายละเอียดตามภาพที่ 2



ภาพที่ 2: จำนวนคำศัพท์ที่ปรากฏในแต่ละช่วงความถี่

Figure 2: The Number of Vocabulary in each Frequency Range

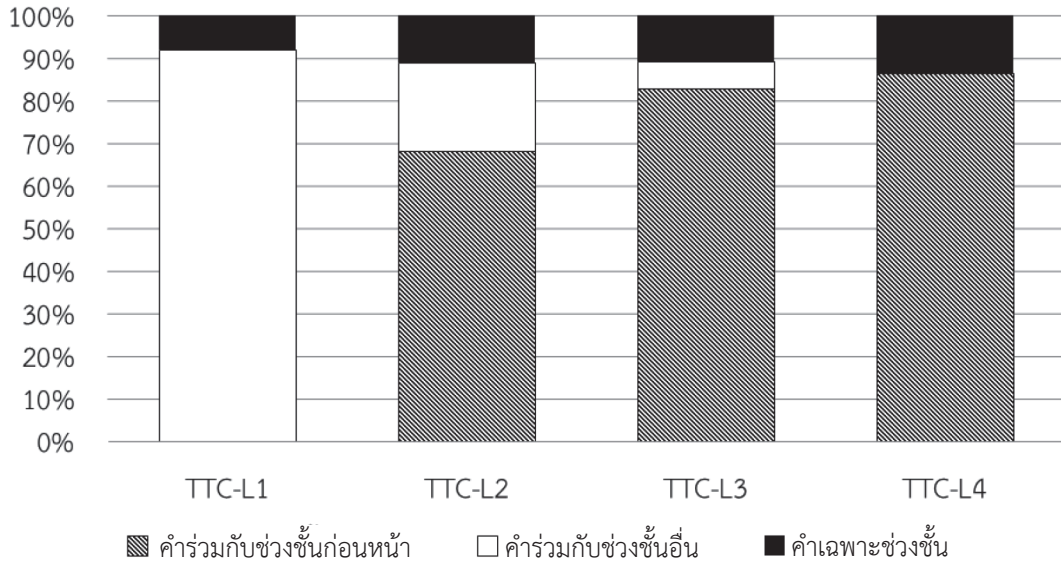
ที่มา: จากการวิเคราะห์ข้อมูล

Source: From data analysis

เมื่อพิจารณาเปรียบเทียบ TTC-L1-4 พบว่ามีคำศัพท์ที่เกิดร่วมกันทั้ง 4 คลังข้อมูลเป็นจำนวน 7,322 คำ นอกจากนี้ ยังมีคำศัพท์ที่เกิดร่วมกันใน 2 หรือ 3 คลังข้อมูล โดยเมื่อเทียบขึ้นไปทีละชั้นจากระดับประถมต้นไปจนถึงมัธยมปลาย จะพบจำนวนคำรวม ดังนี้

TTC-L2 มีรูปศัพท์ร่วมกับ TTC-L1	9,235 คำ	คิดเป็นร้อยละ 68.01 ของ TTC-L2
TTC-L3 มีรูปศัพท์ร่วมกับ TTC-L1+2	10,971 คำ	คิดเป็นร้อยละ 82.70 ของ TTC-L3
TTC-L4 มีรูปศัพท์ร่วมกับ TTC-L1+2+3	11,479 คำ	คิดเป็นร้อยละ 86.53 ของ TTC-L4

จำนวนคำที่เกิดร่วมกันระหว่างช่วงชั้นที่สูงขึ้นกับช่วงชั้นที่อยู่ล่างลงมาเพิ่มมากขึ้นเรื่อย ๆ เนื่องจากในการพิจารณาเปรียบเทียบได้รวมช่วงชั้นที่อยู่ต่ำกว่าช่วงชั้นที่ต้องการเทียบเข้าไว้ด้วยกัน จนฐานใหญ่ขึ้นเรื่อย ๆ โดยตั้งอยู่บนหลักการที่ว่านักเรียนในช่วงชั้นที่สูงขึ้นนั้นจำเป็นหรือควรจะได้เรียนรู้คำศัพท์ในช่วงชั้นที่ตนเองได้เคยผ่านมาแล้วทั้งหมด อย่างไรก็ตาม พบว่าคำศัพท์ที่ปรากฏในช่วงชั้นที่สูงกว่าไม่ได้ครอบคลุมคำศัพท์ทั้งหมดจากช่วงชั้นที่ต่ำกว่า โดยครอบคลุมเพียงบางส่วน เช่นเดียวกับคำศัพท์เฉพาะที่ปรากฏเฉพาะในคลังข้อมูลย่อยคลังใดคลังหนึ่งเท่านั้น ก็มีปรากฏเพิ่มมากขึ้นเช่นกัน คำศัพท์ที่ปรากฏร่วมกับช่วงชั้นอื่น ๆ ที่ไม่ใช่ช่วงชั้นก่อนหน้า (แถบสีขาว) ใน TTC-L2 และ TTC-L3 มีจำนวนไม่มากนัก ซึ่งสัดส่วนปริมาณของคำในกลุ่มนี้จาก TTC-L2 มีมากกว่า TTC-L3 รายละเอียดตามภาพที่ 3



ภาพที่ 3: สัดส่วนคำที่เกิดร่วมกันและคำที่เกิดเฉพาะในคลังข้อมูลย่อย
Figure 3: Proportion of Shared and Specific Word between Subcorpora

ที่มา: จากการวิเคราะห์ข้อมูล

Source: From data analysis

ผู้วิจัยนำข้อมูลคำศัพท์ทั้งหมดมาพิจารณาในแง่ของรูปศัพท์ที่เพิ่มขึ้นในแต่ละช่วงชั้นจากคลังข้อมูลย่อยต่าง ๆ เพื่อให้เห็นการขยายของวงศัพท์ที่ชัดเจนยิ่งขึ้น โดยที่รูปศัพท์ตั้งต้นใน TTC-L1 มีจำนวน 11,068 รูปศัพท์ พบว่าจำนวนและร้อยละของรูปศัพท์ที่เพิ่มขึ้นเมื่อเทียบกับช่วงชั้นที่อยู่ก่อนหน้าทั้งหมด เป็นดังนี้

เพิ่มขึ้นใน	TTC-L2	จำนวน	4,344	รูปศัพท์ ²	คิดเป็นร้อยละ 39.25
เพิ่มขึ้นใน	TTC-L3	จำนวน	2,295	รูปศัพท์	คิดเป็นร้อยละ 16.90
เพิ่มขึ้นใน	TTC-L4	จำนวน	1,787	รูปศัพท์	คิดเป็นร้อยละ 13.47

จากข้อมูลที่ได้จะเห็นว่าคำศัพท์เพิ่มขึ้นในช่วงชั้นที่ 2 (ช่วงชั้นประถมศึกษาตอนปลาย) มากกว่าที่เพิ่มขึ้นในช่วงชั้นอื่น ๆ แสดงให้เห็นว่าแบบเรียนสำหรับเด็กในช่วงอายุ 9-11 ปี เป็นแบบเรียนที่มีพัฒนาการในการเพิ่มพูนคำศัพท์สูงสุด

² ตัวเลขของจำนวนรูปศัพท์ที่เพิ่มขึ้นในส่วนนี้ไม่สามารถคำนวณจากตัวเลขซึ่งแสดงในตารางที่ 1 ได้ เนื่องจากรูปศัพท์ที่ปรากฏในแต่ละคลังข้อมูลย่อยมีความแตกต่างกัน เช่น บางคำปรากฏใน TTC-L1 เท่านั้น ไม่ปรากฏใน TTC-L2 จึงไม่สามารถนำตัวเลขมาหักลบกันได้โดยตรง ตัวเลขในส่วนนี้ได้จากการประมวลผลด้วยโปรแกรมเปรียบเทียบคำศัพท์

ความยากง่ายของคำศัพท์ในแง่จำนวนพยางค์และความถี่ใสทางความหมาย

เมื่อพิจารณาการเปลี่ยนแปลงด้านจำนวนพยางค์ พบว่าสัดส่วนรูปศัพท์ของจำนวนคำที่มี 1 พยางค์: คำที่มี 2 พยางค์ : คำที่มีตั้งแต่ 3 พยางค์ขึ้นไป ของ TTC-L1-4 อยู่ที่ราว 3 : 5 : 2 ตามลำดับ ในขณะที่สัดส่วนของรูปปรากฏใน TTC-L1 อยู่ที่ประมาณ 27 : 6 : 1 ตามลำดับ ใน TTC-L2-3 มีสัดส่วน 20 : 5 : 1 และใน TTC-L4 มีสัดส่วน 16.5 : 4 : 1 (โดยจะพิจารณาเฉพาะคำ 3 พยางค์ ไม่รวมชื่อเฉพาะประเภทชื่อคน) นั่นคือในแง่ของรูปศัพท์ คำ 2 พยางค์ > คำ 1 พยางค์ > คำ 3 พยางค์เสมอ ในขณะที่จำนวนรูปปรากฏ คำ 1 พยางค์ > คำ 2 พยางค์ > คำ 3 พยางค์เสมอเช่นกัน ซึ่งจะนำผลส่วนนี้ไปอภิปรายต่อไป นอกจากนี้ หากคิดว่าข้อมูลภาษาที่ยังมีคำหลายพยางค์เป็นจำนวนมากจะยิ่งมีความยากมากขึ้น ข้อมูลภาษาในระดับขั้นที่สูงขึ้นก็ควรจะมีจำนวนรูปปรากฏของคำหลายพยางค์มากยิ่งขึ้น ในที่นี้ ผู้วิจัยจึงได้นำค่ารูปปรากฏของคำที่มีจำนวนพยางค์ 1 2 และ 3 พยางค์ที่นับได้ในแต่ละช่วงชั้นดังแสดงในตารางที่ 2 มาทดสอบไคสแควร์ และพบว่า ช่วงชั้นที่ต่างกันมีผลต่อการปรากฏของจำนวนพยางค์ของคำในข้อมูลที่ระดับนัยสำคัญ 0.01 อีกด้วย โดยได้ค่า $\chi^2 = 4,299.48$ ซึ่งมากกว่าค่าวิกฤตของ $\chi^2_{0.01, 6} = 16.81$

ตารางที่ 2: จำนวนพยางค์ของแต่ละคำ

Table 2: The Number of Syllables in a Word

คลังย่อย	รูปศัพท์			รูปปรากฏ		
	1 พยางค์	2 พยางค์	3 พยางค์	1 พยางค์	2 พยางค์	3 พยางค์
TTC-L1	3,335 30.13%	5,731 51.78%	2,002 18.09%	353,494 79.61%	77,581 17.47%	12,969 2.92%
TTC-L2	3,742 27.56%	6,976 51.37%	2,861 21.07%	546,386 77.58%	131,370 18.65%	26,533 3.77%
TTC-L3	3,740 28.19%	6,652 50.14%	2,874 21.66%	655,629 76.94%	163,806 19.22%	32,664 3.83%
TTC-L4	3,618 27.27%	6,595 49.71%	3,053 23.01%	781,663 75.35%	208,259 20.08%	47,418 4.57%

ที่มา: จากการวิเคราะห์ข้อมูล

Source: From data analysis

ผลที่ได้ในที่นี้ จึงยืนยันสมมติฐานที่ตั้งไว้ว่าช่วงชั้นที่สูงขึ้นจะปรากฏคำศัพท์หลายพยางค์มากกว่าช่วงชั้นที่ต่ำลงมา นอกจากนี้ หากมองในแง่ของรูปศัพท์ก็พบว่าจำนวนรูปศัพท์ในช่วงชั้นที่สูงขึ้นก็มีจำนวนศัพท์ของคำ 3 พยางค์เพิ่มสูงมากขึ้นด้วยเช่นกัน กล่าวคือ จำนวนรูปศัพท์ของคำ 3 พยางค์จากช่วงชั้นที่ 1 ไปยังช่วงชั้นที่ 2 เพิ่มจำนวนขึ้นอย่างเห็นได้ชัด ในขณะที่ช่วงชั้นที่สูงขึ้นอื่น ๆ ปรากฏรูปศัพท์ 3 พยางค์เพิ่มขึ้นบ้างแต่เป็นจำนวนไม่มากนัก ทำให้เกิดข้อสังเกตว่าช่วงชั้นที่ 2 เป็นช่วงชั้นที่มีการขยายตัวทางด้าน

จำนวนคำศัพท์มากที่สุดเมื่อเทียบกับช่วงชั้นอื่น ๆ แสดงถึงวงศ์พท์ที่กว้างขึ้น รายละเอียดของจำนวนและร้อยละที่เพิ่มขึ้นจะกล่าวรวมไว้ในช่วงท้ายของหัวข้อนี้ อย่างไรก็ตาม เมื่อพิจารณาตัวเลขในส่วนร้อยละของรูปปรากฏจากตารางที่ 2 จะเห็นได้ว่าคำ 3 พยางค์มีจำนวนการปรากฏเพิ่มขึ้นตามช่วงชั้นที่สูงขึ้นจริง โดยเพิ่มจำนวนจากช่วงชั้นที่ 1 ไปยังช่วงชั้นที่ 2 ร้อยละ 0.85 จากช่วงชั้นที่ 2 ไปยังช่วงชั้นที่ 3 ร้อยละ 0.06 และจากช่วงชั้นที่ 3 ไปยังช่วงชั้นที่ 4 ร้อยละ 0.74 จะเห็นได้ว่ารูปปรากฏของคำ 3 พยางค์ที่เพิ่มขึ้นในแต่ละช่วงชั้นนั้นมีปริมาณไม่มากนักเมื่อพิจารณาจากค่าร้อยละ ซึ่งหมายความว่าหากความยากของตัวบทแปรผันตรงกับจำนวนรูปปรากฏของคำ 3 พยางค์ ก็จะตีความได้ว่าตัวบทในช่วงชั้นที่ 2 ยากกว่าช่วงชั้นที่ 1 และช่วงชั้นที่ 4 ยากกว่าช่วงชั้นที่ 3 ส่วนตัวบทในช่วงชั้นที่ 2 และ 3 ยากพอ ๆ กัน เพราะมีจำนวนคำ 3 พยางค์ใกล้เคียงกัน

ประเด็นความทึบใสทางความหมาย บทความขึ้นนี้แบ่งระดับความสัมพันธ์ทางความหมายออกเป็น 4 กลุ่ม คือ SEM 0-3 ซึ่งหมายถึงคำเดี่ยว คำใส่ได้ระดับไปจนถึงคำทึบ จากข้อมูลคำพท์ใน TTC พบว่าจำนวนรูปศัพท์และรูปปรากฏของคำที่เป็นคำเดี่ยว คำใส่ได้ คำก้ำกึ่ง และคำทึบ มีสัดส่วนเหมือนกันทุกช่วงชั้น นั่นคือ คำเดี่ยว > คำใส่ได้ > คำก้ำกึ่ง > คำทึบ กล่าวคือ SEM 0 หรือคำเดี่ยวและในที่นี้ต้องรวมคำทับศัพท์ภาษาอังกฤษไว้ในกลุ่มนี้ด้วย³ มีรูปศัพท์และรูปปรากฏสูงสุดในคลังข้อมูล และจากสัดส่วนของรูปปรากฏในตารางที่ 3 จะเห็นได้ว่าคำในกลุ่มนี้เป็นประชากรส่วนใหญ่ราวร้อยละ 87 ของจำนวนรูปปรากฏทั้งหมด ในขณะที่คำในกลุ่มคำประกอบมีจำนวนรวมกันเพียงร้อยละ 13 ในจำนวนนี้เป็นคำประเภทคำใส่ได้มากที่สุด รองลงมาคือคำที่ยังก้ำกึ่ง ไม่ทึบ ไม่ใส่ และคำทึบปรากฏน้อยที่สุด เช่นเดียวกับกรณีคำหลายพยางค์ ในที่นี้หากข้อมูลภาษามีคำทึบเป็นจำนวนมากก็ยิ่งแสดงให้เห็นว่าตัวบทนั้นมีความยาก ข้อมูลภาษาในช่วงชั้นที่สูงขึ้นก็น่าจะมีปริมาณรูปปรากฏของคำทึบมากขึ้น ผู้วิจัยจึงได้นำค่ารูปปรากฏของ SEM 0-3 ที่นับได้ในแต่ละช่วงชั้นดังที่แสดงในตารางที่ 3 มาทดสอบไคสแควร์ และพบว่ามีความเกี่ยวข้องสัมพันธ์กันระหว่างช่วงชั้นที่เปลี่ยนไปกับจำนวนคำที่มีความทึบใสทางความหมายแตกต่างกันที่ระดับนัยสำคัญ 0.01 โดย พบว่าค่า $\chi^2 = 1,591.66$ ซึ่งมากกว่าค่าวิกฤตของ $\chi^2_{0.01,9} = 21.67$

³ เนื่องจากคำทับศัพท์ภาษาอังกฤษมีปรากฏอยู่ไม่น้อย รวมถึงอาจมองได้ว่ามีการรับคำพท์ในกลุ่มนี้เข้ามาทั้งคำจึงส่งผลให้ลักษณะของคำพท์สอดคล้องกับคำเดี่ยวตามการนิยามในงานวิจัยชิ้นนี้

ตารางที่ 3: จำนวนคำที่ระดับความสัมพันธ์ทางความหมายแบบต่าง ๆ

Table 3: Degree of Semantic Transparency of a Word

คลังย่อย	รูปศัพท์				รูปปรากฏ			
	SEM0	SEM1	SEM2	SEM3	SEM0	SEM1	SEM2	SEM3
TTC-L1	5,848 52.84%	4,086 36.92%	639 5.77%	495 4.47%	390,734 87.99%	45,653 10.28%	4,656 1.05%	3,001 0.68%
TTC-L2	6,878 50.65%	5,180 38.15%	847 6.24%	674 4.96%	612,739 87.00%	77,264 10.97%	9,031 1.28%	5,255 0.75%
TTC-L3	6,895 51.97%	4,961 37.40%	823 6.20%	587 4.42%	740,998 86.96%	93,938 11.02%	10,538 1.24%	6,625 0.78%
TTC-L4	6,716 50.63%	5,161 38.90%	823 6.20%	566 4.27%	890,322 85.83%	125,351 12.08%	14,188 1.37%	7,479 0.72%

ที่มา: จากการวิเคราะห์ข้อมูล

Source: From data analysis

จากผลการศึกษาข้างต้น เมื่อนำมาเชื่อมโยงกับแนวคิดที่ว่าช่วงชั้นที่ต่างกันมีผลต่อความต่างเรื่องความทึบใสทางความหมายของคำ แม้ว่าคลังข้อมูลย่อยทั้งสี่จะมีสัดส่วนการปรากฏของคำที่คำใสที่ใกล้เคียงกัน แต่จากค่าสถิติไคสแควร์ข้างต้นก็แสดงให้เห็นว่าทั้งสองปัจจัยนี้มีความเกี่ยวข้องสัมพันธ์กันโดยมีความต่างชัดเจนจากช่วงชั้นที่ 1 ต่อไปยังช่วงชั้นที่ 2 เพราะมีการเพิ่มขึ้นของจำนวนคำศัพท์มากขึ้นอย่างชัดเจน และเมื่อพิจารณาร้อยละของรูปศัพท์และรูปปรากฏของ SEM2 และ SEM3 ในช่วงชั้นที่ 1 ก็จะเห็นว่าการเพิ่มของสัดส่วนในช่วงชั้นที่ 2 ชัดเจนกว่าช่วงชั้นอื่นในส่วนนี้ผู้วิจัยจะได้นำไปอภิปรายต่อไป

ความยากง่ายของคำศัพท์ในแง่ความหมายและการใช้

ในทางการเรียนการสอนภาษามักให้ความสำคัญกับคำศัพท์ที่ใช้บ่อยเป็นลำดับแรก ๆ และในบรรดาคำศัพท์ที่ใช้บ่อยนี้ จำนวนมากเป็นคำที่มีการขยายความหมายและรูปแบบการใช้หลากหลายซึ่งเป็นอีกมิติหนึ่งที่สะท้อนถึงความยากง่ายของศัพท์ได้ การศึกษาคำปรากฏรวมหรือปริบทแวดล้อมของคำศัพท์ต่าง ๆ จะช่วยให้เข้าใจลักษณะความหมายต่าง ๆ ของคำนั้น ๆ ได้ดีขึ้นอีกด้วย

รายการคำที่ได้มาจากคลังข้อมูลภาษาต่าง ๆ ที่เคยสร้างไว้ นั้น พบว่าคำที่ปรากฏด้วยความถี่สูงส่วนใหญ่เป็นคำไวยากรณ์แทบทั้งสิ้น ยกตัวอย่างเช่นคำศัพท์ 5 คำแรกใน The instant words ได้แก่ *the, of, and, a, to* ตามลำดับ (Fry, 1980) หรือตัวอย่าง 5 คำแรกจาก General Service List (GSL) ได้แก่ *the, be, of, and, a* (Bauman) ส่วนคำศัพท์ภาษาไทยที่พบบ่อยที่สุด 5 อันดับแรกอ้างอิงจากคลังข้อมูลภาษาไทยแห่งชาติ (TNC) ได้แก่ *ที่ การ เป็น ใน จะ* ในขณะที่รายการคำที่ปรากฏเป็นอันดับต้น ๆ จาก TTC ซึ่งใช้ในการศึกษาครั้งนี้ อยู่ในตารางที่ 4

ตารางที่ 4: รายการคำศัพท์ที่ปรากฏบ่อยเป็นลำดับต้น ๆ

Table 4: High Frequency Word List

ลำดับที่	คำศัพท์	per million words
1	ที่	20,780
2	จะ	17,013
3	เป็น	16,878
4	ไป	15,329
5	ก็	15,277
6	ไม่	15,108
7	มี	14,780
8	ได้	14,653
9	ว่า	13,263
10	ให้	12,745
11	มา	12,318

ที่มา: จากการรวบรวมข้อมูลและคำนวณ

Source: From data collection and calculation

จากตัวอย่างข้างต้น มีคำศัพท์อยู่จำนวนไม่น้อยที่ขยายขอบเขตการใช้จากคำเนื้อหาไปสู่คำไวยากรณ์ และคำเหล่านี้ก็ปรากฏด้วยความถี่สูงอีกด้วย เช่น คำว่า ไป และ มา (ใน TNC ปรากฏในลำดับที่ 13 และ 14 ตามลำดับ) หลายคนอาจมีความเห็นว่าทั้งสองคำนี้ก็มีความหมายชัดเจนในตัวเองอยู่แล้ว นั่นคือ ไป หมายถึง การเคลื่อนที่ออกจากตัวผู้พูด ในขณะที่ มา หมายถึง การเคลื่อนที่เข้าหาตัวผู้พูด โดยที่ความหมายดังกล่าวจะเป็นลักษณะความหมายของ ไป และ มา ในฐานะคำกริยาหลัก ทว่าในการใช้ภาษาจริงยังพบว่า ไป และ มา สามารถปรากฏในตำแหน่งคำกริยานำและกริยารองเพื่อช่วยเสริมความหมายของคำกริยาหลัก รวมถึง ไป และ มา ที่เป็นส่วนหนึ่งของสำนวนโวหารด้วย จึงจะเห็นได้ว่าคู่คำนี้มีการใช้และความหมายที่หลากหลายไม่น้อย ยิ่งไปกว่านั้นยังน่าสนใจอีกเช่นกันว่าแม้คำกริยาที่พบได้บ่อยและมีความถี่ในการปรากฏสูงจะเป็นคำที่ง่าย แต่ก็ถือได้ว่าค่อนข้างท้าทายในการเรียนรู้ภาษา เพราะในการใช้จริงอาจไม่ง่ายอย่างที่คิด (Gouverneur, 2008: 223) จึงน่าสนใจว่าคู่คำ ไป-มา จะมีรูปแบบการปรากฏที่หลากหลายขึ้นและสะท้อนความยากง่ายเพิ่มขึ้นหรือไม่อย่างไร

เมื่อพิจารณาคำว่า ไป และ มา ในแต่ละคลังข้อมูลย่อย พบว่า ไป ปรากฏบ่อยกว่า มา ในทุกช่วงชั้น รวมถึงทั้ง ไป และ มา มีแนวโน้มจะปรากฏน้อยลงตามช่วงชั้นที่สูงขึ้น

ตารางที่ 5: ลำดับของ ไป และ มา ในคลังข้อมูลย่อย

Table 5: Rank of ‘go’ (/paj/) and ‘come’ (/maa/) in Subcorpora

ลำดับที่	TTC-L1		TTC-L2		TTC-L3		TTC-L4	
	คำ	per mil.	คำ	per mil.	คำ	per mil.	คำ	per mil.
1	ไป	17228	ที่	18176	ที่	21336	ที่	24001
2	ที่	16323	ไป	16580	เป็น	17836	จะ	18248
3	ก็	14728	จะ	15780	จะ	17814	เป็น	18226
4	จะ	14546	เป็น	15468	ก็	16687	มี	16510
5	เป็น	14129	ก็	15332	ไป	15982	ไม่	16035
6	ว่า	14075	ได้	14324	ไม่	15562	ได้	15595
7	ไม่	14039	ให้	13974	ได้	15010	ก็	14317
8	มี	13913	ไม่	13868	มี	14376	ใน	14219
9	มา	13771	ว่า	13733	ว่า	13031	ของ	13659
10	ให้	13665	มา	13529	มา	12643	ไป	13132
11	ได้	12292	มี	13269	ให้	12293	ว่า	12787
12	และ	12024	และ	11779	ใน	11929	ความ	12294
13	ของ	10411	ของ	10905	ของ	11681	การ	12192
14	แล้ว	9294	แล้ว	9553	และ	9965	และ	12010
15	ใน	8785	ใน	9033	แล้ว	9499	ให้	11886
16	อยู่	8765	อยู่	7968	อยู่	8792	มา	10606
17	คน	7887	คน	7859	ความ	8701	คน	8294
18	กัน	7704	ความ	7447	การ	8180	อยู่	8080
19	แม่	6767	กัน	7037	กัน	6680	แล้ว	7535
20	จึง	5999	จึง	6492	คน	6554	กัน	6902

ที่มา: จากการรวบรวมข้อมูลและคำนวณ

Source: From data collection and calculation

จากการสุ่มวิเคราะห์ตัวอย่างจากแต่ละคลังข้อมูลย่อย คลังละ 100 ตัวอย่าง รวมเป็น ไป 400 ตัวอย่าง และ มา 400 ตัวอย่าง พบลักษณะทางความหมายดังนี้

- (1) การเคลื่อนที่เข้าหรือออกจากตัวผู้พูด ปรากฏในตำแหน่งกริยาหลักและเป็นกริยาเดียว เช่น
 - แล้วก็เฉยเสียไม่ไป <NE>ธิดา</NE>จึงคิดว่า (TTC-L1)
 - ยายมายายจะด่า (TTC-L1)
 - ขาดเธอคนเดียว นอกนั้นไปทุกคน (TTC-L2)
 - ไม่แพร้งพรายให้ใครรู้ว่าเรามาที่นี้ละ (TTC-L2)
 - ไปสวรรค์แล้ว (TTC-L3)
 - พิมพ์ว่า “เฮ้! มาจากใคร” แล้วก็หยิบขึ้นอ่าน (TTC-L3)
 - จะไม่ไปก็ไม่กล้า (TTC-L4)
 - ไม่รู้หรือกว่าคนแถวพวกนี้มาจากไหน (TTC-L4)
- (2) เสริมความหมายเกี่ยวกับทิศทางในการเคลื่อนที่ทางกายภาพ ปรากฏนำหน้าหรือตามหลังคำกริยา 1 คำหรือมากกว่า 1 คำก็ได้ เช่น
 - หุ่นยนต์ยี้มเล็กน้อยแล้วหันไปริมน้ำ (TTC-L1)
 - ไม่ให้ศัตรูมาแย่งเอาไป (TTC-L1)
 - <NE>มาลี</NE>จึงไปเก็บดอกพิกุลทุก ๆ คื่น (TTC-L2)
 - แปะจะดลเงินขยอกออกมาให้ (TTC-L2)
 - แก่วนพูด แล้วก็ก้มหน้าเข้าครัวไปนั่งพลา่กึ่งด้วยความวิตก (TTC-L3)
 - แต่ยังไม่เคยมีใครมารับไปรักษาเลย (TTC-L3)
 - มันเดินไปตามถนน (TTC-L4)
 - เป็นหน้าที่ที่จะต้องมาอยู่ใกล้ชิดลูกสาวผม (TTC-L4)
- (3) เสริมความหมายด้านอื่นนอกเหนือจากการเคลื่อนที่ทางกายภาพ ปรากฏนำหน้าหรือตามหลังคำกริยา 1 คำหรือมากกว่า 1 คำก็ได้ เช่น
 - ดินของมัน จึงเจ็บไปอีก (TTC-L1)
 - นิทานอีสป เป็นนิทานที่มีมานานเป็นร้อย ๆ ปี (TTC-L1)
 - ถั่วงาฉิบหายวอดไปแล้ว (TTC-L2)
 - เหมือนอย่างที่ผมคุ้นเคยมาแต่เล็ก (TTC-L2)
 - จะไปขับขี่ซ้อนมิใช้วิสัยผู้มีปัญญา (TTC-L3)
 - ที่ได้กราบทูลมาข้างต้นนั้น (TTC-L3)
 - พอปลุกกันไปไม่เท่าไร ก็เผลอตายหมด (TTC-L4)
 - ครบชุดเมื่อ 2 เดือนมานี้เอง (TTC-L4)

(4) ปรากฏอยู่ในสำนวน พบ 2 ตัวอย่าง จาก TTC-L4 คือ

- ไม่เคยไปมาหาสู่กันที่บ้าน (TTC-L4)
- เดินทางไปมาค้าขายอยู่คับคั่ง (TTC-L4)

สัดส่วนความหมายแต่ละแบบของคำ ไป และ มา ในแต่ละคลังข้อมูลย่อย แสดงในตารางที่ 6

ตารางที่ 6: ความหมายของ ไป และ มา

Table 6: The Meanings of 'go' (/paj/) and 'come' (/maa/)

คลัง ข้อมูล ย่อย	ไป					มา				
	กริยา เดี่ยว	เสริม- ทิศทาง	เสริม- ไม่ใช่ ทิศทาง	สำนวน	รวม	กริยา เดี่ยว	เสริม- ทิศทาง	เสริม- ไม่ใช่ ทิศทาง	สำนวน	รวม
TTC-L1	9	73	18	0	100	15	69	16	0	100
TTC-L2	4	68	28	0	100	5	78	17	0	100
TTC-L3	6	70	24	0	100	11	74	15	0	100
TTC-L4	12	55	33	0	100	5	52	41	2	100
รวม	31	266	103	0	400	36	273	89	2	400

ที่มา: จากการวิเคราะห์ข้อมูล

Source: From data analysis

จะเห็นได้ว่าความหมายในรูปแบบต่าง ๆ ของ ไป และ มา ปรากฏในสัดส่วนที่คล้ายคลึงกัน นั่นคือ กริยาเดี่ยว : เสริมความหมายด้านทิศทาง : เสริมความหมายที่ไม่ใช่ทิศทาง มีสัดส่วนโดยประมาณเป็น 1 : 7 : 2 ตามลำดับ กล่าวคือ ความหมายที่ปรากฏมากที่สุดของสองคำนี้จากทุกช่วงชั้นไม่ใช่ความหมายในฐานะ คำกริยาเดี่ยว หากเป็นความหมายที่ช่วยเสริมเกี่ยวกับทิศทาง รองลงมาเป็นความหมายเสริมที่ไม่เกี่ยวข้อง กับทิศทาง ส่วนการปรากฏเป็นส่วนหนึ่งของสำนวน จากการสุ่มวิเคราะห์ความหมายพบเพียง 2 ตัวอย่าง ตามที่ได้กล่าวไว้ข้างต้น เรียงลำดับความหมายตามการปรากฏมาน้อยได้ ดังนี้

เสริมความเกี่ยวกับทิศทาง > เสริมความไม่เกี่ยวกับทิศทาง > กริยาเดี่ยว > สำนวน

ประเด็นข้อคำถามที่ว่าคู่คำ ไป-มา จะมีรูปแบบการปรากฏที่หลากหลายขึ้นหรือไม่นั้น ผลจากการศึกษาไม่พบว่ามีหลากหลายเพิ่มขึ้นแต่อย่างใด ข้อมูลที่ได้จากแต่ละช่วงชั้นพบลักษณะทาง ความหมายครบทุกรูปแบบ ในสัดส่วนที่คล้ายคลึงกัน

ส่วนเรื่องการสะท้อนความยากง่ายเพิ่มขึ้นหรือไม่อย่างไรนั้น หากพิจารณาจากเกณฑ์ที่วางไว้ใน ตอนต้นว่า ความหมายในตำแหน่งของคำกริยาหลักที่หมายถึงการเคลื่อนที่นั้นจะเป็นความหมายพื้นฐานที่สุด

และง่ายที่สุด ความหมายในฐานะการเสริมความหมายเรื่องทิศทางในการเคลื่อนที่ที่น่าจะยากขึ้นมากอีกชั้น ส่วนการเสริมความหมายที่ขยับออกไปจากเรื่องทิศทางการเคลื่อนที่ที่น่าจะเข้าใจได้ยากขึ้นอีก และสำนวน น่าจะเป็นความหมายที่ยากที่สุด ความง่ายในส่วนนี้จากตัวอย่างที่สุ่มมาศึกษาพบเพียงว่าสำนวนปรากฏ เฉพาะใน TTC-L4 หรือช่วงชั้นมัธยมปลายเท่านั้น อย่างไรก็ตาม อาจยังไม่สามารถยืนยันได้ชัดเจนว่าจะ ไม่พบ ไป-มา ในฐานะส่วนหนึ่งของสำนวนจากช่วงชั้นอื่น ๆ เพราะมีความเป็นไปได้ที่จะมีข้อมูลอยู่ แต่บังเอิญสุ่มไม่พบในช่วงชั้นที่เหลือ ดังนั้น จึงควรพิจารณาลักษณะความหมายเฉพาะ 3 ส่วนแรกก่อน เพราะมีจำนวนปรากฏมากในระดับหนึ่ง ซึ่งจากตัวเลขที่ได้จะเห็นว่าจำนวนตัวอย่างที่แสดงความหมาย เสริมเรื่องทิศทางการเคลื่อนที่มีจำนวนมากพอ ๆ กันในแต่ละช่วงชั้น จนแทบจะแยกแยะความยากง่ายไม่ได้ ซึ่งหากลักษณะความหมายดังกล่าวสามารถสะท้อนความยากง่ายได้แล้ว ความหมายที่ยากที่สุดคือการเสริม ความหมายที่ไม่เกี่ยวข้องกับทิศทางควรพบเฉพาะในช่วงชั้นที่สูงเท่านั้น แต่จากผลการศึกษาพบลักษณะ ความหมายทุกรูปแบบจากทุกช่วงชั้น แสดงว่านักเรียนในทุกระดับชั้นเข้าใจและเรียนรู้ความหมายทั้งหมด แล้วตั้งแต่ในระดับชั้นประถมต้น คำว่า ไป-มา ซึ่งเป็นคำพื้นฐานที่พบบ่อยนี้จึงไม่สามารถนำมาใช้สะท้อน ความยากง่ายผ่านทางความหลากหลายของความหมายได้ นอกจากนี้ ผลการศึกษาในส่วนนี้ยังทำให้เห็นว่า ความหมายที่ใช้บ่อยจริง ๆ อาจไม่จำเป็นต้องเป็นความหมายพื้นฐานเสมอไปก็ได้

สรุปผลการศึกษา (Conclusions)

จากการศึกษาคำศัพท์ที่ปรากฏในแบบเรียนภาษาไทยตั้งแต่ระดับเริ่มเรียน-มัธยมศึกษาปีที่ 6 โดยอาศัยข้อมูลจากคลังข้อมูลขนาด 3 ล้านคำ ซึ่งแบ่งออกได้เป็น 4 ช่วงชั้นหรือ 4 คลังข้อมูลย่อย (TTC-L1-4) พบว่ามีคำศัพท์ที่เกิดร่วมกันทั้ง 4 คลังข้อมูลเป็นจำนวน 7,322 คำ เมื่อพิจารณาเปรียบเทียบ ขึ้นไปที่ระดับพบว่าคำศัพท์ที่ปรากฏในช่วงชั้นที่สูงกว่าไม่ได้ครอบคลุมคำศัพท์ทั้งหมดจากช่วงชั้นที่ต่ำ กว่า และไม่ได้มีจำนวนมากเท่าเสมอไป โดย TTC-L1 (ประถมต้น) และ TTC-L2 (ประถมปลาย) มีคำศัพท์ ที่เกิดร่วมกันมากที่สุด คิดเป็นร้อยละ 83.44 ของ TTC-L1 และร้อยละ 68.01 ของ TTC-L2

ส่วนสมมติฐานที่ว่าช่วงชั้นที่สูงขึ้นจะปรากฏคำศัพท์หลายพยางค์มากกว่าช่วงชั้นที่ต่ำลงมา พบว่า คำ 3 พยางค์ขึ้นไปมีจำนวนการปรากฏเพิ่มขึ้นตามช่วงชั้นที่สูงขึ้นจริง สมมติฐานในส่วนนี้จึงเป็นจริง ซึ่งหากเชื่อตามความคิดที่ว่าคำที่มีจำนวนพยางค์มาก จะยากกว่า ก็อาจสรุปได้ว่าข้อมูลที่พบในแบบเรียน สะท้อนให้เห็นถึงความยากที่เพิ่มขึ้นของคำศัพท์ในช่วงชั้นที่สูงขึ้น

ส่วนเรื่องความถี่ใสทางความหมาย พบว่า คำเดี่ยวและคำทับศัพท์ภาษาอังกฤษ มีรูปศัพท์และ รูปปรากฏสูงสุดในคลังข้อมูล จากสัดส่วนของรูปปรากฏ พบว่า คำในกลุ่มนี้เป็นประชากรส่วนใหญ่ราว ร้อยละ 87 ในขณะที่คำประกอบมีอยู่ราวร้อยละ 13 ซึ่งในจำนวนนี้คำที่พบได้น้อยที่สุด ผลการศึกษา พบว่าปริมาณของคำที่พบเพิ่มจำนวนรูปปรากฏมากขึ้นตามช่วงชั้นที่สูงขึ้นจริงบางส่วน โดยช่วงชั้นที่มีการ เพิ่มขึ้นของคำที่บ่อยชัดเจน คือ ช่วงชั้น TTC-L2

เมื่อพิจารณาความยากง่ายของวงศัพท์ที่เป็นอยู่ในแต่ละช่วงชั้นจากสองมิติข้างต้นแล้ว ผู้วิจัยยังให้ความสนใจ “ความหมายและการใช้” คำศัพท์เหล่านั้นในบทเรียนด้วยว่าจะแสดงถึงพัฒนาการและการเปลี่ยนแปลงทางความหมายอย่างไรได้บ้าง โดยเชื่อว่าคำศัพท์เดียวกันที่ปรากฏในช่วงชั้นที่สูงกว่าจะมีความหมายและการใช้ที่หลากหลายกว่าคำนั้น ๆ ที่ปรากฏในช่วงชั้นที่ต่ำกว่า ผู้วิจัยเลือกคำกริยา *ไป-มา* ซึ่งเป็นคำที่ปรากฏด้วยความถี่สูงมาเป็นตัวแทนของการศึกษาการขยายความหมายในมิตินี้

ลักษณะทางความหมายของ *ไป* และ *มา* เกี่ยวข้องกับการเคลื่อนที่ดังที่ได้กล่าวไว้ในหัวข้อก่อนหน้านี้แล้ว แบ่งออกได้เป็น 3 ความหมายหลัก เรียงจากง่ายไปหายากโดยพิจารณาจากความหมายพื้นฐานที่เข้าใจง่ายที่สุดไปสู่ความหมายที่ซับซ้อนห่างออกจากความหมายพื้นฐานมากที่สุด รวมถึงความซับซ้อนทางโครงสร้างประโยคได้ ดังนี้

- (1) การเคลื่อนที่เข้าหา/ออกห่างจากจุดอ้างอิง (*ไป-มา* เป็นกริยาเดียว)
- (2) แสดงถึงทิศทางการเคลื่อนที่ทางกายภาพ (*ไป-มา* เป็นกริยานำหรือกริยารอง)
- (3) แสดงการเคลื่อนที่ทางเวลา การเปลี่ยนสภาพ การณ์ลักษณะ แสดงถึงความเกินพอดี (*ไป-มา* เป็นกริยานำหรือกริยารอง)

พบ *ไป* และ *มา* เป็นส่วนหนึ่งของสำนวนด้วย แต่จากตัวอย่างที่สุ่มมาได้นั้น พบเพียง 2 ตัวอย่าง จึงเห็นว่าอาจน้อยเกินไปกว่าจะนำมาสรุปหรืออภิปรายต่อได้ ในส่วนของความหมายทั้งสามแบบนี้ พบว่าความหมาย (2) ปรากฏมากที่สุด รองลงมาเป็นความหมาย (3) ผลเป็นเช่นนี้เหมือนกันทุกช่วงชั้น ในส่วนนี้เองอาจพอสรุปถึงแนวโน้มในการใช้คำสองคำนี้ว่าเน้นใช้ในการเสริมความหมายทางด้านทิศทางการเคลื่อนที่ทางกายภาพมากกว่าใช้บอกการเคลื่อนที่ในฐานะคำกริยาหลัก และเพราะลักษณะทางความหมายที่พบไม่ได้แสดงความหลากหลายเพิ่มขึ้นแต่อย่างใด ทุกช่วงชั้นมีปรากฏใช้ทุกความหมาย ดังนั้นจึงไม่อาจสรุปได้ว่าลักษณะความหมายเช่นนี้สะท้อนถึงความยากง่ายของคำตามช่วงชั้นที่สูงขึ้น เพียงแสดงให้เห็นว่าความหมายที่ใช้บ่อยของคำหนึ่ง ๆ อาจไม่จำเป็นต้องเป็นความหมายพื้นฐานที่สุดก็ได้

อภิปรายผล (Discussion)

สิ่งที่อยู่นอกเหนือขอบเขตการศึกษาครั้งนี้แต่เป็นข้อค้นพบในระหว่างที่ศึกษาและผู้วิจัยเห็นว่าน่าสนใจ จึงจะนำมาอภิปรายไว้ในหัวข้อนี้ได้แก่ 1) เรื่องสัดส่วนจำนวนรูปศัพท์ต่อจำนวนรูปปรากฏในข้อมูล (Type-Token ratios หรือ TTR) 2) ช่วงชั้นที่มีความหลากหลายทางรูปศัพท์สูงสุด 3) ลักษณะพยางค์ของคำในภาษาไทย และ 4) ลักษณะทางความหมายของคำในภาษาไทย อภิปรายทีละประเด็นได้ ดังนี้

TTR ถูกนำมาใช้เป็นตัววัดความหลากหลายของศัพท์ (Richards, 1987) โดยคาดว่ายิ่งค่าที่ได้จากการคำนวณสูงเท่าไร ก็ยิ่งแสดงให้เห็นถึงความหลากหลายของคำศัพท์มากเท่านั้น ค่าดังกล่าวสามารถคำนวณได้จาก

$$TTR = (Type/Token)*100$$

ผลปรากฏว่าค่า TTR ของช่วงชั้นที่ 1-4 เรียงตามลำดับคือ 2.49 1.93 1.56 และ 1.28 ซึ่งลดลงตามช่วงชั้นที่สูงขึ้น หรือความหลากหลายของคำศัพท์ลดลงตามช่วงชั้นที่สูงขึ้น แม้อาจจะดูแปลก แต่ผลที่ได้นี้ก็กลับสอดคล้องกับที่ Richards (1987) ได้อธิบายไว้ว่า เมื่อคำนวณค่า TTR จากข้อมูลขนาดใหญ่จะได้ค่าน้อยกว่าเมื่อคำนวณจากข้อมูลขนาดเล็ก เหตุที่เป็นเช่นนี้ เพราะเมื่อเพิ่มจำนวนคำในตัวบทไปเรื่อย ๆ จำนวนศัพท์ที่ใช้ในตัวบทก็จะเพิ่มมากขึ้น แต่เมื่อเพิ่มไปอย่างต่อเนื่อง จำนวนศัพท์ก็จะถูกใช้ไปจนเกือบหมดวงศัพท์ที่มี ทำให้เมื่อเพิ่มจำนวนข้อมูลหรือคำในตัวบทก็จะยิ่งทำให้ค่าที่คำนวณได้ลดลงเรื่อย ๆ ค่าที่ได้ดังกล่าวจึงอาจไม่สะท้อนความหลากหลายของศัพท์ได้อย่างที่คาดการณ์ไว้แต่แรก อย่างไรก็ตาม ผู้วิจัยมีความเห็นว่าหากปรับคลังข้อมูลย่อยจากแต่ละระดับให้มีขนาดเท่ากัน นั่นคือ ทุกคลังข้อมูลย่อยมีจำนวนรูปปรากฏเท่ากันหรือใกล้เคียงกัน ค่า TTR ที่ได้อาจแตกต่างออกไป จากการวิจัยครั้งนี้จึงได้ข้อสังเกตว่าไม่ควรนำค่า TTR ที่คำนวณจากคลังข้อมูลที่มีขนาดไม่เท่ากันมาเปรียบเทียบกัน เนื่องจากไม่สามารถเปรียบเทียบความหลากหลายของคำศัพท์ระหว่างกลุ่มต่าง ๆ ได้จริง

ประเด็นเรื่องช่วงชั้นที่มีความหลากหลายทางรูปศัพท์สูงสุดนั้น เมื่อพิจารณาคำศัพท์ที่เกิดร่วมกันใน 2 หรือ 3 คลังข้อมูลย่อยแล้ว พบว่า ใน TTC-L2 มีคำศัพท์กลุ่มนี้มากกว่าคลังข้อมูลย่อยอื่น ๆ นอกจากนี้จำนวนรูปศัพท์ยังมากที่สุดด้วยคืออยู่ที่ 13,579 รูปศัพท์ และเมื่อพิจารณาเฉพาะคำศัพท์ที่เพิ่มขึ้นในแต่ละคลังข้อมูลย่อยก็ยังพบอีกว่าคำศัพท์เพิ่มขึ้นใน TTC-L2 มากกว่าที่เพิ่มขึ้นในช่วงชั้นอื่น ๆ แม้ว่าสาเหตุหนึ่งที่รูปศัพท์ในช่วงชั้นที่ 2 เพิ่มปริมาณมากขึ้นกว่าช่วงชั้นอื่น ๆ อาจมาจากที่ปริมาณรูปศัพท์ตั้งต้นในช่วงชั้นที่ 1 มีน้อย แต่ด้วยสัดส่วนที่มากกว่าราว 3 เท่า เมื่อเปรียบเทียบกับรูปศัพท์ที่เพิ่มขึ้นในช่วงชั้นอื่น ๆ จึงน่าจะเป็นไปได้ว่าช่วงชั้นที่ 2 เป็นช่วงชั้นที่แสดงถึงพัฒนาการในทางรูปศัพท์แบบก้าวกระโดด หรือเมื่อพิจารณาเฉพาะส่วนที่เป็น “คำยาก” จากการวิเคราะห์ในมิติจำนวนพยางค์และความซับซ้อนทางความหมาย ข้างต้นก็ยังให้ผลที่ยืนยันไปในทิศทางเดียวกัน กล่าวคือ ทั้งคำที่มี 3 พยางค์ขึ้นไปและคำที่มีความหมายซับซ้อนเพิ่มจำนวนมากที่สุด ใน TTC-L2 หรือช่วงชั้นประถมปลาย โดยพบว่า คำที่มีตั้งแต่ 3 พยางค์ขึ้นไปใน TTC-L1 มีจำนวนตั้งต้นที่ 2,002 รูปศัพท์นั้น

เพิ่มขึ้นใน	TTC-L2	จำนวน	1,394	รูปศัพท์	คิดเป็นร้อยละ	9.63
เพิ่มขึ้นใน	TTC-L3	จำนวน	798	รูปศัพท์	คิดเป็นร้อยละ	23.50
เพิ่มขึ้นใน	TTC-L4	จำนวน	746	รูปศัพท์	คิดเป็นร้อยละ	17.79

ในขณะที่คำทึบ (SEM 3) ใน TTC-L1 มีจำนวนตั้งต้นที่ 495 รูปศัพท์นั้น

เพิ่มขึ้นใน	TTC-L2	จำนวน	324	รูปศัพท์	คิดเป็นร้อยละ	65.45
เพิ่มขึ้นใน	TTC-L3	จำนวน	154	รูปศัพท์	คิดเป็นร้อยละ	18.80
เพิ่มขึ้นใน	TTC-L4	จำนวน	129	รูปศัพท์	คิดเป็นร้อยละ	13.26

จึงอาจกล่าวได้ว่าวงคำศัพท์ในช่วงชั้นประถมศึกษาตอนปลาย เป็นช่วงที่มีวงศัพท์กว้างที่สุด มีความหลากหลายสูงสุด ซึ่งอาจพอสรุปได้ว่าช่วงชั้นที่ 2 เป็นช่วงชั้นที่มีบทบาทสำคัญในการขยายตัวของ

วงศ์ศัพท์มากกว่าช่วงชั้นอื่น ๆ แสดงถึงพัฒนาการในเชิงจำนวนรูปศัพท์ได้อย่างชัดเจนที่สุด

การศึกษาเรื่องจำนวนพยางค์ พบว่า แม้รูปศัพท์ส่วนใหญ่ในแบบเรียนภาษาไทยจะเป็นคำที่มี 2 พยางค์ โดยมีจำนวนมากเป็น 2 เท่าของคำ 1 พยางค์ แต่คำ 1 พยางค์ก็มีรูปปรากฏมากที่สุด โดยมากถึงราว 4 เท่าเมื่อเทียบกับคำ 2 พยางค์ จึงอาจพออนุมานได้ว่าคำในแบบเรียนภาษาไทยส่วนใหญ่และอาจหมายรวมถึงคำในภาษาไทยที่ใช้ทั่วไปด้วย มี 2 พยางค์ แต่ในการเข้าประโยคจะพบคำพยางค์เดียวมากกว่า เนื่องจากเป็นคำที่มีหลายหน้าที่ หลายความหมายซึ่งประกอบด้วยคำไวยากรณ์เป็นจำนวนมาก จึงพบในบริบทการใช้ที่มากกว่า

ประเด็นสุดท้ายเรื่องลักษณะทางความหมายของคำในภาษาไทย ขอเริ่มต้นอภิปรายคำว่า *ไป* และ *มา* ซึ่งเป็นคำหลักในงานวิจัยชิ้นนี้ก่อน จากการศึกษาในภาพกว้างพบว่า *ไป* ปรากฏบ่อยกว่า *มา* ในทุก ๆ ช่วงชั้น แต่ลำดับ (rank) ของการปรากฏลดต่ำลงเรื่อย ๆ ตามช่วงชั้นที่สูงขึ้น ในขณะที่คำศัพท์คำอื่น เช่น *ที่* *จะ* *ก็* *ไม่* ฯลฯ เข้ามามีบทบาทมากกว่า *ไป* และ *มา* ข้อแรกแสดงให้เห็นว่า *ไป* มีบริบทในการปรากฏที่หลากหลายกว่า *มา* ทำให้จำนวนของ *ไป* สูงกว่า *มา* นอกเหนือจากนั้น ลำดับของรายการคำศัพท์ที่ปรากฏยังแสดงให้เห็นอีกด้วยว่า คำที่มีความเป็น “คำไวยากรณ์” มากกว่าเริ่มมีปรากฏใช้มากขึ้น เนื่องจาก *ไป-มา* มีความหมายในฐานะที่เป็นคำหลักสูง คือแสดงถึงการเคลื่อนที่บางอย่าง แต่ช่วงชั้นที่สูงขึ้นจะเริ่มปรากฏบทอ่านที่เป็นนามธรรมและเป็นวิชาการมากขึ้น *ไป-มา* ที่มีความหมายถึงการเคลื่อนที่ในเชิงรูปธรรมจึงลดจำนวนลง ในขณะที่ *ที่* *จะ* *ก็* *ไม่* เพิ่มจำนวนมากขึ้น

ส่วนความหมายและการใช้คำนั้น ในที่นี้อภิปรายได้ว่าแม้ว่าความหมายของ *ไป* และ *มา* ในฐานะคำกริยาหลักจะเป็นความหมายพื้นฐานที่สุด เข้าใจได้ง่ายที่สุด แต่ก็ยังเป็นลักษณะทางความหมายที่ผู้ใช้ภาษาหรือในที่นี่คือนักเรียน รับรู้ได้ตั้งแต่ก่อนจะถึงวัยเรียน จึงอาจไม่จำเป็นต้องสอน *ไป-มา* ในความหมายนี้มากนัก ในขณะที่ความหมายแสดงถึงทิศทางการเคลื่อนที่ทางกายภาพในฐานะกริยานำหรือกริยารองที่ปรากฏว่าใช้มากที่สุดนั้น แม้โครงสร้างประโยคอาจจะซับซ้อนขึ้นเพราะมีคำกริยาหลายตัวเรียงต่อกัน แต่ความหมายยังเข้าใจได้ไม่ยาก เพราะยังเกี่ยวข้องกับทิศทางการเคลื่อนที่เช่นเดียวกับคำกริยาหลัก ความซับซ้อนทางโครงสร้างที่เกิดขึ้นจึงอาจไม่ได้ส่งผลต่อการสร้างบทอ่านในแบบเรียนของช่วงชั้นต่าง ๆ นักผู้เขียนแบบเรียนไม่ได้นำปัจจัยเรื่องความมาก-น้อยของกริยาในประโยคหนึ่ง ๆ มาเป็นเกณฑ์ประกอบการเขียนแบบเรียนในระดับต่าง ๆ แต่พิจารณาจากเนื้อเรื่องที่จะนำเสนอเป็นสำคัญ

หากจะพิสูจน์ว่าคำที่มีความหลากหลายทางความหมายมีความหมายที่ซับซ้อนขึ้นในระดับชั้นที่สูงขึ้นอาจจะต้องเลือกคำที่ไม่ใช่คำทั่วไปที่พบใช้บ่อย เพราะคำเหล่านี้ผู้เรียนอาจจะเรียนรู้ทุกความหมายครบถ้วนแล้วตั้งแต่เล็ก หรือหากจะวัดระดับความยากง่ายของภาษาไทยที่แปรไปตามระดับชั้นเรียนอาจดูได้จากเรื่องอื่น ๆ เช่น โครงสร้างทางวากยสัมพันธ์ ซึ่งเป็นเรื่องที่จะต้องศึกษาต่อไป

เอกสารอ้างอิง (References)

- กระทรวงศึกษาธิการ, กรมวิชาการ. (2532). รายงานผลการศึกษาคำพื้นฐานที่ใช้ในการเรียนการสอนภาษาไทย ระดับประถมศึกษา เล่มที่ 1. กรุงเทพฯ: ครูสภาลาดพร้าว.
- กิตติพงษ์ ทวีวงศ์. (2538). การศึกษาเปรียบเทียบระดับความยากง่ายต่อการอ่านของหนังสือพิมพ์บางกอกโพสต์และเดอะเนชั่น. นิเทศศาสตรมหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.
- เกรียงไกร วัฒนาวาสดี. (2544). หน่วยสร้างกริยาเรียง ลักษณะร่วมจำเพาะบริเวณภาษาเอเชียตะวันออกเฉียงใต้. *วารสารมนุษยศาสตร์*, 9, 94-109.
- เฉลิมลาภ ทองอาจ. (2557). ปัญหาอ่านไม่ออก เขียนไม่ได้: แนวทางแก้ไขอย่างเป็นรูปธรรม เข้าถึงเมื่อ 27 เมษายน 2559, จาก <http://www.oknation.net/blog/Chalermlahp/2014/11/02/entry-1>
- ชาภินี มณีนาถ. (2551). การศึกษาสมมูลภาพในการแปลคำกริยา “ไป” และ “มา” ในภาษาไทยเป็นภาษาเยอรมันในงานวรรณกรรมนวนิยายแปล. อักษรศาสตรมหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.
- นวรรณ พันธุเมธา. (Ed.) (2555). คลังคำ (6 ed.). กรุงเทพฯ: อมรินทร์.
- พจนานุกรมฉบับมติชน. (2547). กรุงเทพฯ: มติชน.
- เพียรศิริ วงศ์วิภานนท์. (2531). คำซ้อนในภาษาไทย ตอนที่ 1: ลักษณะพิเศษของคำซ้อน. *ภาษา*, 5(3), 15-28.
- ราชบัณฑิตยสถาน. (Ed.) (2556). พจนานุกรม ฉบับราชบัณฑิตยสถาน พ.ศ. 2554 (2 ed.). กรุงเทพฯ: ราชบัณฑิตยสถาน.
- Bauman, J. About the General Service List Retrieved 2015, from <http://jbauman.com/gsl.html>
- Carroll, J. B., Davies, P., & Richman, B. (1971). *Word Frequency Book*. New York: American Heritage Publishing.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Fry, E. (1980). The New Instant Word List. *The Reading Teacher*, 34(3), 284-289.
- Gouverneur, C. (2008). The Phraseological Patterns of High-Frequency Verbs in Advanced English for General Purposes: A Corpus-Driven Approach to EFL Textbook Analysis. In S. G. Fanny Meunier (Ed.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Loraine, S. (2008). Vocabulary Development Retrieved 25 June 2015, from http://www.superduperinc.com/handouts/pdf/149_VocabularyDevelopment.pdf
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. New York: Routledge.
- Nagy, W. E., & Anderson, R. C. (1984). How Many Words are There in Printed School English? *Reading Research Quarterly*, 19(3), 304-330.

- Richards, B. (1987). Type/Token Ratios: what do they really tell us? *Journal of Child Language*, 14(02), 201-209. doi: 10.1017/S0305000900012885
- West, M. (1953). *A General Service List of English Words*. London: Longman.
- Xue, G., & Nation, I. S. P. (1984). A University Word List. *Language Learning and Communication*, 3(2), 215-229.

Translated Thai References

- Bandhumetha, Navavan. (Ed.) (2012). *Khlang-Kham* (6 ed.). Bangkok: Amarin. (In Thai)
- Department of Curriculum and Instruction Development, Ministry of Education. (1989). *The Studying of Basic Words Using in Thai Language Teaching for the Elementary Education* (Vol. 1). Bangkok: Khurusapha Ladprao. (In Thai)
- Maneenavachai, Chadinee. (2008). *A Study of Equivalence in Translating the Verbs “Pay” and “Maa” in Thai into German in a Translated Novel. Master of Arts (Linguistics), Chulalongkorn University.* (In Thai)
- Matichon Dictionary. (2004). Bangkok: Matichon. (In Thai)
- Thavevong, Kittipong. (1995). *A Comparative Readability Study of the Bangkok Post and The Nation*. Master of Arts (Communication Arts), Chulalongkorn University. (In Thai)
- The Royal Institute. (Ed.) (2013). *The Royal Institute Dictionary B.E. 2554* (2 ed.). Bangkok: The Royal Institute. (In Thai)
- Thong-art, Chalermлах. (2014). Illiteracy problem: The concrete solving method Retrieved 27 April 2016, from <http://www.oknation.net/blog/Chalermлах/2014/11/02/entry-1> (In Thai)
- Vongvipanond, Peansiri. (1988). Semantic Doublets in Thai: Part 1: Special Characteristic of Semantic Doublets. *‘Bhasa’ Journal*, 5(3), 15-28. (In Thai)
- Wattanasawat, Kriangkrai. (2001). Serial Verb Constructions: the Specific Co-dominance of South-East Asia language Area. *Journal of Humanity*, 9, 94-109. (In Thai)