

跨域趨勢——

臺灣華語文語料庫與能力基準整合應用系統



華語文八年計畫「建置應用語料庫及標準體系」

跨域趨勢——
臺灣華語文語料庫與能力基準整合
應用系統



國家教育研究院

National Academy for Educational Research

www.naer.edu.tw

序

本書性質屬臺灣華語文應用語料庫與能力基準整合應用系統的使用手冊，內容涵蓋語料庫索引典，華語文能力基準的漢字、詞語、語法點分級標準檢索系統，語義場關聯詞查詢系統，例句編輯輔助系統，作文錯別字自動批改系統及在辭典編輯應用、近義詞的查詢與辨析、和教材與課堂教學的綜合應用。

2013 年時，教育部為提升臺灣在華語文教育的品質及效能，因此於華語文教育八年計畫中將「建置應用語料庫及標準體系」列為建構永續發展基礎計畫，並由國家教育研究院執行。內容涵蓋建構華語文能力指標，建置華語文應用語料庫，應用語料庫以建置華語文漢字、詞語及語法點分級標準，及辦理相關的應用推廣工作。

至 2020 年，在建置應用語料庫，已完成建置華語文書面語、口語、中介語、華英雙語語料庫及通用詞頻表，定名為「臺灣華語文語料庫」(英文名稱為 Corpus of Contemporary Taiwanese Mandarin, 簡稱 COCT)。在建置標準體系，完成華語文三等七級能力指標與漢字、詞語、語法點分級標準，定名為「臺灣華語文能力基準」。而且，完成第 1~5 級詞語的情境分類和第 1~3 級詞語例句等資訊。另外，在建置臺灣華語文語料庫與能力基準過程，同時發展多個應用系統，並建置「臺灣華語文語料庫與標準體系整合應用系統」入口網站(網址：<https://coct.naer.edu.tw>)，提供使用者方便應用，且是科技整合應用於華語文教學的新發展。本書為計畫部份成果的 5 本專書之一，其中已出版的有《遣辭用「據」—臺灣華語文能力第一套標準》及《臺灣華語文語料庫—華語文教與學的必備工具》。

該計畫已辦理多場應用工作坊，詳細介紹如何應用，並與全球華語文教育專案辦公室合作，將課程製成影片，網址為：簡介及基礎應用 <https://reurl.cc/20Ela4>、<https://reurl.cc/lV7ap6>；進階應用及實作 <https://reurl.cc/4RWbYj>、<https://reurl.cc/O1jRQ3>。而且，國立空中大學數位華語文中心亦製作【華師增能篇】華語教師不可不知的語料庫系統 A Corpus That Chinese Teachers Must Know!! 教學影片(網址為 <https://youtu.be/yojmaAZ897A>)，簡介系統應用。本整合應用系統使用者來自世界各地，平均每月超過 1 萬 3 千人次使用，點擊次數超過 23 萬次，而且今年使用數超過去年的 2 倍以上。

本書能順利完成，感謝「建置應用語料庫及標準體系」計畫各主持人、專家學者、華語文教師、本院研究人員及計畫參與人員對計畫的努力及貢獻。編撰小組同仁的編撰校對。最後，本書雖然力求完善，一定有可以更圓滿的地方，敬請專家指正。

林慶隆 謹識

國家教育研究院語文教育及編譯研究中心研究員兼任中心主任

2020年12月

目錄

第一章、前言.....	1
第二章、語料庫索引典.....	5
第一節、索引典介面.....	5
第二節、萬用字元.....	10
第三節、CQP 檢索語法使用.....	13
第四節、進階功能.....	17
第五節、華英雙語索引典.....	23
第三章、華語文能力基準整合應用系統.....	29
第一節、分級標準檢索系統.....	29
第二節、語義場關聯詞查詢系統.....	34
第三節、例句編輯輔助系統.....	38
第四節、作文錯別字自動批改系統.....	41
第四章、綜合應用.....	43
第一節、辭典編輯應用.....	43
第二節、近義詞的查詢與辨析.....	49
第三節、教材與課堂教學.....	52
第五章、結語.....	58
誌謝.....	59
參考文獻.....	60
附件.....	61

圖目錄

圖 1-1 臺灣華語文應用語料庫與臺灣華語文能力基準整合應用系統入口網頁..	4
圖 2-1 國教院索引典系統 (https://coct.naer.edu.tw/cqpweb/)	5
圖 2-2 COCT 書面語語料庫 2019 的介面	6
圖 2-3 Corpus queries (語料庫查詢)	6
圖 2-4 COCT 書面語語料庫 2019 的 Standard Query (標準查詢) 介面	7
圖 2-5 COCT 書面語語料庫 2019 Restricted Query (限制查詢) 的介面	7
圖 2-6 COCT 書面語語料庫 2019 Word lookup (單詞查詢) 的介面	8
圖 2-7 書面語語料庫 2019 Frequency lists (詞頻列表)	9
圖 2-8 書面語語料庫 2019 Frequency lists (詞頻列表) 的 Frequency list option settings	9
圖 2-9 書面語語料庫 Keywords 關鍵詞.....	9
圖 2-10 2019 書面語語料庫檢索[word="快樂"].....	13
圖 2-11 2019 書面語語料庫檢索[word="台北"][word="捷運"].....	13
圖 2-12 2019 書面語語料庫檢索[pos="Nf"] [word="電話"]檢索結果	14
圖 2-13 2019 書面語語料庫檢索[pos!="Nf"] [word="電話"]檢索結果	14
圖 2-14 2019 書面語語料庫檢索[word="頭"] []檢索結果	14
圖 2-15 2019 書面語語料庫檢索[word="頭"&pos="Nf"] []檢索結果	15
圖 2-16 2019 書面語語料庫檢索[word="頭"&pos="Nf"] [pos="Na"]檢索結果 ...	15
圖 2-17 離合詞「見面」部分檢索結果.....	15
圖 2-18 2019 書面語語料庫檢索[word="見"][] {1,5} [word="面"]檢索結果.....	16
圖 2-19 2019 書面語語料庫檢索[word="見"] [word!="， 。"]* [word="面"]檢索結果	16
圖 2-20 書面語語料庫檢索結果介面.....	17
圖 2-21 查詢結果介面中的「再查詢」	17
圖 2-22 2019 書面語語料庫檢索[word="頭"&pos="Nf"] [pos="Na"]檢索結果	18
圖 2-23 進階搜尋 Thin 介面.....	18
圖 2-24 進階搜尋 Thin 搜尋結果.....	18
圖 2-25 進階搜尋 Frequency breakdown 搜尋結果.....	19
圖 2-26 進階搜尋 Distribution 搜尋結果.....	20
圖 2-27 進階搜尋 Sort 搜尋結果.....	21
圖 2-28 進階搜尋 Sort 介面設定.....	21
圖 2-29 進階搜尋 Collocation 介面	22
圖 2-30 進階搜尋 Collocation 搜尋結果	22
圖 2-31 華英雙語索引典系統	23
圖 2-32 華英雙語索引典以「老師」一詞為例查詢結果	24
圖 2-33 華英雙語索引典以「老師」一詞為例點選圖 2-32 自動抽取搭配詞中的	

「位」結果.....	24
圖 2-34 以英文單字 teacher 為例.....	25
圖 2-35 以「老師」一詞為例結果.....	25
圖 2-36 以「老師」一詞為例觀察自動對應翻譯.....	26
圖 2-37 以「老師」一詞為例觀察自動抽取搭配詞.....	27
圖 3-1 華語分級標準漢字、詞語及語法點檢索系統.....	29
圖 3-2 漢字分級標準檢索系統「果」字查詢結果介面.....	30
圖 3-3 國字標準字體筆順學習網，以「果」字為例查詢結果介面.....	30
圖 3-4 詞語分級標準檢索系統「果」查詢結果介面.....	31
圖 3-5 詞語分級標準檢索系統「蘋果」查詢結果介面.....	31
圖 3-6 詞語分級標準檢索系統「一心一意」查詢結果介面.....	32
圖 3-7 語法點分級標準檢索系統「在」搜尋結果介面.....	32
圖 3-8 語義場關聯詞查詢系統.....	34
圖 3-9 語義場關聯詞查詢系統，正關聯詞「蘋果」為例查詢結果介面.....	35
圖 3-10 語義場關聯詞查詢系統，鍵入正關聯詞「蘋果」和負關聯詞「賈伯 斯」的查詢結果介面.....	36
圖 3-11 截取圖 3-9 和圖 3-10 之輸出語義場關聯詞比較圖.....	36
圖 3-12 例句編輯輔助系統介面.....	38
圖 3-13 例句編輯輔助系統以「突然下了一場暴雨，我的衣服都濕了。」為例	39
圖 3-14 例句編輯輔助系統輸入《當代中文課程 3》課本短文分析結果介面 ..	40
圖 3-15 作文錯別字自動批改系統介面.....	41
圖 3-16 作文錯別字自動批改系統文句修正介面.....	41
圖 4-1 COCT 書面語語料庫 2019 與常用度比較搭配使用結果.....	43
圖 4-2 COCT 口語語料庫 2019 與常用度比較搭配使用結果.....	44
圖 4-3 COCT 書面語語料庫 2019 與常用度比較搭配使用結果.....	44
圖 4-4 COCT 口語語料庫 2019 與常用度比較搭配使用結果.....	45
圖 4-5 COCT 書面語語料庫 2019 與使用頻率搭配使用結果.....	45
圖 4-6 COCT 書面語語料庫 2019 與使用頻率搭配使用結果.....	46
圖 4-7 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果	47
圖 4-8 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果	47
圖 4-9 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果	48
圖 4-10 國教院語義場關聯詞查詢系統以「感覺」一詞為例.....	49
圖 4-11 「感覺」、「感受」和「覺得」搭配詞.....	50
圖 4-12 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果	

.....	52
圖 4-13 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果	53
.....	53
圖 4-14 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果	53
.....	53
圖 4-15 例句編輯輔助系統	54
圖 4-16 例句編輯輔助系統以重修本「高興」釋義為例	54
圖 4-17 語法點分級標準檢索系統以「不但……，反而……」為例	55
圖 4-18 COCT 書面語語料庫 2019 以「不但……，反而……」為例	56
圖 4-19 例句編輯輔助系統以確認用詞之等級	56

表目錄

表 1 「感覺」、「感受」和「覺得」兩兩比對	50
表 2 「感覺」、「感受」和「覺得」比對	51

第一章、前言

臺灣在華語文教學雖然頗具口碑，惟缺乏標準體系、大型語料庫及應用系統，影響教學、教材設計及測驗評量的效能，說明如下：

一、華語文教學缺乏標準體系

華語文標準體系包括能力指標及漢字、詞語及語法點的分級標準。分析臺灣四部最重要的華語文教材中，共同收錄的生詞僅占 9.8%；華語地區四套最常用的教材，共同收錄的語法點也僅為 13%；多份華語文字表，共同收錄的漢字差異亦很大。顯示因為缺乏華語文標準體系，教材編撰者編寫教材時，多依主觀經驗選用漢字、詞語及語法點。作為教學及測驗評量參考使用的教材差異如此大，教學及測驗評量亦會有類似情形。

就如中小學教學需要課程綱要，華語文教學亦需要標準體系做為課程指引，以提供學習者在各學習階段應知和應能為何，學習才能由淺入深、循序漸進有系統的進行，確保並提升學生的學習成效。

二、華語文標準體系建置缺乏大型華語文語料庫及應用系統

華語文標準體系建置若能依據語言實際使用情境，則更能符合學習的需要。語料庫建置大量實際使用的語言資料，是最好的選擇。語料庫經由應用系統分析，統計漢字、詞語、語法點及搭配詞的真實使用語境和頻率，可以應用在建置華語文標準體系；也可以提供詞義分析及例句編寫等應用在華語文教學。然而，2013 年當時國內缺乏大型華語文語料庫，自然也無應用系統。

然而，更重要的是，即使應用語料庫完成華語文漢字、詞語及語法點分級標準，而沒有線上查詢系統，及未與語料庫及應用系統整合，都將影響其在華語文教學、教材設計及測驗評量的有效應用。

上述問題的解決方式如下：

一、建置華語文語料庫及臺灣華語文能力基準

為了建置華語文語料庫及研發具實證基礎的華語文能力指標和漢字、詞語及語法點分級標準，國家教育研究院（以下稱本院）邀請學者專家成立諮詢團隊並組成語料庫組、能力指標組、漢字組、詞語組及語法組等 5 個研究團隊進行研發建置；除了研究團隊人員研究外，也召開超過百場各類學者專家諮詢會議，其中包括全面徵詢華語教學系所、華語文教學中心及華語文學術社團。參與研發諮詢專家之多及廣，在華語文教學界可謂空前；歷經 6 年完成臺灣華語文能力第一套標準體系及建置最新且最大的華語文語料庫及多套應用系統。

在語料庫建置，本院已建置四類型語料庫，包括書面語語料庫約 4 億 1,000 萬字（含新聞語料共 31 億 7,100 萬字），口語語料庫約 3,440 萬字，雙語語料庫約 1,160 萬字，中介語語料庫約 147 萬字。並配合標準體系建置開發多套應用系統。

在標準體系建置，本院至 2020 年 8 月 31 日，已完成聽、說、讀、寫、譯三等七級能力指標，完成三等七級 3,100 個漢字字表及 14,470 個詞的詞語表、73 個類詞綴；並完成第 1 級至第 5 級 496 個語法點及例句編寫。在與國際指標接軌，完成與美國外語教學委員會（American Council on the Teaching of Foreign Language, ACTFL）及國家華語測驗推動工作委員會之「華語文能力測驗」（Test of Chinese as a Foreign Language, TOCFL）（依據歐洲共同語文參考架構指標，Common European Framework of Reference for Languages, CEFR）能力指標的等級對應。

華語文標準體系是依據教育部華語文八年計畫所研發，因此正式名稱參考歐洲漢語能力基準（European Benchmarks for the Chinese Language, EBCL），訂為臺灣華語文能力基準，英文為 Taiwan Benchmarks for the Chinese Language（簡稱 TBCL）。從 TBCL 與 TOCFL 對應圖，發現在聽力、閱讀及寫作，TBCL 的第 1 級及第 2 級都在 TOCFL 入門級（A1）以下；在口說，TBCL 的第 1 級在 TOCFL 入門級（A1）以下；顯示 TBCL 確實達成華語文專家學者深切期待的建構具華語文特性之能力基準的目標。

二、建置華語文語料庫與標準體系整合應用系統

為了提供使用者更容易的應用，本院將華語文分級標準分別建置「漢字分級標準檢索系統」、「詞語分級標準檢索系統」、「語法點分級標準檢索系統」等線上查詢系統。而且，將研發標準體系使用的「國教院索引典系統」、「華英雙語索引典系統」、「華語中介語索引典系統」、「語義場關聯詞查詢系統」、「例句編輯輔助系統」、「作文錯別字自動批改系統」等系統開放使用。

然而，各系統各有自己的網址，使用者不易應用，因此本院將其整合建置「華語文語料庫與標準體系整合應用系統」（網址：<https://coct.naer.edu.tw>），提升使用者應用的效能。

華語文語料庫與標準體系整合應用系統平台除了單一入口服務，並具備下述多面向服務優點，提供華語文教學、教材設計、測驗評量及研究應用。

一、具有華語文語料類別最完整、內容最新、數量最多的語料庫

本平台建置有華語書面語、口語、中介語及華英雙語等四類語料庫，為臺灣華語文語料類別最完整語料庫；這四類語料庫建置的語料都是最新，且數量最多。

其中，書面語語料庫內容涵蓋 2008 年以後，哲學及宗教類、科學類、應用科學類、社會科學類、史地類、語言文學類、藝術類、商業及金融類、休閒類等多元文章，約 4 億 1000 萬字（含新聞語料共 31 億 7,100 萬字）。口語語料庫內容則涵蓋 2008 年以後之法政軍事、財經、時事、科學、生活時尚、文教藝術等多元電視節目的語音及文字，約 3,440 萬字。由於，語料最新且數量最多，可使詞語及語法點的使用頻率、可能義項及前後搭配詞的分析，更具信效度並符合華語文目前實際語用情境。

中介語語料庫內容包括臺灣各大學華語文中心的華語學習者提供的作文，及國家華語測驗推動工作委員會授權的學生考試語料，共約 147 萬字。內容包括題目、性別、文體（口語、書面語）、作者母語、能力等級、性質（限時考試、不限時作業）等資訊。華英雙語語料庫內容涵蓋臺灣特色的文學、科學、財經、藝術、思想、文化、全球、休閒等議題，中文共約 1,160 萬字。

二、具有線上查詢漢字、詞語及語法分級標準

華語文分級標準包括三等七級漢字 3,100 個、詞語 14,470 個及類詞綴 73 個，也包括第 1 級至第 5 級 496 個語法點。數量龐大，不易使用。因此，本平台建置「漢字分級標準檢索系統」、「詞語分級標準檢索系統」及「語法點分級標準檢索系統」，提供使用者查詢漢字、詞語及語法點的等級，漢字、詞語在書面語及口語中的使用頻率，及語法點例句。

三、整合應用語料庫、標準體系及自然語言技術並單一入口服務

除了上述語料庫及標準體系的應用系統，本院更將上述內容整合應用自然語言技術，建置「語義場關聯詞查詢系統」、「例句編輯輔助系統」、及「作文錯別字自動批改系統」等系統。不過，各系統各自有自己的網址，使用者不易記住，因此本院將這些系統整合建置單一入口網站（網址：<https://coect.naer.edu.tw>）（圖 1），提供使用者一站式的服務。

本院為了讓華語教育相關人士了解且應用此系統，除了辦理多場工作坊及競賽活動；並受邀於各大專院校專題演講及於教育部華語教學人員赴美、德、俄任教之行前培訓課程，分享如何應用；目前已有很多華語文教師及研究者使用，使用人次超過百萬及有多篇論文發表。而且，本院與全球華語文教育專案辦公室合作，將北區工作坊課程製作成數位教學影片，詳細介紹與實作應用，網址如下：

I. 簡介及基礎應用

<https://reurl.cc/20Ela4>、<https://reurl.cc/IV7ap6>

II. 進階應用及實作

<https://reurl.cc/R1oKle>、<https://reurl.cc/4RWbYj>、<https://reurl.cc/O1jRQ3>

本書將進一步說明分別如何使用。

圖 1-1 臺灣華語文應用語料庫與臺灣華語文能力基準整合應用系統入口網頁



第二章、語料庫索引典

索引典 (Concordance) 是指在語料庫中，將資料的重要字詞，依某種邏輯順序排列成表，來標示其在原語料的位置，以利於語料的檢索。索引典是搜索語料庫最強而有力的工具，可以找到單詞、短語、標籤、文件、文本類型或語料庫結構 (國家教育研究院，2003 年)。索引典可以分類、分析、計算，以及進一步處理獲得所需的結果。本章依序按索引典介面、萬用字元及 CQP 檢索語法介紹。

第一節、索引典介面

索引典具有查詢功能和分析能力，語料庫基本上具有下列功能：KWIC (關鍵詞索引)、搭配詞 (Collocation)、語料庫查詢語言 (Corpus Query Language, CQL)、詞頻表、近義詞、雙語平行語料庫、詞彙特性素描 (Word Sketch)、詞串。

圖 2-1 國教院索引典系統 (<https://coct.naer.edu.tw/cqpweb/>)

國教院語料庫索引典系統(含國教院華語中介語索引典系統)		
目前使用者: naer [登出]		
Corpora available on this server (click here to view your own corpus access privileges)		
中研院平衡語料庫4.0	COCT 口語語料庫2016	COCT 口語語料庫2017
COCT 口語語料庫2019	COCT 中介語語料庫2016	COCT 中介語語料庫2019
COCT 書面語語料庫2016	COCT 書面語語料庫2017	COCT 書面語語料庫2019
COCT 書面語語料庫2015		

「國教院索引典系統」包括：中研院平衡語料庫 4.0、COCT 口語語料庫 2016、COCT 口語語料庫 2017、COCT 口語語料庫 2019、COCT 中介語語料庫 2016、COCT 中介語語料庫 2019、COCT 書面語語料庫 2016、COCT 書面語語料庫 2017、COCT 書面語語料庫 2019、COCT 書面語語料庫 2015 (排序由左至右、上至下)。COCT 為「國家教育研究院華語文語料庫」(Corpus of Contemporary Taiwanese Mandarin) 的英文縮寫，包括四種不同類型的語料庫：口語、中介語、書面語及華英雙語。以下以「COCT 書面語語料庫 2019」查詢系統為例，說明索引典的介面。本系統語料來源為本計畫截至 2019 年止 (2014 年~2019 年) 所蒐集之書面語，總字數約為 4 億 5 千萬字，其介面如下圖 2-2：

圖 2-2 COCT 書面語語料庫 2019 的介面



索引典介面主要分成 4 區，左邊主要是 Menu（功能選單）：Corpus queries（語料庫查詢）、Saved query data（儲存查詢資料）、Corpus info.（語料庫資訊）。語料庫查詢則具有五種查詢模式：Standard Query（標準查詢）、Restricted Query（限制查詢）、Word lookup（詞語檢索）、Frequency lists（詞頻列表）及 Keywords（關鍵詞），如圖 2-3。

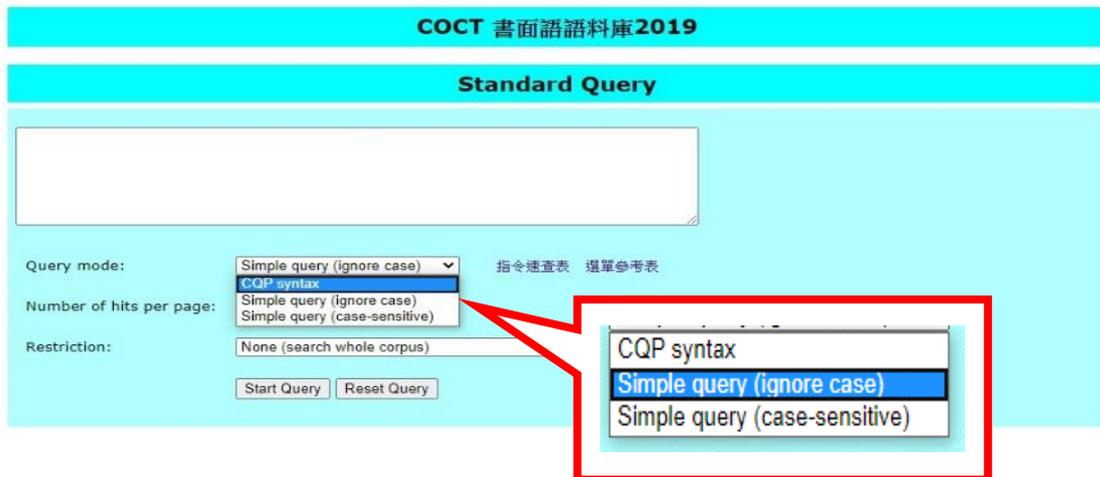
圖 2-3 Corpus queries（語料庫查詢）

Corpus queries	
Standard query	標準查詢
Restricted query	限制查詢
Word lookup	單字查詢
Frequency lists	詞頻列表
Keywords	關鍵字

第一種、Standard Query（標準查詢）

在此模式中 Query mode（查詢模式）有下拉選單，共有三種選項：Corpus Query Processor Syntax（CQP 語法）、Simple query(ignore case)（簡易查詢（不區分大小寫））和 Simple query(case-sensitive)（簡易查詢（區分大小寫）），如圖 2-4。

圖 2-4 COCT 書面語語料庫 2019 的 Standard Query (標準查詢) 介面



第二種、Restricted Query (限制查詢)

此模式的介面分為五區，左邊同樣仍是 Menu (功能選單)：Corpus queries (語料庫查詢)、Saved query data (儲存查詢資料)、Corpus info. (語料庫資訊)。右邊除了 Restricted Query (限制查詢) 的搜尋外，多了條件限制的欄位，也就是選單 Select the text-type restrictions for your query (文本限制設定)，分別為類別和出版年份，此設定可以針對特定的類別和年份縮小搜尋範圍，詳見圖 2-5 中的選單，其類別項下有應用科學類、藝術類、商業及金融類等等，出版年份選項則從 1986 年至 2019 年。

圖 2-5 COCT 書面語語料庫 2019 Restricted Query (限制查詢) 的介面



第三種、Word lookup (單詞查詢)

此模式相較於 Standard Query (標準查詢), 其輸入指令的相關設定也較多, 且輸入指令的介面也全然不同, 如圖 2-6 中的欄位。從圖中發現輸入的指令多了「Show only words...」(僅顯示單字……) 這個欄位, 此欄位有四個選項: starting with (始於), ending with (結束於), containing (包含), matching exactly (完全符合); starting with 是搜尋以關鍵字作為起始的詞表; ending with 則相反, 是以搜尋關鍵字作為結尾; containing 則是會列出有關鍵字的全部詞條; matching exactly 則是要完全符合關鍵字的搜尋模式。此外, 此搜尋模式可以直接於介面上選擇是否顯示詞性, 詳參附件〈中研院平衡語料庫詞類標記集〉¹。

圖 2-6 COCT 書面語料庫 2019 Word lookup (單詞查詢) 的介面

The screenshot shows the 'Word lookup' interface for the COCT 書面語料庫 2019. The interface is divided into several sections: a title bar, a subtitle, a description of the search function, a search input field, a 'Show only words ...' section with radio buttons for 'starting with', 'ending with', 'containing', and 'matching exactly', a 'List results by word-form, or by word-form AND tag?' section with a dropdown menu set to 'List by word-form and tag', and a 'Number of items shown per page:' section with a dropdown menu set to '50'. There are 'Lookup' and 'Clear the form' buttons at the bottom.

第四種、Frequency lists (詞頻列表)

詳見圖 2-7。本項下有「View frequency list for」(詞頻列表) 和「View a list based on」(以~作為基準的詞表) 兩項功能, 「View frequency list for」(詞頻列表) 是選擇來源, 也就是各類的語料庫。另一欄則是「View a list based on」(以~作為基準的詞表), 可選 Word forms 單詞和詞性。下半部的介面 (Frequency list option settings 詞頻列表選項設定) 也有其他的設定, 例如「Filter the list by pattern - show only words/tags」(按照模式分析列表: 僅顯示單字和標誌) 選項則有四: starting with (始於), ending with (結束於), containing (包含), matching exactly (完全符合), 其用法如同 Word lookup (單詞查詢), 詳見圖 2-8。

¹ 檢附中研院表六〈中研院平衡語料庫詞類標記集〉。

圖 2-7 書面語語料庫 2019 Frequency lists (詞頻列表)

圖 2-8 書面語語料庫 2019 Frequency lists (詞頻列表) 的 Frequency list option settings

第五種、Keywords 關鍵詞

藉由兩種不同的語料庫互相比較，此功能可觀察語料在不同語料庫的使用頻率，其比較百萬詞頻為單位，如圖 2-9。

圖 2-9 書面語語料庫 Keywords 關鍵詞

第二節、萬用字元

Query mode 中的 Simple query 檢索可以配合萬用字元使檢索方式更加多元，萬用字元是指：?、*、+等字元。

1. ?：代替一個中文字元，例如：?書→「婚」書、「證」書、「漢」書、「叢」書……。

Solution 1 to 50		Page 1 / 1975
的律令，對於婚姻的規定是要看有沒有設定	婚書	或有沒有授受聘財，但是並沒有限定聘財必須多少
的佼佼者。雖然衛道人士抨擊這是一本「淫書」，父母總是告誡子女不准讀這本書，	淫書	
並購買結婚申請書一份（二元）及結婚公	證書	三份（十五元）即可。對於某些因婚姻
是有史可考的：贅婿的婚制大約是始於秦朝，	漢書	賈誼傳有「家富子壯則出分，家貧子壯則出贅」
一九八四年，我們大膽推出《大眾心理學全集》的系列	叢書	，企圖雄大地編輯各種心理學普及讀物，迄今
了兩個內容：一是「心理學」，指出	叢書	的範圍，但我們採取了更寬廣的解釋，
是「大眾」，我們用它來描述這個	叢書	的「閱讀介面」，大眾，是一種語調
兩個概念經得起考驗，甚至看來加倍清晰。但	叢書	要成交道的讀者組成變了，叢書內容取擇的
清晰。但叢書要成交道的讀者組成變了，	叢書	內容取擇的理念也變了。從讀者面來說
更加廣大、也更加精細（sophisticated）；這個	叢書	同時要了解高度都市化的香港、日趨多元的台灣
而一切實用的卻都是有限的。這個	叢書	將在未來，使「實用的」能夠與時俱進（

2. *：前字重複 0~N 次，例如：哈*→哈、哈「哈」、哈「哈哈」；副*→副、副「手」、副「作用」、副「總統」、副「院長」……。

那並非如願以償，只是自我超越的一種	副	效果而已。」他認為生命的意義雖然常常改變
看，並據此進行認知。母親向來認為電視	副作用	大，禁止林小弟觀看電視。一年級寒假時，
的最典型案例。洛克斐勒（納爾遜）曾任美國	副總統	和數任的紐約州州長，從小患有嚴重的閱讀
不僅敏感，而且害羞，加上脆弱多病，一	副	小小玉樹臨風的樣子，所以在小弟未出生前
很快地乘了直升機，五年之內，登上	副總經理	的寶座。他的老闆，每次在精明出去談
個專門會拍馬屁的職員以外，沒有人喜歡曾	副總	。在公司中，他為什麼這樣令人不喜歡
這些學校寄給我的信，同時也都會將	副本	寄給明大的輔導中心，因為我是外國人，有
會覺得社會充滿了「虛情假意」、「名不	副實	」的現象；他們也會不停地在競爭中
」的聲音就跟到那裡，甚至隨時看到自己一	副	惡貫滿盈的樣子。他們開始不專心原有的工作，
看笑臉，相反地，如果他回來看到的是一	副	晚娘的臉，逃都來不及了，還想準時
主管為了自己的升官和家庭的發財，選用	副主管	時主要的考慮在對方有沒有人事背景，而
廿一日電】在普遍反應高中數學難的今天，臺大	副教授	黃武雄呼籲高中教師，解題不能「太精彩」！

3. +：代替 1~N 個字元，例如：副+→副「手」、副「作用」、副「總統」、副「總經理」……。

名。她為了貼補家用，找了一些家庭	副業	回家做。工資低，工作時間又長，常常
，孩子們的學費卻相對的增加，她做	副業	的眼力和體力都不夠了。有時她希望
，經濟不至於太拮据，也就不再	副業	，同時眼力也不允許。她不打牌，沒有
看，並據此進行認知。母親向來認為電視	副作用	大，禁止林小弟觀看電視。一年級寒假時，
的最典型案例。洛克斐勒（納爾遜）曾任美國	副總統	和數任的紐約州州長，從小患有嚴重的閱讀
很快地乘了直升機，五年之內，登上	副總經理	的寶座。他的老闆，每次在精明出去談
個專門會拍馬屁的職員以外，沒有人喜歡曾	副總	。在公司中，他為什麼這樣令人不喜歡
這些學校寄給我的信，同時也都會將	副本	寄給明大的輔導中心，因為我是外國人，有

除了單獨使用之外，也可以藉由簡易查詢及萬用字元的搭配使用，查到更適合的結果。

1. 字_詞性，例如：把_Nf→一「把」、一「把」蔥。

我記得老師看到可愛的小孩總喜歡捏他一	把	，抓他一下，我當時多希望老師也抓
他一下，我當時多希望老師也抓我一	把	、捏我一下，可是就沒有這樣的機會。
上街買菜，臨別時，如果偷偷地抓了一	把	蔥，在旁的年壽毫無考慮地也順手抓
客廳裡的洋娃娃一樣，高興時可以摸他一	把	，在外面被上司罵了，就回來訓他
的可愛，她可愛得讓人想捏他一	把	。那位顧客就是因為捏她一把，

2. 搭配萬用字元

A. 例如：「把_N*」→手「把」、火「把」、掃「把」、拖「把」。

衝向他們？這些都不重要了，她緊握安全	手把	，聞到毒氣混濁的焦味，曲身抱住頭顱，用
沒有落地，忽然湧現出一群歹徒，他們舉著	火把	、利劍，叛徒猶大（註46）為
的火車飛馳而過，彷彿一個人手舉	火把	從田野向他家跑來。一輛向西行駛
。低飛的雲絮幾乎觸到樹梢，彷彿繫成一柄	掃把	，打掃天空。打在木板牆上的灰白色冷雨
傭人，他們會辛勤勞動一整天，拿著	拖把	、清潔刷、錘子、園藝剪等工具，修補前

B. 例如：「+把_N*」→車「把」、火「把」、手「把」、掃「把」、拖「把」。

，那是一個奴隸的恐懼，但仍抓著	車把	，不要命地緊抓不放... You remain here!
輪船時，我詫異地看到一些小船亮著	火把	和煤氣燈，正繞著大船搜尋，同時有
衝向他們？這些都不重要了，她緊握安全	手把	，聞到毒氣混濁的焦味，曲身抱住頭顱，用
沒有落地，忽然湧現出一群歹徒，他們舉著	火把	、利劍，叛徒猶大（註46）為
的火車飛馳而過，彷彿一個人手舉	火把	從田野向他家跑來。一輛向西行駛
。低飛的雲絮幾乎觸到樹梢，彷彿繫成一柄	掃把	，打掃天空。打在木板牆上的灰白色冷雨
傭人，他們會辛勤勞動一整天，拿著	拖把	、清潔刷、錘子、園藝剪等工具，修補前

C. [,]：內的字詞符合任一即可，例如：[台,臺]灣→臺灣、台灣。

更年期無自我生活的中年婦女等），是目前	臺灣	社會最容易瞭解寂寞的人們。」這些寂寞的
，自宅或其他場所舉行均可」。而且「	臺灣	地區各地方法院皆設有公證處，亦辦理公證結婚事
應征收公證費二十元。除此公證費外，依	臺灣	臺北地方法院所印製《公證結婚登記須知》，申請
鄰居和社區之間就產生相互重疊的影響。以	臺灣	的現況為例，新社區如雨後春筍般出現，
。《大眾心理學全集》的出版，立刻就在	台灣	、香港得到旋風式的歡迎，翌年，論者更
心理學有著密不可分的聯結印象，一方面也解釋了	台灣	社會在群體生活日趨複雜的背景下，人們如何

D. (|)：會搜索語料庫內的所有詞（類），例如：喜歡(我|你)→喜歡我、喜歡你。

「你好乖，你要做乖寶寶，老師	喜歡你	。」一般的孩子聽到老師這樣鼓勵，應會
裡的一種報復心理，他認為「你們不	喜歡我	、恨我，我也恨你們。」所以看到
狗做什麼？」「那別的小朋友就會	喜歡我	。」）（3）提供替代品——例如幫孩子
屬下有六個領班和七十五個男女員工。你	喜歡你	的工作、你的上司、你的下屬，而且

E. (+){數字,數字}：可限定夾帶的詞的數量，例如：一(+){4,6}就。

有什麼花招，便答道「好，你先走	一步，我隨後就	到」阿色走後，喻哪雅烏對
，隨時都想逃跑，只要它的身子稍為□□	一動，整個大地就	會震蕩起來，為了防備鯊魚逃跑
啄它的眼睛，有時候，金雞太疲倦了，	一閉眼睛，鯊魚就	動起來，發生地震，這時候，人們
在箴帽上，戴著遮太陽，可是	一走出門，蠟就	被太陽曬化了，一滴一滴地□□

F. <<數字>>：前後關鍵字字詞的順序不固定，且限制夾帶的詞的數量，例如：一 <<3>> 就。

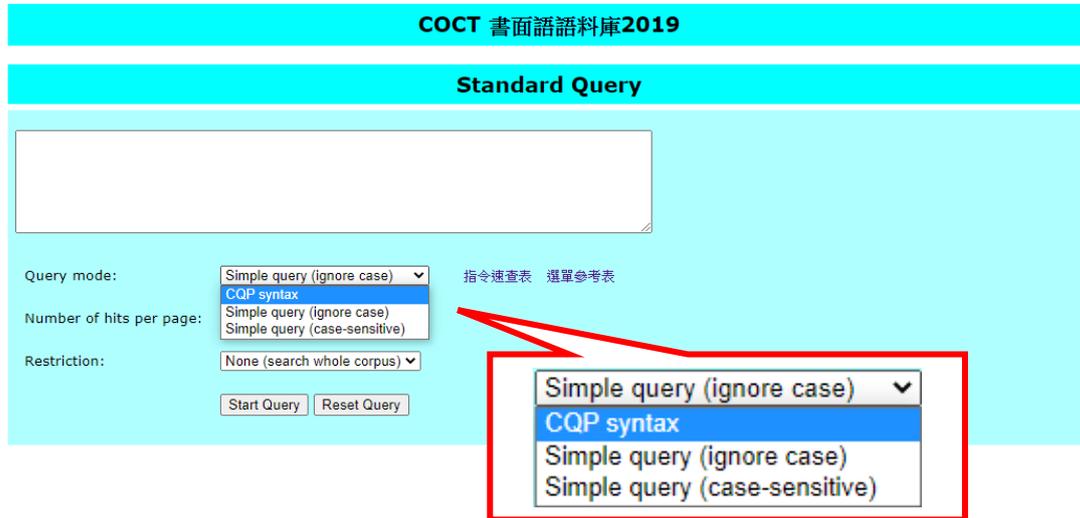
在五倫之中。人們從小學習怎樣結交朋友，等	一結婚就跟朋友之間定下那麼多的禁忌，甚至由於
了。(六)有的男生有征服感，和女朋友有了	一一次性行為就要做第二次，甚至還要換
更密切，就不會失去對方。有極少部分是	一開始就不在乎這件事，認為這沒什麼
事，結果一時衝動就做了，而且有很多	一次就懷了孕。未婚懷孕的女孩有不少

G. >>數字>>：前後關鍵字字詞的順序固定，且限制夾帶的詞的數量，例如：一>>3>>就。

大多數的銀行也有不成文的規定，女職員	一結婚就要辭職；有些讀者投書給我，
歲，而且對什麼都是無所謂的態度。他	一開始就用這種態度約她下班後去喝
不情不願的心情下生下他，所以他	出世就很少得到父母的愛。後來父親雖然「
很乖很乖」。父母都在上班，早上	出門，就要她們鎖起門來做功課，連

第三節、CQP 檢索語法使用

Simple query 檢索在查詢上仍有其限制，因此，就需要使用 CQP 語法，CQP 語法能提供更細緻的查詢功能。例如，CQP 語法可以排除標點符號，讓搜尋範圍限定於一個句子內。



CQP 語法的基本使用原則為一個方框只能使用一個詞，例如：[word="快樂"]，如圖 2-10；如果想要一次查詢兩個詞彙以上，則需鍵入[word="][word="]"，例如[word="台北"][word="捷運"]，詳見圖 2-11。

圖 2-10 2019 書面語料庫檢索[word="快樂"]



圖 2-11 2019 書面語料庫檢索[word="台北"][word="捷運"]



用限制詞性的方式 [pos=" "]調整搜尋的內容，另一種則是[pos!=" "]，此用法是為了搜尋關鍵詞性以外的詞。此處將以[pos="Nf"] [word="電話"]將搜索限制於量詞，可看出量詞有「次、個、通」等，如圖 2-12。或是輸入[pos!="Nf"] [word="電話"]，則是排除量詞的檢索，顯示「張老師、以、的」等非量詞的詞，詳見圖 2-13。

圖 2-12 2019 書面語語料庫檢索[pos="Nf"] [word="電話"]檢索結果

Solution 1 to 50 Page 1 / 96	
，坐在電話機旁等待別人打電話過來，結果整	晚電話 都沒有響，而他也在酒醉中醉倒。
】：讓孩子順利度過自我為中心的語言期每	次電話 一響，小孩子個個都像極拼命三郎似地搶
的交誼和家庭的穩定。七、親近上帝打	個電話 到 W P C B (Western Pennsylvania Christian Broadcasting Company) 有線電
經營者或管理人員從早忙到晚，一	個電話 剛掛斷，另一個電話又響了；一
忙到晚，一個電話剛掛斷，另一	個電話 又響了；一個約會還未結束，已經
這樣的緣，讓我們成了朋友，只要一	通電話 ，他便會成為我的司機。他是個愛

圖 2-13 2019 書面語語料庫檢索[pos!="Nf"] [word="電話"]檢索結果

Solution 1 to 50 Page 1 / 679	
福利系的「張媽媽信箱」，青年服務社的「	張老師 電話 等等都提供教養孩子的指導)，不同「
比起來，顯然好得多了。其後我們連續	以 電話 追蹤三星期，楊小妹怕睡覺或怕自己待在房
C」成績。社交上沒有男朋友的邀約及女朋友	的 電話 ；另外，母親對成績不適當的稱讚、過度
事情，和怎樣幫助你；有困難時媽媽會	用 電話 跟老師討論如何幫忙，或是把帽子領回來。」
和視聽均不成問題。他向來容易忘記人名	和 電話 號碼，也視此為畏途，因此常依賴小

為了要得出更加精準的結果，語法指令也會更加的複雜，例如想要觀察量詞「頭」的使用方式時，有幾種語法指令可以輸入，也會得出不一樣的結果。第一個、[word="頭"] []；第二個、[word="頭"&pos="Nf"] []；第三個、[word="頭"&pos="Nf"] [pos="Na"]，結果分別如圖 2-14、圖 2-15、圖 2-16。

圖 2-14 2019 書面語語料庫檢索[word="頭"] []檢索結果

，父母以孝或不孝的大帽子扣在子女	頭上	，使子女日夜陷於孝與不孝的困惑中。
一個來。」萬物靜下後，乾茅草伸出	頭來	，說「我來哭，準能表達大家的
？」有一個更認為跟異性緊握雙手，	頭靠	頭就會懷孕，這種無知簡直到了匪夷所思
有一個更認為跟異性緊握雙手，頭靠	頭就	會懷孕，這種無知簡直到了匪夷所思的地步
告訴我，她比她的男朋友矮差不多兩個	頭，	雖然她每次跟他出去都穿起碼五寸
我的感情也全放在他身上。結婚	頭兩	年們好像生活在蜜裡一樣，我認為我

圖 2-15 2019 書面語語料庫檢索[word="頭"&pos="Nf"] []檢索結果

Solution 1 to 50		Page 1 / 357
是從頭人撮高家的牛群裡挑來的一	頭大	白牛，比賽結果是白牛鬥敗，黑牛得勝
尋求階段，他是在丹麥出生的猶太人，一	頭金髮	、魁梧的身材，看來就是一個典型的
。有一次，我要從餐廳的這一	頭走到	另一頭；那時，餐廳裡都是女生在
，我要從餐廳的這一頭走到另一	頭；	那時，餐廳裡都是女生在打掃地板，他們
練習句子，例如：「這湯匙比○○大」「這	頭豬	比○○小」「這棵樹跟○○一樣綠」。
是先摘下兩枝帶穗的牛筋草，兩	頭相向	重疊排列，然後彼此把草莖頂端的穗綁

圖 2-16 2019 書面語語料庫檢索[word="頭"&pos="Nf"] [pos="Na"]檢索結果

Solution 1 to 50		Page 1 / 139
尋求階段，他是在丹麥出生的猶太人，一	頭金髮	、魁梧的身材，看來就是一個典型的
練習句子，例如：「這湯匙比○○大」「這	頭豬	比○○小」「這棵樹跟○○一樣綠」。
沒想到也有灰心喪志的一天。他以一	頭亂髮	，訴說心中的無助。「一無所有了，我
信不信由你，我們學校差點要養一	頭豬	，只是礙於水源區水質保護的規定，三位想
是很想呀，你知道夏天的時候，這	頭長髮	真的好煩喔！短髮看起來清爽又好整理，
· 育】2學習照顧別人常聽見人安慰那	頭胎	生女的沮喪母親說：「第一個生女兒

如前述所言，圖 2-14、圖 2-15、圖 2-16 都是以[word="頭"]為例。第一組指令[word="頭"] []，因為「頭」本身具有名詞跟量詞的特性，又因指令不夠精準，故無法判定是要搜索名詞的「頭」、還是量詞的「頭」，所以得到圖 2-14 的列表，並於此列表中同時觀察到兩種詞性的使用方式及搭配。第二組指令[word="頭"&pos="Nf"] []，因有設定 pos="Nf"，也就是將詞性限定為量詞，所以圖 2-15 出現量詞的比例較高，例如：「一頭金髮」、「這頭豬」等。但和第三組指令([word="頭"&pos="Nf"] [pos="Na"]) 相較，給予更多指令的結果如圖 2-16，明顯能看出僅有一條詞條並非量詞。換而言之，給予的指令越詳細，便能得出越精確的結果，例如：「一頭金髮」、「這頭豬」、「一頭亂髮」、「這頭長髮」等。

因為中文的特殊性，以離合詞為例，如：「畢不了業」、「見了面」、「睡了一覺」……等詞。當想要運用語料庫搜尋離合詞時，僅需鍵入[word="見"] []*[word="面"]，便會出現下圖 2-17，檢索結果有「見過面」、「見了面」、「見人愛、不好看等等，將它們分為正、中、負三面」、「見過他的頑皮樣，他成了城市隱士，只在吃飯時出來露個面」、「見過面」。

圖 2-17 離合詞「見面」部分檢索結果

見過面
見了面
見了面
見人愛、不好看等等，將它們分為正、中、負三面
見過他的頑皮樣，他成了城市隱士，只在吃飯時出來露個面
見過面

圖 2-17「見面」有嚴重跨句的現象，例如圖 2-17 中的第四句例句：「見人愛、不好看等等，將它們分為正、中、負三面」，所以為了要避免如圖 2-17 中的狀況，便要利用較為複雜的指令達成較細緻的成果，以下方法有二：

1. [word="見"][]{}{1,5}[word="面"]

第二個[]後面加上大括號{,}並填上數字，用意為限制中間詞組的字數，以便達到以下這些成果，例如：「見過面」、「見了第二次面」、「見不到面」等等，詳可見圖 2-18。

圖 2-18 2019 書面語語料庫檢索[word="見"][]{}{1,5}[word="面"]檢索結果

時間，我還在台和美都先後和她	見過面	，她都明顯地適應得很好，也很
每月，或竟然每半年、一年才	見個面	的養育方式。「內製外包」做節目的方式
積極的父母介入並與治療師討論如何處理問題，	見了第二次面	之後就有極大的進步，整個人情緒
看到阿諾回部落去找死黨羅迪，羅迪明明在家卻不願	見他，羅迪的父親面	無表情的說「羅迪不在」，他離開
空間的約會資料，它就非常準確，一定不會	見不到面	。透過學生已有的經驗，他也就了解
一驚，我在她女兒考上名校的慶功宴上	見過她女兒一面	，是個十分乖巧聰慧的孩子，怎麼會變成
人就被迫分離。老太太身體越來越虛弱，但渴望	見丈夫一面	，安慰他幾句話，所以她每天練習

2. [word="見"][word!="，|。"]*[word="面"]

此指令的目的是對標點符號設限。藉由限制「見」、「面」兩字當中不出現標點符號，以達成不跨句的成果，例如：「見過面」、「見了面」等等，其他例子可詳如圖 2-19。然此種指令只適用於書面語索引典，因口語索引典語料並未使用標點符號，故不適用。

圖 2-19 2019 書面語語料庫檢索[word="見"][word!="，|。"]*[word="面"]檢索結果

Solution 1 to 50 Page 1 / 64		
「今天你們老師請假，我介紹一位你們從未	見過面	的代課老師，為了了解你們對老師的反應
一位導演外，我全不認識。在日本	見了面	後，他們一致強調「有朋自遠方來」
人逛街去了。有的人使用一種儀式，	見了面	總是問：「嗨！你好嗎？」
困擾。她是一個養女，經常怨恨未曾	見過面	的生父母，為何他們會那般狠心，一出生
。去年春天，與札帕達公司的九人評議委員	見了面	，從此便揭開他朝這個方向發展的序幕
年多裡，我與雅卿仍常相聚首。	見了面	，所談不外乎學苑種種。似乎在雅卿心中

圖 2-18 及圖 2-19 所下的指令方向不太相同，但藉由這兩組指令，均可完成限制例子「見面」當中的文字數量及標點符號，使得成果更加的精細。

第四節、進階功能

使用萬用字元或CQP語法檢索後，便能看到下圖 2-20 語料庫所呈現的介面，此圖是以「出口」一詞為例，可看出檢索基本資料，例如：筆數、文本、百萬詞頻²、總字數及文本等。

圖 2-20 書面語語料庫檢索結果介面

Your query "出口" returned 10,817 matches in 6,649 different texts (in 319,712,694 words [168,090 texts]; frequency: 33.83 instances per million words)

分布在 6649 個文本中

每百萬詞頻

關鍵字

共找到 10817 筆

語料庫總字數、總文本

No	Filename	Solution 1 to 50	Page 1 / 217
1	576a5609326d1b3fb77fe231	閱讀。此時最敏感的壓痛點為後頭第一頸神經 出口 的枕下三角肌群。抬頭時，因對該	
2	576a5609326d1b3fb77fe231	。此時最敏感的壓痛點，為後頭第一頸神經 出口 的枕下三角肌群。抬頭時，因對該	
3	576a5609326d1b3fb77fe27f	不誠實」是相當籠統的形容詞，一般人不經意地 出口 「誠實」、「不誠實」，好像「誠實	
4	576a5609326d1b3fb77fe293	新文化運動。這些新文化運動均以生產和 出口 香薰為目的，而以近十五年來為最。	

除了以上的基本查詢之外，圖 2-20 介面的右上方還有「再查詢」。「再查詢」包含了：Thin 語料隨機取樣、Frequency breakdown 頻率列表、Distribution 詞彙分布、Sort 按前後文排序、Collocation 搭配詞統計、Download 下載查詢結果、Categorise 樣本分類標記、Save current set of hits 儲存關鍵字。

圖 2-21 查詢結果介面中的「再查詢」

Your query "出口" returned 10,817 matches in 6,649 different texts (in 319,712,694 words [168,090 texts]; frequency: 33.83 instances per million words) [3.493 seconds]

New query

- New query
- Thin...
- Frequency breakdown
- Distribution
- Sort
- Collocations...
- Download...
- Categorise
- Save current set of hits...

除了基本詞語的查詢之外，為了展現 CQP 語法結果與 Simple Query 之間的差異，以下以[word="頭"&pos="Nf"] [pos="Na"]作為例子。

² 每百萬字中會出現幾次。

圖 2-22 2019 書面語語料庫檢索[word="頭"&pos="Nf"] [pos="Na"]檢索結果

Your query "[word="頭"&pos="Nf"] [pos="Na"]" returned 6,914 matches in 4,837 different texts (in 319,712,694 words [168,090 texts]; frequency: 21.63 instances per million words) [0.458 seconds - retrieved from cache]			
Show Page: 1		Line View	Show in random order
No	Filename	Solution 1 to 50	Page 1 / 139
1	576a5609326d1b3fb77fc23d	尋求階段，他是在丹麥出生的猶太人，一	頭金髮、魁梧的身材，看來就是一個典型的
2	576a5609326d1b3fb77fc335	練習句子，例如：「這湯匙比○○大」「這	頭豬比○○小」「這棵樹跟○○一樣綠」。
3	576a5609326d1b3fb77fc3e7	沒想到也有灰心喪志的一天，他以一	頭亂髮，訴說心中的無助，「一無所有了，我
4	576a5609326d1b3fb77fc414	信不信由你，我們學校差點要養一	頭豬，只是礙於水源區水質保護的規定，三位想
5	576a5609326d1b3fb77fc525	是很想呀，你知道夏天的時候，這	頭長髮真的好煩喔！短髮看起來清爽又好整理，

從上圖 2-22 中可觀察出此組指令於書面語語料庫中一共找到 6914 筆，分布在 4837 個文本中，其百萬詞頻為 21.63 次/每百萬詞。藉由「再查詢」，能夠更深入搜尋。

1. Thin

在右上角點選 Thin 後，便會看到圖 2-23 的介面，此欄框可輸入數字或百分比，再點選 Thin this query 精簡搜尋，即可看到圖 2-24，此步驟能夠從語料庫中自動選擇出所要的詞條數量，如圖 2-24 所顯示的則為 15 條。

圖 2-23 進階搜尋 Thin 介面

Choose options for thinning your query (current no. of instances: 18,109)

Thinning method:	<input type="text" value="random (selection is reproducible)"/>	<input type="text" value=""/> (number of instances or percentage)	<input type="button" value="Thin this query"/>
------------------	---	--	--

圖 2-24 進階搜尋 Thin 搜尋結果

Your query "[word="頭"&pos="Nf"] [pos="Na"]" returned 6,914 matches in 4,837 different texts (in 319,712,694 words [168,090 texts]; frequency: 21.63 instances per million words), thinned with method random selection to 15 hits [0.156 seconds]			
Show Page: 1		Line View	Show in random order
No	Filename	Solution 1 to 15	Page 1 / 1
1	576a560b326d1b3fb77fe0b4	，另外一個護士卻記得她體型厚重，一	頭金髮，尼勒警官向副局長報告說：「就這樣，
2	576a560c326d1b3fb77ff886	悄悄躲進海邊一個大石洞裡，生下了一	頭渾身金黃的寶角牛。寶角牛長得很快，
3	576a560c326d1b3fb77ff85e	又「吱吱」地合上了門。牛歡轉著兩	頭牛回到二十四畝地裡，套上犁子，他只輕輕
4	576a560d326d1b3fb7800b05	日子長短，都沒法對電車視若無睹，緩慢的一	頭綠色巨獸，早年是單層的，其後變高了，
5	576a560f326d1b3fb7801957	長老來了，大家才知道，於是要賠三	頭牛，雖然很難賠得起，也得賠。」苦
6	595c5e23326d1b284d8e4e4d	件破了洞冬衣的乾瘦小男孩，一	頭黑髮散到了肩，渾身顫抖地蜷縮在樹底。
7	595c5e28326d1b284d8e934	。她身材修長，有如天仙般的姿態，一	頭金髮在陽光中熠熠發光。我自付著，這是
8	595c5e29326d1b284d8eeeb3	在路上走的時候，又見到了我那	頭驢子。我一眼就認出來了，那個騎毛驢的
9	595c5e29326d1b284d8eeebf	面目奇醜的魔鬼，他領著駕車的幾	頭驢子在前頭走著，這輛大車是散著的
10	595c5e2a326d1b284d8ed170	清楚一點。一張粗壯的臉，還有一	頭黏答答的及肩金髮，大約三十歲，身上散發出
11	595c5e2f326d1b284d8f2465	兜風。那是一個炎熱的八月天，瑞秋把一	頭紅髮梳在腦後綁成馬尾，身上穿著泳衣，
12	595c5e2f326d1b284d8f24b7	處所，最慘的是，這隻寵物還是	頭獅子！我們花了好幾天拜訪房地產仲介，想找
13	5b334b1f9027f3ebd2203465	。然關於歐洲事項，大部似已決定於三	頭會議開幕之前，只待特魯曼、邱吉爾與史太林三人
14	5b334b259027f3ebd2204373	部分變得模糊不清，麗英，她曼像一	頭獅子在那裡掙扎不已，〈原始的喊叫〉顯得很
15	5b334b389027f3ebd2205b53	公這樣凶惡的面孔，他雙手平舉，像一	頭猩猩，把大家逼退到營火旁。大聲公盤腿坐下，

2. Frequency breakdown

其步驟同上，右上方點選 Frequency breakdown 後，介面會跳轉至圖 2-25 的介面。從詞頻彙整的部分可以觀察到常用度比例的高低，例如從此圖 2-25 中便會發現常用比例最高的為「頭牛」，次高為「頭豬」，以詞性來判斷，均為量詞。

圖 2-25 進階搜尋 Frequency breakdown 搜尋結果

Your query "[word="頭"&pos="Nf"] [pos="Na]" returned 6,914 matches in 4,837 different texts (in 319,712,694 words [168,090 texts]; frequency: 21.63 instances per million words)

Showing frequency breakdown of words in this query, at the query node; there are 912 different types and 6,914 tokens at this concordance position.

|< << >> >| Breakdown position: Node Frequency breakdown of words only Go!

No.	Search result	No. of occurrences	Percent
1	頭牛	872	12.61%
2	頭豬	325	4.7%
3	頭金猴	219	3.17%
4	頭獅子	178	2.57%
5	頭大象	175	2.53%
6	頭長猴	148	2.14%
7	頭黑猴	148	2.14%
8	頭鹿	138	2%
9	頭羊	124	1.94%
10	頭獅子	127	1.84%
11	頭黑猴	118	1.71%
12	頭椰子	117	1.69%
13	頭紅猴	106	1.53%
14	頭白猴	103	1.49%
15	頭野獸	96	1.39%

3. Distribution

點選 Distribution 詞彙分布後，便會看到如下圖 2-26 的介面。可從此部分觀察出此組指令當中關鍵字的使用類別趨向，例如於圖 2-26 中，類別的部分應用科學類使用量詞「頭」的比例最高，另外還有藝術類、商業金融類等等，另外一部分「出版年份」則是 2011 年最高。

圖 2-26 進階搜尋 Distribution 搜尋結果

Distribution breakdown for query "[word="頭"&pos="Nf"] [pos="Na"]": this query returned 6,914 matches in 4,837 different texts

Categories: Show as:

Category for crosstabs: Show distribution

Based on classification: 類別

Category [1]	Words in category	Hits in category	Dispersion (no. files with 1+ hits)	Frequency [1] per million words in category
應用科學類	19767279	302	200 out of 9074	15.28
藝術類	8420129	100	76 out of 4573	11.88
商業及金融類	21464158	155	124 out of 10774	7.22
總類	3300362	48	39 out of 1274	14.54
史地類	34507282	550	389 out of 18263	15.94
語言文學類	79759661	2692	2029 out of 42143	33.75
哲學及宗教類	31478990	411	283 out of 16293	13.06
科學類	7603892	213	147 out of 3194	28.01
社會科學類	112219233	2403	1518 out of 61827	21.41
休閒類	1191708	40	32 out of 675	33.57
Total:	319712694	6914	4837 out of 168090	21.63

Based on classification: 出版年份

Category [1]	Words in category	Hits in category	Dispersion (no. files with 1+ hits)	Frequency [1] per million words in category
1986	220095	1	1 out of 76	4.54
1988	70028	0	0 out of 35	0
1989	280484	3	2 out of 145	10.7
1991	58414	0	0 out of 19	0
1992	52376	0	0 out of 32	0
1993	156127	0	0 out of 82	0
1994	608462	2	2 out of 380	3.29
1995	48897	0	0 out of 16	0
1996	150464	1	1 out of 61	6.65
1997	68005	1	1 out of 22	14.7
1998	286579	1	1 out of 138	3.49
1999	690948	3	3 out of 361	4.34
2000	2610776	62	41 out of 1093	23.75
2001	1141451	24	21 out of 522	21.03
2002	2270581	82	61 out of 1042	36.11
2003	3597589	80	75 out of 2059	22.24
2004	7009344	100	82 out of 1911	14.27
2005	4924421	95	64 out of 2074	19.29
2006	19216038	384	287 out of 10300	19.98
2007	16228476	255	189 out of 8565	15.71
2008	16842018	281	196 out of 8346	16.68
2009	20118647	586	347 out of 9324	29.13
2010	32965073	520	411 out of 18011	15.77
2011	25917405	1094	608 out of 14277	42.21
2012	17756773	303	247 out of 9460	17.06
2013	22849476	486	373 out of 12065	21.27
2014	29468929	614	466 out of 14867	20.84
2015	28264325	661	475 out of 14792	23.39
2016	38643503	823	563 out of 22512	21.3
2017	14648843	286	177 out of 8397	19.52
2018	12510624	166	143 out of 7084	13.27
2019	37523	0	0 out of 22	0
Total:	319712694	6914	4837 out of 168090	21.63

4. Sort 按前後文排序

點選 Sort 後，便會進入圖 2-27 的介面，可藉由點選 Position 位置的部分，再搭配 Tag Restriction 標籤限制，能夠觀察出關鍵前（左）後（右）文詞組的使用狀況，如圖 2-27。從圖 2-27 中，可以看出後（右）文是以名詞居多，例如：「牛」、「王」、「野獸」。

圖 2-27 進階搜尋 Sort 搜尋結果

Your query "[word="頭"&pos="Nf"] [pos="Na"]" returned 6,914 matches in 4,837 different texts (in 319,712,694 words [168,090 texts]; frequency: 21.63 instances per million words), sorted on position +1 (6,914 hits) [0.116 seconds - retrieved from cache]			
<div style="display: flex; justify-content: space-between;"> < << >> > Show Page: 1 Line View Show in random order New query Go! </div>			
<div style="display: flex; justify-content: space-between;"> Sort control: Position: 1 Right Tag restriction: None Starting with: Update sort </div>			
No	Filename	Solution 1 to 50	Page 1 / 139
1	5b334b229027f3cbd2203ca1	(Pluto) · 頭圈馬 (Harace Horsecollar) 和一	頭牛 (Clarabelle Cow) · 小布魯托有人叫他
2	576a560c326d1b3fb77ff884	斑枝花：一名攀枝花，似木棉 · 蘭嘎十	頭王 (一)很久很久以前，在西南方蔚藍
3	576a560c326d1b3fb77ff886	四處尋找，看不到噶希納的影子 · 蘭嘎十	頭王 (三) 召喚瑪和蘭嘎四處尋找噶希納，不知
4	576a560c326d1b3fb77ff885	給勳達塔塔帶來了莫大的驕傲和光榮 · 蘭嘎十	頭王 (二) 瘋麻溫達森林裡修行的那
5	576a560c326d1b3fb77ff888	獅子兵趁機殺了過去，大獲全勝 · 蘭嘎十	頭王 (五) 十頭王失掉了英達西打，好比
6	576a560c326d1b3fb77ff887	長毛，據說就是這麼來的 · 蘭嘎十	頭王 (四) 阿努曼火燒了蘭嘎京城，十頭王
7	5b334b399027f3cbd2205e2e	那老頭不止是快要死了，而且快變成一	頭野獸 (我們早先已看到那老頭毛茸茸的臉和獸

如圖 2-28 中，倘若在 Position 選擇了 1Left，Tag Restriction 選擇 Na，便能從此圖表中看出文句裡，置中關鍵字的左方名詞變成了粗體，且有標明其詞性 Na，有助於使用者輕易地分析詞句當中的使用。

圖 2-28 進階搜尋 Sort 介面設定

Your query "[word="頭"&pos="Nf"] [pos="Na"]" returned 6,914 matches in 4,837 different texts (in 319,712,694 words [168,090 texts]; frequency: 21.63 instances per million words), sorted on position -2 with tag-restriction Na (197 hits) [0.195 seconds]			
<div style="display: flex; justify-content: space-between;"> < << >> > Show Page: 1 Line View Show in random order New query Go! </div>			
<div style="display: flex; justify-content: space-between;"> Sort control: Position: 2 Left Tag restriction: Na Starting with: Update sort </div>			
No	Filename	Solution 1 to 50	Page 1 / 4
1	595e5e2f326d1b284d8f224d	了名的放高利貸者，他借了曼多沙上尉_Na 兩	頭驃子，一定會需索無度，甚至會要求整個寶藏的
2	576a5612326d1b3fb7802511	前面提過的，我們以前從來沒有與中年_Na 道	頭怪獸 對戰過，事實上，其他有關中年的迷思
3	576a5612326d1b3fb7802517	前面提過的，我們以前從來沒有與中年_Na 道	頭怪獸 對戰過，事實上，其他有關中年的迷思
4	576a560c326d1b3fb77ffda4	，火種雖有，但只有得到他的主人_Na 一	頭老爺 的同意才能給。三總總沒法，問明了路程
5	5b334b379027f3cbd2205863	大武權設置義塾，並規定牛墟內每交易_Na 一	頭牛 納稅一百文錢，做為辦理義塾的經費。目前
6	576a560c326d1b3fb7800262	官府共同發給每人三兩銀子，每三人_Na 一	頭牛，一下子解決問題。鄭芝龍當時的商業勢力之大
7	576a560a326d1b3fb77fdd4	名十九歲左右女孩的畫像，畫中人_Na 一	頭金絲 般的頭髮，大大的藍眼睛裏笑意迎人

5. Collocation 搭配詞統計

選擇 Collocation 後，介面將會連結至頁面如圖 2-29，再點選 Create Collocation Database，就能夠看到如下圖 2-30 的介面。Collocation 使用主要目的在於觀察詞組與詞組之間關係的強度，例如於此圖 2-30 中可看出與關鍵字「頭」連結性最高的詞為「一」、次之則為「九」，均為數字。反而直到序號五之後才開始有名詞，例如：「驢」和序號九的「牛」，由此可知量詞、數量詞及名詞之間的緊密度。

圖 2-29 進階搜尋 Collocation 介面

Choose settings for proximity-based collocations:

Include annotation: 詞性 Include Exclude

Maximum window span: + / - 5

Create collocation database

圖 2-30 進階搜尋 Collocation 搜尋結果

Collocation controls

Collocation based on: Word form Statistic: Log-likelihood

Collocation window from: 3 to the Left Collocation window to: 3 to the Right

Freq(node, collocate) at least: 5 Freq(collocate) at least: 5

Filter results by: specific collocate: and/or tag: (none) Submit changed parameters Go!

Extra information: Log-likelihood scores collocations by significance: the higher the score, the more evidence you have that the association is not due to chance. More frequent words tend to get higher log-likelihood scores, because there is more evidence for such words.

There are 9,969 different words in your collocation database for "[word="頭"&pos="Nf"] [pos="Na"]". (Your query "[word="頭"&pos="Nf"] [pos="Na"]" returned 6,914 matches in 4,837 different texts) [1.58 seconds - retrieved from cache]

No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed collocate frequency	In no. of texts	Log-likelihood
1	一	3,701,681	480.309	4,213	3372	11186.101
2	九	34,506	4.477	283	68	1793.961
3	那	754,138	97.853	741	598	1724.668
4	用	526,190	68.275	494	371	1108.525
5	種	2,872	0.373	102	88	945.473
6	嘗	1,004,687	130.362	562	524	783.802
7	響	596	0.077	63	63	725.744
8	費	16,614	2.156	104	93	603.453
9	牛	22,594	2.932	108	91	569.634
10	長	2,243	0.291	64	64	564.851

第五節、華英雙語索引典

本計畫所建置之華英雙語語料庫提供使用者中英兩種語言的雙向查詢，外國學生可藉由英文做為檢索的媒介，觀察相對應的中文使用方式。

圖 2-31 華英雙語索引典系統

國教院華英雙語索引典系統

語料庫 關鍵詞:

自動對應翻譯

自動抽取搭配詞

以「老師」一詞為例，華英雙語語料庫可以提供「老師」的中英文例句，如圖 2-32。且於同一介面中，右上方的欄框為英文的對應翻譯，及相關詞的英文單字，從圖中可以看到部分對應英文翻譯，例如 teacher, professor, instructor 等等，除了列舉的這三個英文單字之外，往下捲動可以看到更多的對應選項。右下方的欄位則是搭配詞及其詞性，有助於學習常用搭配詞，從此欄當中可以觀察到系統介紹了「老師」的詞性：名詞，也列出了常搭配使用的 M (量詞)：位～；N (名詞)：學校～、國小～、學生～等；P (介詞)：跟～；Vt (動作使動動詞)：當～、擔任～；N (名詞)：～學生、～們等。此外，右方兩個欄位中的選項均可點選，以下將以右下方的搭配詞為例。藉由點選自動抽取搭配詞，有助於加深學習範圍，例如：點選自動抽取搭配詞當中的量詞「位」，便會看到更多中英文的翻譯對照，如圖 2-33。

圖 2-32 華英雙語索引典以「老師」一詞為例查詢結果

國教院華英雙語索引典系統

語料庫 關鍵詞:

總共找到 3584 << 上一頁 1 下一頁

- The young Chou first revealed his native artistic talent under his father's tutelage; then during his school years two **teachers** served as his artistic mentors at different times: his junior high school art **teacher** Yang Chien-chung and his high school Chinese **teacher** Chiang Chao-shen.
在父執輩的教導下，年少時的周澄已嶄露藝術才華；其後求學過程中，引領他進入藝術殿堂的兩位**老師**，又各自代表兩個重要階段的開始，一是國中時期的美術**老師**楊乾鍾、一是高中時期的國文**老師**江兆申。
- and how at the kindergarten he used Hangzhou dialect to tell his **teacher** that he wanted to go to the toilet, to his **teacher**'s mystification and to everyone's amusement.
兒子在幼稚園裡用杭州話跟**老師**說要「撒稀」（小便）、「撒汗」（大便）時，也弄得**老師**一頭霧水，鬧了不少笑話。
- The children learn about life, customs, and the adult world in general at the **school**.
幼稚班**老師**周嘉玲指出：「蘭生幼稚園每班有兩位**老師**，但我們並不打算灌輸孩子許多知識；而是希望他們能有個愉快的學習過程，養成他們樂於學習的態度。」
- When he first began studying, Lin would take a finished ch'in to his **teacher** Sun to try out.
林立正剛開始向孫毓芹學藝時，一把琴作好了，請**老師**試彈，音色不對便拆開修改再請**老師**試彈，如此折騰，一把琴往往得拆拆改改廿幾回。
- "We have more than 100 female **teachers** and staff in our school," says Luo Li-liang, a **teacher** at the Taipei Municipal Lung An Primary School, "but so far only two have actually asked for child-care leave."
「我們學校女**老師**和女職員加起來有一百多位，但到目前為止，只有二位請過育嬰假」，台北市龍安國小資源班**老師**羅麗鈴表示。
- Even more luckily, Wen-hsiung's **teacher** Ms Weng is also a Hakka, and half by guesswork she has slowly and patiently built up a channel of communication with Wen-

自動對應翻譯

teacher (1106)
teachers (1024)
professors (69)
instructor (58)
instructors (53)
professor (45)
students (45)

自動抽取搭配詞

老師 [N]

[M-]
位 ~,

[N-]
學校 ~, 國小 ~, 學生 ~,
國中 ~, 家長 ~, 小學 ~,
陳 ~, 美術 ~, 輔導 ~,

[P-]
跟 ~,

[Vt-]
當 ~, 擔任 ~,

[~N]
~學生, ~們, ~同學,
~家長,

國家教育研究院 10644臺北市大安區和平東路一段179號 意見信箱: mhbai@mail.naer.edu.tw 語料由外交部台灣光華雜誌提供。

圖 2-33 華英雙語索引典以「老師」一詞為例點選圖 2-32 自動抽取搭配詞中的「位」結果

國教院華英雙語索引典系統

語料庫 關鍵詞:

總共找到 3584 << 上一頁 1 下一頁

- A **Teacher**'s Lens
一位**老師**的鏡頭
- one guidance **counselor** sighs.
打回來」，一位輔導**老師**嘆道。
- On average each **faculty member** **instructs** only two pupils.
平均起來，每位**老師**只教兩名學生。
- Each new **teacher** that comes brings along some new things.
而每位新**老師**進來，總會帶來一些新東西。
- But by no means can all parents understand the **teacher**'s intentions.
但並非每位家長都能體察**老師**的用心。
- After Hsu Ch'in and three **instructors** arrived at their posts, the school's name was changed to Ta-t'ung (Universal Harmony).
徐勳及三位**老師**到任後，即將校名改為「大同」。
- One of her **professors** once even said that he had never seen her teeth!
他的一位**老師**甚至說，從來沒看到她的牙齒！
- There, under the guidance of four **instructors** by turns, they sometimes dance about, and sometimes sing at the tops of their voices.
在四位**老師**的輪流指點下，時而手舞足蹈，時而放懷高歌。
- "That's right," another **teacher** chimes in. " Sometimes you have to look a little farther down the road.
另一位**老師**也在旁補充：「的確，有時候要看長遠一點。」
- Virtually everyone **asks**: " What do I do if the egg breaks? "

自動對應翻譯

teacher (1106)
teachers (1024)
professors (69)
instructor (58)
instructors (53)
professor (45)
students (45)

自動抽取搭配詞

老師 [N]

[M-]
位 ~,

[N-]
學校 ~, 國小 ~, 學生 ~, 國中 ~,
家長 ~, 小學 ~, 陳 ~, 美術 ~,
輔導 ~,

[P-]
跟 ~,

[Vt-]
當 ~, 擔任 ~,

[~N]
~學生, ~們, ~同學, ~家長,

[~Vt]

國家教育研究院 10644臺北市大安區和平東路一段179號 意見信箱: mhbai@mail.naer.edu.tw 語料由外交部台灣光華雜誌提供。

綜合以上對於華英雙語語料庫的論述，可整理出以下三點使用上的優勢：

(一)、雙語對照例句

介面將會出現中文及英文的例句，有助於中、英文非母語者模仿學習例句搭配和句型，以便使學習者能更貼近中、英文母語者句型的使用，如圖 2-34 以英文單字 teacher 作為例子，下圖 2-35 則是以中文的「老師」作為例子。

圖 2-34 以英文單字 teacher 為例

國教院華英雙語索引系統

語料庫 光華雜誌 關鍵詞：

總共找到 3549

<< 上一頁 1 下一頁 >>

- The young Chou first revealed his native artistic talent under his father's tutelage; then during his school years two **teachers** served as his artistic mentors at different times: his junior high school art **teacher** Yang Chien-chung and his high school Chinese **teacher** Chiang Chao-shen.
在父執輩的教導下，年少時的周澄已嶄露藝術才華；其後求學過程中，引領他進入藝術殿堂的兩位**老師**，又各自代表兩個重要階段的開始，一是國中時期的美術**老師**楊乾鍾、一是高中時期的國文**老師**江兆申。
- One prestigious buxiban math **teacher** posted an advertisement in the Mandarin Daily News. One student he had taught before saw the ad overseas and gave him an international phone call. "**Teacher**," he said, "do you want to immigrate here and teach?"
一位補習界的數學**名師**在國語日報登廣告，他過去教過的學生在海外看到了，打越洋電話來向他說：「**老師**，您要不要移民到這邊教？」
- and how at the kindergarten he used Hangzhou dialect to tell his **teacher** that he wanted to go to the toilet, to his **teacher**'s mystification and to everyone's amusement.
兒子在幼稚園裡用杭州話跟**老師**說要「撒稀」（小便）、「撒尿」（大便）時，也弄得**老師**一頭霧水，鬧了不少笑話。
- The new history curriculum stipulates that everybody must study either India, Africa or China, so we hope that these teaching materials will prove to be satisfactory. After all, people will only look at an option if **teachers** have good interesting material to use and it is only through **teachers** that we can really improve this society's understanding of China. 我們希望這份教材能有滿意的結果，因為新訂必選課程中的外國史地包括了印度、非洲、中國，**教師**如果能有有用又吸引人的教材，才會選用，也只有從教育著手，才能真正改善這個社會對中國的瞭解。

自動對應翻譯

老師 (1948)
教師 (244)
師資 (121)
師生 (86)
老師們 (55)
師 (50)
師範 (45)

自動抽取搭配詞

teacher [NNS]

[JJ-]
primary ~, elementary ~,
[NNS-]
student ~, parent ~,
[NN-]
school ~,
[~NNS]
~ student, ~ parent, ~ school,
~ college,
[~VB]
~ teach,
[~VBP]
~,
teacher [NN]

國家教育研究院 10644 臺北市大安區和平東路一段179號 意見信箱: mhbai@mail.naer.edu.tw 語料由外交部台灣光華雜誌提供。

圖 2-35 以「老師」一詞為例結果

國教院華英雙語索引系統

語料庫 光華雜誌 關鍵詞：

總共找到 3584

<< 上一頁 1 下一頁 >>

- The young Chou first revealed his native artistic talent under his father's tutelage; then during his school years two **teachers** served as his artistic mentors at different times: his junior high school art **teacher** Yang Chien-chung and his high school Chinese **teacher** Chiang Chao-shen.
在父執輩的教導下，年少時的周澄已嶄露藝術才華；其後求學過程中，引領他進入藝術殿堂的兩位**老師**，又各自代表兩個重要階段的開始，一是國中時期的美術**老師**楊乾鍾、一是高中時期的國文**老師**江兆申。
- and how at the kindergarten he used Hangzhou dialect to tell his **teacher** that he wanted to go to the toilet, to his **teacher**'s mystification and to everyone's amusement.
兒子在幼稚園裡用杭州話跟**老師**說要「撒稀」（小便）、「撒尿」（大便）時，也弄得**老師**一頭霧水，鬧了不少笑話。
- The children learn about life, customs, and the adult world in general at the **school**.
幼稚園**老師**周嘉玲指出：「蘭生幼稚園每班有兩位**老師**，但我們並不算灌輸孩子許多知識；而是希望他們能有個愉快的學習過程，養成他們樂於學習的態度。」
- When he first began studying, Lin would take a finished ch'in to his **teacher** Sun to try out.
林立正剛開始向孫毓芹學藝時，一把琴作好了，請**老師**試彈，音色不對便拆開修改再請**老師**試彈，如此折騰，一把琴往往得拆拆改改廿幾回。
- "We have more than 100 female **teachers** and staff in our school," says Luo Li-ling, a **teacher** at the Taipei Municipal Lung An Primary School, "but so far only two have actually asked for child-care leave."
「我們學校女**老師**和女職員加起來有一百多位，但到目前為止，只有二位請過育嬰假。」，台北市龍安國小資深班**老師**羅麗玲表示。

自動對應翻譯

teacher (1106)
teachers (1024)
professors (69)
instructor (58)
instructors (53)
professor (45)
students (45)

自動抽取搭配詞

老師 [N]

[M-]
位 ~,
[N-]
學校 ~, 國小 ~, 學生 ~, 國中 ~,
家長 ~, 小學 ~, 陳 ~, 美術 ~,
輔導 ~,
[P-]
跟 ~,
[V-]
當 ~, 擔任 ~,
[~N]
~ 學生, ~ 們, ~ 同學, ~ 家長,
[~V]
~ 說, ~ 教,

國家教育研究院 10644 臺北市大安區和平東路一段179號 意見信箱: mhbai@mail.naer.edu.tw 語料由外交部台灣光華雜誌提供。

(二)、語詞對應翻譯

可運用在同義詞詞典編輯、語言教材編輯，有助於學習者能同時接觸同義詞及近義詞。而此欄位所提供的不同翻譯，是按照此關鍵字出現在語料庫的次數所計，也就是括號中由高到低的數字排序。(廖柏森，2019)

圖 2-36 以「老師」一詞為例觀察自動對應翻譯

國教院華英雙語索引典系統

語料庫 關鍵詞：

總共找到 3584 << 上一頁 1 下一頁 >>

1. The young Chou first revealed his native artistic talent under his father's tutelage; then during his school years two **teachers** served as his artistic mentors at different times: his junior high school art **teacher** Yang Chien-chung and his high school Chinese **teacher** Chiang Chao-shen.
在父執輩的教導下，年少時的周澄已蘊露藝術才華；其後求學過程中，引領他進入藝術殿堂的兩位**老師**，又各自代表兩個重要階段的開始，一是國中時期的美術**老師**楊乾鍾、一是高中時期的國文**老師**江兆申。

2. and how at the kindergarten he used Hangzhou dialect to tell his **teacher** that he wanted to go to the toilet, to his **teacher**'s mystification and to everyone's amusement.
兒子在幼稚園裡用杭州話跟**老師**說要「撒稀」（小便）、「撒汗」（大便）時，也弄得**老師**一頭霧水，鬧了不少笑話。

3. The children learn about life, customs, and the adult world in general at the **school**.
幼稚班**老師**周嘉玲指出：「蘭生幼稚園每班有兩位**老師**，但我們並不打算灌輸孩子許多知識；而是希望他們能有個愉快的學習過程，養成他們樂於學習的態度。」

4. When he first began studying, Lin would take a finished ch'in to his **teacher** Sun to try out.
林立正剛開始向孫毓芹學藝時，一把琴作好了，請**老師**試彈，音色不對便拆開修改再請**老師**試彈，如此折騰，一把琴往往得拆拆改改廿幾回。

5. "We have more than 100 female **teachers** and staff in our school," says Luo Li-ling, a **teacher** at the Taipei Municipal Lung An Primary School, "but so far only two have actually asked for child-care leave."
「我們學校女**老師**和女職員加起來有一百多位，但到目前為止，只有二位請過育嬰假」，台北市龍安國小資源班**老師**羅麗鈴表示。

6. Even more luckily, Wen-hsiung's **teacher** Ms Weng is also a Hakka, and half by guesswork she has slowly and patiently built up a channel of communication with Wen-

自動對應翻譯

- teacher (1106)
- teachers (1024)
- professors (69)
- instructor (58)
- instructors (53)
- professor (45)
- students (45)

自動搭配詞

老師 [N]

- [M-] 位 ~,
- [N-] 學校 ~, 國小 ~, 學生 ~, 國中 ~, 家長 ~, 小學 ~, 陳 ~, 美術 ~, 輔導 ~,
- [P-] 頭 ~,
- [Vt-] 當 ~, 擔任 ~,
- [~N] ~學生, ~們, ~同學, ~家長,

(三)、搭配詞語的功能

在關鍵字搜尋之際，右下方搭配詞的欄位也會同事出現於介面上，其詞性亦是，且點選搭配詞也能看到搭配詞的中英文例句。

圖 2-37 以「老師」一詞為例觀察自動抽取搭配詞

國教院華英雙語索引典系統

語料庫 [光華雜誌](#) 關鍵詞：

總共找到 3584 << 上一頁 1 下一頁

- The young Chou first revealed his native artistic talent under his father's tutelage; then during his school years two **teachers** served as his artistic mentors at different times: his junior high school art **teacher** Yang Chien-chung and his high school Chinese **teacher** Chiang Chao-shen.
在父執輩的教導下，年少時的周澄已嶄露藝術才華；其後求學過程中，引領他進入藝術殿堂的兩位**老師**，又各自代表兩個重要階段的開始，一是國中時期的美術**老師**楊乾鍾、一是高中時期的國文**老師**江兆申。
- and how at the kindergarten he used Hangzhou dialect to tell his **teacher** that he wanted to go to the toilet, to his **teacher**'s mystification and to everyone's amusement.
兒子在幼稚園裡用杭州話跟**老師**說要「撒稀」（小便）、「撒汗」（大便）時，也弄得**老師**一頭霧水，鬧了不少笑話。
- The children learn about life, customs, and the adult world in general at the **school**.
幼稚班**老師**周嘉玲指出：「蘭生幼稚園每班有兩位**老師**，但我們並不打算灌輸孩子許多知識；而是希望他們能有個愉快的學習過程，養成他們樂於學習的態度。」
- When he first began studying, Lin would take a finished ch'in to his **teacher** Sun to try out.
林立正剛開始向孫毓芹學藝時，一把琴作好了，請**老師**試彈，音色不對便拆開修改再請**老師**試彈，如此折騰，一把琴往往得拆拆改改廿幾回。
- "We have more than 100 female **teachers** and staff in our school," says Luo Li-ling, a **teacher** at the Taipei Municipal Lung An Primary School, "but so far only two have actually asked for child-care leave."
「我們學校女**老師**和女職員加起來有一百多位，但到目前為止，只有二位請過育嬰假」，台北市龍安國小資源班**老師**羅麗鈴表示。
- Even more luckily, Wen-hsiung's **teacher** Ms Weng is also a Hakka, and half by guesswork she has slowly and patiently built up a channel of communication with Wen-

自動對應翻譯

- teacher (1106)
- teachers (1024)
- professors (69)
- instructor (58)
- instructors (53)
- professor (45)
- students (45)

自動抽取搭配詞

老師 [N]

- [M-] 位 ~,
- [N-] 學校 ~, 國小 ~, 學生 ~, 國中 ~, 家長 ~, 小學 ~, 陳 ~, 美術 ~, 輔導 ~,
- [P-] 頭 ~,
- [Vt-] 當 ~, 擔任 ~,
- [~N] ~學生, ~們, ~同學, ~家長,

國家教育研究院 10644臺北市大安區和平東路一段179號 意見信箱: mhbai@mail.naer.edu.tw 語料由外交部台灣光華雜誌提供。

國教院華英雙語索引典系統雙語的取材方式同時支援了中文及英文，運用兩段不同語言的文章相互對應，有助於華語文學習者複習或是自習。

第三章、華語文能力基準整合應用系統

台灣華語文能力基準包括華語文能力指標及漢字、詞語及語法點分級標準，本章將分別介紹分級標準檢索系統、語義場關聯詞查詢系統、例句編輯輔助系統及作文錯別字自動批改系統。

第一節、分級標準檢索系統

華語文分級標準包括三等七級（漢字 3,100 個、詞語 14,470 個、類詞綴 73 個及 496 個語法點），數量龐大，本平臺建置「漢字分級標準檢索系統」、「詞語分級標準檢索系統」及「語法點分級標準檢索系統」，提供使用者查詢漢字、詞語及語法點的等級，漢字、詞語在書面語及口語中的使用頻率，及語法點例句，如圖 3-1。

圖 3-1 華語分級標準漢字、詞語及語法點檢索系統



以「果」字為例，運用「漢字分級標準檢索系統」搜尋，會出現以下的介面，圖 3-2。可看出此字級別為第一級，書面語的百萬字頻為 1623 次/每百萬詞，口語的百萬字頻則是 1945 次/每百萬詞。此外，為了有助於學習華語文，最末一欄點選後，則會連結至「國字標準字體筆順學習網」，詳圖 3-3。學習者可了解筆順，並運用「生字練習簿」下載字形與格數練習字體筆順。

圖 3-2 漢字分級標準檢索系統「果」字查詢結果介面



圖 3-3 國字標準字體筆順學習網，以「果」字為例查詢結果介面



因中文詞的特殊性，詞的最小單位可以為一個字，故同樣以「果」為例，查詢「詞語分級標準檢索系統」，如圖 3-4，得到含有「果」字的詞，此系統同樣也會顯示書面詞頻和口語詞頻。不過，最後一欄點選後則會連結至「華語文書面語索引典」或「華語文口語索引典」的頁面。除了單字詞之外，也可搜索多字詞，如「蘋果」和「一心一意」為例，詳如圖 3-5、圖 3-6。

圖 3-4 詞語分級標準檢索系統「果」查詢結果介面

國家教育研究院
National Academy for Educational Research
華語教學標準體系應用查詢系統

漢字分級標準檢索系統 詞語分級標準檢索系統 語法點分級標準檢索系統

果 Search

查詢結果 檢索策略 果 結果共 19 筆

筆數 10 跳至 1 /2 頁

序號	詞語	等別	級別	書面詞頻(每百萬詞)	口語詞頻(每百萬詞)	國語辭典	語料庫
1	水果	基礎	第1級	28	124		書面語 <input type="checkbox"/>
2	如果	基礎	第2級	1392	1604		書面語 <input type="checkbox"/>
3	蘋果	基礎	第2級	41	25		書面語 <input type="checkbox"/>
4	果汁	基礎	第2級	8	10		書面語 <input type="checkbox"/>
5	糖果	基礎	第3級	10	15		書面語 <input type="checkbox"/>
6	結果	基礎	第3級	484	379		書面語 <input type="checkbox"/>
7	芒果	基礎	第3級	5	8		書面語 <input type="checkbox"/>
8	後果	進階	第4級	51	23		書面語 <input type="checkbox"/>
9	成果	進階	第4級	60	52		書面語 <input type="checkbox"/>
10	效果	進階	第4級	155	124		書面語 <input type="checkbox"/>

筆數 10 跳至 1 /2 頁

圖 3-5 詞語分級標準檢索系統「蘋果」查詢結果介面

國家教育研究院
National Academy for Educational Research
華語教學標準體系應用查詢系統

漢字分級標準檢索系統 詞語分級標準檢索系統 語法點分級標準檢索系統

蘋果 Search

查詢結果 檢索策略 蘋果 結果共 1 筆

筆數 10 跳至 1 /1 頁

序號	詞語	等別	級別	書面詞頻(每百萬詞)	口語詞頻(每百萬詞)	國語辭典	語料庫
1	蘋果	基礎	第2級	41	25		書面語 <input type="checkbox"/>

筆數 10 跳至 1 /1 頁

圖 3-6 詞語分級標準檢索系統「一心一意」查詢結果介面



「語法點分級標準檢索系統」建置基本常用的語法點共有 496 個(林慶隆等, 2020)。以「在」為例, 於中文語法中, 「在」可以運用於空間、時間或表持續進行的狀態等。例如可鍵入關鍵字, 「在」可查得其不同的用法與其所在級數, 有助於教材編寫, 如圖 3-7。

圖 3-7 語法點分級標準檢索系統「在」搜尋結果介面



從圖 3-7 中可查到「在」有三種用法：第一種、「在 1」，其例句為：他在學校；第二種、「在 2/正在 V」，例句為：媽媽正在包餃子；第三種、「在 3」，例句為：他在聖誕節的時候去美國，且分別列在第一級和第二級。

第二節、語義場關聯詞查詢系統

語義場是在同一個語義系統中、共時條件下，若干個具有共同義素的義位聚合起來的聚合體。例如「炎熱」和「寒冷」屬於意思相反的負義語義場，「牛」和「馬」屬於（動物）類型語義場。語義場關聯詞查詢系統主要是在分析詞語的正、負關聯詞³，以自然語言處理技術及機器學習中的詞彙內嵌技術(Mikolov, 2013)，自動從華語文語料庫(COCT)中計算詞彙之間的相似度(柯華葳等, 2016)。

圖 3-8 語義場關聯詞查詢系統

詞彙	相似度	等級
感受	0.7342	6,高階級
覺得	0.6458	1,中階一級
感測	0.5878	6,高階級
情緒	0.5664	6,高階級
身體	0.5535	2,中階二級
心管	0.5337	4,基礎級
透樂	0.5082	7,高階級
反應	0.5058	5,中階級 6,高階級
舒服	0.4925	3,入門級
解宜	0.4856	5,中階級

詞彙的近似程度是以語料庫為基礎自動計算語義向量，所以語料庫所收錄的文本特性將影響到詞彙相似性（柯華葳等，2016），如下圖 3-9 介面有不同語料庫的選項：遠流語料、中國時報、國語日報、中央研究院平衡語料庫和 Chinese Giga word Corpus。再下方的欄位是選擇分級詞表：國教院分級詞表和華測八千詞。左下方的欄位則是「選擇詞彙等級」，包含有 unlimited 詞級、第一級到第七級，這些選項均有助於教材編輯者按級別挑選適合的用字。

³ 正關聯詞：加強語義向度；負關聯詞：減弱語義向度。

圖 3-9 語義場關聯詞查詢系統，正關聯詞「蘋果」為例查詢結果介面

國教院語義場關聯詞查詢系統

選擇語料庫

遠流語料
 中國時報
 國語日報
 平衡語料庫
 Chinese Giga Words

選擇分級詞表

國教院分級詞表
 華測八千詞

選擇詞彙等級

不限制
 查 7 級(含)以下詞
 查 6 級(含)以下詞
 查 5 級(含)以下詞
 查 4 級(含)以下詞
 查 3 級(含)以下詞
 查 2 級(含)以下詞
 查 1 級詞

查詢詞(組)語義場關聯詞

輸入

正關聯詞(組):
 蘋果

負關聯詞(組):
 請輸入負關聯詞，例如：賈伯斯

輸出詞數:

輸出

語義場關聯詞:

藍莓	0.6044	--
檸檬	0.5820	4.進階
橘子	0.5804	3.基礎
柳橙	0.5728	--
Apple	0.5664	--
梨子	0.5659	--
桃子	0.5615	5.進階
草莓	0.5603	3.基礎
柳丁	0.5554	--
葡萄	0.5551	3.基礎
Mac	0.5506	--
櫻桃	0.5392	5.進階
葡萄乾	0.5382	--
水果	0.5376	1.基礎
番茄	0.5374	3.基礎

如圖 3-9 以「蘋果」為例，鍵入正關聯詞（組）的欄位中，便能於輸出一欄中觀察語義場相關的詞彙，例如：「藍莓」、「檸檬」、「橘子」、「柳橙」……等詞，且列出詞彙等級。正關聯詞的下方也有一欄負關聯詞（組），負關聯詞最主要的用途在於排除詞彙，如圖 3-10，以「賈伯斯」作為負關聯詞的關鍵字，系統將會將負關聯詞的詞彙排除。所以鍵入正關聯詞「蘋果」（圖 3-11 左）和負關聯詞「賈伯斯」（圖 3-11 右），可發現圖 3-11 設定負關聯詞，已排除 apple、Mac 等詞彙。

圖 3-10 語義場關聯詞查詢系統，鍵入正關聯詞「蘋果」和負關聯詞「賈伯斯」的查詢結果介面

國教院語義場關聯詞查詢系統

選擇語料庫

遠流語料
 中國時報
 國語日報
 平衡語料庫
 Chinese Giga Words

選擇分級詞表

國教院分級詞表
 華測八千詞

選擇詞彙等級

不限制
 查 7 級(含)以下詞
 查 6 級(含)以下詞
 查 5 級(含)以下詞
 查 4 級(含)以下詞
 查 3 級(含)以下詞
 查 2 級(含)以下詞
 查 1 級詞

查詢詞(組)語義場關聯詞

輸入

正關聯詞(組):

負關聯詞(組):

輸出詞數:

輸出

語義場關聯詞:

梨子 0.5032	--
葡萄 0.4911	3.基礎
餅乾 0.4875	3.基礎
藍莓 0.4852	--
香蕉 0.4757	2.基礎
堅果 0.4693	--
橄欖 0.4679	--
番石榴 0.4628	--
乾果 0.4603	--
水果 0.4557	1.基礎
黑麵包 0.4522	--
草莓 0.4509	3.基礎
葡萄乾 0.4484	--
柳橙 0.4469	--
麵包 0.4385	1.基礎

圖 3-11 截取圖 3-9 和圖 3-10 之輸出語義場關聯詞比較圖

輸出

語義場關聯詞:

藍莓 0.6044	--
檸檬 0.5820	4.進階
橘子 0.5804	3.基礎
柳橙 0.5728	--
Apple 0.5664	--
梨子 0.5659	--
桃子 0.5615	5.進階
草莓 0.5603	3.基礎
柳丁 0.5554	--
葡萄 0.5551	3.基礎
Mac 0.5506	--
櫻桃 0.5392	5.進階
葡萄乾 0.5382	--
水果 0.5376	1.基礎
番茄 0.5374	3.基礎

輸出

語義場關聯詞:

梨子 0.5032	--
葡萄 0.4911	3.基礎
餅乾 0.4875	3.基礎
藍莓 0.4852	--
香蕉 0.4757	2.基礎
堅果 0.4693	--
橄欖 0.4679	--
番石榴 0.4628	--
乾果 0.4603	--
水果 0.4557	1.基礎
黑麵包 0.4522	--
草莓 0.4509	3.基礎
葡萄乾 0.4484	--
柳橙 0.4469	--
麵包 0.4385	1.基礎

國教院語義場關聯詞查詢系統藉由區分語義，適合用於編纂課文時詞彙的抽換，有助於按級別挑選適合的教學用字。

第三節、例句編輯輔助系統

例句編輯輔助系統可顯示輸入文本中所用詞語等級。詞語表包括台灣華語文能力基準的三等七級詞語表及華語八千詞介面，詳如圖 3-12。

圖 3-12 例句編輯輔助系統介面



以「突然下了一場暴雨，我的衣服都濕了。」為例，將其輸入於空白欄位當中，點選送出，右方欄位呈現的結果已斷句分詞，並於詞的下方標示詞的等級，如圖 3-13，「突然(4)下(1)了(X)一(1)場(2)暴雨(6)，我(1)的(1)衣服(1)都(1)濕(3)了(X)」，右下方還有圓餅圖(圖 3-13)，可瞭解句子所用詞語的等級。

圖 3-13 例句編輯輔助系統以「突然下了一場暴雨，我的衣服都濕了。」為例

國教院例句編輯輔助系統 例句編輯工具

例句編輯工具

自動斷詞 國教院詞表 華語八千詞

「突然下了一場暴雨，我的衣服都濕了。」

輸出詞彙分級訊息

分級詞表：國教院詞彙分級表(初稿)

「突然下了一場暴雨，我的衣服都濕了。」

突
然
下
了
一
場
暴
雨
，
我
的
衣
服
都
濕
了
。
」

4
1
X
1
2
6
1
1
1
1
3
X

詞彙等級分佈

Level	Percentage
level 1	50%
level 2	8.3%
level 3	8.3%
level 4	8.3%
level 6	8.3%
level X	16.7%

以下再以《當代中文課程3》課本的短文為例：

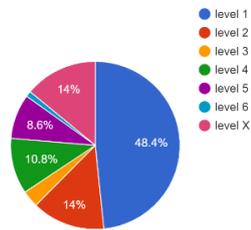
星期天一大早，何雅婷就陪著媽媽到農夫市集去買菜。最近食品安全出了好幾次問題，何雅婷的媽媽為了家人的健康，開始注意食材的產地，也盡量到有機商店和農夫市集買菜。雖然這些地方賣的東西種類沒有市場那麼多，價錢也比較高，可是媽媽常跟農夫聊天，了解小農用友善的方式對待土地和環境，很信任他們，所以寧可每個星期跑一趟，多花一點錢，也要支持小農。(鄧守信主編，王瓊淑、盧翠英、盧德昭編寫，2016，頁 134)

將其輸入欄位中，點選送出，結果如圖 3-14。

圖 3-14 例句編輯輔助系統輸入《當代中文課程 3》課本短文分析結果介面



詞彙等級分佈



從圓餅圖可瞭解此篇文章中，第一級詞條的使用頻率是最高的，佔了將近一半的數量（48.4%），次高則是第二級和未收詞，比例各為 14%。

第四節、作文錯別字自動批改系統

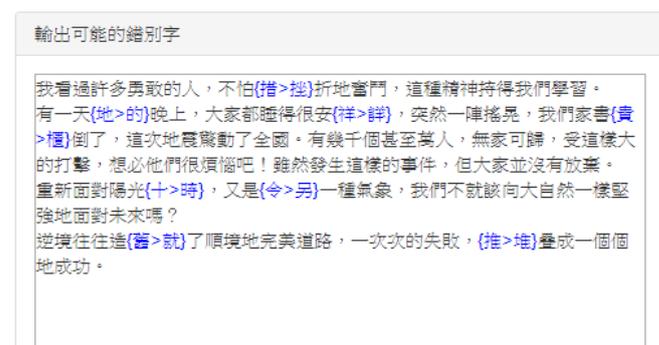
作文批改是相當耗時費力的工作，然而，語言學習者用錯別字的機率相當高，本系統以 Hsieh 等 (2015) 所提出的理論為基礎所開發。經過實際的驗證，錯別字批改的正確率可以達到 78%，為目前線上華語文錯字批改系統中正確率最高的系統 (許添明等，2017)。

圖 3-15 作文錯別字自動批改系統介面



在文章的輸入框中將待偵測錯字的文章貼上，按下送出後，詳如圖 3-16。

圖 3-16 作文錯別字自動批改系統文句修正介面



系統將文句中的錯別字一一標出並提供建議用字，除了挑出同音異字的錯字，能藉由文句上下文判斷詞彙的正確性，如圖 3-20 最後一句例句當中的「堆疊」一詞，在例句當中原本打了「推疊」，但系統仍是能將文句中要表達的「堆疊」校正出來。

第四章、綜合應用

臺灣華語文語料庫 (COCT) 及能力基準 (TBCL) 整合應用系統，已於第二、三章介紹，這些系統可以實際應用在語言教育的許多面向，例如：辭典編纂、教材撰寫，或是直接應用在課堂教學等等。本章節將以實際應用情境為例，使讀者對系統的應用面有初步的概念。

第一節、辭典編輯應用

整合應用系統在辭典編輯的應用，例如：比較字詞的常用度、比較異形詞的常用度、觀察錯別字的使用情形、辨析易混淆字等，以下均使用 2019 年的語料庫作為查詢的語料庫。

1. 字詞的常用度比較

對華語初學者來說，詞彙的學習應以常用字詞優先學習，所以，字詞常用度在語言教學中是個重要的訊息。舉例來說，以「腳踏車」與「自行車」為同義詞。藉由 COCT 書面語語料庫 2019 及 COCT 口語語料庫 2019，可看出「腳踏車」與「自行車」在不同語料庫的常用度比較。輸入 (腳踏車 | 自行車) 後，再利用 Frequency breakdown 的功能，進一步統計兩個詞的個別頻率。便會得出圖 4-1 為 COCT 書面語語料庫 2019 的查詢結果為；圖 4-2 則是 COCT 口語語料庫 2019 的。

圖 4-1 COCT 書面語語料庫 2019 與常用度比較搭配使用結果

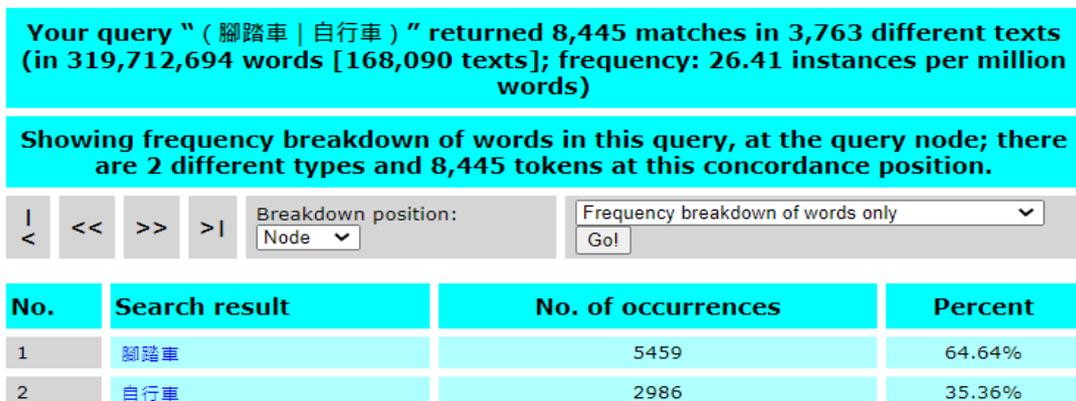
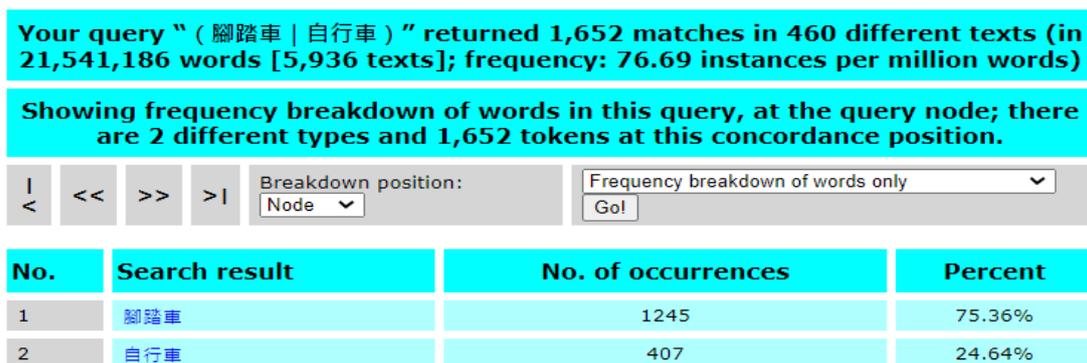


圖 4-2 COCT 口語語料庫 2019 與常用度比較搭配使用結果



從圖 4-1 和圖 4-2 發現不管是在書面語還是口語，腳踏車的使用頻率都明顯高於自行車。意即「腳踏車」不論在書面語或口語，都比「自行車」常用。

再以「桌球」與「乒乓球」這兩組詞彙為例。在 COCT 書面語語料庫 2019 中輸入（桌球 | 乒乓球）後，可以看到圖 4-3、圖 4-4 分別為 COCT 書面語語料庫 2019 及 COCT 口語語料庫 2019 的查詢結果。

圖 4-3 COCT 書面語語料庫 2019 與常用度比較搭配使用結果

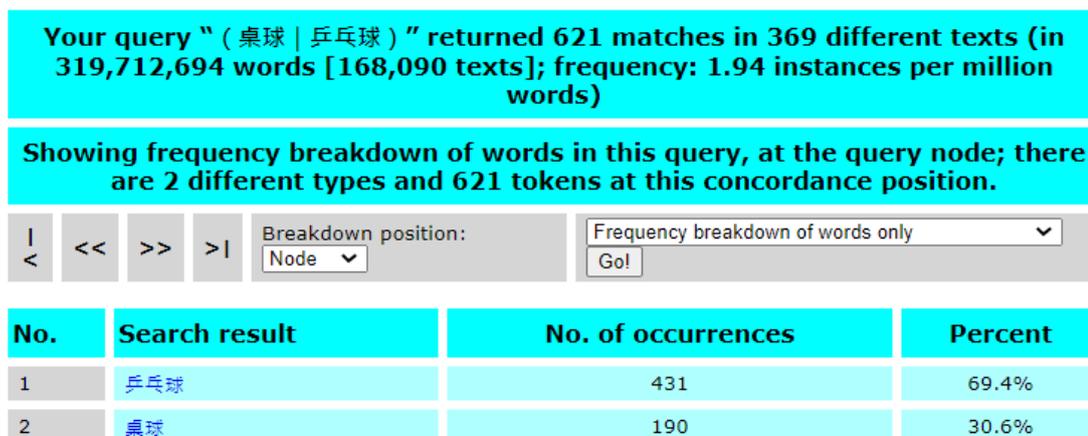
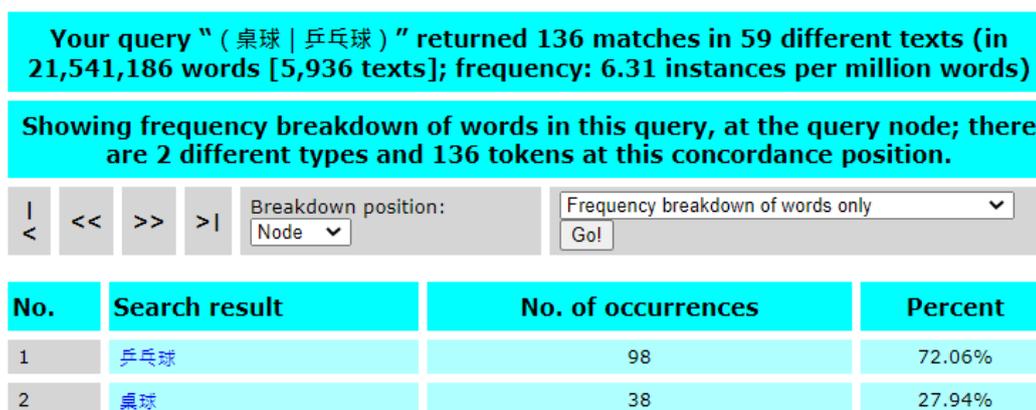


圖 4-4 COCT 口語語料庫 2019 與常用度比較搭配使用結果



從圖 4-3 和圖 4-4 可發現「桌球」與「乒乓球」是「乒乓球」的使用頻率較高。而且在口語使用中，差異比書面語還高。意即「乒乓球」不論在書面語或口語，都比「桌球」常用。

上述的字詞的常用度比較，是透過 (|) 運算及 Frequency breakdown 的使用，系統可以同時呈現次數及百分比，對於常用度的判斷十分便利。

2. 異形詞的常用度比較

異形詞是指同音、同義但書寫形式不同的詞語。辭典使用者在面對異形詞時，通常希望辭典能夠提供比較正確或是比較典型的詞形。所以，異形詞的辨析為辭典編輯的重要研究工作之一。語料庫能夠用來比較異形詞的使用頻率，藉由頻率訊息以瞭解典型的詞形。以「鈕扣」、「鈕釦」與「鈕釦」、「抹煞」與「抹殺」作為例子，以觀察這兩組當中異形詞的使用頻率。

- 輸入[鈕,鈕][釦,扣]，再利用 Frequency breakdown 的功能，進一步統計各種詞形的頻率。

圖 4-5 COCT 書面語語料庫 2019 與使用頻率搭配使用結果

No.	Search result	No. of occurrences	Percent
1	鈕釦	554	50.97%
2	鈕扣	465	42.78%
3	鈕扣	63	5.8%
4	鈕釦	5	0.46%

結果如圖 4-5 所示，在書面語中「鈕釦」和「鈕扣」是最常見的使用詞形（約佔 92%），其中又以「鈕釦」最為常見（約佔 51%）。

圖 4-6 是「抹煞」與「抹殺」書面語語料庫 2019 的檢索結果。結果如圖 4-6 所示，在書面語中「抹煞」比「抹殺」略微常見，但相差只有 10% 左右，可見這兩個詞形在現代書面語中都算常用。

● 輸入「抹[煞,殺]」

圖 4-6 COCT 書面語語料庫 2019 與使用頻率搭配使用結果

No.	Search result	No. of occurrences	Percent
1	抹煞	604	54.71%
2	抹殺	500	45.29%

上述的異形詞常用度比較，可以透過(|)運算，或是[,]運算搭配 Frequency breakdown 功能的使用來比較，和字詞常用度比較的方法差不多。不同的是，異形詞常用度不能使用口語語料來比較，因為異形詞是因為書寫時是用了不同的詞形所造成。而口語語料是從語音來的，在語音中不會有異形詞的問題，只有在轉寫成文字時才会有詞形的差異。但是口語語料的文字是由特定的語音轉寫人員進行，轉寫所造成異形詞差異反應了轉寫人員的書寫習慣，不能代表語言使用社群的分布現象。故口語語料不適用於異形詞的常用度比較。

3. 錯別字的使用觀察

書面語語料庫所蒐集的文本大多來自於書籍、報章、雜誌等出版品，但仍會有部分的錯別字。錯別字可能除了字音相同外，在字形、字義上也都十分的接近。所以，即便是經過精心校對的出版品也仍存有錯別字。例如「口乾舌躁（燥）」及「心無旁鶩（鶩）」，在本院的書面語語料庫或是中研院的平衡語料庫都可以發現這些錯別字。藉由書面語語料庫查詢，我們可以驗證錯別字在日常生活中被誤用的情況。

此部分將以「口乾舌燥／躁／噪」為例，藉由運用萬用字元檢索書面語語料庫，有助於觀察出使用的情況。為了要觀察「口乾舌燥／躁／噪」錯別字的使用，輸入「口乾舌*」後再點選 frequency breakdown，結果如圖 4-7 所示，出現六種結果：「口乾舌燥」、「口乾舌躁」、「口乾舌渴」、「口乾舌」、「口乾舌擗」、「口乾舌啞」，撇除後面四種和「口乾舌燥／躁／噪」沒有相關聯的詞彙，可以看出「口乾舌燥」和「口乾舌躁」這兩種用法最多。其中「口乾舌燥」應為正確的使用字

詞(請參考教育部重編國語辭典修訂本),「口乾舌燥」則為錯別字,比例約 2.5%。以出版品來說 2.5%的錯誤率算是非常高的比例,若是能將錯誤率高的詞收錄在詞典中特別說明,或是在課堂教學中特別說明辨析,就可以減少錯別字的發生率。

圖 4-7 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果

No.	Search result	No. of occurrences	Percent
1	口乾舌燥	371	93.69%
2	口乾舌燥	10	2.53%
3	口乾舌燥	10	2.53%
4	口乾舌	3	0.76%
5	口乾舌燥	1	0.25%
6	口乾舌燥	1	0.25%

4. 易混淆字的辨析

因中國字的特殊性,所以會有相似字使用上的困擾,相似字不僅同音,有些字義也是相當接近,例如以下將以「跡」和「蹟」為例。

根據教育部國語辭典簡編本的定義,「跡」應解為步行時所遺留的印痕或事物的遺痕(教育部);「蹟」的意思則為前人所留下來的事物(教育部)。將這兩個字分別放入語料庫觀察,便會得出以下圖 4-10、圖 4-11。

圖 4-8 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果

痕跡	事跡
軌跡	發跡
足跡	絕跡
血跡	人跡
遺跡	形跡
蹤跡	真跡
字跡	行跡
筆跡	墨跡

圖 4-9 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果

奇蹟	真蹟
事蹟	殘蹟
古蹟	功蹟
神蹟	勝蹟
遺蹟	戰蹟
史蹟	秘蹟
聖蹟	奧蹟
墨蹟	名蹟

在瞭解這兩個字之間的定義差異之後，再觀察語料庫如圖 4-8、圖 4-9，就可以輕易地分別出這兩個字之間的差異。深入了解亦混淆字後，便能有助於老師掌控學生學習漢字時可能出現的困難。

第二節、近義詞的查詢與辨析

近義詞教學是華語詞彙教學相當重要的部分，華語學習者在累積一定的詞彙量後，便會遭遇到近義詞選擇的問題。例如「感覺」、「感受」及「覺得」這三個詞在意義的非常接近，華語學習者不容易掌握到近義詞的細微差別及使用時機。所以在教學上，必須對近義詞做適當的歸納與辨析。「國教院語義場關聯詞查詢系統」可以從大量的華語文語料庫中自動提取語近義詞，以做為華語教學、教材編輯及近義詞詞典編輯的輔助工具。

先設定「國教院語義場關聯詞查詢系統」的語料庫、分級詞表及詞彙等級，再輸入關鍵詞「感覺」至正關聯詞的欄位當中，點選送出後，便會看到以下圖 4-10。

圖 4-10 國教院語義場關聯詞查詢系統以「感覺」一詞為例

國教院語義場關聯詞查詢系統

The screenshot shows the 'National Education Research Institute Semantic Field Association Word Query System' interface. It is divided into three main sections: Selection, Input, and Output.

- 選擇語料庫 (Select Corpus):** Includes options for '遠流語料' (selected), '中國時報', '國語日報', '平衡語料庫', and 'Chinese Giga Words'.
- 選擇分級詞表 (Select Grading Table):** Includes options for '國教院分級詞表' (selected) and '華測八千詞'.
- 選擇詞彙等級 (Select Vocabulary Level):** Includes options for '不限制' (selected), '查 7 級(含)以下詞', '查 6 級(含)以下詞', '查 5 級(含)以下詞', '查 4 級(含)以下詞', '查 3 級(含)以下詞', '查 2 級(含)以下詞', and '查 1 級詞'.
- 查詢詞(組)語義場關聯詞 (Query Word/Semantic Field Association Words):** The '輸入' (Input) section has a '正關聯詞(組):' (Positive Association Word/Group) field containing '感覺' and an '負關聯詞(組):' (Negative Association Word/Group) field with the instruction '請輸入負關聯詞，例如：賈伯斯' (Please input negative association words, e.g., Jobs).
- 輸出 (Output):** The '輸出' section shows the results under the heading '語義場關聯詞:' (Semantic Field Association Words). The results are as follows:

Word	Score	Level
感受	0.7342	5.進階
感覺到	0.6793	--
覺得	0.6458	1.基礎
感到	0.5878	4.進階
情緒	0.5664	5.進階
身體	0.5535	1.基礎
感受到	0.5418	--
感覺	0.5405	7.精熟
心情	0.5337	3.基礎
熟悉感	0.5322	--

從圖 4-10 當中的輸出，即可看出與「感覺」有相關聯的一些詞彙，如：感受、感覺到、覺得、感到……等詞彙。同時列出了這些詞在華語教學上的難度分級，可以提供華語教師在近義詞教學上的參考。近義詞教學的另一個重點在於近義詞的辨析，和關鍵字「感覺」（動詞）相比較，學生必須瞭解這些近義詞使用上的差異，故僅選出了動詞詞性的詞彙。經過挑選之後，則以「感受」和「覺得」這兩個詞的辨析為例來說明。

首先，我們分別將「感覺」、「感受」和「覺得」這三個詞輸入索引典查詢(COCT 書面語語料庫 2019)，並取得三組詞彙的搭配詞，如圖 4-11 所示。

圖 4-11 「感覺」、「感受」和「覺得」搭配詞

感覺	感受	覺得
愉快	深刻	不安
遲鈍	安寧	有趣
舒服	強烈	自在
自在	敏銳	無助
敏銳	正向	舒服
不安	深	寂寞
無比	寧靜	無比
美好	自自然然	沮喪
翩翩	充分	不已
溫馨	用心	無聊
難受	悲傷	好笑
容易	自然	好玩
痛	實際	無力
舒適	爽	容易
平靜	喜悅	可笑
強烈	簡約	乏味

圖 4-11 是三組詞彙的搭配詞，之後再運用「詞表比較工具」，以兩兩相比較的方式，觀察出這三組詞彙之間的差異，如下表 1，而表 2 則是採用三組詞彙的比較方式，較能夠看出三者近義詞的相似度。

表 1 「感覺」、「感受」和「覺得」兩兩比對

感覺	感受
愉快	深刻
遲鈍	安寧
舒服	正向
自在	深
不安	寧靜
無比	充分

感受	覺得
深刻	不安
安寧	有趣
強烈	自在
敏銳	無助
正向	舒服
深	寂寞

覺得	感覺
有趣	愉快
無助	遲鈍
寂寞	敏銳
沮喪	美好
不已	翩翩
無聊	溫馨

從表 1 的搭配詞對比中，我們可以區別出近義詞的細微差異。例如：「強烈」可以做為「感受」的狀語，但不能做為「覺得」的狀語。且表 1 可以進一步整理成表 2。

表 2 「感覺」、「感受」和「覺得」比對

	感覺	感受	覺得
愉快	V	X	X
遲鈍	V	X	X
舒服	V	X	V
自在	V	X	V
敏銳	V	X	X
不安	V	X	V
深刻	X	V	X
安寧	X	V	X
強烈	X	V	X
正向	X	V	X
深	X	V	X
寧靜	X	V	X
有趣	X	X	V
無助	X	X	V
寂寞	X	X	V
沮喪	X	X	V
無比	X	X	V

從表 2 當中可以明確地看出相似性，例如：「感覺」及「覺得」的共同搭配詞有「舒服」、「自在」和「不安」，而「感受」一詞的搭配詞和「感覺」、「覺得」有明顯的差別，故由此可知在使用上「感覺」和「覺得」比較接近。

第三節、教材與課堂教學

由於語料庫是從真實語料取樣，所以在教材的輔助編輯上，可以更貼近生活。下面將舉幾個例子以說明語料庫在教材及課堂教學的應用。

1. 量詞和名詞的搭配

量詞一直以來就是中國文法獨有的特色，這也是許多想學中文他國學生最困擾之處。但藉由語料庫檢索，便能夠輕易地觀察量詞與名詞之間的搭配，以下以「匹」為例。

圖 4-12 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果

Collocation controls						
Collocation based on:	Word form	Statistic:	Log-likelihood			
Collocation window from:	1 to the Left	Collocation window to:	3 to the Right			
Freq(node, collocate) at least:	5	Freq(collocate) at least:	5			
Filter results by:	specific collocate:	and/or tag:	Na Na	Submit changed parameters	Go!	
Extra information: Log-likelihood scores collocations by significance: the higher the score, the more evidence you have that the association is not due to chance. More frequent words tend to get higher log-likelihood scores, because there is more evidence for such words.						
There are 12,698 different words in your collocation database for "[word="匹"]". (Your query "[word="匹]" returned 10,062 matches in 5,516 different texts) <small>(1.463 seconds - retrieved from cache)</small>						
No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed collocate frequency	In no. of texts	Log-likelihood
1	馬	49,895	6.281	3,744	2287	41021.283
2	駿馬	1,290	0.162	296	261	3929.005
3	強	11,042	1.39	232	129	1919.411
4	馬馬	501	0.063	124	110	1666.731
5	馬力	974	0.123	111	69	1303.138
6	戰馬	871	0.11	97	89	1134.068
7	種	72	0.009	64	9	1098.819
8	布	9,262	1.166	129	107	960.719
9	樣子	1,124	0.142	82	61	885.959
10	小馬	766	0.096	72	66	815.985
11	綢	1,107	0.139	70	51	735.552
12	賣馬	438	0.055	58	33	699.473

藉由 CQP 語法，輸入[word="匹"]，搭配 Collocation 搭配詞的進階搜索，即可得出圖 4-12 的搜尋結果，如使用頻率最高的詞為「馬」，次高則是「駿馬」。從上圖 4-12 當中，即可觀察出「匹」大多是作為「馬」的量詞，但到序號第 8 點之後，開始出現根「布」有相關的詞彙，如序號第 11 的「綢」等。故由此可知，「匹」的最常用搭配詞為「馬」，次之則為布品。

2. A-not-A、不 X 不 Y 詞群的觀察

基礎語法典（第一級）當中常用的 A-not-A 適用動詞及不 X 不 Y 的搭配使用，均可藉由語料庫的觀察，找出其常用詞彙，如圖 4-13、圖 4-14。

圖 4-13 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果

No.	Search result	No. of occurrences	Percent
1	是不是	37142	60.62%
2	對不對	5514	9%
3	好不好	2734	4.46%
4	，不，	2694	4.4%
5	想不想	1135	1.85%
6	夠不夠	546	0.89%
7	要不要	531	0.87%
8	值不值	480	0.78%
9	敢不敢	434	0.71%
10	算不算	431	0.7%

從圖 4-13 當中，可以看到在 COCT 書面語語料庫 2019 當中，最常使用的 A-not-A 是「是不是」、「對不對」、「好不好」等，有助於老師教學上的使用。

倘若想要觀察「不 X 不 Y」也是同樣的作法，將語法[word="不.不." & char(word, 1)!=char(word, 3)]輸入，便會得到圖 4-14 的結果，亦可從當中觀察到常用字詞。

圖 4-14 COCT 書面語語料庫 2019 與再運用 Frequency breakdown 搭配使用結果

Your query "[word="不.不." & char(word, 1)!=char(word, 3)]" returned 13,216 matches in 11,328 different texts (in 319,712,694 words [168,090 texts]; frequency: 41.34 instances per million words)

Showing frequency breakdown of words in this query, at the query node; there are 268 different types and 13,216 tokens at this concordance position.

|< << >> >| Breakdown position: Node | Frequency breakdown of words only | Go |

No.	Search result	No. of occurrences	Percent
1	不知不覺	5694	43.08%
2	不折不扣	1144	8.66%
3	不慌不忙	1127	8.53%
4	不慌不擾	702	5.31%
5	不痛不癢	624	4.72%
6	不絕不休	348	2.63%
7	不明不白	347	2.63%
8	不聲不響	328	2.48%
9	不聲不響	304	2.3%
10	不聲不亮	280	2.12%

3. 教材難度的觀察

教材及課堂教學最重要的一部分莫過於要符合學生程度，如同上述所言，本院已建置三等七級能力指標，也有相對應的系統，有助於老師安排教材的詞彙等級使用。例如本院建置的「例句編輯輔助系統」，使用上簡易且一目了然，詳如圖 4-15、圖 4-16。

圖 4-15 例句編輯輔助系統

詞彙等級分佈

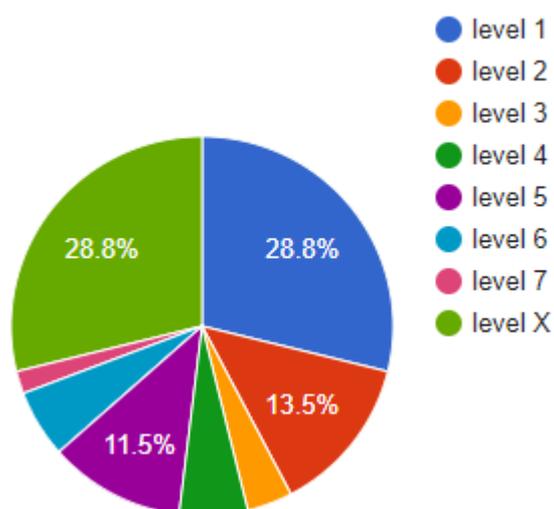


圖 4-16 例句編輯輔助系統以重修本「高興」釋義為例

輸出詞彙分級訊息

分級詞表：國教院詞彙分級表(初稿)

1 興趣甚高。《文選·殷仲文·南州桓公九井作詩》：「
 3 6 1 X X X X 1 5 7
 獨有清秋日，能使高興盡。」唐·杜甫〈北征〉詩：
 X 1 4 X 1 5 2 5 X X X 5
 「青雲動高興，幽事亦可悅。」
 X 2 X 6 3 X

2 歡喜、愉快。如：「今天玩得高興嗎？」
 5 4 X 1 1 1,1,7 2 1

3 以興致作決定事情的根據。如：「他高興，就讓他先
 5 6 X 2 2 1 4 X 1 2 1 2 1 1
 走吧！」
 1 1

圖 4-15 及圖 4-16 是採用教育部重編國語辭典修訂本「高興」釋義(教育部)，再藉由「例句編輯輔助系統」分級，有助於使用者觀察其詞彙的分級。

4. 資料驅動學習

資料驅動學習是一種學習方法，除了藉由傳統的課本及授課方式學習新知識外，學習者可以藉由大數據的使用，分析數據及資料，有效的運用到學習上，以華語文學習者運用 COCT 書面語語料庫 2019 為例，站在華語文學習者的角度，觀察語料庫中的資料，進而了解使用頻率及搭配使用，以下將以「不但……，反而……」為例。

首先先用「語法點分級標準檢索系統」確認「不但……，反而……」的等級。從圖 4-17 當中可看出，此語法點為進階第五級，且還能得知「不但」一詞後面需接否定句。

圖 4-17 語法點分級標準檢索系統以「不但……，反而……」為例

序號	語法點	等別	級別	例句
1	不但……而且……	基礎	第3級	這個故事不但容易了解，而且很有意思。
2	不但……還/也……	基礎	第3級	在超商不但可以買東西還可以寄東西。 林先生不但是一個好老師，也是一位好父親。
3	可2	進階	第4級	他不但可理解老師上課的內容，還能說出重點。
4	除此之外	進階	第5級	他不但是個成功的企業家，也是個學者，除此之外，更是一個好父親。
5	從……的角度來+V	進階	第5級	從使用者的角度來設計商品，不但實用也能增加購買的人數。
6	此外	進階	第5級	今天加入會員不但享有折扣，還可以免費體驗。此外，生日當天還有贈品。
7	(不但)不/沒……反而……	進階	第5級	雖然爬山爬了一整天，但他們不覺得累，反而精神很好。 不小心打破了花瓶，爸爸不但沒罵我，反而安慰我。

選擇「COCT 書面語語料庫 2019」後，再點選 CQP 語法輸入[word="不但"][]{}{1,5}[word="反而"]，便會出現以下的畫面，如下圖 4-18。

圖 4-18 COCT 書面語語料庫 2019 以「不但……，反而……」為例

Your query "[word="不但"][[{1,5}][word="反而"]" returned 1,879 matches in 1,777 different texts (in 319,712,694 words [168,090 texts]; frequency: 5.88 instances per million words) [1.301 seconds]

No	Filename	Solution 1 to 50	Page 1 / 38
1	576a5609326d1b3fb77fc1e9	的。因此，「門當戶對」的「婚姻觀」，	不但沒有落伍，反而因時代的不同而更複雜了。籍貫
2	576a5609326d1b3fb77fc1f1	別人不認識真正的人才。對妻子的辛苦工作	不但感激，反而譏諷說世人只知道用她這樣的
3	576a5609326d1b3fb77fc1f3	父親「救」姐姐和妹妹。但是每次他	不但失望，反而還招來父親的責罵。初中時他才
4	576a5609326d1b3fb77fc1f8	陰沉了。再加上父母對他讀書成績好，	不但沒有一絲欣慰，反而在每次要錢交註冊費的時候，
5	576a5609326d1b3fb77fc1fb	，把他們兄弟當親生孩子一樣疼愛。所以他	不但對後母沒有成見，反而盼望家裡有個母親以後，父親會
6	576a5609326d1b3fb77fc1fb	，什麼事都要靠他一個人做，	不但不能做內助，反而是個累贅。她很想再學那些不
7	576a5609326d1b3fb77fc201	他當然不大懂得怎樣的表達愛意，所以她	不但怪他，反而慶幸自己得到一個忠厚的男人。
8	576a5609326d1b3fb77fc207	可以幫幫忙。但她們又忙著參加各式各樣的活動，	不但幫忙做家事，反而很少在家了。她要是責備她們，
9	576a5609326d1b3fb77fc20d	跟女兒坐下桌)，他可以盡量發展自己，	不但受阻礙，反而會得到他能得到的一切協助。支
10	576a5609326d1b3fb77fc268	能很快指出進行中缺乏效果的團體治療，	不但無法幫助病人反而會帶來傷害；他提供刺激與動力
11	576a5609326d1b3fb77fc288	時，不要立下結論，以為孩子不用功，這樣	不但無補於事，反而增加孩子的困擾，我們大人必須
12	576a5609326d1b3fb77fc2d7	年輕人了。他們說：現在的年輕人看到我們，	不但打招呼，反而瞪我們一眼。你說你跟他打
13	576a5609326d1b3fb77fc2e0	工作卻都是他們所感興趣的事物，難怪他們	不但感到寂寞，反而還樂此不疲呢！所以，當你成著
14	576a5609326d1b3fb77fc2f3	的人經常覺得缺少友伴，沒有親近的人，	不但不被了解，反而被遺忘，沒人可以幫助他，沒

觀察語料庫後，可觀察出「不但……，反而……」的使用語境，即可運用於學生的練習測驗當中，例如從語料庫中（圖 4-18）選擇幾條例句，並刪除例句的上下文，再請學生創作出適合例句的情境（同時也已經藉由「國教院例句編輯輔助系統」確認過例句當中所使用字詞的程度等級，如圖 4-19），例如：

- _____，不但不被了解，反而被遺忘。
- _____，不但不打招呼，反而瞪我們一眼。
- _____，不但受阻礙，反而會得到他的協助。

圖 4-19 例句編輯輔助系統以確認用詞之等級

國教院例句編輯輔助系統 例句編輯工具

例句編輯工具

自動斷詞 國教院詞表 華語八千詞

不但不被了解，反而被遺忘。
不但不打招呼，反而瞪我們一眼。
不但受阻礙，反而會得到他的協助。

送出

輸出詞彙分級訊息

分級詞表：國教院詞彙分級表(初稿)

不但不被了解，反而被遺忘。
 2 1 2 2 5 2 5
 不但不打招呼，反而瞪我們一眼。
 2 1 4 5 5 1 1 3
 不但受阻礙，反而會得到他的協助。
 2 1 4 5 5 1 3 1 1 4

詞彙等級分佈

Level	Percentage
level 1	32%
level 2	24%
level 3	12%
level 4	8%
level 5	24%

本計畫建置華語文語料庫語標準體系整合應用系統，分別有助於不同層面的使用，如同本章節當中的運用，可用於老師的教材編纂、習題使用；同時也可以用於學習者端，例如雙語的翻譯、觀察搭配詞所使用的頻率等等，是一個相當多元的應用系統。

第五章、結語

華語文語料庫與能力基準整合應用系統應用科技，將臺灣華語文語料庫（COCT）及臺灣華語文能力基準（TBCL）整合，並提供華語文教學、學習、教材編撰及測驗評量，更方便、更有系統的應用。將科技導入華語文教育，提升效能。

本整合系統各子系統的功能摘述如下：

- （一）查詢華語文漢字、詞語及語法點的等級，可分別使用「漢字分級標準檢索系統」、「詞語分級標準檢索系統」、「語法點分級標準檢索系統」。
- （二）分析詞語及語法點的使用頻率、可能義項及前後搭配詞，可應用「國教院索引典系統」。
- （三）分析華語非母語人士使用詞語、語法點及前後搭配詞的可能偏誤，可應用「華語中介語索引典系統」。
- （四）分析詞語的近義詞及反義詞，可應用「語義場關聯詞查詢系統」。
- （五）分析例句或文章所用詞語的等級，可應用「例句編輯輔助系統」。
- （六）分析作文錯別字並建議正確用字，可應用「作文錯別字自動批改系統」。
- （七）分析華英雙語的雙向翻譯、使用頻率、可能義項及華英雙語前後搭配詞，可應用「華英雙語索引典系統」。

本整合系統 2010 年 1 月至 2010 年 10 月的使用人次達 130,635，點擊次數達 2,308,341。使用者除了臺灣網域約佔 89% 外，其餘來自包括美國、大陸、香港……等世界各國，而且與 2019 年比較，使用人次及點擊次數都達一倍以上，顯示華語文教育越來越結合科技應用。

誌謝

本書為華語文八年計畫「建置應用語料庫及標準體系」的部分成果，該計畫由教育部委由國家教育研究院執行。感謝教育部的經費補助及感謝研究團隊，109年團隊成員如下：

主持人

許添明、林慶隆、柯華葳、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍

本院研究人員

吳鑑城、白明弘、李詩敏、吳欣儒、許愷容

計畫參與人員

丁彥平、王冠孺、盧昱勳、陳沛璇、蔡岳璋

參考文獻

- 柯華葳、林慶隆、張俊盛、陳浩然、高照明、信世昌、方麗娜、蔡雅薰、張郁雯、范信賢 (2013)。《華語文八年計畫「建置應用語料庫及標準體系」102 年工作計畫【期末報告】》。國家教育研究院。
- 柯華葳、林慶隆、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍 (2015)。《華語文八年計畫「建置應用語料庫及標準體系」104 年工作計畫【期末報告】》。國家教育研究院。
- 柯華葳、林慶隆、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍 (2016)。《華語文八年計畫「建置應用語料庫及標準體系」105 年工作計畫【期末報告】》。國家教育研究院。
- 柯華葳、林慶隆、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍、范信賢 (2014)。《華語文八年計畫「建置應用語料庫及標準體系」103 年工作計畫【期末報告】》。國家教育研究院。
- 國家教育研究院。2003 年 6 月。雙語詞彙、學術名詞暨辭書資訊網。
<http://terms.naer.edu.tw/detail/1275315/>
- 教育部。國語辭典簡編本。<http://dict.concised.moe.edu.tw/cgi-bin/jbdic/gswweb.cgi?o=djbdic&searchid=W00000021268>
- 教育部。國語辭典簡編本。<http://dict.concised.moe.edu.tw/cgi-bin/jbdic/gswweb.cgi?o=djbdic&searchid=W00000021315>
- 教育部。教育部重編國語辭典修訂本。<http://dict.revised.moe.edu.tw/cgi-bin/cbdc/gswweb.cgi?o=dcbdic&searchid=Z00000068945>
- 許添明、林慶隆、柯華葳、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍 (2019)。《華語文八年計畫「建置應用語料庫及標準體系」108 年工作計畫【期末報告】》。國家教育研究院。
- 許添明、林慶隆、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍 (2017)。《華語文八年計畫「建置應用語料庫及標準體系」106 年工作計畫【期末報告】》。國家教育研究院。
- 郭工賓、林慶隆、柯華葳、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍 (2018)。《華語文八年計畫「建置應用語料庫及標準體系」107 年工作計畫【期末報告】》。國家教育研究院。
- 廖柏森 (2019)。《英文系最受歡迎的線上語料庫：用 20 個網站學習英文句型與詞彙搭配》。眾文圖書股份有限公司。
- 鄧守信主編，王瓊淑、盧翠英、盧德昭編寫 (2016)。《當代中文課程 3》。聯經出版社。
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

附件

表六、中研院平衡語料庫詞類標記集

精簡詞類	簡化標記	對應的CKIP詞類標記 ¹	
A	A	A	/*非謂形容詞*/
C	Caa	Caa	/*對等連接詞，如：和、跟*/
POST	Cab	Cab	/*連接詞，如：等等*/
POST	Cba	Cbab	/*連接詞，如：的話*/
C	Cbb	Cbaa, Cbba, Cbbb, Cbca, Cbcb	/*關聯連接詞*/
ADV	Da	Daa	/*數量副詞*/
ADV	Dfa	Dfa	/*動詞前程度副詞*/
ADV	Dfb	Dfb	/*動詞後程度副詞*/
ASP	Di	Di	/*時態標記*/
ADV	Dk	Dk	/*句副詞*/
ADV	D	Dab, Dbaa, Dbab, Dbb, Dbc, Dc, Dd, Dg, Dh, Dj	/*副詞*/
N	Na	Naa, Nab, Nac, Nad, Naea, Naeb	/*普通名詞*/
N	Nb	Nba, Nbc	/*專有名稱*/
N	Nc	Nca, Ncb, Ncc, Nce	/*地方詞*/
N	Ncd	Ncda, Ncdb	/*位置詞*/
N	Nd	Ndaa, Ndab, Ndc, Ndd	/*時間詞*/
DET	Neu	Neu	/*數詞定詞*/
DET	Nes	Nes	/*特指定詞*/
DET	Nep	Nep	/*指代定詞*/
DET	Neqa	Neqa	/*數量定詞*/
POST	Neqb	Neqb	/*後置數量定詞*/
M	Nf	Nfa, Nfb, Nfc, Nfd, Nfe, Nfg, Nfh, Nfi	/*量詞*/
POST	Ng	Ng	/*後置詞*/
N	Nh	Nhaa, Nhab, Nhac, Nhb, Nhc	/*代名詞*/
Nv	Nv	Nv1, Nv2, Nv3, Nv4	/*名物化動詞*/
T	I	I	/*感嘆詞*/
P	P	P*	/*介詞*/
T	T	Ta, Tb, Tc, Td	/*語助詞*/
Vi	VA	VA11, 12, 13, VA3, VA4	/*動作不及物動詞*/
Vt	VAC	VA2	/*動作使動動詞*/
Vi	VB	VB11, 12, VB2	/*動作類及物動詞*/
Vt	VC	VC2, VC31, 32, 33	/*動作及物動詞*/
Vt	VCL	VC1	/*動作接地方賓語動詞*/
Vt	VD	VD1, VD2	/*雙賓動詞*/
Vt	VE	VE11, VE12, VE2	/*動作句賓動詞*/
Vt	VF	VF1, VF2	/*動作謂賓動詞*/
Vt	VG	VG1, VG2	/*分類動詞*/
Vi	VH	VH11, 12, 13, 14, 15, 17, VH21	/*狀態不及物動詞*/
Vt	VHC	VH16, VH22	/*狀態使動動詞*/
Vi	VI	VI1, 2, 3	/*狀態類及物動詞*/
Vt	VJ	VJ1, 2, 3	/*狀態及物動詞*/
Vt	VK	VK1, 2	/*狀態句賓動詞*/
Vt	VL	VL1, 2, 3, 4	/*狀態謂賓動詞*/
Vt	V_2	V_2	/*有*/
T	DE	/*的、之、得、地*/	
Vt	SHI	/*是*/	
FW	FW	/*外文標記*/	
COLONCATEGORY			/*冒號*/

¹ 斜體詞類，表示在技術報告#93-05中沒有定義，即後來增列的。

COMMACATEGORY
DASHCATEGORY
ETCCATEGORY
EXCLAMATIONCATEGORY
PARENTHESISCATEGORY
PAUSECATEGORY
PERIODCATEGORY
QUESTIONCATEGORY
SEMICOLONCATEGORY
SPCHANGECATEGORY

/* 逗號 */
/* 破折號 */
/* 刪節號 */
/* 驚嘆號 */
/* 括弧 */
/* 頓號 */
/* 句號 */
/* 問號 */
/* 分號 */
/* 雙直線 */

跨域趨勢——臺灣華語文語料庫與能力基準整合應用系統

書名：跨域趨勢——臺灣華語文語料庫與能力基準整合應用系統

發行人：許添明

主編：國家教育研究院

編輯小組：林慶隆、陳沛璇、白明弘（通訊作者）、吳鑑城、李詩敏、吳欣儒、
許愷容、丁彥平、王冠孺、盧昱勳

出版機關：國家教育研究院

地址：237201 新北市三峽區三樹路 2 號

電話：(02) 7740-7890

傳真：(02) 7740-7064

網址：<https://www.naer.edu.tw>

出版日期：110 年 1 月

GPN：4711000003

ISBN：9789865461591

電子書設計製作

設計製作：國家教育研究院語文教育及編譯中心

地址：106011 臺北市大安區和平東路一段 179 號

電話：(02) 7740-7890

網址：<https://www.naer.edu.tw>

電子書播放資訊

作業系統：PC/iOS/Android

檔案格式：PDF

檔案內容：電子書

播放軟體：HyRead Library

本書著作財產權為國家教育研究院所有，欲使用本書內容，須徵求同意或書面授權。

