

在職華語教師語料庫素養培訓初探*

張莉萍

國立臺灣大學華語教學碩士學位學程

摘要

語料驅動學習外語的方法在歐美課堂中已經實際應用多年，但這個方式在漢語做為外語的教學中才剛起步。究其原因，語料驅動的教學法對華語教師有一定的門檻，教師需要熟悉語料庫、語料庫工具、檢索技巧以及應用到教學的活動設計等等。為了培養教師擁有語料庫素養，本研究設計了一個 12 個小時針對在職教師的培訓課程，總共有 73 位在職華語教師參加。課程包括介紹相關的華語語料庫、教授基本和進階的檢索技巧、介紹與使用語料庫分析工具 (AntConc, Sketch Engine)、語料呈現方法與教學設計。全部課程結束後請參加者匿名填寫問卷，問卷由 39 題李克特量表問題和 4 題開放性問題組成，本研究將這些問題分為 7 類：「參與課程的動機」、「課前相關先備知識」、「課後對操作語料庫的能力」、「對語料庫的負面態度」、「執行、設計 DDL 教學活動的能力」、「未來在課程中應用 DDL 教學法的可能性」、「對整體課程的評價」，使用 R studio 以斯皮爾曼等級相關係數檢定之間的相關性，從檢定分析可知，不論教師的參與動機高低，皆能透過課程內容建立操作語料庫的能力與信心；課前擁有相關先備知識、擁有較佳語料庫操作能力的教師，未來應用 DDL 的可能性越高。結果顯示，擁有語料庫的操作能力是將語料庫應用於教學的關鍵，建議未來可以持續進行相關培訓。

關鍵詞：相關性檢定分析 問卷 語料庫 語料檢索 語料驅動學習

* 本研究感謝科技部計畫 (MOST 110-2410-H-002-048) 經費補助、期刊匿名審查人與編輯委員給予之建議、參與工作坊教師之回饋意見、工作坊助教施睿庭以及周均亭協助相關性檢定分析部分工作。

1. 研究背景

興起研究語料驅動的學習方法 (Data-driven learning, DDL)，主要是受「基於用法」(usage-based) 語言觀的啟發。「基於用法」的語言觀主張學習一個語言，最早是從個別的語言實例而來，透過語言的頻繁使用，高頻的語言單位逐漸形成語塊，進一步形成固定的構式 (construction) 或型式 (pattern)，而語言的規律性特徵就是在這過程中自然出現的 (Halliday 2003; Tomasello 2003; Bybee 2006)。從頻繁使用的這個概念出發，本研究相信讓學習者接觸大量語料、高頻複現同樣的形式，對語言習得定能有所幫助。而語料庫、語料庫檢索工具正是可以較為輕鬆、快速達到這個目的有利途徑。

然而語料驅動語言學習的教學法在歐美已經行之有年，但在漢語做為外語的教學中才剛起步。儘管國外已有不少研究，但 DDL 在實際課堂中的應用仍受到不少阻力，一般語言教師排斥在課堂中使用是因為 DDL 具有以下限制：(1) 在課堂上使用 DDL 的前提是，學生需要一定的科技使用及語料檢索能力，而這些技能需要教師事先花大量的時間去培養。但在一般教學進度的壓力下，教師沒有多餘的時間在課堂上引導學生；(2) 語料庫取得的語料不適合語言能力較低的學習者；(3) DDL 較著重於對語言形式的觀察與分析，和現代教學主流方法著重在溝通，看重流利度的教學方式有所衝突；(4) 學習者需要很強的歸納及語言分析的能力；(5) DDL 要求學生高度獨立學習的能力，但可能與某些教育單位的性質，或某些背景的學習者學習風格不同 (Leńko-Szymańska and Boulton 2015)。

以上所述的限制多指使用 DDL 直接教學法的關係，實際上應用 DDL 於教學可以有兩種做法，一是直接法、二是間接法。直接法顧名思義就是直接在課堂上讓學生使用語料庫，透過檢索、分析來學習；間接法則是教師事先準備好語料庫驅動得來的學習材料，在課堂上讓學生觀察、分析。因此要降低上述的限制，可以有不同的做法，例如：(1) 教師將語料庫中的語料列印出，做為課堂中的講義，讓學生分析語言，不直接操作電腦 (Boulton 2010)；(2) 讓學習者使用特別設置的語料環境，例如建構符合學生程度或簡易版的語料庫，而非傳統的語料庫 (Babych 2015; Chujo, Oghigian and Akasegawa 2015)；(3) 使用學習者較為熟悉的一般搜尋引擎作為 DDL 的工具 (Boulton 2015)。

對於 DDL 較著重在語言形式上的分析，與現行主流的溝通式教學有不小的分歧的批評，筆者的想法是，其實每一種教學法都有適合應用的對象，

加上，溝通式教學的研究也證實，在語言學習中，聚焦形式 (focus on form) 是學習過程中必要的一個階段，對於程度高的學習者尤其如此 (Doughty and Williams 1998; Norris and Ortega 2000)，而 DDL 正好可以滿足這個要求 (Gaskell and Cobb 2004)。因為 DDL 講究的就是從上下文來理解、分析語言；學習者透過注意形式 (noticing)、演繹、歸納的過程，有助於記憶與習得語言的效率 (Aston 2001:19)。

儘管如此，目前將 DDL 應用於華語教學的研究幾乎沒有，筆者設想，要有一定數量的具備這方面能力的師資，才能實際應用於課堂，否則將停滯在方法論或實驗階段，無法進一步得知 DDL 應用於華語教學或學習的成效究竟如何。不過要將 DDL 帶入教學中，教師需要有基本使用語料庫的能力，現階段華語教學有關 DDL 教學法的實踐或研究不多，究其原因，關鍵還是在接觸過語料庫的現職華語教師不多，教學者得熟悉使用語料庫的這個門檻可能也讓人卻步；而且看不到 DDL 的學習成效，教師也不敢貿然採用這個方法。也就是說，沒有這方面師資，就不可能有關於 DDL 成效的研究。上述動機促成筆者進行 DDL 教師培訓的研究。

本研究欲透過舉辦培訓在職華語教師工作坊，分析在職教師的回饋與問卷以做為未來培訓的基礎。

2. 文獻綜述

由於 DDL 在華語教學中應用的極少，在進行研究前，有必要瞭解國內外的現況，以做為本研究的基礎，包括瞭解 DDL 應用在外語教學的實例、應用在華語的研究、國內外關於 DDL 的師培研究以及 DDL 教學法教師所需的技能。以下筆者就三個面向來陳述與評論：

- (1) DDL 應用在外語教學上的實例研究。
- (2) DDL 應用在華語教學的相關研究。
- (3) 關於 DDL 的師培研究。

2.1 DDL 應用在外語教學上的實例研究

DDL 應用在外語教學中的實例隨著科技的發展自 21 世紀蓬勃展開，許多研究已證明 DDL 對學習外語是有效的。例如，Leńko-Szymańska (2015) 的研究顯示在語料庫中發現的訊息會比傳統的課本來源更豐富；Liu 與 Jiang (2009) 的研究顯示 DDL 能對學習者學習詞彙語法有正面的回饋；Mizumoto

與 Chujo (2016) 的研究顯示 DDL 可以協助學習者在寫作上發展糾錯和校對的能力；O'Sullivan (2007)、Yoon 與 Jo (2014) 的研究顯示 DDL 能協助學習者發展認知和後認知能力。相關的研究不勝枚舉，以下僅針對兩個方面來探討：一是應用 DDL 的方式，比較間接 DDL 與直接 DDL 教學法的成效；二是 DDL 適用的對象。

對於 DDL 的教學方式，前人多有採用直接法對學習者的挑戰相當高的論述，如果學生分析的是教師事先找好的語料，對學生的難度則較低。Smith (2020) 的研究即是為了比較直接 DDL 和間接 DDL 的教學成效是否有差別，這個研究的做法是在一間英國大學的英文學術論文寫作課上招募了 6 組國際商務會計與財務系的學生，共 94 人，這些學生都是第二語言 (L2) 的英語學習者，其中 88 人是中文或粵語母語者，他們的雅思分數介於 6.5~7 之間。94 位受試者共分成 4 組，其中兩組為實驗組，兩組為對照組。這 4 組受試者的任務是學習與會計或財經相關專業領域的詞彙，研究者要求實驗組的參與者自行建置和會計或財經相關的語料庫，用關鍵詞表 (keyword) 的方式提取專業領域的詞表學習。對照組則是直接學習研究者以實驗組相同方式建立語料庫所獲得的詞表。

實驗為期一個學期 (11 週)，實驗組每個禮拜會花 20 分鐘待在電腦教室，學習建置語料庫及產出詞表；對照組則是使用研究者每週直接提供的一個詞表來學習詞彙。為了檢驗學習的成效，受試者們在實驗的前後會接受前測及後測。最後統計出來的數據顯示兩組學生的進步都達顯著性，但實驗組整體只比對照組多進步了一點，似乎隱含不需要教授學生親手建置語料庫與詞表也能達到與實驗組類似的進步空間。

這個研究結果增強筆者支持間接法的信心，因為就筆者教授本國研究生的經驗，研究生要熟悉與操作語料庫至少需要 10 個小時以上的時間，更遑論外籍人士。

第二個研究是關於適用對象的研究，由於多數文獻指出 DDL 適合用在中級以上程度的學習者，Corino 與 Onesti (2019) 為了瞭解應用於初級學習者的效果如何，他們以義大利語母語背景的高職初級英語學習者 (17 位) 為研究對象，教學內容針對美髮科的學生學習構式：“have/get + object + past particle”；作者認為此構式相當不符合義大利語的邏輯，因此義大利語背景的英語學習者鮮少使用該句式，但於美髮相關的場域卻時常看到，如「have his hair cut...」。因要使用與美髮相關的語料，一般語料庫沒辦法達到這個需求，

作者為此先建了一個語料庫（長度共 3,776 個詞，550 個不同詞種），放在免費檢索工具 LancsBox 中，訓練課程共計 12 小時。

教學進行方式是，教師在上課的前幾天以 Whatsapp 傳送給學生，要求學生聽錄音檔。在課堂中，老師要求學生看著圖片，仔細聽錄音檔中所提到的動作和物品，並嘗試將單字填入老師製作的填字遊戲中。學生大概知道老師要他們留意錄音內容的動作及物品後，老師開始教他們如何操作 LancsBox 這個工具。之後，老師要求學生使用關鍵詞檢索功能 (KWIC) 以完成填字遊戲。學習者可以從錄音檔中，猜測自己聽到了什麼單詞，並以關鍵詞檢索功能去驗證自己聽到的是不是自己所想的。

他們完成了填字遊戲，學會了這些詞彙之後，教師開始要求學習者將這些學會的單詞鍵入語料庫中搜尋，像是 cut、wash、brush 之類的動詞。這些動詞多伴隨著“get/have + object + past particle”這個結構出現。之後，教師要求學生去歸納這些動詞前後的結構，歸納出來之後，學生就開始對這個構式進行分析並猜測這個結構使用於什麼情況下。學生最後自行歸納出這個結構用以描述「使……被……」的用法。教學活動結束的一個星期後，學生被帶到美髮工作坊，學生要以英語描述工作坊中正在進行的動作。教師沒有特別要求學生要以“get/have + object + past particle”描述問題，結果發現學生不但能夠自然且正確使用這個結構，而且連發音都進步了。

這篇研究報告讓大家看見語言能力較低的學習者也能受益於 DDL，而且 DDL 可以應用在特定目的的語言教學 (LSP, language for specific purposes) 和結合內容與語言學習 (CLIL, content and language integrated learning) 的教學中，也可以讓學習者獲得鷹架學習 (scaffolding) 及自信。

2.2 DDL 應用在華語教學的相關研究

將 DDL 實際應用在華語教學上的實證研究極有限，多數談的是如何應用、還有語料庫對教學的幫助。屬於前者研究的文獻，包括：(1) Wang (2001) 建置了中英平行語料庫，學生使用平行語料庫後發現，英文的“now”和中文的「現在」，很多時候不能直接對翻，因為這兩個不僅意義上有所差異，兩者使用的語法及語境也不盡相同。(2) Tao (2005) 則是建立了一個三十萬詞的當代漢語口語語料庫，用以教導高級學習者自然對話的一些特徵。(3) 王炳勻等人 (2020) 則是企圖從不同示例與面向提出教學引導與設計原則，研究中分別針對直接與間接 DDL 的模式來論述，使用的語料庫工具則是 Sketch

Engine 和 BCC (北京語言大學語料庫)。

Smith (2011) 的研究是首次將 DDL 應用在華語課堂中，雖然是實驗性質，沒有學分的正式課堂，因此只找到三個志願者參與實驗，但整體設計的課堂活動非常值得參考。教學的對象是西交利物浦大學中級程度的學習者，採用的平台是 Sketch Engine (SkE)。這篇研究設計了五個語料庫為本的任務，這些任務是根據課堂上使用的教材內容來設計。學習者需要使用 SkE 的詞彙對照素描 (Sketch Differences)、詞彙素描 (Word Sketches) 以及一些進階的搜尋功能來完成這些任務。使用 SkE 的詞彙對照素描功能來完成的任務如：請比較「結果」和「後果」的用法，並找出例句。另一種任務是要學習者寫 CQL 語法來檢索出與「石頭」搭配的量詞。

作者於結論指出這些學生成功地完成了這些任務，同時認為這個學習方式很具有挑戰性且相當有趣。最後，作者認為 SkE 對於學習者來說是一個很理想的工具，除了索引典 (Thesaurus) 是一個不錯的工具外，好例句 (GDEx, Good Dictionary Examples) 也是一個很好的功能；在這個功能之下，系統會自動排除不適合學習者學習的句子，例如語料來源為政治、宗教等具有爭議內容，或帶有罕用詞、過長，或是代名詞的指涉不在句內的句子。從這個實作的課程，可以清楚地看到哪些語料庫工具適合應用在華語課堂上以及如何應用。

Yeh 與 Zhang (2018) 則是對照自建的母語者語料庫 (42,182 個詞) 與學習者語料庫 (72,198 個詞)，發現學習者在敘述故事時不使用具有連接篇章功能的「就」，與母語者的頻繁使用差異很大。為了提高學習者使用「就」的意識，研究者設計了 4 堂以語料庫教學為主的課程，每堂 50 分鐘。讓學生在課堂上透過觀察母語者和學習者敘述同一個故事，但有兩個版本的差異，來提高學生對「就」的使用意識，進而引導分析，討論「就」的功能以及設計多種產出活動。該研究顯示實驗組和對照組 (未經過針對「就」的語料庫教學) 這兩組學生在前測的任務中都沒有使用「就」，然而經過語料庫為主的教學後，實驗組產出了「就」，而對照組的學生仍舊連一次「就」的產出都沒有，顯示實驗組對「就」的產出表現比對照組佳。這個實證研究顯示透過語料庫來教學確實有成效，美中不足的是，對照組沒有實施其他教學方法，較難說服華語教師 DDL 的教學法較其他方法成效佳。

2.3 關於 DDL 師資培訓的研究

隨著語料庫工具的日益優化與科技的發達，具備語料庫素養 (corpus literacy) 的語言教師似乎已成為近年來國外培育職前教師或在職教師追求的一個目標 (Leńko-Szymańska 2014; Leńko-Szymańska and Boulton 2015)。關於語料庫素養的概念，最早是 Mukherjee (2006) 為了語料驅動學習所提出的，定義基本的語料庫素養為：(1) 知道什麼是語料庫；(2) 知道語料庫能做什麼、不能做什麼；(3) 知道如何分析語料庫中的語料；(4) 知道如何從語料中得知語言的使用；之後的學者多在這基礎上進行師資培訓的研究 (Heather and Helt 2012; Callies 2016, 2019)，不過，幾乎都以英語教學為主。以下筆者將重點放在職前教師和在職教師這兩種對象的培訓來敘述。

2.3.1 職前教師

針對職前教師的培訓，多數研究結果顯示，職前教師經過語料庫素養培訓，可以有效地提高他們對語言的意識 (awareness) (O’Keeffe and Farr 2003; Farr 2008)，而這些研究很多是針對各國培育小學、中學英語教師的課程 (Abdel Latif 2021; Crosthwaite, Luciana and Wijaya 2021)。本研究的討論僅聚焦在大學應用語言學相關系所為可能從事外語教學的學生所開設的「語料庫」課程。

Leńko-Szymańska (2017) 的研究是以華沙大學應用語言學研究所外語語料庫課程的 98 位參與者為樣本，授課目標是使參與者熟悉如何在語言教學中使用 DDL。該研究蒐集五個學期（每學期上課 13-15 週，每週 90 分鐘）的資料，欲瞭解受訓後的參與者是否習得了使用語料庫基本的技能，是否能有效地結合教學和語料庫。她的研究最後以學生是否能完成以下 3 項任務來評估：(1) 建置特定目的的小型語料庫，最少 3 萬詞；(2) 對於該語料庫進行語言分析，列出關鍵詞清單或是值得注意的搭配，或找出該語料庫的其他特徵，並且分享其與語言教學的相關性；(3) 設計一堂專業英語教學 (ESP, English for specific purposes) 課程，其中需包含製作學習單或是課室活動。

該研究從量化與質化的結果中發現，在為期一學期的訓練中，學生大多僅能學習到語料庫的基本操作技能，無法靈活地檢索與解讀搜查到的訊息，且無法在教學中融合自身專長與 DDL。也就是一學期的授課，無法讓參與者具備應用 DDL 的三種能力。

這個結果與筆者的經驗吻合，筆者連續 4 年在大學研究所開設「語料庫

輔助華語教學」這門課，教授基本與進階操作技能；經過一個學期的授課，還是可以明顯感覺學生對於資訊檢索以及語料解讀的卻步。個人推測這跟學習的動機性與背景有很大的關係，首先是研究生還沒有豐富的華語教學經驗，較沒有動機與概念思考語料庫對教學的幫助；其次是研究所課業的壓力下，可能也沒有充裕的時間常常運用、接觸語料庫，而技能與熟悉度是需要經驗與時間的累積。

2.3.2 在職教師

多數師培的研究都是針對職前教師，以下介紹兩個針對在職教師的研究。第一個是 **Bunting (2013)** 的研究，他與 3 位擁有 20 多年教學經驗的在職英語教師合作，**Bunting** 指導他們使用語料庫設計課堂活動。在觀察與反思後，發現教師對於教學信念、對語料庫的先備知識、是否有操作過語料庫的經驗以及設定的教學目標，都會影響他們繼續使用語料庫的意願。雖然這 3 名資深教師皆對語料庫抱持著積極的態度，但是也提出了使用語料庫來備課費時以及在課堂上如果遇到技術問題，將佔據大量課堂時間的疑慮。其他研究也有相似結果 (**Lin and Lee 2015**)。

第二個研究則是 **Chen、Flowerdew 與 Anthony (2019)**，他們的研究對象是香港在職的英語教師，做法是辦理為時 3 個小時的工作坊，3 位講師，一位是該研究的作者，其餘兩位都是資深的語料庫語言學家與語言教育者。該工作坊舉行了兩次，至少有來自 10 個學校 63 位的在職老師參與。工作坊分成四個部分。(1) 20 分鐘，向參與者介紹語料庫和語料庫語言學的基本概念及語料庫在英語教學中的應用。(2) 75 分鐘，教授者展示如何幫助學生解決寫作中的詞彙語法問題。參與者有兩個任務：使用 **BNCweb** 創立學術子語料庫，透過簡單的檢索與複雜的檢索方法，找出字詞的分類、搭配和頻率，並且對這些結果進行解釋。在教師演示之後，參與者都會拿到一個學生寫作不當的例子，必須和他人分組合作，使用工具在語料庫中檢索正確的句子示例。(3) 75 分鐘，參與者需使用講師建立的特定學科語料庫與 **AntConc** 解決學生在詞彙、篇章的寫作問題。另外，講師還展示該如何使用 **AntFileConverter** (**Anthony 2016**) 建置自己的語料庫。(4) 15 分鐘，問答時間。工作坊結束時，參與者需填寫問卷，調查他們之前的經驗、對於工作坊的評價、對語料庫使用在英語教學中的看法，問卷共有 18 題李克氏特量表的問題以及 2 題開放性問題。

Chen 等人 (2019) 的問卷調查對象有 63 位，同時也是少數針對在職教師的 DDL 訓練，因此筆者多著墨這個研究的問卷結果。作者指出參與者大多都能理解語料庫使用在教學中的好處，但他們也擔心採用 DDL 會造成的技術困難與費時問題，這與其他文獻發現一致；然而卻沒有什麼研究去探究導致教師對語料庫又愛又怕的因素，因此他們的問卷中特別針對教師的特質來探討。在教師背景部分，本次參與的教師有 10 年以上教學經驗的佔了一半以上，有 6 年以上教學經驗的教師佔了三分之二。參加本次工作坊的教師對於語料庫語言學的先備知識少，但是他們對於學習新的教學法抱持了極大的興趣。在 24 位填寫開放性問題的教師中，有 10 位都認為在工作坊中學習如何使用 BNCweb 和 AntConc 是最大的收穫，讓他們理解語料庫促進英語學術寫作的價值。在教師特質部分，教師對於語料庫的了解與過往使用語料庫的經驗出現了統計上的正相關。也就是對語料庫了解越少的老師會對這種教學方式感到恐懼，而與語料庫資源接觸較多的教師則在使用上更有信心。而越有經驗的老師，也越有可能在未來的教學中採用 DDL。教學的熱情也影響了教師使用語料庫的動機。統計結果顯示，對教學更有興趣的老師，對語料庫保持著更積極、想學習的態度，並且想要發掘促進學習和提高學習者程度的方法。另外，教師的工作經驗也與使用語料庫具有相關性。非常資深的教師，大多是擁有超過 11 年教學經驗的老師們，對於語料庫比其他資源更有用的說法保持懷疑。而擁有 5 年以下教學經驗的老師大多強烈同意語料庫在學習上的好處；介於 6-10 年工作經驗的老師們，有一半保持中立，有一半同意語料庫對於教學有效。

由於工作坊僅 3 小時，少數教師也反映三個小時內要完成教學與活動太趕了，也難以全程保持專注。儘管如此，研究結果顯示教師對於語料庫的態度大多是積極的，如何在課程安排上，讓教師減少畏懼與增進學習成就感應該是未來要關注的重點。從問卷結果也可以注意到資深教師較不信任 DDL，或許對於資深教師的培訓必須更費心思的安排，以提高他們使用的動機。對比 Leńko-Szymańska (2017) 的研究可知，3 小時的培訓時間實在不足，筆者建議未來可以分階段、長期培訓，以消除教師的恐懼與增進學習的成就感。

3. 華語教師的語料庫素養

前面提到應用 DDL 於教學，教師至少需要具備四種能力 (Mukherjee 2006; Callies 2016, 2019)，本研究以此做為培訓目標。培訓對象以已經具有華語教學資格或經驗為主，也就是在職教師，如此一來，課程的設置就可以著重在對語料庫的認識和使用，而非基本的教學技巧。

以下本研究分別就教師對華語語料庫相關工具的認識、使用能力、解讀、分析語料的能力、融合語料庫與自身專業能力等方面來說明工作坊課程設計的理念。

3.1 認識華語語料庫與相關工具

國外最早建置的大型語料庫（一百萬詞）當屬 1960 年代中期的 Brown Corpus (Kucera and Francis 1967)，隨著計算機技術的快速發展，自 20 世紀 80 年代末到 90 年代初，各種功用的語料庫紛紛建立，有提供書面語的語料庫、口語的語料庫、為翻譯而做的平行語料庫等等。臺灣的中央研究院則是從 1990 年代開始蒐集語料，於 1995 年完成第一版 200 萬詞的漢語語料庫；1997 年完成 500 萬詞；至目前為止網路上開放的是 4.0 版本 1,000 萬詞語料庫¹。不過，中央研究院官網上的檢索系統乃早期的設計，除了功能較陽春外，也不是那麼易於使用 (user friendly)，但這個缺陷在這幾年國家教育研究院（以下簡稱國教院）所開發建置的臺灣華語文語料庫 (Corpus of Contemporary Taiwan Mandarin，以下簡稱 COCT) 系統中也得到了改善。

國教院自 2013 年接受教育部委託建置華語應用語料庫²，至今 (2022 年) 在官網開放的書面語語料庫已近 3 億 2 千萬詞 (319,712,694)，口語語料庫 2 千 100 多萬詞，中介語語料庫 70 多萬詞；除了這些自建的語料庫外，也把中研院平衡語料庫 4.0 版（一千萬詞）放在這系統平台上（以上數據皆來自官網文件說明）。另外也有華英雙語語料庫 1 千多萬字（吳鑑城、白明弘、林慶隆 2019:30）。也就是使用者從這個平台上可以使用各種不同類型的語料庫，除了中介語的語料庫詞數較少外，其他都在一千萬詞以上。而這個平台所開發的檢索功能或工具，大致涵蓋了國際上大型語料庫所擁有的，也較早期中研院的檢索系統易於使用。但與多數外語教學所廣泛使用的 SkE 平台相比，在資料的視覺化 (visualized) 呈現方面較為落後。這是可以理解的，畢竟

¹ 請參考中央研究院漢語語料庫：<http://asbc.iis.sinica.edu.tw/>

² 國家教育研究院華語文語料庫語能力基準整合應用系統：<https://coct.naer.edu.tw/>

COCT 從建置至今不到 10 年，而且是完全免費提供使用，與 SkE 商業化的做法不同。

SkE 漂亮的外觀與設計、易於使用等優勢獲得國際上多數使用者的青睞，在以往文獻中也顯示使用 SkE 有不錯的成效 (Thomas 2008)。筆者建議如果採間接 DDL，教師學會國教院的檢索與使用技巧即可，如果採直接 DDL，那麼考量學習者的學習動力以及快速獲得的成就感，學習者宜使用 SkE 這樣的平台較佳。這也說明教師在語料庫的學習方面，至少需要學習、熟悉這兩個系統。

上面提到 COCT 中的中介語語料不算多，如果要補充學習者的語料，做為對比分析的教學或學習材料，教師也可以接觸 TOCFL 學習者語料庫³，這個語料庫和 COCT 中介語語料庫的大小差不多，不同的是語料經過人工標記，是個帶有偏誤標記的語料庫，可較快速地找到教學者或學習者的偏誤語料 (張莉萍 2017)。大陸也建置了中介語語料庫－HSK 動態作文語料庫⁴，有 400 多萬字，雖然語料量較大，也標記了偏誤，只是寫作者都是高級程度的考生，檢索介面及語言也都是簡體字，對臺灣教學者或學習者而言，不是那麼實用⁵。

上述的平台或工具都是在他人已經建置好的語料庫上來使用，但在語言教學或是提供學習者輔助材料的時候，有時教學者需要自己建置的語料庫，如同前面文獻所示，不少教師或學習者自建語料庫 (Corino and Onesti 2019; Smith 2020)，而現在網路資料氾濫，想要快速取得文本或材料，最快的方法應該是從網路上擷取。在這方面，筆者推薦 BootCaT，它是個容易上手又免費的軟體。不推薦網路蜘蛛 (Web spider) 或網路爬蟲 (Web crawler) 工具，是因為這類工具較難控制使用者想要的主題內容文本 (吳鑑城、陳浩然、張俊盛 2017)。

一旦使用者建置了自己的語料庫，那麼就需要有分析語料的工具，從文獻中也可以看到，不少研究者使用 AntConc 這個軟體，這個軟體可以做到一般檢索或分析語料所需要的功能，例如，關鍵詞檢索 (KWIC)、排序 (sorting)、文本分布 (text distribution)、詞語搭配 (collocation)，也可以自己

³ TOCFL 學習者語料庫：<http://tocfl.itc.ntnu.edu.tw/>

⁴ HSK 動態作文語料庫：<http://hsk.blcu.edu.cn/Login>

⁵ 也因此本研究所提的華語語料庫基本上只侷限在使用繁體字的資源，其他非在台教學能用得上的，不在討論之列。

輸入參照語料，做出關鍵詞表 (keyword list) (張莉萍 2014)；同性質也為多數人使用的工具還有 WordSmith，只是使用這個工具需要付費。筆者個人使用 AntConc 這個工具的經驗非常好，基本上教學所需的功能都在其中。

以上大致介紹了欲使用 DDL 學習或教學者，需要熟悉的語料庫和分析工具。下一節則進入探討實際需要具備的使用能力。

3.2 與語料庫相關的檢索與使用能力

基本上，從前人的文獻可以得知使用語料庫最方便、簡單的地方就是檢索詞語，本研究以 COCT 為主，SkE 為輔，說明教學應用者應該具備的基本能力。

3.2.1 COCT 平台

3.2.1.1 檢索功能

教師或學習者在進入檢索之前，需要先有中文斷詞的概念，這是中文不同於英文語料檢索的地方，也是多數使用者在一接觸語料庫時容易出錯的地方。因為中文原始文本沒有空格 (space) 清楚的界定什麼是一個「詞」，如果使用者輸入「百貨公司」，「百貨」與「公司」之間沒有空格，那麼將檢索不到所需的語料，這與一般人熟悉的 Google 查詢方式不同。

COCT 檢索方式可以粗分為兩種，一種是簡易查詢 (simple query)；一種是 CQP 語法查詢。雖然 CQP 查詢類似程式語言，有固定的格式與寫法，一般人需要練習後才能上手，但它的功能較強，可以做到簡易檢索無法做到的功能。系統上也提供了實用的檢索範式供使用者參考⁶，可以大大增進檢索能力，是人文領域研究者或語言教師學得來的技巧。

COCT 的簡易檢索，不只能檢索詞，也能檢索詞組。例如，選擇簡易查詢 (見圖 1)，輸入「洗 + 澡」可以直接找出語料庫中所有「洗澡」中插一個詞的使用情況。如圖 2：

⁶ COCT 指令速查表：<https://coct.naer.edu.tw/cqpweb/doc/指令速查表.pdf>

COCT 書面語語料庫2019

Standard Query

洗 + 澡

Query mode: [指令速查表](#) [選單參考表](#)

Number of hits per page:

Restriction:

圖 1：「洗澡」拆開使用的查詢法（COCT 螢幕截圖）

· 嗚呼，新舊唐書上沒有說她給安祿山先生	洗過澡	，就可反證沒洗過澡乎哉？「
她給安祿山先生洗過澡，就可反證沒	洗過澡	乎哉？「正史」也者，是文字詐欺
上都臭，那說明病人很久很久沒有	洗過澡	。很多病人還害着肺結核或其他惡疾，但
二十一歲那一年，還不會說話，還沒有	洗過澡	，他被無情地虐待和羞辱損毀，一生中
詩中「初云政治決，繼云怨道行；三云	洗個澡	，四云待人誠。好話都說盡，
將他趕出門外，然後進了浴室。我	洗了澡	，並且先將一把剪刀放在浴盆旁，
是正確的，可見萬喜良早吩咐了。她	洗過澡	，全身疲憊，精神卻很亢奮，還到各
裡有個數，水我燒了很久，想	洗個澡	。剩下的水洗洗我和卡佳的衣服。
考慮下一步該怎麼辦，不過睡覺前一定得	洗個澡	。」「我馬上把內衣找出來。謝謝。衣櫥
熱水痛痛快快地洗了個澡。拉拉也給卡堅卡	洗了澡	。齊瓦哥渾身清爽，背朝屋裡，舒適地
他，讓他們進門，委婉地打聽回來後	洗過澡	沒有一斑疹傷寒仍然很猖獗。齊瓦哥剛到的那
把水擰乾，在太陽底下一曬就乾。	洗個澡	，換上乾淨衣裳。」「可總不大雅觀吧
沒有。」10清涼的夜晚降臨了。客人們	洗了澡	。女人們在指派給他們的房間裡整理床鋪
，就像她得先吃完羊肉才吃布丁，	洗了澡	才能睡覺一樣。「他知道他在說謊嗎
這會兒不是好好的？」「好，妳先	洗個澡	，我去幫妳沖一杯牛奶，喝完牛奶
就會去幫忙放洗澡水，然後準備和李叔叔	洗皇帝澡	。」「所謂「皇帝澡」就是放水到浴缸
沒有吃，就知道自己還有幾天畢業。	洗戰鬥澡	——學校的浴室很大，一間四四方方的建築物

圖 2：「洗澡」中插一個詞的查詢結果（COCT 螢幕截圖）

如果要中插兩個詞，輸入「洗 ++ 澡」；中插三個詞，輸入「洗 +++ 澡」；以此類推。但如果希望一次可以檢索中插 1~5 個詞，使用 CQP 檢索方式，就可以快速得到結果，不需要執行 5 次動作。如，選擇 CQP 語法檢索方式後（見圖 1 中的選項），輸入[word="洗"] []{1,5} [word="澡"]，可以找出像「洗了一個爽快的澡」這樣的短語⁷。另外像動詞重疊形式 AABB、ABAB 等

⁷ 更多檢索範例可參考張莉萍（2022:46）。

漢語的基本語法，也可以使用 CQP 語法檢索出來，因此簡易檢索與 CQP 檢索，筆者建議兩者都是教學者必備的能力。

3.2.1.2 再處理功能

檢索出來的語料，一般需要進行觀察、分析，如果知道基本的一些再處理功能，可以事半功倍。舉例來說，經過排序，利用靠右排序或靠左排序可以較輕鬆地觀察到固定的形式，例如，關鍵詞左邊都是標點符號，即可推測關鍵詞都出現在句首。在 COCT 語料庫中實用的再處理功能有 7 種，如：(1) 語料隨機取樣 (Thin)；(2) 頻率彙整 (Frequency breakdown)；(3) 觀察詞彙在語料類別中的分佈 (Distribution)；(4) 按關鍵詞左邊或右邊排序 (Sort)；(5) 搭配詞分析 (Collocations)；(6) 下載查詢結果 (Download)；(7) 樣本分類標記 (Categorise)；其中，「隨機取樣、詞彙分布、排序、下載、樣本分類」這幾個功能技術性不高，在此略過不談。下面僅介紹頻率彙整以及搭配詞分析這兩個教學者一定要能掌握的功能。

(1) 頻率彙整

回到前面「洗澡」中插 1~5 個詞的例子，檢索後點選 frequency breakdown 的功能可以得到像表 1 的結果：

表 1：「洗澡」中插 1~5 個詞的詞塊頻率前 10 名

No.	Search result	No. of occurrences	Percent
1.	洗個澡	326	38.17%
2.	洗過澡	157	18.38%
3.	洗了澡	81	9.48%
4.	洗一次澡	66	7.73%
5.	洗了個澡	58	6.79%
6.	洗起澡	12	1.41%
7.	洗三次澡	9	1.05%
8.	洗了一次澡	7	0.82%
9.	洗了一個澡	6	0.70%
10.	洗戰鬥澡	5	0.59%
11.	洗幾次澡	5	0.59%

也就是說，系統自動統計出「洗澡」分離的情況，其中「洗個澡」最高頻，共出現 326 次，佔所有「洗澡」分離使用語料的 38.17%；其次是「洗過澡」，然後是「洗了澡、洗一次澡、洗了個澡……」；這個功能為離合詞的教學提供了非常有用的資訊，教師在設計教學材料時，得以快速地知道教什麼樣分離的方式或是給什麼典型的例子，對學習者最有幫助。因為學習時間有限，教學者不可能要學生在第一次使用時即學會所有分離情況，高頻、典型的用法是教學者或教材編輯者選擇材料時優先的考量。表 1 顯示的是以不同「詞」組合情況為頻率計算依據，這個功能還可以選擇以「類」的概念呈現統計數據，如圖 3 所示，使用者選擇 frequency breakdown of annotation only，則可以得到以「類」為組合的統計頻率，如圖 4 所示。

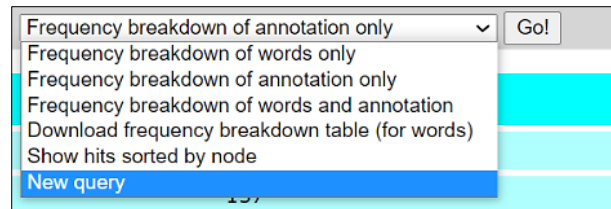


圖 3：頻率彙整功能的選項介面（COCT 螢幕截圖）

No.	Search result	No. of occurrences	Percent
1	VC Nf Na	326	38.17%
2	VC Di Na	254	29.74%
3	VC Neu Nf Na	101	11.83%
4	VC Di Nf Na	58	6.79%
5	VC Di Neu Nf Na	24	2.81%
6	VC Na Na	20	2.34%

圖 4：「洗澡」中插 5 個詞的詞類組合頻率（COCT 螢幕截圖）

從圖 4 可以看到前 5 名的詞類組合為⁸：

- (1) 洗 + 量詞 + 澡
- (2) 洗 + 時貌詞（如：了、著、過） + 澡
- (3) 洗 + 數詞 + 量詞 + 澡
- (4) 洗 + 時貌詞 + 量詞 + 澡
- (5) 洗 + 時貌詞 + 數詞 + 量詞 + 澡

⁸ COCT 採用中央研究院平衡語料庫的詞類標記，以下簡單解釋圖 4 中的標記：VC（動作及物動詞）、Nf（量詞）、Na（普通名詞）、Neu（數詞）、Di（時貌助詞）。

前 5 名組合即佔了接近全體的 90%，在設計教材時，可以以這幾種型態為主。上述簡單地介紹了頻率彙整在教學應用時的優勢，這功能讓使用者對於所要查詢的成分出現頻率的分布高低一目了然。

(2) 搭配詞分析

搭配詞分析幾乎是所有語料庫平台都會有的功能，因為分析語料時，需要依據詞與詞之間共同出現關係強度來解釋或說明典型的用法，而搭配詞分析這個功能在教學上的最大功用就是找出詞語的典型用法，提供教學設計或是學習典範。例如，中高級教材中出現「加上」這個關連詞，當使用者從語料庫中檢索這個關鍵詞時，出現 3 萬 8 千多筆的語料，在這麼多語料中，如何知道什麼是「加上」的典型用法？如何給予學習者典型例句？這時，可以藉著**搭配詞分析**這個功能，一般的做法是觀察關鍵詞左右 5 個詞的搭配關係，不過由於這是個關連詞，需要觀察的範圍更廣。筆者建議設定觀察左邊 10 個詞，結果發現位於左邊和「加上」搭配最緊密的詞是「再」，在操作時，點選螢幕上的「再」，也可以馬上看到 97% 的機率都出現在「加上」左邊的第 1 個位置（下圖中的 -1 處）。

Within the window -10 to -1, 再 occurs 10,596 times in 9,061 different texts (expected frequency: 388.954)			
Distance	No. of occurrences	In no. of texts	Percent
-10	23	23	0.2%
-9	40	38	0.4%
-8	39	39	0.4%
-7	34	34	0.3%
-6	45	45	0.4%
-5	34	31	0.3%
-4	20	20	0.2%
-3	32	32	0.3%
-2	25	25	0.2%
-1	10,304	8,862	97.2%

圖 5：「再」出現在「加上」左邊 10 個詞的位置分布（COCT 螢幕截圖）

左邊搭配緊密的前幾名還有「由於、又、還、若、因為、如果」等等，右邊前幾名有「使得、使、又、所以」等等。利用這些資訊，教師可以構成語言定式，來加強學生篇章連貫銜接的能力。如下例：

定式：（因為/由於）…（再）加上…所以/因此/於是

例句：台北交通的擠和亂，主要是因為人口和車輛增加得太快了，再加上有許多不遵守交通規則的駕駛和行人，於是路上的情況越來越糟糕。

以上的例句來自語料庫，像這種需要長段語境的學習材料，一般人很難憑空想像造出自然的句子，但從語料庫中找尋符合所需的材料不是難事，更可以將找出來的定式應用在寫作教學中。

搭配詞分析這個功能也是研究者在分析近義詞必備的利器。前人有不少利用搭配詞分析方法的文獻，如，蔡美智、黃居仁、陳克健（1999）對於近義詞「快樂、高興」的分析，他們利用詞語搭配關係來分析這兩個詞彙的句法行為，進而解釋語義的差異；Xiao 與 McEnery（2006）也利用搭配關係來分析多組近義詞，如「結果、後果」，透過他們共現的詞語分析出他們的語義韻（semantic prosody）不同，「結果」多用於正面或中性的情境中，「後果」則用於負面情境。

從上述「簡易檢索」、「CQP 檢索」、「頻率彙整」與「搭配詞分析」這些功能的演示，希望能讓讀者理解為什麼本研究選擇使用這些功能做為教學者應該具備的語料庫素養。接下來，筆者介紹另一個常用的語料庫平台。

3.2.2 Sketch Engine

在外語 DDL 的教學中，多數都使用 Sketch Engine (Kilgarriff et al. 2004)，在華語教學中也有不少研究者推薦這個平台（Smith 2011；王炳勻等 2020）。就本文前一小節所介紹的 COCT 與這個平台來比較，檢索方式與功能大同小異，SkE 除簡易檢索外，也提供 CQL 的進階檢索方法，CQL 與 CQP 語法相似度很高，因此不論使用者學會哪一種，對於另一個平台的使用都有相輔相成的效果。這個商業平台的整體設計賞心悅目，而且是英語的介面，對於外籍學習者而言，至少不讓他們覺得畫面索然無趣或覺得困難而畏懼，可以引起學習的興趣。因此，如果採用直接 DDL 的方式教學，筆者建議讓學生使用 SkE⁹；如果採間接 DDL 的方式則建議使用 COCT。筆者建議華語教師先從 COCT 學習起，熟悉各式的檢索技巧後，再採用 SkE 也就輕而易舉。

不過，SkE 中有三個功能是 COCT 沒有的，而且這些功能對於語言教學或學習有不小的助益。一是詞彙對照素描；二是視覺化的功能；三是好例句的功能，以下簡單介紹。

⁹ 但因為 SkE 的中文處理方式，幾乎以簡體字為主，就筆者的使用經驗，不論是現在內建的中文網路語料庫或是筆者自己建置的語料庫，因為都得採用系統所提供的斷詞程式處理，然而他們對於繁體字的斷詞處理得不理想，進而導致分析不準確的問題。

3.2.2.1 詞彙對照素描 (sketch difference)

這個功能可以直接讓使用者輸入兩個相似詞，系統就自動統計出他們的差異處。這快速方便的工具非常適合教學者或學習者使用，如果採用 COCT 的做法，通常得分別檢索這兩個詞，再從搭配詞關係中來分析兩詞語的異同。

圖 6 是使用詞彙對照素描來分析「結果、後果」兩詞彙用法的差異，此處僅擷取充當這兩個詞的修飾語差別。圖中靠近上面（綠色）的詞代表和「後果」有緊密關係的詞，而越靠近下面（紅色）的詞是則是和「結果」高度相關。學習者透過色彩即可清楚地看到兩個詞修飾語的不同。以第一列為例，中間那些數字左邊是「嚴重」和「後果」共同出現的次數，右邊是「嚴重」和「結果」搭配出現的次數。同理，從最下面這列可知母語者說「得出的結果」，不說「得出的後果」。



A_Modifier			
严重	11053	329	...
不良	957	144	...
不利	1185	285	...
造成	10803	5345	...
带来	7279	5175	...
产生	5507	4701	...
导致	3501	7548	...
作用	14	6098	...
想要	23	6520	...
得到	35	10606	...
努力	12	9585	...
得出	0	6137	...

圖 6：「結果、後果」的詞彙對照素描分析畫面（SkE 螢幕截圖）

3.2.2.2 視覺化呈現

前面提過 SkE 是商用平台，視覺化的設計是它的強項。圖 7 是「結果、後果」的視覺化呈現畫面，從視覺化呈現中可以看出「結果」常常跟「公布、篩選、得出」等動詞一起出現；「後果」常跟「造成、承擔、帶來」等動詞一

起出現，視覺化可以清晰呈現兩者的差異，再透過進一步的教學活動來讓學習者鞏固知識，內化習得。

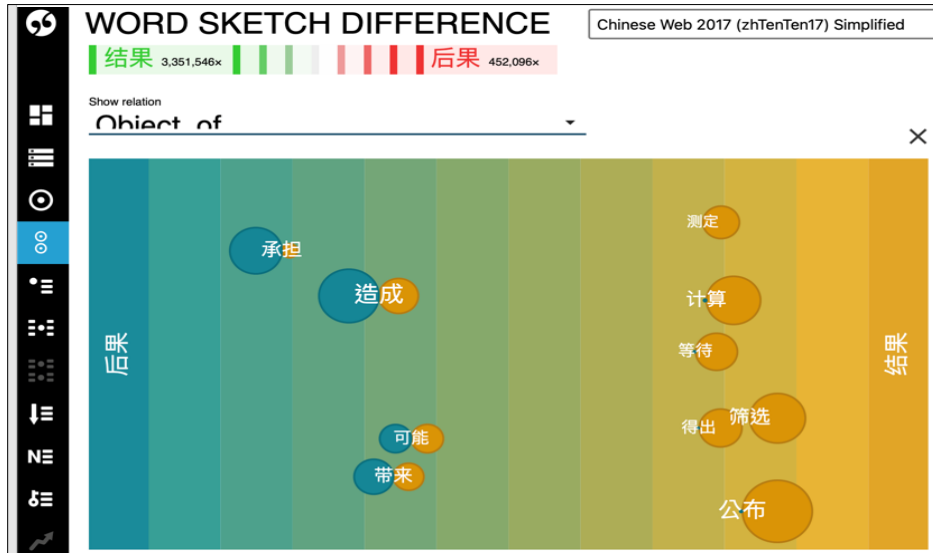


圖 7：與「結果、後果」搭配的動詞視覺化呈現畫面（SkE 螢幕截圖）

圖 8 也是透過視覺化呈現檢索結果。使用者利用索引典功能輸入「得到」這個詞，可以找出和它相近語義的詞語，如「受到、找到、取得、失去、做出……」。圖中圈圈的大小、顏色等等都有代表的意涵，不在此詳述。

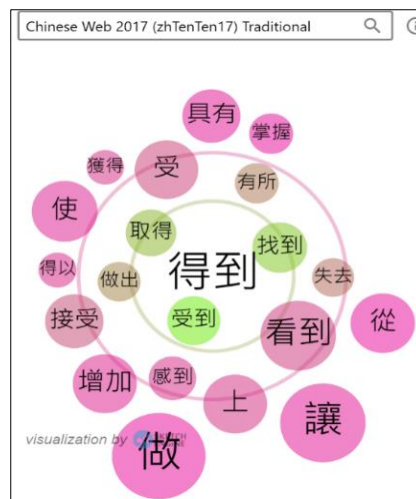


圖 8：和「得到」相近的詞語視覺化效果（SkE 螢幕截圖）

3.2.2.3 好例句 (GDEX)

好例句是 SkE 專門為語言教學設計的一項功能，目的是為了提供教學所需的語料，因此系統考量下面幾個面向（1）-（5），計算並選出適合學習者學習的例句。

- (1) 句子長度
- (2) 句中使用的詞彙難度
- (3) 語料是否來自爭議性主題（宗教、政治）
- (4) 句子有無足夠語境
- (5) 句中代名詞的指涉是否清楚

以「吃……飯」這個結構功能為例，使用者可以自己設定、決定要看的筆數。下圖的結果顯示沒有太複雜、太長的語料，而且和日常生活語境結合，適合學習者觀察。可見這功能的確可以有效地為教師省掉一些篩選例句的時間，教師也可以直接讓學習者使用這個功能來觀察語料。

<input type="checkbox"/> Details	sentence
<input type="checkbox"/> ① vikilife.com	<S> 我和他 吃过一次饭。 </S> PN P PN VV AS CD M NN PU
<input type="checkbox"/> ① hbxuli.com	<S> 不过我得先回家 吃个饭！ </S> AD PN VV AD VV VV M NN PU
<input type="checkbox"/> ① hsw.cn	<S> 觉得无非是 吃个饭，没有意识到背后可能潜藏一些危险信号。” </S> VV AD VC VV M NN PU AD VV VV NN VV VV CD NN NN PU PU
<input type="checkbox"/> ① poco.cn	<S> 偶然认识了他， 吃过几顿饭，挺随意的一个人。 </S> AD VV AS PN PU VV AS CD M NN PU AD VA DEC CD M NN PU
<input type="checkbox"/> ① 1688best.com	<S> 我过去早上 锻炼，后来增添了下午运动，现在晚上 吃过饭 也 散散步。 </S> PN NT NT NT VV PU AD VV AS NT NN PU NT NT VV AS NN AD VV NN PU
<input type="checkbox"/> ① neidian8.cn	<S> 那股气息强大无比今晚一起 吃个饭 吧真是一个有故事的大叔。 </S> DT M NN VA AD NT AD VV M NN SP AD VC CD M VE NN DEC NN PU
<input type="checkbox"/> ① cp.com.cn	<S> 他把车开到他家门口，让我到他家 吃了饭 再走。 </S> PN BA NN VV PN NN PU VV PN VV PN NN VV AS NN AD VV PU

圖 9：「吃……飯」這個結構的 GDEX 結果（SkE 螢幕截圖）

3.2.3 BootCat

雖然上述兩種平台上都有建置好的大型語料庫，但從前人文獻中也可以發現不少教學課堂上或是研究都是需要自建語料庫以符合使用者的需求，如前面所提到的給美髮科學生學習的材料或是財經科學生學習的內容。本研究中推薦使用 BootCaT 這個工具，雖然 SkE 也有可自建語料庫的功能，但因為斷詞結果不理想（請參見註解 9），也不接受使用者自行斷詞的語料，因此筆者推薦使用 BootCat，它有良好的圖形使用者介面（GUI）、清楚簡潔的設計及引導，使用者擁有一般的科技能力即可自建語料庫。只要使用者想好一組（5 個以上）代表自己所要找的文本領域的詞彙，即可以輕鬆取得文本。舉

例來說，輸入「食譜、燉、煮、炒、炸、煎」就可以找到有關烹調主題的文本。如果教師帶著學生做一次，學生大概也能學到怎麼快速的蒐集語料。在這過程中，教師也可以透過讓學生輸入專門領域的代表詞彙，來訓練他們對於關鍵詞彙的學習與掌握，不失為一個可行的教學活動。

3.2.4 AntConc

有了自己建置的語料後，就需要有能分析語料的工具。筆者推薦免費單機版的工具 **AntConc** (Anthony 2020)，下載、安裝容易，語言分析所需要的一般功能幾乎都有，如關鍵詞檢索、詞頻表、排序、搭配分析等等。已經具有一般檢索能力的教師，使用這個工具除了要熟悉操作介面外，唯一需要學習的一個新功能就是製作關鍵詞表，在語料庫為本的研究中，這是一個重要方法，它的原理是拿一個受觀察的語料庫與一個參照語料庫來比較，藉由兩者顯著不同用法（頻率）來發現受觀察語料庫的特徵（張莉萍 2014）。而在教學上，如果要製作專業華語的詞表，就需要利用這個功能，關鍵詞表可以讓使用者清楚看到自己所建置語料庫中的專業詞語，教學者可以更有針對性地去設計教學內容。因此如果 **DDL** 教學內容是專業華語的課程，教師需要有這個能力，反之，也可以略過不學。

要注意的是，在把自己建置的語料放入 **AntConc** 分析之前，必須先把語料做好斷詞的處理，因為如果機器無法辨識什麼是一個詞，也就沒辦法做好語言處理，簡單舉例來說，使用者想檢索「了」的用法，如果沒經過斷詞的語料，機器會把「了解」的用法也計入。雖然 **Ant*** 系列工具也開發了斷詞工具 **SegmentAnt**¹⁰，但經過筆者的實地使用，不管是 **SkE** 或是 **SegmentAnt** 的效果都不佳，筆者建議繁體字斷詞採用中研院的工具¹¹或是 **COCT** 的分詞系統，前者可以批次處理，也就是可以處理大量需要斷詞的文本；後者則必須把文本複製貼上，適用少量斷詞文本。

3.3 語料解讀與分析能力

教師除了要有前述使用語料庫的功能與技巧的能力，實際上教學者還必須具備語料解讀與分析的能力，才能提供給學習者有效的學習內容，因此解讀與分析語料是 **DDL** 教學必備的能力。一般科班出身的華語教師，受過語

¹⁰ **SegmentAnt**: <https://www.laurenceanthony.net/software/segmentant/>

¹¹ 中央研究院中文斷詞系統：<https://ckip.iis.sinica.edu.tw/project/ws/>

言本體與教學法的訓練，基本上應該沒有太大的問題，教學者常常是因為不熟悉語料庫可以協助教學到什麼地步，也就沒有辦法有效的應用語料庫引導學生學習。筆者以「得到、達到」這組詞為演示的例子，首先教學者要知道學習者的困難在哪裡，才能對症下藥。此時，教學者可以透過中介語（學習者）語料庫初步探索，發現學習者的偏誤是發生在動詞與賓語的搭配，如下面例（1）-（4）（引自 HSK 動態作文語料庫）：

- （1）我覺得以罰款的方法 得到* 這個目的，是因為吸煙的人恐怕被罰款，儘量在街上努力不吸煙。
- （2）隨著社會發展，煙的銷售也 得到* 了前所未有的發展。
- （3）而且法律的實施應該嚴格履行才能 得到* 效果，因而作者也贊同對違法抽煙者的罰款措施。
- （4）雖然我沒有這種工作的經驗，但是我在華美服裝學校和文化服裝學校的成績都很好，得到* 了相當高的水平。

這 4 個偏誤顯示的是學習者用「得到」取代了「達到」，這個時候如何讓學習者知道兩者的差別，如果選擇 COCT 做為分析工具，那麼教學者要知道使用哪個功能來分析，應該引導學習者觀察什麼；如果選擇 SkE，那麼以詞彙對照素描功能做出來的結果（如下圖 10），教學者需要有解讀這些資訊的能力，應該知道賓語（Object）這一欄是這兩個動詞關鍵的差異，而這兩個詞的主語或修飾語的差異不是教學的重點。因為學習者的偏誤主要出現在動詞和賓語搭配的錯誤，從圖 10 可以看到「得到」後面幾乎不會接「水準、標準、目標、目的」（後面顯示 0 次），這些詞只能做為「達到」的賓語。

得到 40,039x			達到 82,830x		
Object			SentObject_of		
好處	323	0	想像	47	0
銅牌	243	0	感受	78	0
銀牌	235	0	渴望	21	0
選票	436	7	獲知	14	0
票	562	40	輸	18	0
分	996	155	高興	22	0
效果	124	2316	希望	706	1054
美元	135	4232	期望	14	66
目標	8	4519	以期	27	245
水準	0	1075	可望	83	1295
標準	0	1807	率	0	95
目的	0	4615	不足以	0	24
			本學	43	0
			安打	46	0
			中華隊	65	0
			感受	38	0
			機會	195	79
			我們	306	143
			人數	0	1062
			產量	0	357
			產值	0	353
			總額	0	621
			總數	0	426
			成長率	0	1250
			答案	78	0
			教訓	38	0
			靈感	8	0
			結論	75	0
			好處	24	0
			訊息	73	0
			成果	22	9
			效果	9	30
			成就	0	9
			水平	0	7
			目的	0	23
			目標	0	95

圖 10：「得到、達到」的詞彙對照素描（SkE 螢幕截圖）

3.4 培訓課程的設計與規劃

在上述討論的基礎下，本研究為在職教師設計了 12 個小時的課程「語料驅動教學工作坊」。課程分為 4 週進行，每週上課 3 小時，授課大綱請見表 2¹²。課程主要著重在 COCT 語料庫的檢索技巧、教學應用以及 DDL 的教學活動設計方法。

表 2：「語料驅動教學工作坊」每週課程內容

週數	課程內容
第一週	語料庫基本概念；COCT 語料庫的介紹與操作技巧
第二週	作業檢討；COCT 進階使用技巧；學習觀察語料的方法、分析語料技巧；介紹與操作 Sketch Engine
第三週	作業檢討；學習操作 BootCat；介紹與使用 AntConc
第四週	作業檢討；DDL 語料呈現與教學活動設計

課程受 COVID-19 疫情影響，以 Google Meet 遠距同步教學進行，於 2021 年 10 月舉辦，利用 4 個週末免費授課。因為是線上教學，也吸引了一些海外教師參與，因此，課程錄影檔在當天課後也會上傳到 Youtube 平台，供時間無法配合或希望課後複習的教師觀看¹³。後來發現，這個做法意外地得到受訓教師的喜愛，認為錄影可以幫助他們反覆觀看，對學習很有幫助。

每次課程的相關知識、語料庫的實際操作均是由授課教師（即筆者本人）先行講解、示範，課程中間也設計了練習題，受訓教師再根據授課教師的示範實際進入語料庫中操作、檢索。在同步遠距上課的過程中，每週有兩位助教在旁協助，以排除受訓教師在操作語料庫時遇到的問題。DDL 教學法強調「實作」，因此，前三堂課程結束後均安排回家作業，作業經由授課教師或助教批改，並於下次上課時檢討。

第四堂課程結束後，受訓教師須結合課程中所學與自身華語教學的經驗，設計出語法點或詞彙的 DDL 教學活動，做為整體課程的評測。除了作業及評測外，在四週的課程結束後，筆者也邀請所有受訓教師填寫問卷，以瞭解教師對整體課程的評價，做為日後培訓課程的參考。

¹² 原來設想讓教師具備建置語料庫的能力，規劃了 BootCat 的使用介紹，不過，實際授課後，發現時間不足，只好略過。

¹³ 有興趣的讀者可以觀看工作坊錄影事後的剪輯課程：<https://www.youtube.com/playlist?list=PLX0mYsdwnR6N6Fuj8PPGD7X4vin3aaDJY>

4. 課後問卷與結果分析

「語料驅動教學工作坊」共計有 79 位具有實際華語教學經驗的教師報名參加，73 位於第 1 堂課實際到課，他們的華語教學經驗介於半年至 38 年之間，平均的年資為 9.8 年。課後問卷於最後一堂課程後發送，採匿名填答方式，73 位受訓教師中，共有 50 位填覆問卷，填答率約 68%。

4.1 「語料驅動教學工作坊」問卷設計與分析方法

本研究參考了 Leńko-Szymańska (2015) 研究中使用的問卷，該問卷由 43 個問題組成，針對學生的動機、先備知識、學習的結果、未來使用的態度、對課程的評價這 5 個面向來設計。本研究則針對在職教師的特徵微幅調整了部分問題，最後問卷由 43 個問題組成，其中 4 題為開放式問題，39 題為李克特量表問題。

分析問卷的方法則是將問卷裡 39 題李克特量表問題中的 36 題分為 7 個面向¹⁴，分別是「參與課程的動機」、「課前相關先備知識」、「課後對操作語料庫的能力」、「對語料庫的負面態度」、「執行、設計 DDL 教學活動的能力」、「未來在課程中應用 DDL 教學法的可能性」、「對整體課程的評價」，並使用 R studio 以斯皮爾曼等級相關係數檢定 (Spearman's rank correlation coefficient) 來揭示各個面向之間的關係。

4.2 問卷量化分析結果

4.2.1 問卷填答者的背景資訊

關於填答者的背景共有 4 題，由表 3 可見，這 50 位教師中，多數（約三分之二）擁有 5 年以上教學經驗；70% 擁有碩士以上學歷。有 84% 的教師完整參與 4 週課程，有 62% 的教師完成每週課後作業。在沒有約束力的情況下，多數教師每週到課，並完成作業，可見這些在職教師渴望求知的精神。

¹⁴ 所有 39 題中，只有 36 題用於分析，這是因為其中一題涉及課程的活動，而該活動最終因為課程時間不足沒有進行。另外，本研究也發現原先設計的兩題無法反映受訓教師的真實狀況，如「我剛好可以配合培訓課的上課時間」原先設計在「參與課程的動機」面向中。但經過討論後，我們認為時間能否配合與教師個人參與課程的動機無關。或許有教師時間剛好能配合，但學習動機非常強烈。這些題目便沒有納入分析之中。

表 3：問卷填答者背景資訊

華語教學經驗	人數（比例）	教育程度	人數（比例）
1 年（含）以下	2（4%）	大學	12（24%）
1 年以上-3 年	7（14%）	碩士	35（70%）
3 年以上-5 年	8（16%）	博士	3（6%）
5 年以上-10 年	12（24%）		
10 年以上	21（42%）		
是否完整參與 四週培訓課程	人數（比例）	是否每一次上課 前完成了作業	人數（比例） ¹⁵
是	42（84%）	是	31（62%）
否	8（16%）	否	19（38%）

4.2.2 問卷題目與填答結果

本問卷原始各題的選項分別為「非常同意」、「同意」、「不同意」、「非常不同意」及「沒有意見」5 個。在轉化分數時，略過「沒有意見」這個選項，而「非常同意」、「同意」、「不同意」及「非常不同意」4 個選項，分別轉換為 4、3、2、1 分。表 4 係依照問卷的 7 個面向分類，呈現題目以及每一題的平均分數及標準差資訊。

表 4：問卷題目與量化結果¹⁶

面向	題目	平均	標準差
參與課程的動機	1a. 我想了解語料庫和語料庫語言學	3.78	0.42
	1b. 我想了解新的華語教學法	3.84	0.37
	1c. <u>我報名不到其他工作坊，轉而報名本課程</u>	3.62	0.64
		3.75	

¹⁵ 因發放問卷的時間為第四堂課程後，尚未到達繳交最終評測的時間，因此這邊指的「每一次課前完成作業」是指前三週作業的繳交情形。

¹⁶ 以下底線標示的題目，「非常同意」、「同意」、「不同意」、「非常不同意」分別轉換為 1、2、3、4 分。如「參與課程的動機」面向中題目「1c. 我報名不到其他工作坊，轉而報名本課程」，當初設計的用意是希望瞭解受訓教師是否只是為了滿足語言中心的相關進修規定而報名課程。若教師勾選「非常同意」，本研究認為該師的學習動機較低，因此在分析前先將分數轉置。

表 4：問卷題目與量化結果（續）

面向	題目	平均	標準差
課前 相關 先備 知識	2a. 我以前就聽說過語料庫和語料庫語言學	3.29	0.79
	2b. 我大概了解語料庫和語料庫語言學的內容	3.00	0.82
	2c. 我在其他課程中曾使用過語料庫	2.82	0.95
		3.03	
課後 對操 作語 料庫 的能 力	3a. 我知道不同的語料庫資源和工具	3.44	0.67
	3b. 我知道如何操作不同的語料庫資源和工具	3.16	0.69
	3c. 閱讀和理解索引行 (concordance lines) 對我來說並不難	2.96	0.75
	3d. 我知道如何運用語料庫檢索目標詞彙 (word)	3.38	0.49
	3e. 我知道如何運用語料庫檢索短語 (phrase) 訊息	3.19	0.64
	3f. 我知道如何運用語料庫檢索語法 (grammar) 訊息	3.02	0.65
	3g. 我知道如何運用語料庫檢索篇章 (discourse) 相關訊息	2.76	0.72
	3h. 我會使用語料庫來解決我在其他課程中或日常使用華語時遇到的語言問題	3.37	0.57
	3i. 我會使用語料庫來輔助我寫作 (writing)	2.71	0.65
	3j. 我會使用語料庫來輔助我閱讀 (reading) 承上題，如果我「會」使用語料庫來解決我在其他課程中或日常使用華語時遇到的語言問題，是因為... ¹⁷	2.77	0.69
	3k. 我知道如何運用語料庫檢索目標詞彙 (word)	3.33	0.47
	3l. 我知道如何運用語料庫檢索短語 (phrase) 訊息	3.22	0.55
	3m. 我知道如何運用語料庫檢索語法 (grammar) 訊息	3.04	0.68
	3n. 我知道如何運用語料庫檢索篇章 (discourse) 相關訊息	2.83	0.66
		3.09	
對語 料庫 的負 面態 度	如果我「不會」使用語料庫來解決我在其他課程中或日常使用華語時遇到的語言問題，是因為...		
	4a. 我覺得操作語料庫不太容易	2.87	0.87
	4b. 我覺得語料庫的檢索結果不太有用	1.95	0.72
	4c. 我覺得操作語料庫很麻煩，我比較喜歡傳統查詢的方式	2.14	0.56
		2.33	

¹⁷ 此題組 (3k, 3l, 3m, 3n) 中的 4 題，選項為「非常同意」、「同意」、「不同意」、「非常不同意」、「我不會使用語料庫來解決語言問題」，前四個選項轉換的分數與其他題目相同，而最後一個選項以空缺資料的方式處理。

表 4：問卷題目與量化結果（續）

面向	題目	平均	標準差
執行、設計 DDL 教學活動的能力	5a. 我認為把語料庫應用於語言教學中是有好處的	3.62	0.49
	5b. 我知道在語言教學中應用語料庫的不同方式	3.19	0.57
	5c. 我知道如何在備課時使用語料庫	3.28	0.57
	5d. 我知道該如何將基於語料庫 (corpus-based) 的素材融入教學	2.95	0.61
	5e. 我知道該如何為學習者設計基於語料庫 (corpus-based) 的活動	2.86	0.59
	5f. 我知道該如何把基於語料庫 (corpus-based) 的活動融入語言教學中	2.93	0.54
		3.14	
未來在課程中應用 DDL 教學法的可能性	6a. 備課時，我會使用大型通用語料庫 (large general corpora)	3.08	0.61
	6b. 我會讓我的學生參考大型通用語料庫 (large general corpora) 來解決語言問題	2.50	0.67
	6c. 我會讓我的學生參考我建製的專門語料庫	2.58	0.72
		2.75	
對整體課程的評價	7a. 我覺得課程內容很實用，資訊豐富	3.62	0.49
	7b. 我覺得這個培訓課程的作業量很大	2.67	0.66
	7c. 我喜歡這個課程的上課方式（例：老師先講解，再進行實作練習題）	3.41	0.65
	7d. 我覺得最後的評量方式可以應用所學	3.20	0.51
	7e. 您會如何評價本培訓課程的授課內容？	3.74	0.49
		3.36	

從分數上的呈現，可以看出關於參與課程的動機這個部分平均分數最高 (3.75)，其次是對課程的評價 (3.36)；最低的是對語料庫的負面態度 (2.33)。這可以顯示來參加工作坊的教師具有很強的動機，對語料庫抱持的負面態度相對地低；而對課程的滿意度也相當高。

4.3 相關性檢定分析與討論

除了上述 7 個面向外，本研究將教師的華語教學經驗及個人的教育程度納入分析。所有 9 個面向之間的相關性，經統計檢定後整理如表 5，然後逐項討論。

表 5：與教師應用語料庫於課堂有關各面向相關性之檢定結果^{18,19}

	經驗	學歷	動機	先備	能力	態度	課室	可能	評價
經驗	1								
學歷	.40**	1							
動機	.33*	.21	1						
先備	.23	.37**	.31*	1					
能力	-.12	.22	.12	.49***	1				
態度	.42*	-.08	.13	-.29	-.72***	1			
課室	.04	.23	.24	.28*	.52***	-.59**	1		
可能	.00	-.15	.06	.36*	.45**	-.16	.23	1	
評價	.14	.15	.43**	.22	.29*	-.24	.47***	.16	1

註：* p -value < .05 (低度相關)；** p -value < .01 (中度相關)；*** p -value < .001 (高度相關)

(1)「教師經驗」與其他面向的相關性

從表 5 得知「教師經驗」與「教師教育程度」、「參與課程的動機」、「對語料庫的負面態度」都呈現顯著的相關性，表示教學年資越長的教師，對於想學習 DDL 教學法的動機也越強。但另一方面，雖然資深教師參與課程的動機較高，同時對語料庫抱持負面的態度也較高。

另外，教師的經驗與他們在課前所具備的相關先備知識並沒有顯著的關聯 ($\rho = .23$; p -value = .11)。這或許可以說明資深教師雖然渴望瞭解新的教學法，但在他們長年的教學旅程中，接觸到語料庫的機會並不比年輕的教師

¹⁸ 為了表格的閱讀方便，「教師經驗」、「教育程度」、「參與課程的動機」、「相關先備知識」、「課後對操作語料庫的能力」、「對語料庫的負面態度」、「執行、設計 DDL 教學活動的能力」、「未來在課程中應用 DDL 教學法的可能性」、「對整體課程的評價」在此表中分別以「經驗」、「學歷」、「動機」、「先備」、「能力」、「態度」、「課室」、「可能」、「評價」代表。

¹⁹ 表格中呈現的數字為 ρ 值 (rho)，代表兩因子之間的相關性，介於 1 到 -1 間，1 代表兩因子呈完全正相關；-1 代表兩因子呈完全負相關；0 則代表兩因子完全沒有相關性。根據 Dancey 與 Reidy (2004)，在絕對值後 $\rho > .70$ ，表示兩因子完全相關 (very strong relationship)； ρ 介於 .40-.69，表示兩因子高度相關 (strong relationship)； ρ 介於 .30-.39，表示兩因子中度相關 (moderate relationship)； ρ 介於 .20-.29，表示兩因子低度相關 (weak relationship)； ρ 介於 .01-.19，表示兩因子微弱或無相關 (no or negligible relationship)。

高。而資深教師雖然對語料庫抱持負面的態度，卻與他們在「課後對操作語料庫的能力」（ $\rho = -.12$ ； $p\text{-value} = .40$ ），及「執行、設計 DDL 教學活動的能力」（ $\rho = .04$ ； $p\text{-value} = 0.80$ ）無關。這或許說明雖然資深教師對語料庫操作感到吃力，但培訓後的成效並不小於資淺的教師。

（2）「教師教育程度」與其他面向的相關性

教師受過的教育程度越高，對語料庫也具備了更多的先備知識（ $\rho = .37$ ； $p\text{-value} = .008^{**}$ ）。值得注意的是，教師的教育程度，與反映師培課程成效的兩個面向，分別是「課後對操作語料庫的能力」（ $\rho = .22$ ； $p\text{-value} = .12$ ）及「執行、設計 DDL 教學活動的能力」（ $\rho = .23$ ； $p\text{-value} = .11$ ）皆沒有顯著的相關性，綜合上述兩點的討論，說明了不論教師的學經歷如何，都能在本培訓課程得到相當的收獲。

（3）「參與課程的動機」與其他面向的相關性

越資深、在課前具備越多相關先備知識的教師，參與本課程的動機也越強（ $\rho = .33$ ； $p\text{-value} = .02^*$ ； $\rho = .31$ ； $p\text{-value} = .03^*$ ），而參與課程動機強的教師，給予課程的評價亦越佳（ $\rho = .43$ ； $p\text{-value} = .002^{**}$ ），但與教師課後對語料庫的操作能力（ $\rho = .12$ ； $p\text{-value} = .40$ ）、DDL 教學活動設計、執行能力（ $\rho = .24$ ； $p\text{-value} = .10$ ）皆不具相關性。這個現象說明不論教師在課前的參與動機高低，皆能透過課程內容建立起操作語料庫的能力與信心，推測與課程內容設計有關。

（4）「課前相關先備知識」與其他面向的相關性

在課前對語料庫有相關先備知識的教師，除了對培訓課程有更高的參與動機外（ $\rho = .31$ ； $p\text{-value} = .03^*$ ），在課後對操作語料庫能力（ $\rho = .49$ ； $p\text{-value} = .000^{***}$ ）及執行、設計 DDL 教學活動的能力（ $\rho = .28$ ； $p\text{-value} = .047^*$ ）也越佳，同時，他們也更傾向在日後的實際教學中應用 DDL 教學法（ $\rho = .36$ ； $p\text{-value} = .01^*$ ）。由此可見，推廣語料庫應用方法是相當重要的。另一點值得注意的是，教師在課前對語料庫的先備知識，隨著教育程度提升而增加（ $\rho = .37$ ； $p\text{-value} = .008^{**}$ ），卻與教學經驗多寡無關（ $\rho = .23$ ； $p\text{-value} = .11$ ），這或許可以說明教師接觸語料庫的機會多半是在就學期間，而在職業生涯中較缺乏。這點也與本研究在規劃前的假設一致。因此，本課程在設計之初即是針對在職教師為授課對象。

(5) 「課後對操作語料庫的能力」與其他面向的相關性

「課後對操作語料庫的能力」與「執行、設計 DDL 教學活動的能力」這兩個面向，分別反映了教師在操作語料庫的能力及將語料庫方法應用於教學活動的能力。透過分析的結果，可以看到這兩個能力具有高度相關性 ($\rho = .52$; $p\text{-value} < .001^{***}$)，對操作語料庫較為嫻熟的教師，在設計與執行 DDL 課程上的能力也越佳，也更傾向在日後的課程中實際使用 DDL 教學法 ($\rho = .45$; $p\text{-value} < .01^{**}$)；傾向給本課程更高的評價 ($\rho = .29$; $p\text{-value} = .04^*$)。此外，操作語料庫能力越好的教師，對語料庫的態度也越不負面 ($\rho = -.72$; $p\text{-value} < .001^{***}$)。從相關性分析結果得知擁有良好操作語料庫能力的教師，在設計、執行 DDL 教學活動的能力也隨之增加，同時也更傾向在未來課室應用 DDL 教學法。可見擁有語料庫的操作能力是將語料庫應用於教學的關鍵。

(6) 「對語料庫的負面態度」與其他面向的相關性

前已提及，教師對語料庫的負面態度與較差的語料庫操作能力具有高度相關性 ($\rho = -.72$; $p\text{-value} < .001^{***}$)，教學經驗越長的教師也對語料庫越感到負面 ($\rho = .40$; $p\text{-value} = .04^*$)。此外，對語料庫越感負面的教師，在執行、設計 DDL 教學活動的能力也越差 ($\rho = -.57$; $p\text{-value} = .004^{**}$)。值得一提的是，教師對語料庫的負面態度與他們是否具備相關先備知識無關 ($\rho = -.29$; $p\text{-value} = .18$)，也與日後在課程中使用 DDL 教學法的可能性無關 ($\rho = -.16$; $p\text{-value} = .45$)，推測參加工作坊的教師並非均想應用 DDL 於教學，可能單純為了學習新知。

(7) 「執行、設計 DDL 教學活動的能力」與其他面向的相關性

具有相關先備知識的教師 ($\rho = .28$; $p\text{-value} = .047^*$)、對語料庫操作能力較高的教師 ($\rho = .52$; $p\text{-value} < .001^{***}$)，在課後對執行、設計 DDL 教學活動的能力亦越佳。而執行、設計 DDL 教學活動的能力越高的教師，越傾向給予課程較高的評價 ($\rho = .47$; $p\text{-value} < .001^{***}$)。相反的，對語料庫越感到負面的教師，執行、設計 DDL 教學活動的能力越差 ($\rho = -.57$; $p\text{-value} = .004^{**}$)。

(8)「未來在課程中應用 DDL 教學法的可能性」與其他面向的相關性

課前擁有語料庫相關先備知識的教師 ($\rho = .36$; $p\text{-value} = .01^*$) 及對語料庫操作能力佳的教師 ($\rho = .45$; $p\text{-value} = .001^{**}$)，越傾向在日後的課室中執行 DDL 教學活動。這說明了若希望 DDL 教學法廣泛地使用於華語教學的課室中，需要加強推廣語料庫方法，並提升教師操作語料庫的能力。

(9)「對整體課程的評價」與其他面向的相關性

參與課程動機強烈的教師給予本課程較高的評價 ($\rho = .43$; $p\text{-value} = .002^{**}$)。此外，教師在接受培訓後的成效也反映在對課程的評價上，對操作語料庫的能力較高 ($\rho = .29$; $p\text{-value} = .04^*$) 及執行、設計 DDL 教學活動的能力較高 ($\rho = .47$; $p\text{-value} < .001^{***}$) 的教師，傾向給予課程較高的評價。這或許可以說明本課程教授的內容符合教師的期望，即在培訓過程中建立了良好操作與應用語料庫能力的教師，也對培訓感到滿意。

透過分析 9 個面向之間的相關性，讀者可以得知教師在各個面向的特質與應用 DDL 教學法的關係。相關性結果說明了教師對於語料庫的操作能力為關鍵因素，善於操作語料庫的教師，對 DDL 教學活動設計、執行的能力越佳，且更樂於在課室中嘗試進行 DDL 教學活動。實際上，前人研究也顯示語料庫方法進入語言教學現場其中一個需要克服的困難就是降低教師對使用語料庫的擔憂，而這份擔憂源於教師對資訊能力的信心不足 (Lam 2000; Mauranen 2004; Römer 2009; Lin 2019)，這點與本研究分析的結果一致，即，操作語料庫能力越佳的教師，越不易對語料庫抱持負面的態度，而能力可以透過持續開設相關課程來培養。

4.4 「語料驅動教學工作坊」參與教師的文字反饋

除了問卷的相關性分析外，本研究進一步分析教師的文字反饋。首先，為了瞭解教師的需求，本研究設計了一個問題「您最喜歡哪一週的課程？」，因為 4 週的課程有不同的主題，問卷結果顯示第 2 週的課程最受歡迎，如下圖²⁰，而選擇第 3 週的人最少。

²⁰ 問卷表單連結：<https://docs.google.com/forms/d/1pZP15ZnkVhm18DhIW8R0yHV9GYtjbgatg8S-U8EGU/edit#responses>

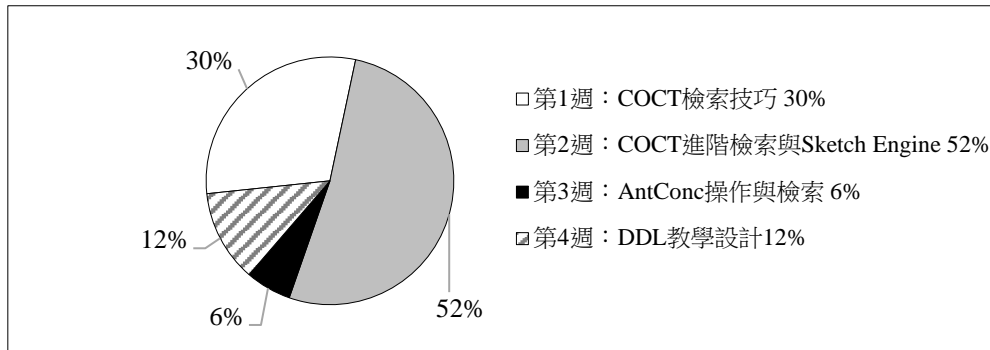


圖 11：「您最喜歡哪一週的課程？」比例圖

選擇第 2 週的原因多數關鍵詞為「實用」，一位老師這樣描述「很實用，以前只會檢索生詞，不知道語法點怎麼做。現在知道了，備課如虎添翼！」；另一位老師的描述是「*sketch engine* 對於自己學習外語也有所幫助。運用 COCT 進階檢索可以節省許多造例句的時間。若語料用詞不合適，稍加修改即可。」可見只要對教學實用、有幫助，大家仍願意投入時間來學習進階的檢索技巧（如 CQP 工具），原來筆者擔心進階技巧對華語教師而言太難，似乎是多慮了。

至於選擇第 3 週的人最少，就筆者與教師的私下互動推測是多數教師認為自己還不需要用到 AntConc 這樣的分析軟體，另外也看到有位老師的文字回饋是「我覺得第三週的課程能減少，對實務教學感覺較用不上。能多講第一、二週的課。因為寫式子真的不容易。」，可見本研究因為課程時間有限，刪除了 BootCat 這個主題課程，對於教師來說，不是太大的損失。從許多文字的回饋看到，在職教師來參加這樣的培訓最重要的考量是實用性，加上這是第一次針對在職教師的培訓，教師預設自己接收的應該是最基礎的訓練，自建語料庫與下載新的語料庫工具，或許不在他們預期內。

以下則節錄數則教師對於 4 週工作坊，分享的感受或建議。

.....培訓之前，原來對平衡語料庫只有非常粗淺的認識，通過四週課程，現在進一步知道怎麼更有效率地檢索。希望.....將來實際應用在課堂上，甚至能在規劃高級班的時候，能融入 DDL 教學理念，提供學生更有用的課程。（教師 F）

我覺得**老師講解得非常詳細而且語速適中**，也給我們時間實際操作，這幾堂課受益良多，也讓我進一步思考怎麼應用語料庫並融入我的教學中。謝謝老師豐富的課程內容，也謝謝助教即時的回覆以及批改作業。(教師 G)

.....最後的驗收聽到要做教案，反而覺得壓力有點大，因為平常已經要備課了，再加上我個人主要是將語料庫做為輔助備課的材料使用，不太會作為「教學」內容，所以雖然了解老師設計這個作業的目的，但還是會覺得寧願直接做考試檢測，可能會更快一點.....(教師 H)

從上述最後一則文字可以知道為何不是每位教師都能繳交作業，雖然筆者認為語料庫素養非一朝一夕可以達成，課後實作練習是很重要的條件，不動手做，很難習得。然而，多數華語教師幾乎每天有課，平日要備課，仍願意連續 4 個週末來參加這個工作坊，無庸置疑的，參加的教師都有很強的學習動機，如何給予有意思的作業或評量或許是未來辦理這類工作坊可以思考加強之處。雖然，教師對於課程多給予正面回饋，也有幾則負面評論值得注意：

.....在教學課堂上讓學生分析語料感覺**不符合經濟效益**。語料中的**超綱詞彙也太多**，基本上教師還是要整個大改，學生才能「分析語料」，但是這也就不是真實語料了。(教師 J)

.....在做第四週期末評量作業時，**我發現語料庫找到的語句絕大部分都不合用**，除了語料中有太多超出學生識字範圍的生詞以外，語料庫中的句子大多不包含學生前幾課剛學過的生詞，若直接引用語料庫的句子，就無法幫助學生複習前面學過的生詞，所以我給學生練習的句子都是我自己重新編寫的。(教師 K)

花費太多心力篩選句子，似乎是普遍語言教師的心聲，這個問題本研究期望未來能透過群體合作建置教學語料以及建置適合教學、學習的語料庫來排除。

5. 結論

本研究乃基於華語教學課堂鮮少人實踐 DDL 的教學法開始，探討欲實踐 DDL，教師需要具備的能力為何，從而規劃課程，為在職的華語教師開設

工作坊。工作坊透過講解、教授語料庫相關知識、示範操作語料庫、分析語料，並帶領教師練習、解答教師疑惑、放置課程錄影於 YouTube 上，期待能培養在職教師應用語料庫的教學能力，進而在課堂上實踐。

經過課後問卷的量化與質化分析與討論，本研究從相關性檢定分析得知具備語料庫的先備知識以及擁有良好使用技巧的教師，越有可能在未來課堂上運用語料庫來教學，這也鼓勵研究者可以持續進行這樣的培訓工作。唯有教師實際應用 DDL 授課後，教師或研究者才能得知學習者的感受與學習成效，因此，本研究只是一個開始，期待看到開花結果的一天。

引用文獻

- Abdel Latif, Muhammad M. M. 2021. Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL* 33.1: 34-48.
- Anthony, Laurence. 2016. Computer Software: AntFileConverter (Version 1.2.0). Tokyo, Japan: Waseda University. Accessed online, <http://www.laurenceanthony.net/>
- Anthony, Laurence. 2020. Computer Software: AntConc (Version 3.5.9). Tokyo, Japan: Waseda University. Accessed online, <http://www.antlab.sci.waseda.ac.jp/>
- Aston, Guy. 2001. Learning with corpora: An overview. *Learning with Corpora*, ed. by Guy Aston, 6-45. Houston, TX: Athelstan.
- Babych, Svitlana. 2015. Textual cohesion patterns for developing reading skills: A corpus-based multilingual learning environment. *Multiple Affordances of Language Corpora for Data-driven Learning*, eds. by Agnieszka Leńko-Szymańska, and Alex Boulton, 155-176. Amsterdam: John Benjamins.
- Boulton, Alex. 2010. Data-driven learning: Taking the computer out of the equation. *Language Learning* 60.3: 534-572.
- Boulton, Alex. 2015. Applying data-driven learning to the web. *Multiple Affordances of Language Corpora for Data-driven Learning*, eds. by Agnieszka Leńko-Szymańska, and Alex Boulton, 267-296. Amsterdam: John Benjamins.

- Bunting, John David. 2013. *An Investigation of Language Teachers' Explorations of the Use of Corpus Tools in the English for Academic Purposes Class*. Atlanta: Georgia State University Ph. D. dissertation. Accessed online, March 5, 2017. http://scholarworks.gsu.edu/alesl_diss/26
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.4: 711-733.
- Callies, Marcus. 2016. Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. *Angewandte Linguistik in Schule und Hochschule*, eds. by Rolf Kreyer, Steffen Schaub, and Ann Gulderning, 391-415. Frankfurt, Germany: Peter Lang.
- Callies, Marcus. 2019. Integrating corpus literacy into language teacher education. *Learner Corpora and Language Teaching*, eds. by Sandra Götz, and Joybrato Mukherjee, 245-263. Amsterdam: John Benjamins.
- Chen, Meilin, John Flowerdew, and Laurence Anthony. 2019. Introducing in-service English language teachers to data-driven learning for academic writing. *System*, 87. <https://doi.org/10.1016/j.system.2019.102148>
- Chujo, Kiyomi, Kathryn Oghigian, and Shiro Akasegawa. 2015. A corpus and grammatical browsing system for remedial EFL learners. *Multiple Affordances of Language Corpora for Data-driven Learning*, eds. by Agnieszka Leńko-Szymańska, and Alex Boulton, 109-128. Amsterdam: John Benjamins.
- Corino, Elisa, and Cristina Onesti. 2019. Data-driven learning: A scaffolding methodology for CLIL and LSP teaching and learning. *Frontiers in Education* 4: 7. <https://doi.org/10.3389/educ.2019.00007>
- Crosthwaite, Peter, Luciana Luciana, and David Wijaya. 2021. Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning*, 1-29. <https://doi.org/10.1080/09588221.2021.1995001>
- Dancey, Christine, and John Reidy. 2004. *Statistics without Maths for Psychology: Using SPSS for Windows*. London: Prentice Hall.
- Doughty, Catherine, and Jessica Williams. (eds.) 1998. *Focus on Form in*

- Classroom Second Language Acquisition*. Cambridge: CUP.
- Farr, Fiona. 2008. Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness* 17.1: 25-43.
- Gaskell, Delian, and Thomas Cobb. 2004. Can learners use concordance feedback for writing errors? *System* 32.3: 301-319. DOI: 10.1016/j.system.2004.04.001
- Halliday, Michael A. K. 2003. *On Language and Linguistics*. New York: Continuum.
- Heather, Julian, and Marie Helt. 2012. Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education* 20.4: 415-440.
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž, and David Tugwell. 2004. The sketch engine. Paper presented at *XI EURALEX International Congress*, July 2004. France: Lorient.
- Kucera, Henry, and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. Providence: Brown University Press.
- Lam, Yvonne. 2000. Technophilia vs technophobia: A preliminary look at why second-language teachers do or do not use technology in their classrooms. *Canadian Modern Language Review* 56.3: 390-420.
- Leńko-Szymańska, Agnieszka. 2014. Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL* 26.2: 260-278.
- Leńko-Szymańska, Agnieszka. 2015. A teacher-training course on the use of corpora in language education: Perspectives of the students. *Insights into Technology Enhanced Language Pedagogy*, eds. by Anna Turula, Beata Mikołajewska, and Danuta Stanulewicz, 129-144. Frankfurt am Main: Peter Lang.
- Leńko-Szymańska, Agnieszka. 2017. Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology* 21.3: 217-241.
- Leńko-Szymańska, Agnieszka, and Alex Boulton. 2015. Introduction: Data-driven learning in language pedagogy. *Multiple Affordances of*

- Language Corpora for Data-driven Learning*, eds. by Agnieszka Leńko-Szymańska, and Alex Boulton, 1-14. Amsterdam: John Benjamins.
- Lin, Ming Huei. 2019. Becoming a DDL teacher in English grammar classes: A pilot study. *Journal of Language Learning and Teaching* 9.1: 70-82.
- Lin, Ming Huei, and Jia-Ying Lee. 2015. Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal* 69.3: 264-274.
- Liu, Dilin, and Ping Jiang. 2009. Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal* 93.1: 61-78.
- Mauranen, Anna. 2004. Speech corpora in the classroom. *Corpora and Language Learners*, eds. by Guy Aston, Silvia Bernardini, and Dominic Stewart, 197-213. Amsterdam: John Benjamins.
- Mizumoto, Atsushi, and Kiyomi Chujo. 2016. Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System* 61: 55-64.
- Mukherjee, Joybrato. 2006. Corpus linguistics and language pedagogy: The state of the art—and beyond. *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, eds. by Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, 5-24. Frankfurt am Main: Peter Lang.
- Norris, John M., and Lourdes Ortega. 2000. Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50: 417-528. DOI: 10.1111/0023-8333.00136
- O’Keeffe, Anna, and Fiona Farr. 2003. Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly* 37.3: 389-418.
- O’Sullivan, Ide. 2007. Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL* 19.3: 269-286.
- Römer, Ute. 2009. Corpus research and practice: What help do teachers need and what can we offer? *Corpora and Language Teaching*, ed. by Karin Aijmer, 83-98. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.33.09rom>
- Smith, Simon. 2011. Corpus-based tasks for learning Chinese: A data-driven

- approach. *The Asian Conference on Technology in the Classroom Official Conference Proceedings 2011*, 48-59.
- Smith, Simon. 2020. DIY Corpora for Accounting & Finance Vocabulary Learning. *English for Specific Purposes* 57: 1-12.
- Tao, Hongyin. 2005. The gap between natural speech and spoken Chinese teaching material: Discourse perspectives on Chinese pedagogy. *Journal of the Chinese Language Teachers' Association* 40.2: 1-24.
- Thomas, James. 2008. Impatience is a virtue: Students and teachers interact with corpus data - now. *Proceedings of the 8th Teaching and Language Corpora Conference*, ed. by Ana Frankenberg-Garcia, 463-469. Lisbon: ISLA.
- Tomasello, Michael. 2003. *Constructing a Language: A Usage Based Theory of Language Acquisition*. Harvard University Press.
- Wang, Lixun. 2001. Exploring parallel concordancing in English and Chinese. *Language Learning and Technology* 5.3: 174-184.
- Xiao, Richard, and Tony McEnery. 2006. Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics* 27.1: 103-129.
- Yeh, Meng, and Xian Zhang. 2018. Corpus-based instruction: Teaching discourse-linking *jiu* (就) in storytelling. *Chinese as a Second Language* 53.1: 1-23.
- Yoon, Hyunsook, and Jung Won Jo. 2014. Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language, Learning and Technology* 18.1: 96-117.
- 王炳勻、許展嘉、龍水水、丁曉穎。2020。〈語料驅動學習融入華語課堂之教學設計〉，《華語文教學研究》，第17卷第3期，103-137。[Wang, Ben Pin-Yun, Chan-Chia Hsu, Shuishui Long, and Xiaoying Liles. 2020. Designing data-driven learning activities for the Chinese as a Second Language classroom. *Journal of Chinese Language Teaching* 17.3: 103-137.]
- 吳鑑城、陳浩然、張俊盛。2017。〈網路語料庫介紹與應用〉，《語料庫與華語教學》，陳浩然（主編），49-70。臺北：高等教育出版社。[Wu, Jian-Cheng, Hao-Jan Chen, and Jung-Sheng Chang. 2017. Introduction and application

- of web corpus. *Corpus and Teaching Chinese as a Second Language*, ed. by Hao-Jan Chen, 49-70. Taipei: Taiwan Higher Education Press Co.]
- 吳鑑城、白明弘、林慶隆。2019。〈臺灣華語文語料庫在華語文教育的應用〉，《華語文教學研究》，第16卷第3期，29-55。[Wu, Jian-Cheng, Ming-Hong Bai, and Ching-Lung Lin. 2019. Applying the corpus of contemporary Taiwanese mandarin in teaching Chinese as a second language. *Journal of Chinese Language Teaching* 16.3: 29-55.]
- 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》，第19卷第2期，53-72。[Chang, Li-ping. 2014. Salient linguistic features of Chinese learners with different L1s: A corpus-based study. *International Journal of Computational Linguistics and Chinese Language Processing* 19.2: 53-72.]
- 張莉萍。2017。〈TOCFL學習者語料庫的偏誤標記〉，《語料庫與華語教學》，陳浩然(主編)，159-196。臺北：高等教育出版社。[Chang, Li-ping. 2017. Error annotation for the TOCFL Learner Corpus. *Corpus and Teaching Chinese as a Second Language*, ed. by Hao-Jan Chen, 159-196. Taipei: Taiwan Higher Education Press Co.]
- 張莉萍。2022。〈從基於用法的理論探討中高級華語教學語法點〉，《華語文教學研究》，第19卷第2期，33-64。[Chang, Li-ping. 2022. A study of pedagogical grammar points for advanced L2 Chinese learners: A usage-based approach. *Journal of Chinese Language Teaching* 19.2: 33-64.]
- 蔡美智、黃居仁、陳克健。1999。〈由近義詞辨義標準看語意、句法之互動〉，《中國境內語言暨語言學·第五輯·語言中的互動》，殷允美、楊懿麗、詹惠珍(主編)，439-459。臺北：中央研究院語言學研究所。[Tsai, Mei-chih, Chu-Ren Huang, and Keh-jiann Chen. 1999. From near-synonyms to the interaction between Syntax and Semantics. *Chinese Languages and Linguistics V: Interactions in Language*, eds. by Yuen-mei Yin, I-li Yang, and Hui-chen Chan, 439-459. Taipei: Institute of Linguistics, Academia Sinica.]

[審查：2022.8.29 修改：2022.9.29 接受：2022.11.10]

華語文教學研究

張莉萍

Li-Ping CHANG

10617 臺北市大安區羅斯福路四段 1 號 國立臺灣大學華語教學碩士學位學程

Graduate Program of Teaching Chinese as a Second Language

National Taiwan University

No.1, Sec. 4, Roosevelt Rd., Da'an Dist., Taipei City 10617, Taiwan

lchang@ntu.edu.tw

The Preliminary Study of Corpus Literacy Training for In-service Chinese Language Teachers

Li-Ping CHANG

**Graduate Program of Teaching Chinese as a Second Language
National Taiwan University**

Abstract

The data-driven learning (DDL) approach has been applied in learning English for over a decade. However, corpus-based pedagogy has not been widely applied to Chinese as a second language (CSL) teaching, the main reason may be that CSL teachers are relatively lacking corpus literacy, such as familiarity with corpus tools, corpus consulting techniques, corpus-based pedagogical activities, etc. To cultivate corpus literacy and promote DDL application in future classrooms, 73 CSL teachers participated in a 4-week, 3-hour/week DDL training workshop. The workshop was divided into four topics: the introduction to Chinese corpora, corpus consulting techniques, hands-on training for corpus analysis tools (e.g. AntConc), and pedagogical designs or activities for guiding learners in inducting or deducting rules from corpus data. After the workshop, a post-workshop questionnaire consisting of 39 Likert scale questions and four open-ended questions was distributed to participants. The 36 quantitative questions were divided into seven variables that reflect the traits of the participants. These variables were self-evaluations of their (1) motivations for participating in the workshop, (2) prior knowledge of corpora, (3) skills toward using corpora after the workshop, (4) negative perceptions of adopting corpora in teaching, (5) skills to implement and design DDL activities, (6) willingness to adopt corpus-based methods in future teaching, and (7) evaluation of the workshop. The results of Spearman's rank-order correlation test indicated that regardless of the level of motivation of the teachers to attend the workshop, they could build corpus-related abilities through lectures and practice; and teachers who had more prior knowledge of corpora and with greater confidence in their

skills to operate corpora, the more likely to perform DDL teaching activities in their future classrooms. These results suggest that cultivating teachers' corpus literacy is effective, and continually offering training courses can facilitate implementing the corpus-based approach in the future.

Keywords: corpus, corpus consulting, correlation test, data-driven learning, questionnaire