

การศึกษาวิเคราะห์คำไทยที่มักเขียนผิด

พลวัฒน์ ไหลมหนู*

วิโรจน์ อรุณมานะกุล**

บทคัดย่อ

การสะกดคำผิดอาจทำให้เกิดคำผิดที่ไม่เป็นคำหรือเป็นคำที่ไม่ต้องการสื่อ ซึ่งการสะกดผิดแบบไม่เป็นคำสามารถตรวจแก้ได้ง่ายกว่าด้วยวิธีการเทียบจากรายการคำ เนื่องจากจะไม่พบในพจนานุกรม แต่การตรวจจับคำสะกดผิดที่เป็นคำจริงเป็นงานที่ยากกว่าเพราะไม่สามารถตรวจแก้ด้วยการเทียบจากรายการคำได้โดยตรง การศึกษาคำนี้ มีวัตถุประสงค์เพื่อศึกษาวิเคราะห์คำที่มักเขียนผิดในภาษาไทยที่สะกดผิดแบบเป็นคำจริง ซึ่งได้มาจากการรวบรวมคำจากหนังสือคำไทยที่มักเขียนผิดแล้วคัดเลือกเฉพาะคำสะกดผิดที่เป็นคำจริง ได้จำนวน 1,674 คำ จากนั้นจึงศึกษาวิเคราะห์คำที่มักเขียนผิดที่รวบรวมไว้ว่ามีรูปแบบหรือลักษณะของการสะกดผิดเป็นอย่างไร โดยการเปรียบเทียบลักษณะของคำที่สะกดผิดนั้นว่าเปลี่ยนแปลงไปจากรูปการสะกดที่ถูกต้องอย่างไรบ้าง ซึ่งจากผลการวิเคราะห์ระบุว่าคำที่มักเขียนผิดและเป็นการสะกดผิดแบบเป็นคำจริงที่นำมาศึกษาในครั้งนี้สามารถจำแนกได้ตามลักษณะของความแตกต่างระหว่างคำที่สะกดผิดกับคำที่สะกดถูกต้องได้เป็น 10 ลักษณะ

* นิสิตระดับมหาบัณฑิต ภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ติดต่อได้ที่: adamas22pl@gmail.com

** รองศาสตราจารย์ประจำภาควิชาภาษาศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ติดต่อได้ที่: awrote@gmail.com

คำสำคัญ: คำที่มักเขียนผิด; การสะกดคำ; การสะกดคำผิด; คำที่สะกดผิด;
การสะกดคำผิดแบบเป็นคำจริง

An Analysis of Commonly Misspelled Thai Words

Ponlawat Laimanoo*

Wirote Aroonmanakun**

Abstract

Misspelling can cause either a non-word error or an error which is a real word but not the intended one. A non-word misspelling error is easier to detect by comparing with a word list in a dictionary, while a real-word misspelling error cannot be detected with this method. This research aimed to analyze 1,674 real-word errors of commonly misspelled Thai words which were collected and analyzed from all examples of spelling errors found in books on commonly misspelled words in Thai. By analyzing the data by comparing each real-word error with the intended word in each case, it was found that Thai real-word spelling errors can be grouped into 10 different categories.

Keywords: misspelled words; real-word errors; spelling errors;
real-word spelling errors

* Masters student, Department of Linguistics, Faculty of Arts, Chulalongkorn University. e-mail: adamas22pl@gmail.com

** Associate Professor, Department of Linguistics, Faculty of Arts, Chulalongkorn University. e-mail: awirote@gmail.com

1. บทนำ

การเขียนนั้นถือเป็นวิธีการสื่อสารที่สำคัญวิธีการหนึ่งที่มีมนุษย์ใช้เพื่อติดต่อสื่อสารกัน และเป็นวิธีการหนึ่งที่สามารถใช้เพื่อถ่ายทอดความรู้ ความคิด และประสบการณ์ระหว่างผู้สื่อสารและผู้รับสาร อีกทั้งยังเป็นวิธีการสำคัญวิธีการหนึ่งที่มีมนุษย์ใช้เพื่อบันทึกข้อมูลต่างๆ ไม่ว่าจะเป็นความรู้ ประสบการณ์ หรือเหตุการณ์ต่างๆ เก็บไว้เพื่อเป็นหลักฐานหรือเพื่อให้คนรุ่นหลังได้ศึกษา (เสนีย์ วิจารณ์, 2544) ในปัจจุบันเทคโนโลยีทางด้านโทรคมนาคมและการสื่อสารก็ได้พัฒนาไปอย่างก้าวไกล ทำให้จำนวนช่องทางและความสะดวกในการติดต่อสื่อสารมีเพิ่มมากขึ้น และหนึ่งในช่องทางการติดต่อสื่อสารที่กำลังได้รับความนิยมเป็นอย่างมากในขณะนี้ก็คือ การติดต่อสื่อสารผ่านทางโลกออนไลน์หรือโซเชียลมีเดียและแพลตฟอร์มต่างๆ เช่น เฟสบุ๊ก อินสตาแกรม หรือเว็บพันทิป เป็นต้น ซึ่งเป็นวิธีการติดต่อสื่อสารที่สะดวกและรวดเร็ว อีกทั้งยังไม่จำกัดคุณลักษณะและวุฒิการศึกษาของผู้ใช้ ทำให้ประชาชนทุกผู้ทุกวัยสามารถเข้าถึงและใช้งานผ่านช่องทางการสื่อสารประเภทนี้ได้ซึ่งถือเป็นข้อดี แต่ด้วยจำนวนผู้ใช้ที่มหาศาลและความต้องการด้านความสะดวกรวดเร็วในการสื่อสารของผู้ใช้บริการนั้นทำให้ข้อมูลหรือข้อความที่เขียนนั้นไม่ได้ผ่านการตรวจสอบความถูกต้องและอาจมีข้อผิดพลาดเกิดขึ้น เช่น การสะกดคำผิด ซึ่งถ้าหากคำที่เขียนผิดนั้นเป็นการสะกดผิดแบบไม่เป็นคำ เช่น คำว่า *สังเกต* เขียนผิดเป็น *สังเกตุ* โปรแกรมตรวจสอบความถูกต้องของภาษาที่ติดตั้งในคอมพิวเตอร์หรือโทรศัพท์สมาร์ทโฟนก็อาจจะช่วยแก้ไขคำผิดนั้นให้อัตโนมัติหรือทำเครื่องหมายแจ้งเตือนว่ามีคำที่สะกดผิดอยู่ในข้อความ แต่คำผิดแบบที่ตรวจพบได้ยากกว่าเป็นคำผิดที่สะกดผิดแบบเป็นคำจริง เช่น สะกดคำว่า *ผู้ชาย* ผิดเป็น *ผู้ชาย* ในข้อความ *ช่วยแนะนำผู้ชายให้ด้วยค่ะ จะเอาไปให้แฟนเป็นของขวัญปีใหม่¹* ซึ่งคำผิดในข้อความนี้เป็นคำที่มีความหมายและปรากฏในพจนานุกรม ถ้าหากว่าในข้อความที่เขียนมีคำที่สะกดผิดแบบเป็นคำจริงนี้ปนอยู่ก็อาจจะส่งผลกระทบต่อความหมายของข้อความที่ต้องการจะสื่อ ทำให้สื่อความหมาย

¹ นำประโยคข้อความนี้มาจาก <https://pantip.com/topic/32504798>

ผิดเพี้ยนไปได้ และนอกจากนี้อาจจะส่งผลให้ผู้หนึ่งจดจำหรือเลียนแบบรูปแบบการเขียนที่ไม่ถูกต้องนี้ต่อกันไปเรื่อยๆ ได้

ด้วยเหตุนี้ ผู้วิจัยจึงต้องการศึกษาวิเคราะห์ลักษณะของคำไทยที่มักเขียนผิด และเป็นการสะกดผิดแบบเป็นคำจริง เพื่อที่จะสามารถนำผลการศึกษามาเป็นข้อมูลพื้นฐานสำหรับระบบการตรวจแก้การเขียนสะกดผิดในภาษาไทยได้

2. ประเภทของคำที่สะกดผิด

ในการตรวจจับคำสะกดผิดนั้น ในทางคอมพิวเตอร์สามารถมองคำที่สะกดผิดออกเป็นสองประเภท ประเภทแรกคือคำที่สะกดผิดแล้วไม่เป็นคำ คือไม่พบคำๆ นั้น ปรากฏในพจนานุกรม ส่วนประเภทที่สองคือคำที่สะกดผิดแล้วกลายเป็นอีกคำหนึ่ง ซึ่งไม่ควรใช้ในบริบทนั้น ตัวอย่างเช่น *ลูกชายผมสอวขบมีไซ่สูงมากให้กินยาลดไซ่ไตใหม่* ข้อความนี้มีคำที่สะกดผิดอยู่สองคำซึ่งเป็นการสะกดผิดในลักษณะที่แตกต่างกัน คำแรกคือคำว่า *สอว* เป็นคำผิดที่เกิดจากการสะกดผิดแบบไม่เป็นคำ เนื่องจากคำนี้ไม่มีปรากฏอยู่ในพจนานุกรมและไม่มี ความหมาย ซึ่งคำที่ถูกต้องคือ *สอง* ส่วนคำที่สะกดผิดคำที่สองในข้อความนี้คือคำว่า *ไซ่* เป็นคำผิดที่เกิดจากความผิดพลาดที่มีความซับซ้อนกว่าคำผิดคำแรก ซึ่งถ้าดูแบบผิวเผินโดยไม่สนใจคำบริบทข้างเคียง จะไม่สามารถทราบได้ว่าคำว่า *ไซ่* นั้นเป็นคำที่สะกดผิดเพราะว่า *ไซ่* เป็นคำที่ปรากฏอยู่ในพจนานุกรมและเป็นคำที่มีความหมาย แต่ในประโยคนี้ *ไซ่* ถือเป็นคำที่เกิดจากการสะกดผิดเป็นคำอื่น คือ คำที่สะกดผิดนั้นยังคงเป็นคำที่ปรากฏในพจนานุกรม แต่เมื่อดูจากคำแวดล้อมอื่นๆ เช่น คำบริบท *กินยาลดไซ่* และคำข้างเคียง *สูง* จะเห็นได้ว่าคำว่า *ไซ่* เมื่อปรากฏกับคำเหล่านี้แล้วทำให้ความหมายของประโยคนั้นแปลกประหลาด ทำให้ทราบได้ว่า *ไซ่* เป็นคำที่สะกดผิด แต่คำที่ถูกต้องคือคำว่า *ไซ่* ซึ่งมีความหมายว่า ‘อาการที่มีอุณหภูมิของร่างกายผิดจากระดับปรกติเนื่องจากความเจ็บป่วย’ จึงจะเหมาะสมกับคำบริบทและคำข้างเคียง

จากตัวอย่างข้างต้นจะเห็นได้ว่า คำที่สะกดผิดนั้นสามารถแบ่งออกได้เป็น 2 ประเภท (Mishra & Kaur, 2013; Mitton, 1987; Murthy S et al., 2012; Verberne, 2002) ซึ่งได้แก่ คำผิดที่เกิดจากการสะกดผิดแบบไม่เป็นคำและคำผิดที่เกิดจากการสะกดผิดแบบเป็นคำจริง

2.1 คำผิดที่เกิดจากการสะกดผิดแบบไม่เป็นคำ (non-word spelling errors)

การสะกดผิดแบบไม่เป็นคำ คือ การนำตัวอักษรของพยัญชนะ สระ และวรรณยุกต์มาประกอบเข้าด้วยกันเป็นคำแต่คำที่ได้นี้ไม่มีความหมายและไม่ปรากฏในพจนานุกรม ซึ่งอาจจะมีสาเหตุมาจากการเติม (insertion) การลบ (deletion) หรือการแทนที่ (substitution) ตัวอักษรผิดพลาดไปจากคำที่ถูกต้องตั้งแต่หนึ่งตัวขึ้นไป (Kaur & Garg, 2014; Stehouwer, 2011) เช่น

ประสบ → ประสนบ เป็นความผิดพลาดที่เกิดจาก การเติม <น>
 ประสบ → ประส เป็นความผิดพลาดที่เกิดจาก การลบ <บ>
 ประสบ → ปรัสบ เป็นความผิดพลาดที่เกิดจาก การแทนที่ <ะ>
 ด้วย <ั>

2.2 คำผิดที่เกิดจากการสะกดผิดแบบเป็นคำจริง (real-word spelling errors)

การสะกดผิดแบบเป็นคำจริง คือ การประกอบตัวอักษรของพยัญชนะ สระ และวรรณยุกต์เข้าด้วยกันแล้วเกิดเป็นคำอื่นที่มีความหมายและปรากฏอยู่ในพจนานุกรม แต่คำนั้นไม่ใช่คำที่ผู้เขียนตั้งใจจะเขียน (Wilcox-O'Hearn, Hirst, & Budanitsky, 2008) จึงถือเป็นการสะกดผิดอีกประเภทหนึ่ง โดยคำที่สะกดผิดนี้อาจมีผลทำให้โครงสร้างหรือความหมายของประโยคผิดเพี้ยนไป (Mitton, 1987) ซึ่งสาเหตุของความผิดพลาดที่เกิดขึ้นนั้นอาจเกิดจากสาเหตุเดียวกันกับการสะกดผิดประเภทแรกด้วยเช่นกัน ซึ่งเกิดจากการเติม การลบ หรือการแทนที่ตัวอักษรตั้งแต่หนึ่งตัวขึ้นไป ผิดพลาด เช่น

- มองนก → มตองนก เป็นความผิดพลาดที่เกิดจาก การเติม <ต>
 มองนก → องนก เป็นความผิดพลาดที่เกิดจาก การลบ <ม>
 มองนก → ลองนก เป็นความผิดพลาดที่เกิดจาก การแทนที่ <ม>
 ด้วย <ล>

จะเห็นได้ว่าการสะกดคำผิดแบบแรกนั้น โปรแกรมตรวจแก้คำผิดสามารถตรวจจับได้ง่ายเพราะใช้วิธีการเทียบหาค่าในพจนานุกรมได้ว่า หากข้อความนั้นมีคำที่สะกดผิดแล้วไม่เป็นคำที่รู้จักก็น่าจะมีข้อผิดพลาดเรื่องการสะกดคำอยู่ แต่การสะกดผิดแบบเป็นคำจริงนั้นเป็นปัญหาที่ทำนายต่อการตรวจจับด้วยคอมพิวเตอร์ เนื่องจากคำที่สะกดผิดก็ยังเป็นคำที่มีใช้อยู่จริง ไม่สามารถตรวจพบได้ด้วยการเทียบหารายการคำในพจนานุกรม แต่ต้องอาศัยบริบทข้างเคียงเพื่อพิจารณาว่าคำนั้นๆ เป็นคำที่เหมาะสมหรือไม่ หากไม่ใช่ก็มีความเป็นไปได้สูงที่จะสะกดผิด

นอกจากโปรแกรมตรวจแก้คำผิดจะสามารถช่วยตรวจจับคำที่สะกดผิดแล้วยังต้องสามารถแก้ไขหรือเสนอคำที่ถูกต้องได้ด้วย ซึ่งวิธีการพื้นฐานที่มักถูกนำมาใช้ในการแก้ไขคำที่สะกดผิดด้วยคอมพิวเตอร์นั้นก็คือ วิธีการที่มีชื่อว่า Minimum Edit Distance (Jurafsky, 2016) ซึ่งเป็นการปรับแก้คำที่สะกดผิดด้วยการเติม การลบ หรือการแทนที่ ทำให้คำที่สะกดผิดกลายเป็นคำที่สะกดถูกต้อง ซึ่งหัวใจสำคัญของวิธีการนี้ก็คือการแก้ไขคำที่สะกดผิดให้ถูกต้องด้วยจำนวนการปรับแก้ที่น้อยที่สุด

3. การศึกษาเรื่องการเขียนคำผิด

การศึกษาเรื่องการเขียนคำไทยผิดมักพบในสองกลุ่มใหญ่ กลุ่มแรกสนใจสาเหตุของการเขียนผิด กลุ่มสองสนใจลักษณะหรือรูปแบบของข้อผิดพลาดที่พบในงานเขียน การศึกษากลุ่มแรก เช่น กุหลาบ มัลลิกะมาส และวิพุธ โสภางค์ (2536) ระบุสาเหตุของการสะกดผิดของนักเรียนชั้นมัธยมศึกษาตอนปลายไว้ว่ามาจากสาเหตุหลักสี่อย่าง คือ การออกเสียงผิด การใช้แนวเทียบผิด ความไม่เข้าใจในความหมายของคำพ้องเสียง แต่สะคำ และการขาดความรู้ในการใช้วิสรรชนีย์ นอกจากนี้ จากผลงานวิจัยของคณาจารย์ภาควิชาภาษาไทยเพื่อการสื่อสาร คณะมนุษยศาสตร์ (2541) พบว่า การสะกดผิดนั้น

อาจเกิดขึ้นได้หากผู้ใช้ภาษามีความรู้เรื่องการใช้ภาษาไม่ดีพอและไม่คำนึงถึงหลักการ
ใช้ภาษาให้ถูกต้อง ทำให้เกิดความผิดพลาดในการสื่อสารหรือสื่อความหมายผิดเพี้ยน
ไป เป็นการสื่อสารที่ขาดประสิทธิภาพ

ประกาศรี สีหอำไพ และคณะ (2544) ได้สรุปสาเหตุต่างๆ ของการสะกดผิด
เอาไว้ในหนังสือประกอบการเรียนการสอนรายวิชาภาษาไทย 4 สำหรับคณะครุศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย ในทิศทางคล้ายกัน คือ เกิดจากการออกเสียงผิด แล้วเขียน
ตามทีออกเสียง เกิดจากการใช้แนวเทียบผิด โดยเทียบว่าคำที่ออกเสียงเหมือนกันก็
เขียนเหมือนกัน เกิดจากการขาดความรู้เรื่องการสะกดคำที่ถูกต้อง เมื่อเห็นคำที่สะกด
ผิดก็ไม่รู้ว่าผิด และเกิดจากความเคยชินและไม่สนใจที่จะเขียนให้ถูกต้อง

สำหรับกลุ่มที่สองซึ่งสนใจลักษณะของข้อผิดพลาดที่พบในงานเขียน โดยทั่วไป
จะไม่ได้จำกัดเฉพาะเรื่องการสะกดคำผิด แต่รวมถึงการใช้เครื่องหมายผิด ใช้คำผิด
หรือไม่เหมาะสมด้วย เช่น ชูศรี ศรีแก้ว (2527) ได้ศึกษาข้อผิดพลาดในการเขียน
ภาษาไทยของนักศึกษาชั้นปีที่ 1 มหาวิทยาลัยเชียงใหม่ ปีการศึกษา 2525 ซึ่งจาก
ผลการศึกษาปรากฏว่านักศึกษาส่วนใหญ่มักจะเขียนตัวสะกดการันต์ผิดบ่อยที่สุด
รองลงมาเป็นข้อผิดพลาดในการใช้คำ ข้อผิดพลาดในการแต่งประโยค ข้อผิดพลาดใน
การใช้เครื่องหมาย และข้อผิดพลาดที่พบน้อยที่สุดในการศึกษาคั้งนี้คือ ข้อผิดพลาด
ในการใช้อักษรย่อ

พยอม ธนมี (2541) ได้ศึกษาเกี่ยวกับลักษณะข้อผิดพลาดในการเขียน
ภาษาไทยของนักศึกษาระดับประกาศนียบัตรวิชาชีพชั้นสูง ปีที่ 1 ในวิทยาลัยเทคนิค
นราธิวาส ปีการศึกษา 2539 ซึ่งผลการศึกษาพบข้อผิดพลาดทั้งหมด 651 ครั้งซึ่ง
จำแนกตามลักษณะข้อผิดพลาดในการเขียนของนักศึกษากลุ่มดังกล่าวทั้งหมด 5
ลักษณะ ดังนี้

- ก. ข้อผิดพลาดในการใช้คำ 226 ครั้ง (34.72%)
- ข. การใช้เครื่องหมายวรรคตอน 142 ครั้ง (21.81%)
- ค. การใช้รูปแบบการเขียน 103 ครั้ง (15.82%)
- ง. การใช้ถ้อยคำและสำนวน 102 ครั้ง (15.67%)
- จ. การใช้โครงสร้างของประโยค 78 ครั้ง (11.98%)

นอกจากนี้ยังพบว่า เกือบครึ่งหนึ่งหรือ 107 ครั้งจาก 226 ครั้ง ของข้อผิดพลาดในการใช้คำนั้นมีสาเหตุจากการสะกดคำไม่ถูกต้อง คือใช้พยัญชนะต้น สระ วรรณยุกต์ การันต์ และพยัญชนะสะกดไม่ตรงตามมาตรฐานตัวสะกดและหลักภาษาไทย

จากงานที่กล่าวมา จะเห็นว่าความสนใจเรื่องการศึกษาการสะกดคำผิด โดยเฉพาะยังมีไม่มาก ส่วนใหญ่สนใจการเขียนผิดในภาพรวมและสนใจที่สาเหตุการเขียนผิด จะมีที่วิเคราะห์คำที่เขียนผิดบ้าง แต่ก็ศึกษาผ่านข้อมูลที่มีขนาดไม่ใหญ่มากนัก จึงอาจจะยังไม่สามารถนำมาอธิบายถึงภาพรวมของข้อผิดพลาดในการเขียนของผู้ใช้ภาษาไทยได้ นอกจากนี้ผู้วิจัยยังไม่พบว่ามีงานวิจัยงานใด นักวิจัยหรือนักวิชาการท่านใดที่ศึกษาวิเคราะห์คำไทยที่มักเขียนผิดประเภทคำที่สะกดผิดแบบเป็นคำจริง ซึ่งจะเป็นพื้นฐานความรู้ที่สำคัญสำหรับการพัฒนาโปรแกรมแก้ไขคำผิดในภาษาไทยด้วยคอมพิวเตอร์ได้ ด้วยเหตุนี้ผู้วิจัยจึงได้ออกแบบการศึกษาวิเคราะห์คำที่มักเขียนผิดแล้วยังคงเป็นคำจริงในครั้งนี้นั้น เพื่อที่จะวิเคราะห์ข้อผิดพลาดที่เกิดจากการสะกดผิดแบบเป็นคำและนำผลที่ได้ไปประยุกต์ใช้ให้เกิดประโยชน์ต่อไป

4. ขอบเขตการวิจัย

การศึกษาครั้งนี้ศึกษาและวิเคราะห์เฉพาะคำที่สะกดผิดแบบเป็นคำจริงเท่านั้น เพราะฉะนั้นข้อมูลที่น่ามาใช้วิเคราะห์ในการศึกษาครั้งนี้เป็นคลังคำที่สะกดผิดแบบเป็นคำจริงที่มีจำนวนคำอยู่ทั้งหมด 1,674 คำ ที่ผู้วิจัยได้รวบรวมเอาเฉพาะคำที่สะกดผิดแบบเป็นคำจริงจากหนังสือคำไทยที่มักเขียนผิดทั้งหมด 5 เล่ม ได้แก่ *อ่านอย่างไรและเขียนอย่างไร ฉบับราชบัณฑิตยสถาน (แก้ไขเพิ่มเติม) พิมพ์ครั้งที่ 22* (ราชบัณฑิตยสถาน, 2557), *280 คำไทยที่มักเขียนและใช้กันผิด* (दनัย เมธิตานนท์, 2549), *ร้อยแปด (108) คำที่มักเขียนผิด* (ธนู ทดแทนคุณ, 2550), *คำไทยที่มักใช้ผิด* (ตระการ เอี่ยมตระกูล, 2554) และ *ภาษาไทย คำที่มักเขียนผิด* (ฝ่ายวิชาการ พีบีซี, 2553) ที่ภายในหนังสือดังกล่าวทุกเล่มบันทึกคำภาษาไทยที่มักเขียนผิดรวมถึงคำที่เขียนถูกต้องเอาไว้

5. วิธีดำเนินการวิจัย

5.1 การเตรียมคลังคำที่สะกดผิดแบบเป็นคำจริง

เนื่องจากการศึกษาครั้งนี้ศึกษาและวิเคราะห์คำไทยที่มักเขียนผิดเฉพาะคำที่มีการสะกดผิดแบบเป็นคำจริงโดยรวบรวมจากหนังสือคำไทยที่มักเขียนผิด 5 เล่ม ดังที่ระบุไว้ข้างต้น แต่หนังสือคำที่มักเขียนผิดทั้ง 5 เล่มนี้ไม่ได้แยกประเภทของคำที่สะกดผิดแบบไม่เป็นคำและคำที่สะกดผิดแบบเป็นคำจริงเอาไว้ ดังนั้นผู้วิจัยจึงนำรายการคำที่มักเขียนผิดทั้งหมดที่พบในหนังสือทั้ง 5 เล่มมาบันทึกรูปคำที่ถูกและรูปคำต่างๆ ที่มักเขียนผิดไป จากนั้นจึงนำข้อมูลคำเหล่านั้นไปผ่านกระบวนการตัดคำด้วยโปรแกรมตัดคำ Thai Segmentation (Aroonmanakun, 2002) แล้วคัดเลือกเฉพาะคำที่เขียนผิดที่ผ่านกระบวนการตัดคำนี้ คำเหล่านั้นถือเป็นคำสะกดผิดที่เป็นคำจริง เพราะคำที่มีความหมายและปรากฏในพจนานุกรมเท่านั้นที่โปรแกรมจะสามารถตัดคำได้สำเร็จ หลังจากที่ได้ผ่านกระบวนการตัดคำแล้วผู้วิจัยก็คัดเลือกเอาเฉพาะคำที่ผ่านการตัดคำสำเร็จเท่านั้น ซึ่งมีจำนวนทั้งหมด 1,674 คำ ที่จะนำมาศึกษาวิเคราะห์ในครั้งนี้ โดยคำในคลังคำที่สะกดผิดแบบเป็นคำจริงมีลักษณะดังตัวอย่าง

คำที่สะกดถูกต้อง	คำที่สะกดผิด	คำที่สะกดถูกต้อง	คำที่สะกดผิด
กากบาท	กา กะ บาท	ทูล หัว	ทูล หัว
อุดมการณ์	อุ ดม การ	กั ม พู ชา	กั พู ชา
บนบาน	บน บ าล	ตะ กร้อ	ตะ กร้อ

โดยทั่วไป คำที่สะกดผิด โปรแกรมตัดคำจะเข้าใจว่าเป็นคำหลายคำมาประกอบกัน เช่น กา|กะ|บาท ซึ่งแต่ละส่วนสามารถเป็นคำได้ หลังจากที่ได้คลังคำที่สะกดผิดแบบเป็นคำจริงมาเรียบร้อยแล้วก็นำไปศึกษาวิเคราะห์คำผิดเหล่านั้นเพื่อหารูปแบบหรือลักษณะของการสะกดผิด

5.2 การวิเคราะห์คำที่มักเขียนผิดที่เป็นการสะกดผิดแบบเป็นคำจริง

ในการศึกษาวิเคราะห์คำที่มักเขียนผิดครั้งนี้ ผู้วิจัยได้ศึกษาเปรียบเทียบความแตกต่างระหว่างคำที่สะกดผิดควบคู่กับรูปคำสะกดที่ถูกต้องตามพจนานุกรมฉบับราชบัณฑิตยสถาน ว่ามีข้อผิดพลาดเกิดขึ้นที่ส่วนใดหรือในลักษณะใดระหว่างคำทั้งสอง เช่น

คำว่า กั้ววาน พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2554 ระบุว่าเป็นคำวิเศษณ์ มีความหมายว่า ก้องอยู่ได้นาน มักจะถูกสะกดผิดเป็น กั้ววาล

ตารางที่ 1

ตัวอย่างข้อมูลลักษณะข้อผิดพลาดในการสะกดผิดที่ผู้วิจัยบันทึกไว้เพื่อนำมาศึกษาวิเคราะห์

คำที่สะกดถูกต้อง	มักสะกดผิดเป็น	จำนวนข้อผิดพลาด	ตำแหน่งข้อผิดพลาด	รูปแบบข้อผิดพลาด
กั้ววาน	กั้ว <u>วาล</u> ¹	1	1 ตัวสะกด	น_ล (น ถูกแทนที่ด้วย ล)
วิกฤตติกาล	วิกฤ <u>ต</u> การ ²	2	1 สระ 2 ตัวสะกด	ิ_0 (ลบสระ อี) ล_ร (ล ถูกแทนที่ด้วย ร)
นานับการ	นา <u>น</u> ับ <u>ป</u> ระ ² <u>ะ</u> ³ การ	3	1 ตัวสะกด 2 พยัญชนะควบกล้ำ 3 สระ	0_บ (เติม บ) 0_ร (เติม ร) 0_ะ (เติมสระ อะ)

จากตัวอย่าง จะเห็นได้ว่าคำที่สะกดผิดและคำที่สะกดถูกต้องนั้นแตกต่างกันอยู่หนึ่งก็คือที่ตัวสะกด จาก <น> เปลี่ยนไปเป็น <ล> ซึ่งผู้วิจัยได้ทำการบันทึกข้อมูลต่างๆ ที่จะสามารถช่วยอธิบายลักษณะของการสะกดผิดเก็บเอาไว้ เช่น จำนวนของข้อผิดพลาด ตำแหน่งที่เกิดข้อผิดพลาด และรูปแบบของอักขระที่เปลี่ยนไปในคำที่สะกดผิดเมื่อเปรียบเทียบกับคำที่สะกดถูกต้อง (ตั้งตัวอย่างในตารางที่ 1) จากนั้นก็นำข้อมูลลักษณะข้อผิดพลาดในการสะกดผิดที่บันทึกไว้ทั้งหมดมาศึกษาวิเคราะห์เพื่อรายงานผล

6. ผลการวิเคราะห์

จากการศึกษาวิเคราะห์ลักษณะข้อผิดพลาดในการสะกดผิดแบบเป็นคำจริงของคำที่มักเขียนผิดจำนวน 1,674 คำ สามารถจำแนกคำที่มักเขียนผิดเหล่านี้ตามจำนวนของลักษณะข้อผิดพลาดที่พบในแต่ละคำได้เป็น 2 กลุ่ม คือ กลุ่มของคำที่สะกดผิดหนึ่งลักษณะ มีจำนวน 1,339 คำ คิดเป็นร้อยละ 80 ของคำที่เขียนผิดทั้งหมด และกลุ่มของคำที่สะกดผิดหลายลักษณะ โดยมีจำนวน 335 คำ คิดเป็นร้อยละ 20 ของคำที่เขียนผิดทั้งหมด

6.1 คำที่สะกดผิดหนึ่งลักษณะ

เมื่อศึกษาวิเคราะห์คำที่สะกดผิดหนึ่งลักษณะจากตำแหน่งที่สะกดผิด พบว่าสามารถแบ่งคำในกลุ่มนี้ออกเป็น 7 กลุ่ม ตามลักษณะของการสะกดผิดในลักษณะต่างๆ ได้แก่

6.1.1 คำที่สะกดผิดที่พยัญชนะสะกด

จากการศึกษาครั้งนี้ พบว่าคำที่มักเขียนผิดที่สะกดผิดแบบเป็นคำจริงมักจะสะกดผิดที่ตัวสะกดทั้งหมด 338 คำ หรือร้อยละ 20.2 จากคำที่มักเขียนผิดในคลังคำทั้งหมด 1,674 คำ ซึ่งเป็นลักษณะการสะกดผิดที่พบมากที่สุด และเมื่อศึกษาวิเคราะห์คำที่มักเขียนผิดในกลุ่มนี้โดยเฉพาะ ก็พบว่าสามารถจำแนกคำที่สะกดผิดในกลุ่มนี้ออกตามการเปลี่ยนแปลงที่เกิดขึ้นระหว่างคำที่สะกดถูกและคำที่สะกดผิด โดยสังเกตจากคำที่สะกดผิดว่ามีการเปลี่ยนแปลงคำไปในลักษณะใดจากคำที่สะกดถูกได้เป็น 3 กลุ่มย่อย คือ

ก. กลุ่มที่แทนที่ตัวสะกดที่ถูกต้องด้วยตัวสะกดอื่นในมาตราตัวสะกดเดียวกัน ซึ่งเป็นลักษณะข้อผิดพลาดที่เกี่ยวข้องกับตัวสะกด ที่พบมากที่สุดคือ 315 คำ หรือ 93.2% จากทั้งหมด 338 คำ ซึ่งรูปแบบที่พบมากเป็นอันดับต้นๆ ได้แก่ การใช้ตัวสะกด <ล> แทน <น> เช่น หูนหัว เขียนผิดเป็น หูลหัว การใช้ <น> แทน <ล> เช่น กังวล

เขียนผิดเป็น กังวน และการใช้ <น> แทน <ณ> เช่น เกษียณอายุ เขียนผิดเป็น เกษียนอายุ

ข. กลุ่มที่เติมตัวสะกดเพิ่มเข้าไปในคำที่ถูกต้องทำให้กลายเป็นคำที่สะกดผิด พบทั้งหมด 12 คำ หรือร้อยละ 3.55 ตัวอย่างเช่น การเติมตัวสะกด <น> ใน คำว่า *ทรมานทรรกรรม* เขียนผิดเป็น *ทรมานทรกรรม* เป็นต้น

ค. กลุ่มที่ลบตัวสะกดของคำที่สะกดถูกต้องทำให้คำนั้นสะกดผิด มีทั้งหมด 11 คำหรือ 3.25% เช่น การลบตัวสะกด <ษ> ออก จากคำว่า *อิฐฐาน* ทำให้เขียนผิดเป็น *อิฐฐาน* เป็นต้น

6.1.2 คำที่สะกดผิดที่ตัวการันต์

การสะกดผิดที่เกิดจากการใช้ตัวการันต์เป็นลักษณะของการสะกดผิดที่พบมากเป็นอันดับที่สอง คือ 270 คำ ซึ่งคิดเป็นเปอร์เซ็นต์ได้ 16.13% หลังจากที่ได้ศึกษาวิเคราะห์คำที่มักเขียนผิดในกลุ่มนี้พบว่า สามารถแบ่งคำผิดในกลุ่มนี้ออกเป็น 3 กลุ่มย่อยตามการเปลี่ยนแปลงที่เกิดขึ้นระหว่างคำที่สะกดถูกและคำที่สะกดผิด เช่นเดียวกันกับคำในกลุ่มแรก ได้แก่

ก. กลุ่มที่เติมตัวการันต์เพิ่มเข้าไปทำให้สะกดผิด กลุ่มนี้เป็นกลุ่มที่พบว่ามี การเขียนผิดมากที่สุด 101 คำ จากทั้งหมด 270 คำ หรือคิดเป็น 37.4% ตัวอย่างเช่น การเติม <ย> ในคำว่า อินทรี (นก) เขียนผิดเป็น อินทรีย (สาร)

ข. กลุ่มที่แทนที่ตัวการันต์เดิมในคำที่สะกดถูกต้องด้วยตัวการันต์ใหม่พบทั้งหมด 90 คำหรือร้อยละ 33.33 เช่น การแทนที่ <ต> ด้วย <ตร> เช่นในคำว่า *น้ำมันต์* เขียนผิดเป็น *น้ำมันตร์*

ค. กลุ่มที่ลบตัวการันต์จากคำที่สะกดถูกต้องทำให้สะกดผิด มีทั้งหมด 79 คำ หรือ 26.33% ยกตัวอย่างเช่น การลบ <ณ> ในคำว่า *ประสบการณร์* เขียนผิดเป็น *ประสบการ* เป็นต้น

6.1.3 คำที่สะกดผิดที่พยัญชนะต้น

ลักษณะของความผิดพลาดในการสะกดผิดที่พยัญชนะต้นของคำ ซึ่งจากการศึกษาพบว่าคำที่มักเขียนผิดที่สะกดผิดในลักษณะนี้มีทั้งหมด 204 คำ หรือร้อยละ 12.18 ซึ่งเป็นลักษณะการสะกดผิดที่พบบ่อยเป็นอันดับที่สาม นอกจากนี้เมื่อสังเกตความแตกต่างระหว่างคำที่สะกดผิดกับคำที่สะกดถูกต้องแล้วก็สามารถแบ่งคำผิดในกลุ่มออกเป็น 3 กลุ่มย่อย คือ

ก. กลุ่มที่แทนที่พยัญชนะต้นด้วยเสียงที่คล้ายกันหรือเหมือนกันแต่รูปเขียนต่างกันเป็นลักษณะข้อผิดพลาดที่พยัญชนะต้น ที่พบบ่อยที่สุดคือ 186 คำจากทั้งหมด 204 คำ หรือเท่ากับ 91.2% ซึ่งรูปแบบของการสะกดผิดในลักษณะนี้ที่พบบ่อยเป็นอันดับต้นๆ ได้แก่ การใช้พยัญชนะ <ล> แทน <ร> หรือ ใช้ <ล> แทน <ร> ซึ่งในทางปฏิบัติแล้วการพูดหรือออกเสียงพยัญชนะทั้งสองตัวนี้มักจะถูกใช้สลับกันอยู่เป็นประจำ แต่ด้วยบริบทในการสนทนาจึงไม่ส่งผลกระทบต่อความหมาย ผู้ฟังยังสามารถเข้าใจสิ่งที่ผู้พูดต้องการจะสื่อแม้จะออกเสียงไม่ถูกต้องตามหลักภาษาไทย ด้วยเหตุนี้วิธีการออกเสียงอาจเป็นปัจจัยสำคัญที่ทำให้เกิดการสะกดผิดในลักษณะนี้ได้ ตัวอย่างเช่นคำว่า ร้างรา เขียนผิดเป็น ร้างลา หรือ เล้าโลม เขียนผิดเป็น เล้าโลม ซึ่งเป็นการใช้ <ล> สลับกับ <ร> และการใช้ <ล> สลับกับ <ร> ตามลำดับ นอกจากนี้ความผิดพลาดในลักษณะนี้ยังสามารถเกิดขึ้นกับพยัญชนะตัวอื่นๆ ด้วย ตัวอย่างเช่น การใช้พยัญชนะ <น> แทน <ณ>, ใช้ <ส> แทน <ศ>, และใช้ <ทร> แทน <ช> ในคำว่า ปรานี เขียนผิดเป็น ปรานี, ประกาศิต เขียนผิดเป็น ประกาศิต และ ชาบซึ่ง เขียนผิดเป็น ทราบซึ่ง ตามลำดับ ซึ่งพยัญชนะต้นแต่ละคู่ที่สะกดผิดในลักษณะนี้มักเป็นรูปเขียนที่ใช้แทนหน่วยเสียงเดียวกัน

ข. กลุ่มที่เติมพยัญชนะต้นเพิ่มไปในคำทำให้เป็นการสะกดผิดมีจำนวน 15 คำ หรือคิดเป็นร้อยละ 7.35 ซึ่งพยัญชนะที่มักจะถูกเติมเข้าไปในคำแล้วทำให้เกิดการสะกดผิดในลักษณะนี้คือ <ณ> เช่น รัญพีช เขียนผิดเป็น รัญญพีช หรือ สัญลักษณ์ เขียนผิดเป็น สัญญลักษณ์

ค. กลุ่มที่ลบบัญชีขณะต้นออกทำให้สะกดผิด พบเพียงแค่ 3 คำ เท่ากับ 1.47% ได้แก่คำว่า *พรรณนา* เขียนผิดเป็น *พรรณา*, *สัมมนา* เขียนผิดเป็น *สัมนา* และ *จิตรลดา* เขียนผิดเป็น *จิตรดา*

6.1.4 คำที่สะกดผิดที่การประวิสรรชนีย์

การสะกดผิดพลาดที่เกิดจากการประวิสรรชนีย์หรือไม่ประวิสรรชนีย์เป็นลักษณะความผิดพลาดที่พบมากเป็นอันดับที่สี่ในการศึกษาครั้งนี้ ซึ่งมีจำนวนทั้งหมด 198 คำหรือเท่ากับ 11.82% ของคำในคลังคำที่มักเขียนผิด ซึ่งจากการศึกษาพบว่าการสะกดผิดในลักษณะนี้สามารถจำแนกได้เป็น 2 กลุ่มย่อย ได้แก่

ก. กลุ่มคำสะกดผิดเนื่องจากการประวิสรรชนีย์ จำนวน 110 คำจาก 198 คำ หรือร้อยละ 55.55 ซึ่งรูปแบบของการสะกดผิดในกลุ่มที่หนึ่งนั้นมักจะเกิดขึ้นกับคำที่ประกอบด้วยพยางค์ที่ออกเสียงว่า *ระ* แต่ไม่ประวิสรรชนีย์ เช่น *ธุรกิจ* เขียนผิดเป็น *ธุระกิจ*, *สรีรวิทยา* เขียนผิดเป็น *สรีระวิทยา*, หรือ *สารประโยชน์* เขียนผิดเป็น *สาระประโยชน์* เป็นต้น นอกจากพยางค์ที่ออกเสียงว่า *ระ* แล้วก็ยังพบพยางค์อื่นๆ อีกที่เกิดข้อผิดพลาดในลักษณะเดียวกัน ได้แก่ *วะ* และ *สะ* ซึ่งความถี่ที่พบก็ลดหลั่นตามลำดับ ตัวอย่างเช่น คำว่า *สภาวะการณ์* เขียนผิดเป็น *สภาวะการณ* และ *สดมภ์* เขียนผิดเป็น *สะดมภ์* เป็นต้น

ข. กลุ่มคำสะกดผิดเนื่องจากการไม่ประวิสรรชนีย์ จำนวน 88 คำ หรือร้อยละ 44.45 ส่วนรูปแบบการสะกดผิดในกลุ่มที่สองนั้นพบว่าพยางค์ที่ประวิสรรชนีย์และออกเสียงว่า *ชะ* *ระ* และ *สะ* ประกอบอยู่ในคำนั้นนั้นมักจะถูกสะกดผิดโดยการไม่ประวิสรรชนีย์ ตัวอย่างเช่น *ชะงัก* เขียนผิดเป็น *ซงัก* *จระเข้* เขียนผิดเป็น *จรเข้* และ *สะอาด* เขียนผิดเป็น *สอาด* เป็นต้น

6.1.5 คำที่สะกดผิดที่พยัญชนะควบกล้ำ

การสะกดผิดที่เกิดจากความผิดพลาดของคำควบกล้ำนี้เป็นลักษณะการสะกดผิดที่เมื่อจัดเรียงตามจำนวนคำผิดที่ถูกจำแนกตามลักษณะการสะกดผิดแล้ว อยู่ในลำดับที่ห้า พบว่ามีจำนวนคำที่สะกดผิดเนื่องจากการผิดพลาดในของการใช้

คำควบกล้ำอยู่ 178 คำ หรือคิดเป็น 10.60% ของคำที่มักเขียนผิดในคลังทั้งหมด ซึ่งการสะกดผิดในลักษณะนี้สามารถจำแนกตามรูปแบบการเปลี่ยนแปลงระหว่างคำที่สะกดถูกและคำที่สะกดผิดได้ 3 กลุ่มย่อย ได้แก่

ก. กลุ่มคำสะกดผิดเนื่องจากการเติมพยัญชนะควบกล้ำ <ร> <ล> หลังพยัญชนะต้นเพื่อให้เกิดเป็นอักษรควบกล้ำ ซึ่งพบว่ามีคำที่มักเขียนผิดในลักษณะนี้ทั้งหมด 65 คำหรือร้อยละ 36.52 และรูปแบบของคำที่สะกดผิดในลักษณะนี้ที่พบมากที่สุดคือ การเติม <ร> หลังพยัญชนะต้น <ก> ได้เป็นอักษรควบกล้ำ <กร> เช่น กะทันหัน เขียนผิดเป็น กระทันหัน และรูปแบบที่พบรองลงมา คือ การเติม <ล> หลังพยัญชนะต้น <ผ> หรือ <ก> ได้เป็นอักษรควบกล้ำ <ผล> หรือ <กล> เช่น ผัดผ้อน เขียนผิดเป็น ผลัดผ้อน และ วางก้าม เขียนผิดเป็น วางกล้าม เป็นต้น

ข. กลุ่มคำที่สะกดผิดเนื่องจากการลบพยัญชนะควบกล้ำ <ร> <ล> ออกไปเหลือแต่พยัญชนะต้น ซึ่งมีคำที่สะกดผิดในลักษณะนี้ทั้งสิ้น 61 คำ เท่ากับ 34.27% และการลบพยัญชนะควบกล้ำ <ร> ที่ตามหลังพยัญชนะต้น <ก> เป็นรูปแบบการสะกดผิดลักษณะนี้ที่พบมากที่สุด ตัวอย่างเช่น กระเพาะ เขียนผิดเป็น กะเพาะ และการลบ <ล> และ <ร> หลังพยัญชนะต้น <ผ> และ <ป> ตามลำดับ เป็นรูปแบบการสะกดผิดลักษณะนี้ที่พบรองลงมาตามความถี่ ตัวอย่างเช่น ผลัดเวร เขียนผิดเป็น ผัดเวร และ ประณิธาน เขียนผิดเป็น ปณิธาน เป็นต้น

ค. กลุ่มคำที่สะกดผิดเนื่องจากการเลือกใช้พยัญชนะควบกล้ำผิดทำซึ่งพบว่ามีจำนวนคำที่สะกดผิดในลักษณะนี้จำนวน 52 คำ หรือ 29.21% ซึ่งรูปแบบที่มักเขียนผิดในลักษณะนี้ ได้แก่ การเลือกใช้อักษรควบกล้ำ <พล> แทน <พร>, <กล> แทน <กร>, และ <คล> แทน <คร> ตัวอย่างเช่น พรางตา เขียนผิดเป็น พलगตา, เกร็ดความรู้ เขียนผิดเป็น เกล็ดความรู้, และ หอยแครง เขียนผิดเป็น หอยแกลง เป็นต้น

6.1.6 คำที่สะกดผิดที่สระ

การสะกดผิดในลักษณะนี้เกิดจากการเลือกใช้สระอื่นๆ ที่ไม่ใช่รูปสระอะ <ะ> ไม่ถูกต้อง ซึ่งพบจำนวนคำที่สะกดผิดในลักษณะนี้จำนวน 108 คำจากคำทั้งหมดใน

คลังคำที่ศึกษา หรือเท่ากับ 6.45% จากการศึกษพบว่าสามารถจำแนกคำที่สะกดผิด ในลักษณะนี้ออกเป็น 3 กลุ่มย่อย ตามลักษณะที่เปลี่ยนแปลงไประหว่างคำที่ถูกต้อง และคำผิด ได้แก่

ก. กลุ่มแรกเป็นการสะกดผิดที่เกิดจากการเลือกใช้สระไม่ถูกต้อง ซึ่งพบว่า มีคำที่สะกดผิดในลักษณะนี้ 56 คำ จาก 108 คำ คิดเป็น 51.85% ซึ่งรูปแบบของการ สะกดผิดในลักษณะนี้มักจะเป็นการเลือกใช้สระผิดระหว่าง สระ <ไ> และ<ใ> ตัวอย่างเช่น *ไต้ฝุ่น* เขียนผิดเป็น *ไต้ฝุ่น* และ *เยื่อใย* เขียนผิดเป็น *เยื่อใย* เป็นต้น

ข. กลุ่มที่สองเป็นการสะกดผิดที่เกิดจากการเติมรูปสระเพิ่มเข้าไปในคำ ซึ่งมี จำนวนคำที่สะกดผิด 34 คำ คิดเป็น 31.48% โดยรูปแบบของคำที่มักเขียนผิดใน ลักษณะนี้ได้แก่ การเติมสระ <า>ตัวอย่างเช่น *คารวะ* เขียนผิดเป็น *คารวะ*

ค. กลุ่มที่สามเป็นการสะกดผิดที่เกิดจากการลบรูปสระออก ซึ่งรูปแบบของ การสะกดผิดในลักษณะนี้ได้แก่ การลบไม้ไต่คู้ออก < ็> ซึ่งเป็นเครื่องหมายที่เป็น ส่วนประกอบของสระเอะ < ็ะ> สระแอะ (< ็>) และสระเอาะ (< ็>) ตัวอย่างเช่น *ผล็อย* เขียนผิดเป็น *ผลอย* เป็นต้น

6.1.7 คำที่สะกดผิดที่วรรณยุกต์

จากการศึกษาพบว่าการสะกดผิดหนึ่งลักษณะที่เกิดจากความผิดพลาดใน การใช้วรรณยุกต์นั้นเป็นการสะกดผิดที่พบน้อยที่สุดคือ 43 คำ จากคำในคลังที่ศึกษา ทั้งหมด 1,674 คำ หรือเท่ากับร้อยละ 2.57 ซึ่งจากการศึกษาพบว่าสามารถจำแนก การสะกดผิดที่เกิดจากความผิดพลาดที่วรรณยุกต์โดยสังเกตจากลักษณะของการ เปลี่ยนแปลงระหว่างคำที่ถูกต้องกับคำที่ผิดได้ 3 กลุ่มย่อย ได้แก่

ก. กลุ่มที่หนึ่งซึ่งพบว่ามีจำนวนคำที่มักสะกดผิดในลักษณะมากที่สุด คือ การสะกดผิดเนื่องจากการเติมรูปวรรณยุกต์ตัวใดตัวหนึ่งซึ่งในคำที่ผิดพลาดที่ วรรณยุกต์ทั้งหมด 43 คำ พบว่ามีคำที่สะกดผิดในลักษณะนี้ 22 คำ หรือเท่ากับ 51.16% โดยรูปวรรณยุกต์ที่มักถูกเติมในการสะกดผิดลักษณะนี้คือ วรรณยุกต์เอก ตรี โท และจัตวา ไล่เรียงตามความถี่ที่พบ ตัวอย่างเช่น *เล็กลัก* เขียนผิดเป็น *เล็กลัก*, *จักจั่น* เขียนผิดเป็น *จักจั่น*, *ตั้งไ้อ้* เขียนผิดเป็น *ตั้งไ้อ้* และ *กวยจี๊* เขียนผิดเป็น *ก้วยจี๊*

ข. กลุ่มที่สองเป็นคำที่มีการสะกดผิดเนื่องจากใช้รูปวรรณยุกต์ผิดเพี้ยนไปจากเดิม ซึ่งมีทั้งหมด 15 คำ คิดเป็น 34.88% ซึ่งการสะกดผิดในลักษณะนี้มักจะมีรูปแบบดังนี้ ใช้รูปวรรณยุกต์ตรีแทนที่รูปวรรณยุกต์โท หรือใช้รูปวรรณยุกต์เอกแทนที่รูปวรรณยุกต์โท ตัวอย่างเช่น *ปลั๊กไฟ* เขียนผิดเป็น *ปลักไฟ* หรือ *เลื่อยเจ็ย* เขียนผิดเป็น *เลื่อยเจ็ย* เป็นต้น

ค. กลุ่มที่สามเป็นคำที่สะกดผิดเนื่องจากการไม่เติมรูปวรรณยุกต์ซึ่งมีคำที่สะกดผิดในลักษณะนี้ 6 คำ หรือเท่ากับ 13.96% และรูปแบบการสะกดผิดที่มักพบในลักษณะนี้ ได้แก่ การไม่ใส่รูปวรรณยุกต์ตรี ตัวอย่างเช่น *ก๊าช* เขียนผิดเป็น *กาช* เป็นต้น

6.2 คำที่สะกดผิดหลายลักษณะ

คำที่มักเขียนผิดที่ถูกจำแนกไว้ในกลุ่มนี้เป็นคำที่มีความผิดพลาดมากกว่า 1 ลักษณะ เช่นคำว่า *จันท์น้กพ้อ* เขียนผิดเป็น *จันท์กะพ้อ* จะเห็นได้ว่ามีข้อผิดพลาด 2 ลักษณะ ได้แก่ ความผิดพลาดในการเลือกใช้ตัวกรันต์และความผิดพลาดในการประวิสรรชนีย์ หรือ คำว่า *ขโมย* เขียนผิดเป็น *โขมย* ซึ่งเป็นการสะกดผิดในลักษณะของโครงสร้างอักษรนำ โดยผู้วิจัยจัดว่ามีข้อผิดพลาด 2 ลักษณะ คือ การลบพยัญชนะต้น <ข> หรือ สระ <โ> แล้วจึงเติมสระ <โ> หน้าพยัญชนะ <ข> หรือเติมพยัญชนะ <ข> หลังสระ <โ> ซึ่งจากการศึกษาพบว่า มีคำที่สะกดผิดในลักษณะดังกล่าวข้างต้นอยู่ทั้งหมด 335 คำ หรือร้อยละ 20.01 ของคำทั้งหมดในคลังคำที่ใช้ศึกษาวิเคราะห์ โดยสามารถจำแนกออกได้เป็น 3 กลุ่ม ดังนี้

6.2.1 คำที่สะกดผิดหลายลักษณะแต่เสียงคงเดิม

คำที่สะกดผิดหลายลักษณะในกลุ่มแรกเป็นกลุ่มของคำที่มีการสะกดผิดโดยเขียนพยางค์ในรูปแบบอื่นที่มีเสียงคงเดิม ทำให้ผิดมากกว่าหนึ่งตำแหน่ง ได้แก่ การใช้รูป <ัน> แทน รูป ร ्हัน <รร> ซึ่งความผิดพลาดในลักษณะนี้มักเกิดขึ้นกับคำที่ขึ้นด้วยคำว่า *กรร* หรือ *บร* เช่นคำว่า *กรรแสง* เขียนผิดเป็น *กันแสง* และคำว่า *บรรด* เขียนผิดเป็น *บันรด* เป็นต้น นอกจากนี้ยังมี คำว่า *หม้าย* กับ *ม้าย* และ *สูญ* กับ *ศูนย์* ซึ่งทั้งหมดเป็นคำพ้องเสียงที่มีรูปเขียนไม่เหมือนกัน หากใช้ผิดก็อาจจะทำ

ให้ความหมายผิดเพี้ยนไปได้ ตัวอย่างเช่นคำว่า *ศูนย์หน้า* เขียนผิดเป็น *สูญหน้า* และ *พ้อม่าย* เขียนผิดเป็น *พ้อม้าย* ตามลำดับ กลุ่มนี้พบมากสุดในกลุ่มคำที่สะกดผิดหลายลักษณะ คือมีจำนวน 220 คำ เท่ากับร้อยละ 65.67 ของคำที่สะกดผิดหลายลักษณะ และ 13.14% ของคำผิดทั้งหมด

6.2.2 คำที่สะกดผิดหลายลักษณะที่ออกเสียงเปลี่ยนไป

กลุ่มที่สองเป็นกลุ่มของคำที่ผู้วิจัยคาดว่าเป็นการสะกดผิดที่มีสาเหตุมาจากการออกเสียงคำที่ผิดเพี้ยนไปทำให้รูปเขียนของคำเหล่านี้ผิดเพี้ยนตามไปด้วย จากการศึกษาพบว่า มีคำที่สะกดผิดในลักษณะนี้จำนวน 98 คำจาก 335 คำ หรือร้อยละ 29.25 ของคำที่สะกดผิดหลายลักษณะ และ 5.85% ของคำผิดทั้งหมด ตัวอย่างเช่น คำว่า *กระบวนกร* เขียนผิดเป็น *ขบวนกร* หรือ คำว่า *พั้น* เขียนผิดเป็น *ควัน* เป็นต้น

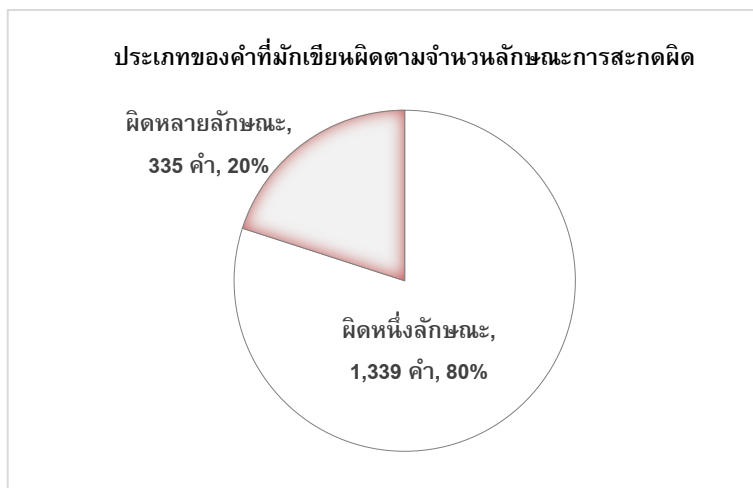
6.2.3 คำที่สะกดผิดหลายลักษณะที่อักษรนำ

กลุ่มสุดท้ายเป็นกลุ่มของคำที่มีการสะกดผิดในลักษณะของโครงสร้างอักษรนำ คือ โครงสร้างของคำที่มีพยัญชนะต้นสองตัว แต่อ่านแยกกันเป็นสองพยางค์ (สุนันท์ อัญชสิณกุล, 2552, น. 88) ซึ่งไม่สามารถระบุได้แน่ชัดว่าคำที่สะกดผิดนั้นจะออกเสียงเช่นเดิมหรือเปลี่ยนไป โดยพบว่ามีคำที่สะกดผิดในลักษณะเช่นนี้จำนวน 17 คำ หรือร้อยละ 5.07 ของคำที่สะกดผิดหลายลักษณะ และ 1.02% ของคำผิดทั้งหมด ตัวอย่างเช่น คำว่า *ทแยง* (อ่านว่า ทะ-แยง) เขียนผิดเป็น *แทยง* (อาจจะอ่านว่า ทะ-แยง หรือ แท-ยง ก็ได้) หรือ คำว่า *แสลง* (อ่านว่า สะ-แสลง) เขียนผิดเป็น *สแสลง* (อาจจะอ่านว่า สะ-แสลง หรือ สะ-แลง) เป็นต้น

7. สรุปและอภิปรายผล

จากผลการศึกษาวิเคราะห์คำที่มักเขียนผิดที่เป็นการสะกดผิดแบบเป็นคำจริงจำนวน 1,674 คำ โดยนำคำที่มักเขียนผิดที่เป็นคำที่สะกดผิดแบบเป็นคำจริงมาวิเคราะห์เปรียบเทียบลักษณะของคำที่สะกดผิดนั้นว่าเปลี่ยนแปลงไปจากรูปการสะกดที่ถูกต้อง

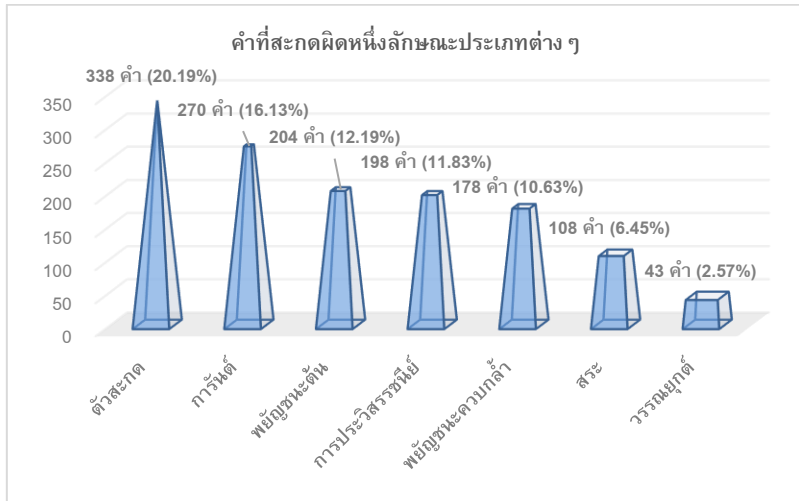
อย่างไรบ้าง จากผลการวิเคราะห์สามารถสรุปได้ว่าคำที่มักเขียนผิดที่นำมาศึกษาวิเคราะห์ห้ในครั้งนี้อาจจำแนกตามจำนวนของลักษณะข้อผิดพลาดที่พบในคำที่สะกดผิดแต่ละคำได้เป็น 2 กลุ่มใหญ่ คือ กลุ่มของคำที่สะกดผิดหนึ่งลักษณะ 1,339 คำ และกลุ่มของคำที่สะกดผิดหลายลักษณะ 335 คำ หรือประมาณร้อยละ 80 และร้อยละ 20 ตามลำดับ ดังที่แสดงในรูปภาพที่ 1



ภาพที่ 1. ประเภทของคำที่มักเขียนผิดจำแนกตามจำนวนลักษณะข้อผิดพลาดที่พบในการสะกดผิด

เมื่อพิจารณาตำแหน่งของการสะกดผิดในคำที่สะกดผิดหนึ่งลักษณะพบว่าสามารถจำแนกออกเป็นกลุ่มย่อยๆ ได้อีกทั้งหมด 7 กลุ่ม (ดังรูปภาพที่ 2) ได้แก่

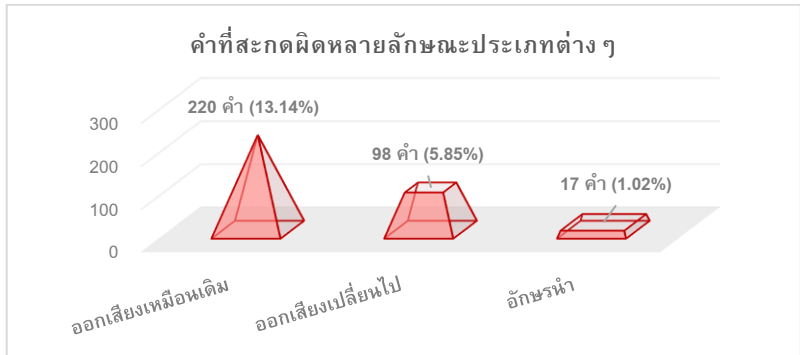
1. คำที่สะกดผิดที่ตัวสะกด จำนวน 338 คำ คิดเป็นร้อยละ 20.19
2. คำที่สะกดผิดที่ตัวการันต์ จำนวน 270 คำ คิดเป็นร้อยละ 16.13
3. คำสะกดผิดที่พยัญชนะต้น จำนวน 204 คำ คิดเป็นร้อยละ 12.19
4. คำที่สะกดผิดที่การประวิสรรชนีย์ จำนวน 198 คำ คิดเป็นร้อยละ 11.83%
5. คำที่สะกดผิดที่คำควบกล้ำ จำนวน 178 คำ คิดเป็นร้อยละ 10.63
6. คำที่สะกดผิดที่สระ จำนวน 108 คำ คิดเป็นร้อยละ 6.45
7. คำที่สะกดผิดที่วรรณยุกต์ จำนวน 43 คำ คิดเป็นร้อยละ 2.57



ภาพที่ 2. แผนภูมิแท่งแสดงคำที่สะกดผิดหนึ่งลักษณะประเภทต่าง ๆ

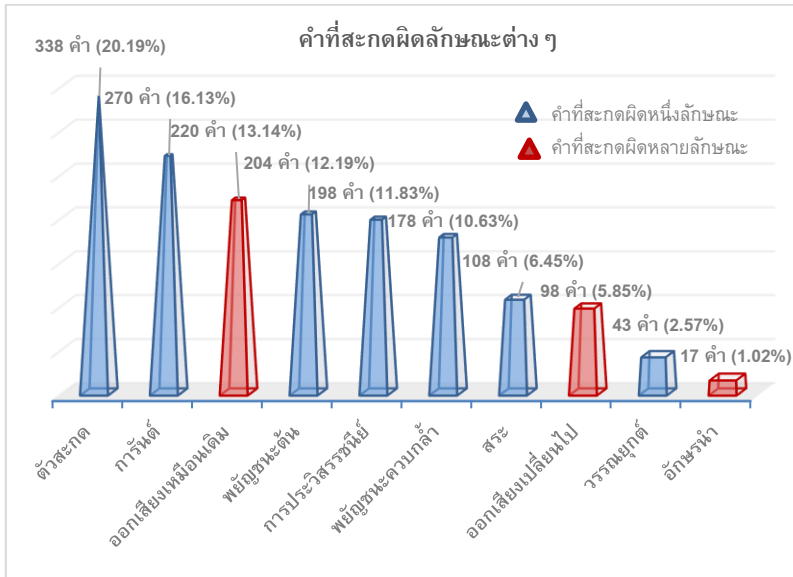
นอกจากนี้ยังพบว่าสามารถจำแนกคำที่สะกดหลายลักษณะออกเป็น 3 กลุ่มย่อย (ดังรูปภาพที่ 3) ได้แก่

1. คำที่สะกดผิดหลายลักษณะแต่เสียงคงเดิม จำนวน 220 คำ คิดเป็นร้อยละ 13.14
2. คำที่สะกดผิดหลายลักษณะที่ออกเสียงเปลี่ยนไป จำนวน 98 คำ คิดเป็นร้อยละ 5.85
3. คำที่สะกดผิดหลายลักษณะที่อักษรนำ จำนวน 17 คำ คิดเป็นร้อยละ 1.02



ภาพที่ 3. แผนภูมิแท่งแสดงคำที่สะกดผิดหลายลักษณะประเภทต่าง ๆ

จากผลการวิเคราะห์สามารถสรุปได้ว่าคำที่มักเขียนผิดที่นำมาศึกษาวิเคราะห์ในครั้งนี้สามารถจำแนกตามลักษณะของคำที่สะกดผิดซึ่งเปลี่ยนแปลงไปจากการสะกดเดิมที่สะกดได้ถูกต้องออกเป็น 2 กลุ่มใหญ่ คือ คำที่สะกดผิดหนึ่งตำแหน่งและคำที่สะกดผิดหลายตำแหน่ง โดยคำที่สะกดผิดหนึ่งตำแหน่งสามารถจำแนกออกเป็น 7 กลุ่มย่อย และคำที่สะกดผิดหลายตำแหน่งสามารถจำแนกออกเป็น 3 กลุ่มย่อย รวมเป็น 10 ลักษณะดังในรูปภาพที่ 4 (ด้านล่าง) โดยเรียงลำดับจากลักษณะที่พบว่ามีคำที่สะกดผิดในลักษณะนั้นจากมากที่สุดไปหาน้อยที่สุด แสดงให้เห็นว่าคำที่มักเขียนผิดส่วนใหญ่นั้นมีรูปแบบที่สามารถนำมาใช้เป็นเกณฑ์เพื่อจำแนกหรือจัดกลุ่มได้



ภาพที่ 4. แผนภูมิแท่งแสดงคำที่สะกดผิดประเภทต่าง ๆ ที่พบทั้งหมดในการศึกษาครั้งนี้

ผลการวิเคราะห์ที่ได้จากการศึกษาในครั้งนี้ ผู้วิจัยพบว่า คำที่สะกดผิดส่วนใหญ่ มักมีการสะกดผิดเพียงหนึ่งตำแหน่ง คือ พิมพ์หรือเขียนอักษรหนึ่งตัวผิดเป็นตัวอื่น หรือดักอักษรหนึ่งตัว หรือใส่อักษรเกินมาหนึ่งตัว ดังตัวอย่างในหัวข้อ 3.2 เพราะฉะนั้น ในทางกลับกันคำที่สะกดผิดหนึ่งตำแหน่งเหล่านี้ก็สามารถปรับแก้ให้เป็นคำที่สะกดถูกต้องได้ด้วยวิธีการปรับแก้เพียงครั้งเดียว โดยอาศัยรูปแบบของการสะกดผิดที่ได้จากการวิเคราะห์คำไทยที่มักสะกดผิดในครั้งนี้ ซึ่งรูปแบบของการสะกดผิดเหล่านี้จะเป็นประโยชน์ต่อการแก้ไขการสะกดผิดด้วยระบบคอมพิวเตอร์ โดยทั่วไปแล้วมักจะทำการแก้ไขการสะกดผิดด้วยโมดูล *minimum edit distance* (Neil Bowers, 2015) เป็นวิธีการปรับแก้คำที่สะกดผิดให้เป็นคำที่สะกดถูกต้องด้วยการปรับแก้ให้น้อยที่สุดซึ่งส่วนใหญ่แล้วจะถูกกำหนดให้เป็นหนึ่งครั้ง ตัวอย่างเช่น เมื่อให้ระบบแก้ไขการสะกดผิดของคำว่า *เกร็ดปลา* และ *หอมหวล* โดยปกติแล้วระบบก็จะค้นหาว่ามีคำใดที่เป็นไปได้บ้างจากการปรับแก้คำสะกดผิดเพียงหนึ่งครั้ง ซึ่งอาจจะพบว่า *เกร็งปลา* และ *หอมหวง* เป็นคำที่เป็นไปได้เพราะมีจำนวนการปรับแก้เพียงครั้งเดียวแต่ยังไม่ใช่คำที่ถูกต้องที่สุด

ซึ่งก็คือคำว่า *เกล็ดปลา* และ *หอมหวน* ซึ่งจากรูปแบบที่ได้จากการวิเคราะห์จะทำให้ทราบว่าตัวอักษร <ร> และ <ล> มักถูกใช้สลับกัน และตัวสะกดในมาตราเดียวกันมักจะถูกใช้สลับกัน ในตัวอย่างนี้ก็คือ <ล> แทน <น> เมื่อฝึกฝนให้ระบบเรียนรู้ว่าคำใดที่สะกดผิดที่ตัวอักษร <ร> ก็ควรแทนที่ด้วย <ล> หรือในทางกลับกันหากสะกดผิดที่ตัวอักษร <ล> ก็ควรแทนที่ด้วย <ร> เพราะทั้งสองตัวอักษรนี้มักจะถูกใช้สลับกัน หรือควรแทนที่ตัวสะกด <ล> ด้วยตัวอักษร <น> เพราะ ทั้งสองตัวอักษรนี้เป็นตัวสะกดในมาตราแม่กน เหมือนกัน และ <น> มักถูกใช้สลับกับ <ล> เป็นต้น รูปแบบของการสะกดผิดเหล่านี้จะช่วยให้คาดเดาแนวทางการแก้ไขคำที่สะกดผิดให้ถูกต้องได้อย่างมีประสิทธิภาพยิ่งขึ้น โดยการจำกัดการปรับแก้ที่เป็นไปได้ให้แคบลงแต่ความแม่นยำเพิ่มมากขึ้น ผู้วิจัยคาดว่า การนำเอารูปแบบของการสะกดผิดเหล่านี้มาประยุกต์ใช้ปรับแก้การสะกดผิดจะสามารถเพิ่มประสิทธิภาพและความรวดเร็วในการแก้ไขคำที่สะกดผิดให้ถูกต้องแม่นยำและรวดเร็วมากยิ่งขึ้น

อย่างไรก็ดี การปรับแก้คำสะกดผิดตามที่กล่าวมาข้างต้นเพียงอย่างเดียวคงยังไม่เพียงพอต่อการเลือกคำที่ถูกต้องได้ เพราะทั้งคำที่สะกดผิดและคำที่สะกดถูกนั้นต่างก็เป็นคำที่มีปรากฏในพจนานุกรมทั้งคู่ การตรวจแก้คำสะกดผิดแบบเป็นคำจริงนี้จึงจำเป็นต้องอาศัยบริบทช่วยทั้งในการระบุคำที่สะกดผิดและเลือกคำที่ควรจะเป็นวิธีการหนึ่งที่คอมพิวเตอร์จะทำได้คือดูสายคำที่ปรากฏในข้อความว่าเป็นสายคำที่พบเป็นปกติหรือไม่ เพราะหากเป็นการสะกดคำผิดเป็นคำอื่นแล้ว คำที่ผิดนั้นก็ไม่ต้องจะพบเห็นบ่อยอยู่ในสายคำนั้น ในขณะที่คำที่ถูกควรเป็นตัวเลือกคำที่เมื่อนำมาแทนที่คำนั้นแล้วทำให้ได้สายคำที่เป็นปกติมากกว่า การตรวจแก้คำสะกดผิดแบบเป็นคำจริงจึงเป็นงานที่จำเป็นต้องพัฒนาต่อโดยอาศัยแนวคิดที่อิงบริบทในการระบุคำที่ผิดและคัดเลือกคำที่ถูกต้องมาแทน ผลของการวิเคราะห์ในงานนี้เป็นจุดเริ่มต้นที่ทำให้เข้าใจการสะกดผิดแบบเป็นคำจริงนี้มากขึ้น

รายการอ้างอิง

- กุหลาบ มัลลิกะมาส และวิพوخ โสภางค์. (2536). *การเขียน*. กรุงเทพฯ: โอเดียนสโตร์.
- คณาจารย์ภาควิชาภาษาไทยเพื่อการสื่อสาร คณะมนุษยศาสตร์. (2541). *ภาษาไทยเพื่อการสื่อสาร* (วิทยานิพนธ์มหาบัณฑิต สาขาครุศาสตร์). มหาวิทยาลัยทักษิณ, สงขลา.
- ชูศรี ศรีแก้ว. (2527). *การศึกษาข้อผิดพลาดในการเขียนภาษาไทยของนักศึกษาชั้นปีที่ 1 มหาวิทยาลัยเชียงใหม่ ปีการศึกษา 2525* (วิทยานิพนธ์มหาบัณฑิต สาขาศิลปศาสตร์). มหาวิทยาลัยเชียงใหม่, เชียงใหม่.
- จรรย์ เมธิตานนท์. (2549). *280 คำไทยที่มักเขียนและใช้กันผิด*. กรุงเทพฯ: มิติใหม่.
- ตระการ เอี่ยมตระกูล. (2554). *คำไทยที่มักใช้ผิด*. กรุงเทพฯ: คลื่นอักษร.
- ธนู ทดแทนคุณ. (2550). *ร้อยแปด (108) คำไทยที่มักใช้ผิด*. ปทุมธานี: สกายบุ๊กส์.
- ประกาศรี สีอำไพ และคณะ. (2544). *การใช้ภาษาไทย*. กรุงเทพฯ: ไทยวัฒนาพานิช.
- ฝ่ายวิชาการ พีบีซี. (2553). *ภาษาไทย คำที่มักเขียนผิด*. กรุงเทพฯ: พีบีซี.
- พยอม ธนมี. (2541). *การศึกษาลักษณะข้อผิดพลาดในการเขียนภาษาไทยของนักศึกษา ระดับประกาศนียบัตรวิชาชีพชั้นสูง ปีที่ 1 ในวิทยาลัยเทคนิคคนราธิวาส ปีการศึกษา 2539* (วิทยานิพนธ์มหาบัณฑิต สาขาการศึกษา). มหาวิทยาลัยทักษิณ, สงขลา.
- พัชรี วงจรสร้างสี. (2540). *การวิเคราะห์คำสะกดยากที่เขียนผิดของนักเรียนชั้นประถมศึกษาปีที่ 3 โรงเรียนสังกัดกรุงเทพมหานคร*. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ราชบัณฑิตยสถาน. (2557). *อ่านอย่างไรและเขียนอย่างไร ฉบับราชบัณฑิตยสถาน (แก้ไขเพิ่มเติม) ฉบับพิมพ์ครั้งที่ 22*. กรุงเทพฯ: ราชบัณฑิตยสถาน.
- สนม ครุฑเมือง. (2557). *คำที่มักเขียนผิดในภาษาไทย: การวิเคราะห์จากการตรวจผลงานการสอน. วารสารการจัดการ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏลำปาง, 7(1), 42-55.*
- สุนันท์ อัญชลีบุญกุล. (2552). *ระบบคำภาษาไทย*. กรุงเทพฯ: โครงการเผยแพร่ผลงานวิชาการ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

- เสนีย์ วิลาวรรณ. (2544). พัฒนาทักษะภาษา เล่ม 3 (กลุ่มสาระการเรียนรู้ภาษาไทย ช่วงชั้นที่ 3 ชั้นมัธยมศึกษาปีที่ 2 ตามหลักสูตรการศึกษาพื้นฐาน พุทธศักราช 2544). กรุงเทพฯ: วัฒนาพานิช.
- Aroonmanakun, W. (2002). *Collocation and Thai word segmentation*. Paper presented at the The Fifth Symposium on Natural Language Processing & The Fifth Oriental COCODSA Workshop, Pathumthani.
- Bowers, N. (2015). neilb/Text-Levenshtein. Retrieved from <https://github.com/neilb/Text-Levenshtein>
- Jurafsky, Dan. (2016). Minimum edit distance. Retrieved from <https://www.stanford.edu/class/cs124/lec/med.pdf>
- Kaur, J., & Garg, K. (2014). Hybrid approach for spell checker and grammar checker for Punjabi. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(6).
- Mishra, R., & Kaur, N. (2013). A survey of spelling error detection and correction techniques. *International Journal of Computer Trends and Technology*, 4(3), 372-374.
- Mitton, R. (1987). Spelling checkers, spelling correctors and the misspellings of poor spellers. *Information Processing and Management*, 23(5), 495-505.
- Murthy S, R., Madi, V., D, S., & Kumar P, R. (2012). A non-word Kannada spell checker using morphological analyzer and dictionary lookup method. *International Journal of Engineering Sciences & Emerging Technologies*, 2(2), 43-52.
- Stehouwer, J. H. (2011). *Statistical language models for alternative sequence selection*. Tilburg: TICC Dissertation Series 19.
- Verberne, S. (2002). *Context-sensitive spell checking based on word trigram probabilities*. (Unpublished master's thesis). University of Nijmegen, Netherlands.
- Wilcox-O'Hearn, A., Hirst, G., & Budanitsky, A. (2008). *Real-word spelling errection with trigrams: A reconsideration of the Mays, Damerau, and Mercer model* Paper presented at the CICLing-2008, Haifa.