

## 兩岸華語詞表之比較及選詞建議\*

楊惠媚 陳浩然<sup>1</sup> 潘依婷

國立臺灣師範大學英語系

### 摘要

常用詞表是語言學習的核心內容，廣受學習者、教師及教材編撰人員的關注。目前兩岸皆提出供華語學習使用的詞表，但分析比較兩岸詞表異同的研究迄今相當少見。本研究首先比對 TOCFL 和 HSK 詞表，得到以下幾項重要發現：（1）兩岸選詞相當不同。儘管有 3,700 個詞相同，但 TOCFL 有 3,619 個詞不在 HSK 詞表中，而 HSK 詞表有 1,296 個詞不在 TOCFL 詞表中。（2）兩岸詞表的詞彙呈現及收詞原則不同。（3）HSK 四字詞多出 TOCFL 詞表 4.3 倍的數量。（4）TOCFL 詞表「兒化」詞的數量高於 HSK 詞表。此外，本研究觀察發現部分少用詞彙也被納入兩詞表之中。因此，本研究利用兩岸大型語料庫之前 1 萬筆高頻詞分別檢視兩詞表詞彙。發現 TaiwanWac 語料庫中有 4,770 筆語料庫高頻詞未被納入 TOCFL 詞表中，而 Tenten 語料庫中有 6,357 筆語料庫高頻詞未收錄至 HSK 詞表中。顯示語料庫高頻詞中有部分詞彙可建議納入 TOCFL 及 HSK 詞表中，以提升兩詞表品質及涵蓋率。此研究結果及建議可提供華語學習者、教師、教材編輯人員及考試單位作為參考。

**關鍵詞：**中文學習 華語詞表 核心詞 網路語料庫 高頻詞

## 1. 緒論

### 1.1 研究背景與動機

隨著中文在國際間受矚目的程度與日俱增，學習中文的人數也逐漸增加。然而中文的浩瀚詞彙，往往是造成學習者學習困難的原因之一。因此，許多教育學者與語言學家開始積極地探討中文學習者應具備的詞彙能力為

---

\* 本研究得到「教育部邁向頂尖大學計畫」與「國科會跨國頂尖研究中心計畫」之經費補助，特此感謝。

<sup>1</sup> 本文通訊作者。

何？哪些才是學習者必要掌握的詞彙？為提供中文詞彙教學之使用，兩岸分別推出中文詞彙表。大陸的漢語水平考試（HSK）於 2010 年對外發布了新 HSK 詞彙大綱五千詞，而臺灣的華語能力測驗（TOCFL）則於 2007 年公布了華語八千詞表。兩詞表在兩岸皆被視為華語教學及中文能力分級測驗的重要參考依據。

漢語水平考試（Hanyu Shuiping Kaoshi, HSK）是由大陸國家漢辦規劃執行的中文能力水平考試。1984 年正式對外辦考，但過去舊 HSK 龐大的詞彙量對於學習者而言，特別是初學者來說是「高不可攀」的。很多學習者參加了 HSK 後，感到十分挫折，影響了進一步學習中文的積極性。有鑑於此，中國大陸國家漢辦在 2009 年推出了新漢語水平考試，同時也推出新 HSK 詞彙大綱，將詞彙量從原本的 8822 詞大幅縮減為 5000 詞。（張晉軍等 2010）。

新 HSK 考試內容吸收了舊 HSK 的優點，並借鑒了近年來國際語言測驗的研究成果，實際充分調查、了解中國海外漢語學習情況編寫而成，但未將詞的使用頻率納入選詞依據。在這一方面，張晉軍等人（2012）對新 HSK 詞彙表收詞依據也提出了一些問題及反思：新 HSK 詞彙表研制的時間較短，雖然具時代性、實用性、經濟性等方面的優點，但在收錄哪些詞、哪些詞應歸入哪個等級等方面還有待進一步研究討論。

華語文能力測驗（Test of Chinese as a Foreign Language, TOCFL）則由臺灣國家華語測驗推動工作委員會針對母語非華語人士研發而成，於 2003 年 12 月正式對外開辦考試。測驗以張莉萍於 2003-2004 年發表的八千詞作為參考詞表，於 2007 年對外公布 TOCFL 八千詞表<sup>2</sup>。

TOCFL 詞表及 HSK 詞表不僅在兩岸深受矚目，在全球間華語學習的影響力亦是有目共睹。然而，TOCFL 詞表有八千詞，HSK 詞表則有五千詞。就詞彙的數量而言，兩岸詞表間相差了三千詞。而在詞表內容方面，若不經分析比對兩岸詞表的收詞情形，無法得知兩岸詞表間相似或相異之處為何。因此，本研究首先將探討並比對兩岸詞表內容之異同。

近半世紀以來，為求更加客觀地彙整出語言核心詞彙，並隨者語料庫技術的發展，運用語料庫整理出語言的核心詞及高頻常用詞彙，以作為語言教

---

<sup>2</sup> 引自國家華語測驗推動工作委員會（2013），《華語八千詞表說明》  
[http://www.sc-top.org.tw/download/8000\\_description.pdf](http://www.sc-top.org.tw/download/8000_description.pdf)（華測會已於 2013 年 5 月 10 日公佈新版 TOCFL《華語八千詞表》，然本研究於 2013 年 2 月 27 日完成投稿，故本文所使用之《華語八千詞表》為 2013 年 5 月之前版本。）

學依據，在國際間早已受到相當的重視。1953 年 Michael West 自語料庫<sup>3</sup>取詞頻建立 General Service List of English Words (GSL) 詞表首開先例，該詞表列出英文常見的 2000 個核心詞提供給英語學習者及 ESL (English as Second Language) 教師使用。Engels (1968) 更針對 GSL 詞表研究發現：如果學習者掌握了此詞表的核心詞及其相關詞族，英文閱讀時，可理解約 80% 的內容。著名詞彙教學專家 Nation (2001) 亦表示，第二語言的詞彙教學建議以詞頻作為詞彙教學參考。由此可知，自語料庫歸納出的詞表有助於學習者更有效率地掌握語言的核心詞。

然而，無論是中文核心詞彙或中文詞表，現階段皆較缺乏由語料庫為基礎編纂之詞彙表，難以客觀地彙整出中文詞彙的核心詞及高頻用詞。對此，本文將藉大型網路語料庫以詞頻計算出的高頻詞，檢視 TOCFL 和 HSK 詞表內容與語料庫高頻詞的對應情況，並探究是否尚有語料庫高頻詞未被納入於兩岸詞表之中，研究結果可作為兩岸詞表增修建議。

## 1.2 研究問題

兩岸著名的華語能力測驗——TOCFL 及 HSK，皆推出對應測驗的詞表，且兩詞表在華語教學上深受重視，是教師協助學習者準備能力測驗及學習者自學的主要內容之一。然而，迄今少有研究比對 TOCFL 詞表及 HSK 詞表之詞彙及選詞原則。本研究欲自比對 TOCFL 詞表及 HSK 詞表內容，探究兩岸詞表在選詞上之異同，及兩詞表內容相同及相異之處各為何？

此外，根據 Nation (1997) 的說法，學習者所需的詞彙應從母語者在使用語言時所必用到的詞彙去挑選。著名英語詞表如：GSL、BNC (The British National Corpus)、COCA (Corpus of Contemporary American English) 等詞表便是以英語母語者語料庫為基礎，彙整高頻詞而產生的。而兩岸 TOCFL 詞表及 HSK 詞表內容與華語母語者經常使用的詞彙相比，其相同或相異之處為何？兩詞表內含詞彙是否貼近學習者使用語言時所應具備的詞彙呢？

針對以上的問題，本研究將透過對比 TOCFL 及 HSK 詞表內容探究兩詞表的異同。另外分別自兩岸大型網路語料庫中依詞彙頻率擷取高頻詞，並將語料庫高頻詞分別與 TOCFL 詞表及 HSK 詞表作比對，查驗 TOCFL 詞表及 HSK 詞表與語料庫高頻詞相同之詞彙為何？兩詞表中有哪些詞不在語料庫

---

<sup>3</sup> General Service List of English Words (GSL)是自 1940 年代所建置的 5 百萬字書面語料庫中，依詞頻及各方考量列出前 2,000 筆核心詞。(Nation 1990 & 1997)

高頻詞中？並檢視語料庫高頻詞中，是否有部分高頻詞彙可建議收錄到 TOCFL 詞表及 HSK 詞表中？本文期研究成果能提供兩岸中文教師、教材及考試編輯人員參考。下節針對研究問題提出相關文獻並加以討論。

### 1.3 文獻探討

本節主要就以下兩大要點探討相關文獻：(1) 詞彙對於語言教學的重要性。(2) 語料庫詞頻對詞彙教學的貢獻。

#### 1.3.1 詞彙對於語言教學的重要性

詞彙是語言的重要組成部件、三大要素之一，也是語言的建築材料（賈穎 2001；王麗亞 2011）。Nation（1997）認為有系統及原則地學習詞彙在二語學習中是重要關鍵因素，且詞彙知識有助於語言使用更加流暢。李如龍及吳茗（2005）認為詞彙教學在語言教學中貫穿始終，起著承上啟下的重要作用，詞彙教學的效果並直接影響學習者語言的整體水平。俞靜（2012）同樣指出詞彙教學是教學中的一個重要環節。如前所述，詞彙在語言教學上不容小覷。

有鑑於詞彙對於語言教學的重要，專家學者亦著手進行整合及歸納學習者所應習得之詞彙量。Nation（1997）提到，二語學習者所應習得的詞彙量可由三個方面得知：(1) 目標語中包含的所有詞彙。(2) 目標語母語者所擁有的詞彙量。(3) 語言使用者在語言使用中所需的詞彙量。

首先，取得目標語中包含的所有詞彙，最直接的方法就是查詢目標語現有詞彙量最大的詞典（Dupuy 1974; Goulden, Nation & Read 1990）。然而，專家學者也發現語言隨著時代的演進，亦不斷地產生新的詞彙，因此難以斷定目標語確切的詞彙量。其次，自母語者所知曉的詞彙量評斷學習者所應習得之詞彙量。許多研究針對此進行分析，但評斷母語者詞彙量的方法多自個人的教育背景、智能及閱讀能力方面進行。如 Goulden, Nation & Read（1990）大略估計大學畢業生的詞彙量約是 20,000 個詞族。然而，Jamieson（1976）對非母語學習者及母語學習者詞彙習得增加率作比較，發現因兩者之間的起始點不同，即使兩者在某一階段詞彙習得增加率相當，但兩者間詞彙量上的差距並未因此拉近。因此，自母語者的詞彙量推估學習者的詞彙量或許可行，但若二語學習者的詞彙量要達到母語者所擁有的語彙量，對於大多數外語學習者而言挑戰甚大（Nation 1997）。第三方面，自語言使用者使用語言時所需詞彙，歸納出學習者所應掌握詞彙。該論點秉持著：即便語言中包含大量詞

彙，但非所有詞彙皆常使用。因此，Nation（1997）提出可依據詞彙在語言使用上出現的次數來判斷詞彙是否為常用詞，此一概念進而引導出詞彙頻率的概念。然而詞頻難以自生活中計算，因此許多專家學者投入語料庫建置上。期望自語料庫中歸納出的詞彙使用次數，以提供詞彙教學之參考。下節便將針對語料庫對於詞彙教學的貢獻加以深入討論。

### 1.3.2 語料庫詞頻對詞彙教學的貢獻

近年來，語料庫在語言研究中扮演的重要角色不言而喻，藉由語料庫可觀察字詞頻率及句法結構（Hunston 2002；王建新 2005）。同時，因語料庫提供了大量來自真實語境的語言資料，除了增加分析研究的可信度並可探究複雜的語言架構。Hunston（2002）進一步表示，即使沒聽過語料庫的人也在使用語料庫研究的產出。她進一步解釋，許多字典和參考書籍因語料庫而在五大方面：詞頻、詞彙的搭配組合、詞彙的變化、詞彙的真實性及語彙在語法上的作用等獲得了幫助。Kilgariff（1997）也提出透過大型語料庫可以計算出詞彙的頻率，而詞彙出現頻率越高，越能顯示出詞彙的重要性。並且在語言教學時越是高頻的詞彙越是需要先教。

以語料庫作為詞彙表制定依據最早可追溯回 1953 年 Michael West 以英文寫作文本語料庫作為依據，發表二千個英文常用詞表（General Service List, GSL）。該詞表為英語學習者建立出最實用的兩千個英語詞彙，其選字標準是將詞彙使用頻率、代表性等列為評量元素之一。其後 Engels（1968）以 GSL 作為研究素材，發現如果學習者掌握了此詞表的核心詞及其相關詞族，在英文閱讀時，可理解約 80% 的文章內容。

Time Benbow 於 1991 至 1994 年期間，集合各領域語料建置了英國國家語料庫（The British National Corpus, BNC）。該語料庫以口說（spoken）與書寫（written）英語為主，涵蓋了一億個字詞。其中，90% 為書寫部分，收集來自區域、國家的報紙，期刊、學術書籍、流行小說、已出版或尚未出版之信件或備忘錄、大專院校的論文、其他形式的文件等等。10% 為口說部分，取材自不同年齡、地區和社經背景，內容包括正式的商業或政府機關會議，以及廣播節目、電話訪問等（Kilgariff 1997）。雖然該語料庫的口語語料僅佔語料庫內容的 10%，然而，口語語料在建置上相當費時費力。不同於書面語語料取得較為容易，口語語料需先收集語音檔，再經由人工轉寫成為文字，進而集結成口語語料庫。

美國當代英語語料庫(Corpus of Contemporary American English, COCA)亦是相當著名的大型英語語料庫。由 Mark Davies 主持，自 1990 年至今(更新至 2011 年)收錄了約 4.5 億字。語料內容包括美國當代的口語、小說、流行雜誌、報紙和學術期刊五大類型，並平均分布於五大類型中，能反映出美國當代各類型語言的趨勢。

語料庫不僅能反映語言趨勢，亦可透過語料庫計算出各個詞彙使用頻率及詞彙之搭配關係。Adam Kilgariff 在 1997 年時，將 BNC 的詞彙，取頻次超過 800 次以上的詞，列出 BNC 最常用 6,318 筆詞詞表。Mark Davies 與 Dee Gardner 透過語料庫詞頻及詞彙搭配等特性，於 2009 年依詞頻排列，取 COCA 前 5,000 筆高頻詞編輯成美國當代高頻詞典(A Frequency of Dictionary of Contemporary American English)。該詞典於 2010 年出版，內容包含 COCA 前 5,000 筆高頻詞及其詞意、用法和搭配詞彙。

上述 BNC 及 COCA 語料庫不僅收錄了書面語語料，亦包含了建置費時且採集不易的口語語料，避免了一般語料庫僅能提供書面語詞彙卻不見口語用詞的現象。因此，以 BNC 及 COCA 語料庫詞頻編列的詞表，常受英語學習者及教師的重視。

除了一般通用詞語之外，語料庫也被運用到各類型專用語言詞表中。例如，Averil Coxhead (2000) 收集了 400 多位研究人員的文章，約 350 萬個字的學術用語，建置了學術英語語料庫。在此基礎上，Coxhead 以詞頻為依據發表了學術字彙表(Academic Word List)，組成學術英文最常用的 570 個字彙家族(word family)。該詞表常作為論文寫作的參考依據。

綜上所述，語料庫詞頻編列出的詞表在語言教學中佔有舉足輕重的地位和重要性，依據詞頻有系統地進行詞彙教學已成為當代的趨勢。然而，相較於語料庫詞頻在英文詞彙教學運用蓬勃發展，中文詞彙教學方面，使用語料庫詞頻作為選詞依據之研究及書籍仍處開發及發展階段。本研究將運用大型華語語料庫詞頻檢視 TOCFL 詞表及 HSK 詞表內容，期望能從中揭示兩詞表內容於語料庫中之詞彙使用頻率，提供詞彙教學作為選詞參考依據。詳細介紹於下節研究方法中加以論述。

## 2. 研究方法

如前所述，本研究有兩個主要目標：（1）比對兩岸著名的華語能力測驗－TOCFL 詞表及 HSK 之詞表收詞異同。（2）自大型網路語料庫中取高頻詞作為檢視詞表內容之參考。下面將逐步介紹本研究內容及工具：

### 2.1 研究內容

#### 2.1.1 兩岸華語詞彙表

華語文能力測驗，英文為 Test of Chinese as a Foreign Language（簡稱 TOCFL）。原 Test Of Proficiency-Huayu（簡稱 TOP）。自 2012 年開始，測驗分為五級：入門級、基礎級、進階級、高階級、流利級。入門級詞彙量為 500 個詞、基礎級 800 個詞、進階級 1500 個詞、高階級 5000 個詞、流利級 8000 個詞。張莉萍（2007）表示，華語八千詞主要選詞來源為中研院語料庫核心詞彙、通用詞彙、參考詞彙表、TOP 詞彙表、HSK 詞彙表，運用相對頻率、加權值的方式算出每一個詞的比重。詞彙量八千主要是依據鄭錦全（1998）「詞涯八千」的概念。

新漢語水平考試，又稱新 HSK。中國大陸漢辦於 2009 年推出了新漢語水平考試，同時也推出新 HSK 詞彙表，將詞彙量從原本的 8822 詞大幅縮減為 5000 詞。新 HSK 的五千詞分布於六級數中：第一、二級各為 150 個詞、第三級為 300 個詞、第四級有 600 個詞、第五級涵蓋 1300 個詞、第六級有 2500 個詞。以下稱新版 HSK 為 HSK。

#### 2.1.2 兩岸大型網路語料庫

Kilgariff（2004）提到語料庫的資料越大越能表現出完整及準確的語言行為，也越能計算出該語言中的常用詞彙。且詞頻的可靠性及穩定性取決於語料庫的大小及內容（張亦政、張文芝 1995）。因此，利用大型語料庫所產生的詞表較能體現出該語言詞彙的使用頻率及分布。近年來，隨著網路資源的大幅成長，以網路作為語料收集的來源可大量匯集各類用法，且網路語料最大的優勢在於隨著網站內容的不斷更新，語料量也得以擴增，所檢出的資料不僅是有大量語言範例作為後盾，亦可觀察當前語言的使用情形。

透過網路語料庫，可分析該地區常用字詞的使用頻率，以及相互搭配的使用情形。對於編纂字典、辭典來說是一個相當好的基礎，因為其所需要的大量例句、字頻及詞頻統計，都能在語料庫中獲得。同時，華語教師及學習者也能透過語料庫找出高頻常用詞彙作為教學順序的依據。教材編輯人員也

能採用語料庫的語料、詞組、句式當作教材編輯參考依據。以下介紹本研究使用的兩岸現行大型網語語料庫。

大陸簡體網路語料庫 zhTenten 是由 Adam Kilgarriff 自 2006 年開始，以網路語料庫的概念建立的 Tenten 網路語料庫系列，該系列語料庫包含了：德文、義大利文、西班牙文、英文、（簡體）中文等，且其他不同語言仍持續開展當中。Kilgarriff 透過詞彙速描系統（Sketch Engine）自網路上大量收集語料並經斷詞後集結成網路語料庫。該語料庫以 10 的 10 次方，約百億字為語料蒐集目標，故名為 Tenten。截至 2011 年，Tenten 語料庫的簡體字語料字數達 21 億餘字，稱 zhTenten（以下簡稱 Tenten）。由於網路語料自動大量收集網頁內容，蒐集過程中難免兼含中英文字詞，且由於中文的詞邊界不明顯，機器自動斷詞工具斷詞時，容易出現斷詞錯誤的情況。因此，本研究在取得 Tenten 語料後，先依頻率排序。而後人工刪除英文字、因斷詞而造成的詞意不完整（例：際關係）、特殊紀念日（例：六四）、專有名詞（例如：人名、地名、校名、路名、公司名、政府機關及政黨名稱等）後，排出前 1 萬高頻詞（第 1 萬詞之頻率為 10,386 頻次），作為檢視 HSK 詞表的語料庫高頻詞。

臺灣繁體大型語料庫方面，本研究採用臺灣師範大學建置的臺灣網路語料庫（TaiwanWac）作為對比參照的依據。該語料庫是以網路為語料（Web as Corpus, WAC）的概念建構而成，語料取材於臺灣網頁內容，再經系統斷詞，集結成大型網路語料庫，目前共蒐集了超過 3 億 8 千 4 百 97 萬字。比照 Tenten 語料庫的處理方式，本研究自 TaiwanWac 語料庫中，依照頻率高低節選詞彙。首先人工刪除英文字、因斷詞而造成詞意不完整之詞彙（例：公務人、們為、相關單）、用字錯誤（例：蕭炎）、簡體字（例：从、这个）、特殊紀念日（例：二二八、九二一）、專有名詞（例如：人名、地名、校名、路名、公司名、政府機關及政黨名稱等，不包含國家名），而後排出前 1 萬筆高頻詞（第 1 萬詞之詞頻為 1,519 頻次），作為檢視 TOCFL 詞表詞頻的對照。

### 2.1.3 研究內容前置處理

本研究首先取得 HSK 及 TOCFL 官方詞表，發現 TOCFL 將每一級的詞都加註了詞性，同時也將多音字列為不同詞條，因此，造成 TOCFL 詞表字詞上的重複情形。然而 HSK 詞表僅於第一級及第二級詞彙大綱標注詞性，HSK 三級以上詞彙則不標注詞性。避免兩岸在詞性等級上的看法影響比較結



果，本研究僅比對 TOCFL 及 HSK 詞條，不另比對各詞彙的詞性及分布級數。本研究觀察 HSK 詞表後發現，僅有四筆重複詞彙：「得」、「等」、「對」、「過」。因此，將 HSK 重複的詞彙刪除後，原 HSK 的 5,000 個詞降為 4,996 個詞。

而 TOCFL 詞表中含有不少多義詞及多音詞，本研究參考國外詞表（如 BNC 及 COCA 詞表）中不考慮一字多義的情形，對多義詞及多音詞進行前置處理。在多義詞方面，保留最先出現的詞，刪去後出現之相同詞彙。例如：「把」字在入門級詞彙中出現為介詞，之後在高階級詞彙中作為數量詞。因此刪去高階級中的「把」。TOCFL 詞表中的多音詞也是比照此作法。例如：「長」zháng（入門級）和「長」zhǎng（分別出現在進階級、高階級），則保留入門級中的「長」，刪去進階級和高階級中的「長」。因此，TOCFL 詞表中原 8,013 個詞減至 7,319 個詞。

另外，因兩岸詞表詞彙呈現方式不同，為使系統更能完整對照結果，也將兩詞表呈現方面作了統一。在 TOCFL 詞表中，部分詞彙以括號將詞彙所有的表達方式一併呈現。例如：「唱歌」一詞在 HSK 詞表中獨為「唱歌」，但 TOCFL 詞表則以「唱歌（兒）」呈現在詞表中。此兩個詞所代表的意思皆為「唱歌（to sing）」，為避免比對時的誤差，便統一調整為「唱歌」。

## 2.2 分析工具

本研究使用由程式設計師設計的對比工具進行詞表內容之比較，網址：<http://140.122.83.245/diff/>。下圖 1 為對比工具畫面。

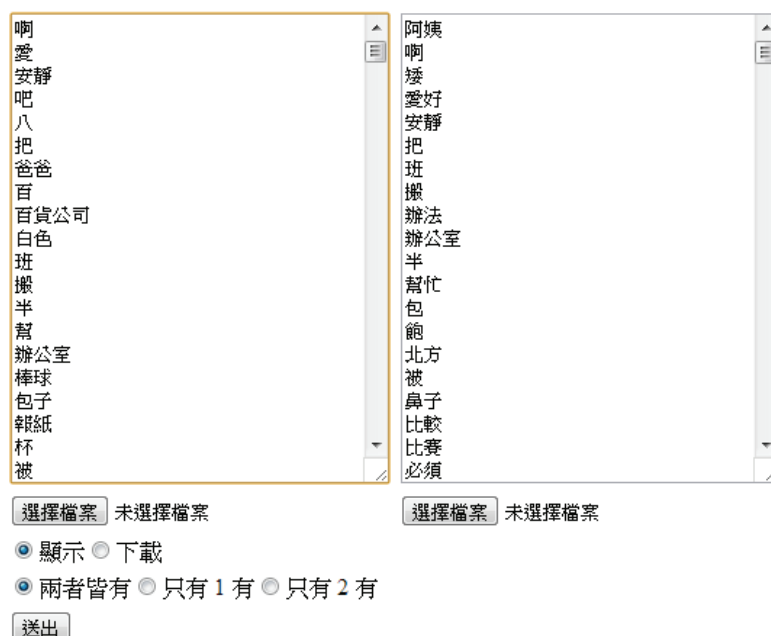


圖 1：對比工具

比對方式是將欲對比的兩詞表各別放在對比工具的左右欄位。若想得到兩詞表共同詞，點選「兩者皆有」。如欲檢視各別詞表中，不在兩者共同詞中，僅單獨出現在該詞表中的詞可點選「只有 1 有」或「只有 2 有」便可得知。1 代表左邊的詞表，2 代表右邊的詞表。

### 3. 研究結果與討論

#### 3.1 TOCFL 詞表及 HSK 詞表之異同

##### 3.1.1 兩詞表數量上的差異

因 TOCFL 及 HSK 測驗級數不同，詞彙量亦不同，實難比較其中差異。此外，TOCFL 雖然為每筆詞彙都標註了詞性，但 HSK 僅於第一級及第二級的詞彙大綱中列出詞彙詞性，其餘第三級以上則未逐一標示出各詞彙之詞性，難以對照兩詞表各級詞彙間的差異。因此迄今為止，少有文獻比對 TOCFL 及新 HSK 各級別程度之間的對照情形。

然而若單以詞條不加上詞性比對 TOCFL 詞表及 HSK 詞表總詞彙量仍可觀察出兩詞表間之異同。本研究取 TOCFL 詞表的 7,319 個詞與 HSK 詞表的 4,996 個詞對比後，結果得到兩詞表中，有 3,700 個詞為共有詞。但在 HSK

詞表的 4,996 個詞中有 1,296 個詞不在 TOCFL 詞表中，TOCFL 則有 3,619 個詞未收在 HSK 詞表中。本研究進一步觀察兩詞表間之相異詞，發現 TOCFL 及 HSK 在收詞上各具特色，下文針對 TOCFL 及 HSK 詞表間收詞之異同提出討論。

### 3.1.2 兩詞表收詞上的差異

自詞表內容探究兩詞表之異同。本研究歸納 TOCFL 和 HSK 在詞彙選擇及詞表編排上有下列五項差異：

#### 3.1.2.1 兩詞表詞彙用字差異

本研究發現 TOCFL 和 HSK 的 3,700 個共有詞中，某些詞所指的物品或意思是相同的，但在選詞用字上卻不同，比如：大陸稱「potato」為「土豆」，而臺灣稱為「馬鈴薯」。又比如「空調」與「冷氣（機）」，臺灣雖也說「空調」一詞，但大多講述「中央空調」一詞，一般日常使用以「冷氣（機）」為多，因此「空調」未收在 TOCFL 詞表中。因篇幅的關係，本研究不就兩岸詞語涵義上的差異作詳細深入地研究。僅列出 HSK 及 TOCFL 詞表中詞意相同但用字不同的例子於下表 1。

表 1：兩岸詞表中意思相同用字不同之詞彙

HSK	TOCFL	英文意義
服務員	服務生	waiter; waitress
空調	冷氣（機）	air conditioner
堵車	塞車	traffic jam
積累	累積	accumulate; accumulation
班主任	導師	class teacher
本科	大學	undergraduate
釐米	公分	centimeter
土豆	馬鈴薯	potato
乘務員	空服員/車掌（小姐）	crew member; train conductor
攝氏度	攝氏	Celsius degree
掙錢	賺錢	earn money
登機牌	登機證	boarding pass
互聯網	網路	internet

## 3.1.2.2 兩詞表詞彙呈現差異

本研究檢視 TOCFL 及 HSK 詞表的同時，發現不僅兩岸在詞彙的選擇上略有不同，詞彙呈現的方式也稍有差異。大陸漢辦宣稱 HSK 詞表遵循「字本位」及「詞本位」原則處理詞彙大綱，但在收錄詞彙時，部分詞彙是以詞組方式呈現，例如：HSK 收錄了「打籃球」、「做生意」等類型詞彙，在詞表中共發現十組，列於下表 2。

表 2：HSK 詞表中的動賓結構組合

詞彙	級別	詞彙	級別	詞彙	級別
打電話	L1	打籃球	L2	踢足球	L2
彈鋼琴	L2	放暑假	L4	做生意	L4
打噴嚏	L5	打招呼	L5	繫領帶	L5
涮火鍋	L6				

除了以上十組詞組外，其餘 HSK 的詞彙則是以獨立詞彙形式呈現，例如：「音樂」、「太極拳」、「牢騷」、「羽毛球」等。因此，HSK 對於詞彙列表方式不甚一致。而 TOCFL 詞表，則是將詞依詞性分開處理，例如：新 HSK 的「打籃球」在 TOCFL 詞表中分為「打」、「籃球」。又如「做生意」，在 TOCFL 中則拆為「做」、「生意」兩詞。因此，在詞彙的處理方面 TOCFL 較 HSK 一致化。

## 3.1.2.3 HSK 詞表獨有詞彙

TOCFL 詞表有 7,319 個詞，HSK 詞表有 4,996 個詞。因 TOCFL 詞彙量較大，那麼 HSK 的 5000 詞應包含在 TOCFL 詞表中，但是兩詞表共同詞僅 3,700 筆。HSK 詞表中有 1,296 個詞不在 TOCFL 詞表中。本研究發現在這些詞中除了上述兩岸詞彙用字及編排方式不同之外，尚有部分 HSK 收錄詞彙類別在 TOCFL 詞表中屬少見之詞彙類別。本研究歸納出下列八項類別，詳見表 3。（詞彙後括號內數字代表該詞在 HSK 詞表中的級別）

表 3：HSK 詞表中不同於 TOCFL 詞表之詞彙類別

類別	例子（選自 HSK 的 1,296 個詞）
影視網路用語	電子郵件（3）、充電器（5）、光碟（5）、滑鼠（5）、數碼（5）、登錄（6）、搖滾（6）、電源（6）...
大陸官方用語	大使館（4）、人民幣（4）、國務院（6）、領事館（6）、公安局（6）、省會（6）、書記（6）...
地名（風景名勝）	北京（1）、長城（4）、長江（4）、赤道（6）...
特殊地方現象	胡同（5）
自然資源/現象	雪花（4）、雪人（5）、銅礦（6）、冰雹（6）、海濱（6）、湖泊（6）、氫（6）、二氧化碳（6）
數理用詞	分之（4）、立方（5）、百分點（6）、化驗（6）
醫學用詞	腫瘤（6）、副作用（6）
稱謂	公主（5）、皇后（5）、王子（5）、姥姥（5）

## 3.1.2.4 HSK 詞表的四字詞

本研究在探究 HSK 詞表時，發現 HSK 詞表中的四字詞數目較 TOCFL 詞表的四字詞多，仔細數算後發現五千詞中有 126 筆四字詞（見附件一），比 TOCFL 詞表的 29 組四字詞還高出逾 4.3 倍之多。細查兩詞表中的四字詞後，發現僅有五組詞彙相同：「公共汽車」、「不好意思」、「通貨膨脹」、「無可奈何」、「總而言之」。HSK 詞表的 126 組四字詞中以成語、俗語居多。例如：「喜聞樂見」、「統籌兼顧」、「任重道遠」、「齊心協力」、「鍥而不舍」、「揠苗助長」等。進一步查詢 HSK 四字詞於 Tenten 語料庫之頻率，發現有 38 組四字詞的詞彙低於 1,000 頻次，列於表 4。（詞彙下方的數字代表該詞在語料庫出現的頻次。NA 表示在 Tenten 語料庫中未出現該詞）

表 4：HSK 詞表中在 Tenten 語料庫詞頻低於 1,000 頻次之四字詞

公共汽車 (NA)	L2	電子郵件(NA)	L3	不好意思 (NA)	L5	高速公路 (NA)	L5
名勝古跡 (NA)	L6	一路平安(NA)	L6	愛不釋手 (925)	L6	揠苗助長 (306)	L6
飽經滄桑 (NA)	L6	波濤洶湧(NA)	L6	稱心如意 (995)	L6	得天獨厚 (NA)	L6
丟三落四 (335)	L6	東張西望(975)	L6	飛禽走獸 (NA)	L6	風土人情 (NA)	L6
各抒己見 (114)	L6	根深蒂固(NA)	L6	畫蛇添足 (561)	L6	急於求成 (NA)	L6
繼往開來 (NA)	L6	見多識廣 (905)	L6	錦繡前程 (NA)	L6	空前絕後 (749)	L6
苦盡甘來 (433)	L6	豈有此理 (235)	L6	恰到好處 (NA)	L6	熱淚盈眶 (NA)	L6
深情濃誼 (NA)	L6	素食主義 (153)	L6	討價還價 (NA)	L6	無可奉告 (229)	L6
想方設法 (620)	L6	懸崖峭壁(NA)	L6	煙花爆竹 (NA)	L6	優勝劣汰 (NA)	L6
斬釘截鐵 (684)	L6	知足常樂(NA)	L6				

針對這些較低頻次的 HSK 詞彙，本研究歸納其主要原因可能有下列三點：(1) 語料庫中有些詞彙是以簡化後的形式呈現，如：「公共汽車」在 Tenten 語料庫中是以日常簡化說法—「公車」呈現，詞的頻率是 28,483 次。又如「電子郵件」常被簡稱為「電郵」，詞頻為 2,277 頻次。(2) 電腦自動斷詞 (post tagging) 因素，如：「一路平安」在 Tenten 語料庫中斷詞成「一路」(詞頻：66,257) 及「平安」(詞頻：85,661)。又例如「名勝古跡」在語料中分別斷為「名勝」(詞頻：9,474) 及「古跡」(詞頻：13,300)。(3) 口語常見用詞在書面語語料庫中的頻率相對較低，如：「無可奉告」在語料庫中僅出現 229 次。

除上述因素導致一般認為使用頻率高的詞彙排除在語料庫前 1 萬高頻詞表外，單就詞頻作為參考標準，某些四字詞使用的頻率實為不高，如「各抒己見」(詞頻：114)、「素食主義」(詞頻：153)。此外，HSK 詞表不重覆之詞彙僅有 4,996 個詞，但其中卻包含了 126 筆四字詞。是否有必要納入如此多的四字詞，值得商榷。

### 3.1.2.5 TOCFL 詞表的「兒化」詞

本研究觀察 TOCFL 詞表及 HSK 詞表後，發現 TOCFL 詞表編排原則之一是將「兒化韻」標示出來，但 HSK 詞表中無特別將「兒化」音的詞彙標記出來。呂叔湘 (2005) 主編的《現代漢語八百詞》裡對「兒」的解釋為「加在名詞性成分或其他成分後面，構成名詞。讀時與前面合成一個音節，叫做

『兒化』，書面上有時不寫出來。『兒化』現象和地方方言很不一致，這裡以北京多數居民口語為準。」（呂叔湘 2005，頁 191）即「兒化」現象多半出現在口語或地方方言中，如北京地方方言。但在 HSK 詞表中僅有 11 個「兒化」詞，僅有「一會兒」一詞與 TOCFL 詞表相同。另一方面將這 49 個「兒化」詞對比 TaiwanWac 語料庫前 1 萬詞，僅出現「這兒」、「那兒」。雖說中文詞彙中有部分詞彙已和「兒化」連接緊密，屬於慣用語。例如：「一點兒」、「有點（兒）」、「差（一）點（兒）」、「一塊（兒）」、「待會（兒）」、「個兒」等詞，本身在口語使用上已具備特定的詞意，在教學上不宜省略「兒」，其餘詞彙則非然。因此，詞表中是否需要特別將所有可加或可不加「兒化」的詞彙皆標示出來，本研究認為可多加評估商榷<sup>4</sup>。下表 5 為 TOCFL 詞表「兒化」詞及分布級數。

表 5：TOCFL 詞表中的「兒化」詞

餐廳/飯館兒	L1	點（兒）	L2	飯館（兒）	L4
唱歌（兒）	L1	裡邊（兒）	L2	份（兒）	L4
船（兒）	L1	聊天（兒）	L2	個兒	L4
好玩（兒）	L1	鳥（兒）	L2	口袋（兒）	L4
花（兒）	L1	旁邊（兒）	L2	空（兒）	L4
哪裡/兒	L1	一半（兒）	L2	人情味（兒）	L4
那裡/兒	L1	一會兒	L2	味（兒）	L4
男孩（子/兒）	L1	冰塊（兒）	L3	相片（兒）	L4
女孩（子/兒）	L1	差（一）點（兒）	L3	待會兒	L4
玩（兒）	L1	歌（兒）	L3	多半兒	L4
小孩（子/兒）	L1	好好（兒）	L3	活（兒）	L5
一點兒	L1	沒事（兒）	L3	勁（兒）	L5
有點（兒）	L1	一塊（兒）	L3	開頭（兒）	L5
有空（兒）	L1	一下子/兒	L3	老頭（兒/子）	L5
魚（兒）	L1	茶館（兒）	L4	字眼（兒）	L5
這裡/兒	L1	待會（兒）	L4	邊（兒）	L2
多半（兒）	L4				

<sup>4</sup> 針對 TOCFL 詞表的「兒化」詞收詞情形，匿名審查委員亦指出，「兒化」詞之構詞用法基本上在大陸南方及台灣地區已漸漸減少使用，故 TOCFL 詞表中是否需要將「兒化」的詞彙皆標示出來，值得多加評估商榷。

綜合上述結果，TOCFL 和 HSK 詞表之異同可分別自詞彙量及詞表收詞上觀察。第一，在詞彙量上，兩詞表的共同詞為 3,700 個詞，TOCFL 詞表中有 2,089 個詞未出現在 HSK 詞表中，而 HSK 詞表亦有 1,353 筆詞彙未出現在 TOCFL 詞表中。第二，兩詞表在詞表收詞方面，主要有 5 項差異：(1) 兩岸詞表部分用字不同。(2) HSK 詞表呈現部分動賓結構詞組，在 TOCFL 詞表中則未見動賓結構詞組。(3) 兩詞表收錄之詞彙類別不同。(4) HSK 詞表四字詞數量是 TOCFL 詞表四字詞的 4.3 倍。(5) TOCFL 詞表「兒化」詞的數量高於 HSK 詞表「兒化」詞。

自兩詞表比對結果中可發現，兩詞表無論是數量或收詞內容及類別都呈現出明顯差異，造成此差異的因素可能是兩詞表在選詞原則、選詞依據等方面的不同。比如新版 HSK 詞表是匯集海內外學者專家，就對外漢語教學的實際情況進行調查與選詞。而 TOCFL 詞表則是以中研院平衡語料庫核心詞彙為主，並參考其他詞彙表進行選詞。前者未將詞頻的概念納入選詞依據，後者則是選用了規模相對較小的語料庫。以至於兩詞表中皆出現部分較少用的詞彙，例如：「仙丹」、「號子」、「勞駕」、「殖民地」等。為求詞表能更客觀地呈現母語者詞彙實際使用情形，本研究借鏡國外行之有年的大型語料庫研究方法，分別自簡體及繁體大型網路語料庫抽取前 1 萬筆高頻詞，對比 HSK 及 TOCFL 詞表，或可自對比結果中彙整出未收錄於詞表之高頻詞，以作為日後詞表修正之參考。

### 3.2 TOCFL 詞表與語料庫詞高頻詞之異同

詞表在語言教學和測驗上佔舉足輕重的位置，亦是學習者掌握語言的關鍵因素之一，因此詞彙學習先後順序相對重要，建議可採取常用先教的原則。然而哪些詞彙屬於常用詞，研究者及教學者往往眾說紛紜。對此，本研究檢視大型語料庫的高頻詞，以詞頻的高低作為是否常用之依據。

Kilgariff et al. (2004) 表示，語料資源越大越能體現出完整及準確的語言行為。因此本研究取大型網路語料庫的高頻詞與 TOCFL 詞表對比，以探究詞表中的詞彙是否貼近語料高頻詞彙。

本研究取臺灣大型網路語料庫 TaiwanWac 詞表，先刪除因電腦自動斷詞而造成詞意不完整之詞彙、英文字、用字錯誤、簡體字、特殊記念日（例如：二二八、九二一等）及專有名詞（專有名詞不包含國家名），再自語料中依照詞頻取前 1 萬筆高頻詞彙（第 1 萬詞之詞頻為 1,519 頻次）。將 TOCFL 詞表



及 1 萬筆 TaiwanWac 語料庫高頻詞以對比工具分析，得到 TOCFL 詞表中有 5,230 個詞與語料高頻詞相同，但另有 2,089 筆 TOCFL 詞彙不包含在語料前 1 萬高頻詞當中。本研究進一步探究 2,089 個詞後發現，因 TaiwanWac 屬網路書面語語料庫，故 2,089 筆詞彙中包含部分口語常用詞彙，例如：「做飯」、「不客氣」、「不好意思」、「來得及」等詞。

然而在 2,089 筆 TOCFL 詞彙中部分詞彙的頻率偏低，例如：「仙丹」(詞頻 171)、「號子」(詞頻 190)、「鑲」(詞頻 208)、「狐疑」(詞頻 92)、「等價」(詞頻 124)、「統制」(詞頻 168)。另外，本研究自 2,089 筆詞彙中觀察到 TOCFL 詞表中收錄了詞意相近／相同詞彙，且這些近義／同義詞以不同詞條分布於 TOCFL 詞表中。例如：入門級－「裡面」vs. 基礎級－「裡」vs. 進階級－「裡(面)」。入門級－「冬天」vs. 進階級「冬」vs. 高階級「冬季」。近義／同義詞是否必要以不同詞條納入詞表中，值得商榷。

根據李如龍及吳茗(2005)提到對外漢語詞彙教學若要提高成效，建議實行常用先教的原則。而語料庫高頻詞可作為詞彙教學參考。對此，本研究比對 TaiwanWac 語料庫前 1 萬筆高頻詞與 TOCFL 詞表，其中有 4,770 筆高頻詞未被納入於 TOCFL 詞表中。顯示語料庫高頻詞中仍有部分詞彙可建議納入在 TOCFL 詞表中。因此，本研究按 TaiwanWac 語料庫之詞頻由高至低，取詞彙頻率超過 1 萬次以上之詞彙，刪除因電腦自動斷詞而造成詞意不完整之字詞後，共 270 筆可單獨成詞之建議詞彙列於下表 6 供詞表編輯及學者專家參考。

表 6：建議可納入 TOCFL 詞表的 270 筆語料庫高頻詞

臺灣	這種	很多	年度	系	美元	曰	之外	社區	季
最高	國民	股份	研究所	中華	原本	兩岸	快速	業者	中華民國
研發	整合	品牌	原住民	股票	印度	歐洲	臺幣	事項	回應
回到	愈	考量	市	想要	今日	中共	藉由	展	整個
董事	亞洲	公告	立委	致	族群	契約	期待	身心	卻
出版社	警方	村	創意	留言	理念	團隊	大幅	幾個	大多
資產	組成	紀錄	留下	特性	運作	街	研討會	核心	再次
縣市	預期	數位	獲利	型	特定	慢慢	專案	立法	近年
更新	司法	基於	全國	影像	日前	知名	重建	在一起	臺灣人
訊息	最近	被告	營運	行銷	還沒	新加坡	人文	理事	認同
使用者	平臺	設定	興建	越南	病患	園區	玩家	公頃	建築物
案例	將會	家族	基金會	電信	澳洲	一路	描述	依法	個案
臺商	提昇	績效	訂定	強化	海岸	定位	搭乘	全民	宣導
檢察官	財務	短期	認知	判決	效益	並非	外籍	效應	修行
分局	給付	互動	經由	回覆	股價	市府	配偶	委員會	客家
學系	美食	申報	組合	保育	認證	關注	因應	訴訟	實務
建構	得知	加拿大	高等	話	料理	營收	購物	用地	同仁
本文	完美	董事長	院長	打造	志工	告知	作法	法規	體制
此時	股東	透露	網友	手法	內心	公車	巷	中國人	筆者
功德	捷運	境界	簡介	卻是	上漲	種種	超	評鑑	天地
導彈	素	縣府	臺股	重要性	近期	決策	強制	福	做法
情感	工商	被害人	俄羅斯	健保	尾	投資人	英語	行情	雇主
檢測	性	排放	面板	步道	足夠	條文	版本	非洲	保健
海上	泰國	部位	競爭力	日本人	主持人	餐	完工	含有	生涯
分類	持有	西元	特質	房地產	做為	機車	當事人	配置	建置
常見	約定	昨日	查詢	城	偏	太平洋	層面	工作者	可能性
疫苗	民意	共識	債務	警察局	諮詢	層次	現今	想起	優質

### 3.3 HSK 詞表與語料庫詞高頻詞之異同

本研究另取簡體大型網路語料庫 Tenten 詞表，檢視並刪除因電腦自動斷詞而造成詞意不完整之詞彙、英文字、特殊記念日、專有名詞（專有名詞不包含國家名）後，再依照詞彙頻率取語料前 1 筆高頻詞（第 1 萬詞之詞頻是 10,386 頻次）。將 HSK 詞表及 Tenten 語料庫前 1 萬筆高頻詞作對比。比對結果顯示 HSK 詞表中有 3,643 個詞在語料庫高頻詞中，但有 1,353 個 HSK 詞彙不在語料前 1 筆筆高頻詞之內。

與 TaiwanWac 語料庫相同，Tenten 語料庫來源亦是網路書面語。因此，1,353 筆 HSK 詞彙未在 Tenten 語料庫前 1 萬詞中也包含部分較偏口語之用詞，如：「不客氣」、「打電話」、「不好意思」、「看不起」、「忍不住」、「算了」等詞。此外，在 Tenten 語料庫前 1 萬筆高頻詞與 HSK 詞表的對比中，另有 6,357 筆高頻詞未被納入於 HSK 詞表，顯示 Tenten 語料庫高頻詞中仍有部分詞彙可建議收錄於 HSK 詞表中。因此，本研究取詞彙頻率在 5 萬頻次以上<sup>5</sup>，刪除因電腦自動斷詞而造成詞意不完整之字詞後，依據 Tenten 語料庫詞頻由高至低列出 1,408 筆建議詞，可提供 HSK 詞表編輯及考試單位參考。因篇幅關係，僅先列出前 500 筆建議詞（第 500 筆詞為「綠化」，其詞頻：115,357 頻次）列於下表 7 供詞表編輯及學者專家參考。

表 7 建議可納入 HSK 詞表的前 500 筆語料庫高頻詞

將	及	研究	服務	大學	沒有	他們	市	工程	此
於	教學	學院	加強	國	該	科技	人民	有關	一些
天	這些	進一步	這樣	提出	具有	幹部	不同	以上	美國
推進	即	機制	未	取得	得到	能夠	科研	起	今年
很多	有效	黨員	強	有限	深入	學科	高校	國內	重大
提升	進入	電子	第	委員會	推動	平臺	品牌	只有	村
人們	期	公共	主任	加快	衛生	黨委	大學生	看到	關注
處	增強	受	見	醫療	據	採用	那麼	投入	青年
職工	加大	出來	部	日本	以下	心	汽車	指出	高度

<sup>5</sup> 本研究取 TaiwanWac 詞頻 1 萬頻次以上的詞作為 TOCFL 建議詞彙，然而 TaiwanWac 僅有近 4 億字。相較於此，Tenten 語料庫的 21 億字，約是 TaiwanWac 的五倍。本研究設定 1 萬的五倍作為 Tenten 語料庫詞頻依據標準，自 Tenten 語料庫中挑選詞頻達 5 萬頻次以上之詞彙作為 HSK 建議詞彙。

應當	大會	基層	手	這裡	負責人	網	間	帶來	自身
者	畢業生	大力	小組	理念	體制	約	高等	做到	美
仍	辦	強化	校園	抓	社會主義	大量	英語	全球	長期
院	打	僅	監管	早	執法	線	無法	實驗室	員工
周	審計	話	學會	團隊	最終	專項	創建	改造	全體
依法	考生	居民	研究所	美元	主體	有些	普通	市委	快速
中華	道路	那些	招生	車	手段	如此	學者	有的	自我
內部	令	做出	法	一下	不少	生	先後	句	法規
環保	啟動	你們	代	思路	辦學	初	相互	下來	畝
消費者	優先	總體	最高	近年	今後	公路	犯罪	院校	產權
引進	分子	若	講話	施工	款	職責	邊	研發	同樣
之前	為主	常	打造	步	構建	在於	平方米	師生	臺灣
下降	審批	頭	這麼	等等	它們	發展觀	綠色	園區	信息化
城鄉	節能	建	住房	家長	報	商務	英國	副書記	尚
抓好	違法	酒店	整合	佳	人數	國外	省級	院長	建成
靠	鎮	購買	變	小學	檢測	現代化	連續	國有	推出
應急	氛圍	西方	章	出版社	山	找到	含	圖	測試
課堂	上述	上市	定位	市民	醫學	收費	編制	對外	中學
店	怎樣	父母	房地產	省委	班子	監測	感到	具	自治區
深化	高速	人文	優	難以	必	工藝	再次	優質	整治
支付	審核	有利於	行銷	政協	車輛	副主任	使得	營造	校區
發放	工會	房	如下	局	群體	德國	相比	處於	帶動
有所	新型	基於	物法	廉政	公安	僅僅	國內外	到位	出臺
文字	爭	稅收	同比	類型	大賽	工商	加入	配置	易
條例	城鎮	分為	高效	評審	招聘	面向	文	論	人力
通信	輪	藥品	招商	轉化	評選	中共	眼	面	會員
隨後	上升	提升	局長	拓展	管	公眾	做法	適用	剛
民生	村民	高中	聽取	現有	少數	分類	眾多	電力	院士
導師	座談會	升級	學員	上漲	納入	互動	有力	著力	部長
競爭力	立法	赴	須	地點	認證	那樣	鐵路	股	流動
設	變成	幼兒	當年	統籌	雖	藥物	法國	所在	志願

身	知名	晚	圖書	流程	詞	新生	工作者	名稱	扶持
法制	主持人	支部	讀者	適當	真的	近期	實際上	瘦身	化工
海外	度	版	證券	古	抓住	房屋	加	收集	只是
轉型	積極性	辦公	階級	視頻	國民	網友	國家級	近日	研討會
運作	垃圾	各位	例	加上	預計	面試	強大	某些	農產品
資料庫	監控	病	留下	走向	表	韓國	病人	意味	讀書
筆	有機	團委	法學	怕	城區	誠信	層面	外國	一行
最新	在一起	哪些	期刊	遠程	累計	推行	產業化	城	美好
種植	幫	班級	處罰	填寫	投資者	遊客	通報	那裡	用地
社團	組長	求	研究員	收到	回到	科學家	監察	融資	綠化

#### 4. 結論

中文詞彙教學時，TOCFL 詞表及 HSK 詞表常被視為詞彙教學指南。然而，現今少有研究針對兩岸重要詞表內容作分析及比對。因此，本研究將 TOCFL 詞表及 HSK 詞表內容以對比工具進行對比，藉以觀察兩詞表收詞之異同。其次，以大型網路語料庫之高頻詞對比兩詞表，以作為建議收錄詞彙之參考。以下依序摘要研究主要成果，並就所得研究結果分別對 TOCFL 詞表及 HSK 詞表提出詞表增修建議。

##### 4.1 TOCFL 詞表及 HSK 詞表間的異同

兩岸重要華語能力測驗——TOCFL 及 HSK 廣受華語教學研究相關人員的重視。由兩測驗所延伸出的詞表在華語教與學上亦佔關鍵角色。然而對於 TOCFL 詞表及 HSK 詞表內容對照及探討之相關研究較為少見。因此，本研究以目前 TOCFL 詞表 8000 詞及 HSK 詞表的 5000 詞作為研究素材，對比兩岸詞表，並以質化和量化結果展現兩岸詞表之間的異同。

在詞彙數量上，TOCFL 詞表的 8000 詞與 HSK 詞表的 5000 詞之差距顯而易見。初步觀察發現 TOCFL 詞表及 HSK 詞表內皆出現重複詞彙。有鑑於國外英語詞表多不考慮詞性，本研究以不重複詞條數量對比兩詞表（刪除多義詞及多音詞之重複詞彙後，TOCFL 詞表有 7,319 筆詞條，HSK 詞表則有 4,995 筆詞條）。兩詞表比對結果顯示 TOCFL 詞表及 HSK 詞表間的共同詞為 3,700 筆詞。HSK 詞表的 4,996 筆詞中有 1,296 筆詞不在 TOCFL 詞表中。而

TOCFL 詞表中的 7,319 筆詞彙則有 3,619 筆詞不在 HSK 詞表之中。顯然兩岸在收詞的選擇及理念上不太一致。

TOCFL 詞表與 HSK 詞表除了在詞彙數量上有顯著的差異外，在詞彙的選擇及編排上也各有所偏好。TOCFL 多以詞的最小意義歸納每個詞彙，並將各個詞彙的詞性及詞意編列成冊。另一方面，HSK 雖以「字本位」及「詞本位」編詞表，但部分詞彙是以詞的搭配組合列於詞表中，舉例來說，「彈鋼琴」、「打籃球」、「繫領帶」等都是 V+N 的搭配組合。然而，這樣的分類在整體看來是少數，只有十組搭配組合，顯示 HSK 詞表在詞彙呈現上較缺乏一致性。

HSK 詞表中的四字詞，如：「通貨膨脹」、「喜聞樂見」、「統籌兼顧」等的數量較 TOCFL 多出 4.3 倍之數。而 HSK 的四字詞中雖包含口語常見詞語，例：「不好意思」、「總而言之」等詞，但總量上還是以成語、俗語居多。成語、俗詞在詞意上多包含文化及典故，使用上也多需考量場合及文體，部分四字詞在語料庫出現的頻次較低，且現行 HSK 詞表僅有 5,000 詞，詞表內卻有 126 筆四字詞。因此，建議 HSK 詞表於四字詞的收詞上可再多加考量。

另一方面，TOCFL 的兒化音詞數量較 HSK 多 38 筆。雖然中文詞彙中有部分詞彙已和「兒化」連接緊密，屬於慣用語，例如：「一點兒」、「有點(兒)」、「差(一)點(兒)」、「一塊(兒)」、「待會(兒)」、「個兒」等詞，本身在口語使用上已具備特定的詞意，在教學上不宜省略「兒」。然而，其餘詞彙則非無「兒」字不可，例如：「魚(兒)」、「花(兒)」。

雖然 TOCFL 詞表特別將「兒」字以括號分別開來或以斜線劃分，但建議 TOCFL 詞表可簡化詞表呈現方式，將常用詞彙使用形式提供給詞表使用者。例如：「魚(兒)」可簡化為「魚」，「一塊(兒)」簡化為「一塊兒」。

#### 4.2 TOCFL 詞表及 HSK 詞表與大型語料庫高頻詞間的異同

TOCFL 詞表及 HSK 詞表雖在詞彙量及收詞原則上不大一致，但兩詞表皆包含部分低頻詞彙，如：TOCFL 八千詞表中的「仙丹」、「號子」、「鑲」、「狐疑」、「等價」、「統制」等及 HSK 五千詞表中的「勞駕」、「殖民地」等。HSK 詞表雖已經由專家學者基於海外漢語教學實際情況，自原本的八千多詞刪減至五千詞。然而缺少詞頻的輔助，較難反應出語言的實際使用情形。另一方面，TOCFL 詞表的八千詞是以鄭錦全（1998）「詞涯八千」的概念出發，並以語料庫計算出各詞的使用頻率。然而 TOCFL 詞表所採用的語料庫是中央研究院平衡語料庫，就現今語料庫規模而言，屬於規模較小之語料庫。而網

路語料庫的優點之一是網路上文體及語料內容皆較為豐富，且得以隨著網路內容的更新不斷累積語料。因此，本研究自兩岸大型網路語料庫—繁體語料庫採 TaiwanWac 語料庫（語料庫大小約 4 億字），簡體語料庫取用 Tenten 語料庫（語料庫大小近 21 億字）中依頻率取高頻前 1 萬詞分別與兩詞表作比對。比對結果顯示：（1）TOCFL 詞表中的 7,319 筆詞彙中，有 5,230 筆詞與 TaiwanWac 語料庫前 1 萬筆高頻詞相同，但有 2,089 筆 TOCFL 詞彙不在語料庫高頻詞之中。（2）HSK 語表的 4,996 筆詞中有 3,643 個詞出現在 Tenten 語料庫前 1 萬筆高頻詞中，其餘 1,353 個 HSK 詞表則落在語料庫前 1 萬筆高頻詞之外。

此外，TOCFL 詞表和 HSK 詞表分別與大型語料庫前 1 萬筆高頻詞比對後，另有以下兩項發現：（1）TOCFL 詞表及 HSK 詞表中皆有部分詞彙在網路語料庫中使用頻率較低。（2）大型網路語料庫中，有不少高頻詞未納入兩岸現有詞表，因此建議可考慮納入兩詞表當中。針對以上兩項發現，本研究取語料庫詞彙依照詞頻排序，分別自 TaiwanWac 語料庫及 Tenten 語料庫中取詞頻達 1 萬頻次（TaiwanWac 語料庫）及 5 萬頻次（Tenten 語料庫）以上的詞彙，分別提出 270 筆（選自 TaiwanWac 語料庫）及 1,408 筆（選自 Tenten 語料庫，因篇幅的限制僅列出前 500 筆詞彙）建議詞供兩詞表作為增修參考。

TOCFL 詞表及 HSK 詞表常作為華語教學之課程分級依據及編輯方針，但若以語言教學中常用的先教的概念檢視，建議兩詞表可參考詞彙使用頻率調整各級詞彙，並可考量將語料庫高頻詞代替部分低頻詞彙優先納入詞表中。

然而，因本文在研究上仍有許多的限制和侷限需加以改進。諸如所用的大型網路語料庫的內容雖為龐大，但語料來源以書面語為主，若用於詞表上便難以完整呈現口語常用詞彙。未來若能學習 BNC 及 COCA 等語料庫作法，加入口語語料以提升高頻核心詞的涵蓋面。期待未來研究能加入更多華語口語語料以提升詞彙涵蓋量及多元性。

詞表是語言學習的核心內容，且在語言學習上佔舉足輕重的角色。兩岸因應華語教學的需求，皆推出重要詞表。然而現今少有針對兩岸詞表相關研究。因此，本研究首先藉著比較兩岸重要詞表找出異同，以提供兩岸詞表編輯人員參考。由於兩詞表數量及收詞內容呈現諸多差異，且其選詞皆未能參考大型語料庫詞頻，為求更客觀、更符合母語者語言使用情形的選詞依據，本研究自大型語料庫計算出高頻詞，並運用這些高頻詞檢視 TOCFL 詞表及 HSK 詞表。再就未收入兩岸詞表之高頻詞，依詞頻高低提出建議可納入兩岸

詞表之詞彙。儘管研究仍有許多不足之處，但期望以拋磚引玉的方式，對兩岸詞表在編列及應用上提出建議，使詞表能更貼近學習者的需求，對華語教學提供更大的幫助。



附件一 HSK 四字詞及級別分布

公共汽車 (NA)	(二級)	電子郵件 (NA)	(三級)	不好意思 (NA)	(五級)	高速公路 (NA)	(五級)
名勝古跡 (NA)	(五級)	一路平安 (NA)	(五級)	愛不釋手 (925)	(六級)	安居樂業 (6,287)	(六級)
揠苗助長 (306)	(六級)	半途而廢 (2,006)	(六級)	飽經滄桑 (NA)	(六級)	波濤洶湧 (NA)	(六級)
博大精深 (4,472)	(六級)	不可思議 (6,418)	(六級)	不相上下 (1,382)	(六級)	不屑一顧 (2,298)	(六級)
不言而喻 (7,540)	(六級)	不擇手段 (2,594)	(六級)	層出不窮 (10,221)	(六級)	朝氣蓬勃 (6,559)	(六級)
稱心如意 (995)	(六級)	川流不息 (1,395)	(六級)	從容不迫 (1,602)	(六級)	當務之急 (10,005)	(六級)
得不償失 (3,375)	(六級)	得天獨濃 (NA)	(六級)	丟三落四 (335)	(六級)	東張西望 (975)	(六級)
斷斷續續 (2,459)	(六級)	二氧化碳 (24,942)	(六級)	飛禽走獸 (NA)	(六級)	風土人情 (NA)	(六級)
各抒己見 (114)	(六級)	根深蒂固 (NA)	(六級)	供不應求 (8,713)	(六級)	歸根到底 (4,427)	(六級)
後顧之憂 (12,428)	(六級)	畫蛇添足 (561)	(六級)	恍然大悟 (2,901)	(六級)	急功近利 (7,955)	(六級)
急於求成 (NA)	(六級)	繼往開來 (NA)	(六級)	家喻戶曉 (7,955)	(六級)	見多識廣 (905)	(六級)
見義勇為 (8,251)	(六級)	竭盡全力 (4,293)	(六級)	津津有味 (2,944)	(六級)	錦繡前程 (NA)	(六級)
兢兢業業 (9,095)	(六級)	精打細算 (2,478)	(六級)	精益求精 (9,840)	(六級)	舉世聞名 (2,381)	(六級)
舉世矚目 (5,648)	(六級)	舉足輕重 (9,551)	(六級)	聚精會神 (4,424)	(六級)	刻不容緩 (5,188)	(六級)
空前絕後 (749)	(六級)	苦盡甘來 (433)	(六級)	理所當然 (8,396)	(六級)	理直氣壯 (3,737)	(六級)
力所能及 (9,927)	(六級)	絡繹不絕 (6,321)	(六級)	名副其實 (9,407)	(六級)	莫名其妙 (5,688)	(六級)
難能可貴 (4,077)	(六級)	迫不及待 (5,778)	(六級)	齊心協力 (12,830)	(六級)	豈有此理 (235)	(六級)
迄今為止	(六級)	恰到好處	(六級)	千方百計	(六級)	潛移默化	(六級)

(9,249)		(NA)		(22,125)		(11,826)	
鏗而不捨 (4,017)	(六級)	輕而易舉 (3,905)	(六級)	全力以赴 (21,401)	(六級)	熱淚盈眶 (NA)	(六級)
任重道遠 (3,886)	(六級)	日新月異 (10,955)	(六級)	深情濃誼 (NA)	(六級)	實事求是 (41,687)	(六級)
肆無忌憚 (3,292)	(六級)	素食主義 (153)	(六級)	滔滔不絕 (2,895)	(六級)	討價還價 (NA)	(六級)
天倫之樂 (1,354)	(六級)	通貨膨脹 (18,870)	(六級)	統籌兼顧 (13,700)	(六級)	微不足道 (5,223)	(六級)
無動於衷 (2,657)	(六級)	無精打采 (1,087)	(六級)	無可奉告 (229)	(六級)	無可奈何 (4,908)	(六級)
無理取鬧 (1,297)	(六級)	無能為力 (4,503)	(六級)	無窮無盡 (1,966)	(六級)	無微不至 (4,578)	(六級)
無憂無慮 (2,373)	(六級)	物美價廉 (2,769)	(六級)	喜聞樂見 (9,113)	(六級)	相輔相成 (8,702)	(六級)
想方設法 (620)	(六級)	小心翼翼 (6,621)	(六級)	欣欣向榮 (3,725)	(六級)	新陳代謝 (11,987)	(六級)
興高采烈 (3,016)	(六級)	興致勃勃 (6,680)	(六級)	懸崖峭壁 (NA)	(六級)	雪上加霜 (3,717)	(六級)
循序漸進 (10,988)	(六級)	煙花爆竹 (NA)	(六級)	咬牙切齒 (1,098)	(六級)	一帆風順 (4,785)	(六級)
一舉兩得 (1,751)	(六級)	一目了然 (7,109)	(六級)	一如既往 (19,052)	(六級)	一絲不苟 (3,267)	(六級)
優勝劣汰 (NA)	(六級)	有條不紊 (10,465)	(六級)	與日俱增 (3,332)	(六級)	再接再厲 (7,698)	(六級)
斬釘截鐵 (684)	(六級)	爭先恐後 (4,548)	(六級)	知足常樂 (NA)	(六級)	眾所周知 (17,990)	(六級)
自力更生 (5,551)	(六級)	總而言之 (4,416)	(六級)				

## 引用文獻

- Bauer, L., and I.S.P. Nation. 1993. Word families. *International Journal of Lexicography* 6.3:1-27.
- Coxhead, Averil. 2000. A new academic word list. *TESOL Quarterly* 34: 213-238.  
Averil Coxhead's website: <http://language.massey.ac.nz/staff/awl/index.shtml>
- Davies, M., and D. Gardner. 2010. *A frequency dictionary of contemporary American English*. New York: Routledge.
- Davies, M. 2012. Corpus of Contemporary American English [Online corpus].  
Retrieved from <http://corpus.byu.edu/coca/>
- Dupuy, H. J. 1974. *The rationale, development, and standardization of a basic word vocabulary test*. US Government Printing Office.
- Engels, L.K. 1968. The fallacy of word counts. *IRAL* 6:213-231.
- Goulden, R., P. Nation, and J. Read. 1990. How large can a receptive vocabulary be? *Applied Linguistics* 11.4:341-363.
- Jamieson, P. 1976. The acquisition of English as a second language by young Tokelau children living in New Zealand.
- Kilgariff, A. 1997. Putting Frequencies in the dictionary. *International Journal of Lexicography* 10.2:135-155.
- Leech, G. 1997. Teaching and Language Corpora. A convergence.(title??)  
*Teaching and Language Corpora*, eds. by A. Wichmann, S. Fligelstone, A. McEnery, 1-23. London: Longman.
- Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge University Press.
- Nation, P., and R. Waring. 1997. Vocabulary size, text coverage and wordlists. *Vocabulary:Description, Acquisition and Pedagogy*, eds. by N. Schmitt, & M. McCarthy, 6-19. Cambridge: Cambridge University Press.
- Sinclair, J. 1996. Beginning the study of lexis. *In Memory of J. R. Firth*, eds. by C. E. Bazell, J. C. Catford, M.A.K. Halliday, & R. H. Robins, 410-430. London: Longman.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

- 尤芳。2012。〈語料庫與外語教學〉，《科技信息》，第2期，195。[You, Fang. 2012. Corpra and Foreign Language Teaching. *Science* 2:195.]
- 王建新。2005。《計算機語料庫的建設與應用》。北京：清華大學出版社。  
[Wang, Jianxin. 2005. *The Development and the Application of Computerised Corpora*. Beijing: Qinghua University Press.]
- 王麗亞。2011。〈中學英語新彙教學藝術〉，《教苑薈萃》，第8期，37。  
[Wang, Li Ya. 2011. High school English vocabulary teaching. *Education Collection* 8:37.]
- 李如龍、吳茗。2005。〈略論對外漢語詞彙教學的兩個原則〉，《語言教學與研究》，第2期，41-47。[Li, Ju Long, and Ming Wu. 2005. On the two principles of vocabulary teaching in TCSL. *Journal of Language Teaching and Research* 2:41-47.]
- 李英。1997。〈關於“漢語水平詞彙與漢字等級大綱”的幾個問題〉，《中山大學學報論叢》，第四期，1-3。[Li, Ying. 1997. Some problems about the outline of graded vocabulary for HSK. *Sun Yatsen University Forum* 4:1-3.]
- 俞靜。2011。〈論對外漢語教學中的詞彙教學〉，《文學教育》，下輯第8期，64-66。[Yu, Jing. 2011. On lexical teaching in HCFL. *Literature Education B*. 8:64-66.]
- 許巧齡、吳慧萍。2010。《華語文評量規劃與測驗設計》，載僑委會華文種子師資班線上課程，[http://www.huayuworld.org/wp-course/97\\_huayu\\_seed\\_plan\\_digital\\_course/Course\\_9\\_wu/download/Manual\\_Test.pdf/](http://www.huayuworld.org/wp-course/97_huayu_seed_plan_digital_course/Course_9_wu/download/Manual_Test.pdf/)。[Xu, Qiao Ling & Wu, Hui Ping. 2010. Designing Chinese Language Testing and Assessment, retrived from HuayuWorld website, [http://www.huayuworld.org/wp-course/97\\_huayu\\_seed\\_plan\\_digital\\_course/Course\\_9\\_wu/download/Manual\\_Test.pdf/](http://www.huayuworld.org/wp-course/97_huayu_seed_plan_digital_course/Course_9_wu/download/Manual_Test.pdf/).]
- 張亦政、張文芝。1995。〈主觀與客觀：教學詞表選詞論旨〉，《外語界》，第57期，37-43。[Zhag, Yi Zheng & Zhang, Wen Zhi. 1995. The word selection for vocabulary teaching: subjectivity and objectivity. *Foreign Language World* 57:37-43.]

- 張能甫。1999。〈漢語基本詞彙研究的回顧與展望〉，《四川師範大學學報（社會科學版）》。第2期，88-93。[Zhang, Neng Fu. 1999. Chinese basic lexicon research: retrospect and prospect. *Journal of Sichuan Normal University Social Sciences Edition* 2:88-93.]
- 張莉萍。2007。〈華語文能力測驗發展現況〉，《外語能力測驗之動向與展望國際研討會論文集》，185-195。[Zhang, Li Ping. 2007. The development and status of TOP—Hanyu. *Proceedings of the International Conference of Foreign Language Proficiency Tests Trends*, 185-195.]
- 張莉萍、陳鳳儀。2005。〈華語詞彙分級初探〉，《第六屆漢語辭彙語義學研討會會後論文集》，250-260。新加坡：新加坡中文與東方語言信息處理學會出版。[Zhang, Li Ping, and Feng Yi Chen. 2005. The exploration on grade classification of Chinese lexicons. *Proceeding of 6th Chinese Lexical Semantics Workshop*, 250-260.]
- 張晉軍、李佩澤、李亞男、解妮妮、黃蕾。2012。〈對新漢語水準考試（HSK）的新思考〉，《中國考試》，2012年第2期，50-53。[Zhang, Jin Jun, Pei Ze Li, Ya Nan Li et al. 2012. New thinking on the new HSK. *China Examinations* 2012.2:50-53.]
- 張晉軍、解妮妮、王世華、李亞男、張鐵英。2010。〈新漢語水平考試（HSK）研製報告〉，《中國考試》，2010年第9期，38-43。[Zhang, Jin Jun, Ni Ni Xie, Shi Hua Wang, et al. 2010. The report of researching and producing the new HSK. *China Examinations* 2010.9:38-43.]
- 賈穎。2001。〈字本位與對外漢語教學〉，《漢語學習：學術版》，第4期，78-80。[Jia, Y. 2001. Character-based theory and TCSOL. *Chinese Language Learning Academic Edition* 4:78-80.]
- 劉成博。2012。〈語料庫語言學理論指導下的詞匯教學〉，《江蘇教育》，第4期，27-28。[Liu, C. B. 2012. Lexicon approach from corpus linguistics theory. *Jiangsu Education* 4:27-28.]
- 顧曰國。1998。〈語料庫與語言研究——兼編者的話〉，《當代語言學（試刊）》，第1期，1-3。[Gu, Y. G. 1998. Corpus and language research—from the editor. *Contemporary Linguistics (prototype edition)* 1:1-3.]

- 國家華語測驗推動工作委員會。2013。《華語八千詞表說明》，載國家華語測驗推動工作委員會網站，[http://www.sc-top.org.tw/download/8000\\_description.pdf](http://www.sc-top.org.tw/download/8000_description.pdf)。[Steering Committee for the Test Of Proficiency—Hanyu. 2012. 8000 words Introduction—Hanyu. Retrieved from [http://www.sc-top.org.tw/download/8000\\_description.pdf](http://www.sc-top.org.tw/download/8000_description.pdf)]
- 胡明揚。1997。〈對外漢語教學中語彙教學的若干問題〉，《語言文字應用》，第1期，12-17。[Hu, M. Y. 1997. Some questions about lexical approach in TCSOL. *Applied Linguistics* 1:12-17.]

[審查：2013.3.22 修改：2013.6.26 接受：2013.7.3]

楊惠媚

Hui-Mei YANG

106 臺北市大安區和平東路一段 162 號

162, Hoping E. Rd., Sec. 1, Taipei 106, Taiwan

huimei.yang2009@gmail.com

陳浩然

Hao-Jan Howard CHEN

106 臺北市大安區和平東路 1 段 162 號

162, Hoping E. Rd., Sec. 1, Taipei 106, Taiwan

hjchen@ntnu.edu.tw

潘依婷

I-Ting PAN

106 臺北市大安區和平東路 1 段 162 號

162, Hoping E. Rd., Sec. 1, Taipei 106, Taiwan

leopardpan@gmail.com

## **Comparison of TOCFL and HSK Wordlists and Suggestions on Word-selection**

**Hui-Mei YANG      Hao-Jan Howard CHEN      I-Ting PAN**  
**Department of English, National Taiwan Normal University**

### **Abstract**

General service list is at the core of language learning, and has received the attention of language learners, teachers and even textbook writers. In the field of Chinese learning, the TOCFL wordlist and HSK wordlist are the most well-known wordlists. However, few studies tried to compare these two important wordlists. In this study, we compared the TOCFL and HSK wordlists and found (1) There are 3,700 words shared by both the TOCFL and HSK wordlists while there are 3,619 words in TOCFL different from HSK. In contrast, HSK has only 1,296 words that differ from TOCFL. (2) The TOCFL and HSK wordlists are laid out differently. (3) The HSK contains more than 4.3 times the number of four-character words than TOCFL does. (4) The TOCFL wordlist contains a larger number of “er” words than the HSK does. Furthermore, we noticed that some low-frequency words were included in these two wordlists. With the help of two large corpora based on web pages of Mainland China and Taiwan, two top 10,000 high-frequency wordlists were compiled. These two wordlists were then compared with TOCFL and HSK wordlists separately. Based on the results of the comparison, we found that 4,770 high-frequency words from the TaiwanWac corpus were not included in the TOCFL wordlist. In addition, 6,357 high-frequency words from the Tenten corpus were not included in the HSK wordlist. To further improve the quality and coverage of TOCFL and HSK wordlists, it is suggested that some of these high-frequency words should be included in the two wordlists. The findings and recommendations of this study can be a

華語文教學研究

useful reference for Chinese learners, teachers, textbook writers, and language test agencies.

**Keywords:** Mandarin learning, Chinese wordlist, core vocabulary, web corpus, high-frequency words