



การใช้คลังข้อมูลภาษาและการเรียนรู้จากข้อมูลในการศึกษาคำปรากฏร่วม รูปแบบของคำ และอรรถสัมผัสของคำ: ตัวอย่างสื่อการสอนภาษาไทยเป็น ภาษาต่างประเทศ

The Use of Language Corpora and Data-driven Learning in the Study of Collocation, Language Pattern, and Semantic Prosody: Sample Materials for Teaching Thai as a Foreign Language

ภาสพงศ์ ศรีพิจารณ์¹
Passapong Sripicharn

บทคัดย่อ

บทความทางวิชาการชิ้นนี้มีวัตถุประสงค์เพื่อชี้ให้เห็นว่าคลังข้อมูลภาษาและการเรียนรู้จากข้อมูลสามารถนำมาใช้สอนภาษาไทยเป็นภาษาต่างประเทศได้อย่างมีประสิทธิภาพโดยเฉพาะการสอนเรื่องคำปรากฏร่วม รูปแบบของคำ และอรรถสัมผัสของคำ บทความเริ่มต้นด้วยการอธิบายความหมายและแหล่งของคลังข้อมูลภาษาทั้งคลังข้อมูลภาษาอังกฤษและคลังข้อมูลภาษาไทย และชี้ให้เห็นว่าผู้เรียนภาษาสามารถวิเคราะห์รูปแบบหรือการใช้ภาษาได้อย่างชัดเจนมากขึ้นการสังเกตตัวอย่างของคำหรือวลีที่ดึงมาจากคลังข้อมูลภาษา บทความในส่วนต่อไปอภิปรายความรู้ที่เกี่ยวข้องกับคำศัพท์ 3 เรื่อง คือ คำปรากฏร่วม รูปแบบภาษา และอรรถสัมผัสของคำ ซึ่งสามารถศึกษาวิเคราะห์ได้อย่างลึกซึ้งมากขึ้นโดยการวิเคราะห์คลังข้อมูลภาษา ตลอดจนอธิบายวิธีการเรียนรู้ภาษาจากการข้อมูลหรือ Data-driven Learning (DDL) ซึ่งเน้นให้ผู้เรียนวิเคราะห์ข้อมูลภาษาเพื่อหาคำตอบหรือข้อสรุปในประเด็นภาษาที่สนใจด้วยตัวเอง บทความในส่วนสุดท้ายนำเสนอตัวอย่างแบบฝึกหัดที่ออกแบบตามแนวคิดของการเรียนรู้จากข้อมูลโดยใช้ตัวอย่างข้อมูลจากคลังข้อมูลภาษาไทยแห่งชาติ พร้อมคำอธิบายและแนวทางการวิเคราะห์ข้อมูล บทความชิ้นนี้จบด้วยการนำเสนอแนวทางในการใช้แบบฝึกหัดดังกล่าวในการสอนที่เน้นการเรียนรู้จากข้อมูลและยกตัวอย่างงานวิจัยที่อาจต่อยอดได้จากการนำแบบฝึกหัดไปใช้ในการสอนภาษาไทยสำหรับชาวต่างประเทศ

¹อาจารย์ประจำสาขาวิชาภาษาอังกฤษ คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์ Email: passapong.s@arts.tu.ac.th



คำสำคัญ: การเรียนรู้จากข้อมูล (DDL), คลังข้อมูลภาษาไทยแห่งชาติ, การสอนภาษาไทยเป็นภาษาต่างประเทศ, คำปรากฏร่วม, รูปแบบภาษา, อรรถสัมผัสของคำ

Abstract

This academic article aims at proposing the use of language corpora and the Data-Driven Learning approach in teaching Thai as a foreign language, particularly in teaching collocation, language pattern, and semantic prosody. To start with, the article gives a definition of language corpora with information about the sources of language corpora in English and in Thai and explains how learners can examine language pattern and use more clearly by observing the data generated from the language corpora. The next part of the paper discusses three lexical features: collocation, language pattern, and semantic prosody, which can be deeply understood by a corpus observation. The paper goes on to describe a learning approach called Data-driven Learning or DDL in which learners are driven by the corpus data to study and generalize about language features that they are interested in. In the final part of the article, sample materials based on the Thai National Corpus designed for learners of Thai as a foreign language are presented with explanations and analysis guidelines. Finally, applications of the materials in the DDL context and some research ideas based on the use of the materials are also mentioned.

Keywords: Data-driven Learning (DDL), Thai National Corpus, Teaching Thai as a foreign language, collocation, language pattern, and semantic prosody.



1. บทนำ

1.1 คลังข้อมูลภาษากับการเรียนการสอนภาษา

เนื่องจากการเรียนรู้จากข้อมูล (Data-driven Learning) ซึ่งต่อไปนี้จะเรียกว่า ‘DDL’ มุ่งเน้นให้ผู้เรียนสังเกตรูปแบบและการใช้ภาษาจากตัวอย่างจริงของภาษาที่ต้องการศึกษาที่รวบรวมไว้ในฐานข้อมูลหรือคลังข้อมูลภาษา (Language Corpora) บทความวิชาการชิ้นนี้จึงเริ่มต้นด้วยการอธิบายความหมาย แนวคิดสำคัญ และทรัพยากรคลังข้อมูลภาษาที่ใช้ในการเรียนการสอนภาษา

คลังข้อมูลภาษา (Corpus หรือใช้ในรูปแบบพหูพจน์ Corpora) คือ การสร้างฐานข้อมูลภาษาโดยการรวบรวมบทความหรือข้อความในหัวข้อที่สนใจศึกษา และบันทึกข้อมูลเหล่านั้นในรูปแบบอิเล็กทรอนิกส์ เช่น บันทึกในรูปแบบไฟล์ข้อความ (Plain Text Format) เพื่อให้สามารถใช้โปรแกรมวิเคราะห์คลังข้อมูลภาษา (Corpus Analysis Software) ในการศึกษาวิเคราะห์หรือประมวลผลในรูปแบบใดรูปแบบหนึ่ง เช่น การสร้างรายการคำ (Word List) ตามความถี่ของคำที่ปรากฏหรือการค้นหา และแสดงคำหรือวลีพร้อมบริบทหรือที่เรียกว่าบรรทัดคอนคอร์ดแอนซ์ (Concordance Lines) คลังข้อมูลภาษามักมีขนาดใหญ่เพื่อให้ครอบคลุมหัวข้อที่หลากหลายและช่วยให้ผู้ศึกษาสามารถค้นหาคำและพบตัวอย่างการใช้คำในบริบทจริงให้มากที่สุดพอที่จะหาข้อสรุปที่น่าเชื่อถือ แต่บางครั้งคลังข้อมูลภาษาอาจมีขนาดไม่ใหญ่มากแต่เป็นตัวแทนของข้อความหรือบทความในหัวข้อเฉพาะที่ต้องการศึกษา และที่สำคัญการสร้างคลังข้อมูลภาษาต้องมีวัตถุประสงค์ที่ชัดเจนว่าจะศึกษาลักษณะภาษาในประเด็นใดเพื่อจะได้สามารถสร้างและใช้เครื่องมือที่เหมาะสมในการวิเคราะห์ข้อมูลเหล่านั้น (Hunston, 2000; Baker et.al, 2006; Crawford Csomay, 2016; วิโรจน์ อรุณมานะกุล, 2553)

คลังข้อมูลภาษาขนาดใหญ่โดยเฉพาะคลังข้อมูลภาษาอังกฤษสามารถสืบค้นได้รูปแบบของฐานข้อมูลออนไลน์ เช่น <http://corpus.byu.edu/> ซึ่งประกอบด้วยฐานข้อมูลย่อย เช่น Corpus of Contemporary American English (COCA), British National Corpus (BNC), Wikipedia Corpus ข้อดีของฐานข้อมูลประเภทนี้คือเป็นฐานข้อมูลที่มีขนาดใหญ่ มีการอัปเดตข้อมูลสม่ำเสมอ เช่น ล่าสุดมีฐานข้อมูลบทความเกี่ยวกับ Covid-19 และมีเครื่องมือค้นหาวิเคราะห์ข้อมูลที่หลากหลาย หรือถ้าผู้ศึกษาต้องการสร้างคลังข้อมูลเฉพาะทางเพื่อศึกษาคำศัพท์หรือลักษณะภาษาเฉพาะทางก็สามารถสร้างคลังข้อมูลภาษาของตัวเองได้โดยค้นหาและบันทึกข้อความที่ต้องการ



ศึกษาในรูปแบบของไฟล์ข้อความ (plain text หรือไฟล์ที่มีนามสกุล .txt) และใช้โปรแกรมวิเคราะห์คลังข้อมูลภาษา เช่น Antconc เพื่อวิเคราะห์และแสดงผลการศึกษาลักษณะทางภาษาที่ต้องการ (Anthony, L. (2021). Antconc (Version 4.0.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software/antconc/>)

ในการประมวลผลคลังข้อมูลออนไลน์หรือคลังข้อมูลที่สร้างขึ้นเอง ผู้ใช้จะต้องเลือกเครื่องมือวิเคราะห์คลังข้อมูลภาษาที่เหมาะสมกับประเด็นที่ต้องการศึกษาวิจัย เช่น ถ้าต้องการระบุหรือศึกษาคำศัพท์ที่พบบ่อยในคลังข้อมูลควรใช้เครื่องมือเพื่อการแสดงรายการคำตามความถี่ (Word List) หรือรายการคำสำคัญเมื่อวิเคราะห์เทียบกับคลังข้อมูลทั่วไปที่นำมาเปรียบเทียบ (Keyword List) ถ้าต้องการศึกษารูปแบบหรือการใช้ภาษาโดยการสังเกตจากตัวอย่างจำนวนมาก ควรเลือกใช้เครื่องมือที่เรียกว่า การแสดงตัวอย่างคำพร้อมบริบทหรือบรรทัดคอนคอร์ดแดนซ์ (Concordance Lines) ตัวอย่างเช่น การค้นหาและแสดงบรรทัดคอนคอร์ดแดนซ์ของคำว่า ‘capable’ จะช่วยให้ผู้วิจัยสามารถสังเกตรูปแบบและการใช้คำที่ต้องการศึกษาได้อย่างชัดเจน ในกรณีนี้จะสังเกตได้ชัดเจนว่าคำว่า ‘capable’ มีรูปแบบการใช้ที่พบบ่อยคือ ‘capable’ + of + noun/V-ING (ไม่ใช่ ‘capable + to + V1’ อย่างที่ผู้เรียนภาษาหลายคนใช้ผิด) ทั้งนี้เนื่องจากเครื่องมือดังกล่าวจะค้นหาตัวอย่างของคำว่า ‘capable’ จากแหล่งข้อมูลต่างๆที่อยู่ในคลังข้อมูล และเมื่อพบตัวอย่างแล้วจะนำเสนอบรรทัดคอนคอร์ดแดนซ์ของคำที่ค้นหาโดยนำคำที่ค้นหาไว้กลางหน้าจอพร้อมกับเรียงคำในบริบทที่อยู่ด้านขวามือหรือซ้ายมือตามลำดับตัวอักษรเพื่อให้คำที่ซ้ำกันเรียงอยู่ติดกันเพื่อให้สังเกตได้ง่าย ในกรณีนี้จึงเห็นคำว่า ‘of’ อยู่ในตำแหน่งด้านขวามือของคำว่า ‘capable’ อย่างชัดเจน



FREQUENCY	CONTEXT	CONTEXT +
long as the Heat are around , but it certainly looks	capable at this point : Portland I rail blazers # Christian	
if their existence depended on it . not for not being	capable but for the fear that something might go wrong , which	
but as a fast , high energy player and physics :	capable defender He said " the most important side is the def	
Sometimes , the whole point of experience is to make you	capable enough to change : # I do agree with Laura that the	
a number of staff changes , including the resignation of its	capable former chairman Stratis Stratisis . He left after a scrap	
new model will become freely accessible by any player with a	capable gaming rig and internet connection . This includes all e	
Jericho . And I'm willing to entrust it to your	capable hands What ? I just said I'd give you the	
in the world , but I think he 's shown how	capable he is at all of these things while working with a budget	
the hands are always in motion , conveying all the complexities	capable in any spoken language . And I believe even beyond th	
on the skidpad , but they felt refined , comfortable and	capable in real-world driving situations . # GM could design a v	
the following highlights : # Metro browser registration	capable installer # Zip builds which do registration upon settir	
plan is that Coalition forces will use the known number of	capable Iran units to * clear , hold and build " a peaceful	
. We're used to that . There are many perfectly	capable it folks that would rather get the download and possit	
people these days want leaders . They want people that are	capable leaders that have their best interests at heart and in th	
the Brits got involved in this key juncture with some very	capable military adventurers They played the competing indie	
diagonal side play . Using this configuration , the TV-50is is	capable of 1.9 ' ' resolution . Useful visual powers range betw	
the i90c , which weighs a mere 4 pounces , is	capable of 120 minutes of talk time with 45 hours of standby o	
resonated and , quite frankly , the idea that anyone is	capable of a suicide prank is n't really on my radar . #	
to Cray Research 's timetable , it will have a computer	capable of a sustained terafoips on real applications by 1997 .	
uttering the monosyllable " Ouf ! ouf ! " which is	capable of a variation of tone and meaning truly extraordinary	
with new roads and expressways , while a modern airport :	capable of accommodating wide-body jets , is near completion	
suffer substantial change , it would no longer be an instrument	capable of accomplishing the work for which God called it in to	
a medication similar to Penegra given that they need to be	capable of achieve and make a more extraordinary tougher er	
in networked and stand-alone motion control applications :	capable of achieving video stepping performance of up to 51.2	
world plan to launch high-resolution imagery satellites , some	capable of achieving resolutions fine enough to detect objects	
are actually many states in international society that are	capable of acting as humanitarian agents . In contrast to ordin	
is perfect . That is not to say we are not	capable of acts of great virtue , but at the end of the	
respondents indicated that they did not see themselves as	capable of affecting change , their open-ended responses sugg	
. It must have " universal quantum gates " logical elements	capable of arranging any desired logical relationship between t	
who have received advanced training in moral theology , are most	capable of attaining accurate moral knowledge consistently . A	

ภาพที่ 1 ตัวอย่างของคำว่า ‘capable’ จากคลังข้อมูลภาษาออนไลน์

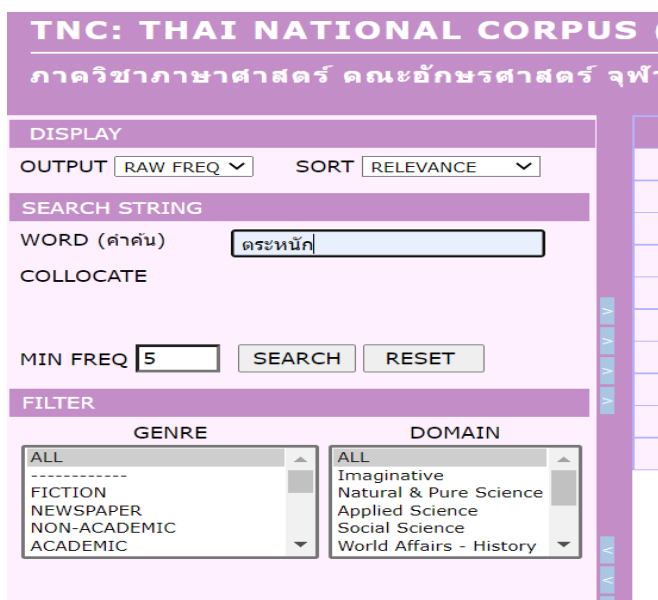
<http://corpus.byu.edu/>

แม้คลังข้อมูลภาษาส่วนใหญ่จะเป็นคลังข้อมูลภาษาอังกฤษและภาษาต่างประเทศอื่น ๆ และนำมาประยุกต์ใช้ในการวิจัยหรือการเรียนการสอนภาษาต่างประเทศ เช่น คลังข้อมูลออนไลน์ภาษาฝรั่งเศส <https://www.sketchengine.eu/frtnten-french-corpus/> หรือคลังข้อมูลออนไลน์ภาษาเยอรมัน <https://www.sketchengine.eu/detenten-german-corpus/> แต่ในปัจจุบันมีการพัฒนาคลังข้อมูลภาษาไทยที่สามารถสืบค้นตัวอย่างของคำพร้อมบริบทและนำข้อมูลมาประยุกต์ใช้ในการสอนภาษาไทยได้เช่นกัน คลังข้อมูลภาษาไทยที่ใช้กันอย่างแพร่หลายแหล่งหนึ่งคือคลังข้อมูลภาษาไทยแห่งชาติ <http://www.arts.chula.ac.th/~ling/tnc3/>) ซึ่งเป็นโครงการที่จัดทำโดยภาควิชาภาษาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ประกอบด้วยข้อมูลภาษาไทยจำนวนประมาณสี่แสนคำที่รวบรวมจากบริบทที่หลากหลาย เช่น หนังสือพิมพ์ สุนทรพจน์ รายงานข่าว บทความวิชาการ นิยายและเรื่องสั้น โดยผู้ใช้งานสามารถสืบค้นและแสดงผลการค้นหาลงทางออนไลน์ ตัวอย่างเช่น ถ้าต้องการค้นหาตัวอย่างพร้อมบริบทของคำว่า “ตระหนัก” สามารถทำตามขั้นตอนดังต่อไปนี้

ขั้นตอนที่ 1: ไปที่หน้าเว็บไซต์ของคลังข้อมูลภาษาไทยแห่งชาติ

<http://www.arts.chula.ac.th/~ling/tnc3/>

ขั้นตอนที่ 2: พิมพ์คำว่า “ตระหนัก” ในช่องสำหรับพิมพ์คำที่ต้องการค้นหา ผู้ใช้งานสามารถเลือกประเภทของข้อมูลที่ต้องการได้ในช่องเมนูประเภทของข้อความ (GENRE) เช่น วรรณกรรม (FICTION) หนังสือพิมพ์ (NEWSPAPER) หรือ งานเขียนทางวิชาการ (ACADEMIC) เป็นต้น แต่โดยปกติแล้วจะเลือก “ทั้งหมด” (ALL) เพื่อให้ได้ตัวอย่างมากที่สุด ยกเว้นต้องการแยกแยะความแตกต่างระหว่างจำนวนหรือการใช้คำในประเภทของสารที่แตกต่างกัน



ภาพที่ 2 หน้าจอแสดงคำที่ต้องการค้นหบนหน้าหลักของคลังข้อมูลภาษาไทยแห่งชาติ

ขั้นตอนที่ 3: กดปุ่มค้นหา หรือ SEARCH เพื่อให้โปรแกรมแสดงตัวอย่างของคำที่ต้องการค้นหา หน้าจอจะแสดงจำนวนคำที่พบทั้งหมดในประเภทสารต่าง ๆ จากนั้นให้เลือกประเภทของสารที่ต้องการหรือเลือกที่ TOT (ทั้งหมด)



TNC: THAI NATIONAL CORPUS (Third Edition)ในพระราชูปถัมภ์สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี

ภาควิชาภาษาศาสตร์ ดอนเมืองนครศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

DISPLAY

OUTPUT RAW FREQ SORT RELEVANCE

SEARCH STRING

WORD (คำค้น)ตระหนัก

COLLOCATE

MIN FREQ5 SEARCH RESET

	TOT	FICTION	NEWSPAPER	NON-ACADEMIC	ACADEMIC	LAW	MISC
1ตระหนัก	2572	159	336	432	1274	3	368
TOTAL	2572	159	336	432	1274	3	368

ภาพที่ 3 หน้าจอแสดงจำนวนคำที่พบทั้งหมดของคำว่า “ตระหนัก” จากหน้าค้นหาของคลังข้อมูลภาษาไทยแห่งชาติ

เมื่อศึกษาตัวอย่างที่ปรากฏจะพบว่าคำว่า “ตระหนัก” มักใช้ร่วมกับคำว่า “ทำให้” “ทำให้ (บุคคล) “เป็นที่” “ย่อมจะ” ในบริบททางด้านซ้าย และมักปรากฏร่วมกับคำว่า “ถึง” “ใน” หรือ “ว่า” ในบริบทด้านขวามือ สังเกตได้ว่าหน้าจอจะแสดงตัวอย่าง 100 ตัวอย่างแบบสุ่มในเบื้องต้นก่อน ถ้าต้องการดูตัวอย่างเพิ่มเติมผู้ใช้สามารถสั่งให้โปรแกรมแสดงข้อมูลที่เหลือเพิ่มเติมได้

ตัวอย่างประโยคที่ปรากฏ	
1	ACHM005 เรื่องเป็นอดีตกาล แต่ก็มิใช่ออกเว้นสำหรับงานหลังสมัยใหม่ที่ปรากฏการเล่าเรื่องเป็นปัจจุบันกาล แต่การอ่านงานประเภทนี้ทำให้ผู้อ่านตระหนักตลอดเวลากว่ากำลังฟังเรื่องราวที่ผู้เล่าเรื่อง สร้างสรรค์ ในขณะนั้น ลองดูตัวอย่างต่อไปนี้ ตอนที่ สุวัฒน์ ตื่นขึ้นมาหลังจากหลับ
2	ACHM025 แลสิ่งกลายเป็นคนคลุ้มคลั่งความดีความงามที่ปรากฏแก่ผู้อ่านและเป็นตัวแทนของอากัปกริยาที่พิสดารไม่ปรากฏนางกรรพาลด้วยตนเอง แต่ตระหนักว่ากรรพาลในใจ คือการแหวก วน ขึ้นไปในท้องฟ้า อันกว้างไกล เพื่อเสาะหา สุนทรโฆษ นรา
3	ACHM025 แลอดีตในทางโลกจากปัจจุบันออกไปอีก บรรทัดต่อมา สารพันทั้งปวง ยิ่งโรยรา เป็นปรากฏการณ์ทุกอย่างหนึ่งในธรรมชาติที่บุคคลผู้เขียนตระหนักว่าดำเนินควบคู่ไปพร้อมๆ กันมาเวลาส่วนหนึ่งนั้น การกล่าวถึงความเลื่อมพรายได้เกิดขึ้นและยังไม่สิ้นสุดลง จนกว่าจะ
4	ACHM025 ช่วงที่สามของบทกวีได้ขยายความกว้างที่เห็นไว้ เมื่อบุคคลผู้เขียนตระหนักในความเป็นจริงของธรรมชาตินั้น เขาได้พบ โลกใหม่ การพบเห็นนี้เกิดขึ้นภายหลังที่ เครื่องปั้นกลางทางจึงที่หลายของ ซึ่งเป็นผลจากการ คำนี้ถึง ขณะสุดท้ายแห่งวัน
5	ACHM025 โดยขั้นนี้ หากทำความเข้าใจต่อการแบ่งความตายได้เร็วเพียงไร ก็ย่อมจะตระหนักในสมมติปริศนาก็ได้เร็วเพียงนั้น แลขั้นนี้คือที่มาของการมองเห็นทวนกลับถึงสิ่งที่ผ่านมาทั้งหมดในชีวิตรวดผ่านไปอย่างรวดเร็ว ดังสองบรรทัดสุดท้ายของกวีนิพนธ์กล่าวไว้อย่างมีนัยนัยใจว่า
6	ACHM054 อัลได้เล็ดลอดหนีจากที่ของวราห์เมื่อปี 4 ปีก่อนคือ ก. ทำให้ความจริงเป็นที่ยอมรับเห็นความเท็จ เช่น ซอเครตีส แสดงให้เห็นว่าปรัชญาของเขาเป็นจริงเพียงแท้
7	ACHM054 กอรจัส มีอะไรหรือ ซอเครตีส ซอเครตีส ข้าพเจ้าขอร้องให้ ข้าพเจ้าตระหนักได้ว่า ฟังจากท่านแล้ว ข้าพเจ้าไม่ทราบว่าจะไรเป็นลักษณะที่แท้จริงหรือไรเป็นหัวข้อของการโน้มน้าวที่ท่านพูดถึง หรือว่าวาทศิลป์ประกอบคลุม ถึงแม้ข้าพเจ้าจะ
8	ACHM055 27 เดือนเริ่มในสรรพนาม จัน ซึ่งเป็นกรสร้างความคิดเกี่ยวกับ self ในลักษณะที่แยกตนเองจากคนอื่นอย่างเด่นชัด ทำให้ตระหนักในความมีตัวตนของตนเองได้อย่างแท้จริง หลังจากสร้างอัตมโนทัศน์ในลักษณะดังกล่าวแล้ว ลักษณะอื่นๆ จดตามมา
9	ACHM056 ในบ้าน หมายถึงการบรรณาการความรู้สึกของตนมารู้สิ่งที่อยู่รอบๆ ตัวในขณะนั้นอย่างเต็มที่และชัดเจน ความสำคัญที่เห็นชัดเป็นการตระหนักถึงพลังที่รู้สึกที่อยู่รอบๆ ตัวใน ขณะนี้ อย่างลึกซึ้งและชัดเจน ความสำคัญทำให้เกิดความเป็นเนื้อเดียวกันระหว่างตนเองกับสิ่งแวดล้อมรอบ
10	ACHM056 จำเป็นที่สวดของจิตที่กลายกลายเป็น การสร้างความรู้สึกตัวให้เกิดขึ้น งานของนักจิตวิทยาคือการเอื้ออำนวยให้ผู้รับการบำบัดมีความรู้สึกตัวของเขา ซึ่งการดังกล่าวทำให้เขาไม่สามารถไปศึกษาภาพที่เขาเผชิญได้อย่างเต็มที่ เมื่อไรที่ 1970

ภาพที่ 4 ตัวอย่างของคำว่า ‘ตระหนัก’ จากหน้าค้นหาของคลังข้อมูลภาษาไทยแห่งชาติ

จะเห็นได้ว่าการวิเคราะห์คลังข้อมูลภาษาไทยโดยเฉพาะการแสดงผลตัวอย่างของคำพร้อมบริบทของคำที่ต้องการศึกษาไม่ว่าจะเป็นคำภาษาใดก็ตามสามารถช่วยให้ผู้ศึกษาสังเกตรูปแบบและการใช้คำในภาษานั้น ๆ ได้ชัดเจนมากขึ้น นอกเหนือจากจะเป็นข้อมูลและเครื่องมือในการศึกษาวิเคราะห์ลักษณะภาษาในเชิงภาษาศาสตร์แล้ว ยังสามารถนำไปประยุกต์ใช้ในการสอนภาษาต่าง ๆ



โดยเฉพาะการสอนภาษาไทยเป็นภาษาต่างประเทศ บทความขึ้นนี้จะเน้นที่การใช้คลังข้อมูลภาษาไทยไปใช้ในการพัฒนาแบบฝึกหัดเพื่อสอนคำศัพท์ภาษาไทยแก่ผู้เรียนที่เรียนภาษาไทยเป็นภาษาต่างประเทศโดยเน้นที่การใช้คำปรากฏร่วม (Collocation) รูปแบบภาษา (Language Pattern) และความหมายแฝงของคำ (Connotation) ซึ่งจะอภิปรายในรายละเอียดในหัวข้อต่อไป

1.2 คลังข้อมูลภาษากับการวิเคราะห์คำ (Corpora and Lexical Analysis)

คลังข้อมูลภาษาและรูปแบบการนำเสนอข้อมูลที่เกี่ยวข้องกับการสังเกตการใช้คำในบริบทการใช้งานจริงช่วยให้ผู้ศึกษาเรื่องคำ (Lexicology) สามารถวิเคราะห์และสังเกตเห็นปรากฏการณ์ทางภาษาที่อาจจะไม่สามารถเห็นได้ชัดเจนจากตัวอย่างที่จำกัด ลักษณะทางภาษาลักษณะหนึ่งที่สังเกตได้ชัดเจนจากการวิเคราะห์คลังข้อมูลภาษาคือ คำปรากฏร่วม (Collocation) ซึ่งหมายถึงการที่คำบางคำมักใช้ร่วมกันหรือปรากฏในบริบทที่ใกล้เคียงกันในข้อความไม่ว่าจะภาษาใดก็ตาม ตัวอย่างเช่น ในภาษาอังกฤษคำปรากฏร่วมแบ่งออกเป็น 2 ประเภทคือ คำปรากฏร่วมระหว่างคำศัพท์กับคำทางไวยากรณ์ (Grammatical Collocation) เช่น Noun + Preposition หรือ Adjective + Preposition (เช่น ‘admiration for’ หรือ ‘amazed at’ และคำปรากฏร่วมระหว่างคำศัพท์กับคำศัพท์ (Lexical Collocation) เช่น Verb + Noun หรือ Adjective + Noun (เช่น to override a veto หรือ a formidable challenge) (Benson et.al. 1997) มีงานวิจัยจำนวนมากชี้ให้เห็นถึงความสำคัญคลังข้อมูลภาษาในการศึกษาคำปรากฏร่วม เช่น แนวคิดเรื่อง Lexical Priming ของ Hoey (2012) ที่แสดงผสมผสานแนวคิดเรื่องคลังข้อมูลภาษาและภาษาศาสตร์จิตวิทยา (Psycholinguistics) ในการทดสอบการรับรู้คำปรากฏร่วม (Collocation Recognition) และผลการทดลองสรุปได้ว่าการที่ผู้ใช้ภาษามีประสบการณ์ทางภาษา เช่น เคยอ่านหรือใช้คำในภาษาจะสามารถเชื่อมโยงหรือระบุคำปรากฏร่วมหรือไวยากรณ์ที่เกี่ยวข้องกับคำนั้นได้อย่างถูกต้องเป็นธรรมชาติและผลการวิเคราะห์ของผู้ทดสอบสอดคล้องกับผลการวิเคราะห์คำปรากฏร่วมของคำที่ศึกษาจากคลังข้อมูลภาษา แสดงให้เห็นว่าถึงแม้ผู้เรียนภาษาจะไม่ใช่ว่าเจ้าของภาษาที่มีประสบการณ์ในการใช้ภาษาอย่างลึกซึ้ง แต่การมีตัวอย่างจำนวนมากจะช่วยให้ผู้เรียนภาษาสามารถใช้คำได้อย่างเป็นธรรมชาติในภาษานั้นมากขึ้นและยังสามารถเห็นความเชื่อมโยงระหว่างความหมายของคำปรากฏร่วมในกลุ่มเดียวกันด้วย ด้วยเหตุนี้จึงมีผู้นำคลังข้อมูลภาษามาประยุกต์ใช้ในการสอนหรือสร้างความตระหนักรู้เกี่ยวกับคำปรากฏร่วมแก่ผู้เรียนภาษา เช่น



Timmis (2015) สร้างแบบฝึกหัดที่ให้ผู้เรียนค้นหาคำปรากฏร่วมของคำว่า ‘carry out’ ใน British National Corpus ซึ่งเป็นฐานข้อมูลหนึ่งของ <http://corpus.byu.edu/> ถ้าผู้เรียนค้นหาคำดังกล่าวโดยใช้เครื่องมือที่ชื่อ collocate จะพบว่าคำนามที่ทำหน้าที่เป็นกรรมของกริยา ‘carry out’ มักเป็นคำที่เกี่ยวข้องกับงานที่ต้องใช้เวลาในการศึกษาหรือดำเนินการ เช่น ‘task’ ‘research’ ‘assessment’ ‘inspection’ หรือ ‘survey’ ตามที่ปรากฏในรูปภาพต่อไปนี้

ON CLICK: [CONTEXT](#) [TRANSLATE \(TH\)](#) [GOOGLE](#)

HELP		
1	☐	THEIR
2	☐	RESEARCH
3	☐	REPAIRS
4	☐	TASKS
5	☐	CERTAIN
6	☐	INSPECTIONS
7	☐	IMPROVEMENTS
8	☐	EXPERIMENTS
9	☐	VARIOUS
10	☐	ATTACKS
11	☐	INSTRUCTIONS
12	☐	REGULAR
13	☐	ASSESSMENTS
14	☐	CHECKS
15	☐	SURVEYS
16	☐	MAINTENANCE
17	☐	TESTS
18	☐	SPECIFIC
19	☐	URGENT
20	☐	EMERGENCY
21	☐	ORDERS
22	☐	SEARCHES
23	☐	DISCLOSED
24	☐	CALCULATIONS
25	☐	INVESTIGATIONS
26	☐	EXTENSIVE

ภาพที่ 5 ผลการค้นหาคำปรากฏร่วมของ ‘carry out’ ใน British National Corpus

การศึกษาการใช้คำปรากฏร่วมให้เป็นธรรมชาติสามารถประยุกต์ใช้ได้กับการเรียนการสอนภาษาต่างประเทศอื่น ๆ ด้วย ตัวอย่างเช่น Vyatkina (2016) สร้างแบบฝึกหัดจากคลังข้อมูลภาษาเยอรมันเพื่อให้ผู้เรียนสังเกตการใช้คำปรากฏร่วมระหว่างคำกริยากับคำบุพบท และนำไปทดลองใช้กับผู้เรียนภาษาเยอรมันเป็นภาษาต่างประเทศที่มีระดับภาษาแตกต่างกัน ผลการวิจัยพบว่ากลุ่มผู้เรียนที่เรียนรู้จากคลังข้อมูลภาษามีผลคะแนนการทดสอบการใช้คำปรากฏร่วมสูงขึ้นอย่างมีนัยยะสำคัญเมื่อเทียบกับกลุ่มที่เรียนคำปรากฏร่วมแบบดั้งเดิม งานวิจัยของ Pereira & Manguilimotan & Matsumoto (2016) เป็นอีกตัวอย่างหนึ่งที่แสดงให้เห็นว่าการใช้คลังข้อมูลภาษาสามารถช่วยให้ผู้เรียนภาษาต่างประเทศสามารถใช้คำปรากฏร่วมได้อย่างเป็นธรรมชาติมาก



ยิ่งขึ้น ในกรณีนี้เป็นการใช้คลังข้อมูลภาษาญี่ปุ่นเพื่อช่วยให้ผู้เรียนใช้คำปรากฏร่วมระหว่างคำกริยาและคำนามได้อย่างถูกต้องโดยเฉพาะคำที่มีความหมายใกล้เคียงกันจนทำให้เกิดความสับสนในการใช้ นอกจากนี้มีงานวิจัยที่ศึกษาเรื่องคำปรากฏร่วมจากคลังข้อมูลภาษาไทย ตัวอย่างเช่น เกียรติชัย เดชพิทักษ์ศิริกุล (2561) ศึกษาคำปรากฏร่วมของกริยา “เป็น” และ “คือ” จากคลังข้อมูลภาษาไทยแห่งชาติ และศึกษาปัญหาในการใช้คำปรากฏร่วมของกริยาสองคำนี้โดยการสร้างคลังข้อมูลภาษาผู้เรียน (Learner Corpus) จากการรวบรวมงานเขียนที่เป็นบันทึกประจำวันสัปดาห์ของนักศึกษาชาวจีนที่เรียนภาษาไทย โดยผู้วิจัยได้นำผลการวิเคราะห์มาออกแบบสื่อการเรียนการสอน และจากการวัดผลสัมฤทธิ์ของแบบฝึกหัดดังกล่าวพบว่าผู้เรียนภาษาไทยมีความเข้าใจและใช้คำปรากฏร่วมของคำกริยาที่ศึกษาได้ดีขึ้นโดยผลการสอบหลังจากการทดลองสูงกว่าผลสอบก่อนการทดลองอย่างมีนัยยะสำคัญ

รูปแบบการใช้ภาษา (Language Pattern) เป็นลักษณะทางภาษาอีกประการหนึ่งที่ผู้เรียนสามารถศึกษาวิเคราะห์ได้โดยการศึกษาวเคราะห์คลังข้อมูลภาษา แนวคิดที่สนับสนุนการใช้คลังข้อมูลภาษาในการสังเกตรูปแบบการใช้คำที่มีความหลากหลายในภาษาอังกฤษ เช่น แนวคิดเรื่องไวยากรณ์รูปแบบภาษา (Pattern Grammar) ของ Hunston and Francis (2000) หรือแนวคิดเรื่อง (Corpus Pattern Analysis) ของ Hanks and Ma (2021) แสดงให้เห็นว่าตัวอย่างจำนวนมากจากคลังข้อมูลภาษาอังกฤษช่วยให้สามารถระบุรูปแบบการใช้คำที่หลากหลายได้จากการสังเกตตัวอย่างการใช้งานจริง เช่น รูปแบบการใช้คำว่า ‘explain’ ที่นำเสนอในงานของ Hunston and Francis (2000, pp. 45-46)

The pattern of ‘explain’ (p. 45-46)

Patterns	Examples
V n	...trying to explain Britain's thinking on...
V with quote	...Monte Fresco explains: 'He's so consistent'
V wh	This helps explain why he was not better known...
V about n	...then went on to explain about...
V that	She explained that she never paid...
V to n	Dentists have to explain to their patients that...
V	Let me explain.

งานวิจัยของ Jones and Waller (2015) เป็นอีกตัวอย่างหนึ่งที่ศึกษารูปแบบภาษา ‘Pronoun’ + was + wondering ในคลังข้อมูล British National Corpus (<https://www.english-corpora.org/bnc/>) ผลการวิเคราะห์พบว่าคำว่า ‘wondering’ มีรูปแบบ



การใช้ที่หลากหลายมากกว่ารูปแบบ ‘wondering if’ ที่ผู้เรียนส่วนใหญ่คุ้นเคย เช่น เจ้าของภาษาใช้รูปแบบอื่น ๆ เช่น ‘I was wondering if’ ‘I was wondering whether’ ‘I was wondering about’ หรือ ‘I was wondering how’ ถึงแม้รูปแบบ 3 รูปแบบหลังจะมีความถี่ต่ำกว่ารูปแบบแรก

การศึกษาคำศัพท์อีกแง่มุมหนึ่งที่น่าจะมีความลุ่มลึกชัดเจนขึ้นจากการสังเกตและวิเคราะห์คลังข้อมูลภาษาคือเรื่องความหมายแฝงของคำ (connotation) โดยเฉพาะความหมายแฝงที่เรียกว่า “อรรถสัมผัสของคำ” (Semantic Prosody) งานวิจัยที่สำคัญเช่น Partington (1998) Partington et.al. (2013) และ Stewart (2010) ชี้ให้เห็นว่าคลังข้อมูลภาษาสามารถช่วยแยกแยะความแตกต่างระหว่างความหมายแฝงทั่วไปกับอรรถสัมผัสของคำได้ กล่าวคือในขณะที่คำบางคำมีความหมายแฝงในทางบวกหรือทางลบอย่างชัดเจนในตัวคำเองหรือบางครั้งอาจขึ้นกับการตีความจากประสบการณ์และสังคมวัฒนธรรมของผู้ใช้คำนั้น แต่อรรถสัมผัสของคำมักจะต้องสังเกตจากบริบทหรือคำปรากฏร่วมของคำโดยที่คำนั้นโดยตัวเองมีความหมายเป็นกลางไม่บวกหรือลบอย่างชัดเจน ตัวอย่างเช่น คำว่า ‘set in’, ‘undergo’, หรือ ‘break out’ โดยตัวคำเองมีความหมายเป็นกลางๆ แต่ถ้าสังเกตคำปรากฏร่วมของคำกลุ่มนี้จะพบว่าเป็นคำที่มีความหมายเป็นลบเป็นส่วนใหญ่จึงทำให้คำกลุ่มนี้มีความหมายแฝงในเชิงลบไปด้วย อรรถสัมผัสของคำสามารถสังเกตได้ในภาษาอื่น ๆ ด้วยเช่นกันรวมถึงภาษาไทยด้วย ตัวอย่างเช่น ชาตบุษย์ ลีละรัตนานนท์ (2551) ศึกษาคำปรากฏร่วมและอรรถสัมผัสของคำว่า “เกิด” จากคลังข้อมูลภาษาไทยที่รวบรวมจากข่าวหนังสือพิมพ์ ผลการศึกษาพบว่าคำว่า “เกิด” มีแนวโน้มที่มีอรรถสัมผัสทางความหมายในทางลบมากกว่าทางบวกหรือเป็นกลางถึงแม้ว่าในบางบริบท เช่น สุนทรพจน์ของผู้นำ คำว่า “เกิด” อาจมีความหมายแฝงทางบวกมากกว่าบริบทอื่น นอกจากนี้คลังข้อมูลภาษาไทยสามารถนำมาใช้ในการศึกษาความหมายของคำตามแนวทางของอรรถศาสตร์ (semantics) ได้อีกด้วย เช่น งานของสรณ์ แอบเงิน และ กิ่งกาญจน์ เทพกาญจนา (2560) ที่ศึกษาความหมายของคำบุพบท “ของ” โดยใช้ข้อมูลจากคลังข้อมูลภาษาไทยแห่งชาติ

กล่าวโดยสรุปจะพบว่าการวิเคราะห์ศึกษาตัวอย่างของคำหรือวลีจากคลังข้อมูลภาษา โดยเฉพาะคลังข้อมูลภาษาอังกฤษช่วยให้เข้าใจธรรมชาติของคำหรือวลีที่ต้องการศึกษามากขึ้น เช่น คำปรากฏร่วม รูปแบบการใช้ และความหมายแฝงหรืออรรถสัมผัสของคำ ถึงแม้ว่ากรณีศึกษาส่วนใหญ่ที่พบในงานวิจัยจะใช้ตัวอย่างการวิเคราะห์จากคลังข้อมูลภาษาอังกฤษ แต่แนวทางการวิเคราะห์ดังกล่าวสามารถนำมาประยุกต์ใช้กับการวิเคราะห์คำในภาษาไทย และใช้เป็นข้อมูลในการ



สอนภาษาไทยเป็นภาษาต่างประเทศ โดยเฉพาะอย่างยิ่งเมื่อนำหลักการเรียนรู้จากข้อมูล (Data-driven Learning) มาใช้ในการส่งเสริมการเรียนรู้ด้วยตนเองของผู้เรียน

1.3 ข้อมูลเบื้องต้นเกี่ยวกับการเรียนรู้จากข้อมูล (Data-driven Learning หรือ DDL)

การเรียนรู้จากข้อมูล (Data-driven Learning หรือ DDL) เป็นแนวทางการเรียนภาษาที่มีที่มาจากการเรียนภาษาอังกฤษเป็นภาษาต่างประเทศโดยใช้ข้อมูลจากคลังข้อมูลภาษา วิธีการเรียนภาษาดังกล่าวมีจุดมุ่งหมายให้ผู้เรียนได้สังเกตและศึกษาวิเคราะห์ข้อมูลจากคลังข้อมูลภาษาในรูปแบบของบรรทัดคอนคอร์แดนซ์เพื่อหาคำตอบหรือข้อสรุปเกี่ยวกับรูปแบบการใช้ภาษาด้วยตัวของผู้เรียนเอง ตามที่นักภาษาศาสตร์และอาจารย์สอนภาษาอังกฤษชื่อ Tim Johns ซึ่งเป็นผู้ริเริ่มและตั้งชื่อแนวทางการเรียนรู้แบบ DDL ได้ให้คำจำกัดความไว้

‘The use in the classroom of computer-generated concordances to get students to explore the regularities of patterning in the target language, and the development of activities and exercises based on concordance output’ (Johns and King (eds.), 1991: iii).

ลักษณะเฉพาะของแนวทางการเรียนแบบ DDL ที่สรุปได้จากงานของ Johns (1991) และ Friginal (2018) คือเป็นแนวทางการเรียนแบบอุปนัย (Inductive Learning) ที่เน้นผู้เรียนเป็นศูนย์กลาง คือเน้นให้ผู้เรียนตั้งคำถามหรือคิดประเด็นการเรียนรู้ด้านการใช้ภาษาด้วยตนเอง และหาคำตอบจากข้อมูลที่มีรูปแบบเฉพาะซึ่งหมายถึงข้อมูลจากคลังข้อมูลภาษาที่นำเสนอในรูปแบบของบรรทัดคอนคอร์แดนซ์ที่ผู้สอนเตรียมไว้หรือการวิเคราะห์คลังข้อมูลด้วยตัวเองแทนที่ผู้สอนจะอธิบายหรือสอนการใช้ภาษาโดยตรง (ดังนั้นจึงเรียกแนวทางนี้ว่า ‘Data-driven’) ลักษณะเฉพาะที่สำคัญอีกประการหนึ่งของวิธีการเรียนแบบ DDL คือ เป็นวิธีการที่ไม่จำเป็นต้องมีการกำหนดเป้าหมายหรือวัตถุประสงค์การเรียนรู้ที่ชัดเจนตายตัวเสมอไปตามแนวทางแบบ Serendipity Learning ซึ่งหมายความว่าผู้เรียนอาจจะเริ่มต้นด้วยการตั้งคำถามหรือกำหนดประเด็นการค้นหาก็หรือวิเคราะห์ข้อมูลแบบหนึ่ง แต่ในกระบวนการการวิเคราะห์ข้อมูลอาจจะพบประเด็นหรือข้อมูลอื่น ๆ ที่น่าสนใจและเปลี่ยนเป้าหมายการวิเคราะห์เป็นประเด็นอื่นได้ในตอนท้ายของบทเรียนหรือการสังเกตข้อมูล นอกจากนี้บทบาทของผู้สอนในกระบวนการเรียนรู้แบบ DDL จะเปลี่ยนผู้สอน



(Teacher) ที่อธิบายกฎไวยากรณ์หรือคำศัพท์โดยตรงเป็นวิทยากรกระบวนการ (Facilitator) ที่ช่วยให้คำแนะนำในกระบวนการวิเคราะห์ข้อมูลของผู้เรียน

กระบวนการเรียนรู้แบบ DDL อาจจะปรับเปลี่ยนได้ตามบริบทของการเรียนการสอนและความสามารถของผู้เรียน จะเห็นได้ว่าในช่วงที่ Tim Johns เริ่มต้นใช้วิธีการเรียนรู้แบบ DDL ในการศึกษาการใช้คำศัพท์ในการเขียนทางวิชาการนั้น ผู้สอนใช้วิธีการที่ไม่ได้เน้นผู้เรียนเป็นศูนย์กลางแบบเต็มรูปแบบ คือ ‘Identify-Classify-Generalize’ โดยผู้สอนระบุประเด็นการเรียนรู้และคัดเลือกบรรทัดคอนคอร์แดนซ์ที่เหมาะสม (Identify) จากนั้นนำข้อมูลดังกล่าวมาจัดกลุ่มตามรูปแบบการใช้ (Classify) และจากนั้นให้ผู้เรียนได้วิเคราะห์ความแตกต่างของการใช้ภาษาในแต่ละกลุ่มและหาข้อสรุปด้วยตัวเอง (Generalize) ต่อมามีการใช้กระบวนการแบบ DDL แบบเน้นผู้เรียนมากยิ่งขึ้น เช่น Claire Kennedy and Tiziana Miceli (2001) ใช้ขั้นตอน 4 ขั้นตอนในการทำกิจกรรมตามแนวทาง DDL ได้แก่ การตั้งคำถามหรือประเด็น (Formulating a question) การค้นหาข้อมูลจากคลังข้อมูลภาษา (Devising a search strategy) การสังเกตหรือวิเคราะห์ข้อมูล (Observing the examples) และการหาข้อสรุป (Drawing conclusion) โดยแต่ละขั้นตอนจะเน้นผู้เรียนเป็นผู้ดำเนินการเองเป็นส่วนใหญ่โดยมีผู้สอนเป็นผู้ให้คำแนะนำและช่วยให้การเรียนรู้เป็นไปในแนวทางที่ถูกต้องและช่วยไม่ให้ผู้เรียนได้ข้อสรุปที่ไม่ถูกต้องหรือข้อสรุปจากข้อมูลที่จำกัดเกินไปหรือไม่คำนึงถึงข้อยกเว้น (Over-generalization)

แม้ว่าการเรียนตามแนวทาง DDL จะใช้ในบริบทของการเรียนภาษาอังกฤษเป็นภาษาต่างประเทศเป็นส่วนใหญ่ แต่ก็สามารถปรับใช้กับการสอนภาษาต่างประเทศอื่น ๆ เช่น ภาษาเยอรมัน (Vyatkina, N., 2016) ภาษาฝรั่งเศส (ดนยา ด้านสวัสดิ์, 2562) รวมถึงภาษาไทยด้วย บทความชิ้นนี้จะนำเสนอตัวอย่างแบบฝึกหัดที่ออกแบบตามแนวทางการเรียนรู้แบบ DDL ที่ใช้คลังข้อมูลภาษาไทยจากคลังข้อมูลภาษาไทยแห่งชาติเป็นข้อมูลในการส่งเสริมให้ผู้เรียนภาษาไทยเป็นภาษาต่างประเทศศึกษาวิเคราะห์การใช้คำโดยเน้นที่การศึกษาคำปรากฏร่วม รูปแบบการใช้คำ และการวิเคราะห์ความหมายแฝงโดยเฉพาะอรรถสัมผัสของคำ

2. ตัวอย่างแบบฝึกหัดพร้อมคำอธิบายและแนวทางเสริมการวิเคราะห์และอภิปรายผล

2.1 คำปรากฏร่วม (Collocation)



กิจกรรมที่ 1 “โอกาส” และ “ความหวัง”

คำสั่ง: ค้นตัวอย่างการใช้คำพร้อมบริบท (Concordance lines) ของคำว่า “โอกาส” และ “ความหวัง” จากคลังข้อมูลภาษาไทย <http://www.arts.chula.ac.th/~ling/tnc3/> และระบุคำกริยาที่ใช้ร่วมกับคำสองคำนี้พร้อมยกตัวอย่างประกอบจากคลังข้อมูลภาษา

คำกริยา + โอกาส	ตัวอย่าง
(ตัวอย่าง) เปิด + โอกาส	เพื่อ <u>เปิด</u> โอกาสให้นักเรียนสมัครเป็นสมาชิกในชมรมใดชมรมหนึ่ง
(ตัวอย่าง) หมด + โอกาส	(ตัวอย่าง) เข้ามาปิดกั้นตัวเองให้ <u>หมด</u> โอกาสได้สัมผัส ความสด แห่งโภชนาอาหาร

คำกริยา + ความหวัง	ตัวอย่าง
(ตัวอย่าง) มี + ความหวัง	ได้ทำให้ผู้คนมีความสุขกับปัจจุบัน และมี <u>ความหวัง</u> ในอนาคต
(ตัวอย่างเพิ่มเติมจากผู้เรียน)	(ตัวอย่างเพิ่มเติมจากผู้เรียน)

คำอธิบายกิจกรรมและแนวทางเสริมการวิเคราะห์และการอภิปรายผล

ตัวอย่างกิจกรรมที่ 1 มีวัตถุประสงค์เพื่อให้ผู้เรียนตระหนักถึงคุณลักษณะที่สำคัญประการหนึ่งของการศึกษาคำศัพท์ คือ ประเด็นเรื่องคำปรากฏร่วมของคำนามที่มีความหมายคล้ายคลึงกันในภาษาไทย ในกรณีนี้ผู้เรียนจะค้นหาตัวอย่างของคำว่า “โอกาส” และคำว่า “ความหวัง” ผู้สอนอาจจะเริ่มต้นให้ผู้เรียนยกตัวอย่างประโยคของคำนามดังกล่าวโดยเน้นที่คำกริยาที่มักใช้ร่วมกับคำนาม ในบางกรณีผู้สอนอาจจะตั้งคำถามนำเพื่อช่วยให้ผู้เรียนสังเกตความเหมือนหรือต่างของคำปรากฏร่วม เช่น เราพูดว่าเรามีโอกาสได้ แต่เราบอกว่าเรามีความหวังได้หรือไม่ เราพูดว่าเราเปิดโอกาสให้คุณได้ แต่เราพูดว่าเราเปิดความหวังให้คุณได้หรือไม่ อย่างไรก็ตามผู้สอนไม่ได้คาดหวังว่า



จะได้คำตอบที่ถูกต้องชัดเจน เนื่องจากหลักการสำคัญของวิธีการเรียนการสอนแบบ DDL คือการให้ผู้เรียนศึกษา วิเคราะห์ และหาคำตอบจากคลังข้อมูลภาษาด้วยตัวเองก่อน ในกิจกรรมนี้เมื่อผู้เรียนอ่านตัวอย่างประโยคหรือบรรทัดคอนคอร์แดนซ์ของคำว่า “โอกาส” จะพบว่าสามารถใช้กริยาได้หลากหลาย เช่น “มีโอกาส” “ต้องการโอกาส” “เสียโอกาส” “คือโอกาส” “เพิ่มโอกาส” “สร้างโอกาส” “ใช้ในโอกาส” หรือใช้กับคำกริยาที่มีลักษณะเป็นสำนวน เช่น “ฉวยโอกาส” ในขณะที่ข้อมูลตัวอย่างของคำว่า “ความหวัง” ชี้ให้เห็นว่าสามารถใช้กับคำกริยาเดียวกันกับคำว่า “โอกาส” เช่น “มีความหวัง” หรือ “สร้างความหวัง” แต่ในขณะเดียวกันมีคำกริยาบางตัวที่ใช้ร่วมกับคำว่า “ความหวัง” อย่างเป็นธรรมชาติในภาษาไทยแต่ไม่นิยมใช้กับคำว่า “โอกาส” เช่น “ตั้งความหวัง” “ฝากความหวัง” หรือในทางกลับกันมีคำกริยาที่มักพบบ่อยกับคำว่า “โอกาส” แต่ไม่ค่อยปรากฏร่วมกับคำว่า “ความหวัง” เช่น “เปิดโอกาส” “คว้าโอกาส” หรือ “ช่วงชิงโอกาส” เป็นต้น

ตามที่ได้อภิปรายไว้ในหัวข้อที่ผ่านมาว่าคุณลักษณะประการหนึ่งของการเรียนแบบ DDL คือความยืดหยุ่นในการตั้งคำถามและการค้นหาคำตอบ ดังนั้นผู้สอนอาจจะตั้งคำถามเพิ่มเติมเพื่อช่วยให้ผู้เรียนค้นหาหรือสังเกตลักษณะทางภาษาเพิ่มเติมจากที่กำหนดไว้ในวัตถุประสงค์หลักของกิจกรรมได้ เช่น นอกเหนือจากคำกริยาที่อยู่หน้าคำนามที่ทำหน้าที่เป็นกรรมของคำกริยาแล้ว มีกริยาอะไรที่อยู่หลังคำนามหรือที่ทำหน้าที่เป็นกริยาของประธานบ้าง เช่น “โอกาสลดลง” ซึ่งมักใช้ในโครงสร้างวลีที่มีความซับซ้อน เช่น “ทำให้โอกาสในการตั้งครุภลดลง” นอกจากนั้นยังอาจให้ผู้เรียนสังเกตคำกริยาที่หลากหลายที่ใช้กับกริยาลี “มีโอกาสต้อนรับ” “มีโอกาสช่วย” “มีโอกาสเจอ” “มีโอกาสเกิดต่อลม” ถึงแม้เป้าหมายหลักของการวิเคราะห์ในกิจกรรมนี้คือเรื่องคำปรากฏรวม แต่ในช่วงท้ายของกิจกรรมผู้สอนสามารถตั้งคำถามเพื่อเชื่อมโยงกับเรื่องความหมายได้ เช่น ให้ผู้เรียนจัดกลุ่มคำกริยาที่พบตามความหมาย เช่น คำกริยาที่สื่อถึงความหมายบวก ความหมายลบ หรือความหมายที่เป็นกลาง ๆ ไม่บวกหรือลบอย่างชัดเจน

2.2 การศึกษารูปแบบการใช้ภาษา (Language Pattern)

กิจกรรมที่ 2 รูปแบบการใช้คำว่า “โอกาส”

คำสั่ง: ระบุรูปแบบการใช้ภาษาของคำว่า “โอกาส” กล่าวคือ มีลักษณะการสร้างประโยคอย่างไรได้บ้างเพื่อจะอธิบายขยายความว่ามีโอกาสเรื่องอะไรหรือมีโอกาสที่จะทำอะไรบ้าง ยกตัวอย่างประกอบ



รูปแบบหรือลักษณะการสร้างประโยค	ตัวอย่าง
(ตัวอย่าง) มีโอกาส + คำกริยา	ขอแนะนำให้ท่านที่มีโอกาสไปทดลองและขอ เรียกร้องให้ประเทศไทยมีตัวเลือกแก่ผู้บริโภคม ากขึ้น
(ตัวอย่าง) “มีโอกาส” + “ที่จะ” + คำกริยา	การซื้อรถยนต์เก่าจากศูนย์ให้เข้ารถยนต์มี <u>โอกาสที่จะได้</u> รถยนต์ที่มีคุณภาพดี
(ตัวอย่างเพิ่มเติมจากผู้เรียน)	(ตัวอย่างเพิ่มเติมจากผู้เรียน)

คำอธิบายกิจกรรมและแนวทางเสริมการวิเคราะห์และการอภิปรายผล

กิจกรรมที่สองมีวัตถุประสงค์เพื่อให้ผู้เรียนสังเกตรูปแบบภาษาและการใช้คำ (Language Pattern) ในการเรียนภาษาต่างประเทศนอกจากผู้เรียนจะรู้ความหมายของคำศัพท์แล้วยังควรใช้คำศัพท์ให้ถูกต้องและสามารถแต่งประโยคให้เหมาะสมกับบริบทอีกด้วย โดยกิจกรรมนี้มีคำศัพท์เป้าหมายคือคำว่า “โอกาส” ซึ่งผู้เรียนได้ศึกษาตัวอย่างของคำนี้ในประเด็นเรื่องคำปรากฏมาแล้ว ในแบบฝึกหัดที่ 1 ในกิจกรรมนี้จะเน้นให้ผู้เรียนสังเกตการใช้คำว่า “โอกาส” ในรูปแบบต่าง ๆ ตัวอย่างรูปแบบที่ให้ไว้เป็นแนวทางคือ การใช้กริยาวลี “มีโอกาส” ร่วมกับคำกริยา และการใช้กริยาวลีเดียวกันตามด้วยคำกริยา “ที่จะ” และตามด้วยกริยาอีกตัวหนึ่ง เมื่อผู้เรียนได้ศึกษาตัวอย่างรูปแบบที่ให้ไว้แล้วสามารถค้นหาตัวอย่างเพิ่มเติมเพื่อสังเกตและระบุรูปแบบการใช้อื่น ๆ ได้ ผู้สอนอาจจะแนะนำให้ผู้เรียนเลือกรูปแบบการนำเสนอตัวอย่างแบบ ‘concord’ ซึ่งจะมีลักษณะเหมือนกับบรรทัดคอนคอร์ดแดนซีในภาษาอังกฤษคือ นำเสนอคำที่ต้องการค้นหาไว้ตรงกลางหน้าจอเพื่อให้สามารถสังเกตบริบทด้านซ้ายและขวาของคำที่ศึกษาได้ชัดเจนขึ้นดังที่แสดงในภาพข้างล่างนี้

[illegible]

ภาพที่ 6 บรรทัดคอนคอร์แดนซ์ของคำว่า “โอกาส” ในคลังข้อมูลภาษาไทยแห่งชาติ

การนำเสนอตัวอย่างในรูปแบบดังกล่าวจะช่วยให้ผู้เรียนสามารถสังเกตรูปแบบการใช้คำที่หลากหลายมากขึ้น เช่น กริยา + โอกาส + ให้(แก่) + อนุประโยค เช่น จัดโอกาสให้แก่นักเรียนในชุมชนทุกเพศทุกวัยได้เรียนรู้อย่างต่อเนื่องตลอดชีวิต นอกจากนั้นผู้เรียนยังอาจจะสังเกตรูปแบบการใช้ในลักษณะที่เป็นวลีหรือกลุ่มคำ (Lexical Chunking) ที่มีลักษณะค่อนข้างตายตัวหรือมีความเป็นสำนวน เช่น “ถือว่าเป็นโอกาสที่ดีของ” “ทันทีที่มีโอกาส” “ท่ามกลางวิกฤติมีโอกาส” “เสียโอกาสโดยใช่ที่” และที่สำคัญถ้าผู้เรียนได้รับการฝึกฝนให้อ่านตัวอย่างคำในบริบทที่กว้างขึ้นจะช่วยให้เห็นรูปแบบหรือโครงสร้างประโยคที่มีความซับซ้อนมากยิ่งขึ้น เช่น ความสัมพันธ์ระหว่างคำหรือวลีกับรูปแบบภาษาด้วย เช่น “เปิดโอกาส” มักจะใช้กับกริยาวิเศษณ์ “ให้มีการ...” หรือ “ไม่มีโอกาส” มักตามด้วยกริยาวิเศษณ์ที่ขึ้นต้นด้วย “ได้” หรือ “ที่จะ” หรือการใช้วลีในประโยคที่มีความซับซ้อน เช่น “และเมื่อมีโอกาส.....ก็.....” เป็นต้น

2.3 ลักษณะทางภาษาที่เป็นเป้าหมายการเรียนรู้: ความหมายแฝง (Connotation) และอรรถ
สัมผัสของคำ (Semantic Prosody)

คำสั่ง: ค้นตัวอย่างตัวอย่างของคำว่า “ประสบ” และ “เผชิญ” จากคลังข้อมูลภาษาไทยแห่งชาติ จากนั้นวิเคราะห์ตัวอย่างและสรุปความแตกต่างระหว่างคำสองคำดังกล่าวโดยเฉพาะในด้านความหมายแฝง อาจใช้ประเด็นวิเคราะห์ที่แสดงไว้ในตารางเป็นแนวทางการนำเสนอ



ประเด็นการวิเคราะห์	“ประสบ”	“เผชิญ”
ประเภทของคำ	เป็นคำกริยา	เป็นคำกริยา
รูปแบบของกริยาวิเศษณ์ที่พบบ่อย	“ประสบ” + คำนาม	“เผชิญ” + คำนาม หรือ เผชิญ + กับ + คำนาม
คำนามที่มักทำหน้าที่เป็นกรรมในกริยาวิเศษณ์ (ความหมายบวก)		
คำนามที่มักทำหน้าที่เป็นกรรมในกริยาวิเศษณ์ (ความหมายลบ)		
คำนามที่มักทำหน้าที่เป็นกรรมในกริยาวิเศษณ์ (ความหมายเป็นกลาง)		
ความแตกต่างหรือข้อสังเกตอื่นๆ		

คำอธิบายกิจกรรมและแนวทางเสริมการวิเคราะห์และการอภิปรายผล

กิจกรรมสุดท้ายที่นำเสนอเป็นตัวอย่างมุ่งเน้นการวิเคราะห์ความหมายแฝงของคำ โดยเฉพาะความหมายแฝงที่มีลักษณะเฉพาะที่เรียกว่า อรรถสัมผัสของคำ (Semantic Prosody) โดยคำที่เป็นเป้าหมายในการวิเคราะห์คือคำว่า “ประสบ” และ “เผชิญ” ซึ่งเป็นคำที่มีความหมายคล้ายคลึงกันคือกล่าวถึงเป็นสิ่งที่ผู้พูดหรือประธานของประโยคได้กระทำหรือมีประสบการณ์กับสิ่งนั้น โดยมีคำเทียบเคียงที่มีผู้ศึกษาเรื่องอรรถสัมผัสของคำในภาษาอังกฤษ เช่น ‘face’ หรือ ‘encounter’ เป็นที่มาของการศึกษาคำว่า “ประสบ” กับ “เผชิญ” ในแบบฝึกหัดนี้ อย่างไรก็ตาม คำดังกล่าวไม่ใช่คำเหมือน (Synonym) ที่สามารถใช้ทดแทนกันได้ในทุกบริบทเนื่องจากมีอรรถสัมผัสต่างกัน กิจกรรมนี้เริ่มต้นด้วยการชี้ให้เห็นถึงความแตกต่างของรูปแบบการใช้คำ กล่าวคือ “ประสบ” มักตามด้วยคำนามได้เลยโดยทันที ในขณะที่ “เผชิญ” มักจะตามด้วยคำบุพบท “กับ” และตามด้วยคำนามหรือนามวลี จากนั้นผู้เรียนจะต้องหาตัวอย่างของคำนามหรือนามวลีที่มีความหมายบวก ลบ หรือเป็นกลางไม่บวกหรือลบ ถ้าผู้เรียนวิเคราะห์ข้อมูลแล้วจะพบว่าคำว่า “ประสบ” มีอรรถสัมผัสทั้งในเชิงบวกและลบ เช่น “ประสบความสำเร็จ” “ประสบความล้มเหลว” ในขณะที่คำว่า “เผชิญ” มีอรรถสัมผัสในเชิงลบมากกว่า เช่น “เผชิญกับปัญหา” “เผชิญกับวิกฤติ”



“เผชิญกับอุปสรรค” ถึงแม้จะมีคำปรากฏร่วมที่มีความหมายเป็นกลางอยู่บ้างแต่ก็เป็นคำที่มีนัยยะแฝงที่บ่งบอกถึงความท้าทายหรือความเปลี่ยนแปลง การที่ผู้เรียนภาษาไทยเป็นภาษาต่างประเทศสามารถสังเกตและใช้คำที่มีความหมายแฝงได้อย่างถูกต้องจะช่วยให้การพูดหรือการเขียนเป็นธรรมชาติและสื่อสารความหมายได้ตรงกับความต้องการสื่อสารมากขึ้น

กิจกรรมการสังเกตความหมายแฝงหรืออรรถสัทสามารถนำไปต่อยอดหรือใช้ในการศึกษาภาษาเชิงเปรียบเทียบ (Contrastive Analysis) หรือประยุกต์ใช้ในการสอนการแปลได้ด้วย ตัวอย่างเช่น ผู้เรียนอาจจะค้นหาตัวอย่างของคำในภาษาอังกฤษที่มีความหมายใกล้เคียงหรือมักจะเป็นคำที่เข้าใจว่าเป็นคำเทียบเคียง (Equivalence) ของคำว่า “ประสบ” หรือ “เผชิญ” เช่นคำว่า ‘to face’ ‘to encounter’ ‘to experience’ ‘to witness’ เป็นต้นจากคลังข้อมูลภาษาอังกฤษจากนั้นให้วิเคราะห์คำปรากฏร่วมโดยเฉพาะคำนามที่ทำหน้าที่เป็นกรรมของกริยากลุ่มนี้ โดยดูว่าคำนามที่พบบ่อยมีความหมายแฝงในเชิงบวกหรือลบ จากการวิเคราะห์บรรทัดคอนคอร์ดแดนซ์ของคำภาษาอังกฤษกลุ่มนี้พบว่าคำว่า ‘to face’ และ ‘to encounter’ มีอรรถสัทสในเชิงลบมากกว่าคำว่า ‘to experience’ หรือ ‘to witness’ ดังนั้นการแปลคำว่า “ประสบ” หรือ “เผชิญ” เป็นภาษาอังกฤษจึงใกล้เคียงกับคำว่า ‘to face’ หรือ ‘to encounter’ มากกว่าถึงแม้ว่าจะมีกริยาบางตัวที่เป็นธรรมชาติมากกว่าในการแปลกลุ่มคำบางกลุ่มคำ เช่น ประสบความสำเร็จ อาจจะแปลเป็น ‘to achieve success’ มากกว่าที่จะแปลเป็นคำสองคำดังกล่าว

3. บทสรุป

บทความชิ้นนี้ชี้ให้เห็นว่าคลังข้อมูลภาษา รวมถึงการนำเสนอตัวอย่างของคำหรือวลีจากคลังข้อมูลภาษาในรูปของตัวอย่างคำพร้อมบริบทหรือบรรทัดคอนคอร์ดแดนซ์ในบริบทการเรียนรู้ที่เน้นการเรียนรู้จากข้อมูล (หรือการเรียนรู้แบบ DDL) ที่มีพัฒนาการและการนำไปใช้อย่างแพร่หลายในการสอนภาษาอังกฤษนั้นสามารถนำมาประยุกต์ใช้กับการเรียนการสอนภาษาไทยเป็นภาษาต่างประเทศได้ในระดับหนึ่งถึงแม้ว่าการสร้างและการประมวลผลภาษาไทยในฐานข้อมูลภาษาจะทำได้ไม่สมบูรณ์ครอบคลุมทุกเครื่องมือการวิเคราะห์เช่นเดียวกับภาษาอังกฤษเนื่องจากลักษณะของตัวอักษรและการแบ่งคำภาษาไทยทำได้ไม่ชัดเจนและได้ผลถูกต้องเท่าภาษาอังกฤษ อย่างไรก็ตามคลังข้อมูลภาษาไทยสามารถใช้เป็นข้อมูลที่ช่วยให้ผู้เรียนภาษาไทยเป็นภาษาต่างประเทศ



สามารถสังเกตและเรียนรู้ลักษณะที่สำคัญของคำ เช่น คำปรากฏร่วม รูปแบบการใช้คำ และความหมายแฝงหรืออรรถสัมผัสของคำได้ด้วยตัวเองโดยมีผู้สอนเป็นผู้ให้แนวทางการวิเคราะห์ข้อมูล ผ่านทางการออกแบบกิจกรรมหรือการตั้งคำถามในกระบวนการวิเคราะห์ข้อมูล

ตัวอย่างของกิจกรรมที่นำเสนอในบทความชิ้นนี้เป็นเพียงสื่อหรือแบบฝึกหัดที่จัดเป็นส่วนหนึ่งของกระบวนการเรียนการสอนแบบ DDL เท่านั้น การนำแนวทางการเรียนรู้แบบ DDL ไปใช้ให้ได้ผลในการเรียนภาษาไทยหรือภาษาต่างประเทศอื่น ๆ ขึ้นอยู่กับปัจจัยอื่น ๆ ด้วย เริ่มจากการเตรียมความพร้อมของผู้เรียนเพื่อให้เข้าใจถึงธรรมชาติและลักษณะของคลังข้อมูลภาษาและการทำงานของโปรแกรมวิเคราะห์ข้อมูล เข้าใจลักษณะเฉพาะของตัวอย่างของคำที่นำเสนอในรูปแบบของบรรทัดคอนคอร์แดนซ์ และต้องได้รับการฝึกฝนในการวิเคราะห์ ตีความ และสรุปประเด็นจากข้อมูลที่ศึกษา (ดูรายละเอียดการเตรียมความพร้อมผู้เรียนได้จาก Sripicharn (2010) ในส่วนของผู้สอนต้องมีความรู้ทางเทคนิคด้านคลังข้อมูลภาษาด้วยเช่นกันเพื่อจะได้สามารถออกแบบกิจกรรมที่ใช้คลังข้อมูลภาษาเป็นข้อมูลหลักได้อย่างมีประสิทธิภาพและเหมาะสมกับความสามารถและความแตกต่างของผู้เรียน เช่น อาจจะเริ่มต้นด้วยกิจกรรมที่มีคำสั่งหรือคำแนะนำการค้นหาหรือวิเคราะห์ข้อมูลที่ชัดเจนและค่อยๆพัฒนาไปสู่กิจกรรมที่เน้นผู้เรียนเป็นศูนย์กลางอย่างเต็มรูปแบบ เช่น การให้ผู้เรียนคิดประเด็นการเรียนรู้ ค้นหาตัวอย่างคำในคลังข้อมูล และสรุปประเด็นการเรียนรู้ด้วยตนเองโดยให้ข้อมูลเป็นตัวขับเคลื่อนกระบวนการเรียนรู้มากกว่าประเด็นที่กำหนดโดยผู้สอนเพียงอย่างเดียว

ท้ายสุดนับบทความชิ้นนี้สามารถนำไปพัฒนาเป็นงานวิจัยที่เกี่ยวกับการเรียนการสอนภาษาไทยเป็นภาษาต่างประเทศได้ ทั้งการวิจัยในเชิงปริมาณ เช่น การศึกษาประสิทธิผลของการใช้คลังข้อมูลภาษาหรือแนวทางการเรียนแบบ DDL ในการเรียนรู้คำศัพท์หรือไวยากรณ์ภาษาไทย โดยแบ่งเป็นกลุ่มทดลองกับกลุ่มควบคุมและทดสอบวัดระดับความรู้ด้านไวยากรณ์หรือคำศัพท์ก่อนและหลังการสอน หรือวัดคุณภาพของงานเขียน เพื่อวัดว่ากลุ่มทดลองที่ใช้แนวทางดังกล่าว จะมีระดับความสามารถทางภาษาหรือการเขียนเพิ่มขึ้นอย่างมีนัยยะสำคัญหรือไม่เมื่อเทียบกับกลุ่มควบคุม นอกจากนี้ยังสามารถศึกษาวิจัยในเชิงคุณภาพหรือแบบผสมผสานได้ เช่น นอกจากการวัดระดับความสามารถทางภาษาแล้ว ยังสามารถศึกษากระบวนการและพฤติกรรมการค้นหาและวิเคราะห์ข้อมูลจากคลังข้อมูลภาษาของผู้เรียน ศึกษาทักษะหรือกลยุทธ์ที่ใช้ในการเรียนแบบ DDL และ



สำรวจความคิดเห็นหรือข้อเสนอแนะของผู้เรียนที่มีต่อแนวทางการเรียนภาษาซึ่งสอดคล้องกับ
แนวโน้มหรือทิศทางของการศึกษาศสมัยใหม่ที่เน้นการใช้เทคโนโลยีเพื่อส่งเสริมการเรียนรู้ด้วยตัวเอง
ของผู้เรียน

เอกสารอ้างอิง

- เกียรติชัย เดชพิทักษ์ศิริกุล (2561). หลักการใช้คำว่า “เป็น” และ “คือ” สำหรับนักศึกษาชาวจีน.
Veridian E-Journal, Silpakorn University ฉบับภาษาไทย สาขามนุษยศาสตร์
สังคมศาสตร์ และศิลปะ. 11(2). 381-394.
- ชาตบุษย์ ลีละรัตนานนท์ (2551). คำปรากฏร่วมจำเพาะและอรรถสัมผัสของคำกริยา “เกิด” ใน
ภาษาไทย: การศึกษาเชิงภาษาศาสตร์คลังข้อมูล. *Language and Linguistics*. 27(1), 61-74.
- ดนยา ด่านสวัสดิ์ (2562). ความสามารถในการเขียนโปรแกรมการท่องเที่ยวภาษาฝรั่งเศสของ
นักศึกษาสาขาวิชาการจัดการอุตสาหกรรมการท่องเที่ยวและบริการโดยใช้คลังข้อมูล.
วารสารสมาคมครูภาษาฝรั่งเศสแห่งประเทศไทย ในพระราชูปถัมภ์ฯ. 42(1), 21-39.
- ธารทอง แจ่มไพบูลย์ และ วิโรจน์ อรุณมานะกุล (2018). วงศัพท์ในแบบเรียนภาษาไทย: การศึกษา
ตามแนวทางภาษาศาสตร์คลังข้อมูล. *NIDA Development Journal*. 58(3), 23-51.
- วิโรจน์ อรุณมานะกุล (2553). *ภาษาศาสตร์คลังข้อมูล: หลักการและการใช้*. (พิมพ์ครั้งที่ 2):
กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- สรณ์ แอบเงิน และ กิ่งกาญจน์ เทพกาญจนา (2560) . ความหมายของคำว่า ของ ในฐานะคำบุพบท
ในภาษาไทย. *วารสารอักษรศาสตร์*. 46(2), 247-294.
- Anthony, L. (2021). Antconc (Version 4.0.2) [Computer Software]. Tokyo, Japan:
Waseda University. Available from <https://www.laurenceanthony.net/software/antconc/>.
- Baker, P., Hardie, A. & McEnery, T. (2006) *A Glossary of Corpus Linguistics*. Edinburgh
University Press.
- Benson, M., Benson, E. & Ilson, R. (1997). *The BBI Dictionary of English Word*
Combinations. John Benjamins.
- Friginal, E. (2018). *Corpus Linguistics for English Teachers: New Tools, Online*
Resources, and Classroom Activities. Routledge.



- Hanks, P. & Ma, W. (2021) Meaning and Grammar in the light of Corpus Pattern Analysis. *International Journal of Lexicography*, 34(1), 135–149.
- Hoey, M. (2012). Lexical priming. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. John Wiley & Sons Inc.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins.
- Johns, T. (1991) Should you be persuaded: Two examples of Data-driven Learning. in Johns, T. and P. King (Eds) *Classroom Concordancing*, *ELR Journal* 4. Birmingham: Centre for English Language Studies, University of Birmingham, pp. 1-16.
- Jones, C., & Waller, D. (2015). *Corpus Linguistics for Grammar: A guide for research*. Routledge.
- Kennedy, C. & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation, *Language Learning and Technology*, 5(3), 77-90.
- Partington, A. (1998). *Pattern and Meaning: Using Corpora for English Language Research and Teaching*. John Benjamins.
- Partington, A., Dugiud, A., & Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. John Benjamins.
- Pereira, L., Manguilimotan, E., & Matsumoto, Y. (2016). Leveraging a Large Learner Corpus for Automatic Suggestion of Collocations for Learners of Japanese as a Second Language'. *Calico journal*. 33(3), 311–333.
- Sripicharn, P. (2010). 'How to prepare learners to use corpora'. In M. McCarthy & A. O'Keeffe. *The Routledge Handbook of Corpus Linguistics*. Routledge.
- Stewart, D. (2010). *Semantic Prosody: A Critical Evaluation*. Routledge.
- Timmis, I. (2015). *Corpus Linguistics for ELT: Research and Practice*. Routledge.
- Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*. 28(2), 207–226.
- Wu, Y.-J.A. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning and Technology*. 25(2), pp. 192-214.