

## 從教科書建置國中學術詞彙表

李岳霖 鍾涵瀾 吳昭容

學術詞彙是學術場域中進行交流的重要工具，是用以解釋專業術語涵義之輔助性詞彙。相同的學術詞彙可能出現在各領域，但出現頻率會因領域而異。本研究使用國民中小學九年一貫課程綱要之國中階段七科（數學、理化、生物、地科、地理、歷史、公民）教科書以編製學術詞彙表。對語料進行斷詞處理、對照華語八千詞表刪除基本詞彙、再以均勻度等相關指標篩選，整理為包含 779 個詞彙的「國中學術詞彙表」。儘管篩選過程考量跨領域特性，但各學術詞彙出現於數學科的單元比率與其他各科顯著不同，而依據出現單元比率進行集群分析所得五組學術詞彙，除單純反映出現單元比率高、低之組外，亦有歷史公民組、生物地科地理組、自然領域組。最後本文重新論述學術詞彙的特性、提出國內國中學術詞彙表建置程序的反思、對未來的研究與教學實務提出建議，並指出本研究的限制。

關鍵詞：均勻度、國中教科書、集群分析、語料庫、學術詞彙

收件：2022年3月2日；修改：2022年7月7日；接受：2022年7月19日

---

李岳霖，國立臺灣師範大學教育心理與輔導學系研究助理

鍾涵瀾，國立臺灣師範大學教育心理與輔導學系碩士

吳昭容，國立臺灣師範大學教育心理與輔導學系教授，E-mail: cjwu@ntnu.edu.tw

# Compiling a List of Middle School Academic Vocabulary From Textbooks

Yueh-Lin Li     Han-Rong Zhong     Chao-Jung Wu

Academic vocabulary is a crucial communication tool in academia. Academic vocabularies are supplementary vocabularies used to define technical terms. The same academic vocabulary might appear in various disciplines, but the frequency will vary. This study compiled an academic vocabulary list using textbooks from the Grade 1-9 Curriculum Guidelines for seven disciplines-mathematics, physics/chemistry, biology, geoscience, geography, history, and civics. After segmenting the corpus, removing the fundamental terms, and filtering by indicators such as dispersions, we compiled a “Middle School Academic Vocabulary List” with 779 vocabulary words. The proportion of academic vocabulary occurring in each unit of mathematics was significantly different from that of other disciplines, despite considering cross-discipline features. The cluster analysis was derived from the following five categories: historical and citizen, biology and geography, natural science, and two groups reflecting a higher and lower proportion of occurrence units. Finally, the study examined the characteristics of academic vocabulary and reflected the construction procedure for an academic vocabulary list for middle school students. The study also gives suggestions for future research and teaching, as well as possible limits to the research.

Keywords: dispersion, middle school textbook, cluster analysis, corpus, academic vocabulary

Received: March 2, 2022; Revised: July 7, 2022; Accepted: July 19, 2022

---

Yueh-Lin Li, Research Assistant, Department of Educational Psychology and Counseling, National Taiwan Normal University.  
Han-Rong Zhong, Master, Department of Educational Psychology and Counseling, National Taiwan Normal University.  
Chao-Jung Wu, Professor, Department of Educational Psychology and Counseling, National Taiwan Normal University, E-mail: [cjwu@ntnu.edu.tw](mailto:cjwu@ntnu.edu.tw)

## 壹、前言

「學術詞彙」(academic vocabulary)為學術語言(academic language)中較為明顯且關鍵的元素之一(Coxhead, 2000; Schleppegrell, 2004; Schuth et al., 2017; Townsend et al., 2012)。不只有學者主張,妥善運用學術詞彙可增進使用者於多個專業學科的學習歷程、知識習得、閱讀理解等認知活動的品質(Bailey & Heritage, 2008; Schleppegrell, 2004),而且實徵研究也支持,對學術詞彙的掌握程度確實顯著預測學生之學業成就表現(Schuth et al., 2017; Townsend et al., 2012)。建立學術詞彙的詞表,能發展詞彙學習目標、評量詞彙知識、分析文本難度、設計詞彙學習工具、發展學習活動,亦能引導學生獨立學習。雖然學術詞彙之意涵可能因學科不同而有些許差異,但學者認為學生根據詞表學習其核心語義,有助於在不同的學科達成更佳的理解(Coxhead, 2000; Nation, 2013; Nation & Webb, 2011)。基於學術詞彙對學術發展與學業成就的重要性(Masrai & Milton, 2018; Schuth et al., 2017; Townsend et al., 2012),學者建置了不同的學術詞彙表(Coxhead, 2000; Gardner & Davies, 2014; Greene & Coxhead, 2015)以輔助現場教師的教學品質以及教科書編製者的運用。

然而中文對於學術詞彙的相關研究極為稀少。劉貞好等人(2016)提出「中文學術常用詞表」,但編製目的在於補充針對外語使用者的華語文能力測驗(Test of Chinese as a Foreign Language, TOCFL)參考詞表「華語八千詞表」的不足。Snow與Uccelli(2009)主張學習學術語言不僅對於外語人士來說是很大的挑戰,對學習階段之母語人士亦如是,因此編製適用於中文母語且於學習階段之學生所使用之學術詞彙表應是重要的研究方向。在眾多學習階段中,Townsend等人(2012)以實徵研究指出中學階段之學術詞彙掌握程度可預測學生學業成就表現,而國內學生在國中結束時即面臨第一場影響生涯發展的重大學業評量——國中教育會考,建立國中學習階段適用之中文學術詞彙表有其重要性。

基於學術詞彙不同於日常用語與專業術語（Beck et al., 2013），建置「國中學術詞彙表」須斟酌語料庫的選擇與詞彙之排除與納入的指標。本研究以文獻回顧釐清學術詞彙的定義，彙整以往學術詞彙表建置時的指標，也指出中文詞彙須經斷詞軟體處理的特性，作為研究方法的基礎。此外，雖學術詞彙具備跨領域之特性，但 Hyland 與 Tse（2007）指出學術詞彙出現於各領域之情形並不相同，因此本研究亦將進一步描述國中學術詞彙於不同科目的分布狀況。

## 貳、文獻探討

### 一、學術詞彙的定義與特性

學術詞彙的意義可從 Beck 等人（2013）提出的三階層詞彙框架（the three tiers framework）來了解。Beck 等人指出教學現場存在三種詞彙，分別為階層一（tier one）之基本詞彙（basic words）、階層三（tier three）之專業術語（technical terms），以及兩者間的階層二（tier two），即學術詞彙。基本詞彙指在各種文本或日常生活中都會出現的高頻詞，人們會於非正式言談中透過潛移默化的方式逐漸學會語義，不需直接教導。專業術語指除了相關學科文本外，較少於其他場域出現的低頻詞彙，通常代表學術文本中各單元的核心觀念，往往會於第一次提及時給予定義和實例，並常出現於搭配的圖表中，需要教師專門介紹並特別給予教導，並在教學後評量學生的掌握程度。而介於兩者之間的即為學術詞彙。出現頻率是區分不同階層的標準之一，這使得階層與階層之間的分界是模糊的、武斷的，研究者選擇的頻率標準會決定一個詞彙的歸類。除出現頻率介於基本詞彙與專業術語之間外，學術詞彙亦存在幾項特性使其不同於另外兩者，分別為跨領域（across disciplines）、輔助性（supportive）、多義性（polysemous）、抽象性（abstraction）。

相較於專業術語，學術詞彙為跨領域詞彙，學生於學習過程中會於各科目接觸到同樣的詞彙，用以形容、描述、介紹專業術語。因存在此特性，學者於編製學術詞彙表時常以「該詞彙出現於多少個次語料庫」作為篩選標準，只有超過一定數量之詞方可通過篩選（Coxhead, 2000; Gardner & Davies, 2014; Greene & Coxhead, 2015）。而儘管學術詞彙為跨領域之詞彙，但在各科目出現的次數會隨不同科目而有所不同（Hyland & Tse, 2007）。

學術詞彙常被用來定義及介紹專業術語，因而產生輔助性的特性。例如，「分解」、「作用」、「演變」等學術詞彙會出現於介紹專業術語「化石」的前後文。學術詞彙的輔助角色使其較不顯眼（salient），且不若專業術語與篇章或教學主題有直接且密切的關聯，教師往往忽略學術詞彙的指導，或者假設學生已明瞭學術詞彙而未加說明（Hiebert & Lubliner, 2008; Ogle et al., 2015; Townsend et al., 2012）。因此學者提醒教學現場中應關注學生對學術詞彙的掌握程度，以免阻礙專門學科的知識學習（Coxhead, 2000; Greene & Coxhead, 2015; Nation, 2013; Townsend et al., 2012）。

因學術詞彙的跨領域與輔助性特性，延伸出另一特性——多義性，即學術詞彙之語義可能會隨學科的不同而有所不同（Baumann & Graves, 2010; Greene & Coxhead, 2015; Harmon & Wood, 2018; Hiebert & Lubliner, 2008; Ogle et al., 2015）。例如，「分析」（analyze）一詞，在自然科學領域時常指確認物質組成的方法，在社會領域出現時則常指將事情考慮清楚（Hyland & Tse, 2007）。此外亦可能因搭配的詞彙不同，所指涉的事物及詞彙的內涵就略有差異。例如，「基礎」（foundation）之核心語義為「讓其他事物建立於其上」（base on which something else can be built），但表面語義可能會因為搭配的詞而有些許不同，例如，建築物的「基礎」、抽象理論的「基礎」，或甚至是信任的「基礎」（Crosson et al., 2019）。學術詞彙之多義性特性，也反映學生應該在具有脈絡的情境下理解及學習學術詞彙（Hyland & Tse, 2007; Nagy & Townsend, 2012）。

學術詞彙亦具抽象性。抽象性通常被解釋為較不具知覺特性、難以形成心像，無法在語文編碼之外形成圖像的內在表徵（Kounios & Holcomb, 1994），或者不容易連結多元脈絡到記憶網絡中，故而較難從先備知識中提取到相關訊息（Crosson et al., 2019; Schwanenflugel et al., 1988）。例如，英文“see”除可指「視覺上看到」外，亦可指「理解」之義，而後者的抽象性高於前者。前後文提供的脈絡可以降低抽象性所造成的理解難度（Schwanenflugel et al., 1988），因此在有脈絡的情況下學習與運用學術詞彙，可以突破其抽象性所導致的障礙。

以下以國中自然與生活科技課本之課文為例：

生物死亡後，遺體大多會被分解或破壞，若遺骸被包埋在泥沙中，隨著時間推移，上方的泥沙層層堆疊，會逐漸形成堅硬的岩層，將遺骸保存在裡面。在此過程中，如果有礦物質滲入這些遺骸的骨骼或牙齒等堅硬部位，經過結晶及壓力等作用，這些遺骸就會漸漸形成化石。科學家可藉由化石推估這些生物生前的樣貌、生存的方式和演變的情形。（陳世煌等人，2014，頁 66-67）

專業術語「化石」為此段課文的重點。然而要確實理解化石的意涵，必須同時也掌握如「分解」、「壓力」、「作用」、「演變」等詞彙的語義。這些輔助詞彙在日常生活上的出現頻率偏低，具有多義與抽象性，且老師們未必會花時間說明其義，例如，「演變」是指演變成什麼呢？具體語義為何？在生物科中和歷史科中的語義是一樣的嗎？由於學生可能不清楚這些輔助性詞彙的語義，進而阻礙學生理解專業術語（Hiebert & Lubliner, 2008）。

綜合文獻回顧，學術詞彙存在「跨領域」、「輔助性」、「多義性」、「抽象性」等特性，並廣泛出現於學術場域中之不同學科領域，雖核心語義不變，但表面語義可能因學科不同而不同，且為介紹專業術語概念時使用的抽象詞彙。因此，學者認為學習者對學術詞彙的掌握程度與閱讀理解和學業成就表現有密切的關聯（Harmon & Wood, 2018; Lawrence

et al., 2019; Masrai & Milton, 2018; Mokhtari & Velten, 2015; Schuth et al., 2017; Townsend et al., 2012)。

## 二、學術詞彙表之編製

學術詞彙表的編製涉及編製目的、語料庫範圍、篩選詞彙的標準。Coxhead (2000) 編製的「學術詞彙表」(academic word list, AWL) 以大學階段的教科書文本為語料庫，總字數約 350 萬字，涵蓋藝術、商業、法律、科學此四個學門，其下再各自細分為七個科目。為滿足學術詞彙之定義，Coxhead (2000) 採用三個選詞標準：詞頻 (frequency)、專業性 (specialized occurrence)、領域廣度 (range)。詞頻標準為詞彙必須在全語料庫中至少出現 100 次，以避免選用過於冷僻之詞彙；專業性標準為透過排除 West (1953) 所編纂的「一般使用詞彙表」(general service list, GSL) 中前 2,000 個常用詞彙，以確保詞彙非基本詞彙；領域廣度標準為詞彙必須在四個學門中皆至少出現 10 次，且在全部科目中至少有 15 個科目出現該詞，以確保詞彙具備跨領域之特性，而非學科專業術語。最後從語料庫中擷取以共同字根為單位，包含字首字尾變化衍伸之 570 個詞族 (word family)。

Gardner 與 Davies (2014) 使用新進且更龐大的「美國當代英語語料庫」(Corpus of Contemporary American, COCA) 其下之學術語料庫為基礎語料庫，以編製新學術詞彙表 (academic vocabulary list, AVL)。學術語料庫總詞數約為 1 億 2,000 萬詞，並涵蓋教育、人文、歷史、社會科學、哲學宗教心理、法政、理工、醫藥衛生、商管等九個學門領域。選詞標準為比率 (ratio)、領域廣度、均勻度 (dispersion)、領域指數 (discipline measure)。比率之標準為該詞彙出現於學術語料庫之頻率必須為出現於 COCA 之下的其他非學術語料庫 (如：小說、流行雜誌等) 之頻率的 1.5 倍以上，以確保所選詞彙非基本詞彙；領域廣度標準為詞彙應出現於至少七個學門領域之中，以滿足跨領域之特性；均勻度之標準為，詞彙於各學門領域之分布均勻度應為 .80 以上，愈趨近 1 表示其

在各學門領域出現之頻率一致、愈趨近 0 則表示其僅出現於特定學門領域且頻率差異甚大 (Juilland & Chang-Rodriguez, 1964)，詳細計算公式請參考附錄；領域指數之標準為避免詞彙出現於某一學門領域之頻率過高，以確保所選詞彙非該科目之專業術語。與前述 AWL 僅以共同字根計算之詞族不同，AVL 將字根之語義及詞性納入考量，共收錄 3,015 個詞群 (lemma)，如 proceedings (名詞的訴訟程序)、procedure (名詞的程序)、procedural (形容詞的程序)，三者為同一詞族、但各自獨立為三個詞群。

劉貞好等人 (2016) 為華語學習者編製之中文學術常用詞表。其蒐集人文社會科學中文電子期刊之 1,000 篇文章為語料庫，共約 900 萬字，涵蓋教育學、管理學、文學、歷史學、政治學、法律學、社會學、區域研究與地理學、語言學、經濟學等 10 個學術領域。選詞標準為詞頻、專業性、領域廣度。詞頻標準為參考 Hyland (2008) 之高頻詞標準，詞彙須出現九次以上，以確保所選詞彙為學術領域之高頻詞彙；專業性之標準與 AWL 的做法類似，乃參考「華語八千詞表」，刪除入門級與基礎級的詞彙，保留進階級、高階級、流利級的詞彙，以確保所選詞彙非基本詞彙；領域廣度之標準為詞彙須至少出現於 8 個學術領域，以確保所選詞彙具備跨領域之特性。最後共收錄 2,405 個詞彙。

Greene 與 Coxhead (2015) 則針對中學學習階段，編製中學學術詞彙表 (Middle School Vocabulary List, MSVL)。其以美國公立中學各科教科書作為語料庫，共約 1,800 萬字，涵蓋英語文法寫作、健康、數學、科學、社會與歷史五個學科。詞表目的在於為五個學科個別編製內容領域 (content area) 詞表，而五項詞表內容同中有異，以回應 Hyland 與 Tse (2007) 應考量領域不同之主張。Greene 與 Coxhead 排除「閱讀與文學」乃因這一科目內容多為小說體裁，不適合納入學術語料庫中。詞表編製流程與 AWL 相同，選詞標準仍為詞頻、專業性、領域廣度，按語料庫大小之比例採用與 AWL 同樣的詞頻標準及領域廣度標準，亦同樣排除 GSL 中前 2,000 的常用詞彙外。最終五張詞表分別含有 300~400

多不等的詞族，每張均兼含跨領域性質的核心學術詞彙，以及具特定學科性質的專業術語。

綜上所述，編製學術詞彙表之標準不外乎為詞頻、專業性及領域廣度。為確保所選詞彙為常出現於學術語料庫之詞彙，因而設定標準篩選（劉貞好等人，2016；Coxhead, 2000; Greene & Coxhead, 2015）。為確保所選詞彙之專業性，因而排除一般使用詞表之基礎詞彙（劉貞好等人，2016；Coxhead, 2000; Greene & Coxhead, 2015），或者如 Gardner 與 Davies（2014）設定比率標準，確保詞彙於學術語料庫中的出現頻率為非學術語料庫的倍數以上，以同時滿足常出現於學術語料庫及專業性兩個標準。最後為確保所選詞彙之跨領域特性或非專業術語，因而排除未滿足領域廣度標準之詞彙。亦可使用均勻度指標，計算詞彙於各科目之均勻度並設定標準，以篩選雖在多數領域均有出現之詞彙但分布不均、難以認定滿足跨領域特性之詞彙（Gardner & Davies, 2014）。

由於書寫特性不同，編製中文詞表必須考量斷詞（segmentation）。中文以「字」（characters）作為文字的基本單位，但「詞」（word）才是表達語義的最小獨立單位（黃居仁，2005）。而一個詞內所包含的文字數量並不固定，可能為雙字詞（如：「電腦」）、可能為三字詞（如：「惡作劇」）、甚至可能為單字詞（如：「花」）。由於中文並未存在如拼音文字的空格以區隔不同的詞，因而不易判別文句中之字是否為詞。例如，「花」，其可以為一個單字詞，但亦可能為「花費」、「花俏」的首字。因此製作中文詞表與拼音文字最大的不同處在於必須先進行斷詞，使斷詞成為中文語料庫分析的起點。本研究使用國內學者發展、適合人文社會學科研究者使用的語料庫分析軟體「庫博中文語料庫分析工具」。該軟體針對中文的特性進行設計開發，適合用以分析中文文本，且工具簡單易用，毋須資訊工程人員的協助，幫助人文社會科學研究者獨立操作、突破以往難以處理大量文本資料的限制，並可藉由辭典校正功能，增補新詞彙以提高斷詞的正確度（闕河嘉，2018；闕河嘉、陳光華，2016）。

### 三、研究問題

本研究以我國國中課本的語料進行「國中學術詞彙表」的編製。科目包含數學、理化、生物、地科、地理、歷史、公民共七科，乃排除國文科與英文科以外的所有國中會考科目。而排除國文科及英語科乃參考 Greene 與 Coxhead (2015) 在建置中學詞表時排除「閱讀與文學」的考量。由於國文科包含新詩、散文及小說等體裁，甚至不在現代生活使用的文言文，而英語科的英文非我國國中生之母語，因而排除國文與英文兩科。

過往文獻在編製學術詞彙表的步驟可歸納為三原則：一、選用常出現於學術語料庫之詞彙。雖劉貞好等人 (2016) 使用 Hyland (2008) 的詞頻標準，但本研究中由於各次語料庫詞數較少，故改採出現單元數以篩選過於冷僻之詞彙。二、避免選用基本詞彙。由於國內缺乏適用於國中生的非學術詞彙語料庫，我們無法採用 Gardner 與 Davies (2014) 的比率標準，而如同多數學術詞彙表的作法 (劉貞好等人, 2016; Coxhead, 2000; Greene & Coxhead, 2015)，採取排除基本詞彙列表之詞彙。三、避免選用學科之專業術語。本研究與文獻同樣採用領域廣度指標，並於操作過程中調整適用本次研究之標準。

因 Hyland 與 Tse (2007) 發現學術詞彙在不同科目存在有不同的分布情況，本研究也將檢驗國中學術詞彙在各科目的分布狀況。同一學術詞彙在各科的出現單元比率之相關 (詳見下節「研究方法」) 可以作為分布狀況一致與否的指標之一，而且也可依詞彙在各科的出現單元比率進行分群，以了解哪些科目會較一致地常用某些學術詞彙，而罕用另一些學術詞彙。例如，《國民中小學九年一貫課程綱要》(2008) 分為數學領域、自然與生活科技領域、社會領域等不同的教學領域別，而學術詞彙出現的狀況可能於領域內較一致，而跨領域間會較不一致。

綜上所述，本研究之研究目的：(一) 以教科書語料建立國中學術詞彙表。(二) 探討國中學術詞彙出現在各科的分布狀況。

## 參、研究方法

### 一、國中教科書全詞彙語料庫

本研究採用《國民中小學九年一貫課程綱要》（2008）國中各版本之課本內容作為建置學術詞彙表之語料來源。因研究進行時，依據十二年國民基本教育課程綱要編製之課本尚缺九年級課本，未完整發行。依據前文對課本科目選擇的說明，語料來源包含七至九年級之數學科、理化、生物、地科、地理、歷史、公民共七科，各包含三個版本（N、K、H），總計 54 本課本、597 個單元（見表 1 第 2 欄）。

表 1 國中教科書詞彙語料來源

科目	單元數 (三版本相加)	斷詞詞次	全詞彙語料庫 詞彙數	詞彙保留比率
數學	185	1,302,480	121,328	9.32%
理化	54	485,209	103,678	21.37%
生物	35	250,650	59,420	23.71%
地科	15	126,496	32,707	25.86%
地理	104	295,223	79,669	26.99%
歷史	102	287,079	65,567	22.84%
公民	102	270,713	82,248	30.38%
總數	597	3,017,850	544,617	18.05%

掃描 54 本教科書內容、轉存文字檔之後，以庫博中文語料庫分析工具進行語料斷詞。除去目錄、前頁、附錄，得到課本內文共計 301 萬 7,850 詞次 (token) (闕河嘉、陳光華，2016)，內容包含中文方塊字、多字詞、數字、英文字母、標點符號，甚至亂碼以及不具語義的中文方塊字組合。

接著以下列四個條件進行詞彙的整理。首先，刪除明顯不為中文詞彙之詞次，如英文字、阿拉伯數字、亂碼等。其次，為去除斷詞系統的錯誤斷詞及過於冷僻之詞彙，以各科目單元平均數 ( $M = 85.29$ ) 減去一個標準差 ( $SD = 56.71$ ) 得到之標準 28，刪除出現總單元數未滿 28 之詞彙。其三，由於詞表獨立於文本脈絡之外，單詞較難有明確語義，因此刪除單字詞。最後，刪除語義不明之詞次，例如，完全不明其義且於教育部重編國語辭典修訂本中查無資料者，可判定不應為詞 (如：「的實」、「的決」、「畫一」等)。剩餘的皆被視為合理、有實質語義之詞彙，以此作為國中教科書全詞彙語料庫 (表 1 第 4 欄)。

國中教科書全詞彙語料庫僅占初始斷詞系統所斷詞次的 18.05% (表 1 第 5 欄)。而各科目最後保留的詞彙比率以數學科最少，僅占斷詞後詞次的 9.32%，乃因數學科課文包括相當多之阿拉伯數字、符號所致。

## 二、國中學術詞彙表

為確保學術詞彙表之詞彙專業性，以下列三種來源刪除非學術語言之詞彙，一為 Beck 等人 (2013) 所指之基本詞彙；二為姓名、國名、地名等專有名詞；三為兩種概念以上之詞彙。為篩選基本詞彙，本研究參考劉貞好等人 (2016) 之作法，將國家華語測驗推動工作委員會所編製並最新修訂之「華語八千詞表」(國家華語測驗推動工作委員會，無日期) 準備級一級、準備級二級、入門級、基礎級視為基本詞彙，從詞表中刪除這些詞彙，但保留可在 AWL 或 AVL 中對應之中文詞彙。本研究亦擴充華語八千詞表中「單詞」的應用，例如，準備級「來」的單詞，

則將詞表內包含「來」的詞亦一併刪除（如：「用來」、「帶來」、「來自」、「來看」、「以來」、「來說」等）；第二，由於專有名詞不符合學術詞彙之抽象性特性（Townsend et al., 2012），因此不予列入學術詞彙表中；第三，黃居仁（2005）認為表達語義的最小單位為詞，但於本次研究中斷詞系統錯誤地將兩種概念以上之組合視為詞彙，由於可再行拆分，不符合黃居仁所指之詞的定義，因此於學術詞彙表中排除，例如，「環境問題」、「工商業」等。最後參考 Farrell（1990）編製「半術語詞表」（semi-technical list）運用直覺刪除基本詞彙之步驟，本研究亦透過主觀認定刪除明顯可歸類於基本詞彙之詞，包含時間（如：「西元」、「民國」、「昔日」、「終年」等）、以人作為群體之代名詞（如：「科學家」、「消費者」、「原住民」等）、課本編輯用語（如：「動動腦」、「想一想」、「箭頭」、「示意圖」等）。除此之外，亦刪除「由於」、「位於」、「何者」、「進而」等不表示實際語義之虛詞。

學術詞彙表亦需符合學術詞彙之「跨領域」特性，因此本研究採用領域廣度與均勻度兩指標作為篩選標準，若詞彙在七個科目中僅於四個科目（含以下）出現，則未滿足跨領域之定義，予以刪除（僅保留四個可在 AWL 或 AVL 中對應之中文詞彙）。亦參考 AVL 計算各詞彙之均勻度，並設定標準以進行篩選。若均勻度愈趨近 1 表示其在各學門領域出現之頻率一致、愈趨近 0 則表示其僅出現於特定學門領域且頻率差異甚大（Juilland & Chang-Rodriguez, 1964）。若參照 Gardner 與 Davies（2014）以 .80 為標準進行篩選，僅包含 25 個詞彙，且未達標準之詞仍有不少可於 AWL 或 AVL 中找到對應的英文詞彙，顯示 .80 應過於嚴苛。因此，研究者檢視詞表之均勻度分布，發現低於 .30 的詞彙多為特定科目出現之專業術語，例如，「遺址」（均勻度為 .14）、「開墾」（均勻度為 .27）等，因此將篩選標準訂為 .30。刪除均勻度未達 .30 的詞彙，但保留 12 個可於 AWL 或 AVL 中找到對應的英文詞彙，包含「協調」（coordinate，均勻度 .29）、「反應」（reaction，均勻度 .25）、「對應」（counterpart，均勻度 .08）等詞。

爲回應研究問題二，觀察學術詞彙之出現狀況是否會因爲科目不同而有不同，分別計算各學術詞彙於七個科目之出現單元數，將其除以該科目的全部單元數，得到「出現單元比率」。例如，「利用」一詞在歷史科的出現單元比率爲 .41，表示該詞出現在歷史科的 42 個單元，占歷史科總數 102 個單元的 41%；但「利用」於理化科的出現單元比率爲 1.00，表示該詞在理化科中所有單元皆有出現。同一詞彙於不同科目的出現單元比率可逕行比較，而不受單元數量、語料庫大小影響。此外，本文也依據學術詞彙在各科目之出現單元比率進行集群分析（cluster analysis），後以推論統計分析比較不同組別之學術詞彙均勻度是否不同。分群方式採階層性集群分析（hierarchical cluster analysis）的 Ward's 法及計算平方歐氏距離（squared Euclidean distance）擷取適當之距離後決定分組之組數，並以 K 平均數集群分析法將學術詞彙進行分組。

## 肆、研究結果

通過對語料庫進行斷詞處理，並依照出現單元數、專業性、領域廣度之各項指標篩選，最終得到包含 779 個詞彙的「國中學術詞彙表」（<https://tinyurl.com/ywstm346>）。表 2 列出序號 1 至 48 之學術詞彙作爲範例，欄位包含序號、詞彙、出現總單元數（單元數）、領域廣度（R）、缺少科目（缺）、均勻度（D）。此 48 個學術詞彙中，有 20 個詞彙於 AWL 存在可對應之英文詞族，例如，「利用」可對應到“utilise”、「主要」可對應到“major”；45 個詞彙於 AVL 中存在可對應之英文詞群，如「形成」可對應至“form, formation”、「分別」可對應至“respectively, separate”。

表 2 序號前 48 之國中學術詞彙

序號	詞彙	單元數	R	缺	D
1	利用	447	7		.73
2	方式	391	7		.81
3	影響	383	7		.76
4	根據	375	7		.92
5	關係	368	7		.75
6	主要	361	7		.73
7	分別	358	7		.54
8	形成	349	7		.62
9	使用	336	7		.77
10	進行	335	7		.69
11	發展	333	7		.55
12	過程	331	7		.80
13	活動	330	7		.82
14	部分	328	7		.87
15	代表	315	7		.85
16	發現	310	7		.75
17	增加	309	7		.83
18	比較	306	7		.81
19	造成	303	7		.69
20	發生	303	7		.63
21	相同	301	7		.58
22	表示	299	7		.60
23	結果	288	7		.72
24	說明	286	7		.64

(續)

表 2 序號前 48 之國中學術詞彙 (續)

序號	詞彙	單元數	R	缺	D
25	逐漸	280	7		.68
26	通常	267	7		.78
27	需要	264	7		.74
28	產生	263	7		.59
29	環境	261	7		.55
30	情形	259	7		.79
31	判斷	256	7		.79
32	提供	256	7		.67
33	改變	250	7		.73
34	能力	250	7		.60
35	面積	250	7		.38
36	方法	249	7		.71
37	變化	249	7		.68
38	觀察	247	7		.59
39	正確	246	7		.82
40	閱讀	246	7		.67
41	經濟	242	5	數、理	.55
42	包括	239	7		.70
43	組成	237	7		.80
44	依據	235	7		.80
45	現象	227	7		.66
46	政府	227	6	數學	.41
47	導致	225	7		.75
48	位置	225	7		.72

## 一、學術詞彙表性質描述

表 3 為國中學術詞彙全部七科總單元數及領域廣度之分布表，其中詞彙平均總單元數為 90.06。縱軸依據學術詞彙在七科共 597 個單元中出現之總單元數高低分成六類，出現總單元數不足全科目總單元數 25%（150 單元）之學術詞彙共 85.24%（664 詞彙），符合 Coxhead（2000）的研究結果，即學術詞彙並非日常生活常用之基本詞彙。橫軸為學術詞彙之領域廣度分布，即學術詞彙在幾個科目中出現，領域廣度為 7 或 6 的占 80%，大致符合跨領域的特性。領域廣度為 6 的 227 個詞彙中，有 125 個詞彙獨缺數學科，其次依序獨缺地球科學科的為 42 個詞、公民科 19 個詞、歷史科 18 個詞、理化科 13 個詞、生物科 9 個詞、地理科 1 個詞。領域廣度為 5 的 148 個詞彙中，以「數學 + 地球科學」的缺詞組合為最大宗，共有 35 個詞，其次為「數學 + 理化」的缺詞組合，共 31 個詞。計算「數學與其他任一科」的缺詞組合共有 103 個詞，占全部領域廣度為 5 的詞彙中 69.59%。

表 3 國中學術詞彙全科目總單元數及領域廣度分布表

		領域廣度				
		7	6	5	4	總數
全 科 目 總 單 元 數	250 ~	35	0	0	0	35
	200 ~ 249	26	2	1	0	29
	150 ~ 199	44	4	3	0	51
	100 ~ 149	91	30	3	0	124
	50 ~ 99	133	90	43	2	268
	28 ~ 49	71	101	98	2	272
	總數	400	227	148	4	779

表 4 為國中學術詞彙均勻度及領域廣度分布表，其中所有詞彙的平均均勻度為 .58。縱軸將學術詞彙之均勻度分成七類，由表中可觀察到，領域廣度較高之詞彙常是均勻度較高之詞彙。

表 5 為各科目缺失及含有學術詞彙之數量及比率，顯示除數學外其餘六科皆包含超過 86% 之學術詞彙，而數學科僅包含總表 70.35% 之詞彙。例如，「組織」，在除數學科外其餘六科之平均出現單元比率為 .33，平均每三個單元就會有一個單元出現「組織」一詞，但在數學科 185 個單元中一次皆未出現。同樣的狀況亦見於「引發」（其餘六科平均出現單元比率為 .32）及「設立」（其餘六科平均出現單元比率為 .29），僅於數學科一次皆未出現，表示數學科的學術詞彙屬性與其他六科不同。

表 4 國中學術詞彙均勻度及領域廣度分布表

		領域廣度				
		7	6	5	4	總數
均勻度	.80 ~	25	0	0	0	25
	.70 ~ .79	117	25	3	0	145
	.60 ~ .69	115	61	10	0	186
	.50 ~ .59	87	63	47	0	197
	.40 ~ .49	37	48	53	2	140
	.30 ~ .39	14	27	32	1	74
	~ .29	5	3	3	1	12
	總數	400	227	148	4	779

表 5 各科目缺失學術詞彙數、包含學術詞彙比率及單元比率

	缺失學術詞彙數	包含學術詞彙比率	單元比率 平均數 (標準差)
數學	231	70.35%	.06 (.12)
理化	66	91.53%	.23 (.24)
生物	41	94.74%	.26 (.24)
地科	106	86.39%	.31 (.27)
地理	8	98.97%	.17 (.17)
歷史	42	94.61%	.16 (.16)
公民	41	94.74%	.19 (.18)

## 二、學術詞彙之各科目差異

為回應研究問題二，以七科為自變項對全部 779 個詞彙在各科的出現單元比率進行單因子重複測量變異數分析。表 5 最右欄為各科目之單元比率平均數及標準差。由於未滿足 Mauchly 的球形檢定結果 ( $p < .001$ )，統計結果將自由度以 Greenhouse-Geisser 法進行校正。結果顯示主要效果達顯著， $F(3.60, 2803.64) = 199.70$ ， $p < .001$ ， $\text{partial } \eta^2 = .20$ ，顯示七個科目的單元比率不盡相同。後續事後比較，經 Bonferroni 法校正後，除數學科之單元比率分別顯著小於所有其他六科，其餘各科目間兩兩差異也多數達顯著水準 ( $p_s < .010$ )，僅地理科及歷史科、地理科及公民科兩組比較未達顯著 ( $p_s > .050$ )。

由於考量數學科僅包含總學術詞彙之 70% (見表 5)，可能因過多詞彙之單元比率為 0 而影響統計結果，我們另外針對領域廣度為 7 之 400 個學術詞彙進行相似的分析。結果顯示這 400 個學術詞彙出現於七個科目的單元比率亦不相同， $F(3.96, 1580.69) = 173.01$ ， $p < .001$ ， $\text{partial } \eta^2 = .30$ 。事後比較的結果多數達顯著，未達顯著差異的僅理化科及生物科、

地理科及歷史科、地理科及公民科三組比較 ( $p_s > .050$ )，結果相似於全部 779 個學術詞彙的結果。

爲了解各科目兩兩間之關係，表 6 爲同一詞彙於不同科目間之出現單元比率相關係數表（單元比率原始資料見完整列表之 AP 欄至 AV 欄）。若兩科目間之相關係數高，即表示詞彙於兩科目的出現單元比率相似。表中可觀察到各科目間的關係，顯示數理領域（數學科及理化科）、自然領域（理化科、生物科及地球科學科）、地球科學科與地理科、社會領域（地理科、歷史科及公民科）四個子群內的相關係數皆較跨子群的科目高。亦即若一個學術詞彙於理化科有高單元比率，也會在生物科及地球科學科具有高單元比率，反之亦然。

爲進一步回應研究問題二，我們以集群分析法對學術詞彙進行分組。依據各學術詞彙在各科目的出現單元比率集群分析的結果，可根據適當組間距離將學術詞彙分爲五組，見表 7。第一組爲高單元比率的學術詞彙 ( $n = 49$ )，平均出現單元比率爲 .60，表示在全部科目、全部單元中，有 60% 之單元至少出現該詞彙一次，包含「使用」、「發生」等詞彙；第二組爲低單元比率的學術詞彙 ( $n = 400$ )，較不常出現於各

表 6 學術詞彙於各科目單元比率之相關係數表

	數學	理化	生物	地科	地理	歷史	公民
數學	-						
理化	.56***	-					
生物	.39***	.74***	-				
地科	.38***	.69***	.69***	-			
地理	.19***	.29***	.43***	.52***	-		
歷史	.11**	.17***	.27***	.23***	.51***	-	
公民	.22***	.25***	.40***	.26***	.47***	.62***	-

\*\* $p < .01$ ，\*\*\* $p < .001$

科目組，平均出現單元比率為 .10，表示有 10% 之單元至少出現該詞彙一次，包含「工具」、「理由」等詞彙；第三組為常見於歷史公民組（ $n = 123$ ），在歷史科及公民科之平均出現單元比率為 .35，但在其他科目之平均出現單元比率僅為 .15，包含「設立」、「改善」等詞彙；第四組為常見於生物地科地理組（ $n = 71$ ），在生物、地科、地理等三個科目中之平均出現單元比率為 .51，但在其他四個科目之平均出現單元比率僅為 .19，包含「吸引」、「技術」等詞彙；第五組為常見於自然領域組（ $n = 136$ ），在理化、生物、地科等三科目中之平均出現單元比率為 .44，但在其他四個科目之平均出現單元比率僅為 .10，包含「實驗」、「操作」等詞彙。

為進一步增加分組合理性之佐證資料，另以組別為獨變項、均勻度為依變項，進行單因子組間變異數分析以檢定五組之均勻度是否相等，僅在部分科目較常出現、其餘科目較不常出現之組別，均勻度應小於在各科目皆較常出現、或皆較不常出現之組別。由於不滿足變異數同質性（ $p = .001$ ），因此以 Welch 法及 Brown-Forsythe 法進行校正。Welch 統計數之值為  $F(4, 200.11) = 42.00$ ， $p < .001$ ，Brown-Forsythe 統計數之值為  $F(4, 427.45) = 29.07$ ， $p < .001$ ，顯示各組之平均均勻度不盡相同。後以 Games-Howell 檢定法進行事後比較，結果顯示除第二組及第三組、第四組及第五組外（ $p_s > .050$ ），其餘各組組間差異皆達到顯著水準， $p_s < .010$ 。顯示第一組之平均均勻度大於第四組及第五組之平均均勻度，亦大於第二組及第三組之平均均勻度。

集群分析結果顯示第一組（高單元比率組）的詞彙亦是均勻度較高的詞彙。例如，「利用」、「發展」等詞彙，除了在各科目皆較常出現外，各科目的詞頻均差不多，表示第一組確實為常出現於各科目組。第四組（常見於生物地科地理組）與第五組（常見於自然領域組）的詞彙則是均勻度次之的詞彙。例如，第四組的「吸引」、「技術」、與第五組的「實驗」、「形狀」等詞彙，表示詞彙雖常出現於該主要科目，亦會以較低詞頻出現於其他科目中。第二組（低單元比率組）與第三組

表 7 依單元比率進行分組之各組平均數及標準差

	數學	理化	生物	地科	地理	歷史	公民	均勻度
	<i>M</i> ( <i>SD</i> )							
第一組 ( <i>n</i> = 049)	.28 (.24)	.79 (.18)	.83 (.11)	.89 (.12)	.48 (.22)	.44 (.23)	.53 (.18)	.72 (.09)
第二組 ( <i>n</i> = 400)	.02 (.06)	.11 (.10)	.13 (.11)	.15 (.13)	.09 (.08)	.09 (.08)	.11 (.09)	.55 (.13)
第三組 ( <i>n</i> = 123)	.03 (.07)	.11 (.11)	.20 (.16)	.20 (.14)	.22 (.15)	.33 (.18)	.38 (.16)	.54 (.12)
第四組 ( <i>n</i> = 071)	.07 (.11)	.31 (.18)	.43 (.20)	.67 (.16)	.42 (.20)	.17 (.12)	.20 (.14)	.63 (.14)
第五組 ( <i>n</i> = 136)	.11 (.15)	.47 (.19)	.40 (.18)	.46 (.17)	.10 (.07)	.09 (.09)	.12 (.11)	.61 (.15)

註：第一組為高單元比率組、第二組為低單元比率組、第三組為常見於歷史公民組、第四組為常見於生物地科地理組、第五組為常見於自然領域組。

(常見於歷史公民組)的詞彙則具有較低之均勻度，例如，第二組的「工程」、「劃分」與第三組的「促進」、「教育」等詞彙，除在各科目出現的單元比率較少外，科目間的詞頻亦分布不均，較集中出現於特定科目、特定單元。

## 伍、結論與建議

本研究以國中階段共七個學科各三版本的課本作為編製國中學術詞彙表的語料來源。以頻率指標(單元數)、排除基本詞彙、考慮領域廣度與均勻度等標準後，建置含有 779 個學術詞彙之「國中學術詞彙表」。學術詞彙出現於各科目的分布狀況明顯不同，數學科尤為特別，僅包含總表七成之學術詞彙，且出現單元比率也顯著較其餘科目少。隨後依據

出現單元比率進行學術詞彙的集群分析，結果分爲五組最合宜，包括反映出現單元比率高、低的兩組，以及常見於歷史公民組、生物地科地理組、自然領域組，五組間的詞彙均勻度不盡相同，高單元比率組的學術詞彙其均勻度較高，亦即這些學術詞彙普遍出現在較多單元、較多科目；而低單元比率組的均勻度亦較低；其他三組則居中。

## 一、再論學術詞彙的特性

首先，針對學術詞彙的多義性與抽象性。觀察「國中學術詞彙表」，可能有些詞彙直覺上難以被認可爲「學術詞彙」，如表 2 之「方式、主要、分別、形成、部分、組成……」，似乎不具難度。然而這些詞彙多數能對應到 AWL 與 AVL 之詞彙表，而且在建構學科知識的脈絡下存在詞彙意義的多義性與抽象性。以「部分」(part) 一詞爲例，其意義的探討可遠溯至希臘哲學的分體論 (Merology)；在自然科學中，部分—整體關係可分爲六種、部分部分關係至少兩種 (陳世文、楊文金, 2008; 楊文金, 2007; Winston et al., 1987)；再考量心理學上如何看待社會事件的「部分」，以及計算語言學上如何處理「部分」的語義關係，就更顯示其意義並不簡單 (楊文金, 2007)。楊文金的實徵結果也發現，讀了包含「血球分爲紅血球、白血球和血小板」此一類屬 (kind of) 論述的文本後，國、高、大學生整體有 56% 接受「白血球是血球的一部分」此種組成 (part of) 論述的描述。然而後者具有歧義性，若解讀爲「視血球及白血球爲集合名詞，而血球這個集合包含白血球集合」，或可接受；若爲「一個血球中有一部分是白血球」，就屬錯誤解讀。此外，表 2 的「形成」及「造成」兩個學術詞彙也曾被研究過 (林文杰、楊文金, 2008)；而且前述這些研究都是針對使用母語的讀者，顯示學術詞彙對母語使用者而言仍具深究之空間。「國中學術詞彙表」裡這些在日常生活溝通無礙的詞彙，放在傳遞學科觀念的教科書中，因爲學生處於學習該學科、專業知識不足的狀態，掌握這些詞彙在學科脈絡下的意義並協同建構學科知識，並非易事。尤有甚者，由於學術詞彙直觀上

不具難度的特性，教師容易輕忽對學術詞彙的指導（Hiebert & Lubliner, 2008; Ogle et al., 2015; Townsend et al., 2012），進而未能察覺學生可能建構了不精準的學科知識。

其次，本研究從領域廣度及出現頻率的資料呈現了學術詞彙既具跨領域又具分領域的特性。在七科中至少出現六科的國中學術詞彙超過 80%，且除了數學科之外的六科都包括 86%（地科）~99%（地理）的國中學術詞彙，顯示這些詞彙相當普遍地出現在各科。然而在各科都有出現，並不代表出現的頻率類似。針對 779 個學術詞彙考驗在各科的出現單元比率（表 5），結果科目間有顯著差異，事後比較也顯示絕大多數的兩兩比較都達顯著差異；即使僅針對領域廣度為 7 的 400 個詞彙進行考驗，也得到類似的結果，這顯示相同的學術詞彙出現在不同科目的單元比率顯著不同。此一現象也反映在表 6 各科單元比率之相關係數表，雖然相關係數（ $r_s = .11 \sim .74$ ）都達顯著，然中位數 .39 不算高。進一步的集群分析顯示可分成五群詞彙，高單元比率的 49 個詞同時也是高均勻度的詞彙、低單元比率的 400 個詞屬低均勻度的詞彙，另外三群則是較常出現於歷史與公民科的 123 個詞、較常出現於生物、地科，與地理的 71 個詞，以及較常出現於理化、生物、地科等自然領域的 136 個詞，這顯示自然科與社會科的傳統區分，確實也存在學術詞彙的差異。

第三，延續頻率的議題，學術詞彙以出現頻率作為判準之一，並非只為了學者分類，而是詞彙頻率影響語言使用者的認知狀態。因學術詞彙在非學術文本中不如基本詞彙的高頻，使讀者對其意義的掌握不像基本詞彙那麼好，而認知狀態才是學習與教學必須關照的重點。例如，「面積」在小學階段可能是專業術語，必須透過正式教學方能理解其意。但到了國中階段，面積一詞的詞頻變高了，也成為介紹新的專業術語的輔助性語詞，此時面積一詞的角色就從專業術語轉變為學術詞彙。由此可知，詞彙位於何種階層不是固定的，必須考量學習情境以及學習者的認知狀況，這呼應了 Greene 與 Coxhead（2015）詞彙並非絕對屬於基

本詞彙、學術詞彙、專業術語之主張。此一觀點說明了國中學術詞彙表有其獨特性，但也有其局限性。此外，此一思考路線將衍生學術詞彙在專業術語到基本詞彙之光譜間發展的可能性，這也是未來研究可探討的問題。

第四，本研究凸顯了數學科的學術詞彙相當特殊，值得從數學的本質與教科書編輯的角度加以探討。「數學是一種語言」是經常被討論的觀點，不論認同這主張（Usiskin, 1996）或者認為僅是譬喻（Pimm, 2019）的學者，都會承認數學文本相較於其他學科文本存在明顯高比率的數學符號與算式，使得文字的比率較低。而學術詞彙是以文字為標的，較低比率的文字可能是數學科學術詞彙不同於其他學科的原因之一。然而，表 1 顯示在刪除數字、英文字母、標點符號之後的全詞彙語料庫中，數學科有 12 萬多個詞，是七科中最多的，因此並不能僅以數學科的文字較少來解釋。第二種解釋是數學演繹的本質不同於自然科學與社會科學之歸納的特性；數學知識是建立在抽象的假設上，論證所得的結果也未必與現實世界有對應，在真實世界之外所建構的數學世界甚至不需要討論對人類的影響，這與自然與社會科學透過觀察、歸納，以及推論其作用，差異甚大。因為思維與探究方式的差異，可能使不少數學科獨特的學術詞彙無法通過跨領域指標，最後保留在跨領域學術詞彙表中的語詞又非數學科會使用的。第三種解釋是從語文說明的重要性來探討數學科的獨特性；數學專業領域的論證主要仰賴證明，語言的解說無法達到證明的精準要求，因此數學學術期刊的文章主要依賴符號與算式，故而既精簡且文字稀少。國中階段的學生理應需要語言的解說方能掌握數學原理的意義與程序，但教科書常見以運算或圖示為例，致使文字的解說相對簡單。學術詞彙是輔助解釋專業術語概念的語詞，愈強調解釋學術概念的文本就愈需要學術詞彙的輔助，而數學教科書似乎以運算或圖示來承載概念解釋的責任，而降低了學術詞彙出現的機會。

## 二、學術詞彙表建置程序的反思

首先，中文在建置學術詞彙表會有不同於拼音系統的文字議題。AWL 與 AVL 等英文學術詞彙表需要處理詞族或詞群的問題，而中文則需斷詞。中文斷詞系統在專業術語常常需人工校對的步驟（黃仲義等人，2021），但在學術詞彙層次上的正確性則尚不清楚。本研究以庫博中文語料庫分析工具分析國中教材之語料，乃借助該軟體適合人文社會學科研究學者使用的特性與針對中文使用語境的優勢，使人文社會學科之研究者可獨立運用，而不需仰賴資訊工程人員的協助，然而本研究在斷詞合理性的檢查尚有不足：庫博中文語料庫分析工具有內建如自建詞典、同類詞整合等功能以修正斷詞結果，但此種方式需逐科、逐詞地修訂內建詞典，且最易錯誤斷詞的是各科的專業術語而非學術詞彙，因此本研究未進行此一步驟，是研究限制之一。未來比較各種斷詞系統在學術詞彙層次的斷詞正確性，例如，庫博中文語料庫分析工具、線上開源軟體——Jieba 中文分詞（<https://github.com/fxsjy/jieba>），或中央研究院中文詞知識庫小組（Chinese Knowledge and Information Processing, CKIP）之中文斷詞系統（<https://ckip.iis.sinica.edu.tw/>）何者較佳，是個值得進行的研究方向。

其次，排除語料中基本詞彙之程序的檢討。過去研究存在兩種方式，第一種是 AWL（Coxhead, 2000）、中文學術常用詞表（劉貞好等人，2016）、MSWL（Greene & Coxhead, 2015）採取的方式——以具公信力、可對應之合適詞彙表作為基本詞彙的依據，加以排除；優點為對照他人建置之詞表，程序較不費力，然而必須存在合宜的詞彙表。第二種作法為 AVL（Gardner & Davies, 2014）所採用的比率標準——學術詞彙須於學術語料庫中的出現頻率為規模類同之非學術語料庫的倍數以上，未達此倍數的詞彙被視為基本詞彙並排除之；優點為此標準更呼應學術詞彙之基本定義，即學術詞彙出現於學術語料庫之頻率應高於非學術語料庫，然此作法除須建製學術語料庫外，亦須蒐集非學術語料，並進行

分析對照，程序上難度較高。AVL 所選用之非學術語料庫為 COCA，乃當代最大之免費英語語料庫，除學術語料庫外亦包含如口語、小說、雜誌等非學術語料庫。而本研究於國內未能搜尋到基於國中階段所建置之非學術語料庫，因此無法採用第二種的比率標準。同時本研究也未找到官方或具公信力且適用於國中階段之基本詞彙表：《十二年國民基本教育課程綱要國民中小學暨普通型高級中等學校——語文領域：國語文》（2018）雖規範第三學習階段（小學五、六年級）學生應認念 4,500 個常用語詞、使用 3,700 個常用語詞，然未明列語詞內容以對照，因此各版本國語課本的習寫與認讀語詞也不盡相同：《國小學童常用字詞調查報告書》（教育部國語推行委員會，2000）雖依照詞頻排列共收錄 46,666 個相異詞，詞頻或許能提供基本詞彙界定之參考，然該詞表之高頻字詞與目前教科書語詞與新聞用語（如：國語日報）的重疊性有限（吳鑑城、邱重毅，2021；李詩敏等人，2020），甚至「所列常用字在近期語料中都有超過 30% 的字已非屬於常用字」（吳鑑城、邱重毅，2021，頁 28），因此《國小學童常用字詞調查報告書》並非適合用來篩選基本詞彙之詞表。本研究僅能選用為華語學習者所建立的華語八千詞表，雖未能配合研究對象為華語母語者的狀態，但其具有將詞彙明確區分為準備級、入門級、基礎級、進階級……而便於據此刪除基本詞彙之優點。此外，本研究排除華語八千詞表之基本詞彙還發現有另外兩個問題：一是華語八千詞表未必能涵蓋中文語詞的變化，例如，單詞「來」，可延伸出華語八千詞表上未包含之基本詞彙（如：「用來」、「來自」等）；二是斷詞後出現虛詞附著在實詞上，難以與華語八千詞表的詞彙完全吻合。上述兩者情況皆需仰賴人工校對才可決定詞彙的去留，是採取排除基本詞彙之程序會面臨的主觀性問題。未來研究國內若因應不同年齡層之非學術語料庫逐漸被建置，採用比率標準不只能避免上述需要人工校對的狀況，亦更能符合學術詞彙在學術場域詞頻較非學術場域高之定義。

第三，延續上一小節的第二點，既然學術詞彙同時具備跨領域與分領域性，那就需要重新思考篩選學術詞彙的均勻度指標。以往研究

(Dang et al., 2017; Gardner & Davies, 2014; Lei & Liu, 2016; Paquot, 2007) 並未對「足夠均勻分布」的標準達成共識，甚至也未列明公式。Dang 等人 (2017) 提出之口語學術詞彙表 (academic spoken word list, ASWL) 標準為 .60、Lei 與 Liu (2016) 提出之醫學學術詞彙表 (middle academic vocabulary list, MAVL) 標準亦為 .60，但皆未附上均勻度計算公式，總表資料亦未附上詞頻以協助釐清。本研究編製國中學術詞彙表是以 Juilland 與 Chang-Rodriguez (1964) 提出之均勻度公式來反映詞彙於各科目中之分布情況，並在過程中上下調整標準，最後以 .30 為篩選標準。均勻度會從文獻的 .60 調降到 .30 的主要原因，是因為若以 .60 為標準，將有不少與 AWL 或 AVL 詞表具對應性的中文詞彙無法被納入，而降至 .30 之後，就開始出現較多的專業術語。我們推測本研究的均勻度會降低到 .30，可能與本研究全詞彙語料庫總數僅 54 萬詞有關。未來若擴大語料庫，如納入不同課綱之國中課本，或許可選用較嚴格之標準。

### 三、研究建議與限制

國中學術詞彙表在教育實務上有以下促進教學、學習，與教科書編寫之功能。其一，多數教師側重專業術語的教學，容易忽略學術詞彙的難度。學術詞彙表具有提醒教師關注學生能否掌握此類詞彙意義的效果，也能促使教師留意同一學術詞彙在不同科目中的細微差異，並在教學時多加區辨。而分領域較為獨特的學術詞彙（如：「改善」一詞較常出現於歷史及公民科），也能引導該領域教師著力協助學生理解。其次，學術詞彙測驗能篩選有學習困難可能性之學生以及評估教學介入之效果，進而促進學習。Townsend 等人 (2012) 的研究顯示，學術詞彙的掌握程度可預測學生學業成就表現，若學生未良好掌握學術詞彙，未來的學業發展將落後於同儕。第三，本研究之國中學術詞彙表應可做為教材編製的參考。一個學術詞彙的頻率與其理解難度有關，在特定主題脈絡第一次出現時，可能需要佐以舉例，以協助學生學習。

最後，總結本研究的限制，並提出可能的改善方向：一、本研究語料庫規模不夠大，容易造成指標的偏誤；考量前後課程綱要對學術詞彙影響可能有限，或許同時納入前後期課程綱要下教科書的語料是擴大語料庫的方法；二、斷詞系統在學術詞彙層次的品質未經考驗，此一議題值得獨立進行探究；三、本研究採用排除華語八千詞表基本詞彙，以及自行擴充其表中單詞的衍生詞，此一作法有主觀之嫌。未來改採選一非學術語料庫做為詞頻對照的基礎，可能是較合宜的方法。

### 致謝

本文改寫自鍾涵灝在吳昭容指導下的碩士論文之研究一。感謝楊文金教授、曾玉村教授在論文口試時深具啟發的指點，以及期刊審查委員的建議。此外，感謝行政院科技部補助專題研究計畫「以眼動探討幾何閱讀歷程與發展閱讀技巧教學」(MOST 108-2511-H-003-014-MY3) 研究經費補助。

## 教科書參考書目

- 陳世煌、方崇維、姚珩、許貫中、李通藝（主編）（2014）。國民中學自然與生活科技（再版，第二冊，一下）。翰林。
- [Chen, S.-H., Fang, C.-W., Yao, Y., Syu, G.-Z., & Li, T.-Y. (Eds.). (2014). *National junior high school science and technology* (2nd ed., Vol. 2, 1st grade 2nd semester). Hanlin.]

## 參考文獻

- 十二年國民基本教育課程綱要國民中小學暨普通型高級中等學校——語文領域：國語文（2018）。
- [*Curriculum guidelines of 12-year basic education: Mandarin domain for elementary, junior high school and upper secondary school education.* (2018).]
- 吳鑑城、邱重毅（2021）。第一至四學習階段教學語料庫的建置技術及應用之研究（計畫編號：NAER-2019-029-C-1-1-B4-01）。國家教育研究院。<http://>

- www.grb.gov.tw/search/planDetail?id=13246992
- [Wu, J.-C., & Chiu, C.-Y. (2021). *Research on the construction technology and application of teaching corpus in primary and secondary schools* (Project No. NAER-2019-029-C-1-1-B4-01). National Academy for Educational Research. <http://www.grb.gov.tw/search/planDetail?id=13246992>]
- 李詩敏、李台元、蔡旻穎（2020）。第一至四學習階段國語文領域常用字詞之分析研究（計畫編號：NAER-2019-029-C-1-1-B4-02）。國家教育研究院。  
<http://www.grb.gov.tw/search/planDetail?id=13247015>
- [Li, S.-M., Li, T.-Y., & Tsai, M.-Y. (2020). *An analysis of characters and words frequently used in the area of Chinese language* (Project No. NAER-2019-029-C-1-1-B4-02). National Academy for Educational Research. <http://www.grb.gov.tw/search/planDetail?id=13247015>]
- 林文杰、楊文金（2008）。高一學生對物理文本過程詞隱含之連接關係的理解。《物理教育學刊》，9（1），1-18。
- [Lin, W.-C., & Yang, W.-G. (2008). An exploration of tenth graders' understanding of the relations implied by the processes deployed in physics texts. *Chinese Physics Education*, 9(1), 1-18.]
- 國民中小學九年一貫課程綱要（2008）。  
[Grade 1-9 curriculum guidelines. (2008).]
- 國家華語測驗推動工作委員會（無日期）。華語八千詞。2021年6月24日，取自 TOCFE。  
<http://www.sc-top.org.tw/chinese/download>
- [Steering Committee for the Test of Proficiency-Huayu. (n.d.). *Chinese 8000 vocabulary*. Retrieved June 24, 2021, from TOCFL. <http://www.sc-top.org.tw/chinese/download>]
- 教育部國語推行委員會（2000）。國小學童常用字詞調查報告書。教育部。  
[https://language.moe.gov.tw/001/Upload/files/SITE\\_CONTENT/M0001/PRIMARY/SHINDEX.HTM](https://language.moe.gov.tw/001/Upload/files/SITE_CONTENT/M0001/PRIMARY/SHINDEX.HTM)
- [National Languages Committee. (2000). *The report of frequently used words and vocabulary in elementary school*. Ministry of Education. [https://language.moe.gov.tw/001/Upload/files/SITE\\_CONTENT/M0001/PRIMARY/SHINDEX.HTM](https://language.moe.gov.tw/001/Upload/files/SITE_CONTENT/M0001/PRIMARY/SHINDEX.HTM)]
- 陳世文、楊文金（2008）。學生對科學教科書詞彙關係理解之分析。《教科書研究》，1（2），101-127。  
<https://doi.org/10.6481/JTR.200812.0101>
- [Chen, S.-W., & Yang, W.-G. (2008). An analysis of student comprehension of lexicon relations in science textbooks. *Journal of Textbook Research*, 1(2), 101-127. <https://doi.org/10.6481/JTR.200812.0101>]
- 黃仲義、陳世文、楊文金（2021）。我國國民中小學科學教科書科學詞彙之差異比較。《教科書研究》，14（1），1-29。  
[https://doi.org/10.6481/JTR.202104\\_14\(1\).01](https://doi.org/10.6481/JTR.202104_14(1).01)

- [Huang, C.-Y., Chen, S.-W., & Yang, W.-J. (2021). The comparison of science vocabulary in Taiwan primary and middle school science textbooks. *Journal of Textbook Research*, 14(1), 1-29. [https://doi.org/10.6481/JTR.202104\\_14\(1\).01](https://doi.org/10.6481/JTR.202104_14(1).01)]
- 黃居仁 (2005)。漢字知識表達的幾個層面：字，詞，與詞義關係概論。載於臺北市府文化局 (主編)，*漢字與全球化國際學術研討會論文集* (頁 77-88)。臺北市府文化局。
- [Huang, C.-R. (2005). Knowledge representation with Hanzi: The relationship among characters, words, and senses. In Department of Cultural Affairs, Taipei City Government (Ed.), *The proceedings of the international symposium on Chinese characters and globalization* (pp. 77-88). Department of Cultural Affairs, Taipei City Government.]
- 楊文金 (2007)。學生對「類屬－組成」論述的語意理解——以「血液」文本為例。 *科學教育學刊*，15 (2)，195-214。 <https://doi.org/10.6173/CJSE.2007.1502.04>
- [Yang, W.-G. (2007). Students' semantic understanding of "kind-of/part-of" discourses: An example of the "blood" science text. *Chinese Journal of Science Education*, 15(2), 195-214. <https://doi.org/10.6173/2fCJSE.2007.1502.04>]
- 劉貞好、陳浩然、楊惠媚 (2016)。藉學術語料庫提出中文學術常用詞表：以人文社會科學為例。 *華語文教學研究*，13 (2)，43-87。
- [Liu, C.-Y., Chen, H. H.-J., & Yang, H.-M. (2016). Compiling a Chinese academic wordlist based on an academic corpus. *Journal of Chinese Language Teaching*, 13(2), 43-87.]
- 闕河嘉 (2018)。庫博中文語料庫分析工具的數位人文價值。 *人文與社會科學簡訊*，19 (2)，118-123。
- [Chueh, H.-C. (2018). The digital humanistic value of CORPRO. *Humanities and Social Sciences Newsletter Quarterly*, 19(2), 118-123.]
- 闕河嘉、陳光華 (2016)。庫博中文獨立語料庫分析工具之開發與應用。載於項潔 (主編)，*數位人文：在過去、現在和未來之間* (頁 285-313)。國立臺灣大學出版中心。 <https://doi.org/10.13140/RG.2.2.36253.97768>
- [Chueh, H.-C., & Chen, K.-H. (2016). CORPRO: A Chinese language corpus tool and a case study of media representation of organic agriculture. In J. Hsiang (Ed.), *Digital humanities: Between past, present, and future* (pp. 285-313). National Taiwan University Press. <https://doi.org/10.13140/RG.2.2.36253.97768>]
- Bailey, A. L., & Heritage, M. (2008). *Formative assessment for literacy, grades K-6: Building reading & academic language skills across the curriculum*. Corwin.
- Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent & Adult Literacy*, 54(1), 4-12. <https://doi.org/10.1598/JAAL.54.1.1>
- Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. The Guilford Press.

- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>
- Crosson, A. C., McKeown, M. G., & Ward, A. K. Jr. (2019). An innovative approach to assessing depth of knowledge of academic words. *Language Assessment Quarterly*, 16(2), 196-216. <https://doi.org/10.1080/15434303.2019.1612899>
- Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, 67(4), 959-997. <https://doi.org/10.1111/lang.12253>
- Farrell, P. (1990). *Vocabulary in ESP: A lexical analysis of the English of electronics and a study of semi-technical vocabulary*. Trinity College, Centre for Language and Communication Studies.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327. <https://doi.org/10.1093/applin/amt015>
- Greene, J. W., & Coxhead, A. (2015). *Academic vocabulary for middle school students*. Paul H. Brookes.
- Harmon, J., & Wood, K. (2018). The vocabulary-comprehension relationship across the disciplines: Implications for instruction. *Education Sciences*, 8(3), 101-109. <https://doi.org/10.3390/educsci8030101>
- Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In S. J. Samuels & A. Farstrup (Eds.), *What research has to say about vocabulary instruction* (pp. 106-129). International Reading Association.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly*, 41(2), 235-253. <https://doi.org/10.1002/j.1545-7249.2007.tb00058.x>
- Juilland, A., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. De Gruyter Mouton.
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 804-823. <https://doi.org/10.1037/0278-7393.20.4.804>
- Lawrence, J. F., Hagen, A. M., Hwang, J. K., Lin, G., & Lervåg, A. (2019). Academic vocabulary and reading comprehension: Exploring the relationships across measures of vocabulary knowledge. *Reading and Writing*, 32(2), 285-306. <https://doi.org/10.1007/s11145-018-9865-2>
- Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53. <https://doi.org/10.1016/j.jeap.2016.01.008>
- Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary

- knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, 31, 44-57. <https://doi.org/10.1016/j.jeap.2017.12.006>
- Mokhtari, K., & Velten, J. (2015). Strengthening academic vocabulary with word generation helps sixth-grade students improve reading comprehension. *Middle Grades Research Journal*, 10(3), 23-42.
- Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91-108. <https://doi.org/10.1002/RRQ.011>
- Nation, I. S. (2013). *Learning vocabulary in another language Google eBook*. Cambridge University Press.
- Nation, I. S., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Cengage Learning.
- Ogle, D., Blachowicz, C., Fisher, P., & Lang, L. (2015). *Academic vocabulary in middle and high school: Effective practices across the disciplines*. Guilford.
- Paquot, M. (2007). Towards a productively-oriented academic word list. In J. Walinski, K. Kredens, & S. Gozdz-Roszkowski (Eds.), *Practical applications in language and computers 2005* (pp. 127-140). Peter Lang.
- Pimm, D. (2019). *Routledge revivals (1987): Speaking mathematically: Communication in mathematics classrooms*. Routledge.
- Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Routledge.
- Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, 49, 157-165. <https://doi.org/10.1016/j.learninstruc.2017.01.005>
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27(5), 499-520. [https://doi.org/10.1016/0749-596X\(88\)90022-8](https://doi.org/10.1016/0749-596X(88)90022-8)
- Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112-133). Cambridge University Press.
- Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal*, 112(3), 497-518. <https://doi.org/10.1086/663301>
- Usiskin, Z. (1996). Mathematics as a language. In P. C. Elliott & M. J. Kenney (Eds.), *Communications in mathematics, K-12 and beyond* (pp. 231-243). National Council of Teachers of Mathematics.
- West, M. (1953). *A general service list of English words*. Longman.
- Winston, M. E., Chaffin, R., & Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, 11(4), 417-444. [https://doi.org/10.1207/s15516709cog1104\\_2](https://doi.org/10.1207/s15516709cog1104_2)

## 附錄

均勻度 (Dispersion, D)

於本篇研究中，針對各詞彙在各科目之詞頻（單位：每一萬詞）進行計算。

$$\text{Dispersion}(D) = 1 - \frac{Cv}{\sqrt{N-1}} = 1 - \frac{\sigma/\mu}{\sqrt{N-1}}$$

$N$  為次語料庫總數，本篇研究為數學、理化、生物、地科、地理、歷史、公民七科，因此  $N = 7$ 。

$Cv$  為變異係數 (coefficient of variation)，用以表示資料之分散程度，可比較單位不同或單位相同但資料差異甚大之資料分散情況。計算方式為  $\sigma$  (標準差) 除以  $\mu$  (平均數)。此處各項統計數因用以描述資料分散情況，並未涉及推論統計，因此符號以及計算上皆將資料視為母體，並非  $s$  ( $\sigma$  之不偏估計數) 或  $\bar{x}$  (樣本平均數)。

以「使用」為例：

科目	原始出現次數	全詞彙語料庫 詞彙數	詞頻 (每一萬詞)	詞頻平方值 <sup>a</sup>
數學	146	121,328	12.03	144.7209
理化	515	103,678	49.67	2467.1089
生物	168	59,420	28.27	799.1929
地科	34	32,707	10.40	108.1600
地理	113	79,669	14.18	201.0724
歷史	102	65,567	15.56	242.1136
公民	239	82,248	29.06	844.4836
總數			159.17	4806.8523
平均			22.7386	686.6932

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum(X - \mu)^2}{N}} = \sqrt{\frac{\sum X^2}{N} - \mu^2} = \sqrt{\frac{4806.8523}{7} - 22.7386^2} \\ &\approx \sqrt{686.6932 - 517.0439} = \sqrt{169.6493} \approx 13.0249\end{aligned}$$

$$Cv = \sigma/\mu = 13.0249/22.7386 \approx 0.5728$$

$$D = 1 - \frac{Cv}{\sqrt{N-1}} = 1 - \frac{0.5728}{\sqrt{7-1}} \approx 1 - \frac{0.5728}{2.4495} \approx 1 - 0.23 = 0.77$$

因此，「使用」一詞之均勻度（D）為 .77。