ค่าบ่งชี้ระดับการกลายเป็นศัพท์ของคำในภาษาไทย

ลีนา มะลูลีม[1]

Received: *September 20, 2022*
Revised: *December 14, 2022*
Accepted: *December 19, 2022*

บทคัดย่อ

งานวิจัยนี้ศึกษาวิธีการวัดระดับการกลายเป็นศัพท์ของหน่วยสร้าง โดยเสนอว่าระดับการกลายเป็นศัพท์สามารถวัดได้จากค่าบ่งชี้เชิงปริมาณทั้งสิ้นสามค่า ได้แก่ ค่าระยะเวลาตอบสนองซึ่งแสดงระดับการเปลี่ยนแปลงโครงสร้างหน่วยคำวากยสัมพันธ์ ระดับความหมายเชิงประกอบซึ่งแสดงระดับการเปลี่ยนแปลงความหมาย และความถี่การปรากฏซึ่งแสดงระดับการเปลี่ยนแปลงในเชิงการใช้ ผลการศึกษาแสดงให้เห็นว่าค่าบ่งชี้ทั้งสามสามารถจำแนกหน่วยสร้างที่มีระดับการกลายเป็นศัพท์ต่ำและสูงออกจากกันได้ และข้อค้นพบบ่งชี้ว่าการเปลี่ยนแปลงแง่มุมต่าง ๆ ในกระบวนการกลายเป็นศัพท์นั้นเกิดขึ้นในเวลาและอัตราเร็วแตกต่างกัน ดังนั้นการศึกษาระดับการกลายเป็นศัพท์ของหน่วยสร้างจึงไม่สามารถใช้ค่าใดค่าหนึ่งเพียงค่าเดียวได้แต่จำเป็นต้องศึกษาจากค่าบ่งชี้หลายค่าประกอบกัน

คำสำคัญ: การกลายเป็นศัพท์, ค่าระยะเวลาตอบสนอง, ระดับความหมายเชิงประกอบ, ความถี่การปรากฏ, คำประสม

[1] หน่วยงานผู้แต่ง: ภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
อีเมล: lynamaluleem@gmail.com

Lexicalization Indicators of Thai Words

Lena Maluleem[2]

Received: *September 20, 2022*
Revised: *December 14, 2022*
Accepted: *December 19, 2022*

Abstract

The present study proposes quantitative measurements for the degree of lexicalization. The study suggests that degree of lexicalization can be measured using three quantitative indicators including reaction time, which represents the degree of morphosyntactic change, semantic compositionality, which represents the degree of semantic change, and usage frequency, representing the degree of change in usage. The results show that all three indicators can distinguish between non-lexicalized and highly lexicalized constructions. In addition, the findings indicate that each aspect of change in lexicalization occurs at a varying rate and time. Consequently, the degree of lexicalization of a construction cannot be represented by a single indicator but must be determined using a combination of indicators.

Keywords: lexicalization, reaction time, semantic compositionality, usage frequency, compound words

[2] Affiliation: Department of Linguistics, Faculty of Arts, Chulalongkorn University
E-mail: lynamaluleem@gmail.com

1. ที่มาและความสำคัญของปัญหา

การกลายเป็นศัพท์ (lexicalization) ตามนิยามแบบกว้างหมายถึงกระบวนการที่ศัพท์ใหม่ถูกเพิ่มเข้าไปในคลังศัพท์ (mental lexicon) ของผู้พูดภาษา (Brinton & Traugott, 2005; Lipka, 1992; Lipka et al., 1994) จากมุมมองการเปลี่ยนแปลงภาษา การกลายเป็นศัพท์คือการเปลี่ยนแปลงของภาษาซึ่งหน่วยสร้างซับซ้อน (complex construction) ควบรวมกลายเป็นหน่วยสร้างที่มีความซับซ้อนน้อยลง (simpler construction) และมีลักษณะคล้ายคลึงกับคำเดี่ยว (monomorphemic word) มากขึ้น (Blank, 2001; Bussman, 1996) โดยกระบวนการดังกล่าวเกี่ยวข้องกับการเปลี่ยนแปลงทั้งทางโครงสร้างหน่วยคำวากยสัมพันธ์ ความหมาย เสียงและการเปลี่ยนแปลงความถี่การใช้

ในแง่หน่วยคำและวากยสัมพันธ์ หน่วยสร้างที่ผ่านกระบวนการกลายเป็นศัพท์มักมีการควบรวมของหน่วยคำจากโครงสร้างที่ซับซ้อนประกอบด้วยหลายหน่วยคำ กลายเป็นโครงสร้างเรียบง่ายที่ประกอบด้วยหน่วยคำเดียว ตัวอย่างเช่น วลี *all one* ในภาษาอังกฤษสมัยกลาง (Middle English) ซึ่งประกอบด้วยสองหน่วยคำ ควบรวมเป็นหน่วยศัพท์ *alone 'ตามลำพัง'* ซึ่งประกอบด้วยหน่วยคำเดียวในภาษาอังกฤษปัจจุบัน (Brinton & Traugott, 2005) หรือในภาษาไทย วลี *พรุก/พรูก + นี้* เกิดการหลอมรวมคำเป็นหน่วยศัพท์ใหม่ *พรุ่งนี้* ที่มีการเปลี่ยนแปลงเสียงร่วมด้วย (Vipas, 2014) ในแง่การเปลี่ยนแปลงความหมาย หน่วยสร้างที่ผ่านกระบวนการกลายเป็นศัพท์มักมีระดับความหมายเชิงประกอบ (compositionality) ลดลง หรือความหมายของทั้งหน่วยสร้างแตกต่างจากความหมายของแต่ละองค์ประกอบ เช่น *แกะดำ* มีความหมายว่า *'คนที่ทำอะไรผิดแผกไปจากกลุ่ม, ใช้ในทางไม่ดี'* ซึ่งไม่เกี่ยวข้องโดยตรงกับความหมายของ *แกะ* และ *ดำ หรือ ฟุ้งซ่าน* (Alisa, 2014) ซึ่งมีความหมายว่า *ไม่สงบ, ไม่นิ่ง, พล่านไป, คิดไปไม่หยุดนิ่ง ใช้แก่อาการของจิต* เป็นความหมายที่แตกต่างจากความหมายตรงตัวของคำว่า *ฟุ้ง* และ *ซ่าน* ซึ่งมีความหมายว่า *ปลิวหรือลอยกระจายไป* นอกจากนี้ในเชิงการใช้ หน่วยสร้างที่ผ่านการกลายเป็นศัพท์มักมีความถี่การใช้ (usage frequency) สูงขึ้นและผู้พูดมีความคุ้นเคย (familiarity) กับคำดังกล่าวมากขึ้น

จากการทบทวนวรรณกรรมพบว่าการศึกษาการกลายเป็นศัพท์ในภาษาต่าง ๆ มักอยู่ในรูปแบบของงานวิจัยเชิงคุณภาพที่บรรยายว่าในกระบวนการนี้มีการเปลี่ยนแปลงขึ้นกับหน่วยสร้างอย่างไรบ้าง การศึกษาการกลายเป็นศัพท์ในเชิงปริมาณและเชิงการทดลองยังคงมีจำกัดและยังไม่มีผู้เสนอวิธีการวัดระดับการกลายเป็นศัพท์ที่ให้ผลลัพธ์ในเชิงปริมาณอย่างชัดเจนว่าแต่ละคำมีระดับการกลายเป็นศัพท์มากน้อยเพียงใด ข้อจำกัดนี้ทำให้งานวิจัยที่ต้องการศึกษากระบวนการกลายเป็นศัพท์ในเชิงปริมาณมักเลือกใช้ปรากฏการณ์ใดปรากฏการณ์หนึ่งเท่านั้นในการบ่งชี้ระดับการกลายเป็นศัพท์ของหน่วยสร้าง ตัวอย่างเช่น ใช้ความถี่การปรากฏเพียงค่าเดียวบ่งชี้ระดับการกลายเป็นศัพท์ทั้งหมด งานวิจัยจึงสามารถนำเสนอภาพของกระบวนการกลายเป็นศัพท์ได้เพียงมิติเดียวทั้งที่การกลายเป็นศัพท์เป็นกระบวนการเปลี่ยนแปลงของภาษาที่ซับซ้อนและเกี่ยวข้องกับการเปลี่ยนแปลงในหลายแง่มุมมิใช่เพียงการเปลี่ยนแปลงความถี่การใช้เท่านั้น

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเสนอวิธีการวัดค่าบ่งชี้การกลายเป็นศัพท์ของคำในภาษาไทยโดยใช้ค่าบ่งชี้เชิงปริมาณทั้งหมดสามค่าร่วมกัน ได้แก่ ค่าระยะเวลาตอบสนอง (reaction time) ระดับความหมายเชิงประกอบ (semantic compositionality) และความถี่การปรากฏ (usage frequency) ซึ่งแสดงระดับการเปลี่ยนแปลงหน่วยคำวากยสัมพันธ์ความหมายและการเปลี่ยนแปลงเชิงการใช้ตามลำดับ งานวิจัยต้องการศึกษาว่าค่าทั้งสามสามารถบ่งชี้คำที่มีระดับกลายเป็นคำศัพท์มากและน้อยออกจากกันได้หรือไม่ รวมถึงศึกษาว่าค่าบ่งชี้ต่าง ๆ นั้นมีความสัมพันธ์กันในทิศทางใดและสามารถใช้เป็นค่าบ่งชี้การกลายเป็นศัพท์โดยลำพังหรือจำเป็นต้องใช้ร่วมกับค่าอื่น ๆ ผู้วิจัยมีสมมติฐานว่าค่าบ่งชี้การกลายเป็นศัพท์ทั้งสามสามารถจำแนกคำเดี่ยวซึ่งเป็นตัวแทนของหน่วยสร้างที่ผ่านการกลายเป็นศัพท์สูง (fully lexicalized construction) และวลีซึ่งเป็นตัวแทนของหน่วยสร้างที่มีระดับการกลายเป็นศัพท์ต่ำ (non-lexicalized construction) ออกจากกันได้อย่างมี

นัยสำคัญทางสถิติ และมีสมมติฐานว่าค่าบ่งชี้ทั้งสามมีความสัมพันธ์กัน โดยความถี่การปรากฏแปรผกผันกับค่าระยะเวลาตอบสนองและความหมายเชิงประกอบ (หน่วยสร้างที่มีความถี่สูง มีค่าระยะเวลาตอบสนองและระดับความหมายเชิงประกอบต่ำ) และค่าระยะเวลาตอบสนองแปรผกผันกับระดับความหมายเชิงประกอบ (หน่วยสร้างที่มีค่าระยะเวลาตอบสนองสูง มีความหมายเชิงประกอบสูงเช่นกัน)

2. ทบทวนวรรณกรรม

2.1 การกลายเป็นศัพท์ (lexicalization)

กระบวนการกลายเป็นศัพท์หมายถึงการเปลี่ยนแปลงของภาษาประเภทหนึ่งซึ่งหน่วยสร้างซับซ้อนควบรวมกลายเป็นหน่วยสร้างที่มีความซับซ้อนน้อยลงและมีลักษณะที่ใกล้เคียงกับคำเดี่ยวมากขึ้น[3] หน่วยสร้างประเภทต่าง ๆ ในภาษามีระดับการกลายเป็นศัพท์ต่างกัน วลีและประโยคเป็นตัวแทนของหน่วยสร้างที่มีระดับการกลายเป็นศัพท์ต่ำและคำเดี่ยวเป็นตัวแทนของหน่วยสร้างที่มีระดับการกลายเป็นศัพท์สูง ในขณะที่คำประสมมีระดับการกลายเป็นศัพท์ที่หลากหลาย คำประสมบางคำมีระดับการกลายเป็นศัพท์มากทำให้มีลักษณะคล้ายคลึงกับคำเดี่ยวในขณะที่คำประสมบางคำผ่านการกลายเป็นศัพท์น้อยและมีลักษณะคล้ายคลึงกับวลี ภาพที่ 1 แสดงตัวอย่างของหน่วยสร้างประเภทต่าง ๆ บนแนวต่อเนื่องของกระบวนการกลายเป็นศัพท์ (lexicalization continuum)

วลี/ประโยค ⟷ คำเดี่ยว

ตัวอย่าง: เธอขับรถ....คนขับรถ....แม่ค้า....มุ้งลวด....ผลไม้....กะปิ

ภาพที่ 1 แนวต่อเนื่องของวลี คำประสม และคำเดี่ยว ในภาษาไทย (Anchalee, 2005)

จากพื้นฐานแนวคิดว่าการกลายเป็นศัพท์เป็นปรากฏการณ์ที่เกี่ยวข้องกับการเปลี่ยนแปลงของภาษาในหลายแง่มุมร่วมกัน (Brinton & Traugott, 2005; Lipka, 1992) งานวิจัยของลิปก้าและคณะ (Lipka et al., 1994) จึงเสนอว่าระดับการกลายเป็นศัพท์ของหน่วยสร้างจำเป็นต้องศึกษาจากการเปลี่ยนแปลงที่เกิดขึ้นในระดับต่าง ๆ ของภาษาประกอบกัน ไม่ใช่เพียงการเปลี่ยนแปลงใดการเปลี่ยนแปลงหนึ่งเท่านั้น ผู้วิจัยทบทวนวรรณกรรมเกี่ยวกับการเปลี่ยนแปลงแง่มุมต่าง ๆ ที่เกิดขึ้นในกระบวนการกลายเป็นศัพท์ รวมถึงระเบียบวิธีเชิงคุณภาพและเชิงปริมาณในการวัดระดับการเปลี่ยนแปลง ในหัวข้อ 2.1.1 - 2.1.4

2.1.1 การเปลี่ยนแปลงหน่วยคำและวากยสัมพันธ์

การเปลี่ยนแปลงทางหน่วยคำและวากยสัมพันธ์ที่เกิดขึ้นในกระบวนการกลายเป็นศัพท์ได้แก่การควบรวมขององค์ประกอบทางวากยสัมพันธ์ การควบรวมนี้อาจทำให้เกิดการเปลี่ยนแปลงขึ้นกับสถานะทางหน่วยคำของหน่วยสร้าง เช่น หน่วยทางวากยสัมพันธ์เปลี่ยนแปลงกลายเป็นหน่วยศัพท์ (syntagm > lexeme) หรือหน่วยศัพท์ที่ซับซ้อนกลายเป็น

[3] ตามนิยาม Lexicalization as fusion โดยบรินตันและเทรากอตต์ (Brinton & Traugott, 2005)

หน่วยศัพท์ที่ซับซ้อนน้อยลง (complex > simpler lexeme) (Lessau, 1994; Lipka, 2002; Blank, 2001; Vipas, 2014)

ในเชิงคุณภาพ การเปลี่ยนแปลงหน่วยคำและวากยสัมพันธ์จึงสามารถศึกษาได้จากการวิเคราะห์ความตรึงแน่นระหว่างองค์ประกอบ โดยมีแนวคิดว่าการควบรวมคือกระบวนการที่ความตรึงแน่นระหว่างองค์ประกอบ (internal cohesion) ของคำเพิ่มขึ้น (Harris & Campbell, 1995) กล่าวคือเมื่อหน่วยสร้างผ่านกระบวนการกลายเป็นศัพท์ ความตรึงแน่นระหว่างองค์ประกอบจะเพิ่มขึ้นกระทั่งไม่สามารถแทรกคำไวยากรณ์หรือคำเนื้อหาระหว่างองค์ประกอบ ไม่สามารถเติมคำเพื่อขยายแต่ละองค์ประกอบของหน่วยสร้างและไม่สามารถเติมคำปฏิเสธหรือทำให้อยู่ในรูปกรรมวาจก (passive voice) ได้

แนวคิดนี้มีพื้นฐานจากมุมมองของสมมติฐาน Lexical Integrity Hypothesis (Haspelmath & Sims, 2010; Bhimbasistha, 2015) ซึ่งเสนอว่าหน่วยสร้างทางวากยสัมพันธ์ เช่น วลี มีระดับการกลายเป็นศัพท์ต่ำและมีความตรึงแน่นระหว่างองค์ประกอบต่ำ กฎทางวากยสัมพันธ์จึงสามารถเข้าไปมีผลกับองค์ประกอบใดองค์ประกอบหนึ่งภายในหน่วยสร้างนั้นได้ ตัวอย่างเช่น นามวลี *big house* สามารถใช้คำว่า *very* ขยายคำคุณศัพท์ *big* เพียงองค์ประกอบเดียวโดยไม่ขยาย *house* ได้ (เช่น *very big house*) ในขณะที่คำเดี่ยวหรือคำที่มีระดับการกลายเป็นศัพท์สูงมีความตรึงแน่นระหว่างองค์ประกอบมาก กฎทางวากยสัมพันธ์จึงไม่สามารถมีผลแบบเฉพาะเจาะจงกับองค์ประกอบใดเพียงองค์ประกอบเดียว ตัวอย่างเช่นคำประสม *crispbread* ที่ไม่สามารถใช้ *very* ขยายเฉพาะคำคุณศัพท์ crisp ได้ **very crispbread* อย่างไรก็ตามการวิเคราะห์ความตรึงแน่นระหว่างองค์ประกอบมีข้อจำกัดเช่นเดียวกับการวิเคราะห์ความหมายในแง่ที่ไม่สามารถให้ข้อมูลเชิงปริมาณจึงไม่สามารถนำไปศึกษาในงานวิจัยบางประเภทได้

สำหรับข้อมูลเชิงปริมาณของโครงสร้างหน่วยคำและวากยสัมพันธ์สามารถศึกษาได้โดยใช้การทดลองทางภาษาศาสตร์จิตวิทยาตามแนวคิดเรื่องการประมวลผลคำ (lexical processing) และรูปแทนในระบบปริชาน (mental representation) โดยมีแนวคิดพื้นฐานว่าคำเดี่ยวและวลีมีรูปแทนในคลังศัพท์และการประมวลผลแตกต่างกัน คำเดี่ยวทั้งคำถูกเก็บในคลังศัพท์โดยการจดจำ (memorize) ในขณะที่ประโยคมีการประมวลผลแบบประกอบความ (composite) จากความหมายของแต่ละคำในประโยคนั้น ตัวอย่างเช่น การที่ผู้พูดเห็นคำเดี่ยว *cat* และสามารถเชื่อมโยงไปสู่สัตว์สี่ขาชนิดหนึ่งได้เป็นสิ่งที่เกิดขึ้นจากความจำ โดยคำดังกล่าวถูกเก็บไว้ในคลังศัพท์ (mental lexicon) ของผู้พูด ไม่ได้เกิดจากการประกอบความขึ้นมาจากหน่วยเสียง */k/, /ae/, และ /t/* ซึ่งไม่มีความหมายในตัวเอง ในทางตรงกันข้ามความหมายของประโยคโดยส่วนใหญ่มักเกิดจากการประกอบความจากความหมายของหน่วยที่เล็กกว่าคือแต่ละคำในประโยคนั้น เช่น ความหมายของประโยค *แมวกินปลา* เกิดจากความหมายของ *แมว* คือสัตว์สี่ขาชนิดหนึ่ง ประกอบกับความหมายของ *กิน* คือกริยาการนำอาหารเข้าปากเคี้ยวและกลืน และความหมายของ *ปลา* ซึ่งคือสัตว์น้ำชนิดหนึ่ง

รูปแทนในคลังศัพท์และรูปแบบการประมวลผลนี้สามารถศึกษาได้โดยใช้การทดลองการตัดสินคำศัพท์ (lexical decision task) ซึ่งเป็นการทดลองทางภาษาศาสตร์จิตวิทยารูปแบบหนึ่งซึ่งให้ผู้เข้าร่วมวิจัยตอบให้รวดเร็วที่สุดว่าคำที่ปรากฏเป็นคำที่มีความหมายในภาษาหรือไม่และทำการวัดค่าระยะเวลาในการตอบสนองตั้งแต่คำปรากฏบนหน้าจอจนกระทั่งผู้เข้าร่วมวิจัยกดตอบ โดยคำที่มีค่าระยะเวลาตอบสนองน้อยแสดงให้เห็นโครงสร้างและการประมวลผลที่เรียบง่ายเช่นเดียวกับการประมวลผลของคำเดี่ยว ส่วนคำที่มีค่าระยะเวลาตอบสนองมากแสดงให้เห็นโครงสร้างทางหน่วยคำวากยสัมพันธ์และการประมวลผลที่ซับซ้อนกว่าเช่นเดียวกับวลีและประโยค

สำหรับคำประสม งานวิจัยในอดีตพบรูปแทนและการประมวลผลทั้งสองรูปแบบในคำประสม คำประสมบางกลุ่มมีค่าระยะเวลาตอบสนองใกล้เคียงกับคำเดี่ยวแสดงถึงโครงสร้างที่เรียบง่าย ในขณะที่คำประสมบางกลุ่มมีค่าระยะเวลาตอบสนองมากใกล้เคียงกับวลี (van Jaarsveld & Rattink, 1988) โดยลิบเบน (Libben, 2007) ได้อธิบายว่าสาเหตุที่พบรูปแทนทั้งสองรูปแบบในคำประเภทนี้อาจเป็นผลมาจากอิทธิพลของกระบวนกลายเป็นศัพท์ คือคำประสมที่มีค่าระยะเวลาตอบสนองคล้ายคำเดี่ยวเป็นคำประสมที่ผ่านการกลายเป็นศัพท์สูง ในขณะที่คำประสมที่มีค่าระยะเวลาตอบสนองคล้ายวลีมีแนวโน้มเป็น

คำประสมที่มีระดับการกลายเป็นศัพท์ต่ำ

จากการทบทวนวรรณกรรมยังไม่พบงานวิจัยเกี่ยวกับการกลายเป็นศัพท์ที่นำการทดลองทางภาษาศาสตร์จิตวิทยาและค่าระยะเวลาตอบสนองมาใช้ศึกษาการเปลี่ยนแปลงโครงสร้างหน่วยคำและวากยสัมพันธ์หรือใช้เป็นค่าบ่งชี้เชิงปริมาณของการกลายเป็นศัพท์ อย่างไรก็ตามข้อค้นพบของงานวิจัยในอดีตแสดงให้เห็นว่าระดับการกลายเป็นศัพท์มีความสัมพันธ์กับค่าระยะเวลาตอบสนอง ผู้วิจัยจึงมีความเห็นว่าค่าระยะเวลาตอบสนองสามารถสะท้อนโครงสร้างหน่วยคำและวากยสัมพันธ์ของหน่วยสร้างและมีความเป็นไปได้ที่จะใช้เป็นค่าเชิงปริมาณที่บ่งชี้ระดับการกลายเป็นศัพท์

2.1.2 การเปลี่ยนแปลงความหมาย

สำหรับแง่มุมทางความหมาย การเปลี่ยนแปลงที่สำคัญและเห็นได้ชัดเจนคือการลดลงของระดับความหมายเชิงประกอบของหน่วยสร้างที่ผ่านกระบวนการกลายเป็นศัพท์ (Brinton & Traugott, 2005) หน่วยสร้างที่มีระดับความหมายเชิงประกอบสูง (high compositionality) หมายถึง ความหมายของหน่วยสร้างประกอบขึ้นจากความหมายของแต่ละองค์ประกอบอย่างสมบูรณ์ เช่น วลี *ซักผ้า* มีความหมายเกี่ยวข้องกับ *ซัก* และ *ผ้า* ในระดับสูงมาก ในขณะที่หน่วยสร้างที่มีระดับความหมายเชิงประกอบต่ำ (low compositionality) หมายถึงความหมายของทั้งหน่วยสร้างคล้ายคลึงกับความหมายของแต่ละองค์ประกอบน้อย เช่น คำเดี่ยว *นาที* ซึ่งความหมายไม่เกี่ยวข้องกับ *นา-* และ *-ที*

แบลงก์ (Blank, 2001) ใช้คำว่ากระบวนการกลายเป็นสำนวน (idiomatization) อธิบายลักษณะการเปลี่ยนแปลงความหมายที่เกิดขึ้นในกระบวนการกลายเป็นศัพท์ โดยมองว่าการเปลี่ยนแปลงความหมายในกระบวนการนี้มีลักษณะเป็นระดับและในท้ายที่สุดอาจเปลี่ยนแปลงไปสู่ระดับที่คาดเดาไม่ได้จากองค์ประกอบเช่นเดียวกับสำนวน ระดับที่ 1 ของการเปลี่ยนแปลงความหมายได้แก่คำที่สามารถคาดเดาความหมายได้อย่างสมบูรณ์จากแต่ละองค์ประกอบหรือคำที่มีความหมายเชิงประกอบสูง (high compositionality) ระดับที่ 2 คือคำที่ยังคงเค้าความหมายขององค์ประกอบอยู่บางส่วนและมีการลดหรือขยายความหมายไปบางส่วน ส่วนระดับที่ 3 นั้นคือคำที่ไม่สามารถคาดเดาความหมายจากแต่ละองค์ประกอบได้มีความหมายเชิงประกอบต่ำหรือคำที่มีการใช้ในเชิงอุปลักษณ์หรือนามนัยไปเรียบร้อยแล้ว

ระดับที่ 1 (เยอรมัน) *Apfelkuchen ‘แอปเปิล’ + ‘ขนมพาย’ = ‘ขนมพายที่ทำจากแอปเปิล’*
ระดับที่ 2 (อังกฤษ) *wheelchair ‘ล้อ’ + ‘เก้าอี้’ = ‘รถเข็นคนพิการ’*
ระดับที่ 3 (อังกฤษ) *black sheep ‘แกะ’ + ‘ดำ’ = ‘คนที่ทำอะไรผิดแผกไปจากกลุ่ม’*

คำที่ผ่านกระบวนการกลายเป็นศัพท์น้อยยังคงมีความหมายเชิงประกอบมากเช่นเดียวกับวลีและประโยคที่ความหมายของหน่วยสร้างประกอบขึ้นอย่างสมบูรณ์จากความหมายของแต่ละองค์ประกอบ ส่วนคำที่ผ่านกระบวนการกลายเป็นศัพท์มากแล้ว ระดับความหมายเชิงประกอบของคำจะลดลงและไม่สามารถคาดเดาความหมายของทั้งคำได้จากความหมายของแต่ละองค์ประกอบหรือในบางกรณีพบว่าความหมายของหน่วยสร้างมีการย้ายที่ มีความหมายเพิ่มขึ้นหรือลดลงในเชิงอุปลักษณ์ (metaphor) หรือมีความหมายเชิงนามนัย (metonymy) ได้อีกด้วย ซึ่งเป็นลักษณะที่คล้ายคลึงกับคำเดี่ยวซึ่งความหมายของทั้งคำไม่สามารถทำนายได้จากความหมายของแต่ละองค์ประกอบ

งานวิจัยเกี่ยวกับการเปลี่ยนแปลงความหมายในกระบวนการกลายเป็นศัพท์มักวิเคราะห์ความหมายในเชิงคุณภาพและอภิปรายว่าคำมีระดับการกลายเป็นความหมายมากน้อยเพียงใด ตัวอย่างเช่นซีวิลเลรี (Civilleri, 2013) ได้ศึกษาระดับการกลายเป็นศัพท์ของคำนามที่มีหน่วยคำเติม (derived nouns, DN) ในภาษากรีกจากการวิเคราะห์การเปลี่ยนแปลงความหมาย ตัวอย่างเช่น

1) Non-lexicalized DNs

amphí- ba- sis

around go proc.f:nom.sg.

'surrounding'

ในตัวอย่างนี้ซีวิลเลรี (Civilleri, 2013) วิเคราะห์ว่าคำยังมีความหมายที่เป็นเชิงประกอบอย่างชัดเจน ความหมายของคำนามทั้งคำ *'surrounding'* เกี่ยวข้องกับความหมายของ *'around'* *'go'* และ *'the process of doing V'* ค่อนข้างมาก แสดงให้เห็นตัวอย่างของคำที่ผ่านการกลายเป็นศัพท์ระดับต่ำ

2) Lexicalized DNs

chý- sis

pour proc.f:nom.sg.

'pile'

ตัวอย่างที่ 2 แสดงคำที่ความหมายไม่เป็นเชิงประกอบ (non-compositional) หน่วยคำเติมท้าย – *sis* เมื่อเติมหลังคำกริยามีความหมายว่า *'the process of doing V'* แต่ในตัวอย่างนี้ความหมายของคำนามไม่ได้มีความหมายว่า *'the process of pouring'* หรือ *การเท(น้ำ)* แต่มีความหมายว่า 'กอง' ซึ่งความหมายไม่ได้เกิดจากการประกอบความของแต่ละองค์ประกอบในหน่วยสร้าง ผู้เขียนจึงวิเคราะห์ให้คำนี้เป็นคำที่ผ่านกระบวนการกลายเป็นศัพท์ในระดับมากแล้ว

นอกจากการวิเคราะห์ความหมาย (semantic analysis) ซึ่งให้ข้อมูลเชิงคุณภาพ ระดับความหมายเชิงประกอบสามารถศึกษาในเชิงปริมาณโดยใช้วิธีการตัดสินคะแนนโดยผู้พูด (native speakers' judgement) เรดดี้และคณะ (Reddy et al., 2011) ได้เสนอวิธีการตัดสินระดับความหมายเชิงประกอบโดยใช้แบบสอบถาม *literality rating* ซึ่งให้ผู้พูดตัดสินว่าคำแต่ละคำที่ต้องการทดสอบมีความหมายตรงตัว (literal meaning)[4] มากน้อยเพียงใดโดยใช้มาตรวัดแบบลิเคิร์ท (Likert's scale) ที่มีคะแนน 1 - 5 ข้อมูลที่ได้จากแบบสอบถามนี้จึงมีลักษณะเป็นตัวเลขซึ่งสามารถนำไปศึกษาเชิงปริมาณได้

แม้ว่าผู้วิจัยไม่พบงานวิจัยในอดีตที่ใช้คะแนนระดับความหมายเชิงประกอบจากการตัดสินของผู้พูดเป็นค่าบ่งชี้ระดับการกลายเป็นศัพท์โดยตรง แต่จากการทบทวนแนวคิดและตัวอย่างข้อมูลที่ได้จากการศึกษาโดยใช้แบบสอบถามดังกล่าวแสดงให้เห็นว่า วิธีการเก็บข้อมูลเช่นนี้สามารถบ่งชี้ระดับความหมายของคำในเชิงปริมาณได้ และมีความเป็นไปได้ในการนำมาใช้เป็นหนึ่งในค่าบ่งชี้การกลายเป็นศัพท์เพื่อศึกษาระดับการกลายเป็นศัพท์ของหน่วยสร้าง (วิธีการเก็บข้อมูลระดับความหมายเชิงประกอบของงานวิจัยนี้ดูหัวข้อ 3.2)

2.1.3 การเปลี่ยนแปลงเชิงการใช้

งานวิจัยในอดีตจำนวนมากพบว่าการกลายเป็นศัพท์สัมพันธ์กับความถี่การปรากฏ (Brinton & Traugott, 2005; Lipka et al., 1994) โดยหน่วยสร้างที่ผ่านการกลายเป็นศัพท์มากมีความถี่การใช้สูงกว่าหน่วยสร้างที่ผ่านการกลายเป็นศัพท์น้อย ความถี่การปรากฏถือเป็นค่าบ่งชี้การกลายเป็นศัพท์เชิงปริมาณเพียงค่าเดียวที่พบการใช้อย่างแพร่หลาย ตัวอย่างงานวิจัยที่ใช้ความถี่การปรากฏเป็นค่าบ่งชี้การกลายเป็นศัพท์ได้แก่แพล็กและคณะ (Plag et al., 2008) ซึ่งศึกษารูปแบบการเน้น

[4] มุมมองของเรดดี้และคณะ (Reddy et al., 2011) มองว่าความหมายตรงตัวและความหมายเชิงประกอบเป็นสิ่งเดียวกัน ผลที่ได้จากแบบสอบถามระดับความหมายตรงตัวจึงสามารถสะท้อนถึงระดับความหมายเชิงประกอบได้

คำประสมในภาษาอังกฤษ งานวิจัยนี้ใช้ความถี่การปรากฏของคำร่วมกับรูปเขียนของคำในการจำแนกคำประสมที่ผ่านการกลายเป็นศัพท์มากออกจากคำประสมใหม่ นอกจากยังมีนี้งานวิจัยภาษาศาสตร์จิตวิทยา เกี่ยวกับการประมวลผลคำที่เลือกใช้ความถี่การปรากฏเป็นค่าบ่งชี้การกลายเป็นคำศัพท์ของคำด้วย เช่น ฟาน จาร์สเวลด์และแรตติงก์ (van Jaarsveld & Rattink, 1988) ซึ่งใช้ความถี่การปรากฏร่วมกับรูปเขียนและระดับความคุ้นเคย และโซสกูธีและเฮย์ (Soskuthy & Hay, 2017) ซึ่งใช้ความถี่การปรากฏเพียงค่าเดียวในการบ่งชี้ระดับการกลายเป็นศัพท์

2.1.4 การกลายเป็นศัพท์ในภาษาไทย

งานวิจัยที่ศึกษาการกลายเป็นศัพท์ในภาษาไทย (Pranee, 2001; Vipas, 2014) ได้แสดงให้เห็นรูปแบบการเปลี่ยนแปลงที่สอดคล้องกับข้อเสนอของงานวิจัยในภาษาอื่น ๆ ทั้งการเปลี่ยนแปลงหน่วยคำวากยสัมพันธ์ซึ่งมีการควบรวมของหน่วยคำ การเปลี่ยนแปลงความหมายซึ่งระดับความหมายเชิงประกอบของหน่วยสร้างลดลง และการเปลี่ยนแปลงเชิงการใช้ซึ่งหน่วยสร้างมีความถี่การใช้เพิ่มขึ้น โดยงานวิจัยในอดีตได้กล่าวถึงการเปลี่ยนแปลงเหล่านี้ไว้แล้ว ทั้งในรูปแบบของการวิเคราะห์เชิงคุณภาพและการตั้งข้อสังเกต อย่างไรก็ตาม ยังไม่พบงานวิจัยในภาษาไทยที่ศึกษาการเปลี่ยนแปลงเหล่านี้ในเชิงปริมาณอย่างละเอียด

วิภาส (Vipas, 2014) ได้ยกตัวอย่างหน่วยสร้างที่สันนิษฐานว่ามีการเปลี่ยนแปลงจากกระบวนการกลายเป็นศัพท์ เช่น *พรูก/พรุก+นี้ > พรูกนี้/พรุกนี้ > พรุ่งนี้* (Pranee, 2001; Vipas, 2014) โดยอธิบายว่าการเปลี่ยนแปลงที่เกิดขึ้นกับหน่วยสร้างดังกล่าวนั้นประกอบด้วย 1) การควบรวมคำจากหน่วยทางวากยสัมพันธ์ที่ประกอบจากหน่วยศัพท์สองหน่วยคือ *พรูก/พรุก* และ *นี้* ได้ควบรวมกลายเป็นหน่วยศัพท์หนึ่งหน่วย คือ *พรุ่งนี้* 2) การเปลี่ยนแปลงความหมาย คือ ความหมายของ *พรูก/พรุก/พรุ่ง* ได้จางหายไป (bleached) และไม่สามารถปรากฏเป็นคำโดยลำพังได้ ทำให้ผู้พูดไม่สามารถตีความความหมายของ *พรุ่งนี้* จากแต่ละองค์ประกอบ และ 3) การเปลี่ยนแปลงเชิงการใช้ ซึ่งผู้เขียนได้ตั้งข้อสังเกตและสันนิษฐานว่า หน่วยสร้างดังกล่าวรวมถึงคำอื่น ๆ ที่เกิดกระบวนการกลายเป็นศัพท์ มีการควบรวมของหน่วยคำและมีการเปลี่ยนแปลงความหมายนี้มักเป็นคำที่มีการใช้บ่อย

จากการทบทวนวรรณกรรมทั้งหมดผู้วิจัยพบว่าการศึกษากระบวนการกลายเป็นศัพท์ในเชิงปริมาณยังคงเป็นช่องว่างของการศึกษาในภาษาไทยและภาษาอื่น ๆ โดยการเปลี่ยนแปลงหน่วยคำวากยสัมพันธ์ การเปลี่ยนแปลงความหมายและการเปลี่ยนแปลงเชิงการใช้นี้สามารถเก็บข้อมูลในเชิงปริมาณได้ โดยการเปลี่ยนแปลงหน่วยคำวากยสัมพันธ์เก็บข้อมูลจากค่าระยะเวลาตอบสนองซึ่งสามารถบ่งชี้ถึงระดับความซับซ้อนของหน่วยคำที่แตกต่างกัน การเปลี่ยนแปลงความหมายสามารถเก็บข้อมูลจากแบบสอบถามระดับความหมายเชิงประกอบ และการเปลี่ยนแปลงเชิงการใช้สามารถเก็บข้อมูลความถี่การปรากฏจากคลังข้อมูล แม้ว่าจากการทบทวนวรรณกรรมไม่พบงานวิจัยในอดีตที่ใช้ค่าทั้งสามประกอบกันเพื่อบ่งชี้ระดับการกลายเป็นศัพท์ แต่หากพิจารณาจากลักษณะและนิยามของค่าทั้งสาม ผู้วิจัยพบความน่าจะเป็นอย่างยิ่งที่ทั้งสามค่าสามารถใช้เป็นค่าบ่งชี้ระดับการกลายเป็นศัพท์ได้ โดยวิธีการเก็บข้อมูลเพื่อทดสอบสมมติฐานจะนำเสนอในหัวข้อที่ 3

3. วิธีดำเนินการวิจัย

ค่าบ่งชี้การกลายเป็นศัพท์ทั้งสามค่าที่ใช้ในงานวิจัยนี้เป็นค่าบ่งชี้เชิงปริมาณทั้งหมดและมีพื้นฐานจากการเปลี่ยนแปลงแง่มุมต่าง ๆ ที่เกิดขึ้นในกระบวนการกลายเป็นศัพท์ดังเสนอในหัวข้อ 2.1.1-2.1.3 งานวิจัยแบ่งออกเป็นสองส่วนเพื่อตอบ

คำถามวิจัยแต่ละข้อ ในส่วนแรกมีเป้าหมายเพื่อทดสอบว่าค่าบ่งชี้เชิงปริมาณทั้งสามค่าที่งานวิจัยนี้เสนอสามารถใช้เป็นค่าบ่งชี้การกลายเป็นศัพท์ได้หรือไม่ หากค่าทั้งสามสามารถจำแนกคำที่มีระดับการกลายเป็นศัพท์มากและน้อยที่สุดออกจากกันได้แสดงว่าค่าดังกล่าวสามารถแสดงถึงระดับการกลายเป็นศัพท์ที่ต่างกันและใช้เป็นค่าบ่งชี้การกลายเป็นศัพท์ได้ ในส่วนนี้ผู้วิจัยทำการเก็บข้อมูลจากคำสองประเภทคือคำเดี่ยวและวลีสองพยางค์ โดยคำเดี่ยวเป็นตัวแทนของคำที่ผ่านการกลายเป็นศัพท์มากและวลีสองพยางค์เป็นตัวแทนของคำที่ผ่านการกลายเป็นศัพท์น้อย

ในส่วนที่สองมีเป้าหมายเพื่อทดสอบสหสัมพันธ์ระหว่างค่าบ่งชี้ทั้งสาม หากค่าบ่งชี้มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติหมายความว่าการระบุระดับการกลายเป็นศัพท์สามารถใช้ข้อมูลจากค่าบ่งชี้ค่าใดค่าหนึ่งเพียงค่าเดียวเนื่องจากเมื่อทราบค่าบ่งชี้หนึ่งสามารถคำนวณระดับของค่าบ่งชี้อื่นได้ แต่หากไม่พบความสัมพันธ์ระหว่างค่าบ่งชี้ทั้งสามก็บอกเป็นนัยว่าการเปลี่ยนแปลงแง่มุมต่าง ๆ ของการกลายเป็นศัพท์เป็นอิสระจากกัน การเปลี่ยนแปลงต่าง ๆ เกิดขึ้นในเวลาและ อัตราเร็วที่แตกต่างกัน ดังนั้นระดับการกลายเป็นศัพท์อาจไม่สามารถบ่งชี้จากค่าใดค่าหนึ่งเพียงค่าเดียวได้ ต้องใช้ค่าบ่งชี้มากกว่าหนึ่งค่าประกอบกัน ในส่วนที่สองนี้เก็บข้อมูลจากคำประสมซึ่งเป็นตัวแทนของคำที่มีระดับการกลายเป็นศัพท์หลากหลาย โดยบางคำมีลักษณะคล้ายวลีและบางคำคล้ายคลึงกับคำเดี่ยว

งานวิจัยได้คัดเลือกคำและวลีเป้าหมายด้วยเกณฑ์ทางหน่วยคำและวากยสัมพันธ์ดังนี้

(1) ผู้วิจัยใช้นิยามคำประสมของฮัสเปิลมัทและซิมส์ (Haspelmath & Sims, 2010) ซึ่งเสนอว่าคำประสมคือคำซับซ้อนที่ประกอบจากหน่วยศัพท์ (lexeme) สองหน่วยขึ้นไป โดยคำประสมที่ใช้ในงานนี้รวบรวมจากงานวิจัยที่เกี่ยวข้องกับคำประสมทั้งสิ้น 5 เล่ม (Alisa, 2014; Anchalee, 2005; Duangchan, 1976; Anong, 1982; Yordkwan, 2008)

(2) งานวิจัยนี้ไม่มีข้อจำกัดในแง่ประเภทของคำประสม คำประสมเป้าหมายสามารถเป็นได้ทั้งคำประสมแบบเท่าเทียม (coordinate compound) หรือคำซ้อน ซึ่งเกิดจากคำในแวดวงความหมายเดียวกัน เช่น *คุ้มกัน* คำประสมแบบไม่เท่าเทียม (subordinate compound) ซึ่งประกอบจากคำในแวดวงความหมายต่างกัน เช่น *งานบ้าน* คำประสมแบบเข้าศูนย์ (endocentric compound) ซึ่งมีองค์ประกอบใดองค์ประกอบหนึ่งทำหน้าที่เป็นส่วนหลักทางความหมาย (semantic head) เช่น *หมาป่า* หรือคำประสมแบบไร้ศูนย์ (exocentric compound) ที่ไม่สามารถระบุได้ชัดเจนว่าองค์ประกอบใดทำหน้าที่เป็นส่วนหลักทางความหมายอย่าง เช่น *พ่อแม่*

(3) ทั้งสององค์ประกอบของคำประสมเป็นคำที่มีการใช้ในภาษาไทยกรุงเทพปัจจุบัน เนื่องจากหากผู้พูดไม่รู้จักคำใดคำหนึ่งอาจส่งผลให้การตัดสินระดับความหมายเชิงประกอบคลาดเคลื่อนได้ ตัวอย่างเช่นคำประสม *ง่ายดาย* ซึ่งคำว่า *ดาย* ไม่มีการใช้เป็นคำอิสระในความหมายว่า *‘ง่าย’* ในภาษาไทยกรุงเทพปัจจุบัน คำประสมนี้จึงไม่สามารถเป็นคำเป้าหมายของงานวิจัยนี้ได้

(4) คำประสมเป้าหมายต้องไม่มีความหมายกำกวมกับวลีโครงสร้างเดียวกันเนื่องจากความกำกวมดังกล่าวอาจส่งผลต่อการตัดสินระดับการกลายเป็นศัพท์ ตัวอย่างเช่นหน่วยสร้าง *ของหาย* อาจเป็นได้ทั้งคำประสมที่มีความหมายว่า *‘lost thing’* และประโยคที่ประกอบด้วยประธานและกริยามีความหมายว่า *‘something disappeared’*

(5) คำประสมที่ใช้ต้องมีโครงสร้างแบบคำนามทั้งสององค์ประกอบ (N+N) หรือกริยาทั้งสององค์ประกอบ (V+V) ส่วนวลีสองพยางค์เป็นโครงสร้าง กริยา+กริยาวิเศษณ์ (V+Adv) หรือ กริยา+นาม (V+N) เท่านั้น เพื่อป้องกันความกำกวมที่อาจเกิดขึ้นระหว่างหน่วยสร้างทั้งสองประเภทและควบคุมอิทธิพลของประเภททางไวยากรณ์ที่อาจส่งผลต่อการตัดสินค่าระยะเวลาตอบสนอง (Xia et al., 2016; Xia et al., 2022)

การเก็บข้อมูลประกอบด้วยการเก็บข้อมูลค่าระยะเวลาตอบสนองจากการทดลองการตัดสินคำศัพท์ เก็บข้อมูลระดับความหมายเชิงประกอบจากแบบสอบถามออนไลน์และเก็บข้อมูลความถี่การปรากฏจากคลังข้อมูลภาษาไทยแห่งชาติ ผู้เข้าร่วมการวิจัยได้แก่นิสิตระดับปริญญาตรีจำนวน 8 คน อายุ 19 - 22 ปี กำลังศึกษาในคณะและสาขาที่ไม่เกี่ยวข้องกับภาษาและภาษาศาสตร์ เป็นผู้พูดภาษาไทยกรุงเทพ อาศัยในเขตกรุงเทพมหานคร นนทบุรี หรือสมุทรปราการ ผู้เข้าร่วมการวิจัยทุกคนเข้าร่วมการเก็บข้อมูลสองครั้งโดยเริ่มจากการตัดสินคำศัพท์ซึ่งเก็บข้อมูลค่าระยะเวลาตอบสนอง จากนั้นจึงทำแบบสอบถามออนไลน์เพื่อเก็บข้อมูลระดับความหมายเชิงประกอบ

3.1 เก็บข้อมูลค่าระยะเวลาตอบสนอง

การเก็บข้อมูลค่าระยะเวลาตอบสนองใช้การทดลองการตัดสินคำศัพท์โดยให้ผู้เข้าร่วมวิจัยตัดสินว่าคำที่ปรากฏนั้นเป็นคำที่มีความหมายในภาษาไทยหรือไม่ ตอบว่า *ใช่* หรือ *ไม่ใช่* โดยกดปุ่มบนแป้นพิมพ์ตามที่กำหนด หาก *“ใช่”* ให้กดปุ่ม *“L”* หาก *“ไม่ใช่”* ให้กดปุ่ม *“A”* ผู้วิจัยสร้างการทดลองโดยใช้โปรแกรม E-Prime (Schneider et al., 2002) คำทดสอบประกอบด้วยคำประสมจำนวน 22 คำ คำเดี่ยวความถี่สูงจำนวน 22 คำ คัดเลือกจากรายการคำความถี่สูง 5,000 คำแรกของคลังข้อมูลภาษาไทยแห่งชาติ (Thai National Corpus) (Aroonmanakun, 2007) และคำสมมติ (non-word) ซึ่งมีโครงสร้างถูกต้องตามสัทสัมผัส (phonotactics) แต่ไม่มีความหมายในภาษาไทย จำนวน 44 คำ รวมทั้งสิ้น 88 คำ ก่อนการทดลองมีคำตัวอย่างจำนวน 5 คำให้ผู้เข้าร่วมวิจัยลองตอบเพื่อตรวจสอบความเข้าใจก่อนทำการทดลองจริง โดยไม่นำค่าระยะเวลาตอบสนองของคำตัวอย่างมาวิเคราะห์ผลการทดลอง (รวมคำในรายการคำทั้งหมด 93 คำ) การเก็บข้อมูลเริ่มต้นจากการนำเสนอคำชี้แจงและวิธีทำแก่ผู้เข้าร่วมวิจัย จากนั้นเป็นการตัดสินคำตัวอย่างจำนวน 5 คำ และจึงเข้าสู่การทดลองจริง การทดลองแบ่งออกเป็น 4 ช่วง นำเสนอคำเป้าหมายช่วงละ 22 คำ ภาพรวมของการทดลองดังแสดงในภาพที่ 2

ภาพที่ 2 ภาพรวมของการทดลอง

เมื่อผู้เข้าร่วมวิจัยกดปุ่มเพื่อตอบคำถามโปรแกรมจะบันทึกค่าระยะเวลาตอบสนองและคำตอบโดยอัตโนมัติ นำเสนอคำทดสอบเป็นระยะเวลาไม่จำกัดเพื่อไม่ให้มีข้อมูลสูญหาย (missing data) ในชุดข้อมูล ผู้เข้าร่วมวิจัยต้องตอบให้เร็วที่สุดเมื่อกดคำตอบแล้วโปรแกรมจึงนำเสนอคำต่อไป

3.2 เก็บข้อมูลระดับความหมายเชิงประกอบ

ระดับความหมายเชิงประกอบเก็บข้อมูลจากแบบสอบถาม โดยผู้เข้าร่วมวิจัยให้คะแนนความเกี่ยวข้องของความหมายของทั้งคำ/วลีและแต่ละองค์ประกอบโดยใช้มาตรวัดแบบเลื่อน (slider scale) ที่มีช่วงคะแนน 0-100 แบบสอบถามประกอบด้วยคำทดสอบทั้งหมด 66 คำ แบ่งเป็นคำประสม 22 คำ คำเดี่ยวสองพยางค์ 22 คำ และวลีสองพยางค์ 22 วลี

ในหน้าแรกของแบบสอบถามให้ข้อมูลเบื้องต้น จำนวนข้อ คำชี้แจงและเวลาที่ใช้ในการทำโดยประมาณ แบบสอบถามหน้าที่สองประกอบด้วยข้อมูลส่วนตัวของผู้ให้ข้อมูล ได้แก่ อายุ เพศ จังหวัดที่เกิด จังหวัดที่อาศัยปัจจุบัน ชั้นปีการศึกษา คณะและสาขาที่ศึกษา รวมถึงประสบการณ์การเรียนภาษาไทยถิ่นอื่น ๆ และภาษาบาลีสันสกฤตและเขมร ซึ่งอาจส่งผลต่อการตัดสินระดับความหมายเชิงประกอบ จากนั้นเป็นตัวอย่างการตัดสินให้คะแนนระดับความหมายเชิงประกอบ ประกอบด้วยตัวอย่างหน่วยสร้างที่มีความหมายเชิงประกอบสูงหนึ่งตัวอย่างได้แก่ วลี *ซักผ้า* ระดับความหมายเชิงประกอบต่ำหนึ่งตัวอย่างได้แก่ คำเดี่ยว *จิงโจ้* และระดับความหมายเชิงประกอบระดับกลางหนึ่งตัวอย่าง ได้แก่คำประสม *แม่ค้า* ซึ่งเป็นคำที่ไม่ได้อยู่ในรายการคำเป้าหมายและไม่นำไปใช้ในการวิเคราะห์ ผู้วิจัยอธิบายวิธีการให้คะแนนแก่ผู้เข้าร่วมการวิจัยโดยพยายามชี้ให้ผู้เข้าร่วมวิจัยตัดสินคะแนนโดยคำนึงถึงความเกี่ยวข้องของความหมายเป็นหลัก

คำถามมีทั้งหมด 155 ข้อ แบ่งออกเป็นคำถามเป้าหมาย 132 ข้อ (66 คำ x 2 ข้อ) และคำถามอื่น ๆ ซึ่งเป็นคำถามแทรก (fillers) จำนวน 23 คำถาม แบบสอบถามแบ่งออกเป็น 7 ส่วน ส่วนละ 20 ข้อ คำถามแทรกอยู่ระหว่างคำถามเป้าหมายทุก ๆ 10 ข้อเพื่อให้ผู้บอกภาษาได้หยุดพักเป็นช่วง ๆ จากการตอบคำถามที่มีเนื้อหาซ้ำกันจำนวนมาก คำถามเป้าหมายเรียงลำดับแบบสุ่ม โดยผู้วิจัยใช้วิธีสุ่มเทียม (pseudorandom) (Arunachalam, 2013) คือสุ่มลำดับของรายการคำแต่ควบคุมไม่ให้คำเดี่ยวปรากฏในลำดับติดกัน ไม่ให้คำประสมแบบเข้าศูนย์ (endocentric compound) ซึ่งสามารถระบุส่วนหลักได้ปรากฏในตำแหน่งติดกันและควบคุมไม่ให้คำถามของคำเดียวกันปรากฏติดกันเพื่อป้องกันอิทธิพลที่แต่ละคำอาจมีซึ่งกันและกัน

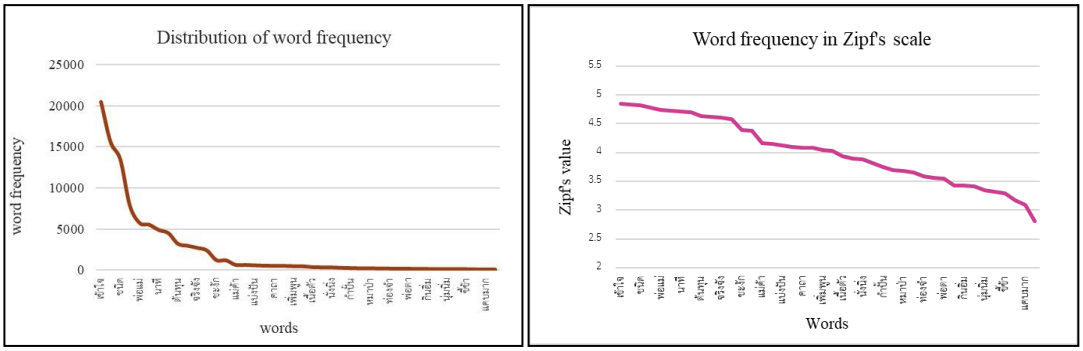
3.3 ความถี่การปรากฏ

ข้อมูลความถี่การปรากฏของคำประสมทั้งคำและแต่ละองค์ประกอบของคำประสมเก็บข้อมูลจากคลังข้อมูลภาษาไทยแห่งชาติ สำหรับความถี่การปรากฏของวลีสืบค้นโดยใช้การปรากฏร่วม (collocation) สำหรับคำเดี่ยวสืบค้นความถี่การปรากฏของคำเดี่ยวทั้งคำเท่านั้นเนื่องจากแต่ละองค์ประกอบไม่มีสถานะเป็นคำจึงไม่สามารถสืบค้นจากคลังข้อมูลได้ เช่น *ชะงัก* เก็บข้อมูลความถี่การปรากฏของ *ชะงัก* เท่านั้นไม่มีข้อมูลของ *ชะ-* และข้อมูลของ *-งัก*

สำหรับความถี่การปรากฏที่ได้จากคลังข้อมูลจะพบการแจกแจงแบบไม่ปกติ (non-parametric distribution) (ภาพที่ 3) โดยคำที่มีความถี่สูงเป็นอันดับที่สองจะมีค่าเท่ากับ 1/2 หรือร้อยละ 50 ของความถี่ของอันดับที่หนึ่ง คำที่มีความถี่ปรากฏมากเป็นอันดับที่สามจะมีความถี่เป็น 1/3 หรือประมาณร้อยละ 33.33 ของความถี่การปรากฏของคำอันดับที่หนึ่ง คำที่มีความถี่การปรากฏลำดับที่สี่จะมีความถี่เท่ากับ 1/4 หรือประมาณร้อยละ 25 ของความถี่การปรากฏของคำอันดับที่หนึ่งและลดลงในลำดับต่อ ๆ ไป ข้อมูลที่มีการกระจายลักษณะนี้อาจเป็นปัญหาในการนำไปทำสถิติบางประเภทเช่นสหสัมพันธ์ของเพียร์สัน (Pearson’s Correlation) ผู้วิจัยจึงทำการปรับค่าความถี่การปรากฏเป็น Zipf’s value คำนวณได้โดยใช้ลอการิทึมฐานสิบและอ้างถึงขนาดของคลังข้อมูล (van Heuven et al., 2014) วิธีการคำนวณค่ามีดังนี้

$$\text{Zipf} = \log_{10}(\text{frequency per million word}) + 3.0$$

นอกจากนี้การปรับค่าเป็น Zipf’s value มีผลให้สามารถตัดสินระดับความมากน้อยของความถี่การปรากฏได้ (มีค่า 1-7 โดยระดับคะแนน 1 คือความถี่การปรากฏต่ำและ 7 คือความถี่การปรากฏสูง) เนื่องจากความถี่การปรากฏเป็นสิ่งที่สัมพันธ์กับขนาดของคลังข้อมูล หากใช้ข้อมูลดิบที่ไม่มีการพิจารณาขนาดของคลังข้อมูลร่วมด้วยจะเป็นการยากที่จะตัดสินว่าความถี่การปรากฏจำนวนดังกล่าวถือว่ามากหรือน้อย โดยขนาดคลังข้อมูลภาษาไทยแห่งชาติ (TNC) ที่ใช้อ้างอิงในการปรับค่าคือ 33,394,574 คำ (ข้อมูล ณ วันที่ 5 ธันวาคม 2560)



ภาพที่ 3 ความถี่การปรากฏของข้อมูลดิบ (ซ้าย) และข้อมูลที่ปรับค่าแล้ว (ขวา)

4. ผลการศึกษา

หัวข้อที่ 4 นำเสนอค่าเฉลี่ยของค่าบ่งชี้การกลายเป็นศัพท์ทั้งสามค่าของหน่วยสร้างแต่ประเภท (หัวข้อที่ 4.1 - 4.3) จากนั้นทดสอบความสามารถในการจำแนกระหว่างคำเดี่ยวและวลีตามสมมติฐานข้อที่หนึ่งของงานวิจัย และผลการทดสอบความสัมพันธ์ระหว่างค่าบ่งชี้ทั้งสามตามสมมติฐานข้อที่สองของงานวิจัยนี้ (หัวข้อที่ 4.4)

4.1 ค่าระยะเวลาตอบสนองเฉลี่ยของคำเดี่ยว คำประสม และวลี

ตารางที่ 1 แสดงค่าระยะเวลาตอบสนองเฉลี่ยของคำประเภทต่าง ๆ คำประสมและคำเดี่ยวซึ่งเป็นคำจริง (real word) มีค่าระยะเวลาตอบสนองสั้นกว่าคำสมมติ โดยคำเดี่ยวมีค่าระยะเวลาตอบสนองเฉลี่ยน้อยที่สุดคือ 879.91 มิลลิวินาที แสดงถึงโครงสร้างหน่วยคำและวากยสัมพันธ์ที่เรียบง่ายที่สุดเมื่อเปรียบเทียบระหว่างคำทั้งสามประเภท จากการทดสอบด้วยสถิติการทดสอบที (T-test) พบว่าค่าระยะเวลาตอบสนองของคำเดี่ยวต่างจากคำประสมอย่างมีนัยสำคัญทางสถิติ ($t(455) = 12.48$, $p < .01$) และคำประสมมีค่าระยะเวลาน้อยกว่าคำสมมติอย่างมีนัยสำคัญทางสถิติเช่นกัน ($t(456) = -9.45$, $p < .001$) คำสมมติมีค่าระยะเวลาตอบสนองมากที่สุดคือ 1,375.82 มิลลิวินาทีและมีส่วนเบี่ยงเบนมาตรฐาน (standard deviation) สูงที่สุด ($SD = 555.10$) แสดงถึงการประมวลผลโดยการประกอบความซึ่งใช้เวลาในการประมวลผลมากกว่าคำเดี่ยวที่ประมวลผลจากการจดจำ ผลการศึกษาค่าระยะเวลาตอบสนองของคำประเภทต่าง ๆ ในภาษาไทยเป็นไปตามสมมติฐานและสอดคล้องกับงานวิจัยเกี่ยวกับการประมวลผลคำในภาษาอื่น ๆ

ตารางที่ 1 ค่าระยะเวลาตอบสนองเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคำเดี่ยวคำประสมและคำสมมติ

ค่าระยะเวลาตอบสนอง	ค่าเฉลี่ย (มิลลิวินาที)	ส่วนเบี่ยงเบนมาตรฐาน
คำเดี่ยว	879.91	293.22
คำประสม	991.04	314.55
คำสมมติ	1375.82	555.10

4.2 ระดับความหมายเชิงประกอบเฉลี่ยของคำเดี่ยวคำประสมและวลี

ตารางที่ 2 แสดงระดับความหมายเชิงประกอบเฉลี่ยของคำแต่ละประเภท รูปแบบของระดับความหมายเชิงประกอบสอดคล้องกับผลของค่าระยะเวลาตอบสนองคือคำเดี่ยวมีระดับความหมายเชิงประกอบน้อยที่สุด มีค่าเฉลี่ยของระดับความหมายเชิงประกอบเท่ากับ 7.70 แสดงให้เห็นว่าความหมายของคำเดี่ยวทั้งคำมีความเกี่ยวข้องกับความหมายของแต่ละองค์ประกอบในระดับต่ำมาก นอกจากนี้คำเดี่ยวแสดงระดับความหมายเชิงประกอบที่ต่ำกว่าคำประสมอย่างมีนัยสำคัญทางสถิติ ($t(245) = 17.10$, $p < .001$) วลีมีระดับความหมายเชิงประกอบเฉลี่ยมากที่สุดและแสดงให้เห็นว่าความหมายของวลีมีความเกี่ยวข้องกับความหมายของแต่ละองค์ประกอบในระดับค่อนข้างสูง คำประสมมีระดับความหมายเชิงประกอบเฉลี่ยอยู่กึ่งกลางระหว่างคำเดี่ยวและวลี โดยมีระดับความหมายเชิงประกอบต่างจากคำทั้งสองประเภทอย่างมีนัยสำคัญทางสถิติ ($t(343) = -4.70$, $p < .001$)

ตารางที่ 2 ระดับความหมายเชิงประกอบเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคำเดี่ยวคำประสมและคำสมมติ

ระดับความหมายเชิงประกอบ	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
คำเดี่ยว	7.70	16.80
คำประสม	55.31	33.37
วลี	71.04	28.96

4.3 ความถี่การปรากฏเฉลี่ยของคำเดี่ยวคำประสมและวลี

ตารางที่ 3 แสดงความถี่การปรากฏเฉลี่ยของหน่วยสร้างทั้งสามประเภท คำเดี่ยวมีความถี่การปรากฏมากที่สุดและวลีมีความถี่การปรากฏน้อยที่สุด โดยคำเดี่ยวมีความถี่ปรากฏมากกว่าคำประสมอย่างมีนัยสำคัญทางสถิติ ($t(23) = -10.55$, $p < .001$) และคำประสมมีความถี่การปรากฏมากกว่าวลีอย่างมีนัยสำคัญทางสถิติเช่นกัน ($t(38) = 4.07$, $p < .001$) ผลการศึกษาสอดคล้องกับแนวคิดเรื่องการกลายเป็นศัพท์ โดยคำเดี่ยวซึ่งมีระดับการกลายเป็นศัพท์สูงมีความถี่การปรากฏเฉลี่ยสูงที่สุด และวลีซึ่งมีระดับการกลายเป็นศัพท์ต่ำมีความถี่การปรากฏต่ำที่สุด

ตารางที่ 3 ความถี่การปรากฏเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคำเดี่ยวคำประสมและวลี

ความถี่การใช้	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
คำเดี่ยว	23906.64	13641.89
คำประสม	1964.50	3635.92
วลี	116.77	140.86

นอกจากค่าเฉลี่ยที่ต่างกันของค่าบ่งชี้ต่าง ๆ ของคำทั้งสามประเภทแล้วผู้วิจัยต้องการทราบว่าทั้งสามค่าสามารถบ่งชี้การกลายเป็นศัพท์และจำแนกคำที่ผ่านการกลายเป็นศัพท์มากและคำที่ยังไม่ผ่านการกลายเป็นศัพท์ได้หรือไม่ การทดสอบสถิติการทดสอบที สามารถบอกได้เพียงว่าค่าเฉลี่ยระหว่างทั้งสามกลุ่มแตกต่างกัน แต่ไม่สามารถบอกได้ว่าค่าทั้งสามนี้สามารถทำนายระดับการกลายเป็นศัพท์ของคำได้หรือไม่และทำนายได้ดีเพียงใด ผู้วิจัยจึงทดสอบเพิ่มเติมโดยใช้สถิติถดถอยโลจิสติก (logistic regression) โดยกำหนดตัวแปรตามคือประเภทของคำ (word type) สองประเภท ได้แก่ คำเดี่ยวและวลี คำเดี่ยวเป็นตัวแทนของหน่วยสร้างที่มีระดับการกลายเป็นศัพท์สูงและวลีเป็นตัวแทนของหน่วยสร้างที่มีระดับการกลายเป็นศัพท์ต่ำ ตัวแปรต้น ได้แก่ ค่าระยะเวลาตอบสนอง ระดับความหมายเชิงประกอบเฉลี่ยของทั้งคำ/วลี และความถี่การปรากฏที่ปรับค่าเป็น Zipf's value

ตารางที่ 4 ค่าการประมาณ (estimates) จากสถิติถดถอยโลจิสติก ของค่าบ่งชี้การกลายเป็นคำศัพท์ ประเภทที่ใช้อ้างอิง (reference category) ได้แก่ คำเดี่ยว

ตัวแปรทำนาย	ค่าการประมาณ (estimates)	*p*	Sig.
ระดับความหมายเชิงประกอบ	0.0102162	<.001	***
ค่าระยะเวลาตอบสนอง	3.998e-04	<.001	***
ความถี่การใช้	-2.536e-05	<.001	***

ผลการศึกษาในตารางที่ 4 แสดงให้เห็นว่าค่าระยะเวลาตอบสนอง ความถี่การปรากฏและระดับความหมายเชิงประกอบสามารถจำแนกคำเดี่ยวและวลีออกจากกันได้ โดยวลีมีระดับความหมายเชิงประกอบและค่าระยะเวลาตอบสนองมากกว่าคำเดี่ยว และมีความถี่การปรากฏน้อยกว่าคำเดี่ยว ผลการศึกษาเป็นไปตามสมมติฐานข้อที่ 1 ของงานวิจัยนี้

4.4 ความสัมพันธ์ระหว่างค่าบ่งชี้การกลายเป็นศัพท์ในคำประสม

หัวข้อที่ 4.1 - 4.3 ได้แสดงให้เห็นว่าค่าบ่งชี้เชิงปริมาณทั้งสามค่าที่งานวิจัยนี้เสนอ สามารถใช้เป็นค่าบ่งชี้ระดับการกลายเป็นศัพท์ซึ่งจำแนกหน่วยสร้างที่มีระดับการกลายเป็นศัพท์ต่ำออกจากหน่วยสร้างที่มีระดับการกลายเป็นศัพท์สูงได้ ในหัวข้อที่ 4.4 ผู้วิจัยทดสอบเพิ่มเติมเกี่ยวกับความสัมพันธ์ของค่าบ่งชี้ทั้งสามตามสมมติฐานข้อที่สองของงานวิจัยนี้

ตารางที่ 5 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficients) ของค่าบ่งชี้การกลายเป็นศัพท์ทั้งสามของคำประสม ผลการศึกษาไม่พบความสัมพันธ์อย่างมีนัยสำคัญทางสถิติระหว่างค่าบ่งชี้การกลายเป็นศัพท์ทั้งหมด ยกเว้นระหว่างความถี่การปรากฏของคำประสมทั้งคำและค่าระยะเวลาตอบสนองที่มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติ อย่างไรก็ตาม

ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าดังกล่าวแสดงระดับความสัมพันธ์ในระดับต่ำเท่านั้น (r = -.24, p < .01) ผลการศึกษาสรุปได้ว่าค่าบ่งชี้การกลายเป็นศัพท์ทั้งสามไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติและไม่เป็นไปตามสมมติฐานของงานวิจัยนี้

ตารางที่ 5 ค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficients, r) ของค่าบ่งชี้การกลายเป็นศัพท์ทั้งสามของคำประสม

ค่าบ่งชี้การกลายเป็นศัพท์	r	p-value
ระดับความหมายเชิงประกอบองค์ประกอบ 1 / ค่าระยะเวลาตอบสนอง	.09	.24
ระดับความหมายเชิงประกอบองค์ประกอบ 2 / ค่าระยะเวลาตอบสนอง	-.09	.22
ระดับความหมายเชิงประกอบเฉลี่ยของทั้งคำ / ค่าระยะเวลาตอบสนอง	0	.98
ความถี่การปรากฏของทั้งคำ / ค่าระยะเวลาตอบสนอง	**-.24**	**<.01**
ความถี่การปรากฏขององค์ประกอบ 1 / ค่าระยะเวลาตอบสนอง	-.10	.20
ความถี่การปรากฏขององค์ประกอบ 2 / ค่าระยะเวลาตอบสนอง	-.13	.07
ระดับความหมายเชิงประกอบองค์ประกอบ 1 / ความถี่การปรากฏของทั้งคำ	-.08	.29
ระดับความหมายเชิงประกอบองค์ประกอบ 2 / ความถี่การปรากฏของทั้งคำ	.13	.09
ระดับความหมายเชิงประกอบเฉลี่ยของทั้งคำ / ความถี่การปรากฏของทั้งคำ	.02	.75

นอกจากนี้ผู้วิจัยได้ทดสอบเพิ่มเติมโดยใช้สถิติถดถอยเชิงเส้น (linear regression) เพื่อทดสอบความสัมพันธ์ของปฏิสัมพันธ์ (interactions) ระหว่างค่าบ่งชี้ต่าง ๆ ที่อาจมีต่อกัน โมเดลสถิติถดถอยทั้งหมดดังตารางที่ 6 ผลการศึกษาแสดงให้เห็นว่าโมเดลสถิติถดถอยทั้งสามมีค่าสัมประสิทธิ์การตัดสินใจ (R-squared) ในระดับต่ำทั้งหมด (R^2 < .31, p > .05) แสดงให้เห็นว่าค่าบ่งชี้ต่าง ๆ มีความสามารถในการทำนายค่าบ่งชี้อื่นได้ประมาณไม่เกินร้อยละ 30 ของข้อมูลเท่านั้น จากผลการศึกษาจึงสามารถตีความได้ว่าค่าบ่งชี้ทั้งสามไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ แตกต่างจากสมมติฐานข้อที่สองของงานวิจัยนี้

ตารางที่ 6 ผลการศึกษาสถิติถดถอยของค่าบ่งชี้การกลายเป็นศัพท์และปฏิสัมพันธ์ระหว่างค่าบ่งชี้ทั้งหมด

โมเดล	R-squared	Model p-value
comp ~ zipf * RT	0.306	0.0803
RT ~ zipf * comp	0.245	0.1584
zipf ~ comp * RT	0.153	0.3825

5. อภิปรายผล

ข้อค้นพบของงานวิจัยนี้แสดงให้เห็นว่าค่าระยะเวลาตอบสนอง ระดับความหมายเชิงประกอบและความถี่การปรากฏสามารถจำแนกวลีและคำเดี่ยวที่เป็นตัวแทนของหน่วยสร้างที่ผ่านกระบวนการกลายเป็นศัพท์น้อยที่สุดและมากที่สุดออกจาก

กันได้ ค่าทั้งสามจึงสามารถใช้เป็นค่าเชิงปริมาณที่บ่งชี้ระดับการกลายเป็นศัพท์ของคำในภาษาไทยได้ ผลการวิจัยเป็นไปตามสมมติฐานข้อที่หนึ่งของงานวิจัยนี้ นอกจากนี้ผลการศึกษายังได้แสดงให้เห็นว่าระดับการกลายเป็นศัพท์ของคำเดี่ยว วลี และคำประสมที่ค้นพบด้วยระเบียบวิธีเชิงปริมาณสอดคล้องกับข้อสังเกตของงานวิจัยในอดีตซึ่งเสนอว่าคำเดี่ยว วลี และคำประสมมีการกลายเป็นศัพท์ที่มีลักษณะเป็นแนวต่อเนื่องลดหลั่นเป็นระดับ โดยวลีมีระดับการกลายเป็นศัพท์ต่ำที่สุด คำเดี่ยวมีระดับการกลายเป็นศัพท์สูงที่สุดและคำประสมมีค่าในระดับกึ่งกลางระหว่างหน่วยสร้างทั้งสองประเภทดังที่อัญชลี (Anchalee, 2005) ได้เสนอไว้ในแผนภาพที่ 1

สำหรับลักษณะความสัมพันธ์ระหว่างค่าบ่งชี้การกลายเป็นศัพท์ทั้งสามตามสมมติฐานข้อที่สองนั้น ผลการศึกษาไม่พบว่าค่าบ่งชี้ทั้งสามมีทิศทางของความสัมพันธ์ที่ชัดเจน ไม่เป็นไปตามสมมติฐานที่ตั้งไว้ ข้อค้นพบนี้แสดงให้เห็นว่าค่าบ่งชี้การกลายเป็นศัพท์แต่ละค่านั้นเป็นอิสระจากกัน การเปลี่ยนแปลงแง่มุมต่าง ๆ ในกระบวนการกลายเป็นศัพท์ ทั้งการเปลี่ยนแปลงโครงสร้างหน่วยคำวากยสัมพันธ์ การเปลี่ยนแปลงความหมายและการเปลี่ยนแปลงการใช้มีเส้นทางการเปลี่ยนแปลงของตนเองซึ่งเกิดขึ้นในระยะเวลาและอัตราเร็วที่แตกต่างกันไป คำศัพท์บางคำมีการเปลี่ยนแปลงหน่วยคำและวากยสัมพันธ์มาก ในขณะที่ยังไม่เกิดการเปลี่ยนแปลงในเชิงความหมายมากนัก หรือคำศัพท์บางคำอาจมีการเปลี่ยนแปลงความหมายเกิดขึ้นก่อน จึงมีการเปลี่ยนแปลงในเชิงการใช้ตามมา ผลการศึกษาของงานวิจัยนี้บอกเป็นนัยว่าระดับการกลายเป็นศัพท์ของหน่วยสร้างหนึ่งอาจไม่สามารถแสดงผ่านค่าใดค่าหนึ่งเพียงค่าเดียว แต่จำเป็นต้องศึกษาจากค่าที่บ่งชี้การเปลี่ยนแปลงในหลายแง่มุมประกอบกันจึงสามารถแสดงให้เห็นระดับการกลายเป็นศัพท์ของหน่วยสร้างอย่างครอบคลุมที่สุด

6. สรุป

งานวิจัยนี้นำเสนอระเบียบวิธีเชิงปริมาณในการศึกษาระดับการกลายเป็นศัพท์ของคำในภาษาไทย โดยเสนอว่าระดับการกลายเป็นศัพท์สามารถศึกษาได้จากค่าซึ่งบ่งชี้การเปลี่ยนแปลงแง่มุมต่าง ๆ ที่เกิดขึ้นในกระบวนการกลายเป็นศัพท์ งานวิจัยเสนอค่าบ่งชี้เชิงปริมาณสามค่า ได้แก่ ค่าระยะเวลาตอบสนองซึ่งแสดงระดับการเปลี่ยนแปลงของโครงสร้างหน่วยคำวากยสัมพันธ์ ระดับความหมายเชิงประกอบซึ่งแสดงระดับการเปลี่ยนแปลงของความหมาย และความถี่การปรากฏซึ่งแสดงระดับการเปลี่ยนแปลงเชิงการใช้ ผลการศึกษาพบว่าค่าบ่งชี้ทั้งสามสามารถใช้เป็นค่าบ่งชี้ที่จำแนกระหว่างคำที่มีระดับการกลายเป็นศัพท์ต่างกันได้ อย่างไรก็ตาม การนำเสนอระดับการกลายเป็นศัพท์จำเป็นต้องศึกษาจากค่าบ่งชี้หลายค่าประกอบกัน เนื่องจากผลการศึกษาแสดงให้เห็นว่าค่าบ่งชี้ต่าง ๆ ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติซึ่งบอกเป็นนัยว่าการเปลี่ยนแปลงในแง่มุมต่าง ๆ ของกระบวนการกลายเป็นศัพท์นั้นเป็นอิสระจากกันและไม่ได้เป็นไปในทิศทางเดียวกันเสมอไป การวัดระดับการกลายเป็นศัพท์ของหน่วยสร้างจึงไม่สามารถใช้ค่าใดค่าหนึ่งเพียงค่าเดียวในการบ่งชี้ได้

7. กิตติกรรมประกาศ

บทความนี้เป็นส่วนหนึ่งของวิทยานิพนธ์หลักสูตรอักษรศาสตรดุษฎีบัณฑิต สาขาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ของนางสาวลีนา มะลูลีม หัวข้อวิทยานิพนธ์ “ความสัมพันธ์ระหว่างโครงสร้างสัทสัมพันธ์ของ

คำประสมภาษาไทยกับการกลายเป็นศัพท์” โครงการวิจัยนี้ได้รับทุนสนับสนุนจากโครงการปริญญาเอกกาญจนาภิเษก (คปก.) สำนักงานการวิจัยแห่งชาติ (วช.) ผู้วิจัยขอขอบพระคุณอาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.พิทยาวัฒน์ พิทยาภรณ์ สำหรับคำแนะนำอันเป็นประโยชน์ตลอดการทำวิจัย และขอขอบพระคุณนิสิตชมรมมุสลิมมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน ที่เอื้อเฟื้อสถานที่และอำนวยความสะดวกในการเก็บข้อมูล

ภาคผนวก ก.

รายการคำ

ลำดับ	คำประสม	คำเดี่ยว	วลี
1.	กำปั้น	คาถา	แคบมาก
2.	ขี้ข้า	กางเกง	บินต่ำ
3.	คุ้มกัน	รักษา	ปิดทึบ
4.	งานบ้าน	ระยะ	นอนนาน
5.	ต้นทุน	ประเทศ	วิ่งเร็ว
6.	ต้านทาน	สามารถ	จับแน่น
7.	ท่องจำ	ระหว่าง	พูดช้า
8.	เนื้อตัว	สำหรับ	เต้นเก่ง
9.	น้ำตาล	กำลัง	หยุดนิ่ง
10.	แบ่งปัน	นาที	ตอบถูก
11.	ป้องกัน	กำหนด	ยิ้มกว้าง
12.	ปูม้า	ระบบ	ตื่นเช้า
13.	เพิ่มเติม	มาตรา	กินเยอะ
14.	เพิ่มพูน	อาหาร	ยืนห่าง
15.	พ่อแม่	ระดับ	เดินช้า
16.	พ่อตา	จำนวน	พูดเพราะ
17.	ภูผา	บุคคล	ไปวัด
18.	หน้าที่	อำนาจ	ขายน้ำ
19.	หมอดู	ภาษา	ตากเสื้อ
20.	หมอนข้าง	สำคัญ	หั่นผัก
21.	หมอผี	กระทำ	สร้างบ้าน
22.	หมาป่า	ชะงัก	หยิกแก้ม

ภาคผนวก ข.

ตัวอย่างแบบสอบถามระดับความหมายเชิงประกอบ

แบบสอบถามนี้มี 8 หน้าประกอบด้วยคำถามทั้งหมด 108 คำถาม โดยใช้เวลาในการทำไม่เกิน 20 นาทีกลุ่มเป้าหมายผู้ตอบแบบสอบถามคือผู้พูดภาษาไทยที่ไม่มีประสบการณ์เรียนภาษาศาสตร์ และผู้ที่ศึกษาในสาขาที่ไม่เกี่ยวข้องกับภาษา ข้อมูลส่วนตัวของท่านจะถูกเก็บเป็นความลับและคำตอบของท่านจะถูกใช้เพื่อการวิจัยเท่านั้น ขอบพระคุณทุกท่านที่ให้ความร่วมมือ

---- (หน้า 1) ----

ข้อมูลส่วนตัว

กรุณากรอกข้อมูลตามความเป็นจริง ข้อมูลส่วนตัวของท่านจะถูกเก็บเป็นความลับ

จังหวัดที่เกิด_____ จังหวัดที่อาศัยปัจจุบัน _____

อายุ _____เพศ_____คณะและสาขาที่ศึกษา_____ ชั้นปี _____

เคยย้ายไปอาศัยที่จังหวัดอื่นหรือไม่ (ถ้าเคย กรุณาระบุจังหวัดและระยะเวลาที่ไปอาศัยอยู่) _____

เคยเรียนภาษาไทยถิ่นหรือไม่ (เช่น ไทยถิ่นอีสาน ถิ่นใต้ คำเมือง ฯลฯ ถ้าเคยเรียนโปรดระบุ) _____

เคยเรียนภาษาเขมรและบาลีสันสกฤตหรือไม่ (ถ้าเคยเรียนโปรดระบุ) _____

---- (หน้า 2) ----

คำชี้แจงและตัวอย่าง

แบบสอบถามนี้มีจุดประสงค์เพื่อสอบถามความคิดเห็นของผู้เข้าร่วมวิจัยเกี่ยวกับคำและวลีในภาษาไทยโดยไม่มี การตัดสินคำตอบถูกผิด ขอความกรุณาเลือกตอบตามความคิดเห็นของท่านโดย**ไม่**ค้นหาคำศัพท์จากพจนานุกรมหรือค้นหาความหมายจากช่องทางใด ๆ ทั้งออนไลน์และออฟไลน์ระหว่างทำแบบสอบถาม

ตัวอย่างการตัดสินคะแนน

ตัวอย่าง 1

ความหมายของ "ซักผ้า" เกี่ยวข้องกับความหมายของ "ซัก" มากน้อยแค่ไหน

คำอธิบาย: *"ซักผ้า"* เป็นกริยาแสดงการทำความสะอาดเสื้อผ้าโดยใช้น้ำ มีความหมายเกี่ยวข้องกับ *"ซัก"* ซึ่งแปลว่า *'การทำความสะอาดโดยใช้น้ำ'* และ *"ผ้า"* ความหมายทั้งสองคำมีความเกี่ยวข้องกันมากที่สุดจึงตัดสินให้คะแนนในระดับสูงแสดงถึงความเกี่ยวข้องอย่างมาก

ตัวอย่าง 2

ความหมายของ"จิงโจ้" เกี่ยวข้องกับความหมายของ "จิง" มากน้อยแค่ไหน

คำอธิบาย: ในข้อนี้คำว่า *"จิงโจ้"* ซึ่งมีความหมายว่าสัตว์เลี้ยงลูกด้วยนมชนิดหนึ่ง ไม่เกี่ยวข้องกับ *"จิง"* ซึ่งไม่มีความหมายในภาษาไทย จึงตัดสินให้คะแนนความเกี่ยวข้องในระดับต่ำมากแสดงถึงความหมายไม่เกี่ยวข้องเลย

ตัวอย่าง 3

ความหมายของ “แม่ค้า” เกี่ยวข้องกับความหมายของ “แม่” มากน้อยแค่ไหน

คำอธิบาย: *“แม่ค้า”* มีความหมายว่าผู้ทำอาชีพค้าขายและเป็นเพศหญิง เกี่ยวข้องกับ “แม่” ในแง่ที่มีความหมายถึงเพศหญิงเหมือนกัน แต่ไม่ได้เกี่ยวข้องอย่างที่สุดเนื่องจาก *“แม่”* ในคำว่า *“แม่ค้า”* ไม่มีนัยยะของการเป็นมารดาจึงตัดสินให้คะแนนระดับค่อนข้างสูง แสดงถึงความหมายเกี่ยวข้องบางส่วน

ในแบบสอบถามนี้ประกอบด้วยคำที่มีความหมายเกี่ยวข้องและไม่เกี่ยวข้องกัน ให้ท่านตัดสินคะแนนมากน้อยตามความเห็นของท่าน ท่านสามารถให้คะแนนมาก ให้คะแนนน้อย หรือให้คะแนนในระดับกลางก็ได้เช่นกัน หากมีคำถามเพิ่มเติมสามารถสอบถามผู้วิจัยได้ตลอดการทำแบบสอบถาม

รายการอ้างอิง

ภาษาไทย

Alisa Injan อลิษา อินจันทร์. (2014). *Kham prasom baep thaothiam nai phasa Thai* คำประสมแบบเท่าเทียมในภาษาไทย [Coordinate compounds in Thai] [Master's thesis, Chulalongkorn University]. CUIR. https://cuir.car.chula.ac.th/handle/123456789/46001

Anchalee Singnoi อัญชลี สิงห์น้อย. (2005). *Kham nam prasom sat lae sin nai kan sang kham Thai* คำนามประสม ศาสตร์และศิลป์ในการสร้างคำไทย [Compound Nouns: Sciences and Arts in Thai Word Formation]. Chulalongkorn University Press.

Anong Iangubol อนงค์ เอียงอุบล. (1982). *Kan seuksa cherng wi kro kham prasom nai phasa Thai* การศึกษาเชิงวิเคราะห์คำประสมในภาษาไทย [An analytical study of compound words in Thai]. [Master's thesis, Chulalongkorn University]. CUIR. https://cuir.car.chula.ac.th/handle/123456789/27665

Bhimbasistha Tejarajanya ภีมพสิษฐ์ เตชะราชันย์. (2015). *Krabuankan klai pen kham waiyakorn khong tuabongchi namwaliplaeng kan lae khwam nai phasa Thai* กระบวนการกลายเป็นคำไวยากรณ์ของตัวบ่งชี้นามวลีแปลง การ และ ความ ในภาษาไทย [Grammaticalzation of KAAN and Khwaam nominalizers in Thai] [Doctoral dissertation, Chulalongkorn University]. CUIR. https://cuir.car.chula.ac.th/handle/123456789/50145

Duangchan Intorn ดวงจันทร์ อินทร. (1976). *Kan sang bot-rian baep prokraem wicha phasa Thai rueang kham prasom samrap chan matthayom sueksa pi thi nueng* การสร้างบทเรียนแบบโปรแกรมวิชาภาษาไทย เรื่อง คำประสม สำหรับชั้นมัธยมศึกษาปีที่หนึ่ง [Construction of a Thai programmed lesson on "compound words" for Mathayom suksa one] [Unpublished master's thesis], Chulalongkorn University.

Pranee Kullavanich ปราณี กุลละวณิชย์. (2001). *Kan plian plaeng khong phasa* การเปลี่ยนแปลงของภาษา [Language change]. In *Ekkasan prakop kan son chut wicha phasa Thai* เอกสารประกอบการสอนชุดวิชาภาษาไทย 3 [Thai coursebook] (Vol. 3, pp. 370-399).÷ Sukhothai Thammathirat Open University Press.

Vipas Pothipath วิภาส โพธิแพทย์. (2014). *Krabuankan koet kham mai nai phasa Thai* กระบวนการเกิดคำใหม่ในภาษาไทย [Lexicalization in Thai]. *Warasarn phasa lae phasa sat* วารสารภาษาและภาษาศาสตร์ [Journal of Language and Linguistics]*, 32*(2), 1-24.

Yordkwan Kaewkhieo ยอดขวัญ แก้วเขียว. (2008). *Kan sueksa priaptiap kham prasom phasa Thai mattrathan kap phasa Thai thin nuea* การศึกษาเปรียบเทียบคำประสมภาษาไทยมาตรฐานกับภาษาไทยถิ่นเหนือ [A comparative study of compound words in Standard Thai and Northern Thai Dialect]. *Warasarn phasa Thai lae watthanatham Thai* วารสารภาษาไทยและวัฒนธรรมไทย [Journal of Thai Language and Thai Culture]*, 2*(3), 139-152.

ภาษาต่างประเทศ

Aroonmanakun, W. (2007). Creating the Thai national corpus. *MANUSYA: Journal of Humanities, 10*(3), 4-17. https://doi.org/10.1163/26659077-01003001

Arunachalam, S. (2013). Experimental methods for linguists. *Language and Linguistics Compass, 7*(4), 221-232. https://doi.org/10.1111/lnc3.12021

Blank, A. (2001). Pathways of lexicalization. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Eds.), Language typology and language universals (Vol. 2, pp. 1596-1608). Waiter de Gruyter.

Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press.

Bussman, H. (1996). Routledge dictionary of linguistics and language [G. Trauth & K. Kazzazi, Eds. & Trans.] Routledge. ~~London.~~

Civilleri, G. O. (2013). Degrees of lexicalization in Ancient Greek deverbal nouns. In P. ten Hacken & C. Thomas (Eds.), *The semantics of word formation and lexicalization* (pp. 203-224). Edinburgh University Press.

Harris, A. C., & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective.* Cambridge University Press.

Haspelmath, M. & Sims, A. (2010). *Understanding morphology* (2nd ed.). Routledge.

Lessau, D. A. (1994). *A dictionary of grammaticalization* (Vol. 1). Universitätsverlag Dr. N. Brockmeyer.

Libben, G. (2007). Why study compound processing? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 1-22). Oxford Academic. https://doi.org/10.1093/acprof:oso/9780199228911.003.0001

Lipka, L. (1992). Lexicalization and institutionalization in English and German. *Linguistica Pragensia/Akademie Ved CR, Ústav pro Jazyk Ceský, 2*, 1-13. https://doi.org/10.5282/ubm/epub.5105

Lipka, L. (2002). *English lexicology: Lexical structure, word semantics & word-formation*. Gunter Narr Verlag.

Lipka, L., Handl, S., & Falkner, W. (1994). Lexicalization and institutionalization. In R. E. Asher & J. M. Y. Simpson (Eds.), *The encyclopedia of language and linguistics* (Vol. 4, pp. 2164-2167). Pergamon Press.

Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language, 84*(4), 760-794.

Reddy, S., McCarthy, D., & Manandhar, S. (2011). *An Empirical Study on Compositionality in Compound Nouns. In Proceeding of the 5th international joint conference on natural language processing* (pp. 210-218). Asian Federation of Natural Language Processing.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime* (Version 2) [Computer software]. Psychology Software Incorporated. https://pstnet.com/products/e-prime/

Soskuthy, M. & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition,* 166, 298-313. https://doi.org/10.1016/j.cognition.2017.05.032

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176-1190. https://doi.org/10.1080/17470218.2013.850521

van Jaarsveld, H. J., & Rattink, G. E. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research, 17*, 447-473. https://doi.org/10.1007/BF01067911

Xia, Q., Gong, W. & Ly, Y. (2016, August 8-10). *Syntactic category of constituent components in the processing of compounds: evidence from noun compounds in Mandarin* [Conference session]. The 8th International Conference in Evolutionary Linguistics, Indiana University, Bloomington, Indiana, USA.

Xia, Q., Wang, T., Wang, J., Meng, Z., & Yu, M. (2022). To combine is to divide: Evidence of the noun-verb distinction at the morphemic level in compound nouns. *Lingua, 271*, 103239. https://doi.org/10.1016/j.lingua.2021.103239