

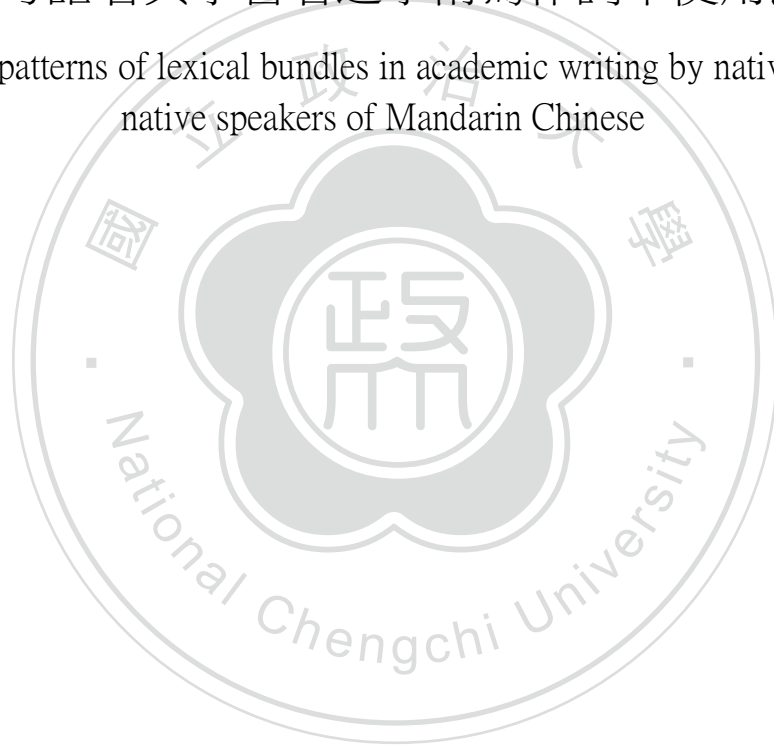
國立政治大學華語文教學碩士學位學程

碩士學位論文

指導老師：張莉萍博士

華語母語者與學習者之學術寫作詞串使用比較

Comparing patterns of lexical bundles in academic writing by native and non-native speakers of Mandarin Chinese



研究生：王莉婷

中華民國 109 年 7 月

## 謝辭

此時此刻還是不敢相信論文真的寫完了。整個碩士生涯就像在「過五關、斬六將」一般，尤其在最後的論文階段，儘管萌生過無數次放棄的念頭、臆想所有中斷研究的可能結果，終究還是把論文完成了。能走到這一步，最要感謝的就是我的指導教授張莉萍教授，不只是老師對研究的專業及嚴謹態度值得學習，生活態度及待人處事的方式更讓人敬佩。有好多次遇到寫作瓶頸，一邊掉淚一邊跟老師通話，又或是想拖到下一個學期在畢業的時候，老師總是溫柔給予支持、提醒我「離終點不遠了」。對老師的感激可能千言萬語也無法準確傳達，謝謝老師一路上推著我前進、更謝謝老師耐心包容學生所有不足。另外，也要感謝神隊友昱靜在半夜花了將近兩個小時教我調整格式、處處提醒我各種細節，讓我這個豬隊友可以跟著妳一起走，我們一起在板橋星巴克寫作的日子可能是我寫作過程中最開心的時候（雖然也可能是最沒效率的一段）。最後還要謝謝汎霓，和妳聊天說八卦吃東西是枯燥日子中的一片綠洲，妳的正能量和生活態度也不時提醒著我要向前進。

寫論文真的是對心智的磨練，謝謝所有陪我度過負面情緒的每個人，尤其是我的姊姊和慧中；謝謝每一個工作機會、每一個學生，讓我在研究所時期可以餵飽自己。總覺得自己當了特別久的學生，尤其在這最後一年半以來，生活中的種種都好深刻，期許自己隨著論文完成，能把負面情緒和事件拋諸腦後、記住自己的成長，以最好的狀態往下一個里程碑前進。

2020 夏

# 華語母語者與學習者之學術寫作詞串使用比較

## 摘要

本研究旨在比較華語母語者與學習者學術寫作詞串之差異。研究採語料庫為本的研究方法，語料來源為臺灣各大學華語教學系所本國籍和外國籍學生撰寫之碩士學位論文各八十篇，共計一百六十篇。本國籍語料共 2,485,285 詞、外國籍語料共 2,104,979 詞。經過文本清理、中文斷詞、抽取詞串等程序，進行詞串使用差異分析。差異顯著性則採 LLR 方法，分析討論學習者顯著多用及少用之詞串；另外將詞串依功能歸類，藉由功能分類探討詞串分布情形。

本研究詞串篩選門檻為每百萬詞 20 次、文本分布率 10%，在功能分類方面採用 Hyland (2008a)、Salazar (2011) 提出的分類，將詞串分為三大類，十五小類，兩語料庫中數量最多的前五個類別相同，分別為：歸類型、比較型、推論型、程序型及描述型。在詞串使用的討論上，本研究有四點發現：第一，學習者的詞串種類和次數都比母語者多，是因為學習者有集中使用特定詞彙或詞串的傾向，反之母語者詞彙較豐富、使用頻次較分散，使詞串之頻次未達篩選門檻。第二，比起母語者，學習者少用連詞和較文言的詞串，如「而」、「所 V 的」結構。第三，與英語詞串研究相比，中英文語言的差異導致詞串分布表現不同，相同意義的詞串在中英文中可能有不同長度，如：英文三詞詞串 a large number of 中文翻譯為兩詞詞串「大量的」，兩詞詞串不在本文討論範圍內。第四，母語者與學習者在學術論文中的指稱詞很不同，母語者偏好使用「我們」、學習者則偏好使用「筆者」或「研究者」。

本研究結果提供常見華語教學領域論文詞串表，供華語學習者撰寫學術論文時參考。

**關鍵詞：**語料庫、詞串、華語教學、詞串功能、學術論文

# 目次

<b>第一章 緒論</b> .....	<b>1</b>
1.1 研究背景及動機 .....	1
1.2 研究目的及問題 .....	2
<b>第二章 文獻探討</b> .....	<b>3</b>
2.1 詞串的定義 .....	3
2.2 詞串與外語學習 .....	9
2.3 學術寫作詞串研究 .....	13
2.4 詞串功能分類 .....	18
<b>第三章 研究方法</b> .....	<b>23</b>
3.1 語料來源與蒐集 .....	23
3.2 語料庫建置 .....	24
3.2.1 文本清理 .....	24
3.2.2 斷詞 .....	28
3.3 語料庫分析工具 .....	29
3.3.1 索引圖 (concordance plot) .....	30
3.3.2 詞表 (word list) .....	30
3.3.3 關鍵詞表 (keyword list) .....	31
3.3.4 詞叢 (Clusters) / N-Gram .....	33
3.4 詞串抽取與統計方法 .....	36
3.5 詞串功能分類步驟 .....	41
<b>第四章 母語者與學習者之用詞與詞串比較</b> .....	<b>44</b>
4.1 關鍵詞表分析 .....	44
4.2 詞串使用分析 .....	49
4.2.1 詞串分布 .....	49
4.2.2 共有詞串分析 .....	52
4.3 功能分析 .....	59
4.3.1 研究導向 .....	61
4.3.1.1 歸類型詞串 .....	61
4.3.1.2 程序型詞串 .....	64
4.3.1.3 描述型詞串 .....	66

4.3.1.4	數量型詞串	68
4.3.2	文本導向	70
4.3.2.1	架構型詞串	70
4.3.2.2	目的型詞串	72
4.3.2.3	引述型詞串	74
4.3.2.4	比較型詞串	76
4.3.2.5	附加型詞串	78
4.3.2.6	推論型詞串	78
4.3.2.7	組織型詞串	80
4.3.2.8	因果型詞串	81
4.3.3	參與導向	82
4.3.3.1	立場型詞串	83
4.3.3.2	吸引型詞串	85
<b>第五章</b>	<b>結論</b>	<b>87</b>
5.1	研究問題的回答	87
5.2	研究限制	93
5.3	未來研究及華語文教學建議	93
<b>參考文獻</b>		<b>95</b>
<b>附錄</b>		<b>100</b>

## 表目錄

表 2-1 詞塊和語塊之比較 .....	5
表 2-2 詞串篩選標準 .....	8
表 2-3 過往研究提及之母語者和學習者詞串差異 .....	12
表 2-4 期刊文章、博士和碩士論文之詞串功能分布 (Hyland, 2008a) .....	15
表 2-5 劉貞好、陳浩然、楊惠媚 (2017) 研究中之詞串數量及舉例 .....	16
表 2-6 劉貞好、陳浩然、楊惠媚 (2017) 研究中之研究導向程序型詞串 .....	17
表 2-7 Cortes (2004) 的詞串功能分類 .....	19
表 2-8 Biber & Barbieri (2007) 之詞串功能分類 .....	20
表 2-9 Hyland (2008a)、Salazar (2011) 提出之詞串功能分類 .....	21
表 2-10 三種詞串功能分類比較 .....	22
表 3-1 母語者語料來源學校分布 .....	23
表 3-2 母語者論文領域分布 .....	23
表 3-3 學習者國籍分布 .....	24
表 3-4 學習者論文領域分布 .....	24
表 3-5 文本清理刪除內容 .....	26
表 3-6 學習者與母語者的詞種、詞數、相異詞比例 .....	31
表 3-7 詞串篩選機制 (筆者整理) .....	36
表 3-8 初步詞串抽取結果 .....	37
表 3-9 頻率調整後詞串抽取結果 .....	37
表 3-10 四詞詞串檢索分類結果 .....	38
表 3-11 五詞詞串檢索分類結果 .....	39
表 3-12 共有四詞詞串 LLR 顯著性分析結果 .....	40
表 3-13 共有五詞詞串 LLR 顯著性分析結果 .....	40
表 3-14 經篩選後之三詞、四詞詞串數量 .....	41
表 3-15 學術詞串功能分類 (Hyland, 2008a, pp.13-14; Salazar, 2011, pp.52-53; 劉貞好、陳浩然、楊惠媚譯, 2017, p.132) .....	42
表 4-1 前 100 個正關鍵詞 .....	44
表 4-2 前 100 個負關鍵詞 .....	45
表 4-3 詞串之種類和頻次 .....	50
表 4-4 母語者及學習者前 20 高頻三詞詞串 .....	51
表 4-5 共有及獨有詞串統計數據 .....	52
表 4-6 L1、L2 共有詞串顯著性分析結果分布 .....	52
表 4-7 學習者多用之詞串 .....	53
表 4-8 學習者少用之詞串 .....	54
表 4-9 「由此可」右邊 1 個詞的檢索結果 .....	56
表 4-10 「在**情況下」**前五高頻搭配詞 .....	57
表 4-11 《當代中文課程》收錄之「而言」解釋 .....	57
表 4-12 各功能類型詞串數量 .....	59
表 4-13 研究導向詞串分布 .....	61

表 4-14 歸類型詞串.....	62
表 4-15 學習者和母語者量詞使用結果.....	63
表 4-16 高頻量詞顯著性比較.....	63
表 4-17 程序型詞串.....	65
表 4-18 「的方式來」之搭配詞彙.....	66
表 4-19 描述型詞串.....	67
表 4-20 「因素」、「部分」左邊搭配詞檢索結果.....	68
表 4-21 數量型詞串.....	68
表 4-22 Cortes (2004) 之數量型詞串.....	69
表 4-23 文本導向詞串分布.....	70
表 4-24 架構型詞串.....	71
表 4-25 劉貞好、陳浩然、楊惠媚(2017) 架構型詞串在本文之分類.....	72
表 4-26 目的型詞串.....	73
表 4-27 引述型詞串.....	75
表 4-28 所 V 的詞串列表.....	76
表 4-29 比較型詞串.....	77
表 4-30 附加型詞串.....	78
表 4-31 推論型詞串.....	79
表 4-32 組織型詞串.....	80
表 4-33 因果型詞串.....	81
表 4-34 母語者與學習者因果型詞串使用頻率.....	82
表 4-35 參與導向詞串分布.....	83
表 4-36 參與導向詞串頻次顯著性檢定.....	83
表 4-37 立場型詞串.....	84
表 4-38 有模糊語的立場型詞串.....	85
表 4-39 吸引型詞串.....	85
表 5-1 母語者及學習者前 14 個共有高頻詞串.....	87
表 5-2 母語者及學習者三、四、五詞共有詞串.....	88
表 5-3 學習者多用及少用的詞串.....	90
表 5-4 母語者及學習者功能分類種類分布比例.....	91



## 圖目錄

圖 3-1 Free my pdf 網站頁面 .....	25
圖 3-2 pdf to doc 網站頁面 .....	25
圖 3-3 Notepad++7.8.3 系統介面.....	27
圖 3-4 段落符號取代前.....	27
圖 3-5 段落符號取代後.....	28
圖 3-6 經斷詞後檔案.....	29
圖 3-7 Antconc3.5.8 (2019) 軟體介面擷取圖.....	29
圖 3-8 Antconc 索引定位功能-檢索「了」為例.....	30
圖 3-9 學習者語料庫前 10 個高頻詞.....	31
圖 3-10 Antconc3.5.8 關鍵詞詞表設定選單.....	32
圖 3-11 學習者負關鍵詞表結果 .....	33
圖 3-12 詞叢功能檢索「而」之結果.....	34
圖 3-13 Antconc 隱藏詞類標記設定介面.....	34
圖 3-14 Antconc N-gram 搜尋設定 (以母語者四詞詞串為例) .....	35
圖 3-15 Antconc N 字元搜尋結果 (以母語者四詞詞串為例) .....	35
圖 3-16 母語者二詞詞串檢索結果.....	37
圖 3-17 「稱之為」 Collocates 檢索結果 .....	43
圖 4-1 Antconc Concordance Plot 檢索「了」結果.....	46
圖 4-2 母語者與學習者詞串分布比例.....	60
圖 4-3 母語者「的目的是」 Collocates 檢索結果.....	74
圖 4-4 學習者「的目的是」 Collocates 檢索結果.....	74



# 第一章 緒論

## 1.1 研究背景及動機

在電腦普及、效能持續提升的環境下，語料庫的發展不斷加速。語料庫蒐集大量語言數據，而文本數位化越趨方便，使語料庫的應用範圍也相對擴展，學者廣泛藉由電腦處理及分析語言，將結果應用在質化及量化研究，提供了分析語言一個不同角度。語料庫研究的先驅代表為 1970 年英國伯明罕大學建立的大型電腦化語料庫（Collins and Birmingham University International Language Database, COBUILD），研究者利用這項計畫的語料讓一系列新派的文法書、辭典、語言教材相繼出版（張俊盛，2017），以 COBUILD 為本而出版的 Collins COBUILD 英語學習辭典（1995）詳細記載了不同詞性的文法規則，進一步提出模式語法（Pattern Grammar）理論 (Francis, Hunston, & Manning, 1996) 這一套詞彙化的文法理論讓詞彙語法開始受到學界的關注。

自 COBUILD 開始發展，其他語料庫的建設及相關研究也不斷增加，在這些研究中，學者們開始關注詞彙和頻率訊息的特殊性（梁茂成、李文中、許家金，2010），詞串（Lexical bundles）的概念就是在這樣的背景下被提出的。詞串的概念首先在 1999 年由美國學者 Douglas Biber 和同僚在 Longman Grammar of spoken and written English 一書提出，其定義為「在一語體（register）中，最常見並反覆出現的詞組排列」。不同領域的研究證實了詞串存在於語言中，心理語言學實驗發現母語者對於詞串的敏感度很高閱讀詞串的反應處理速度快 (Ellis, Simpson-Vlach, & Maynard, 2008) 進一步證實了這些序列都被儲存在人的心理詞彙中，而且這些和我們日常生活的語言有密不可分的關係 (Wray, 2002)。第二外語教學研究也提出，熟悉詞串不只能使學習者提升語言處理效率 (Simpson-Vlach & Ellis, 2010)，也能使學習者的措辭更接近母語者的使用習慣 (Wray & Perkins, 2000)。上述研究都指出詞串在語言中的重要性，並說明學習第二外語時使用詞串的益處，這些研究對語言教學的助益使更多學者投入詞串研究，欲探究更多詞串的特色和功能，並進一步運用在語言教學中。

然而，不同語體的詞串性質也不同，這一點藉由語料庫研究近二十年來的興起獲得證實。Biber, Johansson, Leech, Conrad & Finegant (1999) 透過分析語料

大小超過四千萬詞朗文口語與書面語語料庫 (Longman Spoken and Written English Corpus, LSWE)，他們的研究列舉了高頻口語和寫作詞串，指出口語和學術詞串有相當大的差異，並進一步指出中高級的二語學習者而言或許能在口語交流上能流利應對，但也不一定能夠寫出通順的學術文章，而這點似乎與華語學習者的狀況不謀而合。信世昌 (2001) 在研究中提及儘管許多國際學生的口語已達流利，但仍無法寫出完整、合宜的文章，造成語言偏枯的困境。一般寫作對學習者而言已經不容易，學術寫作的難度層次更是提高。英語為第二外語的領域中學術寫作課程 (English Writing for Academic Purposes, EAP) 行之有年，透過分析架構、提供學術詞彙表及句型結構能幫助學生撰寫學術文章，EAP 教學領域的研究為數不少，其中也不乏學術詞串在 EAP 的應用及比較，而面對學習華語之人數成長的趨勢，推行學術寫作課程也勢必存在其必要性。

由於寫作需要結合不同的認知功能，另外還牽涉繁複的領域專業知識，學術寫作需統和許多技能，這讓外籍生更多了語法、語用與文化方面的偏誤 (彭妮絲, 2013)，對外籍生而言要寫出文意通順、近似母語者風格的文章更是不容易。以英語為第二外語的學術寫作研究、課程發展至今已有一定的成果，然而，華語方面相關的研究為數不多。因此，本研究以比較母語者及學習者使用詞串的角度，提供研究成果作為華語學術寫作課程及教材之參考。

## 1.2 研究目的及問題

英語 EAP 研究與課程已有一定的發展，相較之下，華語教學相關研究不多，坊間相關課程及教材也相當有限。本研究以研究華語學術寫作為初衷，欲協助華語學習者了解母語者的學術寫作特點，進一步撰寫流暢且接近母語者寫作風格的學術文章；同時，也欲將研究結果提供未來學術華語課程及教材發展作為參考。為達到上述目的，本研究蒐集八十篇分別自臺灣華語文教學研究所之華語母語者及學習者的碩士畢業論文作為語料，以自行建置語料庫的方式，篩選出論文中常見詞串，以分析母語者及學習者使用之異同，讓學習者及未來教材有參考依據。透過語料庫分析，本研究探討的問題有以下三點：

- 1) 學習者學術詞串是否與母語者學術詞串的表現相同？若否，差異為何？
- 2) 華語學術詞串的功能分類分布情況為何？
- 3) 外語學術詞串研究的成果與華語的學術詞串特徵的異同為何？

## 第二章 文獻探討

本研究欲探討母語者和非母語者撰寫學術文章時詞串使用之異同，本章分為四節。第一節說明詞串的概念和由來，檢視不同流派如何看待詞串，並討論詞串的定義與篩選標準。第二節為了討論詞串與二語習得的關係，我們將從心理語言學的觀點切入，討論詞串使用對認知及閱讀反應的影響。第三節提出詞串在不同語體中的變異，探討學術文章中詞串的特性。最後，第四節比較前人對詞串功能的不同分類方式，並提出本研究分類的方法。

### 2.1 詞串的定義

傳統的語言學研究常以「詞」為單位進行研究，但是自二十世紀末開始，語言學家開始觀察到詞串在語言中的價值，而語料庫普及與詞串研究興起息息相關。語料庫語言學家 Sinclair 藉由觀察大量的語料提出和傳統語言學截然不同的觀點，他提出了「熟語原則」(principle of idiom) 支持詞串的存在 (Sinclair, 1991)，並認為詞串在語言組成中扮演的角色比過往所認知的重要得多。熟語原則是指語言使用者在記憶中儲存了很多半預組 (semi-constructed) 且固定的組塊，儘管這些組塊在結構上能被切分為更小的單位，但是整個組塊被視為整體，使用時內部元素不會更動，語言使用者在使用這些組合詞時也不需要多加思索，這些序列不如熟語一般有固定形式，而是由共現頻率高的詞彙組合而成。例如，英語中”in other words”是一個常見的多詞組合，由”in”、”other”和”words”三個詞組成，三個詞能夠分開解讀，但是說話者說”in other words”時，不是從記憶中分別提取三個詞彙，而是整個序列直接提取。功能語法學派也提出了相似的概念，他們認為語言使用者可能會在無意識或有意識間選擇特定的詞彙，使詞串中的詞彙相互連結，而創造出詞彙銜接 (lexical cohesion) 的現象。心理語言學家藉由實驗發現，語言使用者對於一些詞串比較敏感，當他們閱讀特定組合時，會有閱讀反應速度較快、注意力較高 (Ellis, Simpson-Vlach, & Maynard, 2008) 的特徵。實驗結果指出語言似乎非以詞彙為單位儲存在我們的記憶中，而是以一連串的序列 (sequence) 為單位儲存。語料庫語言學家 Biber 和同僚在 1999 年透過朗文口語與書面語語料庫 (LSWE) 觀察口語英語和寫作英語的差異，他們以詞串 (lexical bundles) 一詞稱呼多詞組合，該研究提出

後，往後的文獻也大都沿用「詞串」作為術語，而本研究也以語料庫作為媒介，觀察母語和二語學習者論文寫作的差異，故以「詞串」稱之。

Douglas Biber 和同僚對詞串的定義描述如下：

詞串是一種反覆出現的表達方式，而慣用的表達方式和結構不在詞串的討論範圍內，簡單來說，詞串即是在自然言談中時常會一同出現的一串詞。

(p.990)<sup>1</sup>

*(Lexical bundles are recurrent expressions, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse. )*

詞串不僅存在於自然言談中，過往研究也指出詞串在書寫和閱讀文本中佔有一席之地。Altenberg (1998) 分析了 London-Lund 口語語料庫後發現，高達 80% 的資料是由重複的詞串所組成的，也有研究指出使用詞串的流暢度為二語學習者發展的重要指標 (Staples, Egbert, Biber, & McClair, 2013) 可見詞串對於閱讀和寫作都相當關鍵。

上述的文獻皆證實詞串在各方面的重要性，儘管至目前為止已有不少詞串的相關研究，但學者們對於「詞串」的基本定義和特色卻一直沒有定論。過往的詞串研究的共同點為皆是以語料庫為本的研究，研究者藉由語料庫篩選出共現率高的詞彙組合。然而，詞串的篩選頻率和文本分布率的設定卻不盡相同，就如同 Wray (2008) 所言：「要給詞串明確的定義就像是在一間全黑的房間內找一隻黑貓一樣困難。」為何定義詞串是一項艱難的任務，Biber, Conrad & Cortes (2004) 歸納出了五個原因解釋：

- (1) 研究目的的不同。
- (2) 提取詞串的方式不同。
- (3) 定義詞串的準則不同。
- (4) 使用之語料庫的大小。
- (5) 是否比較兩個不同的語體。

---

<sup>1</sup> 譯文為筆者自行翻譯



以上五點互相有關聯。首先，因為不同研究之研究目的不同，研究者對詞串的定義就會有出入，目前詞串研究之目的大約能分為幾種類型：（1）比較兩個語體（如：比較口語跟書面語詞串）（2）在相同語體中比較母語者和非母語者使用的詞串（3）在相同語體中比較生手（例：學生）和專家使用的詞串（4）歸納某一學科中或是語體中的常見詞串。因為研究目的不同，語料庫的選擇、大小和數量會不同，同時也造成語料庫收錄的文本數量和字數的差異。在語料庫大小不同的前提下，提取詞串的條件自然也就不同。儘管各研究之研究目標、範圍以及語料來源之差異性造成了詞串很難有一致的定義，但是核心定義皆為共現率高的字詞，也都是屬於詞組學（phraseology）的討論範圍 (Gries, 2008)。

同屬於詞組學的一類，研究經常將「語塊」（chunk）和「詞串」相提並論，但是在細節的定義卻有出入。「語塊」（chunk）源起於心理學家 Miller (1956) 提出的  $7 \pm 2$  記憶長度組塊（chunk），認知和心理語言學家也將這個概念延伸至第一語和第二語的習得，如 Ellis (2001) 的研究中，即是以「語塊」討論二語學習者記憶詞彙搭配組合的方式，這個記憶長度讓學習者在詞彙搭配、語音、拼字、句法方面奠定了習得的基礎。中國學者蘇丹潔、陸簡明 (2010) 提到「語塊」指的是「一個構式中以一定的句法形式相對獨立地承載構式的一個語義單元的句法語義聚合體」。也就是說，一個語塊僅含有一個語義單位；而詞串則是含有數個語義單位。朱金平、賈益民 (2009) 將「lexical formulae」翻作「詞塊」，並列出「語塊」和「詞塊」在結構、語義和整體功能性上的不同，筆者以下表整理之：

表 二-1 詞塊和語塊之比較

名稱	結構	語義	整體功能
詞塊 Lexical formulae	能自由搭配	構成成分意義相加	有明顯語法功能
語塊 chunk	固定	可能具有引申意義	不具有明顯的語法功能

「詞塊」與本研究討論之「詞串」的概念較相似，相較之下，「語塊」討論的範圍包含了熟語。儘管詞塊和語塊的定義有差異，不過兩者仍有重疊處，

重疊處稱為「半凝固結構」，半凝固結構指的是具有鮮明口語色彩的結構，如：一不小心、老實說等（朱金平、賈益民，2009）。「詞塊」和「語塊」的比較也說明了詞組學子領域內重疊和不可切割的部分，這也是為什麼提出完整定義不容易的原因之一。

除了上述的語塊和詞塊以外，在過往的研究中，我們能看到其他學者以不同的術語稱呼「共現率高的多詞組合」的概念。例如：Altenberg (1998) 以英語口語語料庫 (London-Lund corpus) 為本的調查研究中，將此種序列組合稱為「重複出現的字詞組合」(recurrent word-combinations)，定義為連續的三個詞並至少在語料庫中出現兩次；De Cock et al. (1998) 比較母語和非母語者的面談語料研究中沿用相同的標準，在其研究中除了三字詞的組合以外，也列出兩字、四字和五字的字詞組合；Ellis et al. (2008) 以心理語言學、語料庫語言學、英語為第二外語教學三方面，比較母語者和非母語者使用詞串習慣的異同；Hyland & Milton (1997) 在母語以及學習者的寫作研究中，將共現字詞稱之為 "clusters"；O'Keeffe, McCarthy & Carter (2007) 則在其口語語料庫分析研究中以 "chunks" 稱之。以上文獻是其中較常見的稱呼，另外曾被使用過的術語包含：lexical phrase (Nattinger, 1988)、fixed expressions (Moon, 1992)、cluster (Hyland & Milton, 1997)、lexical bundles (Biber et al., 1999)、recurrent word combinations (De Cock, Granger, Leech, & McEnery, 1998) formulaic language/sequence (Wray, 2002)、chunk (Ellis, 2001)、formulaic expressions (Simpson-Vlach, 2004)、n-grams (Stubbs, 2007) 和 phraseology (Gries, 2008)。雖然核心定義相同，但是因學門和研究目的差異而使用不同術語作為稱呼，筆者整理了以上提及的文獻中所探討的多詞組合，以類型分類，大約能分為四類：

- (1) 慣用語 (idiom) - 序列中詞彙個別的意義和整個序列的意義完全不同。  
(如：kick the bucket)
- (2) 片語 (phrase) - 序列中詞彙個別的意義和整個序列的意義不完全相同。  
(如：in order to)
- (3) 高頻搭配的詞彙 (collocation) - 序列無引申義，但是序列中的詞彙共現率很高。  
(如：to avoid confusion)

(4) 詞串 (lexical bundles) — 共現率高的詞彙組合，非片語也非慣用語，但序列中的詞彙共現率高，語言使用者常從記憶中提取整個組塊。

(如：been seen as a)

慣用語的範例如英文中的「kick the bucket」，照字面翻譯成中文為「踢桶子」，但是實際上是「去世」的意思，「去世」和「踢桶子」照字面意義來看兩個意思完全不同，而華語的慣用語可能是一個詞組或是由短語而組成，如：「吃軟飯」、「潑冷水」、「雞蛋裡跳骨頭」等。第二個種類是片語，大部分的片語是定式結構，且在句中有重要的功能，如英文中的「in order to」表「為了」、「as a result of」表結果，中文的例子如：「從...到....」、「綜上所述」。第三類高頻搭配的詞彙是指共現率高的詞彙，閱讀實驗中常常會以這類的序列測試閱讀時間長短，語言使用者處理共現頻率高的詞彙組合的速度較快。例如，學習者能夠較快閱讀且理解 to avoid confusion 的意思，但是如果將 confusion 換成不常和 avoid 共現的 discovery，學習者可能就得花較多的時間處理這個組合 (Ellis, Simpson-Vlach, & Maynard, 2008)。這類型的搭配組合也時常應用在克漏字考試上。而最後一個分類—詞串，即是本研究欲探討的類別。

綜合前段所述，詞串的特色包含：高頻、非熟語、不一定完整的結構單位（如：I don't know if, I just wanted to）、在篇章和言談中扮演了相當重要的功能角色 (Biber, Conrad, & Cortes, 2004)，而在詞串研究中最關鍵的要素為頻率標準，詞串篩選之頻率因研究而異，下表整理幾篇研究目的不同的詞串研究之篩選標準：



表二-2 詞串篩選標準

作者	語料量 (詞)	詞串出現頻次	文本分布率
Biber et al. (1999)	40,000,000	10 次/一百萬詞	出現在五個文本以上
Biber et al. (2004)	A. 1,248,800 B. 760,600	40 次/一百萬詞	出現在五個文本以上
Hyland (2008a)	A. 632,500 B. 794,100 C. 844,400 D. 1,129,400	20 次/一百萬詞	10% 的文本
Chen & Baker (2010)	A. 164,172 B. 155,781 C. 146,872	25 次/一百萬詞	出現在四個文本以上
劉貞妤、陳浩然、 楊惠媚 (2017)	9,000,000	10 次/一百萬詞	5% 的文本/2% 的文本

Biber et al. (1999) 的研究為詞串研究的先驅之一，他們使用的語料庫為 LSWE，語料性質廣泛，從對話、小說至報紙和學術文章一應俱全。該研究討論口語和學術詞串之差異，所選之語料共計四千萬詞，對於三、四詞詞串篩選頻率為每一百萬詞出現 10 次，初步共篩選出約 6500 組三、四詞詞串。Biber, Conrad & Cortes (2004) 比較大學授課的口語詞串和教科書中的學術詞串之差異，從約兩百萬詞的語料庫中，以每一百萬詞出現 40 次篩選詞串，且每個詞串至少要出現在五個不同的文本中，在這樣的標準下他們得到共約 110 個詞串；Hyland (2008a) 欲比較不同學科在學術文章中使用的詞串，他以每一百萬詞 20 次，且單一詞串需出現在 10% 的文本中作為篩選標準，在不同學科的語料中平均篩選出 150 個四詞詞串。Chen & Baker (2010) 比較專家、母語者和學習者的學術寫作詞串，平均語料量為 15 萬詞，他們以每一百萬詞 25 次、單一詞串至少出現在四個文本以上的頻率篩選出約 110 個詞串；劉貞妤、陳浩然、楊惠媚 (2017) 研究中文人文社會科學常用詞串，從 900 萬詞中以每一百萬詞 10 次、文本分布率 2% 和 5% 篩選出三、四詞詞串共 200 個。從以上研究我們能發現研究者會為了符合自己的研究目的會設定篩選機制，為避免詞串數量大到人力無法負荷以及分析的地步，不論語料量為何，各研究最終篩選出詞串的數量約為 100-200 個。

相較其他語言學研究，詞串仍屬於較新的領域之一，儘管目前英語詞串研究中已經能看到母語者、學習者的詞串比較以及不同語體、領域的詞串研究，但中文詞串研究仍相當有限。英文和中文在語言本質上差異不小，因此中文和英文的詞串研究可能會呈現截然不同的結果，上述的討論皆以英語詞串為主，中文詞串的特徵可能會與上述文獻的敘述有所出入，因此，探討中文詞串的特色也是本研究的研究問題之一。

## 2.2 詞串與外語學習

從日常生活中，我們似乎能夠察覺有時候我們會不加思索地說出一些詞組組合，不論是成語、熟語（如：皇天不負苦心人、解鈴尚需繫鈴人）或是固定形式的詞組組合（如：一般來說、由此可見）相關的例子不勝枚舉，這點引起了心理語言學家的注意。傳統外語教學核心以「詞」作為學習單位，但是在電腦逐漸普及、語料庫數量逐漸增加以後，學者們觀察到一些詞組在語言中共現的頻率極高，因此針對共現頻率高的詞組之研究便越來越多。

為了要觀察語言儲存在大腦的單位，心理語言學家 Pawley 和 Syder (1983) 發現語言使用者有一定的認知限制 (cognitive limitation)，認知限制讓語言使用者很難在有知覺的情況下產出多於 8-10 個詞以上的短語 (clause)，不過語言使用者卻能夠很自然地說出 "You can lead the horse to water, but you can't make him drink" 這類組成超過 10 個詞以上的句子。因此，他們認為大腦並非以詞為單位儲存在記憶中，而是以詞組 (lexical phrases) 作為單位儲存在工作記憶中。母語者能夠自然地產出語言的原因是母語者相當熟悉這些詞組，以整個詞組作為單位從記憶中提取；相較之下，學習者輸出語言時會有不自然的停頓和語言使用不自然的現象，即是因為暴露在外語的時間不足的關係。由於學習者對於語言不夠熟悉，因此無法進行整個詞串單位的提取，而需要自行組織、配對詞彙 (Schmitt & Carter, 2000; Wray, 2002)，也因此會造成語言不自然的狀況。

不少心理語言學實驗證實詞串和閱讀理解有相當重要的連結。過往研究指出，不論母語者或是學習者，都能以較快的速度處理含有慣用詞串的句子 (Conklin & Schmitt, 2008)。Conklin & Schmitt (2008) 以閱讀篇章測試母語者和

學習者閱讀時對慣用詞串的反應速度，研究結果指出母語者處理慣用詞串的速度比處理非慣用型快得多。Ellis et al. (2008) 藉由判斷詞串以及兩個朗讀詞串的實驗，分別紀錄母語者與學習者的判斷和閱讀時間，結果發現，母語者和學習者都能以較快的速度處理共現率高的詞串組合。上述兩個研究的實驗順序皆為先篩選出高頻詞串，再要求學習者進行實驗任務，然而，由於實驗中所套用的詞串為事先設計好的，這些詞串對於學習者而言可能不是熟悉的詞串，因此降低了反應時間。不過，De Cock (2004) 以收集自然言談的方式，分析母語者和學習者的口語語料，也得到了相似的結果。在研究中，De Cock (2004) 以訪談的方式蒐集 50 位母語以及非母語者的口語語料，比較兩群體的詞串使用習慣。在扣除無意義重複以及語助詞後，母語者使用詞串的次數和種類皆高於學習者。除此之外，在研究中也發現儘管是進階級的學習者，詞串使用仍常發生偏誤。上述的實驗證實了不論是在有控制的情況下，還是在自然語境中，學習者對於詞串的掌握仍有進步的空間，甚至是學習英語超過十年的高級學習者，在實驗表現上仍與母語者有差距 (Ellis et al., 2008)，這個結論也凸顯了學習者學習詞串之必要性。

過往研究不但證實了詞串使用能改善學習者的口說能力、加速語言處理速度 (Simpson-Vlach & Ellis, 2010)，也能縮短學習者搜尋適合用語的時間 (Wray & Perkins, 2000)，讓對談和寫作更有效率，但最重要的是，這些研究也不約而同地指出學習者和母語者在詞串使用上仍有一定的差距。為了分辨學習者和母語者使用詞串之差異，學者們開始比較兩個群體以及不同程度學習者的語料。由於詞串在不同語體和學科之間有差異，以下討論三個針對不同語體或學科的研究。

首先討論學習者和母語者在自然情境下口語對談的詞串使用狀況。De Cock (2004) 對 50 位英語學習者和 50 位英語母語者進行約 15 分鐘非正式的開放訪談，訪談的內容大多為大學生活、興趣、未來海外旅行的計畫等與日常生活相關的內容。結果發現，學習者使用的詞串包含大量表猶豫的重複或停頓的語助詞，使用的頻率為母語者的三至四倍，將這些詞串刪除後，母語者使用的詞次 (tokens) 和詞種 (types) 皆較學習者多。從結果得知，絕大多數的學習者以外語表達時仍會表現出不確定的情緒，可能是因為他們需要時間找合適的詞

彙，也可能是對他們而言很難以外語表達自己的想法。在模糊標記 (Markers of Vagueness) 的使用上母語者與學習者也有相當不同的表現，在非正式的言談互動中正確地使用模糊標記 (如：or something、or anything) 不只能讓言談自然、也有拉近和說話者和聽話者距離的作用，母語者使用模糊標記的次數是學習者的兩倍；相反地，學習者顯著地少用和誤用模糊標記，也較常使用正式場合才會使用的模糊標記 (如：for example、for instance)，以上兩個都可能是學習者在對話中表達不如母語者自然的因素。

目前關於寫作的學術詞串研究有兩個取向：一是支持詞串使用和寫作的流暢度有正向關係，如 Chen & Baker (2010) 以三個語料庫分析母語新手、非母語新手以及母語專家之學術寫作詞串，經過分析後發現，越有經驗的作者 (母語專家) 使用的詞串數越多，而非母語的研究生所使用的詞串數量則最少。另外，也有學者發現學習者使用詞串的頻率和語言的流利度有關，程度越好的學習者會使用較多的詞串 (Staples et al., 2013)；但另外一個取向反而認為，越是沒有自信、沒經驗的作者才會在文章中依靠詞串來堆砌內容 (Hyland, 2008a)。不過以上三個研究詞串處理的過程不同，Staples et al. (2013) 將篩選出的詞串分為主題相關 (如：in Hong Kong) 和非主題相關再分別分析；而 Chen & Baker (2010) 以及 Hyland (2008a) 則是排除主題相關詞串，再進行統計。此外，以上三份研究的語料來源也不同，Chen & Baker (2010) 研究之語料來源為學術語料庫，含有不同領域學科的內容；而 Staples et al. (2013) 則是針對 TOEFL 考試學習者的寫作文本進行詞串抽取；Hyland (2008a) 的語料包含了四個學門的論文，這些差異都有可能導致研究結果不同。如同上一節所提及的，詞串的定義尚無定論，研究者會依研究目的和需求自行決定詞串的頻率和篩選門檻，因此在研究結果上可能會出現極端差異。而另一個造成研究結果差異的原因可能是由於語料量的大小不同，語料量越大產出的詞串較多。雖然研究結論不同，但皆指出母語者和非母語者之學術寫作模式仍有差距，筆者將上述文獻整理成表 2-3 進行比較：



表 二-3 過往研究提及之母語者和學習者詞串差異

	母語者	學習者
Hyland & Melton (1997)	適當使用模糊語 (hedging)，能以較間接的方式表達觀點	對模糊語的使用不熟悉，使用數量較少。
De Cock (2004)	使用較多的詞串。	詞串使用會出現誤用、多用、少用的偏誤。
Chen & Baker (2010)	專家使用較多 NP 結構的詞串，並使用較多參考標記。	使用詞串較單一化，容易出現泛化的狀況，但共現詞的數量隨學習者流暢度而上升。
Staples et al. (2013)	指稱詞串 <sup>2</sup> 在使用詞串中佔多數。	在學術文章中使用的詞串多為母語者口說常見的詞串，使用較少指稱詞串。
Amirian, Ketabi & Eshaghi (2013)	包含較多文本導向的詞串，詞串種類也較多元。	包含較多研究導向的詞串。

從上表整理中我們可以發現學習者使用詞串時有以下特點：

- (1) 使用之詞串種類較少、不如母語者的詞串多樣化。
- (2) 可能出現多用和誤用詞串的情形。
- (3) 常將口語常見詞串用於書面寫作中。
- (4) 可能有泛化或過度使用詞串的狀況。

第一點提及學習者使用的詞串種類較少，不如母語者的詞串多樣化，可能與學習者的詞彙量以及教科書的內容有關，學習者熟悉的詞串多來自教科書，所以詞串的多樣性相對較低、較難有創新性。第二點，學習者可能會因為對於片語詞串的理解不夠透澈，因而出現多用及誤用的情形。第三點，學習者在學術文章中使用較多口語常見的詞串，原因可能為學習者對於學術用語還不夠熟悉，也彰顯了學術寫作課程和學術詞串納入教學的重要性。第四點是學習者使用詞串時會出現泛化或過度使用的狀況，可能由於學習者掌握的詞串有限，但

<sup>2</sup> 指稱詞串、文本導向、研究導向皆為詞串之功能分類，於本文 2.4 詳述。

為了達成篇章轉承的目的大量地使用他們已知的詞串，使用時卻未考慮合宜與否。

詞串使用與第二外語的流利度息息相關，透過本節與 2.1 的討論，我們也發現詞串會因語體與文體不同而產生變化，在寫作上使用適當的詞串能夠讓文章更合宜及通順，辨識與使用詞串的能力對學習者而言就格外重要。因此，本文下一節討論學術寫作詞串，從過往研究觀察並整理中、英文學術寫作詞串之特徵。

## 2.3 學術寫作詞串研究

由於英語為學術研究出版品的國際性語言，英語學術寫作課程的必要性和重要性也就不言而喻。EAP 課程的目標為教導所有程度、各年齡學習者學術英語，這使 EAP 課程帶來了一連串跨學科的影響 (Hyland & Hamp-Lyons, 2002, p.3)，而語料庫以及詞串在此領域中受到不少學者的關注。透過語料庫分析能夠呈現詞串在不同語體、文本和學科中的特色，英語教學專家和語料庫語言學家就藉由語料庫分析學術英語和口語英語的不同。Bibier et al. (1999) 使用朗文口語與書面語語料庫 (LSWE) 研究口語和寫作詞串的差異，LSWE 包含四個語體：新聞、學術文章、口語溝通和科幻小說，語料大小超過四千萬詞。透過分析 LSWE，Bibier et al. (1999) 列舉了高頻口語和寫作詞串，並以結構分類詞串再加以討論，最後他們對學術及口語詞串的差異做了以下結論：

- (1) 詞串在兩個語體中有很不同的表現。
- (2) 詞串在口語語料出現的頻率較學術語料高。在口語語料中，包含三個詞和四個詞的詞串共有約 3500 個；在學術語料中則共有約 3000 個。
- (3) 儘管口語語料篩選出較多的詞串，但是在兩語體中詞串出現的頻率無顯著差異。
- (4) 在兩個語體中，最常見的詞串長度為三詞詞串，其出現頻率較四詞詞串多了將近十倍，而四詞詞串及五詞詞串之關係亦是如此。
- (5) 在學術文章中，超過 60% 的詞串由部分的名詞片語或介係詞片語組成；然而，口語語料中超過 90% 的詞串由部分的動詞片語而組成。

從 Biber 等人的研究中得知學術詞串與口語常用詞串有很大的差異，這個研究結果使許多研究者皆對詞串研究產生興趣，此後研究關注不同背景之作  
者、不同領域及不同語體比較的研究便相繼而起。為了幫助學習者增加學術寫  
作的流暢度，EAP 研究比較學習者和母語者所撰寫之學術文章的差異，主要提  
及的差異有三：首先，就寫作風格而言，研究指出學習者撰文時容易有冗贅的  
現象 (Lorenz, 1998)，且表達研究觀點時會有過於主觀、果斷的傾向，造成這個  
主要的原因為學習者對謹慎語言 (cautious language) 的掌握度較低 (Hyland &  
Milton, 1997)。第二點為詞串使用頻率的不同，不少研究發現母語者常在文章  
中大量使用學術詞串，而學習者使用的學術詞串相較不足 (Ädel & Erman, 2012;  
Chen & Baker, 2010; Salazar, 2011)。第三點則是使用詞串之分布，在一些詞串  
研究中學習者使用詞串的數量雖然高於母語者，但卻不如母語者多樣化 (Ädel &  
Erman, 2012)。也就是說，學習者傾向使用特定的詞串，而母語者使用的學術  
詞串則是包含了不同類型，也較有創造力。上述的三點差異說明了母語者和學  
習者的學術文章的不同，也能從之推測詞串使用對二語學習者而言，在學術文  
章的流暢度和準確度的確皆佔有一席之地，若能夠掌握學術詞串，不止能更輕  
鬆地提取所需的語言、增加學術寫作的流暢度和專業性 (Simpson- Vlach & Ellis,  
2010; Chen & Baker, 2010; Hyland, 2008a)、也讓學習者在用字上更接近母語者  
的語用習慣 (Salazar, 2011)。

一些學者藉由分析母語者使用詞串的習慣提出建議，讓學習者的寫作風格  
更趨向母語者；其他學者則持不同的看法，他們認為學術寫作流利度與母語背  
景無關，而是與學術寫作的經驗相關。例如 Hyland (2008a) 的研究，該研究取  
得碩士及博士論文和已出版之期刊研究，首先篩選出三種不同專業度的學術文  
章中的高頻詞串，再進一步以功能角度將詞串歸類，藉以觀察不同種類的詞串  
在不同群組的分布狀況。結果發現，詞串在這三種不同專業度的論文有很不同  
的表現。首先，碩士學生使用的詞串數量最多、專家組的數量則是最少，  
Hyland (2008a) 認為這代表新手研究生會依賴使用詞串來組織其學術文章，而此  
現象也可以在二語學習者撰寫的學術文章中發現。另一點是專家和學生使用的  
詞串類型也有明顯差異。如下表所呈現，碩士學生使用研究導向詞串的頻率相  
當高，專家則是偏向使用文本導向的詞串居多。



表 二-4 期刊文章、博士和碩士論文之詞串功能分布 (Hyland, 2008a)

類型	研究導向	文本導向	參與導向	總和
期刊論文	25.5%	60.3%	14.2%	100%
博士論文	34.1%	54.7%	11.2%	100%
碩士論文	48.6%	42.5%	8.9%	100%
總和	36.1%	52.5%	11.4%	100%

對於詞串類型分布的差異，Hyland 認為是新手研究者將重點放在研究方法，故較頻繁使用研究導向的詞串。Staples et al. (2013) 在研究中也提及新手的學術文章中會使用較多定式詞串，反之專家則會使用較多內部結構能改變的詞串，能夠創造出較多的語言表達。從詞串的功能分布來看，可以看出新手和專家的寫作風格相當不同，透過詞串的分析能夠得到具體的差異，幫助新手研究者增加學術寫作的流暢度，對二語學習者在撰寫學術文章上更是一大助益 (Simpson-Vlach & Ellis, 2010)。

新手的學術文章中出現詞串頻率較高這個現象也在另外一個研究中也得到相似的結論。Staples et al. (2013) 將學習者以托福成績分成三組不同的流利程度，並分析他們的托福考試寫作文章，他們發現托福成績較低的組別使用詞串的頻率反而較高，而且也發現這組學習者使用的詞串多為立即的 (prompt-based)，也就是獨創力較低、和文本主題相關的詞串。他們認為高級學習者已經開始發展自己的語言，不像初中級的學習者那樣依賴現有的詞串，所以使用詞串的頻率不如初中級學習者多；不過，對於低程度學習者而言，詞串的學習是必要的。

不同研究對新手和學習者使用詞串的特徵有不同的結論，主要仍是因為各研究對於詞串定義以及篩選機制之差異所導致。不過，詞串研究皆指出藉由學習詞串學習者或是新手都能增加其學術寫作的流暢性和正確性，也建議將詞串教學納入 EAP 課程大綱中 (e.g., O’Keeffe, McCarthy, & Carter, 2007; Wray, 2000)。

大多數詞串研究仍是以英語為主，中文詞串的研究為數不多。目前針對中文學術詞串的研究僅有由劉貞好、陳浩然、楊惠媚 (2017) 的中文人文社會科

學論文常見詞串研究一文。該研究參考臺灣人文及社會科學引文索引核心期刊和臺灣社會科學引文索引所收錄的期刊，將學門分為十大類，過濾無效電子資源及非中文撰寫的期刊後，篩選出 106 本期刊作為語料來源，最後在十大類期刊內隨機挑選 100 篇文獻，共計約 900 萬字作為語料來源。該研究之詞串篩選標準為每一百萬字至少出現十次、以及文本分布率 5%（三字詞串）和 3%（四字詞串）作為初步篩選標準，再以人工方式排除斷詞錯誤（如：了一種）和含特定領域用語（如：意識形態的）的詞串，最後共篩選出 300 組詞串。下表整理劉貞好、陳浩然、楊惠媚（2017）最終篩選出的詞串數量分布和範例：

表 二-5 劉貞好、陳浩然、楊惠媚（2017）研究中之詞串數量及舉例

詞串分類	三／四字詞串	總數	舉例
研究導向	三字詞串	63	什麼樣的
	四字詞串	33	本研究是以
文本導向	三字詞串	134	主要是因為
	四字詞串	78	所造成的影響
參與導向	三字詞串	28	研究結果顯示
	四字詞串	24	本研究結果發現

根據此統計結果由三個面向說明中文學術論文的特點：（1）使用詞串的原因多數是為了組織文本大綱、連接文本內容以增加文本之連貫性和一致性，代表中文學術寫作中相對重視文本的組織架構。（2）中文學術文章中描述研究程序和方法時，常加入詞串輔助敘述。（3）中文寫作中可能較少出現作者的主觀闡述，多以客觀的敘述為主。該研究以詞串功能分析中文學術文章架構，不只能讓外籍生和華語教師更清楚學術中文之特色，也能夠直接應用該文所篩選出的詞串使其文章詞彙更加豐富及專業，對於學術寫作實為一大助益。

但值得一提的是，從該文獻之結果也能觀察到中文詞串和英文詞串有極大的差異。英文詞串研究中篩選出的詞串能依結構分為「PP型」、「NP型」和「VP型」三大類並進一步分析討論，但是反觀所篩選出的中文詞串發現，並非所有詞串都能進行結構分類。然而，該研究中的詞串似乎能夠進行更進一步的

分類整理。舉例來說，劉貞好、陳浩然、楊惠媚（2017）研究中所列之研究導向程序類別的三字詞串有以下29筆：

表 二-6 劉貞好、陳浩然、楊惠媚（2017）研究中之研究導向程序型詞串

研究導向-程序類別-三字詞				
的過程中	所產生的	本研究以	本研究採用	的方式來
所使用的	為研究對象	所形成的	分析本研究	的相關研究
所得到的	所獲得的	建立一個	所具有的	的實證研究
所建構的	形成一個	所採取的	的方式進行	所產生之
本研究使用	的互動關係	之間的互動	所擁有的	的目的是
發展過程中	的研究方法	形成一種	所採用的	

這29個詞串無法以英語研究中的三個結構種類歸類，然而，我們能仍從中能觀察到一些定式結構，以及常與之出現的搭配詞。如結構「所V的／之」的搭配詞可能有：產生、使用、形成、獲得、得到、具有、建構、採取、產生、採用、擁有。若能夠再進行更深入的分析整理，能夠整理出常見定式和搭配詞，並提供學習者以及未來學術華語教材編纂作為更清楚的參考。

分析學術寫作詞串有必要性的原因，是因為學術文章和口語對談使用的詞串有相當大的差異，若撰寫學術文章時使用過多口語詞串可能使文章失去專業性。除此之外，在比較二語學習者、母語者以及專家的學術文章後發現學習者作者撰文時在表達觀點、詞串使用上仍有不足，因此學者建議將詞串納入EAP教學大綱中，透過詞串教學讓學習者的用詞更貼近母語者，也使其學術文章提升流暢度和專業性。近十五年來已經累積了不少英文學術詞串的相關研究，反觀中文詞串之研究目前數量較少，根據劉貞好、陳浩然、楊惠媚（2017）的研究已分析出中文人文社會科學論文內的詞串特色，但仍舊需要更多分析研究來探討詞串在不同語體、學科中的特點，以提供華語教師及來台留學生的教學及寫作參考。

## 2.4 詞串功能分類

我們先前提到不同文本中的詞串會有不同類型的特性，透過分析詞串的特色能夠更了解詞串通用模式的完整性 (Hyland, 2012)，目前學者們以詞串的結構和詞串的功能來分類，分類過後能夠更具系統性地討論文本差異，對於寫作者、教學者和學習者皆為一大助益。透過功能分類，學習者能夠對照分類直接地使用和檢視所需詞串，使其文章行文流暢、銜接自然，並縮短寫作耗時。另外，也能以詞串類型分布的分析結果觀察母語者碩士論文之撰寫取向，學習者便得以了解母語者寫作的特色，使文章內容更像目標語的表達方式。

功能分類的重要性在於指出語言的意涵和目的，功能分類能夠根據不同的情境和文本背景提供文本的架構和特色，故根據詞串在文本中的含義進行分類是相當重要且有用的 (Cortes, 2004)。Cortes (2004) 比較由專家和學生撰寫的歷史和生物學門學術文章，找出文章中的學術詞串，以結構和功能分類，觀察專家和學生文章的差異。Cortes 以 Halliday & Matthiessen (1994) 提出的語言功能命名不同功能的種類，共提出三大類十一小類的功能類型，其分類如下表：

表 二-7 Cortes (2004) 的詞串功能分類

主分類	子分類
立場表達 (stance expression)	個人認知/可能 (epistemic-impersonal/ probable-possible)
	其他 (other)
語篇組織 (text organizer)	對比/推測 (contrast/ comparison inferential)
	焦點 (focus)
	推測 (inferential)
	架構 (framing)
指稱詞串 (referential)	與主題相關的詞串 (subject-bound bundles)
	時間標記 (time marker)
	數量標記 (quantifying bundles)
	地點標記 (place marker)
	描述 (descriptive)
	其他詞串 (other bundles)

上述三大分類在文本中各有功能：指稱詞串架構作者的經驗、決定他們觀察事件的方式，這樣的詞串包含時間、地點和文本標記；語篇組織詞串則是用來表達和所述之訊息相關的文本內容，例如：對比、推理和聚焦；立場詞串表達對於一事件的態度 (Cortes, 2004)。以學習者的角度來看，Cortes (2004) 的分類稍微簡單了些，立場表達僅有個人認知／可能和其他類；語篇組織僅有四個類別，若學習者欲套用在寫作中，可能仍會有許多模糊地帶。因此 Biber et al. (2004) 以 Cortes (2004) 的分類為基礎，提出更詳細的分類，在這份研究中，Biber 和同僚蒐集大學教科書和課堂授課的語料，從語料庫的索引列表



(concordance listing) 以歸納法分析詞串功能，藉此探究學術口說和學術寫作的差異。接下來 Biber 與 Barbieri 在 2007 年的研究中也繼續沿用此功能分類，他們提出的分類如下表所示：

表 二-8 Biber & Barbieri (2007) 之詞串功能分類

主分類	子分類
立場表達 (stance expression)	認知立場 (epistemic-impersonal)
	態度 (attitudinal modality stance)
語篇組織 (discourse organizing)	主題 (topic introduction/ focus)
	主題闡述 (topic elaboration/ clarification)
指稱詞串 (referential)	定位/焦點 (identification/ focus)
	模糊語 (imprecision)
	特性說明詳述 (specifying attribute)
	時間/地點/文本指稱 (time/ place/ text reference)
特殊對話功能 (special conversational function)	簡單詢問 (simple inquiry)
	報告提出 (reporting)
	表示禮貌 (politeness)

Biber & Barbieri (2007) 的分類較 Cortes (2004) 多了一大主類和四個子類，在子類的分類上也有些許不同。首先，由於該研究包含課堂授課的語料，故分類中多出一大類為「特殊對話功能」，此類詞串多見於口語中。除此之外，他們更細分了立場表達類型的詞串。他們將認知立場詞串再分為兩子類：(1) 個人認知詞串型 (personal epistemic) bundles (2) 非個人認知詞串型 (impersonal epistemic bundles)；態度情態立場則細分四子類：(1) 渴望型 (desire) (2) 義務/指示型 (obligational/directive) (3) 意圖型 (intention) (4) 能力型 (ability)。然而，他們的分類是針對口語詞串，因此對本論文詞串的分類助益不大。

同樣以 Halliday & Matthiessen (1994) 的語言功能為基礎提出詞串功能分類的還有 Hyland (2008a)。Hyland 蒐集電機工程、生物、商業以及應用語言學的學術期刊文章和碩博士學生論文，觀察專家和學生在學術寫作時使用詞串之差

異。儘管依照相同的語言學基礎，Cortes (2004) 和 Biber & Barbieri (2007) 提出的功能分類卻有相當大的差異。Hyland (2008a) 提出三大主類及十一個次類後，Salazar (2011) 在觀察母語和非母語作者的科學學術寫作之研究中，為符合其研究之目的，增減了一些子類，使分類更詳盡。其分類如下表示<sup>3</sup>：

表 二-9 Hyland (2008a) 、Salazar (2011) 提出之詞串功能分類

主分類	子分類	編碼
研究導向 (research-oriented)	地點 (location)	A1
	程序 (procedure)	A2
	數量 (quantification)	A3
	描述 (description)	A4
	<b>歸類 (grouping)</b>	A5
文本導向 (text-oriented)	<b>附加 (additive)</b>	B1
	<b>比較 (comparative)</b>	B2
	<b>推論 (inferential)</b>	B3
	因果 (causative)	B4
	組織 (structuring)	B5
	架構 (framing)	B6
	<b>引述 (citation)</b>	B7
	<b>概括化 (generalization)</b>	B8
	<b>目的 (objective)</b>	B9
參與導向 (participant-oriented)	立場 (stance)	C1
	吸引 (engagement)	C2
	<b>致謝 (acknowledgment)</b>	C3

上述的分類方式使詞串的功能更加清楚，「研究導向」類的詞串能建構作者的行為以及對真實事件的經驗，共包含五個子分類；「文本導向」類的詞串說明文本組成和文本內訊息或是論點之要素的意義，為架構學術文章的重要類

<sup>3</sup> 粗體為 Salazar (2011) 增加的子分類



別；而「參與導向」類的詞串則把焦點放在文本的作者或讀者上，說明作者的立場，亦或吸引讀者的注意。

以上三位不同學者提出的功能主分類和子分類數量如下表所示。Cortes (2004) 的主分類和 Biber & Barbieri (2007) 大致相同，除了後者因在研究中收錄口語語料，故多出「特殊對話功能」一類，還有為分析說話者的評價和態度，在「立場表達」兩個子分類下又共分出五個小類。因本研究非探討口語語料，故不參考 Biber & Barbieri (2007) 的分類；另外比較 Cortes (2004) 以及 Salazar (2011) 兩者的分類，後者子分類數量較多、較詳細，能夠清楚地歸類詞串之功能。相較於前者的分類名稱，Salazar (2011) 提出的類別名稱明確地表達了不同的功能，學習者能夠依照類別對應到到自己所需的詞串。此外，劉貞好、陳浩然、楊惠媚 (2017) 的中文人文社會學科詞串研究也採取 Salazar (2011) 的分類方式，由於該研究為目前唯一一篇中文詞串文獻，參考及比較該研究結果對本研究來說也相當重要，故本研究將採用 Salazar (2011) 所提出之功能分類進行詞串分類及討論。

表 二-10 三種詞串功能分類比較

提出者	主分類	子分類數量
Cortes (2004)	立場表達	2
	語篇組織	4
	指稱詞串	5
	其它詞串	
Biber & Barbieri (2007)	立場表達	2
	語篇組織	2
	指稱詞串	4
	特殊對話功能	3
Salazar (2011)	研究導向	5
	文本導向	9
	參與導向	3

### 第三章 研究方法

本研究探究本國學生和外籍學生撰寫之碩士論文詞串使用之差異，藉由觀察兩群體學術論文中詞串使用頻率，分析母語者及學習者寫碩士論文時使用詞串的異同，進一步提供未來以華語撰寫碩士論文的華語學習者一個參考。本章分五節，分述語料來源與蒐集方式、語料庫建置過程、語料庫分析工具、詞串抽取方式與統計方法以及詞串功能分類步驟。

#### 3.1 語料來源與蒐集

本論文自碩博士論文全文影像系統蒐集從華語文教學研究所畢業之碩士學生論文，本國籍和外國籍學生撰寫之碩士學位論文各八十篇，共計一百六十篇<sup>4</sup>。表 3-1 顯示本研究使用之語料來源學校分布，母語者論文語料分別來自北部 A、B、C 三間國立大學設置之華語文教學系所，A 大學 43 篇；B 大學 21 篇；C 大學 16 篇。論文研究領域包含：偏誤分析、語音、課程設計、學習者分析、教材設計或分析、語用、語法、市場分析、數位應用、對比分析、偏誤分析、漢字、師資培訓、漢字以及篇章結構，詳見表 3-2。

表 3-1 母語者語料來源學校分布

本國人語料	A 大學	B 大學	C 大學	總和
篇數	43	21	16	80

表 3-2 母語者論文領域分布

領域類別	篇數	領域類別	篇數	領域類別	篇數
詞彙研究	20	對比分析	5	師資培訓	1
課程設計	17	語音研究	4	篇章分析	1
語法研究	13	學習者分析	2	測驗研究	1
語用研究	6	數位應用	2	漢字研究	1
教材研究	5	偏誤分析	2		

<sup>4</sup> 其中由本國籍學生撰寫的碩士論文八十篇，已藉由國立台灣大學華語文教學碩士學位學程於 107 學年度第一學期開設之「語料庫輔助華語教學」課程專案蒐集完成。

而外籍學生論文來自全臺五間華語文教學系所之碩士畢業論文，80 篇論文之作者來自 9 個不同國家：俄羅斯、韓國、日本、越南、泰國、法國、西班牙、義大利、德國，分布篇數如下表所示。論文研究包含各領域有：偏誤分析、對比分析、市場與學習者分析、教學應用、語用、語法以及教材分析等，見表 3-4。

表 三-3 學習者國籍分布

國籍	篇數	國籍	篇數	國籍	篇數
韓國	28	泰國	6	法國	2
日本	23	義大利	3	西班牙	2
越南	12	俄羅斯	3	德國	1

表 三-4 學習者論文領域分布

領域類別	篇數	領域類別	篇數	領域類別	篇數
教材研究	14	課程設計	8	語用研究	5
對比分析	9	詞彙研究	8	數位應用	4
語法研究	9	學習者分析	5	市場分析	3
語音研究	9	偏誤分析	5	漢字	1

母語者論文領域種類共 14 類，比學習者多了「師資培訓」、「篇章結構」和「測驗研究」三個領域，蒐集文本數量最多的領域為詞彙研究，共計 20 篇；學習者論文領域共 12 類，其中「教材研究」為收錄篇數最多的領域，共 14 篇，而「市場分析」是母語者論文中沒有的種類。

## 3.2 語料庫建置

語料蒐集完成後，接下來要進行的是語料清理、中文斷詞等步驟，以下分節說明。

### 3.2.1 文本清理

論文自臺灣碩博士論文知識加值系統下載下來後，因文本受密碼保護、不得對文字進行編輯刪減，為後續研究分析需刪去部分文字敘述，故需移除檔案

之編輯限制，以利進行後續分析。本研究選用線上免付費網頁 free my pdf<sup>5</sup>，以檔案為單位移除檔案保護。將原始檔案上傳至 Free my pdf 網站，該網站會提供已移除保護功能的 pdf 檔供下載。雖然下載下來的檔案雖然已無編輯限制，但是由於 pdf 檔無法直接編輯文字，考量直接複製文字再貼上至文字處理程式中會造成格式不一致，故選擇進行二次轉檔。

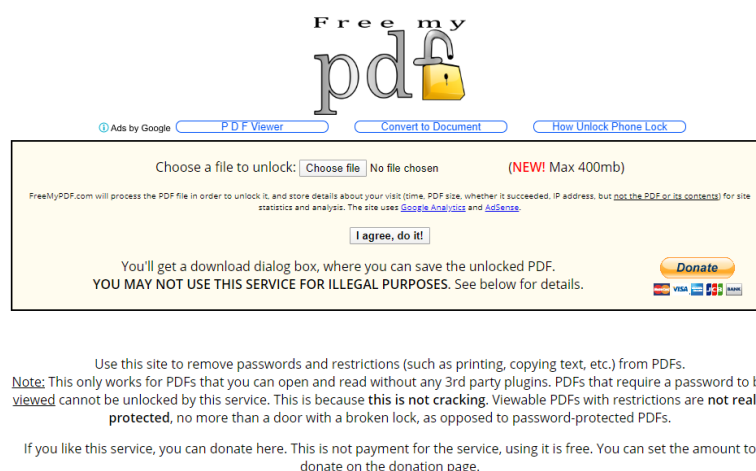


圖 3-1 Free my pdf 網站頁面

目前線上有不少 pdf 檔轉 word 檔的資源，但其中不少有各別檔案大小限制或檔案數量限制，少數資源轉檔後會造成原始文本中的圖表遺失或難以閱讀，經過測試幾個線上網站後，本研究選擇線上免付費網站 pdf to doc<sup>6</sup>，該網站能將 pdf 檔案轉成 word 檔，一次至多能上傳 20 個檔案，並能夠選擇個別下載或是批量下載，檔案經過轉檔後成 word 檔，便能直接對文本進行修改編輯。

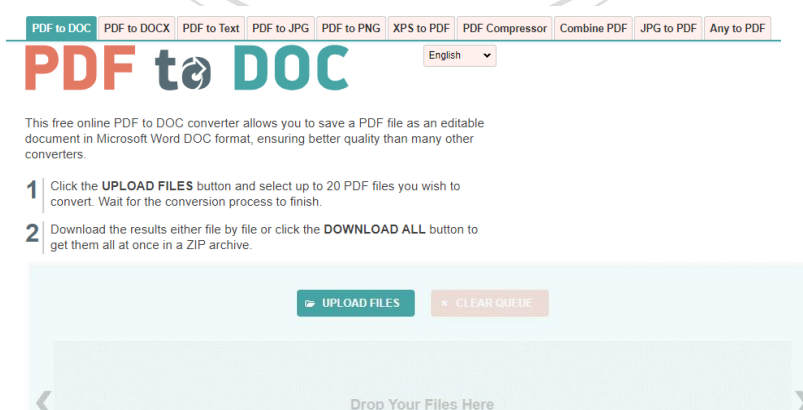


圖 3-2 pdf to doc 網站頁面

Free my pdf<sup>5</sup>網址：<https://freemypdf.com/>

<sup>6</sup> pdf to doc 網址：<https://pdf2doc.com/>

刪除非文字的敘述以及非作者的原創文字之內容以符合本研究的需求。需刪除的內容主要有三大類 (1) 非作者原創之內容，如：引用自其他文獻之敘述、專有名詞定義、例句、中介語和母語者語料、詞彙舉例、字辭典之註釋以及非原創內容之表格。(2) 非學術寫作類型之內容：圖、教案及教材範例、問卷內容、訪談逐字稿、訪談大綱。(3) 論文格式相關內容：作者姓名和校院、指導教授名、論文浮水印、目錄、頁碼及頁首頁尾之訊息。

表 三-5 文本清理刪除內容

	非作者原創內容	非學術寫作類型內容	論文格式相關內容
刪除內容類型	引用自其他文獻之敘述 中介語和母語者語料 非原創內容之表格 字辭典之註釋 專有名詞定義 詞彙舉例 例句	教案及教材範例 訪談逐字稿 問卷內容 訪談大綱	作者姓名 作者畢業之校院 論文浮水印 目錄 頁碼 頁首及頁尾

由於需要排除的內容繁多，且各文本的內容不同，需要以人工過濾所有文本。在過程中，另外發現文本內還有亂碼的問題，推測是檔案轉換所導致的問題，大部分的亂碼是系統性的，如文本中出現：「方陔」、「受誦者」等，與原始檔交叉比對過後，能找出「陔」字型應為「面」；「誦」應為「試」；「貧」為「起」、「帛」為「平」、「討」為「就」。此階段系統性的亂碼錯字問題可以透過免付費的記事本替換工具 notepad++ 批量取代<sup>7</sup>，以避免逐一文本取代的麻煩。Notepad++ 是由一名台灣工程師開發的記事本替換工具，除了有記事本的功能以外，也是程式碼的編輯器，且支援多國語言，其功能繁多，在本研究中最常使用的功能為取代、尋找及標記功能。Notepad++ 最便利的功能為能夠批量處理多個檔案。以本文為例，使用 Notepad++ 一次開啟 80 個記事本文本檔案，再將亂碼字以「在所有開啟的文件中取代」功能批量取代，便可以省

<sup>7</sup> Notepad++ 官方網站：<https://notepad-plus-plus.org/>





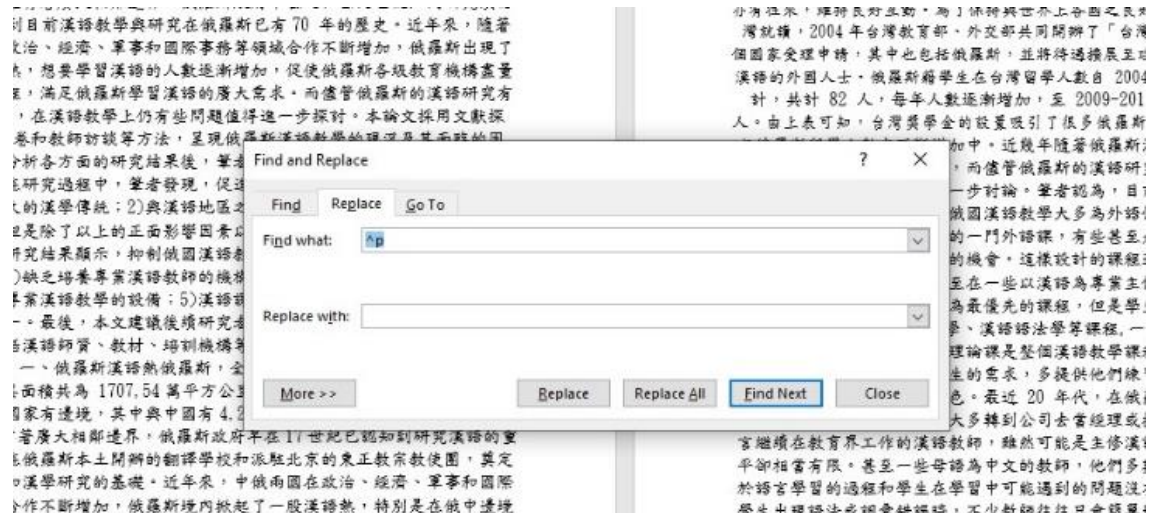


圖 3-5 段落符號取代後

### 3.2.2 斷詞

使用文本分析工具前需要先進行斷詞作業，以利後續詞串的抽取與分析。由於中文詞與詞在句子中無分界，選擇適合的系統進行斷詞相當關鍵，本研究使用中央研究院 CKIP 中文斷詞系統進行斷詞。該系統除了有線上即時分詞功能，亦能夠下載軟體離線斷詞。由於本研究共有 160 個文本，故選擇下載 CKIP 中文斷詞系統學術試用版，以進行批量斷詞。使用者能至 CKIP 中文斷詞系統學術試用版授權申請網站<sup>8</sup>，填妥資料後便能下載軟體、取得解壓縮密碼。

使用軟體前需要將文本存成 ANSI 編碼之 txt 文字檔。將所有文本都存成 ANSI 編碼之 txt 文字檔後，即可使用斷詞系統進行斷詞。斷詞前的文字因經過空格和段落刪減，文字緊鄰無間隔，斷詞後文字會以詞彙為單位被分割，每個詞和標點符號後皆有一個括弧，括弧內顯示前詞詞類（圖 3-6），詞類標記之意義請參考《中研院平衡語料庫詞類標記集》。軟體的作業速度會因文本大小而相異，待斷詞系統處理完所有文本後，即可匯入至語料庫檢索系統進行文本分析。

<sup>8</sup> CKIP 中文斷詞系統學術試用版授權申請網站：<http://ckipsvr.iis.sinica.edu.tw/ckipws/reg.php>  
誌謝中央研究院詞庫小組提供斷詞資源



1 - Notepad  
 File Edit Format View Help  
 俄羅斯(Nc) 漢語(Na) 教學(VA) 之(DE) 現況(Na) 及(Caa) 其(Nep)  
 鍵字(Na) : (COLONCATEGORY)  
 俄羅斯(Nc) 、(PAUSECATEGORY) 漢語(Na) 教學(Nv) 、(PAUSECATEGORY)  
 素(Na) 由於(Cbb) 地理(Na) 上(Ncd) 俄羅斯(Nc) 與(Caa) 中國(Nc)  
 (Nc) , (COMMACATEGORY)  
 俄羅斯(Nc) 政府(Na) 早(D) 在(P) 17世紀(Nd) 已(D) 認知到(VK)  
 (PERIODCATEGORY)  
 到(P) 目前(Nd) 漢語(Na) 教學(Nv) 與(Caa) 研究(Nv) 在(P) 俄  
 (Na) 。(PERIODCATEGORY)  
 近年(Nd) 來(Ng) , (COMMACATEGORY)

圖 3-6 經斷詞後檔案

### 3.3 語料庫分析工具

本研究以 Antconc<sup>9</sup>作為語料庫分析工具。Antconc 在 2002 年由 Laurence Anthony 開發，為一款免付費的語料庫分析研究工具，使用介面如下圖 3-7 所示，功能包含詞彙檢索 (concordance)、索引圖(Concordance Plot)、文件檢視 (File View)、詞叢 (Clusters) / N-Gram、搭配 (Collocates)、詞表 (Word List)、關鍵詞表 (Keyword List) 七個功能，軟體介面擷取圖如圖 3-8 所示。該軟體的網站提供不同電腦作業軟體的適用版本。

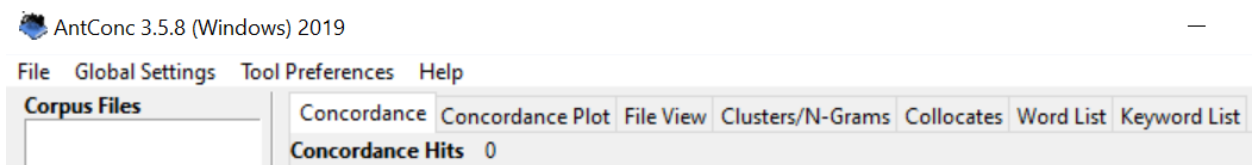


圖 3-7 Antconc3.5.8 (2019) 軟體介面擷取圖

這幾項功能，本研究幾乎都用到，其中詞彙檢索、搭配詞和文件檢視大家比較熟悉，以下各小節僅針對其他四個功能簡要描述。

<sup>9</sup> Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>

### 3.3.1 索引圖 (concordance plot)

索引圖為觀察檢索詞在文本中分布的工具，檢索結果除了可以看到該詞在語料中的次數 (concordance hits)，也有該詞在幾個文本中出現 (total plots) 的訊息，檢索成果以條型碼呈現檢索詞在每個文本中的分布，本研究使用這個功能來判斷關鍵詞表中的詞是否是因論文主題而為不尋常高頻，相關討論將呈現在本文 4.1。圖 3-8 以檢索「了」為例，從下圖得知，有 76 個母語者文本皆使用了「了」，總頻次共 2,466 次，一個 plot 就是一個文本。plot 20 為筆者文本檔案 L1(21)，「了」在 L1(21)被使用了 35 次；plot 22 中「了」的總頻次是 301 次，從條型碼可以看到「了」在 plot 22 中分布相當密集，我們因此能判斷 plot 22 論文內容可能與「了」的使用有關。

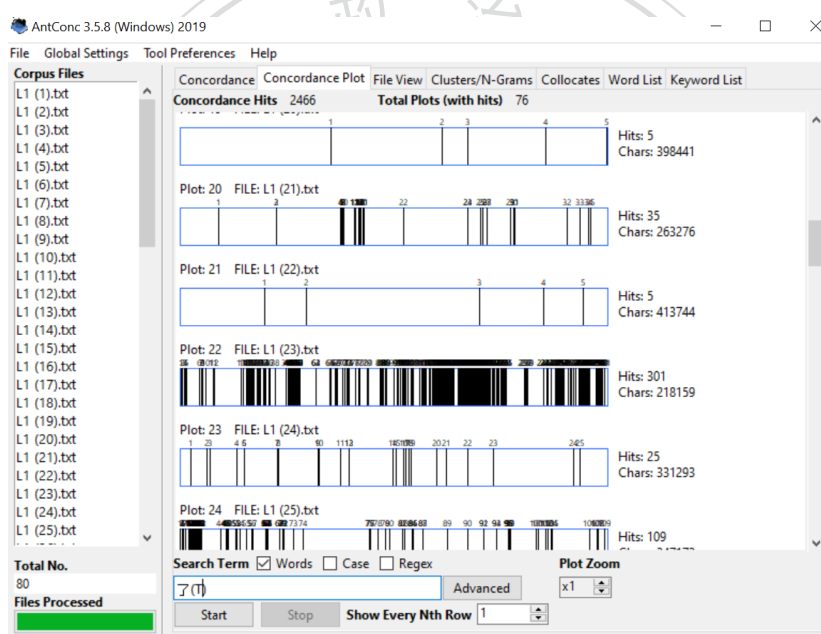


圖 3-8 Antconc 索引定位功能-檢索「了」為例

### 3.3.2 詞表 (word list)

詞表 (word list) 功能是相當方便的工具，我們將斷詞後的語料輸出就能自動計算出該文本每個詞彙出現的次數、頻率等資訊。圖 3-9 顯示學習者語料庫前 10 個高頻詞，圖中的詞種 (types) 為出現在語料庫中不同詞的總數，在學習者的 80 篇論文中共有 33,477 種不同的詞；詞數 (tokens) 為語料庫中所有詞出現的次數，可以視為文本長度，學習者語料長度為 2,104,831 詞。將詞種和詞次相除得到相異詞比率 (type token ratio)，相異詞比率可以視為詞彙的豐富性，越

高代表用字詞變化越大，在學習者語料中相異詞比率為 0.016、母語者為 0.018，表示母語者在學術論文中使用的詞彙較豐富。學習者與母語者的詞種、詞數、相異詞比例資訊請見表 3-6。

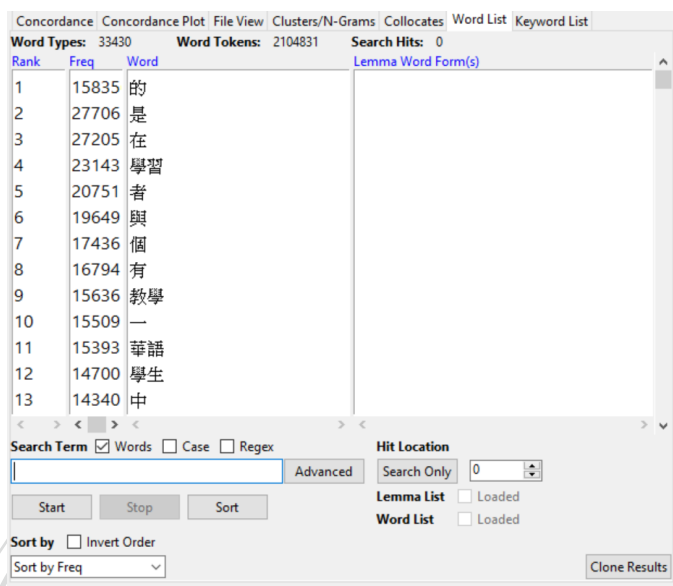


圖 三-9 學習者語料庫前 10 個高頻詞

表 三-6 學習者與母語者的詞種、詞數、相異詞比例

	母語者語料	學習者語料
文本數量	80	80
詞種 (types)	45,577	33,430
詞數 (tokens)	2,485,285	2,104,831
相異詞比率 (type token ratio)	0.018	0.016

### 3.3.3 關鍵詞表 (keyword list)

關鍵詞表為本文使用的分析工具之一，此功能是以主題關鍵性 (keyword-keyness) 統計方法分析文本中的詞彙使用顯著性，可以利用這個功能找出兩個不同語料庫中相對高頻或低頻的關鍵性詞語 (Baker, 2006, p. 139)。張莉萍(2014)的研究已經應用這個方法成功地找出不同母語背景學習者的用詞特徵，我們仿該研究做法將較大的母語者語料則作為參照語料庫 (reference corpus)、學習者語料庫為受觀察語料庫 (observed corpus)。

圖 3-11 為負關鍵詞詞表，負關鍵表示相較於參照語料庫，受觀察語料庫中不尋常低頻的詞彙，也就是學習者少用的詞彙。圖中顯示「指涉」為關鍵值最高的負關鍵詞，學習者顯著最少用的詞彙。至於顯著性的統計方法，我們使用 AntConc 工具中的預設一對數似然值檢定（Log-likelihood Ratio, 之後簡稱 LLR），門檻值設定為預設之  $p < 0.05$ 。相關描述可見 4.1 節。

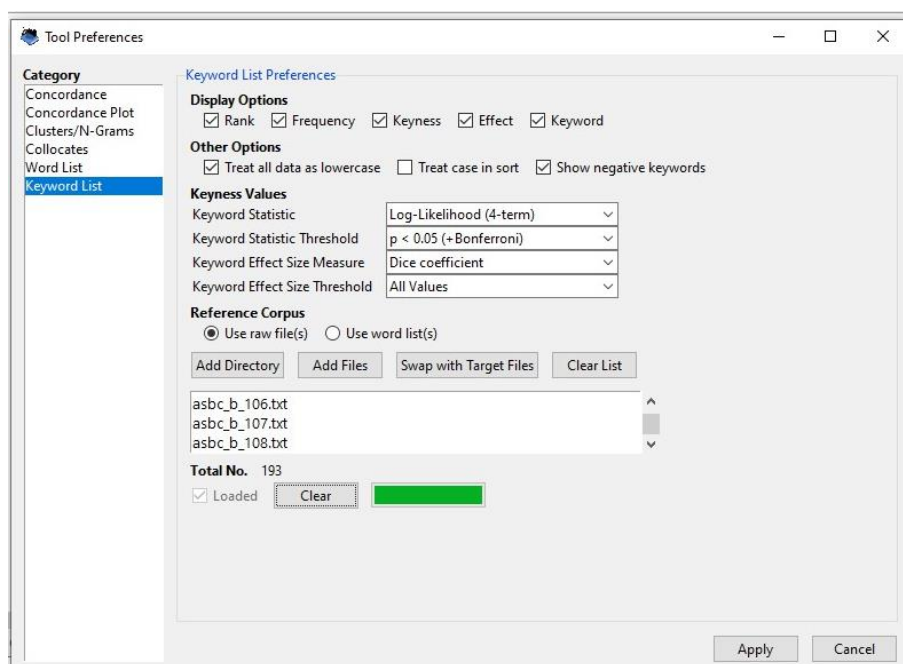


圖 3-10 Antcon3.5.8 關鍵詞詞表設定介面

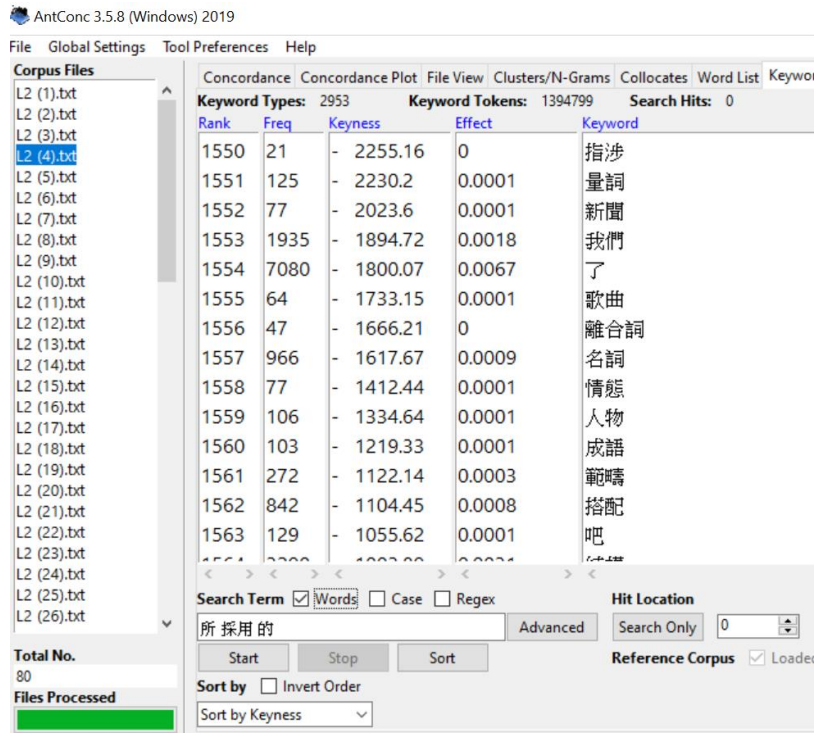


圖 3-11 學習者負關鍵詞表結果

### 3.3.4 詞叢 (Clusters) / N-Gram

詞叢 (Clusters) / N-Gram 為本研究詞串抽取的工具。詞叢用來搜尋檢索詞周圍排列成序的詞叢表，使用者可以設定詞叢的長度、向檢索詞左還是檢索詞右搜尋。與「搭配」功能相同，詞叢檢索可以依照頻率、詞叢最大或最小長度和詞頭詞尾排序，使用者也能指定詞叢的最小頻率 (Min. Freq) 和詞叢必須出現在幾個文本中 (Min. Range)，下圖為以詞叢功能檢索「而」之結果。



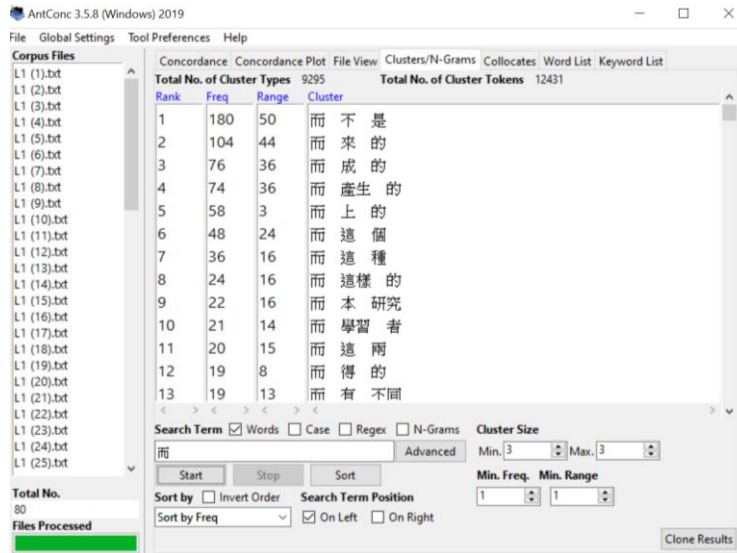


圖 3-12 詞叢功能檢索「而」之結果

詞叢功能中的 n-grams 即為本文抽取詞串之功能，抽取前的基本設定說明如下。先將所有已斷詞的文本匯入 AntconC 中，為了避免系統將詞類標記納入計算，得先進入設定介面（Tag Settings）勾選「隱藏詞類標記」（Hide tags），並將預設的詞類標記起始記號「<>」改為中央研究院的詞類標記記號「（）」。

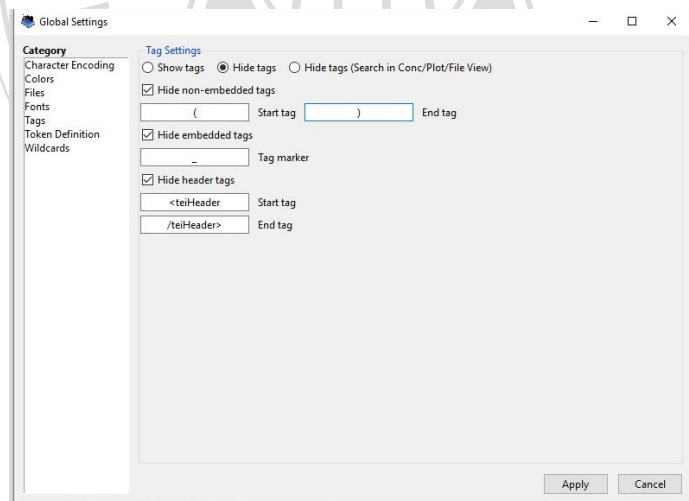


圖 3-13 AntconC 隱藏詞類標記設定介面

以下說明本研究詞串篩選門檻設定。選單進入詞叢／N-grams 列表，在下方搜尋列依照需求設定，圖 3-14 為 AntconC N-gram 設定介面。首先，勾選 N-grams 搜尋後，選擇 N 的最大值與最小值，N 代表詞數，因此若最小值選擇 2、最大值選擇 5，系統會顯示兩詞到五詞的所有詞串。因資料量龐大，筆者選擇

依長度分割，分次進行詞串抽取，以下以抽取母語者四詞詞串設定為例。N-gram 最小值、最大值皆設定為 4，本文以每一百萬詞出現 20 次作為詞串篩選標準，母語者語料共約兩百五十萬詞，因此最小頻率值設定為 50，最小範圍 (Min. Range) 指詞串最少在幾個文本中出現，由於本研究設定至少需要出現在 10% 的文本中，因此設定為 8。設定完成後即能選取開始 (start) 讓系統產出詞串，結果如圖 3-15 所示。

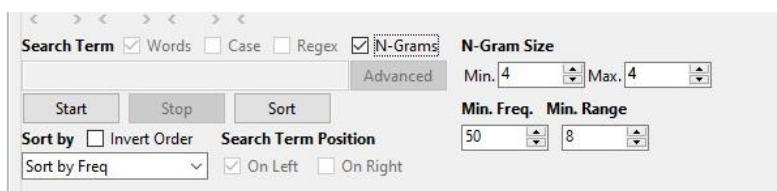


圖 3-14 Antconc N-gram 搜尋設定 (以母語者四詞詞串為例)

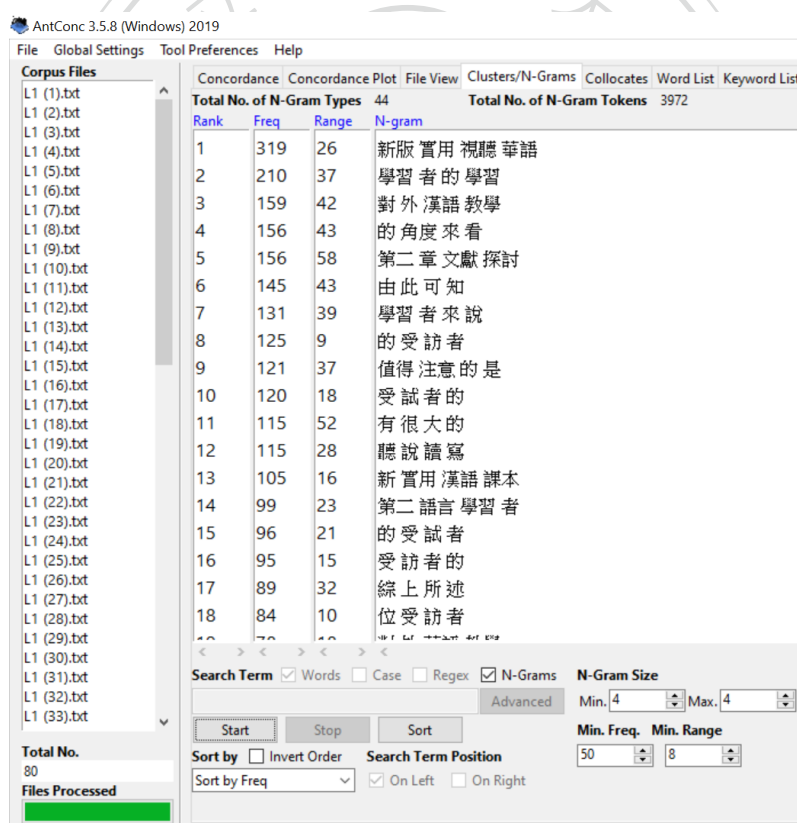


圖 3-15 Antconc N 字元搜尋結果 (以母語者四詞詞串為例)

### 3.4 詞串抽取與統計方法

我們在 2.1 提及不同的研究有不同的詞串篩選機制，過往文獻篩選詞串通常有三個準則。首先為詞串頻率，較常用使用的篩選頻率標準為每一百萬詞 20 至 40 次 (e.g., Biber at al., 2004; Hyland, 2008a)，而針對較小的語料庫，也有研究者選擇以每一百萬詞 2 至 10 次作為標準 (e.g., Altenberg, 1998; De Cock, 1998)。第二個標準為文本分布率，文本分布率是為了避免篩選出單一作者大量使用特定詞串，Biber 和 Barbieri (2007) 建議單一詞串至少要出現在 3-5 個文本中，Hyland (2008a) 則是以單一詞串需要出現在 10% 的文本中作為標準。最後一個篩選機制為是否進行人工篩選，部分研究為比較不同領域的文本，會以人工方式刪除與主題相關的詞彙。舉例來說，「學習者的角度」在本研究中屬於高頻詞串，但在其他領域的文獻中必定相當少見或甚至完全不存在，為達到能與其他研究比較的目標，本文刪除與華語教學主題相關的詞彙。不過誠如 Biber 和 Barbieri (2007) 所言，研究者應自行調整篩選頻率，以符合研究之目的，筆者將本文參考之文獻的篩選標準整理成表 3-7。

表 3-7 詞串篩選機制 (筆者整理)

	篩選標準	過往文獻標準
第一階段	頻率	1. 每一百萬詞 20 次 (Hyland, 2008a) 2. 每一百萬詞 25 次 (Chen & Baker, 2010) 3. 每一百萬詞 40 次 (Biber at al., 2004)
第二階段	文本分布率	1. 單一詞串至少出現在 5 個文本中 (Biber at al., 2004) 2. 單一詞串至少出現在 10% 的文本中 (Hyland, 2008a) 3. 至少出現在 3 個文本中 (Chen & Baker, 2010)
第三階段	人工過濾	1. 是否移除含有領域相關詞彙之詞串。

本研究語料共有 160 篇碩士學位論文，母語者語料詞數 (tokens) 約為兩百四十八萬詞、詞種 (types) 為四萬五千詞；學習者語料詞數約為兩百一十萬詞、詞數為三萬三千詞。因語料來自同一語體、同一學科，故本研究採用相對較嚴格的頻率門檻，首先以每一百萬詞 20 次作為詞串抽取之門檻，另外限制單一詞串必須得出現在 10% 的文本中，以避免詞串來自單一作者過度使用的情況。

依照詞串長度依序抽取完學習者與母語者的詞串，得到的初步結果如表 3-8 所示。我們可以注意到以每百萬詞 20 次的頻率抽取詞串使五詞詞串數量相當少，為利後續分析，研究者將五詞詞串的抽取標準降低至每一百萬詞 10 次、文本分布率 10%。微調後母與者五詞詞串有 22 個、學習者 26 個。

表 三-8 初步詞串抽取結果

詞串	母語者		學習者	
	數量	次數	數量	次數
二詞	4,215	659,433	4,589	643,047
三詞	588	61,736	808	75,789
四詞	44	3,972	77	5,996
五詞	0	0	3	180

表 三-9 頻率調整後詞串抽取結果

長度	母語者		學習者	
	種類	次數	種類	次數
二詞	4215	659,433	4547	636,178
三詞	588	61,736	808	75,789
四詞	44	3,972	77	5,996
五詞	22	711	26	806

從上表我們看到二詞詞串數量非常多，但從圖 3-16 的檢索，我們能看到詞串的組成多為詞彙和詞彙的搭配（如：一個、不同的）或是一個合成詞

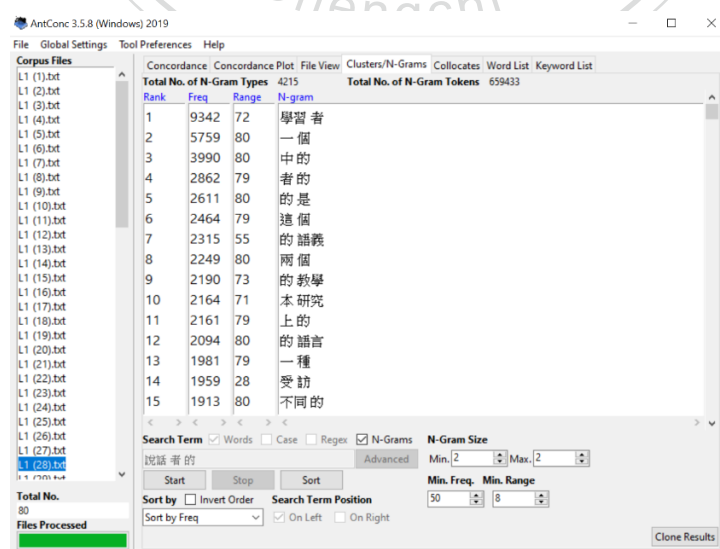


圖 3-16 母語者二詞詞串檢索結果

(如：學習者、本研究)，與本文欲探討之序列組合關聯較低，故本文暫不討論二詞詞串。

排除二詞詞串後，筆者透過 excel 資料篩選工具找出三詞到五詞詞串的共有詞串，再將詞串分類為共有詞串、母語者獨有與學習者獨有三類。由於三詞詞串數量龐大，以下將僅呈現四詞及五詞詞串的檢索分類結果。

表 三-10 四詞詞串檢索分類結果

#	共有詞串	母語者獨有詞串	學習者獨有詞串	
1	值得注意的是	新版實用視聽華語	兩種語言的	與學習者的
2	綜上所述	的受訪者	華語學習者的	學習者的華語
3	受訪者的	新實用漢語課本	在課堂上的	對學習者的
4	的角度來看	這一類的	學生在課堂上	第二語言教學的
5	第二語言學習者	能讓學習者	本論文的研究	這種情況下
6	第二章文獻探討	受訪者對	教師在課堂上	培養學習者的
7	位受訪者	現代漢語八百詞	母語的學習者	受到母語的影響
8	學習者來說	英語為母語的	對外漢語教材	在學習過程中
9	對外華語教學	讓學習者在	為母語的學習	有一個人
10	對外漢語教學	如圖所示	受試者對	第三章研究方法
11	兩者之間的	以英語為母語	學習者為中心	一對一的
12	由此可知	如表所示	學習者的偏誤	受試者認為
13	位受試者	一個完整的	的華語學習者	的一個重要
14	在日常生活中	一個新的	受試者所	華語文教學研究所
15	受試者的	也是一種	在教學過程中	語言教學中的
16	學習者的需求	在現代漢語中	提高學習者的	在對外漢語
17	的受試者	如下表所示	第二語言習得的	學生學習華語的
18	學習者的母語		學習者的興趣	引起學習者的
19	學習者在學習		語言學習中的	本章第一節
20	以學習者為		韓籍學習者的	程度的學習者
21	學習者的語言		的漢語學習者	學習的過程中
22	聽說讀寫		日籍漢語學習者	語言與文化的
23	受試者在		在這種情況	還有一個
24	學習者的學習		表達自己的意見	表四所示
25	兩個以上的		有密切的關係	
26	有很大的		華語教學上的	



表 三-11 五詞詞串檢索分類結果

#	共有詞串	母語者獨有詞串	學習者獨有詞串
1	聽說讀寫四	以英語為母語的	在這種情況下
2	為母語的學習者	實用視聽華語第一冊	大學華語文教學研究所
3	或兩個以上的	對外漢語教學語法	師範大學華語文教學
4	對學習者來說	的學習者來說	學習者的學習動機
5	對外漢語教學中	值得一提的是	學習者為中心的
6	在對外漢語教學	新版實用視聽華語第一	對外漢語教學的
7	兩個或兩個	詞彙與漢字等級大綱	有很大的幫助
8	個或兩個以上	說讀寫四項	兩種一種是
9	以學習者為中心	中所扮演的角色	語言學習的過程中
10		師範大學國語教學中心	第三章研究方法本
11		國家華語測驗推動工作	第一章緒論第一節
12		學習者的語言能力	第二章文獻探討本
13		華語測驗推動工作委員會	台灣師範大學華語文
14			學習者在學習過程
15			章緒論第一節研究
16			者在學習過程中
17			聽說讀寫的

抽取出詞串後筆者進一步將詞串分為獨有詞串及共有詞串，並透過 LLR<sup>10</sup> 顯著性分析母語者及學習者共有詞串使用頻率之差異。表 3-12、13 為四詞及五詞共有詞串的 log-likelihood 分析結果，表中 L1、L2 欄為母語者以及學習者的使用該詞串的頻率，從 sig. 欄位可以得知母語者與學習者在這個詞串的使用頻率上是否達顯著差異，星號 (\*) 越多表示差異越顯著，若無星號即代表統計尚未達顯著，最後一個欄位若為正記號 (+) 表示該詞串在母語者語料庫 (L1) 中頻率較高；負記號 (-) 則代表在學習者語料庫頻率較高，進一步的分析將在 4.2 呈現。

<sup>10</sup> 此處 LLR 結果使用《語料庫應用教程》(梁茂成、李文中、許家金, 2010) 隨書附之 LLR 計算工具

表 三-12 共有四詞詞串 LLR 顯著性分析結果

#	詞串	L1	L2	LLR	Sig.		
1	值得注意的是	121	49	44.61	0.000	***	+
2	綜上所述	89	43	24.92	0.000	***	+
3	受訪者的	95	49	23.59	0.000	***	+
4	的角度來看	156	111	16.93	0.000	***	+
5	第二語言學習者	99	62	15.83	0.000	***	+
6	第二章文獻探討	156	117	13.95	0.000	***	+
7	位受訪者	84	53	13.17	0.000	***	+
8	學習者來說	131	103	9.62	0.002	**	+
9	對外華語教學	78	54	9.28	0.002	**	+
10	對外漢語教學	159	149	4.11	0.043	*	+
11	兩者之間的	69	59	3.32	0.068		+
12	由此可知	145	145	2.00	0.158		+
13	位受試者	50	49	0.86	0.354		+
14	在日常生活中	53	54	0.58	0.446		+
15	受試者的	120	409	122.51	0.000	***	-
16	學習者的需求	52	154	37.28	0.000	***	-
17	的受試者	96	219	31.07	0.000	***	-
18	學習者的母語	50	102	10.56	0.001	**	-
19	學習者在學習	58	113	10.05	0.002	**	-
20	以學習者為	50	77	2.17	0.140		-
21	學習者的語言	68	96	1.28	0.257		-
22	聽說讀寫	115	153	0.94	0.332		-
23	受試者在	72	96	0.61	0.434		-
24	學習者的學習	210	254	0.07	0.796		-
25	兩個以上的	50	61	0.03	0.864		-
26	有很大的	115	138	0.02	0.898		-

表 三-13 共有五詞詞串 LLR 顯著性分析結果

#	詞串	L1	L2	LLR	Sig.		
1	聽說讀寫四	45	24	10.46	0.001	**	+
2	為母語的學習者	40	68	3.44	0.064		-
3	或兩個以上的	28	24	1.33	0.249		+
4	對學習者來說	38	28	3.64	0.057		+
5	對外漢語教學中	32	30	0.82	0.364		+
6	在對外漢語教學	29	27	0.79	0.374		+
7	兩個或兩個	36	30	2.00	0.158		+
8	個或兩個以上	36	30	2.00	0.158		+
9	以學習者為中心	30	65	8.05	0.005	**	-

### 3.5 詞串功能分類步驟

在進入分析前，最後一步就是為我們經過前述步驟得到的詞串進行功能分類，這部分工作只能以人工方式進行。從表 3-9 我們能看到三詞詞串為數量最多的詞串，四詞及五詞詞串數量較少，且對照表 3-10、11，能看到大部分四、五詞詞串皆含有領域相關的詞彙，故不適合進行功能分析。因此，本節分析僅針對三詞詞串討論。

在分類詞串以前，為了能與其他詞串研究比較，以及為提供未來各領域的學習者直接參考的依據，本文需先以人工方式過濾不適合進行功能分類的詞串。需刪減的詞串有兩種：(1) 含有特定或專業領域辭彙之詞串。如：漢語、受訪者、問卷。(2) 語義不完整之詞串。如：的每一、了這個、者之間的。經過刪減後得到的結果如表 3-14 所示，母語者三詞詞串 229 個、學習者論文三詞詞串 272 個。

表 3-14 經篩選後之三詞、四詞詞串數量

長度	母語者		學習者	
	種類	總次數	種類	總次數
三詞	229	24017	272	24555

至於功能分類的架構，我們在 2.4 節已經說明採用 Salazar (2011) 的分類方法，將詞串功能類型分成三主類：研究導向 (research-oriented)、文本導向 (text-oriented)、參與導向 (participant-oriented)，在每一個主類下再細分次類共十七次類，分類的細項和定義見下表：

表 三-15 學術詞串功能分類 (Hyland, 2008a, pp.13-14; Salazar, 2011, pp.52-53; 劉貞好、陳浩然、楊惠媚譯, 2017, p.132)

主類	定義	次類	定義
研究導向	輔助作者架構活動及真實世界的經驗	地點	表地點、界線、方向
		程序	敘述事件、行動及實驗方法
		數量	表計量、數量、比例及相關的變化
		敘述	說明質量、程度以及存在
		歸類	表群組、種類、部分與順序
文本導向	有關傳達訊息與提出論述之文本的組織架構及意義	附加	於不同要素間建立之附加連結
		比較	比較及對比不同要素
		推論	根據資料數據來表示推論與結論
		因果	指出要素間之因果關係
		組織	能組織話題的範圍或將讀者引導至文本它處的文本反思標記
		架構	藉由表明限定條件來提出論證
		引述	說明論文來源或支持論述的數據
		概括化	標示普遍接受之事實或陳述
		目的	介紹作者之研究目的
參與導向	聚焦於作者或讀者	立場	傳達作者之態度與評價
		吸引	直接吸引讀者注意力
		致謝	指出參與研究者或機構之貢獻

進行功能分類時，可能會碰到一個詞串有數種功能的問題 (Biber et al, 2004)。舉例來說，依照分類，詞串「稱之為」可能作為描述型詞串，在語料(1)中，「稱之為」僅說明作者對「了」的分類，為說明存在的一種；而語料(2)的「稱之為」則引用了前人的論文，為引述的一種。

(1) 「為了配合一般的分類，我們稱之為“了3”」 L1-42<sup>11</sup>

<sup>11</sup> L1-42 為研究者對文本的編碼，L1 表「母語者」；L2 為「學習者」，數字為文本編號。

我們依循前人的解決方式——將有多重功能的詞串分類至最頻繁被使用的功能中（Biber et al., 2004; Salazar, 2011）。為了判定「稱之為」在語料庫中較常被作為什麼功能使用，筆者以 AntConc 搭配詞功能觀察與該詞串搭配的词彙。從圖 3-17 我們可以看到最高頻與「稱之為」搭配的词彙左邊為「我們」，用法如語料 (1) 所示。與「稱之為」搭配第二、第三高頻的「等」及「學者」都是引述的用法，不過筆數不如「我們」多。因此，研究者將「稱之為」歸類至描述型詞串。其他詞串如有重疊功能，也是採取此模式來解決。

Concordance					
Concordance Plot		File View		Clusters/N-Grams	
Collocates		Word List		Keyword List	
Total No. of Collocate Types: 87			Total No. of Collocate Tokens: 146		
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	17	17	0	2.78259	我們
2	4	4	0	0.69579	等
3	4	4	0	2.81611	學者
4	4	4	0	0.35571	可以
5	4	4	0	0.72415	並
6	3	3	0	4.24423	說法
7	3	3	0	-0.66046	而
8	3	3	0	0.09591	種
9	3	3	0	5.86438	片語
10	3	3	0	4.63742	大陸
11	3	3	0	0.44859	因此
12	3	3	0	0.34781	可
13	3	3	0	0.46851	則
14	3	3	0	1.06468	他
15	3	3	0	2.19889	亦
16	2	2	0	0.12562	關係
17	2	2	0	-0.37338	都
18	2	2	0	0.36436	詞
19	2	2	0	-0.49278	結構

Search Term <input checked="" type="checkbox"/> Words <input type="checkbox"/> Case <input type="checkbox"/> Regex		Window Span <input type="checkbox"/> Same	
稱之為	Advanced	From... 1L	To... 0
Start	Stop	Min. Collocate Frequency	
Sort by <input type="checkbox"/> Invert Order		1	
Sort by Freq(L)			

圖 3-17 「稱之為」 Collocates 檢索結果



## 第四章 母語者與學習者之用詞與詞串比較

本研究探討母語者和學習者碩士畢業論文中詞串使用的差異，本章從三個面向來討論。第一節以關鍵詞表分析方法來討論學習者和母語者兩群體的用詞差異；第二節則是觀察已篩選出的詞串之使用頻率、兩群體間共有和獨有詞串的特性分析異同；第三節則藉由功能分類 (Hyland 2008a; Salazar, 2011) 觀察兩群體使用之詞串在不同功能類別中的分布情況。

### 4.1 關鍵詞表分析

關鍵詞表時常作為語料風格學 (Corpus stylistics) 的分析依據，學者使用語料庫工具和統計觀察文本用詞，進一步分析作家的寫作風格、語言形式和文學意義 (Mahlberg, 2012; Gao, 2019)，本文也藉關鍵詞表來比較母語者和學習者論文寫作使用的詞彙差異。表 4-1 依照關鍵性 (keyness) 排序，呈現前 100 個正關鍵詞，也就是與母語者語料庫相比，學習者不尋常高頻使用的詞彙；表 4-2 呈現前 100 個負關鍵詞，是學習者顯著少用的詞彙。排列順序為由上至下、由左到右。「華語」為正關鍵詞表中關鍵性最高的詞；「指涉」為負關鍵詞表中關鍵性最高的詞，換句話說，學習者使用「華語」、「指涉」的頻率和母語者使用的頻率差異最大。

表 4-1 前 100 個正關鍵詞

正關鍵詞									
華語	觀光	調查	結果	偏誤	越南籍	試者	旅客	舌尖	越語
韓語	日語	胡志明	漫畫	老師	中國	韓國人	項目	因素	俄語
韓國	學校	日本	形容詞	教師	基頻	聲調	題	年	需要
學生	研究	慣用語	處所詞	陽平	需求	陌生詞	外語	兼語	句式
日籍	俄羅斯	選	去聲	移民	教材	工作	審	服務	中國語
學習	塞音	文詩	專業	大學	佔	課程	語言	比較	裡
越南	網站	同形詞	使用率	高中	漢語	同形	漢字	第二	商務
導遊	他們	疑問句	教育	教科書	上聲	陰平	試	義大利	主要
送氣	音	韓文	漢	漢字詞	日本人	語調	旅遊	筆者	漢日
韓籍	問卷	輔音	韓	讚美	進行	自主	對比	調	策略

表 四-2 前 100 個負關鍵詞

負關鍵詞									
指涉	成語	人家	訪	閱讀	命題	情狀	現代	遊戲	例
量詞	範疇	或是	說話	事件	權勢	能	遠距	高頻	故事
新聞	搭配	語義	規勸	共現	時態	劃分	用法	物	篇
我們	吧	繪本	意義	補語	具	計量	攻擊	必須	複數
了	結構	戰爭	下	重動	文本	小說	成分	門	抽象
歌曲	概念	而	動	後	這麼	道德	重述	完成	副詞
離合詞	標記	新手	討論	時相	義符	詞類	如果說	便	如
名詞	音樂	則	語意	延伸	類	財富	分類	當	詞素
情態	語素	隱喻	指稱	義	歌詞	於	評價	詞	修飾
人物	電影	婚禮	域	報刊	關係	們	尊	然而	習者

正、負關鍵詞表呈現學習者與母語者文本中使用詞彙的差異，我們從詞表中也觀察到斷詞系統的一些問題。儘管中研院斷詞系統的功能已經相當完善、有辨識新詞的能力，但我們仍須特別注意一些斷詞的狀況。例如：表 4-1 的「試者」與表四-2 的「習者」和「訪」原來分別是「受試者」、「學習者」、「受訪者」三個詞組，而斷詞系統分別斷為：學、習者、受、訪、者，因為系統斷詞判斷的緣故，使詞表中出現這幾個不完整的詞。另外，表四-2 中的「如果說」也是因為系統將其辨識為一個詞，本研究詞串之斷詞憑據依照中研院斷詞系統產出之成果，暫不討論系統判定特定詞彙適當與否。

藉由關鍵詞表，我們可以看到幾個現象。首先，文本主題對於關鍵詞表的影響。正關鍵詞表中有不少國家、國籍、語言的專用名詞，如：韓國、韓語、日籍等。另正負關鍵詞表皆有不少與論文主題相關的詞彙，這些詞彙能與 3.1 的表 3-4 相呼應，例：送氣、塞音、教育、量詞、離合詞、語義等。再者，負關鍵詞表中功能詞「了」、「吧」與「這麼」也同樣與論文研究主題相關。這個部份我們可以初步經由 AntConc 索引圖功能看出（第 3 章已介紹），圖 4-1 為筆者以母語者語料庫檢索「了」的結果，我們看到「了」在 plot 22 高度密集，因此我們回頭查閱第 22 本論文的主題，發現為「法籍學習者體標記『了』、『著』及『過』習得之對比分析」，可以說「了」的高頻使用與文本主題有關，表 4-2 中的「吧」與「這麼」，我們也都透過這樣的方式得知這些都是非用詞習慣，而是文本主題的關係。



圖 4-1 Antconc Concordance Plot 檢索「了」結果

雖然正負關鍵詞表中絕大部分都是與文本特性相關的詞彙，但是仍能注意到一些值得討論的詞彙，以下將討論虛詞「而」、指稱詞「我們」和動詞「進行」討論母語者和學習者之差異。負關鍵詞表中有不少虛詞，如（1）連接詞：或是、而、然而（2）副詞：便、則（3）介詞：如、於。虛詞的抽象語法意義讓學習者在使用上時常出現偏誤，是華語學習的難點之一，尤其學術論文為書面語，掌握書面虛詞對學習者而言更為困難。「而」在負關鍵詞表中排第二十

六個，可見與母語者相比，學習者使用「而」的頻率少得多<sup>12</sup>，我們透過語料庫檢索觀察母語者和學習者使用「而」的狀況：

- (3) 「糾錯時機，得視執行階段而定...」 L1-72
- (4) 「包含在雙音組合中，原是能獨立成詞的單音詞因長期使用雙音組合而產生弱化...」 L1-15
- (5) 「而相較之下，語法複數如-s和量詞除了在語意上都標示可數性外...」 L1-6
- (6) 「因此，教學上需要進行替換練習，要求學生恰當地把存現句改成一般語序的句子，而讓他們注意使用介詞。」 L2-37
- (7) 「只顧滔滔不絕地講解，而一點不顧遊客的問題。」 L2-50

「而」是一個多功能詞，語料(3)、(4)、(5)出自於母語者語料，(6)、(7)則是來自學習者語料。其他在負關鍵詞表中類似的關聯詞與還有「則」、「然而」、「便」，這個現象顯示學習者使用關聯詞的頻率相較於母語者較低，若學習者能正確地使用這些關聯詞能使寫作風格更接近母語者，因此未來學術寫作課程及相關教材編纂應多關注這類型詞彙的使用方式及練習。

從人稱代詞的選擇上也能看到母語者和學習者的差異。「我們」在負關鍵詞表排第四，表示母語者使用「我們」的頻率顯著高於學習者。在撰寫學術論文時，母語者時常使用「我們」指稱研究或研究者本身，如下列句子：

- (8) 「我們發現2006年後的臺灣... (略)」 L1-74
- (9) 「我們主張『語句』在交際溝通時... (略)」 L1-75
- (10) 「在第四章我們將提到... (略)」 L1-6

人稱代詞使用與作者的態度有關，使用「我們」含括筆者和讀者，是有禮貌的表達方式 (Brown & Levinson, 1987)，另一觀點是在學術研究中使用「我們」是作者將自己與研究成果連結的手段，突顯自身研究內容的原創性以及作者本人貢獻度 (Salazar, 2011)，也就是說，使用「我們」有吸引讀者與標注研究價值的作用。再對照正關鍵詞表，比起使用「我們」，學習者更常使用「筆者」來指稱自己。從語用方面來看，學習者多使用「筆者」的原因可能是由於

<sup>12</sup> 母語者論文中有一篇為探討「而」的使用

使用第三人稱指稱自己是較為客觀的表達方法，在過往文獻中，也曾顯示學習者撰寫學術文章似乎較重視客觀性 (Amirian, Ketabi, & Eshaghi, 2013; Salazar, 2011) 以及母語者顯著多用「我們」的趨勢 (Salazar, 2011)。

「我們」是學習者不尋常低頻的詞彙，不過「他們」卻是學習者高頻使用的人稱代詞，高頻原因可能有二。第一點是主題的影響。我們從 L2 語料庫中「學習者」為第五高頻、「學生」為第十五個高頻詞，可以推測 L2 語料庫中探討特定的群體的文本較母語者多，因此使用「他們」的機率也相對提高。第二點可能與學習者的偏誤有關，以下兩個例子呈現學習者的使用「他們」的狀況：

- (11) 「他們運用『拒絕』策略 (30%) 的比例反而來得比「修正」還高。德國人在此的表現接近他們所有語料的整體表現」 L2-52
- (12) 「還有他們是否願意在他們的課程中加上漢字字源的教法；他們自己對甲骨文或古代字是否有所理解、是否願意為了增加學生對漢字的了解與興趣而在課堂上特別加上漢字字源的課。」 L2-20

語料(11)的「他們」指學習者，第一個「他們」不能省略，但是第二個「他們」似乎就有些冗贅。相同地，(12)中的「他們」指教師，第一個「他們」同樣不能省，但第二個和第三個省略後能讓句子更通順：

- (11)-1. 「他們運用『拒絕』策略 (30%) 的比例反而來得比「修正」還高。德國人在此的表現接近  $\emptyset$  所有語料的整體表現：」
- (12)-1. 「還有他們是否願意在  $\emptyset$  課程中加上漢字字源的教法； $\emptyset$  自己對甲骨文或古代字是否有所理解...」

以上兩個例句顯示學習者有代詞多用、零代詞少用的現象。零代詞是漢語的特點，且過往研究指出儘管是高級學習者仍有零代詞使用不足的狀況 (周曉芳, 2011)，第三人稱代詞的使用又是其中最複雜的。可見第三人稱代詞的複雜性可能也是造成學習者偏誤的原因之一。因此，零代詞使用的教學對於高級學習者而言仍相當重要。



最後，正關鍵詞表中的「進行」也是值得討論的一個詞。學習者用「進行」的頻率顯著高於母語者，L2 使用例句如下：

- |                            |       |
|----------------------------|-------|
| (13) 「用目的語進行交際... (略)」     | L2-66 |
| (14) 「只針對動詞及物性進行分析... (略)」 | L2-25 |
| (15) 「本研究上第一堂課進行測試... (略)」 | L2-37 |

儘管例(13)-(15)句子都不是病句，但若將「進行」刪除，句子精簡許多：

- |                        |       |
|------------------------|-------|
| (13)-1 「用目的語交際...」     | L2-66 |
| (14)-1 「只針對動詞及物性分析...」 | L2-25 |
| (15)-1 「本研究上第一堂課測試...」 | L2-37 |

漢語語法受到外來語影響的例子越來越多，近幾年語言學家常以「語言癌」討論這個現象（何萬順、蔡維天、張榮興、徐嘉慧、魏美瑤、何德華，2016），但早在 1987 年，余光中先生就提出「漢語歐化」（或漢語西化）的概念，意思是指受到英語語法影響下的中文表現。余光中先生提出了十六點惡性歐化的現象（高大威，2014），其中一點為「進行」的過度使用。受到外語的影響，很多時候人們會將簡單明瞭的動詞分解成「萬能動詞+抽象名詞」，這樣的冗贅結構讓句子看起來較堂皇。關鍵詞表分析的結果看起來，L2 用得比 L1 多，究竟是受目標語泛化的影響還是學習者本身母語語法的影響，有待進一步研究。

## 4.2 詞串使用分析

上一節我們透過關鍵詞表分析發現學習者和母語者用詞上的差異，這一節我們將觀察抽取出之詞串，試討論母語者以及學習者的詞串使用差異。

### 4.2.1 詞串分布

在 3.4 筆者說明了本文詞串篩選的方式，抽取的門檻為文本分布率 10%，兩詞、三詞、四詞詞串頻率為每一百萬詞 20 次；五詞因數量較低故將頻率降低至每一百萬詞 10 次。在 3.4 我們也說明了由於二詞詞串數量過於龐大，且不符合本文欲討論現象，故暫不討論。為方便閱讀，筆者將最終抽取出之詞串結果再次呈現在表 4-3：

表 4-3 詞串之種類和頻次

詞串	母語者		學習者	
	種類	頻次	種類	頻次
三詞	588	61,736	808	74,172
四詞	44	3,972	77	5,996
五詞	22	711	26	806

從表 4-3，我們發現詞串長度越長、種類和頻次皆越少。另外從表中我們也可以看到一個現象，即學習者的詞串種類都較母語者多，這個現象發生在三詞、四詞、五詞；在使用次數上也都多於母語者。這個結果顯示學習者有多用固定詞串的趨勢，因為學習者的語料量小於母語者，而詞串種類和次數卻高於母語者。這個結果似乎支持 Pan, Reppen & Biber (2016) 的調查結果，他們比較母語者和學習者撰寫的電子通訊期刊文章，以每一百萬詞 40 次的頻率作為篩選門檻，發現儘管母語者的平均文長較長，但是詞串の種類數卻較少，因此認為學習者在寫作時較依賴詞串使用。同樣的結果也在 Amirian, Ketabi & Eshaghi (2013) 的研究中出現，他們以每百萬詞 20 次的頻率篩選母語者和學習者的碩士論文詞串，在總和約兩百萬詞的語料庫中篩選出 61 個母語者詞串、211 個學習者詞串，兩群體的詞串數量差異比本文大。該研究認為，第二外語課程強調學習者使用詞串銜接前後句的能力，使學習者使用詞串的頻次明顯高於母語者。

表中三詞詞串的數量最多，筆者刪除與主題相關的內容後，將前 20 高頻的詞串整理成表 4-4<sup>13</sup>。透過下表，我們能觀察到母語者及學習者高頻詞串重疊性很高，在前 20 個詞串中有 14 個相同的詞串，可以說這 14 個詞串為學術論文中最常見的三詞詞串。從排序的角度來看，母語者最高頻的三個詞串是「是一個」、「是一種」、「指的是」皆屬於概念解釋的詞串，這三個詞串也在學習者的高頻詞串中，由此可推測概念解釋在論文中佔很大的比例。學習者最高頻的詞串「本研究的」是與研究相關的內容。藉由高頻詞串表我們觀察到，兩個語料庫重疊性（70%）非常高。

<sup>13</sup> 完整且刪除主題相關詞彙的三、四詞共有詞串詳見本文附錄

表 四-4 母語者及學習者前 20 高頻三詞詞串<sup>14</sup>

排序	母語者		學習者	
	頻次	詞串	頻次	詞串
1	<b>763</b>	<b>是一個</b>	<b>491</b>	<b>本研究的</b>
2	<b>573</b>	<b>是一種</b>	<b>460</b>	<b>是一個</b>
3	<b>490</b>	<b>指的是</b>	<b>390</b>	<b>是一種</b>
4	360	的情況下	350	研究者認為
5	<b>341</b>	<b>這兩個</b>	336	所使用的
6	<b>341</b>	<b>所提出的</b>	<b>335</b>	<b>這兩個</b>
7	<b>326</b>	<b>並不是</b>	<b>295</b>	<b>的一個</b>
8	<b>321</b>	<b>的過程中</b>	<b>293</b>	<b>指的是</b>
9	318	另一個	<b>285</b>	<b>很大的</b>
10	<b>307</b>	<b>本研究的</b>	<b>274</b>	<b>的過程中</b>
11	<b>306</b>	<b>的一種</b>	<b>269</b>	<b>之間的關係</b>
12	<b>275</b>	<b>的一個</b>	<b>246</b>	<b>所提出的</b>
13	<b>273</b>	<b>有一個</b>	237	一個人
14	<b>270</b>	<b>的角度來</b>	<b>236</b>	<b>的角度來</b>
15	<b>249</b>	<b>之間的關係</b>	231	最高的
16	232	同一個	227	個人的
17	223	所說的	<b>215</b>	<b>有一個</b>
18	220	這一類	210	有兩個
19	218	有不同的	<b>203</b>	<b>的一種</b>
20	<b>210</b>	<b>很大的</b>	<b>192</b>	<b>並不是</b>

筆者進一步整理三詞到五詞詞串之語料，將語料分為共有及獨有詞串，藉以觀察母語者和學習者的詞串差異，得到結果如表 4-5。從下表 4-5 我們得知以下兩點：第一點是獨有詞串的差異。母語者三、四詞詞串數量分別為 226、18；學習者則是 466、26，從詞串數量比較，學習者三詞及四詞獨有詞串之數量是母語者的兩倍，這可能是因為學習者受限於教科書和華語課堂的語言 (Amirian, Ketabi, & Eshaghi, 2013)，故多次使用特定的表達方式，而母語者語彙多元，會分散使用不同的表達方式。第二點，從百分比的角度，我們看到共有詞串在詞串總數所佔之比例隨著長度變長而下降，顯示越長的詞串應該越不容易整塊處理、記憶。不過，我們注意到學習者五詞共有詞串的比例略高於四

<sup>14</sup> 粗體為母語者及學習者共有詞串

詞共有詞串約 1%，可能是由於五詞詞串總數少，使共、獨有詞串差距變小的緣故。

表 4-5 共有及獨有詞串統計數據

		母語者		學習者	
		數量	百分比	數量	百分比
三詞	共有	362	61.56%	362	44.80%
	獨有	226	38.44%	466	57.67%
	總數	588	100.00%	808	100.00%
四詞	共有	26	59.09%	26	33.77%
	獨有	18	40.91%	51	66.23%
	總數	44	100.00%	77	100.00%
五詞	共有	9	40.91%	9	34.62%
	獨有	13	59.09%	17	65.38%
	總數	22	100.00%	26	100.00%

#### 4.2.2 共有詞串分析

在 3.4 筆者說明整理母語者與學習者的共有詞串，並以表 3-12、13 呈現母語者及學習者共有詞串的 LLR 分析結果，本節將分別討論學習者在三詞、四詞、五詞多用以及少用的詞串。

先從數量的角度來觀察共有詞串的分布，從表 4-6 我們可以看到母語者和學習者在近半數的共有詞串使用頻率上沒有顯著差異，在有顯著差異的詞串中，有學習者用得比母語者頻率少或是多的不同。學習者用的比較多的詞串我們呈現在表 4-7 中。

表 4-6 L1、L2 共有詞串顯著性分析結果分布

共有詞串				
	L2 多用	L2 少用	無差異	總數
三詞詞串	55(15.19%)	139(38.40%)	168(46.41%)	362(100%)
四詞詞串	5(19.23%)	10(38.46%)	11(42.31%)	26(100%)
五詞詞串	1(11.11%)	1(11.11%)	7(77.78%)	9(100%)

表 4-7 列出學習者多用的三詞、四詞及五詞詞串，表內詞串由左至右依照 LLR 值排序，第一列最左側詞串「受試者」是學習者多用的詞串中 LLR 值最大的詞串、最後一列最右側的詞「最高的」是 LLR 值最小的詞串，LLR 值越大表示兩個群體的差異越大。也就是說，在表 4-7 的詞串中，學習者使用「受試者」的頻率與母語者相差最大。

表 四-7 學習者多用之詞串

詞串長度	詞串
三詞	受試者、在課堂上、兩種語言、試者的、華語學習者、學習華語的、學習者的、 <b>研究者認為</b> 、漢語學習者、者的需求、 <b>研究者在</b> 、 <b>本論文的</b> 、 <b>的研究結果</b> 、對自己的、學習者使用、華語母語者、學生的學習、語言與文化、的受試、第二語言習得、教材編寫原則、第二語言教學、 <b>本研究的</b> 、 <b>所使用的</b> 、學習上的、 <b>研究者將</b> 、 <b>因此研究者</b> 、華語教材的、 <b>的個人</b> 、教材編寫的、的華語教材、母語者的、者在學習、的華語教學、 <b>另外一個</b> 、教學上的、者的母語、學習過程中、學習者所、第二節研究、華語教學上、一種語言、是學習者、教學過程中、 <b>在這種</b> 、 <b>表達自己的</b> 、為母語的、的教學方法、 <b>上的差異</b> 、學習者為、的文化內容、 <b>在本研究</b> 、的華語教師、使用的教材、最高的
四詞	受試者的、學習者的需求、的受試者、學習者的母語、學習者在學習
五詞	以學習者為中心

為方便閱讀，筆者以粗體標示非主題相關（如含有「受試者」、「學習者」等詞彙）之詞串，排除主題領域相關的詞串以後，剩下的詞串數量不多，但是仍能發現多數的多用詞串不約而同有一個明顯的特徵。14 個粗體詞串中，有 8 個是指涉論文本身的詞串，其中 7 個含有「研究」一詞。初步來看，學習者較頻繁地以這些詞串指稱自己或是論文本身，如「研究者認為」、「研究者在」、「本論文的」，這些詞串直接闡述作者的想法、發現以及研究內容，如語料(16)-(18)所舉。

(16) 「**研究者認為**上述的上課方式皆不同，因此需要編寫針對各個教學方式的兒童華語教材」 L2-5

(17) 「**研究者在**這次研究發現日籍學生對教師的依賴度仍是相當高」 L2-52

(18) 「研究者提到**本論文的**試教主要會應用《透明字地部首》來教小朋友」 L2-20



表 4-8 則是學習者少用的詞串，對照表 4-7，我們能發現學習者少用的詞串中，跟學習者多用的詞串相比，有很大部分的比例為非主題相關的詞串，我們一樣以粗體標示出這些詞串。

表 四-8 學習者少用之詞串

詞串 長度	詞串
三詞	<p>受訪者、實用視聽華語、句中的、讓學習者、<b>這一類</b>、說話者的、<b>是一個</b>、遠東生活華語、<b>另一個</b>、<b>指的是</b>、<b>了一個</b>、<b>的情況下</b>、<b>在句中</b>、<b>是一種</b>、教材中的、<b>並不是</b>、<b>較高的</b>、的語言形式、的語用、的教學活動、學習者對於、<b>值得注意的</b>、在同一、高級學習者、<b>的一種</b>、<b>稱之為</b>、<b>注意的是</b>、教學語法的、<b>所做的</b>、<b>有不同的</b>、<b>在這個</b>、<b>所提出的</b>、現代漢語中、<b>所說的</b>、的<b>研究中</b>、<b>表所示</b>、教學中的、華語教學中、中的文化、<b>的兩個</b>、的教學排序、<b>較多的</b>、<b>綜上所</b>、<b>上所述</b>、<b>的另一</b>、<b>是兩個</b>、訪者的、這句話、<b>另一種</b>、<b>一對一</b>、<b>什麼樣的</b>、學習者能、的使用頻率、<b>有一個</b>、第二章文獻、者來說、的語法功能、語言學習者、<b>在不同的</b>、<b>本研究之</b>、<b>所呈現的</b>、<b>同一個</b>、每一課、<b>角度來看</b>、<b>所指的</b>、<b>更好的</b>、<b>的過程中</b>、章文獻探討、引導學習者、在教學上、中<b>的一</b>、幫助學習者、能願動詞、<b>所需的</b>、的教學方式、在<b>一個</b>、<b>兩大類</b>、一書中、學習者而言、位學習者、學習者可以、<b>兩者的</b>、與學習者、初級學習者、<b>本研究以</b>、一個語言、<b>最重要的</b>、<b>的角度來</b>、<b>這兩者</b>、<b>較低的</b>、<b>頻率最高</b>、的語法意義、<b>種不同的</b>、一個詞、<b>而產生的</b>、<b>結構上的</b>、的教學模式、表示動作的、<b>也是一</b>、<b>可以說是</b>、外華語教學、<b>一個是</b>、<b>因此本研究</b>、的語法結構、<b>有可能是</b>、在第三章、<b>分為兩類</b>、的<b>相關研究</b>、在學習者、在教材中、的語義關係、<b>本研究將</b>、學習者有、<b>每一個</b>、位受訪、<b>要表達的</b>、<b>來說是</b>、<b>重要的是</b>、的<b>研究成果</b>、<b>這三個</b>、<b>本研究中</b>、的語料中、學習者來、<b>所產生的</b>、<b>相對應的</b>、<b>這兩個</b>、<b>扮演的角色</b>、<b>這個問題</b>、外籍學習者、的語義功能、外漢語教學、日常生活中、<b>這也是</b>、<b>者之間的</b>、<b>而不是</b>、的<b>對應關係</b>、在課堂中、<b>由此可</b>、中的動詞</p>
四詞	<p><b>值得注意的是</b>、<b>綜上所述</b>、受訪者的、<b>的角度來看</b>、第二語言學習者、第二章文獻探討、位受訪者、學習者來說、對外華語教學、對外漢語教學</p>
五詞	<p>聽說讀寫四</p>

若將這些粗體詞串大略分類，可以分成六大類：

- A. 分類比較型。如：「另一個」、「較高的」、「的一種」。

- B. 概念說明型，如：「稱之為」、「指的是」、「要表達的」、「並不是」、「來說是」、「可以說是」、「有可能是」等。
- C. 研究闡述型，像：「本研究以」、「本研究將」、「本研究中」等。
- D. 吸引讀者型，如：「值得注意的是」與「注意的是」。
- E. 範圍限定型，如：「的情況下」、「的角度來看」、「…而言」。
- F. 結論引導型，如：「綜上所述」、「由此可」

這六大類中 A、B、C 三類和下一節要討論的功能分析重疊，故在本節先不詳述，以下筆者將敘述 D、E、F 三類詞串。首先，我們從表 4-8 可以發現，學習者似乎較母語者少用引起讀者注意的詞串如「值得注意的是」、「注意的是」，而對照 3.4 中表 3-11 也發現五詞詞串「值得一提的是」為母語者的獨有詞串，母語者經常使用此結構標示討論的重點，如(19)、(20)，但這三種吸引讀者的形式學習者都不太常使用。

- (19) 「**值得一提的是**，與聲調記憶測驗的結果相比，受試者的熟習度測驗得分普遍偏高」 L1-22
- (20) 「**然而，必須注意的是**，雖然課本推出了新的版本，但是…」 L1-80

同樣地，結論型詞串如：「綜上所述」、「由此可」學習者也用得比母語者少。「綜上所述」時常用在結論段的開頭，以總結討論之主題；「由此可」也是解釋數據時常見的詞串。可以看到「由此可」除了與「知」搭配以外，母語者共使用 17 種不同的詞彙與「由此可」搭配，總次數 30 次，分布在 49 個文本中；學習者則搭配 14 種詞，總次數 21 次，分布在 33 個文本內。不論詞種、分布的文本數，L2 都比 L1 少，可以看出學習者有集中使用、詞語豐富性較低的特性。

表 四-9 「由此可」右邊 1 個詞的檢索結果

由此可*							
母語者				學習者			
搭配詞	次數	搭配詞	次數	搭配詞	次數	搭配詞	次數
知	146	檢視	1	知	145	取得	1
看出	5	歸納出	1	看出	2	推出	1
推知	3	比較出	1	推知	1	理解	1
推論	3	看到	1	推論	1	看	1
證明	3	說明	1	證明	1	觀察到	1
發現	1	證	1	發現	3	推斷	1
以	1	間接	1	再	4		
察覺	1	推測	4	突出	2		
推斷	1			了解	1		

以上提及學習者用得比母語者少的詞串如：「綜上所述」、「由此可」、「值得注意的是」都出現在中文人文社會學科的論文中（劉貞好、陳浩然、楊惠媚，2017），也就是說，這些詞串在母語者學術論文中都是相當常見的，可見學習者也用，只是用得沒母語者多，未來或許在教學上能訓練學生更有意識地使用。

自學習者詞串表中，我們也能觀察到學習者有集中使用特定詞串的傾向及錯用詞串的現象。首先，儘管表 4-8 顯示學習者「的情況下」用得比母語者少，然而對比 3.4 中表 3-11「在這種情況下」出現在學習者五詞詞串中，於是我們試圖在語料庫中檢索「在\*\*情況下」，並排序前五高頻的搭配詞成表 4-10。如表所示，「在\*\*情況下」母語者有 94 筆，搭配「這種」佔其中 21 筆；學習者「在\*\*情況下」共 97 筆，搭配「這種」有 47 筆，可見學習者有集中使用特定詞串的傾向，母語者則相對較和緩。

表 四-10 「在\*\*情況下」\*\*前五高頻搭配詞

母語者	次數	學習者	次數
在**情況下	94	在**情況下	97
這種	21	這種	47
這樣的	11	一般的	9
很熟	8	這樣的	5
一般的	7	某種	4
某種	4	這個	4

表 4-8 中的詞串「學習者而言」也相當值得注意，就「而言」單詞檢索，母語者語料有 964 筆、學習者語料 983 筆，在使用次數與頻率方面學習者不亞於母語者，不過，從語料(21) - (24)可以看到學習者有錯用「而言」的語料。

(21) 「從以上四點而言，可看出漫畫是一種非常令人容易接近的媒介」。

L2-22

(22) 「.40 到 .60 代表題項與總分之間則有中度相關。後者而言，若該題項刪除後的內部一致性  $\alpha$  係數比整體的  $\alpha$  係數值高」 L2-76

(23) 「根據裝在擔子裡的責任感及心理負擔的量，說『擔子輕』、『擔子重』。進一步而言，中文是將客體擔子視為容器進行隱喻」 L2-72

(24) 「這雖然是諷刺某些人利用老鄉關係搞不正之風，但歸根究底而言，還是源自於中國人的家族觀念和鄉土觀念。」 L2-66

以上四筆語料的「而言」似乎都不太自然，為了探究「而言」的功能，我們將《當代中文課程》中「而言」的結構與定義整理成表 4-11。

表 四-11 《當代中文課程》收錄之「而言」解釋

結構	定義
就...而言	第四冊第十課 as far as are concerned
對於...而言	第四冊第十一課 as far as is concerned- marks the focus of the entire sentence

不論是母語者還是學習者，大部分「而言」或「來說」的語料都是以上述的結構呈現，但是錯用的語料(21)-(24)皆無使用這些結構。其中(21)、(22)的結構是「NP+而言」，若套用「就...而言」似乎就是正確的句子，如下所示：

(21)-1 「就從以上四點而言，可看出漫畫是一種非常令人容易接近的媒介」。  
L2-22

(22)-1 「.40到.60代表題項與總分之間則有中度相關。就後者而言，若該題項刪除後的內部一致性 $\alpha$ 係數比整體的 $\alpha$ 係數值高」  
L2-76

至於(23)、(24)，「而言」前為「近一步」與「歸根究底」，兩筆語料皆欲延展討論之主題，不過「而言」多用於討論一個整體中的部分，如(25)、(26)。(25)的主題為討論學習者，在該筆語料中限定某種條件的學習者，為整體一部分的關係；而(26)的主題為「吧」的用法，語料中以功能的角度討論「吧」，亦為為整體一部分的關係。

(25) 「一對一遠距聽力教學對三位處在不同國家學習的學習者而言」 L1-3

(26) 「就功能而言，「!?:wkan」符合了「吧」的承諾用法，」 L1-75

從這些語料來看，「而言」僅能說明某一整體中部分的關係，那麼延續討論主題的(23)、(24)似乎就不適合使用「而言」。然而，將「而言」改為「來看」就能使其通順許多。

(23)-1 「根據裝在擔子裡的責任感及心理負擔的量，說『擔子輕』、『擔子重』。進一步來看，中文是將客體擔子視為容器進行隱喻」 L2-72

(24)-1 「這雖然是諷刺某些人利用老鄉關係搞不正之風，但歸根究底來看，還是源自於中國人的家族觀念和鄉土觀念。」  
L2-66

本節藉由 LLR 顯著性分析討論母語者與學習者使用三詞、四詞、五詞詞串的差異。首先，從詞串的總數來看我們發現學習者有多用固定詞串的趨勢，學習者各長度詞串之總數皆高於母語者，這很可能是學習者受到語言課程的影響，使學習者的語言較單一化。另外，我們進一步將篩選出之詞串分類為共有以及獨有詞串之後，我們發現學習者的獨有詞串比例高於母語者，表示很多學習者使用的表達法在母語者中不那麼常見。最後，從共有詞串的顯著性分析可



以看到雖然共同使用某些詞串，但在使用頻率上還是有多用和少用的差異，其中也不乏有學習者錯用的語料。

### 4.3 功能分析

上一節，我們發現學習者較母語者少用一些具有文本組織功能的詞串，而在本節，筆者將針對詞串功能，透過功能分類比較母語者和學習者論文的文本架構和特色，並觀察兩個群體在功能類型分布上是否有差異。

承 3.4 所討論，三詞詞串是各詞串長度中數量適中、不至於造成人工處理負擔的長度，因此研究者分析三詞詞串的功能類型，先人工刪除與主題相關之詞串，再依照 Hyland (2008a)、Salazar (2011) 的功能類別定義分類本文篩選出的詞串。篩選後的詞串母語者共 229 個、學習者 272 個，表 4-12 呈現各功能類型的詞串數量、圖 4-2 呈現分布比例：

表 四-12 各功能類型詞串數量

主分類	子分類	L1	L2
研究導向 (research-oriented)	地點 (location)	0	0
	程序 (procedure)	17	27
	數量 (quantification)	12	14
	描述 (description)	22	38
	歸類 (grouping)	53	54
文本導向 (text-oriented)	附加 (additive)	7	8
	比較 (comparative)	29	29
	推論 (inferential)	30	35
	因果 (causative)	5	9
	組織 (structuring)	16	22
	架構 (framing)	6	6
	引述 (citation)	13	13
	概括化 (generalization)	0	0
	目的 (objective)	1	6
參與導向 (participant-oriented)	立場 (stance)	14	9
	吸引 (engagement)	3	2

致謝 (acknowledgment)	0	0
---------------------	---	---

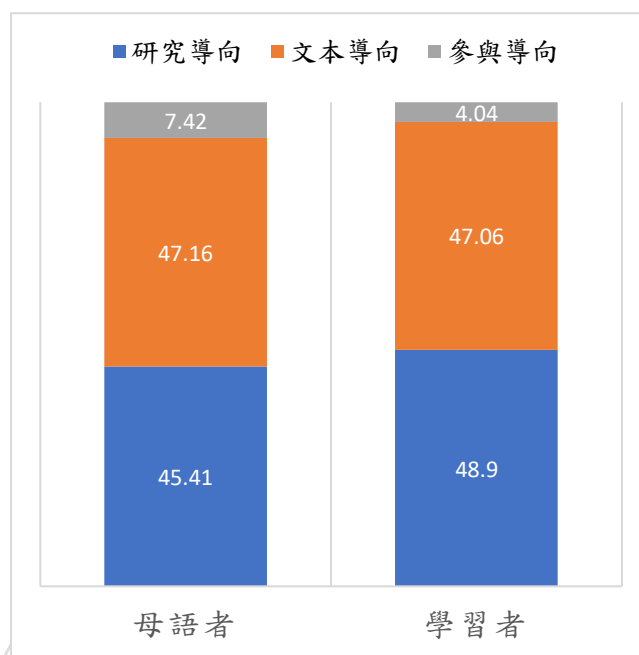


圖 4-2 母語者與學習者詞串分布比例

首先，表 4-12 中「地點」、「概括化」和「致謝」這三個分類的數量為零，由於本研究篩選出的詞串沒有這三個類型，下文分析中便不討論。而藉由圖 4-2，我們看到不論是母語者還是學習者，研究導向與文本導向的比例不但非常接近、將近佔了全部語料的半數，且明顯多於參與導向的詞串。母語者的文本導向詞串所佔比例稍高於研究導向；學習者則是研究導向略高於文本導向，不過兩大類約僅有 1% 的差距。最後，參與導向的比例則是三類中最低的，在母語者語料中約有 7%、學習者語料則僅有 4%。過往學術詞串調查的結果（劉貞好、陳浩然、楊惠媚，2017；Pan, Reppen, & Biber, 2016）與本文詞串呈現一致的分布，文本導向所佔比例最高、研究導向次之、參與導向最低。因此我們可以推測，不論語言為何，學術寫作中架構文本結構和訊息傳達相當重要。由於學術寫作重視研究敘述和程序描寫，故研究導向與文本導向的詞串比例相當高，而參與導向在三者比例中最低的可能原因是學術研究需保持中立，立場表達和吸引讀者注意的詞串就相對少數，不過這類型的詞串在口說或是課堂教學時就多出許多 (Biber et al., 1999)。接下來本文將依照三大功能分類依序討論母語者和學習者的詞串功能差異。

### 4.3.1 研究導向

研究導向詞串的定義為「輔助作者建構在現實世界的活動和經驗」(Help writers to structure their activities and experiences of the real world) (Hyland, 2008a; Salazar, 2011)，這類型的詞串能夠輔助作者帶出跟研究相關的資訊。在 Salazar (2011) 的分類中，他將研究導向詞串分為五個次類：地點、程序、數量、描述、歸類。地點型詞串指出位置、端點和方向，程序詞串敘述研究步驟；數量詞串表達文本內的計量；描述類的詞串大部分為修飾語；歸類為表群組、種類的詞串，讓作者清楚描述內容的分類。在這五個類別中，由於本研究未篩選出地點類詞串，故下表無該分類。

表 4-13 顯示各分類的種類比例及使用總頻次。比較種類的分布能發現，學習者的詞串分布較平均，種類最多的歸類型詞串和最少的數量型詞串相差 30% 左右，母語者種類最多及最少的類型也分別為歸類和數量型，但是兩者相差達 40%，母語者歸類型詞串の種類佔了全部的一半以上，由此可知母語者在歸類時，很多時候使用詞串。母語者與學習者第二和第三多的類別分別為描述型與程序型詞串。不過使用的比例稍有不同，學習者在這兩個分類的詞串種類皆高於母語者。

表 4-13 研究導向詞串分布

研究導向	母語者		學習者	
	種類比例	頻次	種類比例	頻次
歸類 (grouping)	50.96%	5772	40.46%	4995
程序 (procedure)	16.35%	1536	20.30%	2442
描述 (description)	21.25%	2145	28.57%	3151
數量 (quantification)	11.54%	1295	10.53%	1442
總數	100%	10748	100%	12030

#### 4.3.1.1 歸類型詞串

「歸類」型詞串的功能在於指出群組、種類、部分和順序。本類詞串の種類及頻次在母語者和學習者語料中都最高的，在母語者語料中更是超過一半。觀察表 4-14 的詞串可以發現這類詞串在結構上大多包含一個數詞及一個量詞，

也有一部分的詞串包含定詞「這」。因為能使用的詞彙很有限，造成類似組合的詞串出現次數高。

表 4-14 歸類型詞串<sup>15</sup>

歸類 (grouping) : 表群組、種類、部分與順序	
母語者	是一個、是一種、這兩個、同一個、這一類、每一個、兩者的、的各種、在這個、這兩種、這三個、這類的、一類的、這個問題、在不同的、一個是、是兩個、每個人、一種是、的這個、這兩者、兩大類、為一個、其中一個、這件事、一類是、分為兩類、各種不同、某一個、則是以、這三種、兩者在、分為三類、的這種、這方面的、兩個部分、哪一種、在這種、的第一個、這兩類、這四個
學習者	是一個、是一種、這兩個、同一個、每一個、兩者的、這兩種、這三個、一種是、兩個選項、某一個、這個問題、在這種、這種情況、分為兩種、每個人、在這個、一個是、哪一種、這方面的、三個部分、三個階段、分為三個、等方面的、分為三種、各方面的、在不同的、這四個、其中一個、各種不同、的第一個、哪一個、這一類、這三種、分為兩個、這兩者、這種現象、是兩個、不同程度的、兩大類、兩個方面、分為兩類

兩語料庫中前三高頻的「歸類」型詞串相同，分別為「是一個」、「是一種」、「這兩個」，不過使用頻率上有很大的差異。「是一個」和「是一種」是母語者全部詞串中最高頻的詞串（L1/L2 是一個：763/460；是一種：573/390），且次數高了學習者 1.5 倍，這兩個詞串多用來解釋概念，如(27)-(30)所示，這兩個詞串特別高頻表示母語者常以它們說明及解釋概念。

- (27) 「對他們而言會**是一個**一吐未盡之言的機會...」 L1-64
- (28) 「由於拒絕言語行為本質上**是一種**威脅面子的言語行為...」 L1-48
- (29)\* 「主要的原因**是一個** ESP 環境中的「學習環境」與工作環境的條件有所不同...」 L2-4
- (30) 「覺得它們**是一種**帶有濃厚中國風味的習慣用法...」 L2-14

我們從表 4-14 能注意到，母語者的獨有詞串不少都含有量詞<sup>16</sup>「類」，而學習者相對較少，量詞一直以來都是華語教學研究的討論重點，因此本研究也

<sup>15</sup> 粗體詞串為母語者與學習者的共有詞串

<sup>16</sup> 本文量詞判定方式依照中央研究院斷詞系統，斷詞後詞類為(Nf)，在本研究中即作為量詞。詳細詞類標記方法請見《中央研究院平衡語料庫的內容與說明（修訂版）》  
網址：[http://rocling.iis.sinica.edu.tw/CKIP/tr/9804\\_2013.pdf](http://rocling.iis.sinica.edu.tw/CKIP/tr/9804_2013.pdf)

回到文本，使用語料庫觀察母語者和學習者使用的量詞。我們將語料庫檢索結果整理為表 4-15，母語者使用共 245 種量詞、頻次為 117,577 次；學習者使用 195 種量詞、頻次為 113,812 次，雖然學習者使用的量詞類型不如母語者多，但透過 log-likelihood 檢定 (LLR=693.33,  $p < 0.001$ ) 顯示學習者使用量詞詞組的頻次顯著高於母語者。我們進一步比較 L1 和 L2 前 15 個高頻量詞的頻率，並將結果呈現在表 4-16，發現「類」、「篇」、「條」、「冊」、「段」、「項」、「套」是學習者用得顯著較少的量詞。豐富的量詞系統是漢語的特色之一，對於學習者也是學習難點，儘管是高級學習者，對於量詞種類的豐富度仍和母語者有差異。

表 四-15 學習者和母語者量詞使用結果

	母語者	學習者
類型	245	195
頻次	117577	113812

表 四-16 高頻量詞顯著性比較

量詞	L1 使用頻率	L2 使用頻率	Log- likelihood	Sig.	
類	3661	1843	347.89	0.000***	+
篇	727	169	287.91	0.000***	+
條	676	177	234.38	0.000***	+
冊	1157	439	226.64	0.000***	+
段	416	146	183.95	0.000***	+
項	1968	1263	60.24	0.000***	+
套	860	624	8.72	0.003**	+
點	442	334	2.49	0.115	+
筆	609	482	1.24	0.266	+
年	1009	2085	582.15	0.000***	-
些	727	1203	210.97	0.000***	-
位	1701	2254	196.66	0.000***	-
節	2207	2755	185.83	0.000***	-
個	17735	17256	168.02	0.000***	-
種	7306	7606	158.57	0.000***	-
聲	1656	2089	147.88	0.000***	-
本	795	947	50.52	0.000***	-
組	580	654	25.23	0.000***	-



次	1996	1947	19.63	0.000***	-
句	1820	1761	15.84	0.000***	-

另外，從表 4-16 也看到學習者相對多用的量詞，如「個」。研究者檢索語料庫中學習者使用「一＋量詞」短語的情形，發現不少偏誤：

(31) 「C 君希望，於大學畢業之後可以在中國大陸或台灣找一份工作...」

L2-1

(32) 「例如「寫」的教科書裡，為了練習某個句型不讓學習者漫無目的寫一個句子或作文，而讓他們做筆記、寫感謝信等。」

L2-10

(33) 「這個病句是筆者在任教期間所搜集到學生作業上的一個病句...」

L2-8

(34) 「譯者沒有將「明」翻譯出來，反而換上一個動詞「soi」-『照』」。

L2-9

由於漢語「一＋量詞」的結構除了計數的功能外，在語篇上也有參照的功能，學習者使用參照功能的時候常出現病句 (Huang, 2018)，尤其英語母語者最容易出現這類型的偏誤 (張莉萍, 2014; Li, 2004)，(31)-(34)為本研究的語料，我們能觀察到高級學習者仍有多用的問題。Huang (2018, pp. 163) 指出「一＋量詞」的使用取決於被修飾名詞的指稱性，在泛指名詞 (generic noun) 或非指稱名詞 (nonreferential noun) 之前加上「一＋量詞」會造成偏誤。例(31)、(32)中「工作」、「句子」無特指某類型的工作或句子，是泛指用法，不需加「一＋量詞」；(33)、(34)則是因為被修飾的名詞「病句」、「動詞『soi』」指涉的對象在前後文很清楚，因此也不使用「一＋量詞」來修飾。學習者多用的原因可能是受本身母語的影響，也可能是因為華語教科書缺少「一＋量詞」使用時機的說明 (Li, 2004)，從上述四例句來看，儘管是高級學習者仍會有多用這個結構的現象，因此關於「一＋量詞」的功能在未來的課程和教材編纂中應該加強。

#### 4.3.1.2 程序型詞串

程序型詞串的功能為敘述事件、行動及實驗方法，此類詞串能敘述研究步驟和方法，多含有和研究本身相關的詞彙如「本研究」、「本節」、「研究

者」，與和過程及方法有關的詞彙如「過程」、「方式」、「方法」。對照表 4-13 可以看到學習者程序型詞串的種類比例比母語者高一些，母語者有 16.5% 屬於程序型、學習者則是 20.45%，log-likelihood 檢定也呈現學習者使用程序型詞串的頻率顯著高於母語者（LLR=85.10,  $p < 0.001$ ），從以上兩點來看，我們可以說學習者在敘述研究程序時，使用的表達方式較集中，這個結論也呼應到本文在 4.2.2 所提及學習者似乎相對較重視研究架構的現象。

表 四-17 程序型詞串

程序 (procedure)：敘述事件、行動及實驗方法	
母語者	的過程中、所使用的、本研究將、的方式來、本研究以、所討論的、都是以、所需的、本章將、本節將、所設計的、由兩個、過程中的、研究者在、研究者將、的方式呈現、的方式進行
學習者	所使用的、的過程中、的研究對象、研究者的、研究者在、研究者將、的研究方法、的方式來、本研究將、過程中的、本節將、所設計的、研究者以、研究者所、過程中所、最後一個、本章將、本研究以、所需要的、本論文所、的研究問題、為研究對象、的研究者、的方式進行、分析本研究、所需的、透過文獻探討

母語者程序型詞串共 17 個、學習者有 27 個，共有詞串 13 個，由此看來，母語者和學習者使用的程序型詞串類型很相似，母語者僅有 4 個獨有詞串、學習者則有 14 個。學習者的獨有詞串大多數都包含「研究」一詞，除了「研究者」以外，其他還有如「的研究問題」、「的研究方法」、「為研究對象」等，表示學習者多以直接點明研究本身的方式傳達研究程序，我們在 4.2.2 討論學習者多用的詞串時，也發現相同的現象。

(35) 「在觀察個別隱喻詞彙的過程中，我們也注意到」 L1-74

(36) 「可以清楚看出三組受試對象所使用的回應讚美狀況」 L2-75

(37) 「我們將利用同樣的方式來檢視『人家』」 L1-53

(38) 「問卷因本論文的研究對象並非單一的個人，而是群體」 L2-44

「的方式來」也是程序型詞串中高頻的詞串之一，一般而言，「的方式來」前面會有介詞引介，因此筆者檢索「的方式來」左邊的搭配詞，藉此觀察母語者和學習者使用什麼介詞搭配「的方式來」。Antconc 搭配功能結果顯示，

母語者語料共有 96 筆「的方式來」，共有 12 種不同的搭配詞；學習者共有 92 筆，使用 10 種搭配詞，如表 4-18 所列。

表 4-18 「的方式來」之搭配詞彙

	搭配詞	的方式來
母語者	用／以／採用／透過／使用／利用／藉由／運用 ／採取／由／藉／借用	的方式來
學習者	用／以／採用／透過／使用／利用／藉由／運用 ／採／照／通過	的方式來

透過檢索與「的方式來」搭配介詞，能看到母語者使用的搭配詞比學習者多一些。另外，從檢索「的方式來」之搭配詞，我們可以看到中文的句型結構經常不會緊黏在一起，一個結構的元素間能插入不同長度的短語，如(39)「透過」和「的方式來」中間僅隔一個詞「談話」，但(40)中「照」與「的方式來」中間卻隔了一個很長的短語。由於這個特性的緣故，讓我們很難像英語詞串研究一樣，以語料檢索工具篩選出定式或句型，僅能以人工方式逐筆觀察。儘管單看漢語的詞串似乎不如英語能夠提供實用句型參考，不過我們也能夠透過逐筆檢索觀察定式的搭配詞，進一步提供學習者更多的詞彙搭配選擇、增加詞彙多樣性。

(39) 「必須讓學生大量的**透過談話的方式來**熟悉科學語言」 L1-5

(40) 「我們往往**照**老師給作業的內容或一般教材**的方式來**進行練習」

L2-79

#### 4.3.1.3 描述型詞串

描述型詞串的功能為說明質量、程度以及存在，本類詞串有不少都是以「的」結尾的修飾語。學習者使用比較多種的描述型詞串，且使用頻率也顯著高於母語者 (LLR=53.17,  $p<0.001$ )。描述型詞串中的詞彙難度不高，大部分是很常見的修飾語。如(44) 包含詞串「很大的」和「有很大」，經過檢索我們發現母語者使用「有很大的」共 115 筆、學習者 138 筆，是相當基礎且常用的詞串，在觀察「有很大的」語料中我們也發現「對 N 有很大的幫助」是學習者與母語者都常使用的結構。

表 四-19 描述型詞串

描述 (description) : 說明質量、程度以及存在	
母語者	很大的、個人的、什麼樣的、稱之為、日常生活中、有很大、有一定的、有意義的、中出現的、很好的、所處的、在日常生活、對自己的、很重要的、一個完整、就像是、表達自己的、一個很、有這樣的、一個重要、常出現的
學習者	很大的、個人的、對自己的、有很大、中出現的、有一定的、日常生活中、的個人、有意義的、表達自己的、什麼樣的、的日常生活、很重要的、有幫助的、在日常生活、非常重要的、稱之為、很有幫助、一個很、一個重要、不懂的、大的幫助、常使用的、有自己的、大的影響、所存在的、聽不懂、日常生活的、不太會、在一定的、所得到的、重要的角色、常出現的、不好的、太大的、有系統的

另外，我們也發現「自己」在描述型詞串中也相當常見，主要是因為在教學研究和學習者分析研究中常大量討論被觀察對象的回饋，因此許多詞串都有「自己」一詞。「就像是」是描述型詞串中較特殊的詞串，這個詞串後銜接一個生活化的敘述來解釋主題，使描繪之主題更加鮮明，如 (44)。母語者和學習者除了使用「就像是」的頻率相差頗大 (L1:L2 = 61:5) 以外，我們也發現學習者僅有的五筆語料中有一筆錯用語料。「就像是」用於敘述兩事件是相同或相似的概念，但 (45) 的作者將「就像是」誤用成舉例的功能，宜將「就像是」換成「例如」或是「像是」。

- (41) 「對學習者在學習語言上有很大的幫助」 L1-9
- (42) 「語境在詞彙習得上扮演很重要的角色」 L2-61
- (43) 「就是說導遊對自己的工作、所導覽的地方和觀光客該非常熱情」 L2-50
- (44) 「但是不斷的練習**就像是**提供種子成長的養分」 L1-73
- (45)\* 「所蒐集的資料也是由人們的自然行為而來，**就像是**談話、拜訪、觀看、飲食等等的自然行為。」 L2-42

雖然自學習者語料篩選出的詞串種類較多且使用頻率較高，但也可能代表學習者使用這些詞串的重複率高，而母語者使用不同詞彙搭配修飾，然而因頻率分散，故沒被篩選自詞串表內。舉例來說，筆者以 AntConc 搭配詞功能檢索高頻描述型詞串「一個重要」，發現「因素」和「部分」為高頻搭配詞。接著，再分別檢索「因素」和「部分」的搭配詞，從結果挑選出「重要」的近義詞。從表 4-20 的檢索結果可以觀察到學習者似乎集中使用「主要」、「重要」



作為修飾語，而母語者使用不同搭配詞的頻率相較之下較分散。這個結果呼應了學習者有高度集中使用某些詞串的傾向，母語者詞串種類較少是因為使用的詞彙較豐富、頻率相對低而未達本文之篩選門檻所致。

表 四-20 「因素」、「部分」左邊搭配詞檢索結果

因素		部分	
母語者	學習者	母語者	學習者
重要(27)	主要(50)	重要(44)	重要(68)
主要(22)	重要(49)	主要(27)	主要(28)
關鍵(17)	關鍵(12)	核心(9)	核心(6)
決定性(7)		關鍵(8)	不可或缺(6)
		不可或缺(4)	

#### 4.3.1.4 數量型詞串

數量型詞串表計量、數量、比例及相關的變化，數量型詞串和程序型詞串在研究導向中佔的比例相近，母語者和學習者有16個共有詞串，表示母語者和學習者使用的類型差異也不大。然而，以頻率來看，母語者的使用頻率顯著高於學習者（LLR=17.36,  $p < 0.001$ ）。這類型的詞串多數為「有」和「的」與數詞的組合，也有少數詞串和比例描述有關。

表 四-21 數量型詞串

數量 (quantification)：表計量、數量、比例及相關的變化	
母語者	有一個、一個人、有兩個、有兩種、在一個、頻率最高、比例最高、兩個以上、一本書、有一種、一個問題、一種狀態、一件事、的一種、的一個、的兩個、的一部分、的三個、的各個、的幾個、的一環、的一部份、的各項、的四個、的個人
學習者	一個人、有一個、有兩個、有兩種、一本書、有三個、兩個以上、在一個、頻率最高、兩次的、佔的比例、所佔的、有一位、有三種、比率最高、的一個、的一種、的各種、的兩個、的三個、的一部分、的兩種、的四個、兩個部分、的三種、的各個、每一種

語料(46)-(49)顯示學習者和母語者使用數量型詞串：

- (46) 「一個組合裡**有一個**不單用的成分就認為這個組合是詞」 L1-24
- (47) 「當『人家』出現在一個新話輪的開端時」 L1-53
- (48) 「但『介紹』出現**頻率最高**」 L2-5



(49) 「『模仿』寫作的部分應是必要，但所佔的比例不適合太高」 L2-63

不同研究對於數量型詞串的描述有不同的結論，Amirian et al. (2013) 的研究結果指出母語者和學習者使用的數量型詞串比例偏低；然而 Cortes (2004) 指出學術寫作生手缺乏使用數量型詞串。因此，筆者將本文的研究結果與劉貞妤、陳浩然、楊惠媚 (2017) 的研究相比，在其研究中，三詞詞串共195筆，數量型僅有7筆，由此推測在中文研究中不論是專家還是生手，使用數量型詞串的比例可能都不高。欲了解為什麼會有這樣的的不同，我們進一步觀察中英文數量型詞串的形態差異，表4-22列出Cortes (2004) 歸類為數量型之詞串。

表 四-22 Cortes (2004) 之數量型詞串

數量型詞串	
<b>Cortes (2004)</b>	<p>a large number of (大量的...), a measure of the (一定程度的...), a wide range of (廣泛的...), an increase in the (...的增加), an order of magnitude (量級的...), in the number of (的數量...), of the variance of (的差異...), on the number of, (的數量上...) on the order of (按...的順序), one of the most (最...的), the average number of (的平均數), the magnitude of the (的重要性), the mean number of (的平均值), the number of individuals (個體的數量), the number of species (物種的數量), the rest of the (剩下的...), the total number of (的總數...), for each of the (每一個<sup>17</sup>), each of the three (每三個)</p> <p><b>Statistical:</b> was positively correlated with (成正相關), of the distribution of (的分布), not significantly different from (沒有顯著不同), did not differ significantly (沒有顯著差異)</p>

就詞串意義而言，Cortes (2004) 的詞串對應的中文翻譯大部分為二詞詞串，本研究分析之詞串長度為三詞，故未被篩選出。另外一部分的詞串多數與統計相關，這些詞串在本文蒐集的語料中很少見，這樣的差異可能是因為文本所屬學科的不同。Cortes (2004) 所蒐集的語料為歷史學和生物學，生物學屬自然科學，使用統計、數量描述的機率可能較高，目前尚未分析自然科學文本的詞串研究，尚無法確定中文自然科學學術詞串的特徵。

<sup>17</sup> 「每一個」在本文屬於歸類型詞串

### 4.3.2 文本導向

文本導向的功能為傳遞及組織文本內的訊息。這類詞串除了對文章架構很重要，也能讓讀者閱讀時有脈絡可循、判斷文章重點。從表 4-23 能看出兩群體呈現相似的分佈，母語者和學習者前三高頻分別為：推論、比較和組織。

表 四-23 文本導向詞串分佈

文本導向	母語者		學習者	
	種類比例	頻次	種類比例	頻次
架構 (framing)	7.41%	1183	4.69%	807
目的 (objective)	0.93%	93	4.69%	351
引述 (citation)	12.04%	1437	10.16%	1117
比較 (comparative)	26.85%	2775	22.66%	2544
附加 (additive)	6.48%	830	6.25%	627
推論 (inferential)	27.78%	2861	27.34%	2862
因果 (causative)	4.63%	423	7.03%	671
組織 (structuring)	13.89%	1400	17.19%	1961
總數	100.00	11002	100.00	10940

母語者和學習者文本導向詞串種類最多的前三次類相同，分別為推論型、比較型和組織型；使用頻次最高的前兩類也分別同為推論型與比較型。儘管看似相似性很高，但以 Log-likelihood 分析能發現兩者在使用頻率上仍有差距。以下我們將分別討論母語者和學習者在各次類的異同。

#### 4.3.2.1 架構型詞串

架構型詞串的功能為藉由表明限定條件來提出論證，這類詞串說明論證的範圍。儘管架構型詞串之種類在文本導向中僅占 7%、使用頻次在文本導向八個次類中排第五，不過比較母語者及學習者之使用頻次，兩者差異是文本導向中最顯著的，母語者的使用架構型詞串的頻次顯著高於學習者 (LLR=147.63,  $p < 0.001$ )。再者，對照 4.2.1 表 4-4 高頻三詞詞串表，我們能看到母語者有高度集中使用「的情況下」的傾向。「的情況下」在母語者所有詞串中為第四高頻的詞串，使用頻率將近是學習者的兩倍 (L1:L2=360:180)。此外，本類典型詞串似乎相當突出，從表 4-24 我們可以看到母語者與學習者高頻詞的詞串為共有詞串，且也皆收錄在劉貞好、陳浩然、楊惠媚 (2017) 的研究中，其中「的情

況下」、「的角度來」、「角度來看」、「的基礎上」的頻率在全部詞串的前30%，由此可知這幾個詞串在學術論文中的使用頻率相當高。

表 四-24 架構型詞串

架構 (Framing)：藉由表明限定條件來提出論證	
母語者	的情況下、的角度來、角度來看、的基礎上、在這樣的、最基本的、為基礎的、的角度出發
學習者	的角度來、的情況下、角度來看、的基礎上、最基本的、的觀點來

從(50)-(53)顯示，除了「最基本的」是用來表示一個概念中基礎的要素以外，其他詞串都用來劃分話題的限定範圍，且範圍都在詞串前，使用的方式相當一致。本文先前提到「的情況下」在母語者語料中頻率相當高，筆者在語料中搜尋到以下幾個有相同意義的詞串，我們同時將母語者以及學習者使用的頻率呈現在括弧中 (L1/L2)：「的情況中」(16/0)、「的情況之下」

(11/20)、「的情況時」(10/4)、「的情況裡」(8/5)。這些詞串顯示「的情況」在學術論文中的普遍性，且母語者使用的頻率似乎比學習者的來得高許多。

(50) 「當對象為晚輩的情況下選用頻率極低」L1-50

(51) 「在對話能力訓練的基礎上，也要注重成段表達能力」L2-16

(52) 「若從功能的觀點來了解存現句，有助於了解整個華語句型」L2-37

(53) 「詞類及相關的語法標記為語法教學最基本的一環」L1-34

不少過往研究之結果都指出架構型詞串為文本導向類中佔最多數的次類 (Hyland, 2008a; Hyland, 2008b; Pan et al., 2016, Salazar, 2011)，Hyland (2008a) 更指出這類詞串在商業及語言學領域中的普遍性勝於自然科學類的研究，但本文架構型詞串的種類不如其他次類多。筆者比較劉貞好、陳浩然、楊惠媚

(2017) 之研究，該研究所收錄的架構型三詞詞串共有 41 種，遠大於本文母語者的 7 種。經由比較，我們發現數量的差異似乎來自於本文與該研究對詞串的分類方法的不同。表 4-25 整理收錄在該研究中架構型詞串，經筆者對照過後發現該研究 41 種架構型詞串中，本研究歸類為架構型詞串的僅有 6 個，其餘有 2 個屬程序型、8 個屬歸類型、1 個數量型、2 個組織型、22 個未收錄在本文。

表 4-25 劉貞好、陳浩然、楊惠媚（2017）架構型詞串在本文之分類

劉貞好、陳浩然、楊惠媚（2017）架構型詞串	
架構型	的情況下、為基礎的、的角度來、的基礎上、在這樣的、的情形下
程序型	的過程中、所需要的
歸類型	在這個、每一個、這個問題、在這種、作為一種、這種情況、為一種、為一個
數量型	在一個
組織型	視為一個、的問題是
未收錄	發展過程中、為中心的、成為一個、的前提下、作為一個、視為一種、成為一種、從這個、所強調的、在某種、這個概念、此一問題、所面臨的、種情況下、在整個、舉例來說、的理論基礎、直接或間接、所面對的、本研究依據、這段期間、本研究針對

分類方法不同，可能是造成本研究與劉貞好、陳浩然、楊惠媚（2017）之研究結果有差距的原因之一。本文於 3.5 說明我們藉由 collocates 功能觀察詞串的搭配詞以作為功能分類的依據，劉貞好等人僅指出他們是透過對應之英語詞串作為分類參考，未詳細說明歸類方式，故我們推測本研究與該研究歸類準則不同，導致分類成果有差距。另外一個因素可能是與篩選標準有關，劉貞好、陳浩然、楊惠媚（2017）研究之語料共九百萬詞，詞串頻次門檻為每一百萬詞 10 次、文本分布率 2% 以及 5%；而本文蒐集之語料庫個別約兩百多萬詞，詞串頻次門檻為每一百萬詞 20 次、文本分布率 10%；相較於前人之研究，本文語料量較低，但是詞串篩選門檻較嚴謹，也可能是造成篩選出之詞串不同的因素之一。

#### 4.3.2.2 目的型詞串

目的型詞串用以介紹作者之目的，這類詞串是母語者和學習者的使用頻率相差最懸殊的類型(LLR=119.97,  $p < 0.001$ )，學習者在詞串種類數量和使用頻率上都高於母語者。從表 4-26 能看到母語者僅有 1 個目的型詞串「的目的是」，學習者則有 6 個詞串，且這些詞串都含「目的」一詞，在使用頻率方面，學習者使用「的目的是」的頻次為 118，不僅高於母語者的 93 次，也比學習者的第二高頻詞串「目的是為了」高了兩倍之多。

表 四-26 目的型詞串

目的 (Objective) : 介紹作者之研究目的	
母語者	的目的是
學習者	的目的是、目的是為了、的主要目的、的目的在於、主要目的是、研究的目的

從(54)-(57)能看到這類詞串除了引介研究本身的目的外，也會如(57)作為說明概念使用。我們透過搭配詞功能檢索在「的目的是」前的詞彙，以這些詞彙來判斷在本文的語料中，「的目的是」多使用在引介研究相關的目的還是解釋概念。圖 4-3 顯示母語者語料中最常與「的目的是」搭配的词彙為「說話」與「教學」，共有 19 筆，如(54)、(55)所示，這兩個詞彙多數用來解釋概念；而學習者數量最多的搭配詞為「問題」與「研究」，共有 22 筆，如(56)、(57)所呈現，作者多以這樣的組合指出論文的目的。

(54) 「第二語言教學的目的是培養學習者運用目的語進行交際的能力」 L1-14

(55) 「說話者說話的目的是表示預言，而後者的目的是命令。」 L1-75

(56) 「研究者設計本問題的目的是為了瞭解學習者對俄羅斯大學的評價如何。」 L2-1

(57) 「本研究的目的是為了更加瞭解胡志明漢文詩越南語譯文教學的特色與不足」 L2-9



Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	12	12	0	-2.46630	話	11	2	2	0	-6.68830	分類
2	7	7	0	-7.25342	教學	12	1	1	0	-3.99342	闡述
3	6	6	0	-4.91250	主要	13	1	1	0	-7.26054	閱讀
4	5	5	0	-5.08796	話者	14	1	1	0	-6.75122	重要
5	3	3	0	-7.85591	研究	15	1	1	0	-6.06537	遊戲
6	2	2	0	-1.01941	行事	16	1	1	0	-0.53398	身上
7	2	2	0	-8.06107	結構	17	1	1	0	-1.11895	讀書
8	2	2	0	-6.81522	活動	18	1	1	0	1.20298	講解法
9	2	2	0	-3.80077	後者	19	1	1	0	-5.90728	論文
10	2	2	0	-2.57838	增進	20	1	1	0	-7.90641	課程

圖 4-3 母語者「的目的是」Collocates 檢索結果<sup>18</sup>

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate	Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	12	12	0	-4.62985	問題	11	3	3	0	-4.86000	它
2	10	10	0	-6.64681	研究	12	3	3	0	-7.27363	分析
3	7	7	0	-7.39277	教學	13	2	2	0	-5.46598	論文
4	5	5	0	-4.53057	重要	14	2	2	0	-8.94801	語言
5	5	5	0	-5.48186	教育	15	2	2	0	-4.95579	語境
6	4	4	0	-7.57246	教材	16	2	2	0	-6.99886	設計
7	4	4	0	-8.76583	學習	17	2	2	0	-6.47943	練習
8	3	3	0	-5.03683	訪談	18	2	2	0	-5.93465	編寫
9	3	3	0	-9.02347	者	19	2	2	0	-4.99035	測驗
10	3	3	0	-5.43328	本文	20	2	2	0	-2.04890	最終

圖 4-4 學習者「的目的是」Collocates 檢索結果

從上述的搭配和使用頻次來看，我們能發現母語者和學習者使用目的型詞串的方式有些許不同，母語者僅有一種目的型詞串，在使用上不全是為表達研究目，而學習者目的型詞串種類較多、使用頻率亦較高，並且集中使用詞串「的目的是」。Salazar (2011) 也在他的研究中有相同的發現，學習者不但使用目的型詞串的次數較多，更有單一詞串集中高頻的現象。我們似乎可以從以上的結論推測，學習者似乎集中使用某些詞串直接點明研究程序、步驟和目的，而母語者則有較多樣的表達方式。

#### 4.3.2.3 引述型詞串

引述型詞串用來說明論述來源或支持論述的數據，由於這類型詞串引出參考文獻，故詞串內多有「提出」、「說」、「指出」等詞。引述型詞串的種類

<sup>18</sup> 圖 4-3 顯示最高頻搭配為「話」是因為原始語料中「說」因轉碼問題變成亂碼「?」，導致斷詞出現錯誤。經過檢索確認 12 筆「話」的語料皆為「說話」。

和使用頻次在文本導向 8 個次類中排第四，也是常見的詞串類別，母語者使用本類詞串的頻率顯著高於學習者（LLR=110.94,  $p<0.001$ ），且母語者高頻使用的詞串多為共有詞串。

表 4-27 引述型詞串

引述 (Citation)：說明論述來源或支持論述的數據	
母語者	所提出的、所說的、的研究中、所做的、所表示的、的相關研究、一書中、所指涉的、所舉的、所持的、說的話、一文中、舉的例子
學習者	所提出的、所表示的、所說的、所提到的、的相關研究、的研究指出、所提的、的研究中、所採用的、說的話、一書中、所做的、所用的

本類詞串的共有詞串之比例很高，詞串使用方式也相當一致，詞串前普遍連接引用文獻之作者，後提出引用內容，如語料(58)-(61)所示。從表 4-27 我們能看到有引述功能的詞串種類繁多，可供未來學術寫作課程參考，讓作者在引用文獻時可以有多樣的表達方式。

(58) 「且盧 (1991) 所提出的-b 鑑定字法和-d 添加法基本上採取同樣道理」

L1-25

(59) 「就如接受訪問的導遊員所表示的一樣」 L1-50

(60) 「張寶林 (2014) 的研究中也發現，學習者沒辦法區別」 L1-41

(61) 「在課堂上蔡如珮 (2005) 所提到的『字中心詞彙教學』配合起來」 L2-

61

另外一點能從表 4-27 發現的是，本類詞串有不少「所 V 的」結構。事實上，本研究篩選出的詞串有不少都是「所 V 的」結構，母語者共 24 筆、學習者 25 筆，佔總詞串的種類約 10%。「所 V 的」結構總使用頻率母語者為 2434 次、學習者 2201 次，母語者顯著使用頻率較高（LLR=82.34,  $p<0.001$ ），表 4-28 列出本研究中為「所 V 的」結構之詞串。

表 四-28 所 V 的詞串列表

	所 V 的
母語者	所提出的、所說的、所使用的、所指的、所做的、所表示的、所產生的、所呈現的、所表達的、所代表的、所造成的、所討論的、所指涉的、所構成的、所處的、所需的、所提供的、所扮演的、所舉的、所持的、所具有的、所設計的、所表現的、所形成的
學習者	所使用的、所提出的、所表示的、所說的、所提供的、所指的、所提到的、所產生的、所造成的、所表達的、所設計的、所出現的、所提的、所需要的、所佔的、所呈現的、所採用的、所遇到的、所具有的、所涉及的、所做的、所存在的、所得到的、所用的、所需的

從語料(62)-(65)我們可以發現，將句中的「所」刪除都不影響原語義，不過加上「所」似乎讓句子更正式一點。從母語者使用頻率特別高與此結構在本文的詞串所佔的比例來看，這種文言殘留的書面表達方式，母語者用得還是比學習者多。

- (62) 「謝佳玲認為西方學者**所提出的**三種主要語義特質仍待商榷」 L1-46
- (63) 「甚至到課堂上**所使用**的一切資料都可稱之為教材」 L2-30
- (64) 「其實是建立在趙元任**所說的**『社會學上的字』的概念基礎上的」 L1-24
- (65) 「筆者發覺很難找到可以完全符合**所提出**的主要標準」 L2-16

#### 4.3.2.4 比較型詞串

比較型詞串用來比較及對比不同要素，在文本導向中比例佔第二高，在母語者語料佔約 27%、學習者約 23%，在使用頻次方面亦是文本導向中第二高頻的，其中母語者的使用頻率顯著高於學習者 (LLR=85.04,  $p<0.001$ )。由此可知「比較」型詞串在本研究所蒐集的文本中相當常見。「比較」型詞串比較要素間的異同，故詞串中多數有「較」、「最」、「更」等程度副詞，也有「不同」、「同時」、「之間」、「對應」等詞來點出要素間的關係。

表 四-29 比較型詞串

比較 (Comparative) : 比較及對比不同要素	
母語者	之間的關係、有不同的、較高的、而不是、最大的、最高的、一對一、兩者之間、較多的、相對應的、更好的、不一樣的、不同之處、不同的是、較低的、兩種不同、上的差異、兩者都、是相同的、兩個不同、較大的、的對應關係、之間的差異、較強的、同時也是、是不同的、最小的、更深入的、的則是
學習者	之間的關係、最高的、而不是、最大的、有不同的、兩者之間、上的差異、密切的關係、較高的、不同之處、一對一、最常用的、兩者都、最低的、之間的差異、相對應的、大的差異、有密切的、兩種不同、是相同的、兩個不同、較大的、較多的、的對應關係、最常見的、相當高的、較低的、更好的、的最大

語料(66)-(69)呈現比較型詞串的語料，從語料中我們可以看到詞串標示出要素間對比結果。

- (66) 「因而筆者用 ANOVA 進行非母數檢定『Kruskal-Wallis-K-W』，以比較四種聲調之間的關係。」 L2-33
- (67) 「因此教材應該以口語教學為重點，而不是一般的語言綜合課教材。」 L2-16
- (68) 「由以上學者們的看法可得知，同一件事情對不同文化而言有不同的意義」 L1-2
- (69) 「除了『要 5』，初級的學生『要 2』的偏誤是較高的」 L1-21

值得注意的是，儘管比較型詞串在本研究中在文本導向的比例佔最高，但是過往研究中比較型詞串的比例卻不那麼顯著（劉貞好、陳浩然、楊惠媚，2017; Cortes, 2004; Salazar, 2011）。造成不同的可能原因之一是本研究語料的文體特色所致。本研究語料為碩士畢業論文，畢業論文的篇幅長、須參考文獻多，亦須要大量對比參考文獻、比較研究結果來證明研究的能力。另外一個原因可能為單一研究領域限制的關係。本研究僅蒐集華語文教學領域的碩士學位論文，對比過去的詞研究會蒐集不同領域的學術文章，在這些研究發現不同領域的文本所使用的詞串有差異（Hyland, 2008a; Hyland, 2008b; Cortes, 2004），而「比較」型詞串在本文所佔比例較高可能表示華語文教學研究中時常比較不同的要素。第三個可能原因則是跟語言不同有關，若將表 4-29 中的詞串翻譯成英文，大多數詞串在英文中以一詞或兩詞即可完成，未達到研究者收錄的詞串

長度，造成英文文獻中比較型詞串種類較少。未來我們可以進一步分析這些推論是否成立。

#### 4.3.2.5 附加型詞串

附加型詞串的功能為在不同要素間建立的附加連結，從詞串表可以看到「附加」型詞串的結構大部分包含「另」、「也」來標示附加的要素。本類詞串種類與使用頻次不高，但我們仍發現比起學習者，母語者的使用頻率高得多（LLR=72.13,  $p < 0.001$ ）。以語料(70)-(73)來看，使用「附加」型詞串的時機為闡述一個主題中的第二種解釋、可能性或是另一個組成主體的要素。

表 四-30 附加型詞串

附加 (Additive) : 於不同要素間建立之附加連結	
母語者	另一個、另一種、這也是、也可以是、另一類、也不是、另外一個
學習者	另一個、另一種、另外一個、這也是、之外還有、也可以說、另外一種、也有些、但也有

- (70) 「它可以是一個語境的終點，**也可以**是一個語境的起點」 L1-42
- (71) 「義務情態中還有**另一個**系統為保證系統。」 L1-75
- (72) 「這種間接語言是有特定意圖的；**另一種**情況是說話人無法用語言直接地表達自己才無可奈何地使用了間接的表達方式。」 L2-15
- (73) 「**另外一個**原因在於選用原則及理解的適當性。」 L2-61

除了上列的詞串以外，「除此之外」也是表達附加時常用的表達方式。由於中央研究院 CKIP 斷詞系統將「除此之外」標記為一個詞，故未收錄在本研究中，母語者使用 229 次、學習者使用 192 次，母語者的使用頻率顯著高於學習者（LLR=12.30,  $p < 0.001$ ）。

#### 4.3.2.6 推論型詞串

推論型詞串的功能是能根據資料數據來標示推論與結論（Signal inferences and conclusions drawn from data），用來解釋所提的概念或現象。推論型詞串の種類和頻率在兩個群體中皆為最高的，母語者的使用頻率更顯著高於學習者（LLR=39.25,  $p < 0.001$ ），以這兩點來看，推論型詞串在學術型文章的重要性可想而知。推論型詞串包含許多與結果相關的詞彙，如：「呈現」、「構成」、



「產生」、「表示」、「表達」、「結果」、「成果」，仔細觀察學習者詞串表，能發現學習者使用不少含「結果」的詞串，母語者則無。

表 四-31 推論型詞串

推論 (Inferential)：根據資料數據來標示推論與結論	
母語者	指的是、並不是、所指的、所產生的、而來的、所呈現的、所表達的、所代表的、也就是、的研究結果、所構成的、而成的、而產生的、表示的是、可說是、可以說是、所提供的、扮演的角色、所扮演的、是不是、表達的是、的研究成果、來說是、要表達的、所具有的、表現的是、所表現的、的概念是、所形成的、的關係是
學習者	指的是、並不是、的研究結果、調查的結果、的結果是、所提供的、也就是、研究結果顯示、本研究結果、所指的、所產生的、所表達的、分析的結果、的分析結果、結果來看、調查結果顯示、的調查結果、所出現的、研究的結果、可說是、所呈現的、的結果顯示、得出的、扮演的角色、也可以說、之間的互動、而產生的、所具有的、結果可知、這樣才能、可以說是、的研究成果、是指在、來說是、要表達的、的意思是

如(74)-(77)所示，推論型詞串多使用於概念說明和研究結果闡述。(74)、(75)選自母語者語料的「並不是」為最高頻的推論型詞串，同時也是母語者所有詞串中的第七高頻，作者能以該詞串釐清前提的概念，「並」作副詞時與否定副詞連用，有加強否定語氣且略帶反駁的意味（李鑒、施光亨、李行健，2006），母語者使用「並不是」的頻率顯著高於學習者（LLR=16.37,  $p < 0.001$ ），且約有 78% 母語者文本皆使用了「並不是」，可能表示母語者寫作時語氣較強烈。

(74) 「由上可知，教學活動**並不是**只有課堂中的『教』與『學』」 L1-7

(75) 「由於知識建構的教學目的**並不是**為學生打分數」 L1-5

(76) 「教材功能分析結果 A 教材功能**分析的結果**顯示，研究者所組合的 13 個功能中 A 教材裡出現了 8 個」 L2-40

(77) 「本問題**調查結果顯示**，俄羅斯學習者學習漢語的困境五花八門。」 L2-1

另一個學習者詞串的特點是「結果」的使用。學習者的「推論」型詞串共 31 個，其中有 11 個皆含有「結果」；反之，在母語者 31 個「推論」型詞串中僅有一個含有「結果」。以 Antconc 檢索詞彙功能查詢發現學習者使用「結果」的筆數為 5311、母語者為 3311，學習者似乎多使用「結果」點明研究發現

或說明探討文獻的結論，而母語者則會以其他同類詞表達。這個現象我們從前面正關鍵詞表（表 4-1）中「結果」排在前五名也可以得知。

#### 4.3.2.7 組織型詞串

組織型詞串依 Hyland(2008a) 的定義是一種能組織話題的範圍或將讀者引導至文本它處的文本反身標記 (Text-reflexive markers which organize stretches of discourse or direct the reader elsewhere in text)。換句話說，組織型詞串有兩個功能，第一個為引導讀者到文本一處，有帶領讀者閱讀的作用，如：「如下表所」、「結果如表」；第二個功能為指出所有含括細項，如：「以上所述」、「以下幾個」。組織型詞串是文本導向中種類第三多的類型、使用頻次在八個次類中也偏是第三高，在過往研究中組織型詞串一向在文本導向中都有一定的比例 (Hyland, 2008a; Salazar, 2011; Pan et al., 2016)，可見組織型詞串在學術寫作中相當重要。本類詞串學習者使用的頻率顯著高於母語者 (LLR=24.06,  $p<0.001$ )。

表 四-32 組織型詞串

組織 (Structuring)：能組織話題的範圍或將讀者引導至文本它處的文本反身標記	
母語者	本研究的、本研究之、表所示、例中的、在本研究、本論文的、視為一個、本研究中、如表所、本研究在、的分類方式、如下表所、前文所、本文的研究、為本研究
學習者	本研究的、本論文的、在本研究、例中的、本研究之、本研究主要、本文的研究、本論文研究、表所示、在本論文、本研究中、如表三、結果如表、所遇到的、的問題是、為本研究、如下所示、以上所述、以下幾個、在本節、的主要內容

表 4-32 中有許多詞串含有「本論文」、「本研究」的組合，將話題聚焦在論文本身，如(78)-(81)所示，這個使用能夠呼應到本文在 4.2.2 的討論，學習者有多用指稱研究之詞串的傾向，從組織型詞串的使用能更明顯地看到這個趨勢。學習者不僅「本論文」、「本研究」的詞串在種類上多於母語者，使用的頻率也高出許多。

(78) 「由此可知，本研究的受試者中」 L1-22

(79) 「至於『簡體字』的字體內容將放在本研究未來的發展研究」 L2-55

(80) 「本研究主要是透過問卷測驗與訪談的方式，針對日籍中文學習者學習漢語詞彙時的難點進行探討」 L2-46

(81) 「如表所示，高分群學習者對於 VA 的常用排序為第 17」 L1-56

從學習者使用的方式來看，我們能發現學習者組織研究脈絡時，相對集中使「本論文」、「本研究」這類型的詞申明確指出焦點。不過，值得注意的是我們檢索近義詞「本文」發現母語者的使用頻率顯著高於學習者（L1:L2=2679:1723, LLR=398.38,  $p<0.001$ ），由此可知母語者與學習者有不同的詞彙傾向，學習者多用「本論文」、「本研究」；而相對地，母語者則是偏好使用「本文」來指稱研究文本。這現象從負關鍵詞表（表 4-2）也可以看出。

#### 4.3.2.8 因果型詞串

因果型詞串指出要素間之因果關係，這類詞串所佔比例不高（L1: 4.63%; L2: 7.03%）、使用頻率也偏低，但學習者使用因果詞串的頻率顯著高於母語者（LLR=23.06,  $p<0.001$ ）。從表 4-33 我們可以看到，母語者與學習者前五高頻因果型詞串相同，我們可以推斷這五個詞串在學生論文中是相當普遍的。另外，我們也可以觀察到「因為」、「因此」、「原因」這三個詞彙經常用於因果型詞串中。

表 四-33 因果型詞串

因果 (Causative)：指出要素間之因果關係	
母語者	因此本研究、這是因為、所造成的、因此研究者、的原因是
學習者	因此本研究、因此研究者、所造成的、這是因為、的原因是、原因之一、所涉及的、因此本論文、的主要原因

(82) 「因此本研究中的量詞比傳統上所認為的『量詞』範圍要大」 L1-35

(83) 「會造成這種現象的原因是初學華語的時候，學習聲調時間及練習不足」 L2-68

(84) 「因此，語言遷移是所有外語學習者發生偏誤的主要原因之一。」 L2-46

(85) 「根據筆者的判斷，這是因為其語義較不易掌握」 L1-43

由於學術研究相當重視因果關係解釋，劉貞好、陳浩然、楊惠媚（2017）的研究中收錄的因果型詞串數量為文本導向類的第二高，故本研究中母語者這

類詞串的相對低頻令人意外。為進一步了解母語者表達因果的方式，研究者挑選了前幾個學習者高頻使用的詞串，對比母語者的使用頻率。表 4-36 左列詞串是兩群體共有的詞串、右列是學習者的獨有詞串，雖然母語者也使用這些詞串，但是因為頻率不達篩選門檻，所以未列入母語者詞串表。從使用頻率能發現，「的主要原因」和「因此本論文」、「原因之一」三個詞串是學習者常用的詞串，但是在母語者語料中因為次數不到門檻而未收入。

表 四-34 母語者與學習者因果型詞串使用頻率

詞串	L1	L2	詞串	L1	L2
因此本研究	114	88	原因之一	25	49
這是因為	86	82	所涉及的	45	49
所造成的	82	87	因此本論文	21	44
因此研究者	73	149	的主要原因	19	44
的原因是	68	79			

學習者似乎在使用表 4-34 的詞串上表現出種類集中的趨勢，然而，我們進一步檢索其他表達因果關係的方法如連接詞「故」、「因而」，得知母語者使用這兩個詞的次數皆顯著高於學習者（故=2248:613, LLR=1284.59,  $p<0.001$ ; 因而=505:449, LLR=19.16,  $p<0.001$ ）。由此可以推論母語者與學習者對於表因果的方式似乎有不同的傾向，儘管母語者使用本類詞串的頻率顯著低於學習者，但也不能說母語者較少在寫作中表達因果關係。

### 4.3.3 參與導向

參與導向是三個分類中的最後一類，功能是聚焦在文本的作者或讀者上，這類型的次分類僅有吸引及立場兩類，相較於研究導向和文本導向，參與導向的詞串種類非常少，母語者約佔全部的 7%、學習者則僅佔 4%。比較兩個次類的種類數量，我們可以從表 4-35 中看到立場型詞串佔的比例較吸引型詞串高一些。以頻次的角度來看，雖然相較其它類使用頻次不算高，但是兩群體使用參與導向詞串的頻次頗大，如表 4-36 LLR 分析結果顯示，兩個次分類中母語者的

使用頻次顯著高於學習者，尤其是立場型詞串的 LLR 值特別高，表示母語者和學習者的差異顯而易見。

表 四-35 參與導向詞串分布

研究導向	母語者		學習者	
	種類比例	頻次	種類比例	頻次
立場 (stance)	58.82%	760	54.55%	633
吸引 (engagement)	41.18%	744	45.45%	408
總數	100%	1504	100%	1041

表 四-36 參與導向詞串頻次顯著性檢定

類別	L1 使用頻次	L2 使用頻次	LLR	顯著性	
立場	744	408	163.18	0.000	+
吸引	760	633	42.28	0.000	+

吸引及立場詞串的差異能以主觀表達和客觀表達來分辨。立場型詞串直接表達作者的想法；而相對地，吸引型詞串是以客觀的角度吸引讀者的注意 Hyland(2008b)。從表 4-35、36 我們可以看到，不論是主觀敘述還是客觀敘述，母語者的使用頻率皆顯著高於學習者，顯示母語者較常使用這些詞串來闡述想法。母語者使用客觀和主觀的詞串頻率差距不大，只是表達主觀評價的立場型詞串種類較多；學習者在使用頻率上則有差異，立場型詞串較高、吸引型詞串較低。

#### 4.3.3.1 立場型詞串

立場型詞串的功能為傳達作者之態度與評價 (Convey the writer's attitudes and evaluations)，這類詞串能清楚地傳達作者的想法，因此「認為」是這類詞串常見的詞彙。母語者使用的詞串種類和頻率皆高於學習者，Chen & Baker (2010) 的研究中提及母語者表達立場的方式較多元，頻率上的差異更是顯著 (LLR=163.18,  $p<0.001$ )，顯示母語者使用這些詞串表達想法的集中度。另外，這類詞串可以對應我們在 4.1 所提母語者學習者慣以不同詞彙指稱自己的



現象，觀察表 4-37 能夠發現母語者詞串多有「我們」；而學習者則偏向使用「研究者」、「筆者」現象。

表 四-37 立場型詞串

立場 (Stance) : 傳達作者之態度與評價	
母語者	研究者認為、因此我們認為、也可能是、我們認為這、認為這是、有可能是、但我們認為、認為這樣的、我們可以說、我們可以發現
學習者	研究者認為、因此筆者認為、研究者也、本研究發現、研究者發現、有可能是

立場型詞串除了大部分包含「認為」以外，也有如(88)含「可以」或是(89)含「有可能」，從這兩個語料中能夠看出這兩個詞串表達作者立場的方式。

(86) 「因此我們認為狀態動詞中也有不同的次類」 L1-59

(87) 「要掌握好使用日語為媒介的時機**研究者認為**不可時時使用日語」 L2-44

(88) 「**我們可以說**華語慢慢成為世界的主要語言之一」 L1-78

(89) 「VOT 較短的原因**也有可能是**來自母語的影響」 L2-23

另外一點值得我們討論的是模糊語 (hedges) 在本類詞串中的角色，我們在 2.2 曾整理過往文獻對母語者及學習者使用模糊語的狀況根據，以下我們將列出本文模糊語的使用情形。根據 Lakoff (1972) 的定義，模糊語就是一些讓想法表達變得模糊的詞語，語用學時常討論模糊語的使用，適當地使用模糊語可以保留面子。表 4-37 中「也可能是」、「有可能是」、「我們可以說」、「我們可以發現」都是屬於有模糊語的詞串。和其他學科相比，立場型詞串在社會科學中非常普遍，原因是模糊語詞串常作為表達想法的手段，比起其他實證性的自然科學，社會科學使用的頻率自然高得多 (Hyland, 2008a)。本文和過往研究之結果都發現，學習者使用立場型詞串的頻率皆不及母語者 (Salazar, 2011; Amirian, Ketabi, & Eshaghi, 2013; Pan et al., 2016)，尤其學習者使用模糊語的頻率並不高。本文語料也呈現一樣的趨勢，如表 4-38 呈現，含有模糊語的立場型詞串母語者用了四個、學習者則只有一個，使用次數的差異也頗大，這可能顯示學習者對於適當表達立場的方式與母語者仍有差距，這個結果與 Hyland & Milton (1997) 的研究也曾提出學習者寫作時通常會使用較肯定且有限的表達方式來傳達他們的立場。

表 四-38 有模糊語的立場型詞串

母語者		學習者	
也可能是	74	有可能是	42
有可能是	64		
我們可以說	54		
我們可以發現	53		

#### 4.3.3.2 吸引型詞串

作者透過吸引型詞串能直接吸引讀者注意力，作者藉由吸引讀者注意力的方式來表明的要點，我們在 4.3.3 提到相較於立場型詞串，吸引型詞串可以較客觀傳達想法。儘管母語者在僅多一個詞串「強調的是」，但是在使用頻率方面兩群體的差距不小，母語者使用的頻率顯著較高（LLR=42.28,  $p < 0.001$ ），若以詞串為單位分開來看，更能觀察到使用頻率的差距，注意的是

（L1:L2=173:91）、值得注意的（132:57），母語者的使用頻率約為學習者的兩倍。這也對應到前文的分析，表 3-15 五詞詞串「值得一提的是」僅出現在母語者的詞串表中；表 4-8 也顯示「值得注意的是」為學習者少用的四詞詞串之一，再加上本節三詞詞串的探討，皆指出學習者缺少吸引類詞串的跡象。

表 四-39 吸引型詞串

吸引（Engagement）：直接吸引讀者注意力	
母語者	注意的是、最重要的、值得注意的、重要的是、強調的是、最主要的、更重要的
學習者	最重要的、注意的是、重要的是、值得注意的、最主要的

從(90)-(93)我們可以注意到「值得注意的是」是共現頻率相當高的四詞詞串，儘管本文本節因詞串長度設定為三詞所以被拆解成兩個部分，但語料還是提供了「值得注意的是」共現的證據。

(90) 「值得注意的是 Byram 提出兩部分的技能」 L1-10

(91) 「前者強調的是不同個體內在動力的發揮、積極地自由發展」 L1-74

(92) 「另外，值得注意的是來源域-旅行的其他特徵」 L2-72

(93) 「先與兒童建立感情，這是最重要的一環」 L2-65

多數詞串研究皆提及二語學習者有缺少吸引型詞串的現象 (Chen & Baker, 2010; Salazar, 2011; Amirian, Ketabi, & Eshaghi, 2013; Pan, Reppen, & Biber, 2016) , 他們認為學習者表達的方式有限 (Hyland & Milton, 1997) , 另一方面, 學者也推測可能是因為學習者對表達較沒有自信, 故避免使用吸引型詞串主動吸引讀者的注意力, 反而偏好以平鋪直敘的方式呈現他們的研究 (Amirian, Ketabi, & Eshaghi, 2013) 。



## 第五章 結論

中文詞串研究是一個相對新的領域，研究過程及發現對未來研究的發展而言，有相當的參考價值。本研究蒐集來自臺灣各大學華語教學研究所之畢業論文共一百六十篇作為語料，分析母語者及學習者使用的詞串，比較兩者是否呈現差異，本章分為三節敘述。第一節回到本研究一開始設定之三個研究問題來陳述研究之結果；第二節敘述本研究的限制；第三節對未來相關研究與教學提出建議。

### 5.1 研究問題的回答

本文在第四章從不尋常高、低頻詞彙、詞串使用及詞串功能分類三個面向討論母語者及學習者碩士畢業論文的差異。以下就我們一開始在第一章提出之三個問題分別回答。

#### 1) 學習者學術詞串是否與母語者學術詞串的表現相同？若否，差異為何？

學習者與母語者的表現有相同也有不同之處。我們可以從以下三個部分看到兩群體的相同之處：(1) 高頻詞串 (2) 共有詞串 (3) 詞串功能類型的分布。

##### • 高頻詞串

我們於 4.2.1 發現母語者與學習者前 20 個高頻三詞詞串重疊性達 70%，這表示兩群體高頻使用的詞串很相似，也表示重疊的 14 個高頻詞串在中文學術寫作中相當常見，表 5-1 列出這 14 個高頻詞串：

表 5-1 母語者及學習者前 14 個共有高頻詞串

編號	詞串	L1 頻次	L2 頻次	編號	詞串	L1 頻次	L2 頻次
1	是一個	763	460	8	本研究的	307	491
2	是一種	573	390	9	的一種	306	203
3	指的是	490	293	10	的一個	275	295
4	這兩個	341	335	11	有一個	273	215
5	所提出的	341	246	12	的角度來	270	236
6	並不是	326	192	13	之間的關係	249	269
7	的過程中	321	274	14	很大的	210	281

除了知道以上 14 個高頻詞串在中文學術寫作中很常見以外，我們也發現其中又以解釋概念型的詞串「是一個」、「是一種」、「指的是」三個詞串的頻次最高，可以說概念解釋在學術論文中相當重要。

- 共有詞串

母語者與學習者共有之詞串所佔比例不低，對照表 4-5，可以發現共有詞串的比例在母語者三詞及四詞詞串中尤其高，所佔約 60%。這些共有詞串是母語者與學習者語料中皆常見的詞串，可作為未來學術寫作或是教材安排的參考詞串表，筆者將共有詞串整理成表 5-2。

表 5-2 母語者及學習者三、四、五詞共有詞串

<p>三詞</p>	<p>這一種、是一個、另一個、指的是、了一個、的情況下、是一種、並不是、較高的、值得注意的、在同一、的一種、稱之為、注意的是、所做的、有不同的、在這個、所提出的、所說的、的研究中、表所示、的兩個、較多的、綜上所、上所述、的另一、另一種、一對一、什麼樣的、有一個、者來說、在不同的、本研究之、所呈現的、同一個、角度來看、所指的、更好的、的過程中、中的一、所需的、在一個、兩大類、一書中、兩者的、本研究以、最重要的、的角度來、這兩者、較低的、種不同的、一個詞、而產生的、也是一、可以說是、一個是、因此本研究、有可能是、分為兩類、本研究將、每一個、要表達的、來說是、重要的是、的研究成果、這三個、本研究中、所產生的、相對應的、這兩個、扮演的角色、這個問題、日常生活中、這也是、者之間的、而不是、的對應關係、由此可、最大的、是相同的、有兩種、所具有的、的各個、可說是、其中一個、兩種不同、常出現的、較大的、兩個不同、的各種、每個人、這三種、此可知、各種不同、有一定的、兩者之間、最主要的、的方式來、這是因為、本章將、的方式進行、這兩種、所表達的、的一個、為本研究、之間的關係、的一部分、個不同的、有意義的、例中的、的表達方式、最基本的、所造成的、的第一個、的是在、兩個以上、本節將、在日常生活、所設計的、不同之處、的原因是、一個很、這四個、研究者認為、者的需求、研究者在、本論文的、的研究結果、對自己的、本研究的、所使用的、研究者將、因此研究者、的個人、另外一個、在這種、表達自己的、上的差異、在本研究、最高的、某一個、所提供的、很大的、一個人、有兩個、一本書、的四個、過程中的、一種是、兩個部分、所表示的、本文的研究、說讀寫、的三個、很重要的、之間的差異、的目的是、哪一種、的基礎上、四所示、個以上的、也就是、這方面的、兩者都、一個重要、個人的、有很大、中出現的</p>
-----------	--



四詞	值得注意的是、綜上所述、的角度來看、兩者之間的、由此可知、在日常生活中、兩個以上的、有很大的
五詞	或兩個以上的、兩個或兩個、個或兩個以上

- 詞串功能類型的分布

將詞串按照功能分析後，我們能發現母語者與學習者詞串的分布相當類似。圖 4-2 顯示母語者與學習者大部分的詞串都屬於研究導向和文本導向兩大類，參與導向是詞串種類最少的類別，進一步看到三個導向次分類的分布，也會發現母語者與學習者詞串種類的分布相同：研究導向中前三高的次分類分別為歸類、描述、程序型；文本導向中前三高的次分類分別為推論、比較、組織型；參與導向則是立場型詞串多於吸引型。從一致性高的比例分布我們也能推論這些詞串功能在學術論文中的重要性。

我們從以上三點看到母語者與學習者詞串使用表現之相同處，而不同之處我們則能從以下兩點討論：(1)學習者集中使用特定詞彙與詞串(2)學習者的用詞不如母語者多樣化。

- 集中使用特定詞彙和詞串

本文蒐集之母語者語料約比學習者多了將近四十萬詞，但是抽取出之三詞到五詞詞串數量卻學習者卻比較多。因此，我們可以說學習者的陳述方式重複性較高；而母語者因相對會使用不同搭配組合，使用頻率較分散，故許多詞串未達篩選門檻，故詞串數量較少。

從第四章的討論能發現，學習者在闡述研究方面的詞彙和詞串使用頻次似乎特別高，與母語者相比，學習者的詞串包含「研究」、「結果」、「目的」、「原因」幾個詞彙不少，另外，學習者指稱自己研究時的敘述方式重複性也高，如表 5-3 所示。除此上述詞彙外，我們在 4.2 的詞串使用分析中，也注意到在幾個「情況下」變換型的詞串中，學習者集中使用「在這種情況下」，母語者相較之下使用的頻率較分散。

學習者在指稱詞的使用上也呈現了集中的趨勢。4.1 負關鍵詞表中，相較於母語者，學習者少用指稱詞「我們」，且偏向使用「筆者」及「研究者」兩詞。最後，我們也發現學習者缺少吸引讀者的詞串如「值得注意的是」、「值

得一提的是」。我們在 4.2 及 4.3 的分析皆提及，儘管學習者也使用吸引型詞串，但是在頻率上不及母語者，且種類也相當有限。筆者將學習者多用及少用的詞串整理至表 5-3：

表 5-3 學習者多用及少用的詞串

	類別	詞串
多用	共有詞串	研究者認為、研究者在、本論文的、的研究結果、本研究的、研究者將、因此研究者、在本研究
	表目的之獨有詞串	目的是為了、的主要目的、的目的在於、主要目的是、研究的目的
	表因果之獨有詞串	原因之一、所涉及的、因此本論文、的主要原因
少用	共有詞串	注意的是、值得注意的、重要的是、最重要的、值得注意的是
	表強調之獨有詞串	更重要的、強調的是

- 用詞不如母語者多樣

本研究自考察(i)量詞(ii)詞彙搭配兩個部分，發現學習者使用詞彙不如母語者豐富。最直接能觀察到的就是量詞種類的數量，豐富的量詞系統是華語的特色之一，也時常造成學習者學習的困難。本文於 4.3.1.1 考察母語者及學習者的量詞使用，發現學習者使用的種類不只較少，在「類」、「篇」、「條」、「冊」、「段」、「項」、「套」這七個量詞的使用上皆不及母語者。未來的教科書或華語課室可以多補充這些量詞在學術文章中的搭配，讓學習者有更多的選擇，也讓詞彙更豐富。

同樣的現象我們也能在搭配功能檢索結果觀察到，在「的方式來」以及描述型詞串與「因素」和「部份」的搭配上我們都看到學習者使用的種類少於母語者，以頻率的角度來看，也會注意到學習者集中使用特定的搭配。為了使學習者有更豐富的搭配詞彙，我們能夠列出可替換的詞彙，幫助學習者增加詞彙量。如表 4-19 呈現「的方式來」左邊的搭配詞，未來在教材或是課室中，教學者能將這些搭配詞提供給學習者使用，學習者就能夠運用不同的搭配詞。

## 2) 華語學術詞串的功能分類分布情況為何？

對照圖 4-2，以三個大分類來看，研究導向和文本導向的比例相當接近，母語者文本導向詞串的數量略高於研究導向；學習者則是研究導向略高於文本導向，但是兩者的差異皆不超過 2%，多數的學術寫作詞串也都呈現類似的分布（劉貞好、陳浩然、楊惠媚，2017；Pan, Reppen, & Biber, 2016）。

筆者於前文提及，母語者與學習者詞串的功能次類分布非常相似，我們將各分類的數量轉為比例，能更清楚的看出分布的狀況。從表 5-4 來看，歸類型、比較型、推論型、程序型及描述型是前五個數量最多的類別，這說明了這幾個功能類別對碩士論文非常關鍵。相似的結果在劉貞好、陳浩然、楊惠媚（2017）的研究中也能看見，在該研究中歸類、程序、推論在數量上佔有較高的比例，由此可見這三個類型的詞串可能對學術文章而言相當關鍵。

表 五-4 母語者及學習者功能分類種類分布比例

主分類	子分類	母語者	學習者
研究導向 (research-oriented)	程序 (procedure)	7.42%	9.93%
	數量 (quantification)	5.24%	5.15%
	描述 (description)	9.61%	13.97%
	歸類 (grouping)	23.14%	19.85%
文本導向 (text-oriented)	附加 (additive)	3.06%	2.94%
	比較 (comparative)	12.66%	10.66%
	推論 (inferential)	13.10%	12.87%
	因果 (causative)	2.18%	3.31%
	組織 (structuring)	6.55%	8.09%
	架構 (framing)	3.49%	2.21%
	引述 (citation)	5.68%	4.78%
	目的 (objective)	0.44%	2.21%
參與導向 (participant-oriented)	立場 (stance)	4.37%	2.21%
	吸引 (engagement)	3.06%	1.84%

### 3) 外語學術詞串研究的成果與華語的學術詞串特徵的異同為何？

我們在第二章提到，英語學術詞串研究成果有下列五項：

- i. 新手研究生會依賴使用詞串來組織其學術文章。(Hyland, 2008a)
- ii. 學習者使用詞串的數量雖然高於母語者，但卻不如母語者多樣化。(Ädel & Erman, 2012)
- iii. 學習者會出現多用及誤用的情形。(De Cock, 2004)
- iv. 學習者在學術文章中使用較多口語常見的詞串。(De Cock, 2004)
- v. 學習者撰文時容易有過於主觀、果斷的傾向。(Hyland & Milton, 1997)

本研究之結果與上述五個特徵相似。首先，Hyland(2008a)指出新手研究生較依賴詞串使用，可能是因為他們不熟悉學術寫作的技巧，而這也可能是本研究中學習者使用的詞串數量多於母語者的原因之一。相較於母語者，學習者對於目標語寫作較沒有經驗，因此多以詞串來組織學術文章。除了以詞串總數量來看，從第四章的分析結果，我們也發現學習者有集中使用特定詞串的傾向，這也證實了學習者會較依賴詞串使用來組織文章。第二點，本文雖未探討本文詞串之結構，但從檢索結果能發現母語者使用之搭配詞較學習者多樣，我們能將此結果運用在未來研究上，提供學習者近義詞表，使其用詞更豐富。第三點，我們在擷取語料的同時，也觀察到學習者仍有多用及偏誤的狀況，如 4.3.1.1 提及之「一+量詞」的多用、4.3.1.3 「就像是」的誤用，這些發現都能作為未來學術寫作課程及教材內容之參考。第四點，儘管中文詞串研究尚未有口語及書面語詞串比較，但觀察 4.1 關鍵詞表分析，母語者相較之下用了較多單音節功能詞；另外，在 4.3.2.3 我們也提到母語者用了較多「所 V 的」這種文言殘留的書面表達方式。從以上這兩點來判斷，學習者的詞串的確較母語者口語化，但是更深入的分析待未來研究進一步考察。最後一點，本研究在 4.3.3.1 討論立場型詞串使用，也發現學習者寫作容易過於果斷的現象。母語者在種類和頻次上，使用含有模糊語的詞串較學習者還要高，原因可能如 Hyland & Melton(1997) 所言，由於學習者尚未適當掌握模糊語的使用時機，使文章顯得較主觀，可見模糊語的使用也是普遍存在二語學習者間的問題之一。



## 5.2 研究限制

首先，本研究在母語者與學習者的論文蒐集時，限於在網路上能取得論文有限，無法做到主題平均分布或學習者國籍平均分布，導致後續的詞串分析可能受這些因素影響，未必能單純地呈現兩者的差異是本研究的限制。

其次，是詞串篩選的限制。在第四章我們提到華語定式中的元素可能可長可短，目前使用的工具無法篩選出非連續的詞，也因此本研究對於定式結構的討論有限。另外，具篇章功能的詞組（如：一般而言、換句話說），在斷詞系統的分類中屬句副詞(Dk)，整個詞組視為一個詞，不符合詞串定義，故無法列入本文之討論。然而，這類詞組在篇章功能中扮演重要的角色，若能突破目前斷詞限制列入討論，便能有更多數據來分析母語者及學習者之差異。

最後，在功能分類上，這部分是筆者個人歸類，沒找第二人，在歸類的客觀性及準確性都可能有所疏漏。加上對於具有重疊功能的詞串，筆者按過往研究之建議透過檢索工具，以比例最高的搭配功能歸類的這個做法也有待商榷。原來詞串就不限於只有一種功能，僅呈現詞串的單一功能可能不符合語言使用的真實性，或許歸到不同類也是一種做法。我們也發現由於詞串歸類方式不同，使本文與劉貞好、陳浩然、楊惠媚（2017）的研究在結果和數據上有些差異，造成比較和解釋上的困難。未來研究若能突破此限制，能使研究間的比較更加方便且客觀。

## 5.3 未來研究及華語文教學建議

目前中文詞串研究數量相當少，本文目前可取得的中文參考文獻僅有劉貞好、陳浩然、楊惠媚（2017）研究之社會人文科學論文常用詞串乙篇，該研究語料來源自社會人文科學期刊論文，而本文研究語料來自碩士畢業論文，從語料來源與性質的角度來看，未來研究仍有許多方向和議題都值得一探究竟。例如：

- (1) 不同背景作者寫作之比較
- (2) 不同文類、語體詞串比較
- (3) 不同學科之詞串比較
- (4) 不同程度之學習者詞串使用比較



本研究比較學習者及母語者的差異，過往英文詞串研究也比較博士生、碩士生和專家寫作時差異，透過分析可以了解學術寫作專家和生手的差異，能讓研究生和學習者了解專家在論文寫作的脈絡，進一步增進自己的寫作能力；在不同文類、語體方面，過往曾比較包含口語及書面語、學術授課課堂語言及教科書詞串，按照比較結果列出中文口語及書面語詞串，提供學習者使用時的參考；劉貞妤、陳浩然、楊惠媚（2017）研究篩選出中文人文社會學科之詞串，未來研究可以觀察其他學科的詞串是否有獨特性，也能夠透過比較看出不同學科的詞串差異性，更可以提供外籍學習者寫作的參考。

除了以學科作劃分外，依照論文寫作格式、以章節作為單位分析詞串或是針對不同程度的學習者比較使用詞串的方式，也是未來研究能探討的方向。分析論文各章節中常見的詞串，能使學習者了解各章節的寫作要點，也能有更直接的參考依據。而分析各程度學習者多用及少用的詞串，一來能夠檢視學習成果，二來能為華語教師和教科書提供編纂參考。此外，我們也可以分析寫作測驗中拿到高分的文本，探究高分文本是否在詞串的使用上有特殊性。

最後，我們希望本研究之研究結果也能夠應用在華語教學。例如，我們藉由分析少用及多用的詞串觀察到學習者相較母語者缺少的詞串，未來中高級華語寫作課堂或是學術華語教材，能夠針對學習者缺少的詞串類型補充，如前文提及之含有模糊語之詞串、吸引讀者注意的詞串以及常見之功能詞，透過提供範例及情境讓學習者將詞串插入在適當的語句間，能讓行文更通順、也能使文章更接近母語者的寫作風格。除了具有寫出順暢語句的能力以外，豐富的詞彙量在寫作中也相當重要，因此未來教材若能補充常見定式的搭配詞表，對學習者而言會是寫作時相當便利的工具之一。除了課堂及教材編纂的使用外，本研究篩選出的詞串也可供學習者作為自學的參考依據。學習者在撰寫學術文章時，能依自己所需對照詞串表，依照功能分類找到所需的詞串，並加以運用到自己的文章中，使內容更加流暢、語彙更豐富。

## 參考文獻

- 朱金平、賈益民（2009）。詞塊與華文詞塊教學。暨南大學華文學院學報，4，20-26。
- 何萬順、蔡維天、張榮興、徐嘉慧、魏美瑤、何德華（2016）。語言癌不癌？語言學家的看法。台北：聯經。
- 李鑒、施光亨、李行健（2006）。兩岸現代和與常用詞典。台北：中華語文出版社。
- 周曉芳（2011）。歐美學生敘述語篇中的回指習得研究過程。世界漢語教學，25（3），422-432。
- 信世昌（2001）。網路中文應用文寫作課程之設計與實施。第二屆全球華文網路教育研討會。臺北市。
- 高大威（2014）。比較視野下的修辭思維：余光中論現代中文的歐化現象。國立政治大學中國文學系政大中文學報，22，131-158。
- 梁茂成、李文中、許家金（2010）。語料庫應用教程。北京：外語教學與研究出版社。
- 張俊盛（2017）。語料庫輔助華語學習者辭典編撰初探研究報告。國家教育研究院。
- 張莉萍（2014）。不同母語背景華語學習者的用詞特徵：以語料庫為本的研究，中文計算語言學期刊（IJCLCLP），19（2），53-72。
- 彭妮絲（2013）。華語文讀寫讀本暨教學研究—以系統功能語言學理論為基礎之探究。臺北市立大學學報，44（2），33-62。
- 劉貞好、陳浩然、楊惠媚（2017）。中文人文社會科學論文常用詞串之研究。華語文教學研究，14（1），119-152。
- 蘇丹潔、陸簡明（2010）。"構式—語塊"句法分析法和教學法。世界漢語教學，24（4），557-567。

- Ädel, Annelie, and Britt Erman. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92.
- Altenberg, Bengt. (1998). On the phraseology of spoken English. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (101-122). Oxford: Clarendon Press
- Amirian, Zahra, Somaye Ketabi, Hamed Eshaghi. (2013). The Use of Lexical Bundles in Native and Non-native Post-graduate Writing: The Case of Applied Linguistics MA Theses. *Journal of English Language Teaching and Learning*, 11, 2-29.
- Baker, Paul. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Biber, Douglas, & Federica Barbieri. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, Douglas, Susan Conrad, and Viviana Cortes. (2004). If you look at the...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. (1999). *Longman grammar of spoken and written English*. London: Pearson Education Limited.
- Brown, Penelope, Stephen C. Levinson. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press
- Chen, Yu-Hua and Paul Baker. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30-49.
- Conklin, Kathy, & Norbert Schmitt. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72–89.
- Cortes, Viviana. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.
- DeCarrico, Jeanette & James R Nattinger. (1998). Lexical phrases for the comprehension of academic lectures. *English for Specific Purposes*, 7(2), 91-102.
- De Cock, Sylvie, Sylviane Granger, Geoffrey Leech, and Tony McEney. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on Computer* (67-79). London: Addison Wesley Longman Limited.

- De Cock, Sylvie. (2004). Preferred sequences of words in NS and NNS speech. *BELL: Belgian journal of English language and literatures*, 225-246.
- Ellis, Nick C. (2001). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty and M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (63-103). New Jersey: Blackwell Publishing.
- Ellis, Nick C, Rita Simpson-Vlach and Carson Maynard. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375-396.
- Francis, G., Susan Hunston & Elizabeth Manning. (1996). *Collins COBUILD Grammar Patterns 1: Verbs*. London: Harper Collins.
- Gao, Zhao-Ming. (2019) 'Corpus stylistics and Chinese literary discourses: A comparative study of four novels by Shen Congwen and Chang Eileen'. In Chris Shei (ed). *The Routledge Handbook of Chinese Discourse Analysis* (pp. 519-535). London and New York: Routledge.
- Gries, Stefan Th. (2008). Phraseology and linguistic theory: a brief survey. In Sylviane Granger and Fanny Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 3-25). Amsterdam: John Benjamins.
- Halliday, M.A.K, & Christian Matthiessen. (1994). *An Introduction to Functional Grammar*. London: Taylor & Francis Ltd.
- Huang, Tiao-Guan. (2018). Teaching 'Yi+Classifier' to Native Speakers of English and Korean in Intermediate Chinese Classes: Error Analysis and the Designing of a Pedagogical Decision Tree. *Taiwan Journal of Chinese as a Second Language*, 17(1), 153-183
- Hyland, Ken. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21.
- Hyland, Ken. (2008b). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.
- Hyland, Ken. (2012). Disciplinary Identities: Individuality and Community in Academic Discourse. *Applied Linguistics*, 34(1), 112–115
- Hyland, Ken, & John Milton. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183–205.
- Hyland, Ken, & Liz Hamp-Lyons. (2002). EAP: issues and directions. *Journal of English for Academic Purposes*, 1(1), 1-12.
- Lakoff, George (1972). Hedges: a study in meaning criteria and the logic of fuzzy concepts. *The 8<sup>th</sup> Regional Meeting of the Chicago Linguistic Society*, 183-228.

- Li, Wendan. 2004. The discourse perspective in teaching Chinese grammar. *Journal of Chinese Language Teachers Association*, 39, 25-44
- Lorenz, Gus M. (1998). Overstatement in advanced learners' writing: stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on Computer*, 53–66.
- Mahlberg, Michaela. (2012). The corpus stylistic analysis of fiction – or the fiction of corpus stylistics? In: Mukherjee, J, Huber, M (eds). *Corpus Linguistics and Variation in English: Theory and Description* (pp. 77-95). Amsterdam: Rodopi
- Meunier, Fanny, & Sylviane Granger. (Eds.). (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins Publishing.
- Miller, G. A. (1956). The Magical number seven, plus or minus two: limits on our capacity for processing information. *Psychological Review*, 63, 81- 96.
- Moon, Rosamund. (1992). Textual Aspects of Fixed Expressions in Learners' Dictionaries. In Arnaud, Pierre J. L. and Henri Bejoint (Eds.), *Vocabulary and Applied Linguistics* (13-27). London: Palgrave Macmillan.
- Nattinger, James. (1988). Some current trends in vocabulary teaching. In R. Carter, & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 62-82). New York: Longman
- Nattinger, James, & Jeanette S. DeCarrico. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Neely, Elizabeth, & Viviana Cortes. (2009). A little bit about: analyzing and teaching lexical bundles in academic lectures. *Language Value*, 1(1), 17-38.
- O'Keeffe, Anne, Michael McCarthy, and Ronald Carter. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- Palmer, Martha, Daniel Gildea and Paul Kingsbury. (2006). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1), 71-106.
- Pan, Fan, Randi Reppen & Douglas Biber. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.
- Pawley, Andrew & Frances Hodgetts Syder. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In: Jack J. C. Richards and Richard R. W. Schmidt (eds.), *Language and Communication* (pp. 191-225). London: Longman



- Pang, Ping. (2009) A study on the use of four-word lexical bundles in argumentative essays by Chinese English-majors: a comparative study based on WECCCL and LOCNESS. *Teaching English in China*, 32, 25-45.
- Salazar, Danica Joy Lorenzo. (2011). Lexical bundles in scientific English: A corpus-based study of native and non-native writing (Doctoral thesis, Universitat de Barcelona, Barcelona).
- Schmitt, Norbert. (2004). *Formulaic sequences: acquisition, processing, and use*. Amsterdam: John Benjamins.
- Schmitt, Norbert & Carter, Ronald. (2000). The Lexical Advantages of Narrow Reading for Second Language Learners. *TESOL Journal*, 9(1), 4-9
- Simpson-Vlach, Rita. (2004). Stylistic features of academic speech: The role of formulaic expressions. In Ulla Connor and Thomas A. Upton (Eds.), *Discourse in the Professions: Perspectives from corpus linguistics* (pp. 37-64). Amsterdam: John Benjamins Publishing.
- Simpson-Vlach, Rita and Nick C. Ellis. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31(4), 487–512.
- Sinclair, John. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Staples, Shelley, Jesse Egbert, Douglas Biber and Alyson McClair. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for academic purposes*, 12(3), 214-225.
- Stubbs, Michael. (2007). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25 years on* (pp. 89-105). Amsterdam: Radopi.
- Wray, Alison. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4), 463-489.
- Wray, Alison. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, Alison. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.
- Wray, Alison, and Michael Perkins. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, 20, 1-28.

## 附錄

### 華語教學領域論文常見學術詞串表

詞串長度	詞串
三詞詞串	<p>很重要的、研究者認為、本研究的、本論文的、的研究結果、研究者在、對自己的、因此研究者、研究者將、的個人、另外一個、最高的、很大的、一個很、一個人、在本研究、有兩個、上的差異、表達自己的、在這種、某一個、所提供的、所表示的、個人的、一本書、一種是、的目的是、的基礎上、的四個、過程中的、有很大、的一個、的三個、之間的關係、兩個部分、本文的研究、中出現的、之間的差異、哪一種、兩者都、這方面的、一個重要、這兩個、不同之處、的原因是、在日常生活、這四個、所設計的、例中的、本節將、這兩種、兩個以上、有意義的、所造成的、的一部分、兩者之間、的各種、最基本的、而不是、有一定的、最大的、的第一個、所表達的、的方式來、有兩種、這是因為、為本研究、所具有的、每一個、本章將、每個人、日常生活中、的角度來、說的話、的方式進行、最主要的、各種不同、可說是、兩種不同、其中一個、的過程中、這三種、這個問題、兩個不同、較大的、這一類、所使用的、另一個、是一個、的情況下、指的是、較高的、值得注意的、所做的、並不是、稱之為、注意的是、在這個、是一種、的研究中、較多的、表所示、有不同的、是兩個、的一種、所說的、在不同的、也就是、一對一、更好的、的兩個、所需的、什麼樣的、所提出的、另一種、兩大類、本研究之、一書中、所指的、這兩者、較低的、本研究以、在一個、在教學上、而產生的、有可能是、可以說是、分為兩類、角度來看、頻率最高、要表達的、來說是、的研究成果、一個是、同一個、的相關研究、有一個、因此本研究、本研究中、扮演的角色、重要的是、本研究將、所呈現的、兩者的、最重要的、的對應關係、相對應的、是相同的、所產生的、常出現的、的各個、這三個、這也是</p>
四詞詞串	<p>本論文的研究、有很大的、表四所示、的一個重要、一對一的、是一個很、兩個以上的、在本研究中、由此可知、頻率最高的、是很重要的、也是一個、在日常生活中、人與人之間、中所使用的、各種不同的、所要表達的、的不同之處、兩者之間的、兩個不同的、值得注意的是、綜上所述、一個完整的、所扮演的角色、的角度來看、更重要的是、在同一個、兩種不同的、也是一種、一個人的、本研究旨在、最常使用的、的每一個、出現頻率最高、或兩個以上</p>