

# 臺灣華語文語料庫

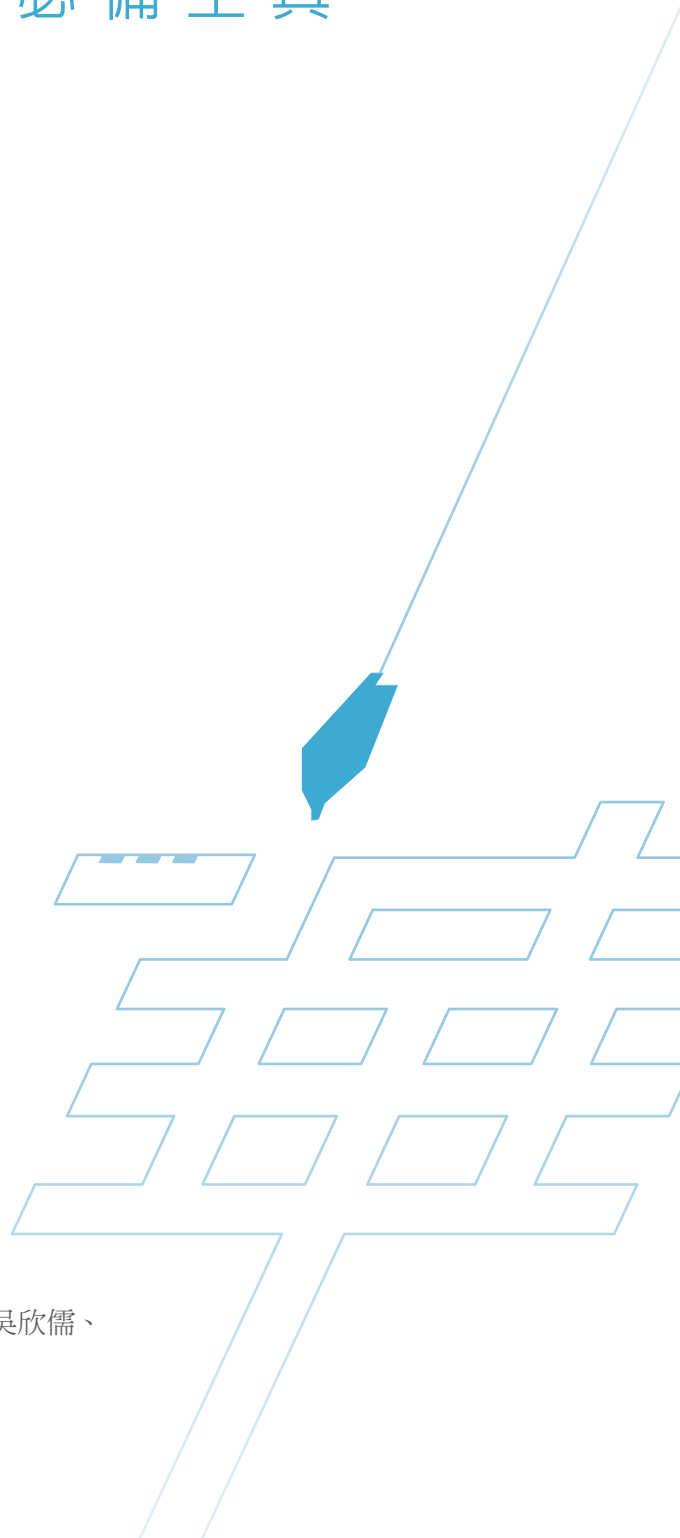
華語文教與學的必備工具





# 臺灣華語文語料庫

華語文教與學的必備工具



編輯小組

林慶隆、白明弘、蔡岳璋、吳鑑城、李詩敏、吳欣儒、  
丁彥平、許愷容、盧昱勳、王冠孺、陳沛璇

# 臺灣華語文語料庫

華語文教與學的必備工具



# 序

本書內容涵蓋臺灣華語文語料庫（英文名稱為 Corpus of Contemporary Taiwanese Mandarin，簡稱 COCT）中的正體字書面語、口語、中介語和華英雙語語料庫之研發建置，及應用書面語和口語語料庫所建置的通用詞頻表。

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（以下稱本院）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，及辦理相關的應用及推廣工作。華語文漢字、詞語及語法點分級標準的研訂若能植基於語文實際使用情境，則更能符合華語文教學需要，因此必須有適合的語料庫可用。然而，臺灣 2013 年當時較具代表性的語料庫為「中央研究院漢語平衡語料庫」。不過，該語料庫以提供語言學研究為主，且語料僅維護至 2007 年。至於國內其他單位建置的語料庫，也以學術研究為主，難以提供建置華語文標準體系所用。因此，本院諮詢專家學者，以「美國當代英語語料庫」（COCA）為目標，進行臺灣華語文語料庫的建置。

至於通用詞頻表，中研院詞庫小組曾於 1998 年研訂中研院詞頻辭典，內容包括中文核心詞彙及通用詞彙，是最早有詞類標記的詞頻表，然而已 20 年，亟需應用臺灣華語文語料庫中最新且大量的語料研訂新的通用詞頻表。

臺灣華語文語料庫是繼「中央研究院漢語平衡語料庫」之後，目前華語文正體字建置類別最完備、語料最新、數量最多的語料庫。通用詞頻表則是繼中研院詞頻辭典的通用詞彙後，最新的詞頻表。

# 臺灣華語文語料庫

華語文教與學的必備工具

本書內容主要摘自計畫的歷年成果報告，這些報告歷經多次語料庫專家諮詢會議及研究人員討論，且各類語料的取得、後設資料及語料庫建置，皆經嚴謹檢查程序，其後又召開多次會議討論才完成報告。本書以這些報告為基礎，再經研究團隊多次討論修訂才定稿。本書成果除了可應用於建置華語文標準體系，更可應用在華語文、國語文、英語文等教學設計、教材編撰、測驗評量、辭典編撰、人工智慧及相關語文研究，提升語文教育的效能。

本書能順利完成，感謝「建置應用語料庫及標準體系」計畫歷年主持人、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，以及編撰小組同仁的編撰校對，策略溝通辦公室的封面和版式設計審核，張華芸小姐的文字校對。也感謝尙曄文化事業股份有限公司的美編設計及排版印刷。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11月

# 目 錄

第一章	前言	1
第二章	語料庫建置技術	3
	第一節 程式開發	3
	第二節 中文分詞系統與詞性標記系統	3
	第三節 中文分詞系統及詞性標記系統 API	9
第三章	華語文書面語語料庫	11
	第一節 前言	11
	第二節 後設資料	12
	第三節 語料蒐集與建置	15
	第四節 小結	18
第四章	華語文口語語料庫	19
	第一節 前言	19
	第二節 後設資料	19
	第三節 語料蒐集與建置	19
	第四節 小結	20
第五章	華英雙語語料庫	21
	第一節 前言	21
	第二節 後設資料	21
	第三節 語料蒐集與建置	21
	第四節 小結	22

# 臺灣華語文語料庫

華語文教與學的必備工具

<b>第六章</b>	<b>華語中介語語料庫</b> .....	<b>23</b>
	第一節 前言 .....	23
	第二節 後設資料 .....	23
	第三節 語料蒐集與建置 .....	23
	第四節 小結 .....	26
<b>第七章</b>	<b>通用詞頻表</b> .....	<b>27</b>
	第一節 前言 .....	27
	第二節 應用的語料 .....	29
	第三節 小結 .....	30
<b>第八章</b>	<b>結語</b> .....	<b>65</b>
	<b>資料來源</b> .....	<b>67</b>
	<b>誌謝</b> .....	<b>71</b>



## 表目錄

表 1	臺灣 2013 年各單位語料庫建置情形 .....	2
表 2	國教院分詞系統與中研院分詞系統分詞結果.....	5
表 3	中研院平衡語料庫詞類簡化標記及代表意義.....	7
表 4	國教院書面語語料庫後設資料資訊類型 .....	12
表 5	臺灣華語文語料庫書面語語料主題分類及相對應的中文圖書分類號 .....	13
表 6	華語文書面語語料庫歷年蒐集的書本總數、篇章數與字數 .....	16
表 7	華語文書面語語料庫歷年蒐集的新聞語料之媒體來源及字數 .....	17
表 8	華語文口語語料庫歷年蒐集的節目集數與字數 .....	20
表 9	華英雙語語料庫歷年蒐集的字數 .....	22
表 10	華語中介語語料庫歷年蒐集的語料來源及數量 .....	25
表 11	通用詞頻表的語料來源、年代、數量及內涵.....	29
表 12	通用詞頻表前 1,000 詞示例 .....	30
表 13	2020 年國內語料庫建置情形 .....	65

## 圖目錄

圖 1	國教院分詞系統.....	6
圖 2	國教院分詞系統詞性標記之結果例示.....	9
圖 3	國教院「臺灣華語文語料庫」(COCT).....	65
圖 4	語料庫、數位科技及標準體系整合應用於華語文教育 .....	66

# 臺灣華語文語料庫

華語文教與學的必備工具

# 第一章 前言

教育部爲了推廣我國華語文教育產業，需研發共同的華語文學習者能力指標及分級標準，以作爲華語文教學、教材編輯及學習評量的共同參照依據。因此將《邁向華語文教育輸出大國八年計畫（102-109）》之「建置應用語料庫及標準體系」，列爲建構永續發展基礎計畫，並由國家教育研究院（以下簡稱國教院）執行。工作項目包括建置華語文語料庫，建置華語文標準體系，及辦理相關應用推廣等活動，以提升華語文教育的品質及效能（柯華葳等人，2013；柯華葳等人，2014）。本書爲建置華語文語料庫的部份研發成果。

語料庫指蒐集大量圖書、雜誌、及新聞等文本文字或口說的聲音及文字，所建置而成的資料庫，目的在於提供字詞語及語法使用的檢索和分析。語料庫有大量實際使用的語言資料，經由統計分析，可以顯示字詞語及語法真實的使用語境和使用頻率。一個大型的華語文語料庫，結合數位科技可應用於建構漢字、詞語及語法點分級標準，編輯教材、文法書及學習者辭典等應用；並可應用在發展華文拼字檢查、文法檢查、易混淆字詞表、搭配詞參考工具、華文作文評分、易讀性自動分級，甚至華語教學應用軟體、E-learning 學習平台、自動產生華文例句及試題等教學學習應用（林慶隆，2019）。

華語文標準體系的建置若能植基於語文實際使用情境，將更能符合華語文教學需要，爲此語料庫是不可或缺的工具。惟 2013 年，臺灣當時較具代表性的語料庫「中央研究院漢語平衡語料庫」（簡稱中研院平衡語料庫），語料的蒐集與維護僅至 2007 年，且以提供語言學研究爲主；至於國內其他單位建置的語料庫，也以學術研究爲主，難以提供建置華語文標準體系所用（詳見表 1）。因此，亟需研發建置華語文語料庫。

# 臺灣華語文語料庫

華語文教與學的必備工具

表 1

臺灣 2013 年各單位語料庫建置情形

		中研院平衡語料庫	其他單位
語料類別	書面語	1,000 萬詞，以新聞語料為主	教育部國語會
	口語	少量	政治大學、高雄師範大學
	華語中介語	無	臺灣師範大學
	華英雙語	無	
維護		2007 年止	視是否繼續獲得補助而定
應用	華語文學習者能力指標及分級標準建置	以語言學和計算語言學研究為主	以學術研究為主且多未取得提供其他利用的授權
	教學、教材編輯及學習評量		
	學習者詞典編輯		

資料來源：整理自柯華葳等人（2013：6）。

本研究團隊分析國內、外語料庫包括：柯林斯語料庫（The Collins Corpus）、英國國家語料庫（British National Corpus, BNC）、美國當代英語語料庫（Corpus of Contemporary American English, COCA）、中研院平衡語料庫、中國大陸國家語委現代漢語語料庫等（柯華葳等人，2013；柯華葳等人，2014）。並且諮詢專家學者後，由於美國當代英語語料庫持續更新，收錄美國當代的口語、小說、流行雜誌、報紙和學術期刊等五大類型的語料，且平均分布於五大類型中，可反映美國當代各類型的語言趨勢，為最新、最大的當代英語平衡語料庫，每天有數十萬人使用。故而，研究團隊以「美國當代英語語料庫」（COCA）為目標，建置包括臺灣正體字書面語、口語、雙語及華語中介語的「臺灣華語文語料庫」（英文名稱為 Corpus of Contemporary Taiwanese Mandarin，簡稱 COCT）（柯華葳等人，2014；柯華葳等人，2015）。以應用於建置華語文標準體系，包括聽、說、讀、寫、譯等能力指標，以及漢字、詞語及語法點等分級標準。

本書內容包括「臺灣華語文語料庫」的研發建置及應用語料庫建置的通用詞頻表（許添明等人，2019b）。

## 第二章 語料庫建置技術

「臺灣華語文語料庫」的建置，包括：文獻分析、專家諮詢、程式開發、設定後設資料、語料蒐集及建置等程序（柯華葳等人，2013；柯華葳等人，2014；柯華葳等人，2016；許添明等人，2017；郭工賓等人，2018；許添明等人，2019b）。而且，語料庫的應用必須依賴語料庫分析工具提供檢索及分析功能，爲了將「臺灣華語文語料庫」應用於建置華語文標準體系之漢字、詞語與語法點分級標準等應用，研究團隊也研發多套語料庫應用系統（吳鑑城、白明弘、林慶隆，2019）。本章僅就程式開發、中文分詞系統與詞性標記系統及其 API 等技術，進行說明。

### 第一節 程式開發

本計畫採用 Python (<https://www.python.org/>) 作爲系統開發的程式語言，資料庫則採用 mongoDB (<http://www.mongodb.org/>)，並採取 Flask (<http://flask.pocoo.org/>) 作爲前期網頁系統開發之框架。

### 第二節 中文分詞系統與詞性標記系統

中文自然語言處理的最基本單位是詞語，然而，與西方語言不同的是，組成中文句子的詞語間沒有空白爲分界。因此，所有中文文件的分析都必先自動切分出詞語的邊界。所以舉凡中文語言的相關研究，包括語料庫語言學研究、全文檢索、圖書資訊檢索、文件自動分類、知識庫系統、資料採礦、大數據分析等，都需仰賴中文分詞系統。中文分詞系統是中文分析相關研究的基礎。

早期分詞技術多半利用詞典中所收錄的詞，與文本進行比對，以找出可能包含的詞。在臺灣，2013 年以前中文分詞系統最著名且最常被使用者，應屬中研院開發的中文分詞系統（Ma & Chen, 2003）。該系統即利用約十萬詞的辭典以及人工撰寫之構詞規則，進行分詞處理，並且在第一階段分詞完成後，針對被分爲單字詞者透過未知詞（Out of vocabulary, OOV）規則進行單字詞合併，處理未知詞。

然而，由於構詞規則需專家撰寫且過程繁複，難免存有矛盾衝突的狀況，同時規則式斷詞的缺點在於十分依賴內建詞典的品質以及數量，因此較無法提升處理新詞的準確率。此外，因該中文分詞系統已停止更新，亦未公開程式原始碼，故無法直接針對特定的分詞需求進行修改。

由於中文分詞與詞性標記為中文語料處理，最基礎亦是最重要的步驟之一，本計畫發展中文分詞系統及詞性標記系統，而且逐年修訂、優化，以符合計畫需求，取得最佳化效果。概述如下：

## 一、中文分詞系統

目前中文分詞系統，已趨向使用機器學習的方式進行系統開發，包括利用最大熵（Maximum Entropy, ME）（Xue, 2003）、支援向量機（Support Vector Machine, SVM）（Asahara et al., 2005）以及條件隨機域（Conditional Random Fields, CRFs）（Peng, F., Feng, F., & McCallum, A, 2004）等。相關研究證實，機器學習式的方法能達到 9 成以上分詞準確性。因此，本計畫研究團隊利用條件隨機域（CRFs）的機器學習方式，以中研院平衡語料庫五百萬詞做為訓練資料，開發自動中文分詞系統與自動詞性標記系統。

在中文分詞系統中，除了利用前後字資訊、是否收錄於辭典中等，作為機器學習的基本特徵外，更加入前後字間的統計資訊，如相互資訊（Mutual Information）作為額外特徵。藉由使用可取得的第三屆國際分詞競賽（2006）的測試資料，進行效能評估，本模組的分詞效能準確率及召回率，已超過 95%，未知詞的召回率亦將近 80%，效能已和現今開放使用的中文分詞模組（如中研院分詞系統），達至同等水準。

表 2 為一文檔經本計畫研發的國教院分詞系統及中研院分詞系統分詞結果比較。中研院分詞系統分為線上展示版與線上服務版，服務版為展示版之簡化版本，僅針對基本詞語進行分詞。

表 2  
國教院分詞系統與中研院分詞系統分詞結果

分詞系統	分詞結果
尚未分詞	冠狀動脈疾病在臺灣十大死因佔第二位，平均每 30 分鐘就有人因心血管問題死亡，其中有一半以上是因為冠狀動脈疾病而導致。臺北市立聯合醫院心臟內科醫師林廷遠表示，高血壓，糖尿病、膽固醇過高、抽菸，以及肥胖的人，其冠狀動脈再發性心血管栓塞，將容易引發心肌梗塞而導致死亡。（梁以辰，2014）
國教院分詞系統	<b>冠狀動脈</b> 疾病在臺灣十大死因佔第二位，平均每 30 分鐘就有 <b>人</b> 因心血管問題死亡，其中有一半以上是因為 <b>冠狀動脈</b> 疾病而導致。臺北市立聯合醫院心臟內科醫師林廷遠表示，高血壓，糖尿病、膽固醇過高、抽菸，以及肥胖的人，其 <b>冠狀動脈</b> <b>再發性</b> 心血管栓塞，將容易引發心肌梗塞而導致死亡。
中研院分詞系統 (線上服務版)	<b>冠狀動脈</b> 疾病在臺灣十大死因佔第二位，平均每 30 分鐘就有 <b>人</b> 因心血管問題死亡，其中有一半以上是因為 <b>冠狀動脈</b> 疾病而導致。臺北市立聯合醫院心臟內科醫師林廷遠表示，高血壓，糖尿病、膽固醇過高、抽菸，以及肥胖的人，其 <b>冠狀動脈</b> <b>再發性</b> 心血管栓塞，將容易引發心肌梗塞而導致死亡。
中研院分詞系統 (線上展示版)	冠狀動脈疾病在臺灣十大死因佔第二位，平均每 30 分鐘就有 <b>人</b> 因心血管問題死亡，其中有一半以上是因為 <b>冠狀動脈</b> 疾病而導致。臺北市立聯合醫院心臟內科醫師林廷遠表示，高血壓，糖尿病、膽固醇過高、抽菸，以及肥胖的人，其 <b>冠狀動脈</b> <b>再發性</b> 心血管栓塞，將容易引發心肌梗塞而導致死亡。

從分詞結果比較，發現對於一般性詞彙，如「平均」、「問題」等，三組系統表現都很優異。但中研院分詞系統線上服務版，在未知詞處理上（如「冠狀動脈」）首次將「冠狀動脈」分為「冠」、「狀動脈」兩詞，第二次則分為「冠狀」、「動脈」。突顯過度依賴辭典，未知詞處理效能較不理想。至於線上展示版，效能雖較線上服務版佳，但由於利用規則法處理未知詞，導致分詞時，較易出現過度合併詞彙或詞彙不完整的狀況：如將「…有人因心血管…」分詞為「…有人因 心血管…」，或「…其冠狀動脈再發性心血管栓塞…」分詞為「…其 冠狀動脈 再發性 心血管 栓塞…」。相對於此，國教院中文分詞系統皆成功地將詞彙完成分詞。

由於本計畫需要處理相當大量的語料，因此，除了分詞效果外，也特別留意分詞速度。測試結果顯示，十萬字的資料只需不到一分半鐘的時間，即可完成分詞。以此效能，顯見足以支持大量語料的快速處理（柯華葳等人，2014）。

為了便於觀察效能並進行改善，研究團隊亦建置製作中文分詞系統網頁版（「國教院分詞系統」如圖 1），以供計畫團隊便於觀察分詞結果。同時為了分享計畫成果，亦撰寫下載、安裝及簡易使用參數等說明，併同原始碼以 GPL（General Public License, Version 3）的方式上傳到 GitHub，開放使用者下載使用，網址詳見：<https://github.com/naernlp/Segmentor>（柯華葳等人，2016）。

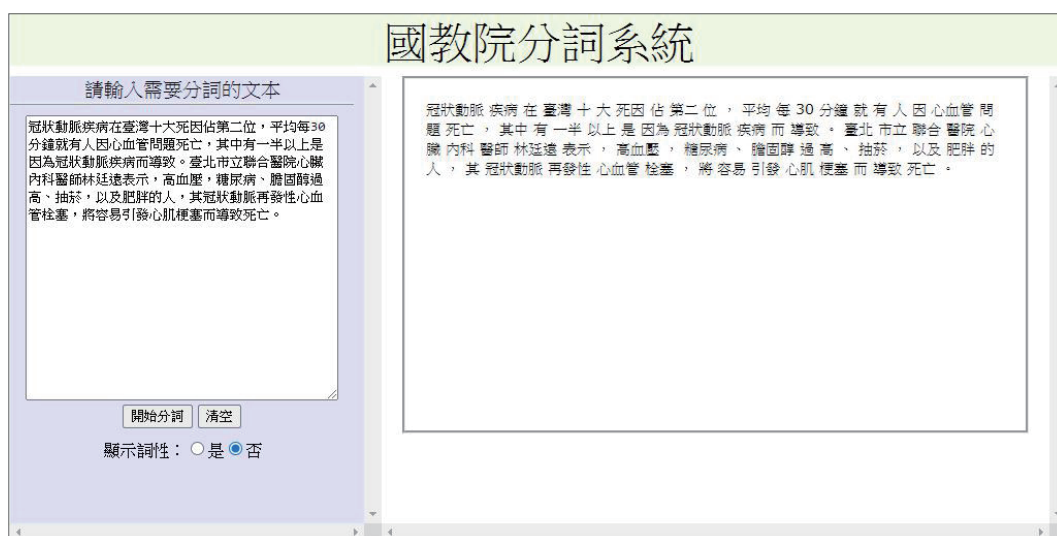


圖 1 國教院分詞系統

## 二、詞性標記系統

詞性標記程式（part-of-speech tagger），顧名思義即能自動將所輸入的句子中的每一個詞，予以標明詞性。正體中文的詞性標記程式的開發，尤以中研院資訊科學研究所中文詞知識庫小組為代表，其分詞程式可以同時處理分詞和詞性標記。國教院分詞系統採用〈中研院平衡語料庫詞類標記集〉的「簡化標記」46 個。C 開頭表示連接詞類、V 表示動詞類、N 表示體詞類、A 表示非謂形容詞、D 表示副詞類、P 表示介詞、I 表示感歎詞、T 代表語助詞。關於 46 個簡化標記及其代表之意義，詳如表 3。



表 3  
中研院平衡語料庫詞類簡化標記及代表意義

序 號	簡 化 標 記	代 表 意 義
1	A	非謂形容詞
2	Caa	對等連接詞，如：和、跟
3	Cab	連接詞，如：等等
4	Cba	連接詞，如：的話
5	Cbb	關聯連接詞
6	D	副詞
7	Da	數量副詞
8	DE	的，之，得，地
9	Dfa	動詞前程度副詞
10	Dfb	動詞後程度副詞
11	Di	時態標記
12	Dk	句副詞
13	FW	外文標記
14	I	感嘆詞
15	Na	普通名詞
16	Nb	專有名稱
17	Nc	地方詞
18	Ncd	位置詞
19	Nd	時間詞
20	Nep	指代定詞
21	Neqa	數量定詞
22	Neqb	後置數量定詞
23	Nes	特指定詞

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

序 號	簡 化 標 記	代 表 意 義
24	Neu	數詞定詞
25	Nf	量詞
26	Ng	後置詞
27	Nh	代名詞
28	P	介詞
29	SHI	是
30	T	語助詞
31	VA	動作不及物動詞
32	VAC	動作使動動詞
33	VB	動作類及物動詞
34	VC	動作及物動詞
35	VCL	動作接地方賓語動詞
36	VD	雙賓動詞
37	VE	動作句賓動詞
38	VF	動作謂賓動詞
39	VG	分類動詞
40	VH	狀態不及物動詞
41	VHC	狀態使動動詞
42	VI	狀態類及物動詞
43	VJ	狀態及物動詞
44	VK	狀態句賓動詞
45	VL	狀態謂賓動詞
46	V_2	有

資料來源：詞庫小組（1998：23）

在詞性標記系統中，除了利用前後字資訊，亦加入詞彙本身組成字的資訊，作為額外特徵。藉由測試資料進行效能評估，本系統的詞性標記準確率，已超過九成。根據文獻，中文詞性標記系統隨著測試資料的不同，準確性約為 83% 至 97%。顯示本系統已有很好的準確率。以前述中文分詞所使用的文本段落為例，經國教院分詞系統進行分詞及詞性標記的結果，詳見圖 2。

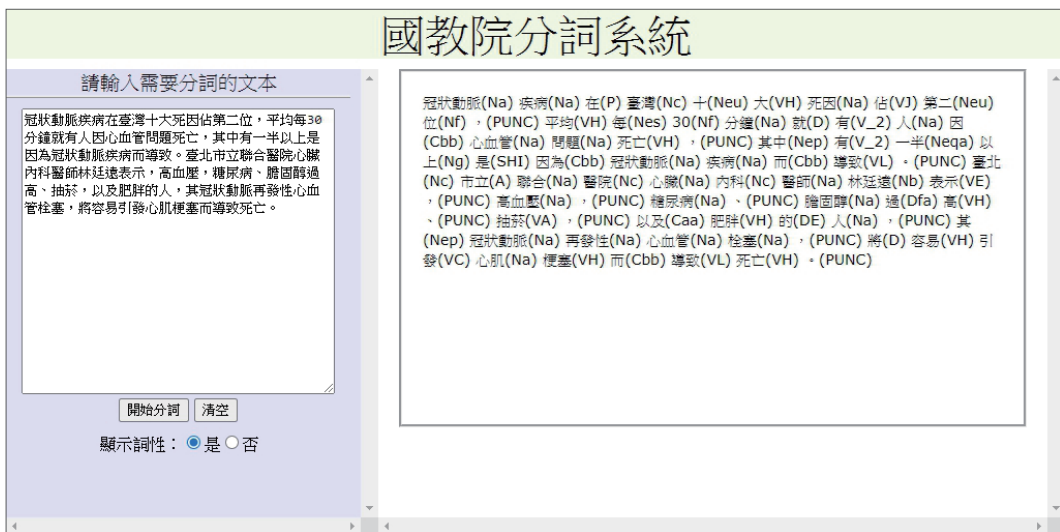


圖 2 國教院分詞系統詞性標記之結果例示

### 第三節 中文分詞系統及詞性標記系統 API

中研院的 CKIP<sup>1</sup> 分詞系統採經驗法則模型，經驗法則模型的好處是可以提供自己的詞典，但執行速度較慢，程式的維護也比較不易。

國教院分詞系統採純統計式模型，純統計式模型好處是模型簡單，執行速度快，程式容易維護。但缺點是使用者不能提供自己的詞典。

若以標記的正確性來說，CKIP 的分詞系統正確率略高於國教院的分詞系統，差別在 0.5% 以下。另外，國教院的詞性標記完全採用中研院的詞性標記系統，唯

<sup>1</sup> CKIP 為“CHINESE KNOWLEDGE AND INFORMATION PROCESSING”之縮寫。

# 臺灣華語文語料庫

華語文教與學的必備工具

一的差別只在於標點符號的詞性：國教院一律標為 PUNC，而 CKIP 則針對每一個標點符號有一個詞性標記。

關於安裝 Segmentor 模組簡易使用方法、命令列參數說明，詳見網址：<https://github.com/naernlp/Segmentor>。

## 第三章 華語文書面語語料庫

### 第一節 前言

關於書面語語料庫，英、美較著名者有柯林斯語料庫（The Collins Corpus）、英國國家語料庫（British National Corpus, BNC）、美國當代英語語料庫（Corpus of Contemporary American English, COCA）。至於華語書面語著名的語料庫，則有：中研院漢語平衡語料庫（Sinica Corpus）、中國大陸國家語委現代漢語語料庫。這些大型語料庫能提供具代表性的、大量的、系統化的語言使用資料，並協助語言學研究者及學習者，驗證其語言使用的假設，而非如傳統作法，僅能仰賴主觀的、個別的語感判斷。換言之，語料庫能提供較為充足的語言使用趨勢，且能協助語言教學者、研究者、學習者等進一步提升其教學、研究與學習的品質。

「中研院漢語平衡語料庫」為我國第一個完整經過人工檢視詞類標記的華語平衡語料庫。該語料庫語料內容分為六大主題為主軸，包括哲學、科學、社會、藝術、生活、文學，共計約 1,000 萬詞。語料來源則是 1981 年到 2007 年間，來自報章雜誌、網路資源、大專院校等機構提供之書面語料，依主題、出處、文類、年代、作者等加以標記。可說是當時華語世界具代表性之樣本（詞庫小組，1998；柯華葳等人，2014）。然而，「中研院漢語平衡語料庫」僅一千萬詞的規模，與英美國家動輒數億，甚至 10 幾億字之書面語語料庫相比，明顯不足。同時，其建置目的在於語言學和計算語言學研究，並非針對對外華語教學而設計；且語料內容以新聞語料居多，缺乏口語語料、中英雙語語料及中介語語料，並自 2007 年以後未再更新。

教育部前國語推行委員會曾兩度執行國語語料庫計畫，蒐集語料並分析新詞彙的研究，共蒐集約 4 百萬詞，由於定位為國語會內部參考及語言學研究，語料未得到授權，故無法開放外界查詢（柯華葳等人，2013）。另外，亦有一小部分散見在各大學的科技部研究計畫中，且僅限計畫人員內部參考外，絕大部分沒有獲得授權，所以也無法公開。

對外華語教學界最迫切需要者，是各文類皆具備且比例平衡的數億書面語語料，當時國內的語料庫，無論在質或量方面，均完全無法支援對外華語教學（柯華葳等人，2013）。

## 第二節 後設資料

華語文語料庫之書面語語料庫的後設資料，研究團隊參考英國國家語料庫（BNC）、美國當代英語語料庫（COCA）以及中研院平衡語料庫等國內、外語料庫相關的後設資料設定方式（柯華葳等人，2013），並根據語料庫之用途，擬定書面語後設資料，共有作者、出版者、出版年、語文、著（譯）作、主題、媒體、語式、著作權等 9 類資訊（詳見表 4）。此外，根據擬定的後設資料以及文本內容需求，本系統之資料庫設計方式，對於每一份文本共設計 12 個欄位進行資訊儲存，包括：標題、內文、作者、出版者、出版年、語文、著（譯）作、主題、媒體、語式、著作權、備註等。

表 4  
國教院書面語語料庫後設資料資訊類型

資訊類型	說明
作者	文檔之作者
出版者	文檔之出版者
出版年	文檔出版年份
語文	文檔主要使用語言（預設華語）
著（譯）作	文檔屬著作，或是其他語言之華語譯作
主題	文檔所屬主題，共分 10 類：總類、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。
媒體	資料來源類型，共分 10 類：報紙、一般雜誌、學術期刊、教材、工具書、學術論著、一般圖書、視聽媒體、會話訪談、其他。
語式	文檔呈現方式（書面語、口語、雙語）
著作權	文檔使用權限（全文、非全文）

其中，關於主題類別，本計畫期使日後文本於應用時，能夠發揮更大的效益，例如分析不同主題類型的文本詞語使用的差異性，或是學習不同主題文本的文句寫作模式等。而且，考量本「臺灣華語文語料庫」主體內容為華文，而中文圖書分類法亦已被採用多年，可視為一穩定的文本類型分類基礎。因此書面語語料主題分類參考中文圖書分類架構，修改並整合為 10 大類：總類、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。國教院書面語語料主題分類與中文圖書分類號之對應，詳見表 5。

表 5

臺灣華語文語料庫書面語語料主題分類及相對應的中文圖書分類號

國教院書面語語料主題分類名稱	中文圖書分類號	備註
總類	000-099[ 總類 ]	
哲學及宗教類	100-199[ 哲學類 ] 200-299[ 宗教類 ]	
科學類	300-399[ 自然科學類 ]	
應用科學類	400-479[ 應用科學類 ]	480-489 『商業：各種營業』 490-499 『商學：經營學』 歸入商業及金融類。
社會科學類	500-549 554-557 570-599[ 社會科學類 ]	550-553 以及 558-569 歸入商業及金融類。
史地類	600-699[ 史地類 ] 710-799[ 世界史地類 ]	700-709 中文圖書分類無資料
語言文學類	800-899[ 語文類 ]	
藝術類	900-989	990-999 『遊藝·娛樂·休閒』 歸入休閒類

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

國教院書面語 語料主題分類名稱	中文圖書分類號	備註
商業及金融類	480-499 550-553 558-568	480：商業總論 481：糧商業 482：其他農產品業 483：畜牧水產品業；飲食相關行業 484：機械業；電機資訊業 485：化學製品業 486：礦產品業 487：製造品業 488：紡織品業 489：其他各種營業 490：商學總論 491：商業地理 492：商政 493：商業實踐 494：企業管理 495：會計 496：商品學；市場學；行銷管理 497：廣告 498：商店 499：企業志；公司行號志 550：經濟學總論 551：經濟學各論 552：經濟史地 553：生產；企業；經濟政策 558：貿易 559：合作 560：財政學總論 561：貨幣；金融 562：銀行 563：金融各論 564：公共財政 565：各國財政狀況 566：地方財政 567：租稅 568：關稅

(續下頁)



國 教 院 書 面 語 語 料 主 題 分 類 名 稱	中 文 圖 書 分 類 號	備 註
休閒類	990-999	990：遊藝及休閒活動總論 991：公共娛樂 992：旅遊；觀光 993：戶外活動 994：水上運動 995：室內遊戲 996：兒童遊戲 997：智力遊戲 998：博戲 999：業餘遊玩；民俗藝術

資料來源：柯華葳等人（2015：6）

### 第三節 語料蒐集與建置

#### 一、語料蒐集

書面語語料分為一般著作語料及新聞語料。分述如下：

##### 1. 一般著作語料

本計畫執行初期，曾蒐集國教院編譯之著作及其他政府機構出版品的書面語語料。前者內容豐富多樣，包括期刊、電子報、叢書與各類書籍等實體或數位出版品。後者包括僑委會的部分資料、國立臺灣科學教育館之月刊、國立科學工藝博物館圖書、部編版國小「國語」、「社會」與「自然」教科書，及國教院與世華會共同出版品。惟因內容與 2015 年以後所收大量語料屬性不同，尙待收入書面語索引典，對外開放查詢。

本計畫自 2015 年起，經由公開招標取得坊間出版業者的書面語語料，年代為 2008 年以後至 2020 年，內容涵蓋哲學及宗教類、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、休閒類等多元文章類型（柯華葳等人，2015）。文章內文以臺灣使用之正體中文為限，每篇文章約 2,000

至 5,000 中文字，且每篇皆須有純文字類型檔案（TXT 格式，UTF-8 編碼）。且每篇文章皆需來自具 ISBN 碼之書籍。本計畫華語文書面語語料歷年蒐集的一般著作語料字數，詳見表 6。

表 6  
華語文書面語語料庫歷年蒐集的書本總數、篇章數與字數

年 度	書 本 總 數	篇 章 數	字 數	累 計
2014	447	20,116	約 5,220 萬	約 5,220 萬
2015	387	20,111	5,000 萬	1 億 220 萬
2016	1,182	63,829	1 億 5,000 萬	2 億 5,220 萬
2017	1,783	67,458	1 億 5,000 萬	4 億 220 萬
2018	未取得			4 億 220 萬
2019	246	9,040	約 2,100 萬	4 億 2,320 萬
2020	78	4,580	約 1,221 萬	4 億 3,541 萬
合 計	4,123	185,134	4 億 3,541 萬	4 億 3,541 萬

## 2. 新聞語料

本計畫為擴充書面語語料的規模，亦積極結合院內其他計畫。如本院歷年來為編撰字、詞典及進行年度語詞調查等工作，所蒐集的資料，包括：《聯合報》（1999-2018 年）、《中國時報》（1999-2018 年）、《大紀元時報》（2016-2018 年）、《蘋果日報》（2016-2018 年）、《國語日報》（2016-2018 年）等，總字數約 30 億字，詳見表 7。

表 7  
華語文書面語語料庫歷年蒐集的新聞語料之媒體來源及字數

年 度	聯 合 報	中 國 時 報	大 紀 元	蘋 果 日 報	國 語 日 報
1999	80,811,113	98,294,638			
2000	120,774,269	98,838,003			
2001	111,789,006	93,245,297			
2002	101,466,860	105,575,760			
2003	94,950,364	104,937,597			
2004	91,113,894	100,694,048			
2005	84,786,928	93,898,158			
2006	81,651,354	89,781,432			
2007	74,013,381	83,889,983	——	——	——
2008	72,923,542	63,670,338			
2009	64,443,013	39,275,377			
2010	64,588,905	42,881,094			
2011	62,577,654	42,486,838			
2012	65,713,357	42,269,652			
2013	63,532,273	45,019,123			
2014	65,229,496	43,658,834			
2015	63,296,395	42,652,212			
2016	52,555,057	37,511,285	21,342,536	39,833,481	1,787,473
2017	38,010,412	34,963,207	22,772,087	27,722,253	1,814,435
2018	38,392,319	37,459,941	15,544,597	30,106,774	1,823,414
小 計	2,996,369,459	1,341,002,817	59,659,220	97,662,508	5,425,322
合 計	2,996,369,459				

## 二、語料建置

以上語料皆經由本計畫中文分詞系統及詞性標記系統處理後，收錄於書面語資料庫，並建置「國教院語料庫索引典系統」（內含 COCT 書面語語料庫 2015、2016、2017 及 2019），提供使用者使用。

### 第四節 小結

華語文書面語語料庫的後設資料，包括作者、出版者、出版年、語文、著（譯）作、主題、媒體、語式、著作權等 9 項資訊。其中主題包括總類、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 10 類。語料來源包含 2008 年至 2020 年的一般著作語料約 4 億 3,541 萬及 5 家新聞媒體語料約 30 億字。使用者可應用「國教院語料庫索引典系統」（<https://coct.naer.edu.tw/cqpweb/>），進行檢索與使用。惟新聞語料限於使用授權，僅限於院內使用，如有使用需求，請與本院聯絡。

## 第四章 華語文口語語料庫

### 第一節 前言

關於口語語料庫，英、美等國家都有一些著名的語料庫（柯華葳等人，2014），然而，2013年時我國各口語語料庫相較於英美各國相形見绌。著名的政大漢語口語語料庫多年成果，雖已收錄約15萬字之語料，然而國外口語語料庫卻早已達到百萬甚至千萬字。中研院的現代漢語語音語料庫則是主要為語言學以及語音研究，非為對外華語教學建置。至於在各大學的科技部研究計畫所建置的語料庫，語料數量少，且僅限計畫人員內部參考，絕大部分沒有獲得授權，所以也無法提供使用。因此，當時國內的語料庫，無論在質或量方面，均完全無法支援對外華語教學（柯華葳等人，2013）。

### 第二節 後設資料

華語文口語語料庫的後設資料，根據本語料庫之用途，並參考書面語語料庫後設資料，共有7類資訊：發行者、發行年、語文（正體中文）、主題、媒體（電視）、語式（口語）、著作權（電視臺）等。其中主題類別包括法政軍事、財經、時事、科學、生活時尚、文教藝術等項。

### 第三節 語料蒐集與建置

#### 一、語料蒐集

本計畫執行初期（2014），曾嘗試蒐集國教院及其他機構口語語料。國教院語料包括愛學網之性別平等、品德教育、社會科之影片，以及291部全國優良教育影片徵集得獎作品。其他機構語料包括國立政治大學英國語文學系徐嘉慧教授之「國立政治大學漢語口語語料庫」及國立高雄師範大學華語文教學研究所王莢芳教授蒐集之口語語料，惟因著作授權等問題，無法取得或取得僅能供內部研究使用（柯華葳等人，2015）。

本計畫自2015年起，經由公開招標取得民間電視臺的口語語料，年代為2008年以後至2020年，內容涵蓋法政軍事、財經、時事、科學、生活時尚、文教藝術

等多元節目類型。節目使用之語言以臺灣使用之國語為限，文字則以臺灣使用之正體中文字為主，且字幕對應之語音，非國語部分以不超過 10% 為原則；每集節目約 20-50 分鐘（不含商業廣告），且每集皆包含語音檔（WAV 格式、Sampling Rate 48kHz[含] 以上）暨逐字稿文字檔（TXT 格式、UTF-8 編碼），兩種類型檔案。本計畫華語文口語語料庫歷年蒐集的節目集數及字數，詳見表 8。

表 8

華語文口語語料庫歷年蒐集的節目集數與字數

年 度	集 數	字 數	累 計 (字)
2014	500	300 萬	300 萬
2015	1,100	660 萬	960 萬
2016	1,270	1,300 萬	2,260 萬
2017	500	290 萬	2,550 萬
2018	750	500 萬	3,050 萬
2019	1,300	895 萬	3,945 萬
2020	1,240	690 萬	4,635 萬
合 計	6,660	4,635 萬	4,635 萬

## 二、語料建置

以上語料皆經由本計畫中文分詞系統及詞性標記系統處理後，收錄於口語資料庫，並建置「國教院語料庫索引典系統」（內含 COCT 口語語料庫 2016、2017 及 2019），提供使用者使用。

### 第四節 小結

華語文口語語料的後設資料，包括發行者、發行年、語文（正體中文）、主題、媒體（電視）、語式（口語）、著作權（電視臺）等 7 類。其中主題涵蓋包括：法政軍事、財經、時事、科學、生活時尚、文教藝術等 6 類。語料來源為 2008 年至 2020 年的電視臺華語口語節目語音檔暨文字檔。至 2020 年為止，華語文口語語料已蒐集 4,635 萬字。使用者可透過「國教院語料庫索引典系統」（<https://coct.naer.edu.tw/cqpweb/>），進行檢索與使用。

## 第五章 華英雙語語料庫

### 第一節 前言

語料庫雖然非常有用，但對於外語學習者而言，傳統的單語關鍵詞語境檢索有相當大的限制，使用者必須先知道所要查詢的詞，但是非母語人士常有不知道如何用外語詞彙表達概念的困擾。雙語關鍵詞語境檢索可以解決這方面的問題，使用者可以用兩種語言進行雙向查詢。例如當外國人士不知道某一個中文概念的表達方式時，可以利用輸入英文關鍵詞，程式將自動從中英對應語料庫找出包含此關鍵詞的句子及其翻譯。反之，當學習者不知道一個中文的意義時，利用輸入中文關鍵詞即可找到對應的英文翻譯。

然而，2013 年時國內外雖然有中研院陳克健教授和黃居仁教授「中文詞知識庫」及黃居仁教授「中文詞彙網路」(<http://cwn.ling.sinica.edu.tw>)，大陸董振東先生「知網 HowNet」(<http://www.keenage.com>) 及北京大學「漢英雙語平行語料庫」，清大資工系張俊盛教授和清大外語系劉顯親教授「TOTALrecall 英中雙語辭彙索引 (Concordance) 查詢系統」，及“Opus Multilingual Corpus”(<http://opus.lingfil.uu.se/bin/opuscqp.pl>)。惟因有些僅有少數雙語語料，且多數都不是近年的語料，有些由英語翻譯而來，裡面的中文未必道地，且這些語料不是生活化的語言，對華語教學而言，並不是理想的材料 (柯華葳等人，2013)。

### 第二節 後設資料

華英雙語語料的後設資料，主要參考前述書面語語料庫及口語語料庫之設定，包含作者、出版年、出版者、出版單位、主題、媒體 (雜誌)、語式 (書面語) 等 7 項。惟因雙語語料之「主題」，涵蓋性不似書面語廣泛，故不採用書面語的主題分類法，而改採文學、科學、財經、藝術、思想、文化、全球、休閒等 8 項分類 (柯華葳等人，2015)。

### 第三節 語料蒐集與建置

#### 一、語料蒐集

本計畫執行初期，曾蒐集政府及民間單位之雙語語料。前者包括僑委會授權之對外華語部分教材及教育部網站訊息，後者如中央社之「新聞 E 點通」、英文中國

郵報之「臺灣原住民族文化」報導等。自 2015 年起，經由招標取得《台灣光華雜誌》(<https://www.taiwan-panorama.com/>) 的華英雙語語料（柯華葳等人，2015、2016；許添明等人，2017；郭工賓等人，2018；許添明等人，2019b）。

本語料庫收錄《台灣光華雜誌》自 1996 年 1 月起至 2018 年 6 月的所有刊期。內容涵蓋文學、科學、財經、藝術、思想、文化、全球、休閒等具教育意義的多元文章類型；文章使用之中文，以臺灣使用之正體中文字為限；每篇文章的文字檔皆包含中、英文字，且為中文與英文段落相互對應的純文字類型檔案（TXT 格式，UTF-8 編碼），不含圖片。本計畫華英雙語語料庫歷年蒐集的字數，詳見表 9。

表 9  
華英雙語語料庫歷年蒐集的字數

年 度	字 數	累 計
2015	340 萬	340 萬
2016	360 萬	700 萬
2017	300 萬	1,000 萬
2018	160 萬	1,160 萬
2019	—	1,160 萬
合 計	1,160 萬	1,160 萬

## 二、語料的建置

華英雙語語料的建置過程，包括開發雙語文字對應模組（柯華葳等人，2013）、人工對應部分雙語語料（柯華葳等人，2015）、開發雙語索引典系統（柯華葳等人，2016），並歷經改良雙語文字對應模組（郭工賓等人，2018）。最後，於 2019 年完成「國教院華英雙語索引典系統」。

### 第四節 小結

華英雙語語料的後設資料，包括作者、出版年、出版者、出版單位、主題、媒體（雜誌）、語式（書面語）等 7 項資訊。其中主題涵蓋：文學、科學、財經、藝術、思想、文化、全球、休閒等 8 類。語料來源為《台灣光華雜誌》1996 年 1 月至 2018 年 6 月出版之所有刊期。至 2019 年，華英雙語平行語料已蒐集約 1,160 萬字（以中文語料計算）。使用者可透過「國教院華英雙語索引典系統」（<https://coct.naer.edu.tw/bc/>），進行檢索與使用。



## 第六章 華語中介語語料庫

### 第一節 前言

中介語語料庫 (interlanguage corpus) 亦稱為學習者語料庫 (learner corpus)，可用於分析語言學習者在目標語 (target language) 上的主要語言特徵與發展。

國際上較著名的英語學習者語料庫，包括國際英語學習者語料庫 (International Corpus of Learner English, ICLE)、英語學習者語料庫 (The Montclair Electronic Language Learner Database, MELD)、日本英語學習者語料庫 (Japanese EFL Learner Corpus, JEFLL)、劍橋學習者語料庫 (Cambridge Learner Corpus) 與中國學習者英語語料庫 (Chinese Learner English Corpus, CLEC) 等。至於國內的華語學習者語料庫，則有臺灣師範大學自 1998 年起陸續建置的各類型臺灣華語中介語語料庫，包括華語病句語料庫、印尼華僑學生的中介語語料庫、對外漢語中介語語料庫、華語文中介語語料庫、華語口語中介語；另外還有華語文能力測驗 (TOCFL) 寫作文本語料庫。大陸則有北京大學 HSK 動態作文語料庫 (柯華葳等人，2013)。

上述國內華語中介語語料庫有使用授權等限制，無法完全支援對外華語教學，因此，本計畫研發建置以正體中文書寫之臺灣華語中介語語料庫，以提供對外華語教學所需資源。

### 第二節 後設資料

本計畫參考國立臺灣師範大學國語教學中心簡章之學生學習背景資料及學者專家諮詢，訂定華語中介語語料庫之後設資料，包括題目、性別、文體 (作文)、母語、能力等級、性質 (限時考試、不限時作業)、建檔方式 (打字、手寫) (許添明等人，2019)。

### 第三節 語料蒐集與建置

本計畫參考教育部國際及兩岸教育司全國大專校院境外生生活資訊網網站之華語文中心國際學生人數統計資料，考量臺灣華語學習者國籍分布情形，進行語料蒐集並確立本計畫所蒐集之語料具代表性。

## 一、語料蒐集

本語料庫語料來源主要有二：一、臺灣各大學附設華語文中心之臺灣華語學習者自願提供的文章。二、國家華語測驗推動工作委員會（以下簡稱華測會）授權之限時考試語料。

### 1. 臺灣華語學習者自願提供的文章

本計畫分別與國立臺灣師範大學國語教學中心、國立臺灣師範大學華語文教學系、國立臺灣大學文學院語文中心中國語文組、國立臺灣大學中國文學系、國立政治大學華語文教學中心、國立臺北教育大學華語文中心、淡江大學成人教育部華語中心、中國文化大學推廣部華語中心、輔仁大學附設語言中心、實踐大學華語中心及開南大學合作蒐集中介語語料。2016 年以前，以研究團隊到場蒐集紙本作文的方式蒐集，共蒐集書面語語料 61 萬 5,000 字。自 2017 年起，由華語教師協助蒐集課堂之寫作文章，取得語料約 22 萬 5,000 字，2018 年蒐集 20 萬 4,000 字。

### 2. 華測會授權之限時考試語料

本計畫自 2015 年起向華測會提出申請，獲其同意授權限時考試語料。2015 年取得「進階高階級寫作測驗」文本，共約 12 萬字；2016 年取得「基礎級記敘文寫作測驗」文本 54 份，「進階級記敘文寫作測驗」文本 117 份，共計 5 萬 5,000 字；2017 年取得「進階級記敘文寫作測驗」文本 99 份及「高階級應用文寫作測驗」文本 52 份，共計 6 萬 2,000 字，而且請華測會於華語文寫作能力測驗試場，協助蒐集「進階高階級」及「流利精通級」受試者之寫作測驗語料，共計取得 4 萬 5,000 字；2018 年取得「高階級應用文寫作測驗」文本 194 份及「基礎級記敘文寫作測驗」文本 49 份，共計 9 萬 6,000 字；2019 年取得「基礎級記敘文寫作測驗」文本 72 份及「進階級記敘文寫作測驗」文本 72 份，共計 5 萬字；2020 年取得「進階級應用文寫作文本」文本 97 份及「流利精通級觀點論述寫作文本」文本 63 份，共計 9 萬 2 千多字。上述兩個來源共蒐集 156 萬 4,000 字（柯華葳等人，2014；柯華葳等人，2015、

2016；許添明等人，2017；郭工賓等人，2018）。本計畫華語中介語語料庫歷年蒐集的語料來源及數量詳見表 10。

表 10  
華語中介語語料庫歷年蒐集的語料來源及數量

年 度	華 語 文 中 心	華 測 會 試 場	華 測 會 授 權	累 計
2015	約 30 萬	—	12 萬	約 42 萬
2016	31 萬 5,000	—	5 萬 5,000	79 萬
2017	22 萬 5,000	4 萬 5,000	6 萬 2,000	112 萬 2,000
2018	20 萬 4,000	—	9 萬 6,000	142 萬 2,000
2019	—	—	5 萬	147 萬 2,000
2020	—	—	9 萬 2,000	156 萬 4,000
小 計	104 萬 4,000	4 萬 5,000	47 萬 5,000	156 萬 4,000
合 計	156 萬 4,000			

分析 2014 至 2017 年本計畫蒐集的華語學習者母語排名依序為：日語、越南語、韓語、印尼語及英語，而我國國際學生國籍人數排名則為日本、越南、南韓、印尼、美國。雖然國際學生國籍人數排名，次序雖非固定，但人數最多的前五個國家皆為日本、美國、南韓、越南、印尼。此前五名和本計畫所蒐集的書面語語料學習者母語之分布大致吻合，足見本計畫蒐集的語料在學生人數上有其代表性（郭工賓等人，2018）。

## 二、語料建置

以上語料皆經由本計畫中文分詞系統及詞性標記系統處理後，收錄於中介語資料庫，並建置「國教院華語中介語索引典系統」（2016 及 2019），提供使用者使用。

## 第四節 小結

華語文中介語語料的後設資料，包括題目、性別、文體（作文）、母語、能力等級、性質（限時考試、不限時作業）、建檔方式（打字、手寫）等 7 項。語料來源為 2015 年至 2018 年臺灣各大學華語文中心之臺灣華語學習者提供之文章，以及 2015 年至 2020 年間，華測會授權之限時考試語料。至 2020 年，華語文中介語語料已蒐集 156 萬 4,000 字。使用者可透過「國教院語料庫索引典系統」（<https://coct.naer.edu.tw/cqpweb/>），進行檢索與使用。

# 第七章 通用詞頻表

## 第一節 前言

傳統語言教材的編寫依賴專家的語言知識，但專家的語言知識有不可觀察驗證的限制（Sampson & Bentham, 1992），以及無法提供詞語的常用度以及語義常用度訊息的缺點（Sinclair, 1991: 39）。過去許多研究者為了解語言教材與實際使用語言之間的差距，將教材內容和語料庫做比較。例如：Ljung（1990）比較瑞典學校中使用的對外英語教學教材與語料庫中的高頻詞彙，發現語言教材中所使用的詞彙和語料庫中最常用的詞彙存在顯著差異。Mindt（1996: 246）發現語言教材中的模態動詞，未來時間取向和條件從句的學習順序與語料庫中的觀察有明顯差異。類似的研究還包括未來式的比較（Mindt, 1986, 1997）連接副詞的比較（Conrad, 2004）情態助動詞的比較（Biber & Reppen, 2002: 199）。這些研究都發現，語言教材中所提供的詞彙和語法知識與學習者在語言真實使用情境的需求常存在極顯著的差異。這是缺乏客觀的語言使用證據及常用度訊息的情況，專家直覺流於主觀與偏離的證據。而基於對實際語言使用的語料庫分析可以幫助改進語言教材所提供的語言知識更貼近實際的需求。

除了比較研究之外，也有許多實際將語料庫應用到語文教材的編輯。例如 West（1953）所編纂的 GSL 詞表（General Service List）是從 250 萬詞的英文書面語語料庫，依據詞彙頻率、易學程度及詞義涵蓋層面等因素，篩選出約 2,000 個中心詞（headwords）編纂而成。GSL 詞表歷年來廣為英語學習者所使用。Coxhead（2000）則自 350 萬字的書面學術語料中，排除 GSL 詞表已收錄的 2000 詞後，再依其在學術語料的分布率及詞頻高低計算出 570 個學術英文常用中心詞，稱作 The Academic Word List。隨著電腦儲存量與計算能力的增加及語料庫技術的發展，語料庫的應用領域與層面也逐漸開展。例如：McCarthy 依據 CANCODE 語料庫為基礎編寫 Touchstone 系列英語學習教材，該教材由於取材自語料庫，並提供詞語及片語的常用度訊息，使學習者避免使用太艱澀的詞語及片語，更貼近母語者的日常用法（Reppen, 2010）。

除此之外，現代詞典的編輯也得力於語料庫的發展。Verlinde & Selva（2001）比較語料庫為本的編撰法和專家直覺式（intuition-based）編撰法的優缺點發現，語料庫為本的詞典能較詳盡的表列重要日常詞彙，而且能得知多少詞彙量足以覆蓋多數

文本，讓學習者能從基礎詞彙出發，較快掌握生活用語。Summers (1996) 認為詞頻統計對詞典編輯的重要性有下列幾項：一、詞義依據詞頻排序；二、詞條標示詞頻訊息；三、詞頻針對不同詞類統計；四、詞組頻率統計；五、高頻慣用語統計詞頻等等。自 1980 年代起蓬勃發展，多部各具特色的英語學習者線上詞典先後出版，且定期更新，如：《牛津高階學習者詞典》(Oxford Advanced Learner's Dictionary)、《朗文當代英語詞典》(Longman Dictionary of Contemporary English)、《柯林斯高階學習者詞典》(Collins COBUILD Advanced Learner's English Dictionary)、《劍橋學習者詞典》(Cambridge Learner's Dictionary)、《劍橋高階學習者詞典》(Cambridge Advanced Learner's Dictionary) 和《麥克米倫高階學習者詞典》(Macmillan English Dictionary for Advanced Learners) 等等。這些詞典以輔助學習為目標，運用現代語言學的研究成果，在大型語料庫的支持下，不斷加強實用性及便利性，為其他語言學習者詞典的編纂樹立了典範。

語料庫頻率是詞語常用性指標，代表每一個詞語在語料庫中的常用性。儘管部份研究者對於語料庫頻率做為常用詞的依據存在疑慮，但是依語料庫詞頻選詞的大原則仍是必然的做法(張莉萍、陳鳳儀，2005)。因為在數量有限的詞表中，收錄較高頻的詞，語料庫覆蓋率也勢必較高，故選擇高頻詞教學關係到語言教學的效率。儘管如此，過去的研究從詞頻統計的角度發現，有部分高頻詞未必是常用詞，這些高頻詞只集中出現在特定領域的文本中重複出現，造成這些詞的頻率雖高卻不是常用詞的現象。為解決此一問題，許多學者提出不同的常用詞計算公式(白明弘、吳鑑城、簡盈妮、黃淑齡、林慶隆，2016; Carroll, 1970; Huang et al., 2005; Juilland & Chang-Rodríguez, 1964; Rosengren, 1971)。這些研究方法主要是透過篇章分佈率的計算，抑制高頻但只分布在少數文本的詞語，使篇章分佈率的值趨近於常用度。

有關中文的核心詞彙及通用詞彙，中研院詞庫小組，曾應用 500 萬詞的語料庫，並參考當時五本坊間通行辭典的收詞內容，於 1998 年訂定包括中文核心詞彙及通用詞彙的中研院詞頻辭典(詞庫小組，1998)，屬於最早有詞類標記的詞頻表，然而距今已 20 年，亟需應用近年來的大量語料，訂出新的通用詞彙表。

建立〈通用詞頻表〉之目的，旨在蒐集真實使用情境且橫跨不同世代中的詞語，探討華語文詞語的歷時發展，及單一詞語在不同語式(書面語、口語)或語體(書籍、新聞)的用法。並建構華語文教學所需的基礎知識，包括字、詞、語法點及搭配詞等等。統計結果不僅能夠提供華語文教學各學習階段常用字、詞、及語法點及搭配的依據，並提供華語教學上詞語常用性的可靠依據，及教材編輯、華語測驗之參考。除了

協助華語教師、教材與測驗編輯者，以此進行詞語的引註與檢索，使特定主題領域詞語的使用更加規範化，從而有助於華語教材及測驗內容的規劃及研發，亦可應用於語言學研究。

## 第二節 應用的語料

通用詞頻表的建置語料來源包括書面語、口語及新聞語料，語料年代、數量及內涵，詳見表 11。

表 11

通用詞頻表的語料來源、年代、數量及內涵

語料來源		語料年代	收錄數量	內涵
書面語		2004 到 2018 年	4 億 1,000 萬	涵蓋包含哲學、宗教、科學、應用科學、社會科學（如法律、政治、軍事等）、史地、語言文學、藝術、商業及金融、及休閒類等文章類型的圖書語料。
口語		2004 到 2018 年	2,230 萬	涵蓋哲學、宗教、科學、應用科學、社會科學、史地、語言文學、藝術、商業及金融、及休閒類等領域電視節目的語料。
新聞語料	中國時報	1999 到 2017 年	12 億 7,167 萬	報紙各版
	聯合報	1999 到 2017 年	14 億 3,683 萬	報紙各版
	國語日報	2010 到 2015 年	1,343 萬	報紙各版
	動腦雜誌	1999 到 2017 年	216 萬	
	遠見雜誌	2011 到 2017 年	88 萬	
	商業周刊	2011 到 2017 年	1,364 萬	
	upaper	2012 到 2017 年	205 萬	

考量本計畫語料庫的書面文本語料多來自書籍資料，口語文本語料主要來自於節目字幕，為避免單一書籍、節目數量導致詞頻分析偏差，因此書面語將從各書籍抽樣、口語將從各節目抽樣，進行統計，提供各類別語料計算出的詞彙及其出現率，並按照詞頻排序至此詞的累積頻率，以及累積百分比等詞頻相關統計資料。

統計資訊依序包括：該詞、書面語詞頻、書面語每百萬詞頻、書面語序位、口語詞頻、口語每百萬詞頻、口語序位、新聞詞頻、新聞每百萬詞頻、新聞序位、每百萬

詞頻（平均）等。書面語總收相異詞數 162,283 詞，每百萬詞頻 10 次（含）以上相異詞，共計 8,370 詞；口語總收相異詞數 95,917 詞，每百萬詞頻 10 次（含）以上相異詞，共計 5,787 詞；新聞總收相異詞數 163,258 詞，每百萬詞頻 10 次（含）以上相異詞，共計 9,990 詞。最後，書面語、口語及新聞語料每百萬詞頻（平均）10 次（含）以上，總詞數共計 8,277 詞。通用詞頻表前 1,000 詞示例，見表 12。詳細資訊詳見光碟檔案。

## 第三節 小結

通用詞頻表所分析之語料來源有三：一、2004 年至 2018 年的一般書面語料；二、2004 年至 2018 年的口語語料；三、新聞語料。書面語每百萬詞頻 10 次（含）以上者，共計 8,370 詞；口語每百萬詞頻 10 次（含）以上者，共計 5,787 詞；新聞每百萬詞頻 10 次（含）以上者，共計 9,990 詞。最後，華語文通用詞頻達每百萬詞頻（平均）10 次（含）以上，總詞數共計 8,277 詞。

表 12

通用詞頻表前 1,000 詞示例

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
1	的	63507.44	1	63558.08	1	36748.51	1	54604.67
2	是	18347.02	2	31103.07	2	11384.47	2	20278.19
3	一	14495.08	3	17168.79	5	7433.13	5	13032.34
4	個	9100.84	8	25784.74	3	3754.93	17	12880.17
5	在	12620.85	5	11892.11	10	10678.63	3	11730.53
6	我	11139.62	6	16568.63	7	3014.14	27	10240.80
7	有	8623.43	10	13615.29	9	8248.43	4	10162.39
8	就	6802.58	12	18464.25	4	4145.25	13	9804.03
9	了	12682.91	4	11789.38	11	4228.78	12	9567.02
10	這	7822.14	11	17101.34	6	3409.26	20	9444.25
11	不	8984.00	9	8836.79	14	6096.87	7	7972.56
12	他	9212.22	7	7550.77	17	4907.33	9	7223.44
13	也	6279.39	13	7451.08	18	6499.36	6	6743.28
14	我們	3641.24	25	14266.35	8	842.95	123	6250.18

(續下頁)



綜合 序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
15	你	5420.22	15	10214.86	13	663.69	157	5432.92
16	都	4454.02	19	7267.20	19	4067.53	14	5262.91
17	說	5072.02	17	5941.53	25	4678.75	10	5230.77
18	會	4016.40	20	7661.33	16	3220.14	22	4965.96
19	要	3336.26	29	8210.36	15	2945.43	30	4830.69
20	人	6107.38	14	3835.10	34	4344.65	11	4762.38
21	那	2850.93	34	10546.70	12	579.26	183	4658.96
22	對	3497.55	27	6593.06	22	2628.02	39	4239.54
23	很	2818.41	35	7110.38	20	1756.41	57	3895.06
24	上	4875.57	18	3212.47	41	3095.25	25	3727.76
25	來	2931.66	33	6032.02	24	1963.14	52	3642.27
26	與	3819.13	24	839.31	150	5436.84	8	3365.09
27	和	5284.36	16	1321.12	97	3355.16	21	3320.21
28	她	3854.07	23	2790.56	49	2761.07	33	3135.23
29	好	1760.22	61	6208.54	23	1331.66	74	3100.14
30	還	2441.81	39	3659.28	35	2682.92	37	2928.01
31	它	1215.40	87	7077.87	21	186.12	701	2826.46
32	可以	2212.78	44	5035.80	27	1023.14	96	2757.24
33	中	4000.54	21	1151.21	111	2811.18	32	2654.31
34	讓	2066.05	51	2817.53	48	3033.31	26	2638.96
35	年	2027.50	55	1945.91	63	3868.01	15	2613.81
36	到	2124.87	48	2934.25	46	2753.18	35	2604.10
37	為	2958.49	31	981.02	125	3655.02	19	2531.51
38	沒有	2329.99	41	3543.23	36	1260.73	77	2377.98
39	著	3982.05	22	1538.17	81	1572.72	63	2364.31
40	因為	1605.05	71	4671.45	28	813.54	126	2363.34
41	所以	858.64	116	5888.43	26	306.32	392	2351.13
42	大	2065.81	52	2230.65	54	2668.04	38	2321.50
43	去	1897.53	56	4298.85	30	665.84	153	2287.41

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
44	自己	2947.06	32	2486.26	52	1359.81	70	2264.38
45	把	2429.57	40	3364.78	39	953.72	104	2249.36
46	他們	2691.19	37	3211.58	42	753.61	134	2218.79
47	種	2777.81	36	2964.41	44	908.46	114	2216.89
48	做	1624.74	69	4005.47	32	931.51	111	2187.24
49	但	2212.03	45	629.11	177	3667.27	18	2169.47
50	而	3612.20	26	1119.12	114	1620.17	61	2117.16
51	看	1748.45	62	3452.22	38	961.06	102	2053.91
52	跟	1023.04	103	4448.68	29	681.28	151	2051.00
53	將	1764.94	60	459.29	233	3762.52	16	1995.58
54	最	1647.16	68	1936.99	64	2200.27	43	1928.14
55	以	2130.92	47	707.02	162	2873.11	31	1903.68
56	能	2299.07	42	988.86	123	2414.40	40	1900.77
57	之	3462.42	28	523.75	203	1681.75	58	1889.31
58	兩	2038.01	54	1657.55	75	1898.32	54	1864.63
59	後	1667.74	67	539.04	198	3214.88	24	1807.22
60	從	1886.79	57	1824.22	68	1647.98	59	1786.33
61	被	2133.80	46	1041.82	116	2145.51	47	1773.71
62	什麼	1699.59	65	3158.43	43	424.78	266	1760.94
63	時候	792.68	126	4225.57	31	147.68	917	1721.98
64	時	2464.43	38	338.73	305	2213.59	42	1672.25
65	得	2270.91	43	1622.75	77	1104.16	92	1665.94
66	地	2987.36	30	684.93	164	1229.86	81	1634.05
67	等	1280.57	82	356.47	293	3217.12	23	1618.05
68	這樣	1107.42	98	3355.17	40	346.99	341	1603.19
69	又	2095.60	49	1530.66	82	1135.34	88	1587.20
70	再	1267.32	84	1982.03	60	1506.16	64	1585.17
71	每	1268.06	83	1651.41	76	1829.95	56	1583.14
72	多	1252.80	85	1304.89	101	2170.77	45	1576.15

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
73	用	1479.42	74	2279.34	53	919.87	112	1559.54
74	次	1312.73	78	1435.34	90	1927.46	53	1558.51
75	才	1558.70	72	1503.50	84	1579.02	62	1547.07
76	然後	573.26	177	3994.59	33	71.61	1927	1546.49
77	更	1415.22	75	1180.90	109	2026.02	50	1540.72
78	想	1774.33	59	1698.22	74	876.85	117	1449.80
79	並	1744.71	63	391.04	264	2181.07	44	1438.94
80	及	1203.10	89	119.62	808	2973.40	28	1432.04
81	其實	391.74	274	3455.32	37	283.42	433	1376.83
82	台灣	502.75	208	879.51	142	2722.90	36	1368.38
83	裡	2062.37	53	1255.49	105	708.79	145	1342.22
84	覺得	724.95	135	2959.48	45	338.84	347	1341.09
85	現在	804.77	125	2589.65	50	585.31	181	1326.58
86	天	1300.14	79	1351.65	95	1304.78	75	1318.86
87	只	1709.56	64	859.06	145	1380.90	69	1316.50
88	或	1886.40	58	471.35	228	1484.62	65	1280.79
89	過	1614.63	70	1269.05	102	948.19	106	1277.29
90	像	1131.35	94	2150.62	56	501.40	218	1261.12
91	前	1068.58	100	594.86	184	2092.01	48	1251.82
92	可能	1067.92	101	1792.93	69	838.12	124	1232.99
93	所	2081.43	50	748.77	159	847.00	122	1225.74
94	月	588.70	172	327.90	313	2755.19	34	1223.93
95	小	1239.70	86	1308.03	99	1113.36	91	1220.36
96	知道	1357.85	76	1782.47	71	404.41	286	1181.58
97	可	1200.43	91	236.52	441	2082.00	49	1172.98
98	啊	530.70	196	2871.43	47	116.67	1210	1172.93
99	下	1294.65	80	856.52	146	1335.68	73	1162.29
100	沒	1195.31	92	1176.63	110	1046.99	95	1139.65
101	表示	294.61	382	164.09	605	2950.77	29	1136.49

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
102	如果	1107.58	97	1748.51	72	521.09	202	1125.73
103	家	932.12	111	1265.53	103	1092.71	93	1096.78
104	嗎	984.98	105	1973.16	62	312.72	380	1090.29
105	比較	441.39	239	2556.25	51	224.40	563	1074.01
106	這些	1175.58	93	1607.78	79	420.61	271	1067.99
107	問題	1100.15	99	1136.48	113	945.74	107	1060.79
108	新	830.62	121	470.46	229	1880.38	55	1060.48
109	開始	969.53	106	1522.31	83	656.73	159	1049.52
110	者	1673.67	66	308.15	331	1135.61	87	1039.14
111	給	948.36	109	1475.82	85	606.36	175	1010.18
112	卻	1315.84	77	435.65	242	1248.83	78	1000.11
113	各	954.56	108	602.74	182	1424.80	66	994.03
114	大家	414.66	259	2050.14	59	515.41	205	993.40
115	三	1203.14	88	774.48	157	998.87	99	992.16
116	但是	857.77	117	1895.06	65	186.08	702	979.63
117	位	1126.89	95	935.94	131	856.53	119	973.12
118	呢	811.31	123	1977.57	61	122.46	1140	970.45
119	已	850.69	118	85.52	1082	1966.05	51	967.42
120	很多	435.51	245	2070.31	58	378.27	308	961.37
121	全	696.80	141	539.65	197	1643.91	60	960.12
122	一些	618.32	164	2091.37	57	162.09	820	957.26
123	非常	600.86	170	1885.06	66	357.65	326	947.86
124	已經	968.87	107	1439.75	88	414.79	273	941.14
125	吃	557.15	184	1733.17	73	503.10	216	931.14
126	元	237.17	483	61.97	1416	2396.86	41	898.67
127	時間	839.66	119	1021.13	119	803.81	128	888.20
128	事	1519.57	73	530.64	200	607.65	173	885.95
129	於	1291.62	81	202.09	510	1159.37	83	884.36
130	工作	868.40	115	938.24	130	809.31	127	871.98

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
131	看到	624.17	163	1548.25	80	440.20	256	870.88
132	日	367.59	295	57.37	1521	2159.54	46	861.50
133	當	1037.19	102	882.18	140	636.46	168	851.94
134	地方	550.45	188	1449.04	87	508.81	212	836.10
135	真的	345.12	314	1872.54	67	239.98	516	819.21
136	孩子	942.58	110	957.85	129	531.34	199	810.59
137	裡面	160.12	705	2221.92	55	42.71	3134	808.25
138	則	1003.77	104	165.17	602	1243.17	79	804.04
139	生活	837.86	120	926.98	134	622.16	170	795.67
140	希望	470.18	217	845.41	149	1058.79	94	791.46
141	一樣	745.43	131	1305.50	100	306.66	391	785.86
142	內	722.85	136	279.07	371	1296.56	76	766.16
143	第一	654.54	153	885.98	139	752.18	135	764.23
144	由	809.48	124	248.90	418	1194.22	82	750.87
145	不會	684.64	145	1025.82	118	510.00	209	740.15
146	還是	605.79	169	1224.34	106	386.59	300	738.91
147	講	276.25	410	1789.93	70	141.27	968	735.82
148	歲	513.71	204	514.08	205	1151.37	84	726.39
149	太	714.46	137	963.43	127	487.64	223	721.84
150	之後	742.18	132	1141.55	112	280.45	439	721.39
151	此	1111.31	96	74.26	1216	976.06	101	720.54
152	東西	552.34	186	1461.47	86	121.96	1147	711.92
153	發現	692.52	143	681.50	165	744.83	136	706.28
154	可是	431.08	249	1612.14	78	57.21	2411	700.15
155	先	530.91	195	997.35	122	562.50	187	696.92
156	老師	345.41	313	1353.63	93	388.86	298	695.96
157	怎麼	561.59	182	1313.19	98	208.79	598	694.53
158	名	537.79	194	171.17	583	1341.85	72	683.60
159	幾	890.90	113	647.54	171	509.43	211	682.62

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
160	高	506.84	206	592.14	185	940.77	109	679.92
161	妳	547.74	189	1352.92	94	138.29	990	679.65
162	應該	651.43	154	1014.89	120	359.95	324	675.42
163	向	912.01	112	162.03	613	913.92	113	662.65
164	認為	758.96	128	288.73	357	885.73	116	644.47
165	因	459.87	224	59.58	1478	1403.00	68	640.82
166	重要	611.13	168	814.96	154	479.72	228	635.27
167	而且	684.18	146	1036.28	117	174.18	755	631.55
168	學生	453.67	233	313.45	324	1119.44	89	628.85
169	你們	400.27	266	1381.21	91	79.63	1752	620.37
170	媽媽	270.82	421	1187.24	108	350.01	334	602.69
171	美國	569.93	178	283.80	361	952.51	105	602.08
172	其	1200.75	90	55.54	1555	535.34	198	597.21
173	需要	658.86	150	718.89	160	405.46	285	594.40
174	今天	272.16	414	1263.89	104	246.99	501	594.35
175	本	750.18	130	339.25	304	690.89	147	593.44
176	一定	545.84	190	901.09	136	333.03	356	593.32
177	比	670.53	149	447.89	239	653.53	161	590.65
178	中國	829.97	122	230.70	447	709.96	144	590.21
179	政府	362.82	299	268.46	386	1113.84	90	581.71
180	水	337.46	321	931.72	133	425.11	264	564.76
181	社會	694.23	142	356.98	290	638.78	167	563.33
182	世界	649.36	155	491.24	218	548.53	193	563.04
183	至	397.27	268	49.58	1695	1231.24	80	559.36
184	今年	40.12	2643	217.52	470	1410.80	67	556.15
185	的話	287.10	396	1331.72	96	49.46	2756	556.09
186	活動	337.05	323	328.98	310	1000.53	98	555.52
187	公司	657.02	151	153.35	646	854.05	120	554.81
188	不過	635.12	159	590.31	186	422.46	270	549.30

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
189	成為	568.40	180	298.86	341	766.84	132	544.70
190	一直	469.03	219	873.65	143	291.37	414	544.68
191	不要	392.48	273	932.19	132	305.24	394	543.30
192	民眾	67.28	1685	193.36	530	1347.66	71	536.10
193	其他	735.62	133	379.59	273	492.47	221	535.89
194	不同	628.40	162	502.54	212	471.96	233	534.30
195	一下	321.66	342	1196.06	107	84.56	1658	534.09
196	書	683.30	147	459.10	234	454.88	243	532.43
197	吧	731.49	134	713.36	161	147.04	924	530.63
198	關係	642.60	158	466.09	231	480.60	227	529.77
199	喔	95.91	1181	1438.95	89	41.33	3222	525.40
200	能夠	379.13	283	1014.75	121	175.52	746	523.13
201	曾	519.83	201	95.74	977	940.76	110	518.78
202	們	556.85	185	553.21	196	441.52	254	517.19
203	不能	642.81	157	496.21	215	410.30	278	516.44
204	事情	438.94	243	982.52	124	119.81	1166	513.76
205	最後	521.33	200	451.04	238	557.57	190	509.98
206	無	618.26	165	176.66	572	726.09	141	507.00
207	整	410.25	261	823.36	151	286.16	430	506.59
208	心	707.45	139	508.13	210	294.05	411	503.21
209	發展	462.76	222	267.39	390	776.93	130	502.36
210	該	576.57	176	193.64	529	733.25	139	501.15
211	國家	429.50	251	361.67	288	710.41	143	500.53
212	愛	485.29	213	501.23	213	507.84	213	498.12
213	因此	655.03	152	247.45	419	587.95	178	496.81
214	方式	526.90	198	511.08	207	449.22	249	495.73
215	帶	416.01	258	637.27	176	432.19	260	495.16
216	一點	358.81	305	976.80	126	141.22	969	492.28
217	許多	611.45	167	274.89	377	585.98	180	490.77

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
218	這麼	458.02	226	862.39	144	151.08	895	490.50
219	必須	673.35	148	376.36	274	409.67	279	486.46
220	件	551.88	187	504.56	211	400.45	288	485.63
221	那麼	487.90	211	852.68	148	113.53	1238	484.70
222	一起	313.98	350	759.84	158	378.32	307	484.05
223	為什麼	395.81	271	960.29	128	85.98	1634	480.69
224	雖然	539.67	192	482.14	222	413.71	274	478.50
225	發生	510.30	205	458.17	235	462.48	240	476.98
226	特別	355.40	307	639.85	174	435.52	258	476.92
227	起	381.24	282	171.08	584	869.39	118	473.90
228	請	457.73	227	571.17	190	378.92	305	469.27
229	啦	203.62	551	1117.72	115	78.56	1771	466.63
230	走	503.30	207	575.35	189	320.12	372	466.25
231	未	400.92	265	35.93	2181	959.16	103	465.34
232	二	569.66	179	141.43	699	684.04	149	465.04
233	影響	377.45	286	431.62	244	579.00	185	462.69
234	環境	279.34	408	637.83	175	470.82	235	462.66
235	文化	454.29	232	280.29	368	650.56	163	461.71
236	甚至	478.19	216	390.19	265	513.32	207	460.57
237	這樣子	13.46	6587	1361.22	92	4.75	16728	459.81
238	當然	401.40	264	818.29	152	159.09	843	459.59
239	大陸	105.87	1082	121.87	799	1150.48	85	459.41
240	問	646.44	156	491.94	216	233.56	533	457.31
241	這裡	384.46	280	880.45	141	93.45	1504	452.78
242	喜歡	449.79	235	647.21	172	253.72	480	450.24
243	如	754.29	129	77.54	1172	514.40	206	448.74
244	仍	363.30	298	37.86	2077	944.01	108	448.39
245	大學	330.09	332	228.59	455	776.32	131	445.00
246	研究	712.82	138	315.84	322	304.90	395	444.52

(續下頁)



綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
247	以及	612.30	166	268.28	388	449.85	247	443.47
248	目前	181.94	614	294.36	352	853.67	121	443.32
249	部分	325.53	339	558.51	194	440.50	255	441.51
250	外	430.92	250	152.60	653	739.02	137	440.85
251	無法	595.16	171	141.48	697	582.21	182	439.62
252	點	469.06	218	302.10	337	546.81	194	439.32
253	快	455.57	231	510.33	208	345.83	342	437.24
254	話	580.30	173	568.17	191	161.60	823	436.69
255	四	540.67	191	344.36	301	424.95	265	436.66
256	手	576.72	175	402.53	254	327.73	362	435.66
257	使用	390.38	275	356.89	291	549.03	192	432.10
258	日本	405.91	262	229.29	453	654.23	160	429.81
259	所有	567.35	181	407.46	253	299.23	404	424.68
260	以後	281.63	400	888.98	138	97.28	1448	422.63
261	透過	240.99	472	362.19	286	661.73	158	421.64
262	未來	212.39	532	249.09	417	801.20	129	420.89
263	見	704.46	140	163.71	608	391.60	295	419.92
264	昨天	47.37	2293	68.11	1309	1139.93	86	418.47
265	只要	371.85	289	480.12	223	391.89	294	414.62
266	台	113.67	1015	117.65	818	1012.43	97	414.58
267	正	687.02	144	152.97	650	396.05	289	412.01
268	進行	396.61	270	231.87	446	607.40	174	411.96
269	學校	292.42	387	422.98	248	518.58	203	411.33
270	除了	276.50	409	475.76	226	478.84	230	410.37
271	找	426.19	254	485.70	220	312.67	381	408.19
272	叫	441.13	240	653.36	170	127.08	1094	407.19
273	如何	579.26	174	291.22	355	348.92	336	406.47
274	包括	292.69	384	280.90	366	644.92	165	406.17
275	經濟	326.97	337	231.97	445	653.02	162	403.99

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
276	分	241.78	469	298.53	343	664.04	156	401.45
277	另	525.22	199	92.74	1008	575.90	186	397.95
278	當時	458.80	225	355.15	296	378.58	306	397.51
279	長	419.05	257	446.06	241	324.45	366	396.52
280	好像	263.06	431	853.99	147	67.94	2038	395.00
281	出現	493.18	210	267.62	389	423.80	267	394.86
282	造成	248.91	452	466.56	230	465.42	238	393.63
283	過程	346.65	311	529.14	201	303.66	399	393.15
284	感覺	359.56	303	645.90	173	169.10	780	391.52
285	些	633.27	160	307.21	333	225.30	558	388.59
286	受	431.75	248	153.21	648	579.01	184	387.99
287	結果	427.51	252	394.46	261	340.23	344	387.40
288	指出	152.01	745	20.83	3362	988.00	100	386.95
289	若	344.71	315	59.67	1475	755.16	133	386.51
290	便	871.76	114	58.07	1503	225.45	557	385.09
291	後來	335.57	325	587.50	188	224.53	561	382.53
292	機會	321.10	344	342.82	302	480.62	226	381.51
293	朋友	397.22	269	487.44	219	257.47	475	380.71
294	過去	340.09	319	325.60	314	475.41	232	380.36
295	寫	448.77	236	477.92	224	214.36	582	380.35
296	對於	369.19	291	420.55	249	349.98	335	379.91
297	牠	271.30	418	787.61	155	79.29	1759	379.40
298	學習	411.56	260	391.55	263	330.05	359	377.72
299	聽	538.81	193	395.54	258	198.08	646	377.48
300	或者	242.90	467	816.51	153	53.85	2553	371.09
301	條	455.66	230	349.95	299	305.82	393	370.47
302	大概	145.63	781	921.02	135	44.03	3064	370.23
303	場	271.80	417	150.53	661	681.73	150	368.02
304	國	286.24	397	142.70	691	673.75	152	367.56

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
305	幫	230.14	499	653.40	168	216.11	577	366.55
306	間	457.31	228	197.82	521	443.67	253	366.27
307	放	329.52	334	567.65	192	192.82	675	363.33
308	約	186.80	599	73.18	1231	822.11	125	360.70
309	邊	486.91	212	298.67	342	289.72	418	358.43
310	提供	287.14	395	165.40	601	622.05	171	358.20
311	故事	395.42	272	400.00	255	276.57	446	357.33
312	應	322.94	341	20.78	3366	726.62	140	356.78
313	變成	289.13	391	598.43	183	169.73	774	352.43
314	連	354.77	308	246.70	421	452.52	246	351.33
315	張	389.54	277	279.82	370	383.00	301	350.79
316	買	235.05	487	457.37	236	359.50	325	350.64
317	一般	309.01	358	510.28	209	230.24	542	349.84
318	五	431.81	247	281.74	364	334.78	353	349.44
319	其中	359.48	304	172.91	577	515.94	204	349.44
320	先生	630.52	161	280.38	367	134.47	1019	348.46
321	或是	296.32	379	587.73	187	158.51	848	347.52
322	同時	439.20	241	213.06	479	389.45	296	347.24
323	使	776.15	127	69.75	1278	194.51	666	346.81
324	去年	34.16	3048	91.57	1018	903.56	115	343.10
325	決定	402.02	263	202.37	508	423.29	269	342.56
326	教育	279.97	407	302.38	335	444.68	252	342.34
327	市場	255.03	443	107.19	891	664.53	155	342.25
328	中心	181.53	616	157.33	627	685.54	148	341.47
329	告訴	483.94	214	395.26	260	141.65	962	340.28
330	國際	160.34	702	155.08	641	704.31	146	339.91
331	完全	456.60	229	321.24	317	238.99	520	338.94
332	主要	308.12	359	395.64	257	310.65	384	338.14
333	身	481.77	215	262.08	399	268.17	462	337.34

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
334	狀況	174.75	649	566.95	193	265.18	466	335.63
335	半	312.42	354	224.74	460	467.79	237	334.98
336	第二	343.20	316	368.90	279	289.40	419	333.83
337	身體	385.09	279	415.62	251	196.30	658	332.34
338	錢	352.43	309	365.43	282	278.16	443	332.01
339	容易	319.88	345	472.61	227	198.43	641	330.31
340	段	379.12	284	345.30	300	263.09	470	329.17
341	以前	228.24	503	617.52	178	139.90	978	328.55
342	您	299.81	369	615.60	180	67.11	2062	327.51
343	健康	196.46	567	430.63	245	348.39	338	325.16
344	同	397.43	267	134.87	729	433.26	259	321.85
345	要求	280.89	404	126.09	782	558.56	189	321.85
346	項	351.21	310	126.70	776	487.51	224	321.81
347	爸爸	167.39	671	653.36	169	140.16	976	320.30
348	生命	361.93	301	385.93	270	208.58	599	318.81
349	服務	149.24	765	182.15	554	619.13	172	316.84
350	計畫	215.07	524	138.15	710	595.58	176	316.26
351	起來	438.81	244	389.44	266	116.38	1216	314.88
352	誰	422.07	256	320.77	318	198.76	640	313.87
353	設計	232.29	493	199.37	518	505.02	214	312.22
354	歷史	374.68	287	204.90	499	356.37	328	311.98
355	這邊	31.42	3263	893.25	137	10.82	9458	311.83
356	往	377.66	285	355.76	295	202.06	628	311.83
357	任何	515.19	203	190.69	537	228.46	549	311.44
358	之間	466.77	220	313.12	325	145.99	933	308.63
359	且	261.13	436	14.78	4371	647.67	164	307.86
360	花	284.68	398	296.94	350	340.13	345	307.25
361	改變	362.53	300	334.23	307	223.39	565	306.72
362	打	301.49	367	281.60	365	336.96	351	306.68

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
363	只是	370.35	290	361.20	289	188.17	689	306.58
364	非	441.49	238	68.21	1305	408.50	281	306.07
365	自然	361.88	302	342.53	303	213.60	587	306.01
366	常	338.31	320	204.62	502	374.40	311	305.78
367	變	316.36	348	388.65	267	198.14	645	301.05
368	別人	426.90	253	369.51	278	105.97	1338	300.79
369	電影	450.42	234	70.55	1270	381.16	303	300.71
370	完成	238.49	477	203.87	505	458.50	241	300.29
371	拿	272.10	415	371.15	275	256.02	477	299.76
372	人員	194.38	575	159.87	618	543.89	195	299.38
373	路	292.44	386	282.72	362	312.03	382	295.73
374	隻	369.16	292	289.90	356	227.33	552	295.46
375	剛剛	80.74	1385	783.49	156	15.30	7251	293.18
376	真	530.25	197	161.70	614	185.74	703	292.56
377	道	559.29	183	152.83	651	163.02	808	291.72
378	繼續	318.15	346	265.79	394	290.23	416	291.39
379	想要	248.05	454	511.69	206	113.70	1237	291.15
380	令	502.61	209	79.28	1148	291.27	415	291.05
381	難	298.66	375	298.20	346	275.36	451	290.74
382	運動	247.20	458	246.32	423	373.83	312	289.12
383	開	245.38	462	274.28	378	347.23	339	288.96
384	接受	333.60	327	171.74	581	360.69	323	288.68
385	家庭	237.31	480	317.63	319	306.89	390	287.27
386	選擇	295.38	380	244.68	427	320.16	371	286.74
387	另外	165.48	680	500.95	214	192.82	674	286.42
388	參加	168.60	667	183.60	550	503.62	215	285.28
389	處	332.58	328	62.62	1406	458.32	242	284.51
390	代表	230.46	497	205.93	496	416.02	272	284.14
391	經驗	341.91	317	256.17	408	251.80	487	283.30

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
392	台北	86.18	1296	89.46	1040	664.64	154	280.09
393	準備	255.57	442	301.30	339	279.09	441	278.65
394	之前	312.93	352	350.18	298	172.35	761	278.49
395	經過	241.64	470	364.16	285	225.85	556	277.22
396	處理	243.09	466	218.13	469	370.10	316	277.11
397	增加	190.54	585	187.59	542	452.54	245	276.89
398	合作	141.32	804	95.74	976	592.90	177	276.65
399	是否	309.56	356	50.62	1672	468.67	236	276.28
400	回	363.51	297	170.28	588	294.42	410	276.07
401	上面	130.29	865	672.36	167	24.02	5064	275.55
402	幫助	247.48	457	370.59	276	206.43	614	274.83
403	塊	214.26	526	491.28	217	117.23	1202	274.26
404	成	298.72	374	116.90	828	406.07	284	273.90
405	能力	368.12	293	197.11	524	255.35	478	273.53
406	滿	182.36	611	382.03	272	253.55	481	272.65
407	第	105.49	1089	1.83	18866	710.55	142	272.62
408	努力	246.84	460	284.27	360	281.00	436	270.70
409	較	309.13	357	44.99	1821	454.09	244	269.40
410	老	287.45	394	221.98	465	298.48	405	269.30
411	小朋友	38.61	2735	609.78	181	157.86	853	268.75
412	小時	192.02	583	229.29	452	382.49	302	267.94
413	政治	312.63	353	60.98	1439	430.06	262	267.89
414	首	176.08	646	66.52	1334	556.45	191	266.35
415	支持	167.44	670	152.27	655	477.44	231	265.71
416	頭	444.57	237	164.09	604	182.98	713	263.88
417	企業	255.59	441	63.33	1394	471.50	234	263.47
418	總統	91.39	1225	57.98	1505	639.27	166	262.88
419	那些	439.17	242	275.50	375	73.18	1885	262.62
420	方法	389.62	276	311.62	328	86.51	1618	262.58

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
421	進入	261.94	434	174.79	575	350.90	333	262.54
422	美	207.83	540	155.27	639	423.62	268	262.24
423	臉	341.14	318	90.07	1034	353.30	331	261.50
424	車	186.49	600	117.32	824	479.48	229	261.10
425	更多	174.12	651	275.50	376	333.04	355	260.89
426	產生	327.15	336	298.30	345	156.94	855	260.80
427	情況	364.12	296	201.66	512	214.39	581	260.06
428	醫院	121.54	937	245.24	425	412.47	277	259.75
429	相當	262.16	432	179.62	561	335.70	352	259.16
430	如此	516.24	202	69.10	1288	191.31	681	258.88
431	即	435.19	246	16.98	3928	322.82	368	258.33
432	安全	139.26	815	186.18	544	449.03	250	258.16
433	傳統	235.96	484	200.68	514	334.28	354	256.97
434	社區	88.15	1263	282.11	363	400.49	287	256.92
435	利用	192.08	582	291.26	354	286.28	429	256.54
436	作品	313.83	351	117.46	822	338.11	350	256.47
437	近	166.25	677	81.01	1130	521.88	201	256.38
438	剛	209.30	538	369.88	277	189.59	685	256.26
439	面對	241.10	471	226.43	457	298.31	406	255.28
440	成功	315.50	349	171.55	582	276.08	448	254.37
441	藝術	317.87	347	116.90	827	325.51	365	253.43
442	哪	250.21	447	384.71	271	118.29	1187	251.07
443	看看	180.71	621	525.39	202	46.49	2932	250.86
444	片	281.13	403	230.42	449	239.78	517	250.44
445	直接	216.26	522	297.50	349	235.76	529	249.84
446	十	298.74	373	256.36	407	193.47	670	249.52
447	相關	163.95	685	87.39	1057	497.19	220	249.51
448	父親	329.77	333	155.74	635	262.40	471	249.30
449	地區	159.36	708	214.10	478	371.44	315	248.30

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
450	不斷	270.37	422	183.04	551	288.74	423	247.38
451	出	330.16	331	181.26	557	227.66	551	246.36
452	昨	1.90	27838	0.09	97204	735.90	138	245.97
453	越	272.88	412	312.61	327	151.93	889	245.81
454	加上	148.89	768	181.82	555	406.07	283	245.59
455	學	189.19	589	356.70	292	190.27	683	245.39
456	某	465.96	221	123.79	790	142.86	956	244.21
457	超過	120.27	955	175.07	574	437.05	257	244.13
458	就是說	19.64	4829	706.88	163	5.29	15632	243.94
459	人生	268.85	425	268.75	385	193.43	671	243.68
460	謝謝	63.07	1800	617.24	179	50.69	2700	243.67
461	系統	265.09	429	158.46	623	307.19	388	243.58
462	方面	297.12	378	266.78	393	165.11	794	243.00
463	坐	367.71	294	214.28	477	143.66	948	241.88
464	空間	206.94	545	179.01	564	339.67	346	241.87
465	少	198.26	562	284.32	359	242.56	509	241.72
466	久	268.88	424	279.02	372	176.58	743	241.49
467	到底	162.86	696	477.17	225	84.21	1665	241.41
468	臺灣	141.99	800	553.58	195	26.01	4757	240.53
469	產業	116.94	992	99.59	941	502.88	217	239.80
470	幾乎	290.52	389	205.51	497	221.23	568	239.09
471	所謂	221.69	514	416.23	250	76.78	1806	238.23
472	辦法	171.31	659	429.50	247	111.04	1271	237.29
473	簡單	204.85	549	387.94	268	118.40	1185	237.06
474	嚴重	155.10	731	268.32	387	287.63	427	237.02
475	照顧	115.25	1002	328.41	312	264.09	467	235.92
476	事實	303.20	364	280.05	369	123.23	1133	235.49
477	字	298.96	372	257.39	404	147.73	916	234.69
478	精神	311.53	355	182.57	553	209.73	595	234.61

(續下頁)



綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
479	住	274.81	411	222.07	464	204.23	621	233.70
480	醫師	117.36	989	329.26	309	249.99	495	232.20
481	口	330.83	330	125.76	783	239.38	518	231.99
482	提出	249.21	450	52.07	1635	394.35	291	231.87
483	政策	122.06	936	60.84	1444	509.73	210	230.88
484	就是	11.54	7457	672.97	166	6.18	14089	230.23
485	常常	152.13	744	484.11	221	53.29	2585	229.84
486	小孩	185.11	602	321.33	316	182.13	722	229.53
487	強調	142.44	796	63.33	1393	481.22	225	229.00
488	得到	334.65	326	230.19	450	121.13	1157	228.66
489	相信	322.96	340	206.17	494	155.91	861	228.35
490	由於	303.79	362	81.06	1129	299.83	403	228.23
491	出來	199.24	561	397.47	256	84.25	1664	226.99
492	六	297.22	377	151.38	659	231.60	539	226.73
493	受到	262.02	433	204.85	500	213.19	588	226.69
494	死	461.01	223	91.10	1022	127.14	1093	226.42
495	舉辦	47.70	2277	44.47	1838	586.20	179	226.13
496	行動	272.03	416	132.90	740	273.10	455	226.01
497	喝	226.79	505	291.40	353	158.57	847	225.59
498	站	267.69	427	136.88	717	272.05	456	225.54
499	知	425.72	255	52.49	1624	198.18	644	225.46
500	原因	263.12	430	227.14	456	185.53	704	225.26
501	時代	291.88	388	155.74	634	227.81	550	225.14
502	靠	201.89	555	224.79	459	248.54	497	225.07
503	句	346.28	312	211.66	485	116.17	1218	224.70
504	獲得	216.49	520	91.47	1019	362.92	321	223.63
505	參與	159.42	707	131.07	746	379.34	304	223.28
506	兩岸	27.26	3669	19.42	3546	622.39	169	223.02
507	清楚	301.58	366	196.27	526	169.15	778	222.33

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
508	慢慢	141.23	805	454.55	237	69.52	1995	221.77
509	共	125.59	911	28.24	2609	511.11	208	221.65
510	找到	257.62	439	238.44	438	162.11	819	219.39
511	解決	219.25	518	170.47	587	268.41	461	219.38
512	尤其	189.12	590	254.16	412	214.48	580	219.25
513	母親	321.51	343	98.70	951	235.42	530	218.54
514	土地	126.61	900	269.36	381	253.45	482	216.47
515	現場	62.10	1820	179.80	560	406.96	282	216.29
516	父母	303.18	365	184.03	548	159.40	838	215.54
517	成長	166.91	673	161.37	615	317.62	374	215.30
518	基本	201.82	556	302.10	338	140.76	972	214.89
519	個人	299.49	370	104.61	906	240.24	515	214.78
520	表現	265.54	428	103.44	915	275.10	452	214.69
521	全球	82.18	1355	199.13	519	362.63	322	214.65
522	玩	150.42	754	308.67	330	182.16	721	213.75
523	真正	301.37	368	207.67	489	131.22	1048	213.42
524	有時候	84.90	1309	533.55	199	20.19	5835	212.88
525	管理	212.58	531	93.30	1001	332.74	357	212.87
526	調查	136.92	829	107.89	889	393.35	292	212.72
527	兒子	237.29	481	142.61	693	256.80	476	212.23
528	討論	232.32	492	130.78	748	273.28	454	212.13
529	行為	384.27	281	93.77	997	158.23	849	212.09
530	回來	178.82	635	387.90	269	69.23	2005	211.98
531	分鐘	207.50	543	180.13	559	247.44	499	211.69
532	聲音	337.07	322	186.28	543	110.96	1272	211.44
533	不少	138.49	820	63.80	1388	431.97	261	211.42
534	建議	179.05	634	127.27	770	325.93	364	210.75
535	同學	119.18	971	338.08	306	173.60	759	210.29
536	早	240.79	473	185.81	545	204.12	622	210.24

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
537	原來	176.99	642	353.14	297	99.63	1419	209.92
538	以上	136.58	833	142.56	694	348.48	337	209.20
539	指	331.80	329	85.61	1079	209.71	596	209.04
540	了解	326.02	338	68.68	1297	230.17	543	208.29
541	人類	303.67	363	242.85	429	75.79	1828	207.44
542	產品	200.43	557	114.55	841	302.09	401	205.69
543	認識	202.68	554	254.16	411	159.26	839	205.37
544	資源	108.44	1060	222.73	463	283.14	435	204.77
545	持續	133.04	850	110.00	869	368.89	318	203.98
546	怕	184.37	603	267.34	392	158.83	844	203.51
547	組織	272.63	413	145.80	671	191.82	680	203.42
548	力量	235.38	486	211.89	482	162.86	810	203.38
549	協助	113.05	1021	109.63	871	386.91	299	203.20
550	女兒	180.35	624	157.71	625	271.02	458	203.03
551	底	147.10	774	87.25	1059	374.56	310	202.97
552	擔心	162.01	697	201.05	513	243.82	505	202.29
553	夠	193.91	577	263.77	397	145.36	937	201.02
554	技術	187.18	596	135.43	724	279.97	440	200.86
555	僅	120.08	958	17.59	3823	462.92	239	200.20
556	畫	227.57	504	269.12	382	102.60	1385	199.77
557	推出	39.47	2680	18.01	3752	541.05	196	199.51
558	食物	165.05	681	314.01	323	117.85	1192	198.97
559	送	128.84	878	142.89	689	321.78	370	197.83
560	度	98.35	1157	182.62	552	310.98	383	197.32
561	原本	137.16	825	170.52	586	283.41	434	197.03
562	怎麼樣	61.09	1845	520.04	204	9.84	10153	196.99
563	業者	26.86	3718	35.18	2215	528.81	200	196.95
564	雖	187.03	597	11.82	5146	389.40	297	196.08
565	而是	292.51	385	135.10	727	160.57	830	196.06

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
566	不僅	166.31	676	98.65	952	322.59	369	195.85
567	一切	372.69	288	93.35	1000	121.30	1154	195.78
568	教	163.42	692	322.88	315	100.65	1410	195.65
569	目標	245.94	461	98.42	954	242.07	511	195.47
570	人家	99.16	1150	446.67	240	39.74	3334	195.19
571	哪裡	157.75	715	365.24	283	61.37	2260	194.79
572	依	142.27	798	25.00	2873	413.41	275	193.56
573	治療	175.82	648	206.73	493	195.15	662	192.57
574	遠	226.67	506	132.90	741	217.93	574	192.50
575	壓力	159.05	710	204.24	504	212.63	590	191.97
576	多少	198.15	563	273.72	379	101.92	1400	191.26
577	年輕	189.23	588	168.45	591	215.54	578	191.07
578	言	385.46	278	28.38	2599	159.24	840	191.02
579	願意	193.34	580	215.22	476	164.45	798	191.00
580	回到	230.78	496	205.32	498	136.70	998	190.94
581	音樂	189.66	586	129.24	759	253.09	483	190.66
582	民進黨	10.85	7854	0.23	63916	560.31	188	190.46
583	中央	180.06	627	35.79	2186	353.46	330	189.77
584	根本	261.48	435	137.45	711	169.75	773	189.56
585	山	213.28	530	220.24	468	135.10	1015	189.54
586	敢	260.06	437	146.26	670	160.94	828	189.09
587	資料	206.40	546	133.79	737	227.05	553	189.08
588	加	110.10	1044	330.43	308	126.34	1098	188.95
589	當中	111.62	1034	409.94	252	45.06	2999	188.87
590	錯	182.91	606	328.88	311	53.35	2580	188.38
591	價值	231.51	495	138.15	709	194.78	664	188.15
592	只有	163.12	694	298.91	340	101.96	1398	188.00
593	自	222.55	510	28.90	2559	309.95	385	187.13
594	投資	150.25	756	37.39	2102	373.25	313	186.96

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
595	市長	17.55	5268	7.46	7259	535.66	197	186.89
596	事件	232.24	494	77.45	1173	250.59	494	186.76
597	曾經	206.11	548	263.21	398	90.66	1545	186.66
598	動物	144.85	788	253.03	413	159.62	835	185.84
599	意義	281.29	402	144.72	679	130.58	1054	185.53
600	於是	357.92	306	102.87	918	92.40	1519	184.40
601	網路	103.76	1099	93.16	1005	354.26	329	183.73
602	習慣	166.25	678	254.20	410	129.91	1061	183.45
603	共同	144.06	790	110.89	862	295.35	408	183.43
604	媒體	125.75	909	72.05	1247	352.40	332	183.40
605	歡迎	76.66	1467	269.03	383	204.26	620	183.32
606	分享	76.70	1466	267.39	391	205.24	619	183.11
607	離開	281.45	401	134.40	733	130.87	1051	182.24
608	生	237.35	479	138.90	705	170.27	770	182.18
609	意思	199.52	560	313.12	326	33.44	3851	182.03
610	經	145.21	783	43.49	1874	356.78	327	181.82
611	女	199.88	559	40.20	1985	303.91	398	181.33
612	國民黨	44.90	2411	1.64	20145	497.23	219	181.26
613	來自	164.04	684	128.72	764	250.90	492	181.22
614	推動	58.49	1914	90.25	1029	394.64	290	181.13
615	作為	224.22	507	71.49	1259	247.40	500	181.04
616	環保	15.92	5686	368.48	280	158.18	850	180.86
617	志工	8.73	9297	361.81	287	171.85	763	180.80
618	似乎	336.36	324	104.61	907	100.63	1411	180.53
619	座	197.46	565	123.70	792	219.76	570	180.31
620	想到	188.81	591	265.51	395	86.32	1623	180.21
621	擔任	136.91	830	64.41	1375	338.65	348	179.99
622	份	210.08	535	187.64	541	141.42	964	179.71
623	成立	112.16	1028	111.13	861	314.17	377	179.15

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
624	聽到	238.69	476	191.49	536	106.19	1335	178.79
625	別	299.42	371	108.31	885	128.47	1075	178.73
626	家人	120.63	951	199.98	516	215.53	579	178.71
627	遭	73.73	1533	30.68	2449	430.05	263	178.15
628	本身	174.50	650	297.74	348	61.15	2266	177.80
629	當地	89.76	1241	189.61	539	252.99	485	177.45
630	有點	171.20	660	286.99	358	73.22	1884	177.14
631	病人	105.61	1085	356.00	294	68.77	2022	176.79
632	樣	123.36	918	365.90	281	40.35	3292	176.54
633	注意	193.91	578	193.13	531	142.19	960	176.41
634	重新	175.91	647	144.53	681	208.44	602	176.29
635	實際	203.46	552	190.69	538	134.39	1020	176.18
636	鼓勵	105.93	1080	168.78	590	251.41	490	175.38
637	同樣	238.84	475	137.02	716	149.09	908	174.98
638	擁有	219.49	516	98.79	948	206.58	612	174.96
639	科技	119.51	964	98.09	958	304.88	396	174.16
640	專業	129.89	871	132.00	743	259.36	474	173.75
641	內容	216.45	521	61.22	1433	243.07	508	173.58
642	穿	181.03	619	160.48	617	179.13	734	173.55
643	行	247.00	459	101.93	923	171.38	765	173.44
644	嘛	73.87	1529	429.88	246	15.82	7054	173.19
645	那邊	50.32	2178	459.81	232	9.12	10767	173.08
646	最近	108.84	1054	160.76	616	249.10	496	172.90
647	存在	297.38	376	112.07	857	107.55	1325	172.33
648	有關	271.24	419	108.92	879	136.81	997	172.32
649	賣	108.76	1055	200.35	515	207.20	608	172.11
650	幫忙	83.49	1334	294.69	351	138.13	991	172.10
651	作	306.32	361	82.09	1117	127.33	1089	171.92
652	部	280.79	406	68.77	1294	164.64	797	171.40

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
653	單位	91.22	1227	80.22	1139	342.28	343	171.24
654	菜	70.27	1602	308.10	332	134.59	1018	170.99
655	顆	125.94	907	242.38	430	144.39	943	170.90
656	突然	268.08	426	103.76	912	138.39	989	170.08
657	稱	237.27	482	55.78	1550	214.15	585	169.07
658	觀眾	123.41	917	238.86	436	144.86	939	169.05
659	功能	151.91	746	181.21	558	173.97	756	169.03
660	城市	126.16	905	109.35	876	271.50	457	169.00
661	來到	176.76	644	252.19	414	77.81	1789	168.92
662	加入	113.71	1013	123.89	789	268.60	460	168.73
663	周	38.37	2752	71.82	1253	393.25	293	167.81
664	例如	250.95	445	90.16	1032	161.22	826	167.44
665	女性	232.54	491	80.31	1138	188.04	690	166.96
666	情緒	259.50	438	131.16	745	110.10	1284	166.92
667	期間	130.26	866	54.65	1585	315.21	375	166.71
668	通常	214.70	525	212.88	481	71.71	1922	166.43
669	動作	171.44	658	215.64	475	112.04	1257	166.37
670	教授	153.01	740	163.01	609	182.80	716	166.28
671	北市	3.12	20028	5.49	8981	489.31	222	165.97
672	達	98.64	1155	28.76	2573	370.03	317	165.81
673	國小	18.65	5037	113.71	847	364.83	320	165.73
674	門	307.11	360	76.84	1177	112.95	1245	165.63
675	低	128.60	881	138.71	706	229.19	547	165.50
676	想法	213.41	529	199.79	517	82.47	1703	165.22
677	經營	156.39	722	93.21	1002	245.81	503	165.13
678	即使	248.00	455	95.74	975	151.20	894	164.98
679	第三	179.39	632	141.34	700	173.87	757	164.87
680	快樂	178.03	639	225.21	458	90.95	1541	164.73
681	不管	150.90	750	256.50	406	86.66	1613	164.69

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
682	七	228.70	502	112.25	856	152.76	881	164.57
683	根據	182.60	607	131.21	744	179.71	731	164.50
684	終於	230.20	498	111.13	860	151.54	891	164.29
685	之外	181.33	618	217.00	472	93.44	1505	163.92
686	晚上	163.66	688	194.82	527	132.50	1037	163.66
687	眼睛	256.72	440	158.93	622	74.97	1850	163.54
688	帶來	159.51	706	151.47	658	178.82	736	163.26
689	思考	247.70	456	121.97	798	119.61	1169	163.09
690	態度	222.79	509	108.64	880	157.62	854	163.02
691	角色	206.26	547	106.34	898	174.81	753	162.47
692	八	212.19	533	115.44	836	159.43	837	162.35
693	警方	30.70	3316	9.43	6098	446.87	251	162.33
694	以為	232.68	490	128.95	761	124.79	1118	162.14
695	全部	133.86	844	254.53	409	97.54	1443	161.98
696	有沒有	83.49	1336	365.10	284	36.55	3590	161.71
697	光	163.83	686	154.80	642	166.04	791	161.56
698	代	155.21	730	145.70	673	182.85	714	161.26
699	有的	118.11	984	298.06	347	67.49	2056	161.22
700	眼	283.76	399	74.21	1218	125.18	1114	161.05
701	不用	90.32	1236	302.15	336	90.45	1550	160.97
702	高雄	17.99	5159	51.27	1654	413.09	276	160.78
703	記得	181.86	615	235.16	444	65.33	2116	160.78
704	好好	120.08	957	264.62	396	97.54	1444	160.75
705	在一起	180.88	620	237.41	440	63.79	2169	160.69
706	通過	115.64	999	59.39	1481	306.99	389	160.67
707	叫做	76.95	1460	395.31	259	8.42	11377	160.23
708	提到	127.00	895	239.76	433	112.94	1246	159.90
709	自由	221.77	513	57.32	1525	199.08	636	159.39
710	辛苦	53.02	2088	316.03	321	108.53	1314	159.19

(續下頁)



綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
711	至少	161.87	699	118.63	813	196.97	654	159.16
712	減少	89.83	1240	153.54	644	233.32	535	158.90
713	樹	136.29	837	193.03	532	147.08	923	158.80
714	她們	195.67	570	207.90	488	71.95	1919	158.51
715	困難	148.18	770	211.80	483	115.22	1227	158.40
716	針對	88.77	1255	72.29	1243	312.76	379	157.94
717	笑	270.93	420	82.84	1110	119.17	1177	157.65
718	香港	184.31	604	37.25	2107	251.08	491	157.55
719	訓練	154.79	734	130.78	747	186.93	696	157.50
720	反而	155.85	726	172.39	578	143.03	955	157.09
721	銀行	132.73	853	50.05	1684	287.33	428	156.71
722	談	173.02	655	178.30	567	118.76	1181	156.69
723	自我	289.90	390	77.92	1169	102.07	1393	156.63
724	感到	293.46	383	47.89	1738	128.39	1077	156.58
725	觀察	173.43	654	155.46	636	140.74	973	156.54
726	保護	128.50	883	171.78	580	169.20	777	156.49
727	建築	122.56	932	137.40	712	208.56	600	156.17
728	慈濟	3.20	19659	435.60	243	29.17	4348	155.99
729	智慧	180.49	623	109.86	870	176.88	742	155.74
730	下去	150.04	759	236.47	442	79.42	1756	155.31
731	對方	244.19	464	45.27	1813	175.14	750	154.87
732	像是	196.21	568	146.45	669	121.89	1148	154.85
733	本來	126.32	902	259.74	401	78.16	1782	154.74
734	長期	110.84	1042	114.60	840	238.77	522	154.74
735	工程	56.08	1984	79.18	1149	328.25	361	154.50
736	然而	287.88	393	62.58	1407	112.15	1255	154.20
737	總是	248.08	453	121.40	805	92.76	1514	154.08
738	類	228.92	501	90.82	1025	142.45	959	154.06
739	屬於	155.67	727	197.40	523	109.07	1307	154.04

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
740	也許	249.02	451	157.80	624	55.10	2503	153.98
741	而已	170.52	662	249.28	416	41.88	3189	153.89
742	強	136.61	832	118.07	816	205.44	618	153.37
743	英國	170.36	663	103.06	917	186.39	700	153.27
744	博士	96.99	1169	310.59	329	52.01	2640	153.20
745	腳	188.36	592	167.33	595	102.25	1391	152.64
746	男	92.24	1212	25.99	2777	338.42	349	152.22
747	改善	84.79	1310	94.38	987	275.74	450	151.64
748	團體	121.25	940	93.73	998	239.35	519	151.44
749	比賽	77.78	1442	100.53	934	275.90	449	151.40
750	市府	1.68	30138	2.30	16101	449.66	248	151.21
751	開心	70.18	1606	209.22	486	173.83	758	151.08
752	團隊	66.56	1700	94.15	991	292.28	412	151.00
753	結合	88.03	1266	102.08	920	262.36	472	150.82
754	極	242.08	468	43.25	1886	166.39	789	150.57
755	耶	40.14	2642	392.21	262	18.44	6269	150.27
756	下午	76.94	1461	60.94	1440	312.88	378	150.25
757	幸福	103.86	1097	203.40	507	143.34	952	150.20
758	節目	84.13	1325	166.11	599	198.96	638	149.73
759	群	154.99	732	168.17	592	125.30	1113	149.49
760	報導	107.50	1072	66.75	1330	270.23	459	148.16
761	茶	85.97	1299	223.52	461	134.99	1016	148.16
762	正在	211.73	534	103.53	914	127.22	1091	147.49
763	形成	187.39	594	140.35	702	114.45	1232	147.40
764	生態	35.88	2922	216.68	473	189.14	686	147.23
765	回家	138.43	821	170.70	585	132.26	1042	147.13
766	植物	75.10	1508	306.84	334	58.84	2350	146.93
767	知識	249.39	449	118.31	814	72.28	1908	146.66
768	說明	207.72	541	63.19	1397	169.00	781	146.64

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
769	附近	126.55	901	125.44	785	187.32	693	146.43
770	結束	160.20	703	90.07	1033	188.63	687	146.30
771	看起來	163.52	691	230.61	448	44.24	3053	146.12
772	變化	167.76	669	162.78	611	107.20	1330	145.91
773	心情	132.53	856	181.63	556	122.13	1144	145.43
774	拍	113.41	1018	142.32	695	179.72	730	145.15
775	馬	192.48	581	48.18	1732	194.67	665	145.11
776	吸引	81.42	1366	64.92	1368	288.78	422	145.04
777	維持	123.26	919	87.11	1060	224.48	562	144.95
778	越來越	107.55	1071	244.77	426	82.50	1701	144.94
779	演出	81.45	1365	49.07	1711	303.98	397	144.83
780	科學	215.13	523	127.92	767	91.40	1532	144.82
781	呀	187.40	593	223.15	462	23.65	5111	144.73
782	解釋	235.86	485	73.70	1226	124.38	1123	144.65
783	創作	121.07	942	78.15	1165	234.70	531	144.64
784	算	133.75	845	192.70	533	107.43	1326	144.63
785	具有	239.21	474	69.24	1286	125.38	1109	144.61
786	人民	143.13	794	51.84	1639	237.58	527	144.18
787	導致	119.51	966	74.77	1204	238.24	525	144.17
788	公園	54.79	2017	88.99	1045	288.72	424	144.17
789	建立	207.66	542	65.77	1351	158.63	846	144.02
790	生產	141.57	802	126.94	773	163.37	805	143.96
791	年輕人	96.70	1172	167.80	594	167.28	788	143.92
792	除	59.90	1880	24.39	2935	347.11	340	143.80
793	主義	327.25	335	26.74	2720	76.88	1804	143.62
794	總	181.46	617	67.17	1323	181.78	723	143.47
795	分析	229.46	500	51.13	1659	149.66	904	143.42
796	不但	157.00	720	129.66	756	143.21	953	143.29
797	充滿	173.71	652	108.41	883	147.57	920	143.23

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
798	反應	209.50	536	109.35	875	110.20	1282	143.02
799	人們	294.81	381	74.87	1201	59.03	2346	142.90
800	馬上	155.25	728	192.61	534	80.80	1729	142.89
801	白	196.97	566	75.81	1192	155.54	863	142.77
802	亦	270.20	423	11.02	5441	147.02	926	142.75
803	資訊	148.31	769	69.75	1280	209.88	594	142.65
804	舉行	69.72	1622	24.96	2879	332.31	358	142.33
805	避免	130.39	864	72.33	1241	223.68	564	142.14
806	需求	114.29	1009	70.46	1271	241.21	513	141.99
807	情形	138.84	816	202.27	509	84.61	1657	141.91
808	效果	150.70	752	133.69	738	140.64	975	141.68
809	報告	146.10	780	80.50	1137	197.09	653	141.23
810	選舉	32.41	3187	18.11	3740	373.01	314	141.17
811	達到	141.15	806	128.86	762	152.27	886	140.76
812	公尺	76.26	1478	119.53	809	226.40	555	140.73
813	具	163.55	690	28.01	2622	229.95	546	140.50
814	新聞	157.87	714	90.49	1026	172.01	762	140.12
815	前往	107.04	1074	36.26	2158	276.52	447	139.94
816	不再	204.47	550	67.64	1316	147.23	922	139.78
817	跑	138.76	817	159.16	620	121.35	1152	139.76
818	例	222.48	511	20.97	3338	175.42	747	139.62
819	需	109.33	1051	21.53	3268	287.88	426	139.58
820	挑戰	107.79	1065	98.42	955	212.53	591	139.58
821	讀	180.16	626	121.50	803	117.00	1205	139.55
822	規定	97.12	1167	40.06	1989	280.80	437	139.33
823	日前	6.67	11561	1.78	19207	409.34	280	139.26
824	絕對	166.75	674	134.82	730	115.99	1221	139.19
825	品牌	64.32	1758	23.60	3017	329.58	360	139.16
826	換	100.59	1132	165.59	600	151.23	893	139.14

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
827	開放	68.82	1642	38.84	2049	308.97	386	138.88
828	太多	121.39	938	191.63	535	102.98	1380	138.66
829	概念	163.63	689	143.59	682	108.30	1319	138.51
830	方向	156.34	724	124.31	787	134.69	1017	138.45
831	堅持	120.66	948	123.94	788	170.53	768	138.38
832	旁	156.27	725	50.52	1676	207.64	605	138.14
833	課程	118.31	981	96.16	971	199.59	635	138.02
834	交通	45.58	2371	72.90	1235	295.15	409	137.88
835	黨	91.75	1218	6.61	7909	315.11	376	137.83
836	實在	202.96	553	121.07	806	88.00	1593	137.34
837	海	131.04	862	127.03	772	153.48	878	137.18
838	溝通	123.55	916	106.72	894	180.96	724	137.08
839	現象	187.02	598	135.66	723	88.14	1590	136.94
840	製作	102.21	1110	94.01	994	214.34	583	136.85
841	樣子	149.94	760	236.14	443	23.45	5151	136.51
842	深	157.68	716	110.43	866	141.34	966	136.48
843	想像	178.39	637	143.12	687	87.68	1596	136.40
844	體驗	62.80	1810	94.05	993	251.76	488	136.20
845	十分	249.96	448	39.78	2002	118.19	1189	135.98
846	電腦	149.51	763	127.78	768	130.05	1059	135.78
847	協會	52.95	2093	64.41	1373	288.80	421	135.38
848	積極	95.24	1190	67.74	1315	243.17	507	135.38
849	適合	101.17	1125	157.05	629	147.78	915	135.33
850	醫生	168.97	666	178.02	569	58.25	2372	135.08
851	台中	14.24	6262	24.91	2881	366.01	319	135.05
852	步	179.19	633	117.13	826	108.62	1313	134.98
853	或許	221.00	515	93.16	1003	90.67	1544	134.94
854	動	180.59	622	142.61	692	81.44	1719	134.88
855	人士	109.52	1049	34.53	2248	260.53	473	134.86

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
856	有機	18.80	4996	298.39	344	87.03	1608	134.74
857	它們	252.59	444	128.34	766	22.90	5257	134.61
858	電視	160.14	704	94.52	986	148.99	909	134.55
859	醫療	54.79	2018	144.86	676	203.96	624	134.53
860	不錯	91.57	1221	220.33	467	91.60	1530	134.50
861	居民	53.25	2078	117.32	823	232.83	537	134.47
862	層	143.39	792	129.89	755	129.29	1065	134.19
863	控制	187.22	595	111.41	859	103.87	1367	134.17
864	會議	119.66	961	36.54	2147	245.74	504	133.98
865	系列	75.32	1502	28.85	2563	297.17	407	133.78
866	經常	193.77	579	97.01	967	109.60	1296	133.46
867	重	120.76	946	129.89	754	148.93	910	133.19
868	初	163.36	693	37.81	2081	198.33	642	133.17
869	階段	140.06	811	97.48	963	161.20	827	132.91
870	負責	144.98	786	98.70	950	153.95	875	132.54
871	並且	238.46	478	90.02	1036	69.14	2009	132.54
872	遇到	115.05	1004	169.02	589	113.43	1240	132.50
873	確定	158.67	713	67.74	1314	170.79	767	132.40
874	反	172.82	656	39.31	2024	184.52	709	132.22
875	明顯	140.77	808	100.90	930	154.58	869	132.08
876	看見	222.09	512	55.54	1559	118.47	1184	132.03
877	衣服	129.64	874	206.12	495	59.87	2315	131.87
878	等等	150.79	751	204.43	503	40.37	3288	131.86
879	活	195.03	574	108.46	882	91.93	1524	131.80
880	紅	136.29	838	74.02	1220	184.20	710	131.51
881	觀念	157.43	717	162.78	610	74.18	1866	131.46
882	季	31.88	3221	39.36	2023	323.06	367	131.43
883	員工	124.40	915	50.94	1661	218.62	572	131.32
884	漂亮	77.45	1449	268.84	384	47.38	2879	131.22

(續下頁)

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
885	電話	150.26	755	68.86	1291	174.43	754	131.18
886	選	67.56	1676	117.60	819	208.24	603	131.13
887	皆	179.78	629	16.14	4092	197.15	651	131.02
888	標準	122.15	934	78.39	1161	192.38	677	130.97
889	綠	57.98	1923	96.07	973	238.39	523	130.81
890	時期	189.25	587	81.90	1119	121.29	1155	130.81
891	那樣	179.86	628	185.20	546	26.95	4639	130.67
892	嗯	69.76	1620	316.97	320	4.89	16393	130.54
893	屆	40.51	2626	23.78	3001	327.23	363	130.51
894	高中	46.17	2348	64.03	1383	280.80	438	130.33
895	英文	59.95	1876	64.27	1378	266.25	465	130.16
896	北京	140.89	807	26.97	2704	222.20	566	130.02
897	人物	207.44	544	65.25	1363	117.13	1203	129.94
898	魚	137.05	827	115.40	837	135.55	1007	129.33
899	那裡	218.90	519	142.70	690	26.22	4721	129.28
900	顯示	129.35	875	37.62	2089	220.79	569	129.25
901	介紹	104.27	1095	193.97	528	89.36	1568	129.20
902	左右	132.92	851	152.74	652	101.86	1401	129.17
903	熱	77.37	1453	199.13	520	110.68	1275	129.06
904	期待	84.95	1308	105.50	902	196.65	657	129.03
905	至於	122.58	930	42.92	1898	221.30	567	128.94
906	投入	74.12	1525	112.40	854	199.06	637	128.52
907	原	129.23	876	18.20	3722	237.91	526	128.45
908	萬	51.07	2143	14.50	4432	318.87	373	128.14
909	專家	101.82	1118	141.24	701	140.70	974	127.92
910	畢業	77.35	1455	108.31	886	197.64	648	127.77
911	立委	4.13	16496	1.08	26342	377.03	309	127.41
912	創意	79.23	1414	105.12	904	197.30	649	127.22
913	酒	182.14	613	47.52	1753	150.28	900	126.64

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
914	倍	90.34	1235	88.19	1052	201.37	630	126.63
915	唯一	151.85	748	78.86	1154	149.16	907	126.62
916	表演	80.87	1382	90.49	1028	208.20	604	126.52
917	地球	84.01	1327	250.69	415	44.68	3019	126.46
918	確實	144.21	789	104.89	905	128.85	1073	125.99
919	部落	68.99	1638	149.78	664	159.12	842	125.96
920	安排	118.63	977	60.98	1437	198.20	643	125.94
921	百	96.52	1174	55.73	1552	225.22	559	125.82
922	特色	63.20	1797	73.51	1230	238.88	521	125.19
923	開發	75.33	1501	93.82	996	206.37	615	125.17
924	基礎	156.46	721	54.70	1581	163.93	800	125.03
925	房子	123.20	920	179.10	563	72.48	1904	124.93
926	米	100.99	1129	117.23	825	155.88	862	124.70
927	國內	68.31	1658	51.18	1657	254.28	479	124.59
928	平常	67.51	1677	238.58	437	67.50	2055	124.53
929	引起	145.17	784	89.32	1041	138.75	984	124.41
930	程度	173.66	653	97.43	965	101.94	1399	124.34
931	興趣	135.17	843	134.58	731	102.56	1386	124.10
932	北	78.34	1436	46.07	1790	247.65	498	124.02
933	遊客	19.27	4910	59.90	1472	291.82	413	123.66
934	女人	280.88	405	28.76	2572	61.34	2261	123.66
935	邀請	56.57	1970	50.66	1670	263.73	468	123.65
936	正常	113.64	1016	145.37	674	111.77	1258	123.59
937	心理	233.25	489	66.00	1347	71.52	1928	123.59
938	逐漸	152.73	741	85.47	1083	132.57	1036	123.59
939	接著	194.13	576	85.33	1084	91.12	1539	123.53
940	感謝	70.10	1611	116.66	830	182.77	717	123.18
941	煮	50.77	2155	245.48	424	73.03	1889	123.09
942	照片	99.41	1144	117.56	820	152.25	887	123.07

(續下頁)



綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
943	條件	143.38	793	85.00	1088	139.64	980	122.67
944	影片	129.72	873	108.41	884	129.76	1062	122.63
945	空氣	70.80	1588	184.31	547	112.12	1256	122.41
946	朝	233.75	488	32.88	2326	100.59	1412	122.41
947	申請	41.40	2579	35.18	2216	289.83	417	122.14
948	阿嬤	21.92	4413	238.35	439	106.03	1337	122.10
949	公里	65.88	1717	99.03	947	201.31	631	122.07
950	隨著	119.26	970	110.19	868	135.78	1004	121.74
951	太太	183.71	605	123.51	793	57.96	2380	121.73
952	差	93.66	1201	143.22	686	128.12	1080	121.66
953	結構	170.07	664	94.62	983	100.18	1414	121.62
954	以來	105.59	1087	63.84	1387	195.41	660	121.62
955	位置	127.08	894	153.44	645	84.06	1668	121.53
956	農民	60.39	1862	126.28	779	177.80	739	121.49
957	氣	209.36	537	68.30	1299	86.41	1621	121.36
958	品質	76.20	1481	95.32	981	192.46	676	121.32
959	分別	100.71	1131	36.59	2145	226.55	554	121.28
960	大量	111.87	1032	116.20	831	135.51	1008	121.19
961	表達	136.51	834	69.85	1276	156.58	857	120.98
962	遊戲	123.13	925	91.80	1014	147.99	914	120.98
963	進	154.06	737	80.54	1136	128.13	1079	120.91
964	感受	119.86	959	121.82	800	120.62	1162	120.77
965	垃圾	28.29	3541	207.48	490	125.93	1102	120.57
966	平均	54.29	2036	95.65	980	211.65	592	120.53
967	校長	55.13	2010	76.28	1187	230.08	544	120.50
968	如今	103.70	1101	50.47	1678	207.25	607	120.47
969	旁邊	100.49	1134	229.29	454	31.63	4031	120.47
970	責任	117.29	990	73.84	1224	170.25	771	120.46
971	建設	45.89	2357	36.68	2134	278.77	442	120.45

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

綜合序位	詞	書面語 每百萬詞頻	書面語 序位	口語 每百萬詞頻	口語 序位	新聞 每百萬詞頻	新聞 序位	每百萬詞頻 (平均)
972	下來	130.18	867	178.59	565	52.07	2633	120.28
973	美元	107.82	1064	19.61	3519	233.26	536	120.23
974	提醒	81.90	1359	97.99	959	180.73	725	120.21
975	農業	36.72	2867	167.33	596	155.08	865	119.71
976	永遠	178.55	636	107.05	892	73.04	1888	119.55
977	集團	81.19	1373	8.96	6326	267.75	463	119.30
978	黑	146.95	775	64.55	1370	146.13	932	119.21
979	死亡	135.49	840	85.75	1078	135.44	1009	118.89
980	呈現	105.71	1084	87.39	1056	163.47	804	118.86
981	回答	250.82	446	43.77	1866	61.70	2250	118.76
982	主題	111.87	1031	54.60	1587	189.72	684	118.73
983	取得	101.19	1124	48.08	1734	206.82	609	118.69
984	穩定	78.50	1433	92.04	1010	185.46	706	118.67
985	豐富	95.92	1179	134.30	735	125.63	1105	118.62
986	感動	59.84	1883	177.08	571	118.33	1186	118.42
987	雙方	88.14	1264	23.97	2979	242.39	510	118.17
988	均	109.61	1047	12.24	5014	232.61	538	118.16
989	保持	154.69	736	86.41	1065	112.70	1248	117.93
990	數	149.84	761	28.71	2580	175.16	749	117.90
991	狗	142.06	799	99.73	939	111.64	1260	117.81
992	高興	152.66	742	156.58	632	44.11	3061	117.78
993	觀光	17.66	5233	49.58	1694	285.89	431	117.71
994	目的	178.31	638	85.84	1073	88.66	1579	117.61
995	消費者	58.93	1901	83.17	1106	210.47	593	117.53
996	信	223.76	508	57.46	1519	71.19	1936	117.47
997	味道	56.09	1983	247.40	420	48.43	2817	117.31
998	支	100.56	1133	82.09	1116	169.14	779	117.26
999	手機	35.94	2916	84.44	1094	231.12	540	117.17
1000	藥	99.17	1149	148.23	666	103.54	1371	116.98

## 第八章 結語

本計畫臺灣華語文語料庫 (COCT) 是繼 2007 年「中研院平衡語料庫」之後，目前華語文世界建置最完備、語料數量最多、內容最新的應用語料庫，語料類別及數量，詳見圖 3。自 2013 年至 2020 年，國教院華語文語料庫業已完成蒐集書面語語料約 4 億 3,541 萬字（含新聞語料共 34 億 3,177 萬字），口語語料 4,635 萬字，雙語語料約 1,160 萬字，中介語語料 156 萬 4,000 字。至 2020 年國內語料庫的建置情形，詳見表 13。



圖 3 國教院「臺灣華語文語料庫」(COCT)

表 13

2020 年國內語料庫建置情形

		臺灣華語文語料庫	中研院平衡語料庫	其他單位
語料庫特性	書面語	34 億 3,177 萬字	新聞為主，1000 萬詞	
	口語	4,635 萬字	少量	政大、高師大
	華語中介語	156 萬 4,000 字	無	臺灣師大
	華英（雙語）	1,160 萬字	無	

(續下頁)

# 臺灣華語文語料庫

華語文教與學的必備工具

		臺灣華語文語料庫	中研院平衡語料庫	其他單位
維 護		~ 持續	2007 年	
整合應用	標準體系建置	設計上兼顧研究與應用，並持續進行應用推廣活動	語言學研究為主	以學術研究為主，除臺灣師大外，規模多不大，故難以整合應用
	教材編輯			
	詞典編輯			
	資料驅動學習			

本語料庫，除了提供研發建置華語文能力指標及漢字、詞語及語法點分級標準，更可應用於華語文、國語文、英語文等教學設計、教材編撰、測驗評量、辭典編撰、人工智慧及語料應用之軟體設計、開發，提升華語文教育的效能（圖 4）。

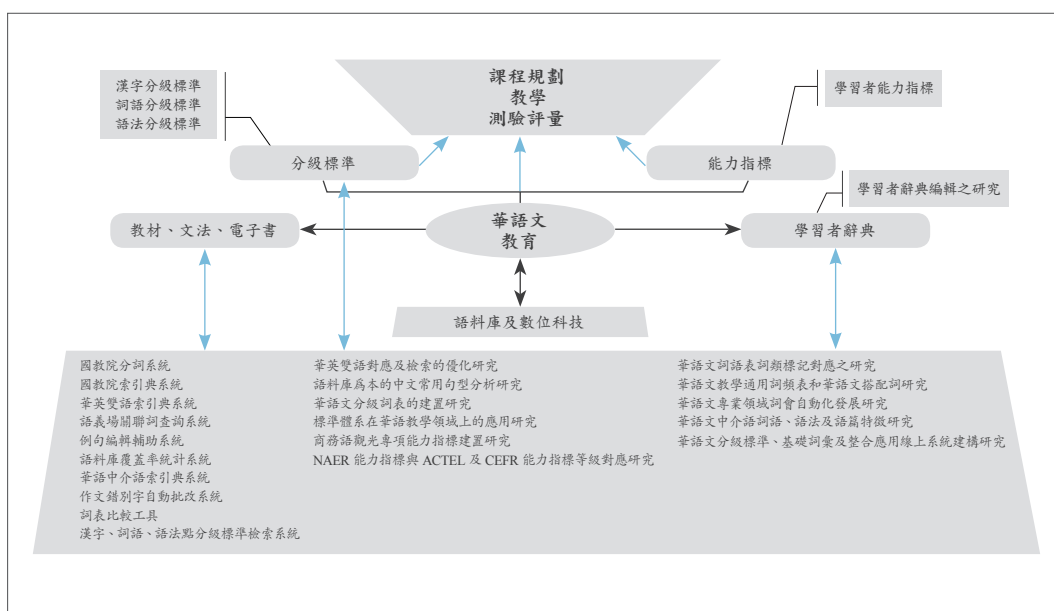


圖 4 語料庫、數位科技及標準體系整合應用於華語文教育

## 資料來源

- 白明弘、吳鑑城、簡盈妮、黃淑齡、林慶隆（2016）。基於詞語分布均勻度的核心詞彙選擇。《中文計算語言學期刊》，21（2），1-17。
- 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》，16（3），29-55。
- 林慶隆（2019年，11月）。語料庫科技在語文教學的應用。《國家教育研究院電子報》，189。取自 <https://reurl.cc/0odANb>
- 柯華葳、林慶隆、張俊盛、陳浩然、高照明、信世昌、…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。教育部委託計畫報告。新北市：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、…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。教育部委託計畫報告。新北市：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。教育部委託計畫報告。新北市：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。教育部委託計畫報告。新北市：國家教育研究院。
- 張莉萍、陳鳳儀（2005年，4月）。華語詞彙分級初探。第六屆漢語辭彙語義學研討會論文集，中國：廈門。
- 梁以辰（2014年11月14日）。冠狀動脈心臟病／引發血管栓塞機率高。《健康醫療網》。取自 <https://reurl.cc/arYl0Z>
- 許添明、林慶隆、柯華葳、張俊盛、陳浩然、高照明、…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。教育部委託計畫報告。新北市：國家教育研究院。
- 許添明、林慶隆、柯華葳、張俊盛、陳浩然、高照明、…張莉萍（2019a）。華語文八年計畫「建置應用語料庫及標準體系」108年工作計畫期中報告。教育部委

- 託計畫報告。新北市：國家教育研究院。
- 許添明、林慶隆、柯華葳、張俊盛、陳浩然、高照明、…張莉萍（2019b）。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。教育部委託計畫報告。新北市：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、…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。教育部委託計畫報告。新北市：國家教育研究院。
- 詞庫小組（1998）。中央研究院平衡語料庫的內容與說明（修訂版）。臺北市：中央研究院資訊科學所。
- Asahara, M., Fukuoka, K., Azuma, A., Goh, C. L., Watanabe, Y., Matsumoto, Y., & Tsuzuki, T. (2005, October 14-15). *Combination of Machine Learning Methods for Optimum Chinese Word Segmentation* [Conference session]. Proceedings of Fourth SIGHAN Workshop on Chinese Language Processing, Jeju Island, Korea. <https://www.aclweb.org/anthology/I05-3000.pdf>
- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24(2), 199-208.
- Carroll, J. B. (1970). An alternative to Juillard's usage coefficient for lexical frequencies and a proposal for a standard frequency index. *Computer Studies in the Humanities and Verbal Behaviour*, 3(2), 61-65.
- Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 67-85). Amsterdam: John Benjamins.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Huang, C.-R., Zhang, H. & Yu, S.-W. (2005). On predicting and verifying a basic lexicon: proposals inspired by distributional consistency. In Dah-an Ho & Ovid J.L. Tzeng (Eds.), *POLA forever: festschrift in honor of Professor William S-Y. Wang on his 70th birthday* (pp. 57-69). Taipei: Language and Linguistics, Academia Sinica.
- Juillard, A. G., & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. The Hague: Mouton & Co.
- Ljung, M. (1990). *A study of TEFL vocabulary*. Almqvist & Wiksell International.
- Ma, Wei-Yun, & Chen, Keh-Jiann. (2003). Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff. In Association

- for Computational Linguistics (Eds.), *SIGHAN '03: Proceedings of the second SIGHAN workshop on Chinese language processing*, pp. 168-171. <https://www.aclweb.org/anthology/W03-1726.pdf>
- Mindt, D. (1986). Corpus, grammar and teaching English as a foreign language. In Gerhard Leitner (Ed), *The English Reference Grammar: Language and Linguistics, Writers and Readers* (pp. 125-39). Tübingen, Niemeyer: Max Niemeyer Verlag GmbH & Co KG.
- Mindt, D. (1996). English corpus linguistics and the foreign language teaching syllabus. In J. Thomas & M. Short (Eds.), *Using Corpora for Language Research* (pp. 232-247). Harlow: Longman.
- Mindt, D. (1997). Corpora and the teaching of English in Germany. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 40-50). New York: Addison Wesley Longman.
- Peng, F., Feng, F., & McCallum, A. (2004, August 23-27). Chinese Segmentation and New Word Detection using Conditional Random Fields. In COLING(Eds.), *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland. <https://www.aclweb.org/anthology/C04-1081.pdf>
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Rosengren, I. (1971). The quantitative concept of language and its relation to the structure of frequency dictionaries. *Études de linguistique appliquée (Nouvelle Série)*, 1, 103-127.
- Sampson, G., & Bentham, Y. (1992). Probabilistic parsing. In J. Svartvik (Ed), *Directions in Corpus Linguistics* (pp. 425-447). New York: Mouton de Gruyter.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Summers, Della. (1996). Corpus lexicography - The importance of representativeness in relation to frequency. *Longman Language Review* 3, 6-9.
- Verlinde, Serge, & Selva, Thierry. (2001, March 29-April 2). *Corpus-based versus intuition-based lexicography: defining a word list for a French learners' dictionary* [Paper presentation]. Proceedings of the Corpus Linguistics 2001 conference, Lancaster, UK. <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/verlinde.pdf>
- West, M. P. (1953). *A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Longmans, Green.
- Xue, N. (2003). Chinese word segmentation as character tagging. *Computational Linguistics and Chinese Language Processing*, 8(1), 29-48.

# 臺灣華語文語料庫

華語文教與學的必備工具



## 誌謝

華語文八年計畫「建置應用語料庫及標準體系」由教育部委由國家教育研究院執行。自 102 年迄今，本院邀集相關領域專家學者、資深華語文教師及本院人員組成研究團隊，共召開超過百場諮詢會議研擬討論，同時結合語料庫科技，歷經數年完成華語文能力指標、漢字分級表、詞語分級表及語法點分級表徵求意見稿，感謝教育部的經費補助及感謝研究團隊，成員如下：

### 歷年主持人

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、郭工賓、范信賢

### 專家學者

方麗娜、王萸芳、江敏華、李明懿、李嘉真、杜昭玫、林青蓉、周美慧、屈承熹、信世昌、柯華葳、張于忻、張麗麗、曹逢甫、陳子璋、陳立元、陳克健、陳定國、陳玟妤、陳麗宇、黃沛榮、黃淑君、彭妮絲、曾建銘、廖柏森、劉水深、劉昭麟、劉顯親、賴慈芸、蔡宜妮、鄧守信、謝佳玲、謝舒凱、戴浩一、鄭錦全、蘇以文（依姓氏筆畫排列）

### 資深華語文教師

孫懿芬、張黛琪、陳懷萱、楊尤媛、盧翠英（依姓氏筆畫排列）

### 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、蔡岳璋、陳沛璇、陳茹玲、林明佳、蔡曉楓、張玳維、余昱瑩、陳威佑、李韻如、王琳、林晏宇、陳鵬妃、黃淑齡、劉寶琦、簡盈妮、蔡雅如、羅紫瑄、張洪瑄、林佳綺、古宜臻、莊惠雯、李詠梅、張馨文、張簡筠珊、張婉婷、范蕊啟、謝佳伶、馬睿琳、陳佩羽、陳威宇、夏儀芬、吳安璿、張婷婷、單師凡、游翔霖、游俊龍、曾偉傑、許鈞惟、陳冠中

國家教育研究院院長 許添明 謹誌

中華民國 109 年 11 月

臺灣華語文語料庫 - 華語文教與學的必備工具 / 國家教育研究院主編. -- 初版. -- 新北市：國家教育研究院, 民110.01

面；公分

ISBN 978-986-5461-63-8 (平裝)

1.漢語教學 2.語料庫

802.03

109020957

## 臺灣華語文語料庫

華語文教與學的必備工具

發行人：許添明

主編：國家教育研究院

編輯小組：林慶隆、白明弘、蔡岳璋、吳鑑城、李詩敏、吳欣儒、丁彥平、許愷容、盧昱勳、王冠孺、陳沛璇

設計審核：策略溝通辦公室

美編設計：尚暉文化事業有限公司

文字校對：張華芸

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>

排版印刷：尚暉文化事業有限公司

展售處：

國家書店松江門市

地址：104472 臺北市中山區松江路 209 號 1 樓

電話：(02)2518-0207

網址：<https://www.govbooks.com.tw>

五南文化廣場

地址：400002 臺中市區中山路 6 號

電話：(04)2226-0330

網址：<https://www.wunanbooks.com.tw>

出版日期：110 年 1 月初版一刷

定價：新臺幣 130 元

GPN：1011000034 ISBN：9789865461638

### 電子書設計製作

設計製作：尚暉文化事業有限公司

地址：220742 新北市板橋區板新路 103 號 4 樓之 1

電話：(02)2958-6010

網址：<http://www.sun-wei.com.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：FlipViewer Xpress



教育部為提升華語文教育的品質及效能，因此在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，並交由本院執行。工作項目包括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，及辦理相關的應用與推廣工作。

本書為建置應用語料庫的部分成果，內容包括書面語語料庫、口語語料庫、華英雙語語料庫、中介語語料庫及通用詞頻表等建置，並涵蓋上述各語料庫的收集範圍、年代、類別、方式及累計數量等；而且建置語料庫與標準體系整合應用系統，提供讀者在字詞語及語法使用的檢索與分析，是華語文教與學的必備工具。



國家教育研究院

National Academy for Educational Research

[www.naer.edu.tw](http://www.naer.edu.tw)

ISBN 978-986-5461-63-8

定價 130 元



9 789865 461638

GPN 1011000034

