

# 分析華語口語語料庫高頻詞之特點並對 TOCFL 詞表提出建議\*

楊惠媚 陳浩然<sup>1</sup> 潘依婷

國立臺灣師範大學國語中心暨英語學系

## 摘要

詞表主要功能在於列出學習者應學習的詞彙，能作為第二語言詞彙教學之參考。臺灣著名的華語詞表—TOCFL 詞表為測驗及學習者常用詞表。但因 TOCFL 詞表選詞以參考書面語語料庫詞頻為主，口語語料所佔的比例較低。然而書面語與口語詞彙為不同語體詞彙，是故語言教學不應僅限於書面語，選詞亦應涵蓋口語詞彙。為探究是否部分口語高頻詞可建議納入 TOCFL 詞表，以使詞表書面語語料及口語語料的比例較為平均，本研究蒐集具有對話性質的華語連續劇及電影對白作為母語者口語語料，藉詞頻排列得出口語高頻詞彙，並與 TOCFL 詞表進行對比。對比後發現有 713 筆口語語料高頻詞並未收錄於 TOCFL 詞表中。本研究提出詞頻最高且最具口語關鍵詞特點之 238 筆詞彙供 TOCFL 詞表增修參考詞彙，亦依據自前人文獻整理出的 6 項口語詞彙特點針對 713 筆口語高頻詞進行分類並歸納出以下特點：(1) 口語詞彙多以詞塊合成詞的形式呈現；(2) 口語詞彙中包含較多多音節熟語；(3) 以「不」及「沒」的組合型式詞塊數量偏多。本研究最後將依詞頻及口語詞彙特色提出建議可納入 TOCFL 詞表之詞彙，期望提高詞表中口語詞彙以增加 TOCFL 詞表的豐富性及多元性，並提供華語教學更貼近口語交際使用之詞彙參考。

**關鍵詞：**華語學習 華語詞表 口語語料庫 高頻詞

---

\* 本研究成果得到「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。

## 1. 緒論

在第二語言詞彙教學中，詞表（wordlist）多用以作為教學大綱擬定時的參考，亦可顯示出各級學習者應學習的詞彙內容。過去詞表中該收錄哪些詞，收錄多少詞，多是由教學者或研究者主觀認定（Nation 2004）。然而，為能提供學習者更貼近母語者所用的詞彙內容，眾多專家學者建議，第二語言的詞彙教學可以詞頻作為詞彙教學參考（Nation 2001; 張莉萍、陳鳳儀 2005）。例如：英文教學當中著名的詞表 GSL (General Service List) 是 1953 年 Michael West 以英文寫作文本語料庫作為依據，取前 2000 筆英文常用詞所編列出的詞表，開啓了後續以詞頻作為詞表選詞的先例。近年來，隨著語料庫建置的發展，以書面語及口語語料的詞頻作為選詞依據之詞表也逐漸增加。例如 Kilgarriff (1997) 提出的 BNC (British National Corpus) 常用詞表；Mark Davies 與 Dee Gardner (2013) 的以詞頻排列的 COCA (Corpus of Contemporary American English) 高頻詞表等，皆是自書面語與口語結合之語料庫中依據詞頻高低排列出的詞表，BNC 與 COCA 語料中所收錄的口語語料都超過 1 百萬字，其中 COCA 約有 9 百萬字。

華語教學中著名的語言能力測驗詞表—TOCFL 詞表，亦參考詞頻作為選詞依據之一。儘管如此，TOCFL 詞表所採用的中央研究院平衡語料庫亦是以 500 萬字的書面語語料為主之詞頻作為選詞依據<sup>2</sup>，其中口語語料僅佔 10%<sup>3</sup>，約 50 萬字。若與 BNC 或 COCA 等著名英文高頻詞表相比，TOCFL 詞表內所涵蓋之口語詞彙比例偏低。

然而學習者實際在使用語言時，除了書面語詞彙，口語詞彙亦不可或缺，特別是在日常生活的口頭交際場合（Saussure 1959; 顧安達、朱志平 2005; 尹惠貞 2006; 張金蘭 2012）。李如龍（2007）更進一步提出，書面語和口語的表達各有不同的要求，兩者都是社會生活中不可或缺的。現今在書面刊物或書信上可見口語詞彙，或是在口語對話方面也可見書面語詞彙，可歸因於口語產生在文字之前，書面語是在口語的基礎上產生和發展（王希杰 1991; 葉蜚聲、徐通鏘 1993; 高名凱、石安石 2002），兩者間的關係密切且通用詞彙多（楊俊萱 1984; 宋婧婧 2013）。但書面語與口語詞彙畢竟是兩種不同語體（周祖謨 1959; 楊俊萱 1984; Biber & Finegan 1991），楊俊萱（1984）表示，口語詞彙具有通俗、自然、親暱的口語色彩，而書面語詞彙則較為正式，在不同的語言環境下需有所分別（曹煒 2003; 謝智香 2011；

<sup>2</sup> 引自張莉萍、陳鳳儀 (2005) 〈華語詞彙分級初探〉

<sup>3</sup> 中研院平衡語料庫中的口語語料來源包含：演講稿、劇本、會話、會議記錄（引自「現代漢語語音語料庫研究」網頁，[http://mmc.sinica.edu.tw/resources\\_c\\_02.html](http://mmc.sinica.edu.tw/resources_c_02.html)）

張金蘭 2012；宋婧婧 2013）。若是使用不當，反而容易造成反效果；特別是對第二語言學習者而言，能夠掌握得體的使用語言尤其重要（楊俊萱 1984；張金蘭 2012）。因此，第二語言教學詞表若能增加口語詞彙便能提升詞表內涵之廣度，更能貼近學習者口頭交際上的使用。

徐素萍（2012）表示透過蒐集日常自然口語錄音，可讓學習者知道口語詞彙與書面詞彙的差異，讓他們更加了解固定短語、口頭慣用語的自然運用。尹惠貞（2006）亦說明口語詞彙對於詞彙教學的重要性已在語言教學中受到重視及肯定。眾多學者們（Nation & Waring 1997；楊惠媚、陳浩然、潘依婷 2014）也同意詞表內容是影響詞彙教學的重要因素之一，因此詞表內所收錄的詞彙內容也是語言教師及相關研究人員關切的課題。

以 TOCFL 詞表為例，雖然 TOCFL 詞表為少數在選詞上涵蓋口語語料作為參考依據的華語詞表。然而該詞表是參考中央研究院平衡語料庫 3.0 版<sup>4</sup>作為選詞依據，此版本之語料庫內容為 500 萬字，口語語料比例不及 10%。張莉萍及陳鳳儀（2005）表示，詞表應以頻率為選詞依據，並避免語料書面語料偏多的態勢，然 TOCFL 詞表所採用的中研院平衡語料庫口語語料不到 50 萬字，這可能會影響 TOCFL 詞表內所涵蓋的口語詞彙及詞彙的豐富度。

要取得常用的口語高頻詞彙，勢必需透過口語語料庫。現今臺灣華語口語語料庫，除中央研究院「現代漢語語音語料庫」（40 萬 5 千多字）外，尚有政治大學「政大漢語口語語料庫」（15 萬 5 千多字）<sup>5</sup>。上述兩語料庫的口語語料量皆小於中研院平衡語料庫的近 50 萬字的口語語料量。若要針對 TOCFL 詞表提出口語高頻詞建議，兩者皆不太適用。本研究在時間與人力限制之各方面考量下，參考 L.K. Engels 和 Dirk Geens 於 1965 年至 1972 年以英語戲劇對白建置 The Leuven Drama Corpus

（Halliday et al., 1990）的方法，將具有口語對話特色的華語電影及電視劇字幕作為語料來源，收集總計超過 150 萬字的口語語料，擬藉口語語料詞頻排列出華語母語者口語常用詞彙。將所得之常用口語詞彙表比對 TOCFL 詞表。按兩者比對所得結果，提出 TOCFL 詞表中未收錄，但可供詞表編纂人員參考的口語高頻詞彙。期望這些口語語料能為口語詞彙教學提供更貼近真實語境的材料，以提升華語詞彙教學

<sup>4</sup> 中央研究院平衡語料庫 3.0 版，1997 年完成。當中取劇本、會話、演講、會議記錄之內容作為口語語料。在 500 萬字的語料量中佔有 76.5 萬字，佔語料庫的 9.7%。引自 <http://app.sinica.edu.tw/kiwi/mkiwi/98-04.pdf>

<sup>5</sup> 有關「現代漢語語音語料庫」及「政大漢語口語語料庫」建置方式及內容將於下節文獻探討的 2.3 提出。

之成效。

## 2. 文獻探討

本節主要就以下三大要點探討相關文獻：（1）口語與書面語詞彙的差異；（2）口語詞彙的特色及重要性；（3）口語語料庫的建置與運用。

### 2.1 口語與書面語詞彙的差異

口語與書面語都是語言存在的形式，兩者間的關係密切（楊俊萱 1984），然而兩者為不同語體，不同的語言環境下使用上有所差異（周祖謨 1959；楊俊萱 1984；Biber & Finegan 1991）。最早提出口語詞和書面語詞的概念的是周祖謨（1959），認為詞彙在語體風格上的區別在於詞的適用範圍需透過場合的不同來判定。Halliday（1989）提到口語是動態的（dynamic），所呈現的方式是向前推進的過程（processes）；書面語則是靜態的（static），其文章本身就是一項物件（object）。然而，語言中使用的詞彙是非常紛繁的，有些詞口頭上常用，書面上比較少用；有些書面上常用，口頭上比較少用；還有很大一部分的詞，口頭上和書面上都常用。口頭上常用的詞和書面上常用的詞，並沒有嚴格的界限，而且二者有接近一致的趨勢。但二者在語體風格上的區別多少還是存在的（周祖謨 1959）。曹煒（2003）在《現代漢語口語詞和書面語詞的差異初探》一文中也提出口語詞和書面語詞是詞彙著眼於不同的適用語域所進行的一種分類：口語詞是在日常口頭交際中所使用的詞，往往具有通俗、俏皮的風格；書面語詞是在正式的交際場合所使用的詞，往往具有典雅、莊重的色彩（謝智香 2011）。此外，李如龍（2007）提出口語是用聲音來實現的，通常包含聽、說雙方和特定的及不斷變化的語境，一來一往往往往不待思索的交流，因而比較粗糙、蕪雜，但是可以隨時修改、更正；反觀書面語是用文字記錄的，可以超乎時空，在預設的、靜止的語境中慢修斯理地推敲和提煉，因二者於交際方式的不同也造成了詞彙上的差異。然而造成口語及書面語之間的差異因素甚廣，亦需追索至古代語音、語素及漢字造詞系統等，本研究因篇幅關係不再此方面多加以著墨。

如上文所述，漢語書面語及口語詞彙中確實存在著差異性，且使用上需切合語言環境，若該用口語而用了書面語不免讓人感覺咬文嚼字，而若該用書面語卻用了口語則顯得不莊重（楊俊萱 1984）；特別是對二語學習者而言，若能分辨兩者的區別，可提升語言的理解力。顧安達與朱志平（2005）提到以往二語教學多以書面語及語法為主旨，但近年來教學思想有所轉變，教師開始著眼於學生口語交際能力的培養。雖然如此，對於口語詞彙的相關研究仍不及書面語詞彙。有鑑於此，許多語

言學家開始針對口語及書面語詞彙的差異進行研究。如：Biber (1988) 藉語料庫分析了書面語與口語的語體差別，提出多特徵、多維度的研究方法，這也影響了華語語言學界對詞彙研究的方式，且更廣泛地將語料庫應用於詞彙研究中，書面語詞彙與口語詞彙間的異同也逐漸分別出來（宋婧婧 2013），且越來越多專家學者投入區分口語詞彙及書面語詞彙的相關研究，例如：吳麗君（2004）〈口語詞彙與書面語詞彙教學研究〉從語體的角度探討了對外漢語的口語詞彙教學與書面語詞彙教學，並討論分析各個不同的教學階段口語詞彙及書面語詞彙教學的重點。

楊俊萱（1984）的〈口語和書面語〉一文中提到口語詞彙具通俗、自然、親暱的口語色彩，而書面語詞彙則典雅、莊嚴、帶有文彩，例如：正好（口）v.s. 恰巧（書）；不吭聲兒（口）v.s. 沉默（書）等。口語也多常用俚語、諺語、歇後語、慣用語等；書面語則多術語、虛詞、文言文和成語等。例如：口語—碰釘子、踢皮球等；書面語—彙報、錦上添花等。口語中也多帶前後綴的詞、重疊詞、象聲詞等，如：老王、小孩兒、通紅通紅的、唧唧喳喳等。

曹煒（2003）〈現代漢語口語詞和書面語詞的差異初探〉以 1983 年版的《現代漢語詞典》所標注的 802 個口語詞和 821 個書面語詞作為研究內容，對口語及書面語詞的基本特徵作一個全方位的考察。他認為現代漢語中的口語詞語和書面語詞之間的差異極其明顯，可從（1）音節結構來看：口語中雙音節詞占絕對優勢，而在書面語詞中，單音節詞的數量比雙音節詞的數量少。（2）詞的內部構成情況來看：口語詞中合成詞是單純詞的 6 倍強，而書面語詞中單純詞的數量接近合成詞。口語詞以偏正型居多書面語詞則以聯合型居多。（3）詞義架構上來看：口語詞同書面語詞均以單義詞為主，多義詞很少，但多義口語詞中雙音詞占主體，而多義書面語詞中單音詞占主體。

對於中文書面語及口語詞彙間的差異的研究還有李世文與陳秋梅（1993）發表的〈中文口語與書寫語的比較研究〉、李如龍（2007）的〈關注漢語口語詞彙與書面語詞彙的研究〉及宋婧婧（2013）〈漢語口語與書面語詞彙使用對比分析〉等。上述研究中皆提出口語詞彙與書面語詞彙在語音、語義、語法屬性及語用功能上皆有所差異。下文將彙整前人對口語詞彙的相關研究進一步探討口語詞彙的特色。

## 2.2 口語詞彙的特色及重要性

口語是日常生活用語，是社會交際的重要工具，也是最早為人類普遍使用的語言形式，人們把用於口頭交際的詞彙稱為口語詞彙（常敬宇 1986；王芳智主編 1990）。

尹惠貞（2006）表示口語語體適用於交際，而交際一般是非正式的。口語交際不僅僅是語言的交際，還有很多非語言手段配合語言來進行交際，例如手勢、動作、面部表情等在口語交際中都會發揮很大的作用。且口語還能借助語音表達出很多東西，如重音、語調、句詞、語速等都是口語中每個句子不可缺的。例如：語氣除了狹義的陳述、疑問、祈使、感嘆四種以外，口語還能表達多種廣義的預期，如強調、委婉、決斷、詞意、溫柔等。上述提及之語言交際手段也影響了口語詞彙的特色。然而大多學者或研究人員對對於口語詞彙的相關研究，多以口語與書面語詞彙作對比分析，對口語詞彙特點直接進行相關研究之文獻較少（謝智香 2011；宋婧婧 2013）。不僅如此，在僅有的文獻中對於口語詞彙的特點無一統整的歸納。

因此，本研究選取 9 篇相關文獻，有 5 篇討論書面語與口語差異（楊俊萱 1984；李世文、陳秋梅 1993；曹煒 2003；李如龍 2007；宋婧婧 2013）。有 4 篇討論口語詞彙相關研究（常敬宇 1986；尹惠貞 2006；謝智香 2011；張文賢、路雲、李曉琪 2012）。探討詞彙本需分析內容結構並從語音、語義、語法、語用等面向一一進行探討，然而因篇幅限制及研究方向非探討詞彙組成過程，本研究將不針對華語口語詞彙形成的語音、語義、語法、語用等面向一一進行討論，僅自前人文獻中所提出的華語口語詞彙特點彙整成表。下表 1 羅列自前人文獻彙整出的華語口語詞彙 6 項特點。

表 1 口語詞彙 6 項特點

| 口語詞彙特點        | 前人提出論點及定義   | 範例               | 引用出處                                  |
|---------------|---|------------------|---------------------------------------|
| 1. 較多雙聲及疊韻詞   | 口語詞彙以雙音節為主，其次是多音節詞，單音節詞最少。歸因於漢語歷史發展過程中，原以單音節詞為基礎，不斷地組成的複音節詞，從單音節詞演變為複音節詞，且亦可能是受白話語影響，因此現代漢語中以雙音節詞佔優勢。此外，古今的雙聲、疊韻等連綿詞、切腳詞、疊音詞、合音詞，應多是由口語衍生出來的。 | 盼著、疙瘩、高高興興地、綠油的等 | 楊 俊 萱 1984；常敬宇 1986；尹惠貞 2006；李如龍 2007 |
| 2. 較多以詞綴衍生出的詞 | 口語中常用前綴詞和後綴詞，是在詞頭或詞尾加上詞綴而成的詞，常見前綴有  | 老公、小販、第          | 楊 俊 萱 1984；尹惠貞                        |

| 口語詞彙特點                     | 前人提出論點及定義  | 範例                             | 引用出處   |
|----------------------------|--|--------------------------------|--|
|                            | 「老」、「小」、「初」、「第」等，後綴有「子」、「兒」、「頭」、「巴」等   | 一、初五、孩子、臉蛋兒、木頭等                | 2006；李如龍<br>2007；謝智香<br>2011；張文賢、路雲、李曉琪 2012 |
| 3. 較多語氣詞/嘆詞/擬聲詞            | 口語詞彙中常使用較多的語氣詞、嘆詞及擬聲詞。   | 坐 下<br>吧、吧<br>嗒、哪哪<br>喳 喳<br>等 | 楊 俊 萱<br>1984；常敬宇<br>1986；宋婧婧<br>2013        |
| 4. 較多詞塊 <sup>6</sup> 合成詞型式 | 口語詞彙當中合成詞較單純詞多，合成詞如：白搭、北邊等。而兩個詞組合在一起又有確定意義，或具有一定的交際功能，這類的詞口語化程度亦顯著，例如：挺好、打電話、各地、是不是、你好、不能不、夠了、才能、只能等。若進一步觀察以詞塊方式呈現的詞，可發現以「不」、「這」、「那」開頭的詞亦佔多數，比方：不少、不大、這樣子、那時候、那是等詞彙。 | 夠了、什麼的、說真的、年前、這兩天、不夠、這時候等      | 張文賢、路雲、李曉琪<br>2012                           |
| 5. 較多多音節熟語 <sup>7</sup>    | 口語在交際中是以聲音呈現，聽與說雙方多以詞義較單一的雙音節詞語作為交   | 碰 釘<br>子、巴不                    | 楊 俊 萱<br>1984；常敬宇                            |

<sup>6</sup> 詞塊，亦有人稱為語塊。張文賢、路雲和李曉琪（2012）提到：『如果兩個詞常常在一起出現，並且這兩個詞組合一起又意義確定，或者具有一定的交際功能，就將之作為語塊出現在詞表中。這樣的詞有「挺好、打電話、打球、那天、十幾、各種．．．慢慢來、是不是、你好、不能不、說真的．．．這兩天、男女平等、好吧、夠了．．．什麼的、也就是說」等。兩個單音節副詞總是連用，如「才能、只能、再不」，也作為一個詞處理。以「不」、「這」、「那」開頭的詞，如「不少、不夠、不大、這樣子、這時候、那時候、那是」等也作為詞收到詞表。從這些詞中能較明顯地反映出原始語料高頻詞的特點：意義明確、結構固定、凝固性強、口語化明顯。』

<sup>7</sup> 在語法分類下熟詞亦屬於詞塊類型之一，然而因口語交際上以熟語型式的詞塊頻率較多，因此將這一類型詞彙自詞塊中提出來自成一類。

| 口語詞彙特點      | 前人提出論點及定義  | 範例              | 引用出處                         |
|-------------|--|-----------------|------------------------------|
|             | 際主要用詞口語中也多使用俗語、諺語、歇後語、慣用語和口頭上常用的成語，上述詞彙亦稱為熟語。      | 得、姑奶奶、過河拆橋、靠邊站等 | 1986；李如龍<br>2007；謝智香<br>2011 |
| 6.較多方言詞/外來語 | 方言詞具濃厚的口語色彩，如：姥姥、把勢，應視作為口語詞。而口語詞彙中亦受外來語影響而產生了新的景象。 | 技術、姥姥、芒果、尷尬等    | 曹煒 2003；<br>李如龍 2007         |

口語詞彙為日常口頭交際所使用的語言，若學習者能清楚掌握口語詞彙，將有助於他們在語言使用上的掌握能力，且若能更進一步將口語及書面語詞彙分類出來，便可更針對性列出兩語體之常用詞彙，有助於提升華語詞彙教學成效。

自 Biber (1988) 藉助語料庫分析了書面語與口語的語體差異開始，語料庫便更廣泛地運用於詞彙的研究；特別是書面語與口語詞彙間異同。近年來在華語語言學及教學界也逐漸重視兩語體間的差異。然而，因書面語較具規範性且語料範本取得較容易，對於書面語詞彙的相關研究較口語詞彙的相關研究多（李如龍 2007）。在華語詞彙教學及研究上也多以書面語作為主要對象。然而，宋婧婧（2013）表示口語及書面語通用的高頻詞是語言教學的核心，學習者若能掌握語體色彩鮮明的口語及書面語詞彙，將有助於提升學習者語言能力。

而為能客觀地得知母語者日常口語使用情形，語言教學學者與研究人員紛紛投入人力建置應用口語語料庫的建置工作。口語語料庫不僅可用於探討詞彙使用情形，亦可運用於教學上，對於課堂口語教學或教材編寫可提供符合符合真實語境的語言材料，使語言教學更貼近日常生活。下文將進一步探討口語語料庫的建置及運用。

### 2.3 口語語料庫的建置與運用

口語語料庫是集結口語內容，轉寫語音檔建置而成，得以反映出日常口語的使用情形，對於課堂口語教學或教材編寫可提供符合真實語境的語言材料（陳露、韋漢 2005；車豔秋、楊虹、曹明 2013）。著名的劍橋大學 Touch Stone 系列教材便是以 CANCODE（Cambridge and Nottingham Corpus of Discourse in English）口語語料庫為基礎編輯而成（McCarthy 2004）。陳露、韋漢（2005）指出口語語料庫具有真



實性，且涵蓋了交際時的多元內容與情感層面。徐素萍（2012）亦表示口語語料庫可以彌補教材對發音部位、發音方法純文字描述的不足，透過蒐集日常自然口語錄音可以讓學習者知道口語詞彙與書面詞彙的差異，讓他們更加瞭解固定短語、口頭慣用語的自然運用。

最早出現的口語語料是 1959 年由英國倫敦大學 Randolph Quirk 教授所主持建置的 Survey of English Usage (SEU) 語料庫。其語料收自 2000 個小時的談話及廣播錄音等口語素材，經由轉寫而成的書面語料。該批書面語料爾後又經由瑞典隆德大學 (Lund University) J. Svartvik 教授電腦化並加入新的口語語料後，於 1975 年建置成 London-Lund Corpus (LLC)，該語料量約為 45 萬字，至今仍是英語口語研究的重要資料來源 (Greenbaum & Svartvik 1990；曹合建主編 2008)。

Tim Benbow 於 1991 至 1994 年期間，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) 英語為主並集合各領域語料，涵蓋了 1 億個字詞。其口語語料取材自 124 名分別來自各個地區，不同社經背景，且年齡層也分佈廣闊的語料提供者，其中最少年齡為 15 歲。口語語料內容分成四個類別：教育、商業、公共部門/機構、休閒等，在此項下再細分為兩小類：40% 的獨白和 60% 的對話。目前口語語料量約為 1 千萬字 (Kilgariff 1997)。

另一知名口語語料庫 The Cambridge and Nottingham Corpus of Discourse in English (CANCODE) 是由英國劍橋大學和諾丁漢大學共同建置，該語料庫包含書面語及口語語料庫，其中口語語料來源為生活會話、寒暄、尋求資訊和討會等，約有 500 萬字。隨後，McCarthy、Carter 等人以 CANCODE 口語語料庫為基礎發展了多套教材及文法書，如 English Vocabulary/Idioms in Use (McCarthy & O'Dell 2003)、Exploring Grammar in Context (Carter, McCarthy & Hughes 2002)，期望藉由口語語料資源補足口語教學及教材之真實性和提供更生活化的材料 (McCarthy & Carter 2001；Carter 2004)。

由 Mark Davies 於 2008 年著手建置的美國當代英語語料庫<sup>8</sup> (The Corpus of Contemporary American English, COCA) 收集了自 1990 至 2012 年間的美式英語語料。目前語料超過 4 億 5 千萬字。其語料內容包含了口語用語、小說、報章雜誌及學術文章。其中口語語料取自電視及廣播節目內容，例如：NPR 的線上新聞節目 *All Things Considered*、PBS 電臺的 *Newshour* 節目對話內容、ABC 新聞網的 *Good Morning America*、Fox 新聞臺的 *Hannity and Colmes* 等節目內容，其口語語料超過 8 千萬字。

---

<sup>8</sup> 引自“Corpus of Contemporary American English” <http://corpus.byu.edu/coca/>

除了以收集各式口語語料方式集結成口語語料庫外，另一種方式是以戲劇對白作為口語語料庫，例如：The Leuven Drama Corpus。該語料庫是由 Catholic University of Leuven 的 L.K. Engels 和 Dirk Geens 蒐集 1965 至 1972 年間，61 齣當代英語戲劇對白建置成英語口語語料庫，共約 1 百萬字的語料（Halliday et al. 1990）。

相同地以戲劇對白作為口語語料來源的有，康乃爾大學於 2012 年公開的 Cornell Movie-Dialogs Corpus，是自 617 部電影中抽取對白字幕所建置而成的語料庫，其中包含了約 22 萬組對話式交談記錄（conversational exchange），共計 304,713 筆話語單位（utterance）（Danescu-Niculescu-Mizil & Lee 2011）。

在華語口語語料庫方面，發展雖不及英語語料庫，但近年來亦漸漸廣受到研究人員重視。由北京語言大學建置的「當代北京口語語料庫」<sup>9</sup>是透過實際採訪調查，記錄了具有不同社會特徵的北京地區居民在自然狀態下的日常口語。語料大多數內容多為受試者的個人即興敘述，交談之話題共分為 6 大類、28 個話題，皆屬日常生活中的家常話題或當時熱門議題，如：居住條件、社會治安、學習就業、婚喪嫁娶等等。談話內容經錄音及轉寫成文字後，共累計約 186 萬字之口語語料。

此外，臺灣政治大學自 2006 年起著手建置「政大漢語口語語料庫」<sup>10</sup>，分別搜集華語、閩南語及客語口語語料。該語料庫透過錄製參與人員的對談內容或觀感陳述，再收集到的音檔轉寫為文字檔並建構語料庫和網路檢索平臺，使語料內容得以廣泛運用，為教學及研究提供真實的口語語料。其中，華語口語語料已累積 27 篇不同年齡層的日常對話內容，總計約 15 萬 5 千多字。

中央研究院語言研究所於 2005 年執行「新世紀語料庫——多媒體的呈現與珍藏」計畫，建置了「現代漢語語音語料庫」<sup>11</sup>，其語料庫是由三個子語料庫組成，分別是「現代漢語主題對話語音語料庫」、「現代漢語連續口語對話語音語料庫」及「現代漢語地圖導引口語語音語料庫」，目前已完成音頻語料的轉寫和標註。共收集了 85 個對話內容，170 名發音人，約 42 個小時的對話內容，其語料大小約 40 萬 5 千個詞次。研究團隊並自該語料庫中按詞頻編纂「漢語口語語料庫詞頻表」<sup>12</sup>。

透過口語語料庫不僅能觀察該語言使用者真實的語言內容，也能藉由分析並歸納母語者口語語用特徵及語言的搭配現象，提供教師在教學上的應用。口語語料庫亦能夠補足書面語語料庫不易觀察出口語語用及詞彙上的不足。且口語語料庫在研

<sup>9</sup> 引自「當代北京口語語料庫」，[http://www.blcu.edu.cn/yys/6\\_beijing/yuliao/shuoming.htm](http://www.blcu.edu.cn/yys/6_beijing/yuliao/shuoming.htm)

<sup>10</sup> 引自「政大漢語口語語料庫」，<http://140.119.172.200/chinese/cmain.php>

<sup>11</sup> 引自「現代漢語語音語料庫」，[http://mmc.sinica.edu.tw/corpus\\_c\\_01.html](http://mmc.sinica.edu.tw/corpus_c_01.html)

<sup>12</sup> 引自「漢語口語語料庫詞頻表」，[http://mmc.sinica.edu.tw/resources\\_c\\_02.html](http://mmc.sinica.edu.tw/resources_c_02.html)

究上所帶出來的成果可帶給語言教學上更廣泛的幫助。例如，徐素萍（2012）表示口語語料庫可彌補現有現代漢語教材的不足。在詞彙、語法等方面，可對比口語和書面語的差異，了解固定短語、口頭慣用語的自然運用，以及虛詞在口語和書面語中的使用頻率差異。亦可探究不同情境、不同對話人物背景間的語言使用情形，如禮貌、合作原則等。更能對詞彙教學提出實質建議。

Shirato(2005)比對了日本以英語為第二外語之教材與實際口語語料中的單字，其研究使用 BNC 語料庫及 West 的 The General Service List (GSL) 詞表。研究成果提出，詞彙教學中應強調 2000 個高頻字及與日常生活相關但不在 2000 個高頻字中的詞彙等作為教學建議。

綜上所述，口語語料庫詞的建置可提供有別於書面語的口語詞彙、固定短語、口頭慣用語等具口語特色詞彙，不僅可針對口語教學提出建議，亦有助於提供詞表編纂或詞彙教學更貼近日常交際用語。基於此因，本研究將藉口語語料中出現頻率較高之詞彙，對比著名華語能力測驗詞表—TOCFL 詞表，期能對 TOCFL 詞表內容提出更貼近日常交際的口語常用詞彙。詳細介紹於下節研究方法中加以論述。

### 3. 研究方法

如前所述，本研究有兩個主要目標：（1）自華語口語語料庫中取口語高頻詞對比檢視臺灣著名的華語能力測驗—TOCFL 詞表，（2）自口語高頻詞中提出可建議納入 TOCFL 詞表之口語高頻詞。下面將逐步介紹本研究語料及工具：

#### 3.1 研究之語料

##### 3.1.1 華語電影、電視劇口語語料庫

由於口語語料庫建置相當不易，且為收集大量母語者日常口語語料，多須仰賴大量人力、時間與金錢。雖然北京語言大學已建置了具 186 萬字的「當代北京口語語料庫」，然而因地域性的不同，其語言使用詞彙有所差異，故本研究將不採用。另一方面，儘管 Kilgariff et al. (2004) 提出語料資源越大越能體現準確及完整的語言行為。但中央研究院和政治大學所建置之口語語料庫內容大小皆未達 50 萬字。在考量語料大小及內容會影響詞表內容後，本研究並不採用。為更客觀地取得口語高頻詞，在時間及人力資源的考量下，本研究參照 The Leuven Drama Corpus 及 Cornell

Movie-Dialogs Corpus 的建置方式，取具有對話性質的華語電影及連續劇字幕對白作為母語者口語語料，建置較大型的口語語料庫。

在華語電影及電視劇的挑選上，以劇情背景為現代為主，且對白貼近現代日常口語為優先考量，共蒐集 63 部華語電影及 5 部華語電視劇<sup>13</sup>（名單詳列於附件一）。此 5 部在臺灣地區製作及拍攝的電視連續劇，是以現代連續劇為素材，包括「光陰的故事」、「赴宴」、「熱血青春」、「痞子英雄」、「我在墾丁天氣晴」。

本研究首先對 63 部華語電影及 5 套電視連續劇進行字幕抽取工作。一般字幕在 DVD 光碟中為圖形格式，因此本研究先以 VobSub<sup>14</sup>軟體將字幕圖檔抽出，再利用 SubToSrt<sup>15</sup>軟體對所截取出的字幕圖檔進行文字識別。由於每套電影、電視劇的字型有所差異，字庫未收錄之文字亦無法辨識，因此文字識別的過程中還需輔以人工輸入和校對。經過錯別字校對、缺漏字增補後，方能一一儲存為純文字檔。本次語料蒐集共累計 128 個純文字檔，近 153 萬 6 千多字的字幕語料。而後，本研究以中央研究院中文斷詞系統對 153 萬 6 千字幕語料進行自動斷詞，斷詞處理完成後統整為本研究之華語母語者口語語料。口語語料庫建置流程如下圖 1。

---

<sup>13</sup> 華語電影及電視劇以國語（普通話）為主，因部分電影、電視劇包含臺語、日語或原住民語，然而在字幕顯示上可作區分，因此不影響取詞。為避免粵語詞彙與國語（普通話）詞彙之混淆，本研究取樣語料，不包含粵語或包含粵語之電影（感謝匿名審查委員之悉心提醒考量普通話與粵語的差異）。因刪除包含粵語的電影語料，因此不僅語料庫字數更動，在口語高頻詞上的頻率及數量也受影響，詳細內容請見內文。

<sup>14</sup> VobSub（又名：VSFilter）是一套免費軟體，可進行字幕顯示、製作工具，亦可抽取 DVD 影片中的字幕圖檔，儲存成 idx、ifo、sub 三種字幕檔案，以對字幕檔案進行翻譯及編修。

<sup>15</sup> SubToSrt 為一免費軟體，其功用是將 DVD 字幕由 SUB 格式轉換成 srt 格式字幕，srt 格式的字幕為純文字檔，檔案體積小，有利於編輯及傳播。

### 分析華語口語語料庫高頻詞之特點並對 TOCFL 詞表提出建議

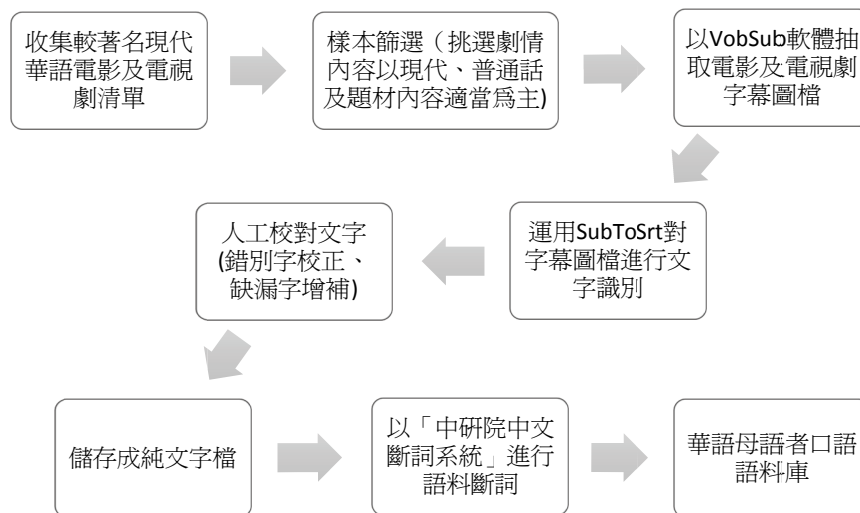


圖 1 華語母語者口語語料庫建置流程

#### 3.1.2 TOCFL 八千詞詞表

華語文能力測驗，英文為 Test of Chinese as a Foreign Language(簡稱 TOCFL)。由國家華語測驗推動工作委員會參照張莉萍 2003 年到 2004 年的研究成果編製，並於 2007 年提供線上下載。2013 年華測會研究人員將原詞表內容與國際上中文教學大綱中所列詞彙，同常見華語教材、中研院平衡語料庫及 TOCFL 學習者語料庫詞頻彙整成大詞表，進一步統計詞彙出現次數及專家學者意見後，調整詞表內容及所屬等級。新版 TOCFL 華語八千詞於 2013 年 5 月置於網上<sup>16</sup>。而詞表的詞彙量八千主要是依據鄭錦全在 1998 年提出的「詞涯八千」概念出發(張莉萍 2007)。

新版詞表按華語文能力測驗(TOCFL)等級—入門級、基礎級、進階級、高階級、流利級分列，分別對應《歐洲共同語文參考架構》(CEFR)的 A1、A2、B1、B2、C1。根據華測會表示，新版 TOCFL 詞表與前一版本相比最大的改變是確立了入門級、基礎級的詞彙內容，以及每個等級的詞彙量：入門級詞彙量為 500 個詞、基礎級 1000 個詞、進階級 2500 個詞、高階級 5000 個詞、流利級 8000 個詞。然而，在華測會官方下載之新版 TOCFL 詞表中實際上入門級為 500 個詞、基礎級 998 個詞、進階級 2501 個詞、高階級 4997 個詞、流利級 7989 個詞。

<sup>16</sup>發布於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)

新舊版 TOCFL 詞表的差異除了在確立各級詞彙量之外，新版詞表將原有各詞彙的英文解釋刪去，並在入門級和基礎級將詞彙依照「任務領域」分成 12 個類別：(1) 個人資料；(2) 工作；(3) 教育；(4) 房屋與家庭、環境；(5) 日常生活；(6) 閒暇時間、娛樂；(7) 與他人關係；(8) 旅行；(9) 健康及身體照護；(10) 購物；(11) 飲食；(12) 其他。

### 3.2 前置處理

本研究首先取得 TOCFL 官方詞表及口語語料庫詞表，發現 TOCFL 將每一級的詞都加註了詞性，同時將破音字列為不同詞條，造成 TOCFL 詞表字詞上的重複情形。再者，TOCFL 詞表就詞彙程度分成 5 級，且將入門級和基礎級的詞彙依全務領域分 12 個類別。為避免詞表比較上的差異，本研究僅就 TOCFL 的詞條比對口語語料庫的高頻詞表。

此外，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 GSL、BNC 及 COCA 詞表，在收詞時無一字多義的情形。因此，本研究對 TOCFL 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系詞，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」cháng（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 7,989 個詞減至 7,281 個詞。

另一方面，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」一詞在口語語料庫高頻詞表中為「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」。

在對比 TOCFL 詞表及口語語料庫高頻詞前，口語語料亦須進行前置處理以取得口語語料高頻詞表。本研究透過英國語言學家 Mike Scott 研發的語料檢索分析軟體—Wordsmith Tools（簡稱 WordSmith）對斷詞後的華語母語口語語料進行分析及計算，按詞頻排列出口語詞表，得 54,975 個詞彙。因詞彙量龐大，本研究僅擷取詞彙出現頻率達 20 次以上且在出現在 10 篇純文字檔以上出現的詞彙，共 3,349 個詞。然而因華語詞與詞間不存在明顯界線，在斷詞時易產生斷詞不完整的情況<sup>17</sup>，因此所

<sup>17</sup> 感謝匿名審查委員的悉心提點，對於斷詞系統的處理方式。本研究以中研院的斷詞系統對

得的 3,349 筆詞以人工再加以檢視因斷詞所產生是不完整詞彙。刪除內容包括：(1) 電腦自動斷詞錯誤，如「的是」、「了吧」、「是最」等。(2) 英文字，如「KA」、「B」等。(3) 姓氏，包括「李」、「蘇」、「郭」等後，篩選後得 2,717 筆口語高頻詞，以下稱為「口語高頻詞表」。

### 3.3 分析工具

本研究使用由程式設計師設計的對比工具進行 TOCFL 詞表與口語高頻詞表間內容之比較，對比分析工具網址：<http://140.122.83.245/diff/>。下圖 2 為對比工具畫面。

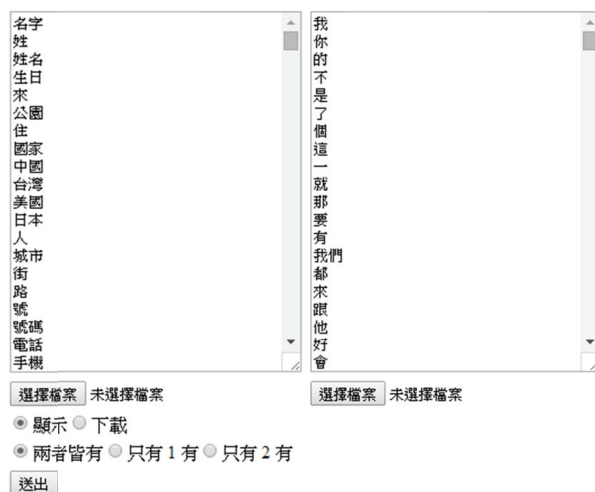


圖 2 對比分析工具

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。

---

所集結的口語語料進行斷詞。而中研院的斷詞系統在偵測未知詞的部份是利用基於語料庫的方法找出句法規則，再利用此規則來分辨單音節元素，(Chen & Bai 1998)。斷詞後的範例：這(Nep) 是(SHI) 一(Neu) 家(Nf) 擁有(VJ) 五百(Neu) 名(Nf) 員工(Na) 的(DE) 中型(A) 企業(Na) (徐立人 2011)。本研究對於斷詞後的詞彙，僅以人工方式審視刪除斷詞不完整內容，不針對詞頻作人工調整。

## 4. 研究結果與討論

### 4.1 TOCFL 詞表與口語高頻詞表數量上的差異

TOCFL 詞表為世界著名華語能力測驗用詞表之一，其內容是依測驗級數對詞彙進行分級，且詞表內每筆詞彙都標註了詞性。而口語語料庫所擷取出的口語高頻詞表無詞性標註。因此，本研究單以詞條不加上詞性比對 TOCFL 詞表及口語高頻詞表，藉以觀察出兩詞表間之異同。

本研究取 TOCFL 詞表經刪除重複出現詞及破音字詞後的 7,281 個詞與口語語料庫高頻詞表的 2,717 筆詞彙對比後，結果得到兩詞表中有 2,004 個詞為共有詞。而在口語語料庫高頻詞表的 2,717 個詞中，有 713 筆高頻口語詞（請見附件二）不在 TOCFL 詞表中。本研究進一步觀察 713 筆未納入 TOCFL 的口語高頻詞，發現其中包含大量詞塊合成詞<sup>18</sup>及多音節詞彙，這類詞彙型式在 TOCFL 詞表中較為少見，初步探究原因可能為 TOCFL 詞表選詞是以詞為最小單位。然而本研究發現 TOCFL 詞表當中其實亦包含例如：「我們」、「他們」、「小學生」、「大學生」等常見詞彙合成型式。另一方面，TOCFL 詞表中具口語詞彙特色的詞彙數量較少，因此本研究期透過自前人文獻中彙整出的 6 項口語詞彙特色對 713 筆口語高頻詞進行分類，並提出可作為 TOCFL 詞表增修之建議口語高頻詞彙。

### 4.2 口語建議詞

#### 4.2.1 前 238 筆詞頻達 60 頻次之口語高頻詞彙

根據 Biber, Conrad & Cortes (2004)對英語高頻詞的定義是每百萬字出現 40 次以上。然而，尚未有研究實際對中文高頻詞提出定義，因此本研究參考 Biber 等人的定義，取每百萬詞出現 40 次以上為口語高頻詞之挑選原則。因本研究語料超過 150 萬字，故按比例取詞頻達 60 頻次以上的詞彙視為高頻詞彙進行探討。因此，在本研究所得的 713 筆未納入 TOCFL 詞表之口語高頻詞中共有 238 筆的詞頻超過 60 的高頻詞，羅列於下表 2，詞彙下方的數字為該詞在口語語料庫中出現的頻率。

---

<sup>18</sup> 感謝匿名審查委員的用心指點，詞塊（語塊）的形式與詞不同。雖然 TOCFL 詞表以詞的型式作為收詞基本原則，但本研究也察驗了 TOCFL 詞表，當中包含了詞與詞塊的形式，例如：「我們」、「他們」、「小學生」、「大學生」等詞的組合。同時，本研究亦因篇幅關係未能更加深入探究詞與詞塊本質性問題，僅先於詞頻作為挑選建議詞的依據，然將在未來研究上進一步申論兩者間的差異，並提出更針對性的建議。



同時，為印證這 238 筆詞彙不僅是高頻且具口語特色的關鍵詞彙，本研究以 WordSmith 的關鍵字（keywords）<sup>19</sup>抽取工具取口語詞彙之關鍵詞，將口語語料庫高頻詞詞表（2,717 筆詞彙）作為研究語料（study corpus），中研院平衡語料庫 5 百萬字所得出的詞表作為參考語料（reference corpus），得出 2,480 筆口語關鍵詞。另將此 2,480 筆關鍵詞與 TOCFL 詞表比對後，得出有 823 筆關鍵詞不在 TOCFL 詞表中。在這 823 筆詞彙當中，詞頻 60 以上的詞彙共有 226 筆，而這 226 筆詞彙皆出現在表 2 的 238 筆高頻詞中（灰階標示出 12 筆口語高頻詞未出現在 226 筆口語關鍵詞中的口語高頻詞），兩者相近度達 95%，亦可表示 238 筆高頻詞為具口語特色關鍵詞彙。

表 2 238 筆詞頻達 60 以上的口語高頻詞

|            |            |            |            |            |             |             |            |            |            |            |            |
|------------|------------|------------|------------|------------|-------------|-------------|------------|------------|------------|------------|------------|
| 不要<br>3772 | 不是<br>2322 | 耶<br>2124  | 不會<br>2040 | 一下<br>1957 | 回來<br>1312  | 好不好<br>1176 | 不能<br>1108 | 沒事<br>939  | 看到<br>932  | 有沒有<br>871 | 等一下<br>703 |
| 怎樣<br>673  | 想要<br>608  | 阿公<br>592  | 不行<br>538  | 阿爸<br>529  | 快點<br>527   | 要不要<br>524  | 囉<br>510   | 很多<br>500  | 回家<br>488  | 過來<br>459  | 走了<br>418  |
| 小小<br>415  | 吃飯<br>413  | 在一起<br>386 | 找到<br>345  | 聽到<br>342  | 都是<br>335   | 來了<br>327   | 會不會<br>319 | 跆拳道<br>297 | 不好<br>291  | 那麼多<br>290 | 談戀愛<br>290 |
| 跆拳道<br>277 | 女朋友<br>272 | 跆拳道<br>265 | 這麼多<br>259 | 不對<br>256  | 可不可以<br>249 | 知不知道<br>243 | 社長<br>241  | 很好<br>237  | 臺北<br>235  | 不可以<br>233 | 好的<br>229  |
| 笨蛋<br>226  | 不可能<br>225 | 黑客<br>216  | 咧<br>204   | 有事<br>204  | 一輩子<br>202  | 阿嬤<br>202   | 沒錯<br>200  | 去了<br>200  | 老爸<br>199  | 村子<br>195  | 唸書<br>194  |
| 死了<br>193  | 金牌<br>193  | 忘了<br>189  | 大姊<br>187  | 上面<br>184  | 墾丁<br>178   | 爸媽<br>175   | 好多<br>172  | 太多<br>170  | 到時候<br>170 | 後面<br>164  | 你好<br>164  |
| 前面<br>159  | 小孩子<br>159 | 男朋友<br>158 | 拿去<br>156  | 有人<br>156  | 你看<br>155   | 沒辦法<br>152  | 沒問題<br>152 | 還不是<br>149 | 跑去<br>148  | 想想<br>147  | 家裡<br>146  |
| 麵攤<br>144  | 交給<br>142  | 總動員<br>141 | 不准<br>138  | 學長<br>136  | 說了<br>133   | 女孩<br>132   | 叫做<br>131  | 局長<br>131  | 小時候<br>130 | 分局<br>129  | 心裡<br>125  |
| 通通<br>124  | 女孩子<br>121 | 還給<br>119  | 年輕人<br>118 | 怎麼說<br>118 | 會長<br>118   | 想說<br>117   | 不見了<br>117 | 過了<br>116  | 考上<br>115  | 們<br>112   | 哭了<br>110  |
| 上來<br>109  | 唸<br>109   | 越來越<br>107 | 打了<br>106  | 做了<br>106  | 組長<br>105   | 收到<br>104   | 超<br>104   | 變得<br>101  | 待會兒<br>101 | 牛肉麵<br>101 | 走開<br>100  |
| 說清楚<br>100 | 過分<br>99   | 上去<br>97   | 賺錢<br>97   | 多久<br>97   | 試試看<br>96   | 好人<br>95    | 能不能<br>94  | 還可以<br>94  | 打給<br>93   | 上車<br>93   | 放在<br>93   |
| 嫁給<br>92   | 參賽<br>91   | 帶走<br>88   | 自個兒<br>87  | 不在<br>87   | 在家<br>87    | 好久<br>87    | 好不容易<br>87 | 回到<br>86   | 喝酒<br>86   | 聯絡<br>86   | 好朋友<br>85  |

<sup>19</sup> WordSmith 的關鍵字工具主要是透過比對研究語料（study corpus）詞表與參考語料（reference corpus）詞表。研究語料的語料量大小必須小於參考語料之語料量。關鍵詞（Keywords）被視為較具專業性/特殊性詞彙的主要原因是這些詞彙在研究語料中的詞頻高於參考語料中的詞頻，具高度顯著性且具信息檢索相關性（aboutness）（Phillips 1989; Berber-Sardinha 2000）。

|            |            |           |           |           |           |           |          |          |           |          |          |
|------------|------------|-----------|-----------|-----------|-----------|-----------|----------|----------|-----------|----------|----------|
| 便當<br>85   | 韓國<br>85   | 準備好<br>84 | 拿到<br>84  | 搞不好<br>84 | 雞婆<br>84  | 閉嘴<br>83  | 不只<br>83 | 開門<br>82 | 沒用<br>82  | 跑到<br>82 | 拿來<br>82 |
| 都在<br>82   | 打贏<br>81   | 不應該<br>81 | 聽話<br>81  | 嚇死<br>80  | 南區<br>79  | 到了<br>79  | 寫信<br>79 | 好玩<br>79 | 跑來<br>79  | 姊夫<br>79 | 多了<br>78 |
| 送給<br>78   | 白癡<br>78   | 馬子<br>78  | 乖乖<br>77  | 說過<br>77  | 壞人<br>77  | 留下來<br>77 | 不理<br>77 | 衝浪<br>76 | 都有<br>75  | 高雄<br>75 | 胡說<br>75 |
| 想過<br>74   | 站在<br>74   | 當成<br>74  | 小聲<br>74  | 不停<br>74  | 吃了<br>73  | 打死<br>73  | 放開<br>72 | 長官<br>72 | 一邊<br>71  | 氣死<br>71 | 偏偏<br>70 |
| 留下<br>70   | 做生意<br>70  | 待在<br>70  | 真相<br>70  | 對手<br>70  | 有夠<br>70  | 很久<br>69  | 吃完<br>68 | 不可<br>68 | 說真的<br>68 | 留在<br>67 | 路上<br>67 |
| 混蛋<br>67   | 見到<br>66   | 做錯<br>66  | 神經病<br>66 | 坐下<br>65  | 還在<br>65  | 來到<br>64  | 放下<br>64 | 打敗<br>64 | 出賽<br>64  | 拿給<br>64 | 長得<br>63 |
| 好聽<br>63   | 好久不見<br>63 | 夥伴<br>63  | 騙人<br>62  | 董事長<br>62 | 總務長<br>62 | 老頭<br>61  | 做好<br>61 | 走到<br>61 | 退學<br>61  | 噁心<br>61 | 帶著<br>61 |
| 生日快樂<br>61 | 等等<br>60   | 一塊兒<br>60 | 沒有用<br>60 | 說出來<br>60 | 早餐<br>60  | 善良<br>60  | 廁所<br>60 | 才是<br>60 | 十幾<br>60  |          |          |

本研究進一步觀察上表中，238 筆詞頻達 60 以上的口語詞彙，發現當中有 25 筆詞彙是「不」開頭的組合。這 25 筆以「不」開頭詞彙當中有 5 筆之詞頻排序在前 10 名，且詞頻皆超過 1 千頻次。這部分符合了張文賢、路雲和李曉琪（2012）對於口語詞塊中「不」開頭的詞彙居多之論點。另外上表二中亦可觀察到部分與「不」組成的詞彙是以肯定加上否定的形式呈現，例如：「會不會」、「知不知道」，這一類的詞組在句型中多用於疑問句。

此外，在前 80 筆口語高頻詞彙當中亦包含了與「沒」開頭或相關問句，比如：「沒事」、「有沒有」、「沒錯」等詞彙。無論是以「不」或「沒」組合的詞彙，例如：「不要」、「不是」、「不會」、「不能」、「沒事」、「沒問題」、「有沒有」、「會不會」、「知不知道」等，在日常口頭交際中使用的頻率皆偏高。而 TOCFL 詞表卻未納入之原因，本研究觀察並歸納，推論可能是 TOCFL 詞表分別收錄「不」、「要」、「是」、「會」、「能」、「行」、「沒」、「問題」、「有」、「沒有」，但未以詞塊型式呈現。然而 TOCFL 詞表內卻也收錄了「不便」、「不用」、「不同」等詞塊型式。相較於「不便」在口語交際上的使用，「不要」、「不是」、「不會」反而較常見。因此，建議可將這些口語高頻常用詞組納入詞表中，以提供學習者在交際中常見的語言模式，增加學習者使用語言的掌握能力。

在前 238 筆頻率超過 60 頻次以上的口語高頻詞中亦可觀察到稱謂相關的詞彙，比方：「阿公」、「阿爸」、「阿嬤」等，這類稱謂用詞中多數詞彙受閩南方言的影響，本研究亦將於下文 4.2.2 中的「較多方言詞/外來語」更進一步探討。另外，還有部分詞彙與男女情感相關的用詞，例如：「在一起」、「談戀愛」、「女朋友」、

「男朋友」等。其中「在一起」的用法除了表示男女之間交往的關係外，亦可作為兩人或兩人以上同處於一個地方，在口語交際中亦常見。此外，因受華語電影、電視劇內容及情節之影響，這些口語高頻詞當中包含了部分專有名詞或地名，例如：「跆拳道」、「跆拳道」、「墾丁」、「韓國」、「高雄」等，這類詞彙雖詞頻偏高，但偏屬專業類用詞，因此本研究建議可作為補充詞彙使用。

本研究根據李如龍及吳茗（2005）提及，對外漢語詞彙教學若要提高成效，建議以常用先教為原則。學習者在詞彙學習上若可多接觸或學習母語者常用詞彙，想必可提高詞彙學習成效。因此，本研究首先於表 2 羅列了 238 筆未納入 TOCFL 詞表但具口語詞彙特色之高頻常用詞作為建議詞彙，提供詞表編纂人員及華語教師於進行詞表選詞或詞彙教學上運用，期以提升華語詞彙教學的豐富性及全面性。

#### 4.2.2 口語高頻詞彙之特色

本研究除了提出 238 筆詞頻高達 60 次以上口語常見高頻詞作為詞表增修優先建議參考詞彙外，另將對比 TOCFL 詞表與口語詞彙表後所得的 713 筆未出現在 TOCFL 詞表的口語高頻詞，參照文獻探討中自 9 篇前人文獻（詳細內容請見前文 2.2 口語詞彙的特色及重要性）所彙整而出的 6 項口語詞彙特色進行分類，此 6 項口語詞彙特色分別為：

- (1) 較多雙聲及疊韻詞
- (2) 較多以詞綴衍生出的詞
- (3) 較多語氣詞／嘆詞／擬聲詞
- (4) 較多詞塊合成詞型式
- (5) 較多多音節熟語
- (6) 較多方言詞／外來語

本研究分析 713 筆口語高頻詞，發現具「詞塊合成詞型式」的筆數最多，有 410 筆口語高頻詞。其次為「多音節熟語」的 51 筆詞彙，接續是「以詞綴衍生出的詞」有 17 筆、「雙聲及疊韻詞」的 25 筆詞彙及「語氣詞／嘆詞／擬聲詞」類別有 13 筆，最後是「方言詞／外來語」的 9 筆。

由於本研究僅針對口語詞彙特色進行討論，而 713 筆自語料庫歸納出的口語高頻詞，有部分詞彙(188 筆)未能歸類於上述 6 項口語詞彙特色中，因篇幅關係，本研究 188 筆未能分類於 6 項口語詞彙特色之詞彙列於附件二，供教師及詞表編輯人員參考。圖 3 為 716 筆口語特色詞彙分佈於上述 6 項口語詞彙特點之數量，其中亦包含 188 筆未能歸類於 6 項口語詞彙特色之詞彙數量。

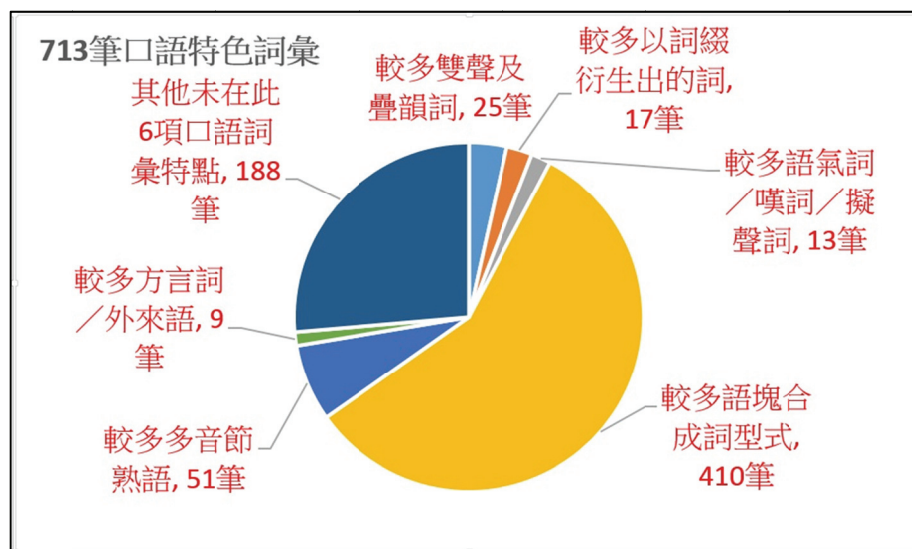


圖 3 713 筆口語特色詞彙分佈於口語詞彙特點之數量及比例

下文將按圖 3 所呈現的口語特色詞彙所得詞彙量高低，逐一介紹 713 筆口語高頻詞中具 6 項口語詞彙特色之詞彙。

#### (1) 較多詞塊合成詞型式

在 713 筆口語高頻詞當中，共有 410 筆詞彙歸為詞塊合成詞型式，佔全數一半以上。詞塊 (chunk) 型式之所以常用於語言交際中，可歸因於語言是由詞彙的組合串連而成，而詞塊是為詞彙組合的模式 (胡元江 2011)。Wray (2002) 定義詞塊是由兩個或兩個以上詞彙構成連續的或不連續的序列，整體儲存在記憶中，使用時整體提取，為一種預制語言單位 (錢旭菁 2008)。Wray (2000) 亦表示，詞塊包含多詞單位 (multiword units 又可視為合成詞)、搭配 (collocations)、預制模塊 (prefabricated patterns) 及詞塊 (chunk) 等，均可統稱為詞塊 (Formulaic Sequence)。另有研究人員認為詞塊包括：合成詞、短語動詞、成語、固定詞組及預制詞塊 (徐海美 2009；翟瑩 2012)。

錢旭菁 (2008) 亦進一步指出，詞塊包含搭配、慣用語、成語、歇後語、諺語、格言、名言、警句、會話套語、兒歌、歌詞、宗教經文等，都是由多個詞構成，具整體儲存、提取及使用的語言結構。其中慣用語、歇後語、諺語等約定俗成之熟語為口語詞彙的特點之一 (楊俊萱 1984；常敬宇 1986；李如龍 2007；謝智香 2011)，因此本研究將口語常用詞中的熟語自詞塊中分別出來在口語詞彙特點「(2) 較多多音節熟語」中更進一步討論，詳細內容請見後文。

本研究所得未納入 TOCFL 之口語詞彙中以詞塊型式合成詞數量較多，佔全數 713 之半數以上。有鑑於詞塊型式合成詞所涵蓋的範圍廣泛，為能更具體地提出具口語特點之建議詞彙，本研究參照曹煒（2003）、尹惠貞（2006）對口語詞彙的研究結果，討論口語合成詞彙中的「動賓」、「偏正」、「並列」、「動補」類高頻詞中的詞塊合成詞<sup>20</sup>。此外，張文賢、路雲、李曉琪（2012）提出以「不」開頭的詞在口語高頻詞中的數量較多，因此本研究亦特別將與「不」/「沒」相關的詞組另闢一類進行討論<sup>21</sup>。而未能分類於在上述「動賓」結構、「偏正」結構、「並列」結構、「動補」結構及以「不」/「沒」相關詞組之詞塊則歸類於其他型式詞塊進一步探討。

本研究將 410 筆口語高頻詞塊依下列 6 項結構型式進行分類，a.「動賓」結構之詞塊合成詞，49 筆；b.「偏正」結構之詞塊合成詞，66 筆；c.「並列」結構之詞塊合成詞，6 筆；d.「動補」結構之詞塊合成詞，110 筆；e.與「不」及「沒」相關之詞塊合成詞，50 筆；f.其他型式詞塊合成詞，129 筆。詳細內容於下文一一陳述。

#### a.「動賓」結構之詞塊合成詞

動賓結構合成詞是動詞與賓語的搭配情形，例如：「回家」、「賺錢」、「上車」、「做生意」、「殺人」、「花錢」、「看醫生」等。由於詞塊的組合結構常多與詞彙間的搭配情形有關，且在語言交際上影響學習者掌握語言的準確性（劉玫芳 2013），因此建議若能在詞表中納入以口語詞彙常見的搭配型式組合而成的詞塊，不僅可幫助提升學習者對於詞彙間搭配的認知能力，亦能增加學習者語言的流利度及準確性。雖然詞彙在搭配上不見得具有嚴謹的規則，但也不可任意組合，通常須遵循既有的搭配形式。不同詞語間的搭配有時所帶出的意思亦有差異，例如英文中的「go home」若以直接翻譯的形式會出現「去家」的情形，因此建議詞表可將詞彙在語言使用上常見型式視為單一詞條列於詞表中，將有助於學習者掌握目標語。表 3 詳列了 49 筆「動賓」結構之詞塊合成詞。

<sup>20</sup> 根據曹煒（2003）、尹惠貞（2006）對口語詞彙的研究顯示，口語詞彙以「動賓」、「偏正」、「並列」、「動補」等結構的詞數居多，而「主謂」結構的詞塊在口語詞彙當中較為少見。因此，本研究僅就「動賓」、「偏正」、「並列」、「動補」結構之詞塊合成詞進行討論。

<sup>21</sup> 因「不」與「沒」在語意及組合結構上相近，因此本研究將「不」與「沒」相關詞彙放在同一類別進行討論。

表 3 49 筆「動賓」結構之詞塊合成詞

|     |    |    |    |    |    |     |     |    |    |
|-----|----|----|----|----|----|-----|-----|----|----|
| 回家  | 吃飯 | 唸書 | 賺錢 | 上車 | 參賽 | 喝酒  | 閉嘴  | 開門 | 寫信 |
| 做生意 | 騙人 | 退學 | 出國 | 殺人 | 騎車 | 吃醋  | 教我  | 喝醉 | 借我 |
| 拚命  | 頂嘴 | 下車 | 開槍 | 喝茶 | 花錢 | 喝水  | 看醫生 | 找人 | 減肥 |
| 打人  | 作夢 | 闖禍 | 被人 | 讓人 | 煮飯 | 通電話 | 停車  | 打球 | 幹活 |
| 求你  | 下山 | 抽煙 | 借錢 | 坐車 | 罵人 | 存錢  | 拿錢  | 抽菸 |    |

## b. 「偏正」結構之詞塊合成詞

偏正結構合成詞則多為表達詞彙間的修飾情形，例如：「很多」、「女朋友」、「好玩」、「真好」、「太晚」等。不論是副詞與形容詞或動詞間的修飾關係，亦或是形容詞與名詞間的修飾搭配，學習者若能掌握此常用語言形式，對於語言表達上可增添豐富色彩，例如：「很好」與「更好」間的差異，若能在詞表中分別條列出來，應可增加學習者對於「很」及「更」使用上的分辨能力。建議若詞表斟酌將口語詞彙的常用偏正結構詞塊納入詞表，相信將有助於提升學習者對於詞彙修飾上的認知能力。本研究將 66 筆口語高頻詞中「偏正」結構之詞塊羅列於表 4。

表 4 66 筆「偏正」結構之詞塊合成詞

|    |     |     |    |    |    |    |    |     |     |    |
|----|-----|-----|----|----|----|----|----|-----|-----|----|
| 很多 | 女朋友 | 很好  | 上面 | 好多 | 太多 | 後面 | 前面 | 男朋友 | 年輕人 | 多久 |
| 好人 | 好久  | 好朋友 | 南區 | 好玩 | 壞人 | 小聲 | 一邊 | 很久  | 好聽  | 噁心 |
| 好棒 | 好事  | 亂講  | 美女 | 亂說 | 帥哥 | 外人 | 年前 | 偷看  | 很忙  | 最愛 |
| 很像 | 更多  | 好心  | 很難 | 很棒 | 很大 | 好笑 | 很愛 | 好想  | 下面  | 壞事 |
| 難聽 | 亂來  | 很少  | 急事 | 站好 | 好喝 | 好痛 | 太大 | 好意  | 亂動  | 真愛 |
| 新來 | 多好  | 慢走  | 死心 | 趁熱 | 偷吃 | 更好 | 很痛 | 亂丟  | 很快  | 真好 |

## c. 「並列」結構之詞塊合成詞

並列結構合成詞，意指兩個意義相同、相近、相關或相反的詞並列組合而成的。713 筆口語高頻詞中有 6 筆「並列」結構之詞塊合成詞：「出走」、「好歹」、「詛咒」、「男女」、「搶劫」及「稀罕」。

並列結構合成詞之特色為兩個詞前後順序一般不可隨意調換。例如「好歹」（表：好和壞）不能說成「歹好」。然而，儘管曹煒（2003）及尹惠貞（2006）對口語詞彙的研究皆表示口語詞彙以「動賓」、「偏正」、「並列」、「動補」等結構的詞數居多，而本研究實際分析口語高頻詞中語塊型式合成詞發現「並列」式結構之語

塊合成詞並不多，只有 6 筆。儘管筆數不多，但因並列結構組合的順序不可隨意調換，因此在學習上可多讓學習者熟悉此結構以避免出現詞彙偏誤的情形。

#### d. 「動補」結構之詞塊合成詞

410 筆詞塊合成詞中，以「動補」結構合成詞所佔的數量最高，共有 110 筆（詳見下表 5）。動詞+補語的組合型態在語言交際上多作為表達動作的結果或狀態，若學習者可多學習這類相關的詞彙，可增加語言的複雜性。本研究自 110 筆「動補」結構之詞塊合成詞當中觀察，歸納出 3 項主要組合型式：(a) V.+趨向補語：「回來」、「拿去」等。(b) V.+狀態補語：「說清楚」、「準備好」等。(c) V.+結果補語：「打贏」、「嚇死」。這類「動補」結構合成詞多因動詞後面所接的補語不同而影響詞彙意涵，若在詞彙教學上可多加以分辨動詞後接補語的差異，讓學習者以詞塊型式學習詞彙，相信可提升學習者在語句理解上的能力，例如：「回來」、「回去」與「回到」詞意上的異同。因此建議詞表可參考詞頻並斟酌納入高頻常用口語「動補」結構合成詞，以協助學習者更廣泛地掌握母語者常用語言型式。

表 5 110 筆「動補」結構之詞塊合成詞

|     |      |     |     |     |     |     |     |     |    |     |
|-----|------|-----|-----|-----|-----|-----|-----|-----|----|-----|
| 回來  | 看到   | 過來  | 找到  | 聽到  | 拿去  | 跑去  | 考上  | 上來  | 收到 | 走開  |
| 說清楚 | 上去   | 放在  | 帶走  | 回到  | 準備好 | 拿到  | 跑到  | 拿來  | 打贏 | 嚇死  |
| 跑來  | 留下   | 打死  | 放開  | 氣死  | 留下  | 吃完  | 做錯  | 見到  | 坐下 | 放下  |
| 打敗  | 來到   | 做好  | 說出來 | 拿出來 | 做到  | 坐下來 | 讓開  | 愛上  | 打來 | 說到  |
| 抓到  | 站起來  | 打到  | 想起  | 解散  | 看清楚 | 講到  | 坐好  | 聽清楚 | 撞到 | 拿出  |
| 死掉  | 帶回來  | 看完  | 聽得懂 | 拿走  | 丟掉  | 停下來 | 跑掉  | 拿過來 | 搞錯 | 害死  |
| 解開  | 看好   | 想清楚 | 壞掉  | 搞清楚 | 嚇到  | 送去  | 帶去  | 收好  | 看錯 | 找出  |
| 生下來 | 想通   | 查到  | 躲起來 | 搞定  | 帶回去 | 說出  | 吵醒  | 想好  | 笑死 | 吃點  |
| 丟下  | 收下   | 累死  | 約好  | 往後  | 聽好  | 做不到 | 搬走  | 抓住  | 送來 | 拿回去 |
| 煩死  | 看得出來 | 拍到  | 講好  | 掉下來 | 賣給  | 關掉  | 講出來 | 找出來 | 輪到 | 感受到 |

#### e. 與「不」及「沒」相關之詞塊合成詞

以「不」及「沒」開頭的詞彙多具否定意，在語言表達上清楚回應說話者對於某件事的拒絕或否定情緒，比如：「不要」、「不是」、「不會」、「不行」、「沒事」、「沒辦法」等。此類否定用語亦常見於日常交際中。

此外，在語言使用中常將肯定及否定的詞彙組合在一起作為問句型式，例如：「會不會」、「知不知道」、「有沒有」等。本研究進一步檢視 TOCFL 詞表發現，詞表編纂人員僅以不同詞條列出肯定及否定詞彙，比如：TOCFL 詞表分別收錄了「有」及「沒有」。然而在日常交際上，不論是肯定、否定或是問句型式的詞塊合成詞皆常見，建議若可將詞彙的否定及問句型式納詞表中，可幫助學習者在表達拒絕、否定情緒的語言掌握情形，亦可增加學習者對於問句型式的運用情況。下表 6 詳列了 50 筆與「不」及「沒」相關之詞塊合成詞。

表 6 50 筆與「不」及「沒」相關之詞塊合成詞

|     |     |      |      |      |     |     |     |     |      |
|-----|-----|------|------|------|-----|-----|-----|-----|------|
| 不要  | 不是  | 不會   | 好不好  | 不能   | 沒事  | 有沒有 | 不行  | 要不要 | 會不會  |
| 不好  | 不對  | 可不可以 | 知不知道 | 不可以  | 不可能 | 沒錯  | 沒辦法 | 沒問題 | 不准   |
| 能不能 | 不在  | 不只   | 沒用   | 不應該  | 不理  | 不停  | 不可  | 沒有用 | 行不行  |
| 對不對 | 不夠  | 不一定  | 不敢   | 記不記得 | 不懂  | 沒空  | 不怕  | 不要緊 | 不配   |
| 不算  | 找不著 | 沒錢   | 不想   | 不說   | 沒出息 | 懂不懂 | 不像  | 不多  | 喜不喜歡 |

#### f. 其他型式詞塊合成詞

本研究將 410 筆語塊合成詞中非上述 5 項詞塊合成詞類別的 129 筆詞彙歸納為「其他型式詞塊」，並列於下表 7。而在此類別當中，本研究觀察到以下幾種主要型式：(a) 動態助詞「了、著、過」<sup>22</sup>，此三個助詞多與動詞所表示的時間相關。例如：「走了」、「來了」、「拿著」、「說過」等。(b) 以「這麼」、「那麼」、「怎麼」開頭的詞塊型式，這類詞彙在語言表達上為副詞形式，多用於表達「程度」上的差異。如「這麼多」與「那麼多」在口語中常用於表示數量之多，超過說話者的想像範圍。同樣地，「怎麼說」在口語表達上可替代「為什麼」用以詢問原因。(c) 助詞「的」詞組，如「好的」、「買的」等。(d) 結構助詞「之」詞組，比如：「之外」、「之中」等。(e) 連接詞形式，如「更何況」亦可用於語氣轉折的情境中。(f) 表條件用語，如「才行」亦用於表達事情達成所需要的條件。(g) 能願動詞「想」詞組，比如：「想要」、「想說」等。以上述 7 個類別中的詞彙雖未在前人文獻所歸納出口語詞彙特點中，但亦可常見用於口語交際中，建議詞表可斟酌納入作為增修參考。

<sup>22</sup>朱慶明(2005)《現代漢語實用語法分析》中提到，「了、著、過」這三個詞，在漢語語法中視為動態助詞，因為它們都跟動詞所表示的時間有關。



值得一提的是，由助詞「的」與「之」衍生出的詞彙，例如：「好的」、「假的」、「之外」、「之中」等，亦常見於口語交際當中。然據本研究觀察 TOCFL 詞表，對於這類詞彙的排列，有部分歸在一個詞條，但亦有部分詞彙是以兩個詞條呈現，前者例如：進階級的「別的」；後者比方：「好的」在 TOCFL 中收錄為「好」（入門級）、「的」（入門級）。同樣的，在 TOCFL 詞表的進階級分別收錄了「之後」、「之間」、「之前」。而「之外」、「之中」在 TOCFL 詞表中分別以不同詞條呈現，「之」（高階級）、「外」（基礎級）、「中」（進階級）。上述詞彙為詞彙的衍生形式，多於在日常交際中出現。若詞表納入詞彙常見詞塊型式，可提高學習者對詞組的理解能力。

表 7 129 筆其他型式詞塊合成詞

|    |     |    |    |    |     |     |    |    |    |    |
|----|-----|----|----|----|-----|-----|----|----|----|----|
| 想要 | 快點  | 走了 | 都是 | 來了 | 那麼多 | 這麼多 | 好的 | 去了 | 死了 | 忘了 |
| 你好 | 家裡  | 交給 | 說了 | 叫做 | 心裡  | 還給  | 想說 | 過了 | 哭了 | 打了 |
| 做了 | 變得  | 過分 | 打給 | 嫁給 | 在家  | 都在  | 到了 | 多了 | 送給 | 說過 |
| 都有 | 想過  | 站在 | 當成 | 吃了 | 待在  | 留在  | 出賽 | 拿給 | 長得 | 走到 |
| 帶著 | 一家人 | 拿著 | 夠了 | 晚安 | 留給  | 行了  | 也有 | 來過 | 您好 | 住在 |
| 老了 | 之外  | 妳好 | 出了 | 完了 | 弄得  | 大了  | 買的 | 錯了 | 聽過 | 睡了 |
| 鬧了 | 借過  | 買了 | 輸了 | 想吃 | 上了  | 沒了  | 有了 | 晚了 | 見過 | 手上 |
| 喝了 | 生了  | 找了 | 拿了 | 贏了 | 眼前  | 懂了  | 忙著 | 打完 | 離家 | 只能 |
| 毀了 | 無關  | 躲在 | 今晚 | 去哪 | 說完  | 在忙  | 殺了 | 瘋了 | 掉了 | 帶到 |
| 那有 | 又是  | 忙了 | 壞了 | 講了 | 活著  | 一早  | 坐著 | 重來 | 累了 | 讓給 |
| 急著 | 抱著  | 帶了 | 去過 | 開了 | 吃到  | 想了  | 少了 | 留著 | 做完 | 想去 |
| 想聽 | 昨晚  | 並肩 | 寫了 | 賣了 | 跟上  | 寄給  | 傻了 |    |    |    |

## （2）較多多音節熟語

因口語詞彙通常為約定熟成之語言形式，因此熟語型式常見於口語交際當中。若學習者可直接學習此類詞彙組合，應有助於擴展其語言表達及交際能力，並可於口語表達時更貼近母語者使用之語言。本研究觀察 TOCFL 詞表，當中雖也收錄了部分熟語或成語，如：「不好意思」、「無可奈何」、「瞧不起」、「形形色色」等。然而若學習者可多掌握口語常見熟語，例如：「等一下」、「在一起」、「談戀愛」、「生日快樂」、「好不容易」、「胡思亂想」等，可幫助學習者更多理解

日常交際語言中的文化內涵，且應有助於擴展其語言表達及交際能力更貼近母語者交際語言。下表 8 羅列了 51 筆口語高頻多音節熟語。

表 8 51 筆多音節熟語

|      |      |     |      |      |      |      |      |      |
|------|------|-----|------|------|------|------|------|------|
| 等一下  | 在一起  | 談戀愛 | 一輩子  | 到時候  | 還不是  | 總動員  | 小時候  | 怎麼說  |
| 不見了  | 越來越  | 還可以 | 好不容易 | 搞不好  | 說真的  | 好久不見 | 生日快樂 | 莫名其妙 |
| 有一點  | 一陣子  | 你們倆 | 更何況  | 看不出來 | 亂七八糟 | 不對勁  | 有出息  | 親愛的  |
| 負責任  | 三更半夜 | 所有人 | 不要臉  | 喜歡上  | 有的時候 | 胡思亂想 | 對得起  | 這陣子  |
| 從小到大 | 哥兒們  | 自己人 | 一模一樣 | 比如說  | 一家子  | 上廁所  | 就是了  | 在一塊兒 |
| 最大的  | 今兒個  | 再一次 | 我幫你  | 怎麼著  | 找麻煩  |      |      |      |

### (3) 較多雙聲及疊韻詞

在雙聲及疊韻詞方面，口語使用上因受白話語影響，以雙音節詞佔優勢（楊俊萱 1984；常敬宇 1986；尹惠貞 2006；李如龍 2007）。單音節動詞、形容詞及副詞在口語表達上則常見重疊型式，例如：「想想」、「試試看」、「乖乖」、「等等」、「聽聽」、「瞧瞧」、「小小」、「一點點」、「剛剛好」、「輕輕」等。上述重疊型式詞彙常使用於日常交際，若能在詞表中參考納入，應可協助學習者增進日常口語的聽、說能力。而「囉唆」在口語語料中屬於高頻詞，在口語使用上多使用於表達某人說話絮絮叨叨；辦事不痛快等（取自《新華字典》）。建議可於教學中做為附加教材，亦可視詞表的收詞需求收錄於詞表中。口語高頻詞中具雙聲及疊韻詞羅列於下表 9。

表 9 25 筆雙聲及疊韻詞

|    |    |    |     |      |     |    |     |     |
|----|----|----|-----|------|-----|----|-----|-----|
| 小小 | 想想 | 通通 | 試試看 | 乖乖   | 偏偏  | 等等 | 一點點 | 聽聽  |
| 囉唆 | 瞧瞧 | 問問 | 說說  | 輕輕   | 剛剛好 | 聊聊 | 尿尿  | 聊聊天 |
| 傻傻 | 玩玩 | 找找 | 試試  | 辛辛苦苦 | 勸勸  | 遠遠 |     |     |

### (4) 較多以詞綴衍生出的詞

根據朱亞軍、田宇（2000）對詞綴的定義：『詞綴必須是定位的虛詞素，只能放在其他詞素之或之後，如「老」、「阿」、「初」、「兒」、「子」、「頭」等。』。而在口語高頻詞裡以詞綴衍生出的詞彙，包含以下幾個類別：一、以前綴「老」開頭的詞，如：「老爸」、「老頭」、「老外」、「老媽」、「老天」等。二、以後綴「子」、「兒」結尾的詞，如：「村子」、「女孩子」、「事兒」、「條子」、

「大夥兒」等。這類以詞綴衍生出的詞彙，其口語色彩較為強烈（楊俊萱 1984；尹惠貞 2006；李如龍 2007；謝智香 2011；張文賢、路雲、李曉琪 2012），例如：「老爸」、「老外」、「事兒」、「男孩子」、「條子」（口語上延伸表示「警察」的意思）等。而在 TOCFL 詞表中已收錄許多以「兒」結尾的詞，楊惠媚、陳浩然及潘依婷（2014）對此提出建議認為「兒化」現象多出現在口語或地方方言中，除詞彙本在口語使用上已具備特定詞意之「兒化」詞，不宜省略「兒」外，在詞表中可多加評估斟酌「兒化」詞的收詞內容。建議詞表可多參考書面及口語高頻詞中常見以詞綴衍生出的詞彙，以提供學習者更具日常交際使用之詞語。下表 10 詳列口語高頻詞中以詞綴衍生出之詞彙。

表 10 17 筆以詞綴衍生出的詞

|     |    |     |     |    |     |    |    |    |
|-----|----|-----|-----|----|-----|----|----|----|
| 老爸  | 村子 | 女孩子 | 自個兒 | 老頭 | 一塊兒 | 今兒 | 事兒 | 條子 |
| 大夥兒 | 老外 | 男孩子 | 明兒  | 老媽 | 老太婆 | 男子 | 膽子 |    |

#### （5）較多語氣詞／嘆詞／擬聲詞

713 筆口語語料高頻詞中，有 13 筆語氣詞／嘆詞／擬聲詞，如：「咧」、「啥」、「哈」、「噓」、「唉唷」、「吶」等列於下表 11。其中「啥」在口語交際中，多代替「什麼」作為問句形式的使用。例如詢問他人中餐要什麼，在口語中可常聽到「中午要吃啥？」。而「咧」一詞，根據漢語字典的解釋，其讀音及用法有 4 種，如：「咧(lie3)嘴」；「大大咧(lie1)咧」；「瞎咧(lie2)」；「好咧(lie)」。「咧」當助詞用時，與「了、啦、哩」意思相近。此外，「哈」、「噓」和「吼」是為擬聲詞所演化而成的，「哈」、「哈哈」多用於代表笑聲[ha]；而「噓」則是用於表示制止或驅逐之意，例如：噓！別出聲（引自《漢語字典》<sup>23</sup>）；「吼」的原義為〔（野獸）大聲叫〕（引自《漢語字典》），在口語中已延申為表示憤怒或情緒激動而大聲呼喊。「吶」本義為吶喊、大聲呼喊等，亦可作為表示提醒注意用，例如：吶，我們可以開始了<sup>24</sup>。「喏」用於句尾有兩種涵義，一、表示「罷了」，如：不懂就學喏。二、表示同意、命令，意思與「吧」相近，如：去就去喏（引自《漢語字典》）。最後「唉唷」及「唉呦」兩者為感嘆詞，意思與英文的“oh”同，兩者僅在字的寫法上不同，可互用。

<sup>23</sup> 《漢語字典》<http://www.zdic.net/>

<sup>24</sup> 引自中國百科網 <http://www.chinabaike.com/dir/zidian/n/538760.html>

雖然上述詞彙多作為語助詞使用，但若學習者多增加此類詞彙相關使用方式，應可在日常交際中更貼近母語者所使用的語言型式。

表 11 13 筆語氣詞／嘆詞／擬聲詞

|   |   |   |    |    |   |   |   |   |   |    |
|---|---|---|----|----|---|---|---|---|---|----|
| 咧 | 啥 | 哈 | 哈哈 | 唉唷 | 啞 | 喂 | 掰 | 噓 | 吼 | 唉呦 |
|---|---|---|----|----|---|---|---|---|---|----|

#### (6) 較多方言詞／外來語

方言詞與外來語中較多反映出口語化的稱謂詞，如「阿公」、「阿嬤」、「阿爸」、「阿伯」、「阿母」等，上述詞彙在閩南方言中表示對祖父、祖母、父親、伯父及母親的稱呼，較具地方文化意涵，建議可在文化用語教學上提供給學習者參考。另外，「哈囉」已是日常口語交際中常用的招呼語，是由英文中的「Hello」譯音而成。同時，「便當」一詞則是以日文譯音而來，此兩筆詞彙亦多出現於日常生活中，建議可參考納入詞表中。下表 12 詳列 9 筆方言詞及外來語。

表 12 9 筆方言詞／外來語

|    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|
| 阿公 | 阿爸 | 阿嬤 | 便當 | 哈囉 | 阿伯 | 甘願 | 甘心 | 阿母 |
|----|----|----|----|----|----|----|----|----|

### 4.3 小結

本研究以口語對話語料為基礎，以詞頻排列出口語高頻詞表，並以此比對 TOCFL 詞表內容，自對比結果所得未收錄於 TOCFL 詞表之 713 筆口語高頻詞彙，提列出可作為 TOCFL 詞表增修之口語建議詞彙。本研究首先自比對結果中，提出詞頻超過 60 頻次的前 238 筆口語高頻詞彙作為給 TOCFL 詞表未來編纂參考詞彙，期藉以提升詞表的多元性。

為了提出具口語特色之高頻詞彙，本研究另自前人文獻所整合出的 6 項口語詞彙特色將 713 筆詞彙進行分類，發現口語詞彙較多以詞塊型式合成詞呈現，有超過半數的 410 筆口語高頻詞為詞塊型式合成詞。正如張文賢、路雲、李曉琪（2012）提到的口語詞彙當中合成詞較單純詞多。而詞塊型式合成詞中又以動補結構詞塊型式合成詞最多，有 110 筆，例如「說清楚」、「很多」等。

眾多學者多贊同詞塊對詞彙教學的重要性，如：Willis（1990）提出基於「詞塊」的語言教學方法可幫助學習者更好地促進運用語言的能力。Nattinger 與 Decarric（1992）亦詳盡論述了詞塊教學法可以更好地認識和產出不同文體且合乎規範的語篇結構和框架。錢旭菁（2008）進一步聲明詞塊的使用確保了語言的自然性，也提

供道地的語言，並且具社會交際功能。此外，詞塊型式亦常出現於口語交際中，同時也影響著口語能力的流利度（Pawley & Syder 1983；胡元江 2011）。Krishnamurthy（2003）更是提到語言教學已從語法教學轉向詞彙教學，並進而朝向詞塊教學。而張文賢、路雲和李曉琪（2012）對於詞表是否應收錄詞塊也提出相關建議，認為如果兩個詞常常在一起出現，並且這兩個詞組合一起又意義確定，或者具有一定的交際功能，建議將之作爲詞塊出現在詞表中。然而，在 TOCFL 詞表中卻僅有少部分的詞塊型式，建議可增加高頻常用詞塊，可協助提升學習者在語言使用中的準確性及流利度。

熟語亦屬於詞塊型式，爲口頭交際中約定熟成之語言形式，在口語中常見慣用語、歇後語、諺語等形式，常具備特殊文化涵義，在口語交際主要用語中亦較常出現（楊俊萱 1984；常敬宇 1986；李如龍 2007；謝智香 2011）。這部分也在本研究的成果中得到印證。在 713 筆口語高頻詞當中，有 51 筆多音節熟語，詞彙數量僅次於詞塊型式合成詞，居 6 項口語詞彙特色中的第二大類。學習者若可多方理解並接觸母語者常用多音節熟語，可提高日常交際語言中的文化內涵，應有助於擴展其語言表達及交際能力更貼近母語者交際語言。

除了詞塊型式合成詞數量最多以外，本研究亦發現以「不」及「沒」的組合型式之詞彙頻率多半較高，如「不要」有 3,772 頻次、「不要」有 2,322 頻次、「沒事」有 939 頻次。呼應了張文賢、路雲和李曉琪（2012）的論點：以「不」開頭的詞在口語高頻詞中數量較多。且因以「不」及「沒」衍生出的詞塊合成詞通常表示拒絕或否定之意，若將肯定及否定形式詞彙結合可作爲問句使用，例如：「不要」、「不是」、「不會」、「不能」、「不行」、「沒事」、「沒問題」、「有沒有」、「會不會」、「知不知道」等。然而 TOCFL 詞表中少見以「不」及「沒」組合成的否定或問句詞塊合成詞。倘若學習者可多掌握這類詞彙，可增加其交際時語言表達之運用能力。

## 5. 結語

詞表常被視爲測驗學習成果及能力的依據，也是語言教與學的核心內容。學習者實際在語言使用時，除了書面語詞彙，口語詞彙亦不可或缺。透過口語語料分析，以詞彙的使用頻率列表，可提供教師、學習者及教材編輯人員等較爲客觀、更符合母語者口語使用情形的選詞建議。本研究自口語語料高頻詞彙與 TOCFL 詞表對比中發現口語高頻詞中有 713 筆詞彙不在 TOCFL 詞表中。本研究進一步觀察這些口語高頻詞彙，發現絕大部分詞彙多爲「詞塊型式」、「多音節熟語」或是以「不」或

「沒」開頭或組合的詞塊。這類的詞彙在口語使用上較常出現，雖然 TOCFL 詞表收詞多以詞為最小單位，僅有少部分的詞塊型式合成詞或熟詞，但對於高頻常用之詞塊型式合成詞若可斟酌納入詞表內，相信應可提供詞表使用之教師及學習者更貼近母語者語言使用之詞彙內容，亦可較為平衡詞表中口語及書面語詞彙涵蓋量。

本研究仍有許多侷限及受限之處需加以改進，例如：母語者口語語料取自於臺灣本地所製作的電視劇及電影中對白，因此所取用的口語語料內容不僅受限於戲劇對話型式及劇情特色，且亦出現臺灣地方方言詞彙、專有名詞或地名，例如「阿公」、「甘心」、「跆拳道」、「高雄」等。但因口語語料收集及轉寫誠屬不易，需大量的金錢、時間及人力。本研究自華語電影、電視劇對白共蒐集約 150 萬 6 千字的口語語料，相較於現今臺灣華語口語語料庫（「政大漢語口語語料庫」約有 15 萬多字；中研院的「現代漢語語音語料庫」約有 40 萬多個詞）的規模屬較為大型口語語料庫。本研究雖不及由著名出版社出資建置的大型英語口語語料庫之規模（如 CANCODE 語料庫，目前已收錄 500 萬字的口語語料；或是 BNC 語料庫，已具 1 千萬字以上的語料內容），但期許未來可參照 CANCODE 及 BNC 語料庫，多方面蒐集華語口語語料，例如對話式節目訪談內容、或是母語者在日常口語交談、寒暄等。一方面可作為分析母語口語交際特點及語用情形，作為口語教材參照使用。另可作為母語者及學習者在口語字詞、語句及語用上對比分析之參考。

本研究主要以詞頻作為建議詞彙參考依據，未來亦將進一步針對詞彙功能及語用形式進一步更深入探討。儘管研究仍有許多不足之處，但期望本研究成果對 TOCFL 詞表在收詞及應用上之建議，能讓詞表能更多元且更貼近學習者的需求，並提供華語教學更貼近口語交際使用之詞彙參考。

## 引用文獻

- Berber-Sardinha, T. 2000. Comparing corpora with WordSmith Tools: How large must the reference corpus be? *Proceedings of the workshop on Comparing corpora-Volume 9*:7-13. Association for Computational Linguistics.
- Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press
- Biber, Douglas, and Edward Finegan. 1991. *On the Exploitation of Computerized Corpora in Variation Studies*. eds, by K. Aijmer and B. Altenberg, English Corpus Linguistics. London: Longman.

- Biber, Douglas., Susan Conrad, and Viviana Cortes. 2004. If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics* 25.3: 371-405.
- Carter, Ronald. 2004. *Language and Creativity: The Art of Common Talk*. London: Routledge.
- Carter, Ronald, Michael McCarthy, and Rebecca Hughes. 2002. *Exploring Grammar in Context: Upper-intermediate and Advanced*. Ernst Klett Sprachen.
- Chen, Keh-Jiann, and Ming-Hong Bai. 1998. Unknown Word Detection for Chinese by a Corpus-based Learning Method. *International Journal of Computational linguistics and Chinese Language Processing* 3.1:27-44.
- Danescu-Niculescu-Mizil, Cristian, and Lillian Lee. 2011. Chameleons in Imagined Conversations: A new Approach to Understanding Coordination of Linguistic Style in Dialogs. *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, 76-87. Association for Computational Linguistics.
- Davies, Mark, and Dee Gardner. 2013. *A Frequency Dictionary of American English: Word Sketches, Collocates and Thematic Lists*. Routledge.
- Greenbaum, Sidney, and Jan Svartvik. 1990. *The London-Lund Corpus of Spoken English Vol. 7*. Lund University Press.
- Halliday, Michael. A. K. 1989. *Spoken and Written Language*. Oxford University Press.
- Halliday, Michael. A. K., Gibbons, John, and Nicholas, Howard (eds.). 1990. *Learning, Keeping and Using Language: Selected papers from the Eighth World Congress of Applied Linguistics, Sydney, 16-21 August 1987*. Vol. 2. John Benjamins Publishing.
- Kilgariff, Adam. 1997. I don't believe in word senses. *Computers and the Humanities* 31.2: 91-113.
- Kilgariff, Adam, Rychly Pavel, Smrz Pavel, and Tugwell David. 2004. The Sketch Engine. *Proc. EURALEX*, 105-116. Lorient. France.
- Krishnamurthy, Ramesh. 2003. Language as Chunks, Not Words. *JALT2003 Proceedings. Tokyo, Japan, JALT*, 288-294.
- McCarthy, Michael, and Ronald Carter. 2001. Size isn't everything: Spoken English, Corpus, and the Classroom. *Tesol Quarterly* 35.2: 337-340.
- McCarthy, Michael, and Felicity O'Dell. 2003. *English Vocabulary in Use. Advanced*. Cambridge: Cambridge University Press

- McCarthy, Michael, and Felicity O'Dell. 2003. *English Idioms in Use. Intermediate to Upper-intermediate. with Answers*. Cambridge: Cambridge University Press
- McCarthy, Michael. 2004. *Touchstone: From corpus to course book*. Cambridge University Press.
- Nation, Paul, and Robert Waring. 1997. Vocabulary Size, Text Coverage and Word Lists. *Vocabulary: Description, Acquisition, Pedagogy* 14:6-19.
- Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge University Press.
- Nation, I. S. P. 2004. A Study of the Most Frequent Word Families in the British National Corpus. *Vocabulary in a Second Language: Selection, Acquisition, and Testing* 3-13.
- Nattinger, James R., and DeCarrico. Jeanette. S. 1992. *Lexical Phrases and Language Teaching*. Oxford University Press.
- Pawley, Andrew, and Frances H. Syder 1983. Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. *Language and communication*: 191, 225.
- Phillips, Martin. 1989. *Lexical Structure of Text*. No. 12. English Language Research.
- Saussure, Ferdinand de. 1959. *Course in General Linguistics*, edited by Charles Bally and Albert Sechehaye, in collaboration with Albert Riedlinger. Translated by Wade Baskin. New York: Philosophical Library.
- Shirato, J. 2005. A Corpus-based Analysis of Basic Spoken Vocabulary in EFL Textbook Conversations. *Hokusei Gakuen University Graduate School Literature Review* 2: 15-31.
- Willis, Dave. 1990. *The Lexical Syllabus: A New Approach to the Language Teaching*. Collins ELT.
- Wray, Alison. 2000. Formulaic Sequences in Second Language Teaching: Principle and Practice. *Applied linguistics* 21.4:463-489.
- Wray, Alison. 2002. Dual Processing in Protolanguage: Performance Without Competence. *The transition to language*, 113-137.
- 尹惠貞。2006。〈現代漢語口語詞彙研究〉。北京語言大學碩士論文。[Yin, Hui-zhen. 2006. A Study on Lexicon in Colloquial Style in Xiandai Hanyu. Beijing Language and Culture University MA Thesis]
- 王芳智（編）。1990。《漢語口語學》。山西教育出版社。[Wang, Fang-zhi (ed). 1990. *Hanyu Kouyu Xue*. Shanxi Educational Pub. Co.]



- 王希杰。1991。《語言學百題》。上海：上海教育出版社。[Wang, Xi-jie. 1991. *Yuyanxue baiti*. Shanghai: Shanghai Educational Press]
- 朱亞軍、田宇。2000。〈現代漢語詞綴的性質及其分類研究〉，《學術交流》，第 2 期，134-137。[Zhu, Ya-jun and Tian-Yu. 2000. The Study on the Affixation Characteristics and its Classifications of Modern Chinese. *Academic Exchange*. 2:134-137.]
- 朱慶明。2005。《現代漢語實用語法分析》。北京：清華大學出版社。[Zhu, Qing-ming. 2005. *Modern Chines Practical Grammar Analysis*. Beijing: Qing Hua University Press.]
- 李如龍、吳茗。2005。〈略論對外漢語辭彙教學的兩個原則〉，《語言教學與研究》，第 2 期，41-47。[Li, Ru-long, and Ming Wu. 2005. On the Two Principles of Vocabulary Teaching. *Journal of Language Teaching and Research* 2:41-47]
- 李如龍。2007。〈關注漢語口語詞彙與書面語詞匯的研究〉，《陝西師範大學學報》，第 36 卷第 2 期，110-116。[Li, Ru-long. 2007. Concern to be Shown with Study of Oral and Written Chinese Vocabulary. *Journal of Shaanxi Normal University* 36.2:110-116]
- 李世文、陳秋梅。1993。〈中文口語與書寫語的比較研究〉，《教學與研究》，第 15 期，63-96。[Li, Shi-wen, and Qiu-mei Chen. 1993. Comparative Research of Chinese Spoken and Written Language. *Teaching and Study* 15:63-96.]
- 宋婧婧。2013。〈漢語口語與書面語詞彙使用對比分析-基於傳媒語料庫〉，《廈門理工學院學報》，第 21 卷第 3 期，88-92。[Song, Jing-jing. 2013. A Media-Corpus-Based Comparative Study on the Use of Vocabulary in Spoken and Written Chinese. *Journal of Xiamen University of Technology* 21.3:88-92.]
- 車豔秋、楊虹、曹明。2013。〈外語焦慮與大學英語口語教學設計〉，《遼寧行政學院學報》，第 11 期，81-83。[Che, Yan-qiu, Hong Yang, and Ming Cao. 2013. Foreign Language Anxiety and Oral English Teaching Design for College Students. *Journal of Liaoning Administration College* 11:81-83]
- 吳麗君。2004。〈口語詞彙與書面語詞匯教學研究〉，《雲南師範大學學報（對外漢語教學與研究版）》，第 3 期，14-19。[Wu, Li-jun. 2004. The Study of Vocabulary Teaching in the Spoken and Written Chinese Language. *Journal of Yunnan Normal University: Teaching and Research on Chinese as a Foreign Language Edition*. 3:14-19.]

- 周祖謨。1959。《漢語詞彙講話》，人民教育出版社。[Zhou, Zu-mo. 1959. *Hanyu Cihui Jianghua*. People's Education Press.]
- 胡元江。2011。〈口語產出中的詞塊研究：回顧與展望〉，《外語教學理論與實踐》，第2期，55-63。[Hu, Yuan-jiang. 2011. Studies on the Formulaic Sequences in Speech Production: Retrospect and Prospect. *Foreign Language Learning Theory and Practice* 2:55-63.]
- 高名凱、石安石。2002。《語言學概論》。北京：中華書局。[Gao, Ming-Kai and An-Shi Shi. 2002. *Yuyuanxue Gailun*. Beijing: ZhongHua Book Company]
- 徐素萍。2012。〈淺議建立與現代漢語課程配套的口語語料庫的意義〉，《長沙大學學報》，第26卷第1期，125-126。[Xu, Su-ping. 2012. Discuss in the Value of Establishing Oral Corpus Supporting Modern Chinese Course. *Journal of Changsha University* 26.1:125-126]
- 徐立人。2011。〈基於全球資訊網的華語文詞彙語法誘答選項正答力計算方法〉。中央大學網路學習科技研究所學位論文。[Xu, Li-Ren. 2011. A Web-based Method for Computing the Correctness of Distractors in Chinese Multiple Choice Cloze. *National Central University Electronic Theses & Dissertations*.]
- 徐海美。2009。〈詞彙組塊在二語習得的實踐分析〉，《吉林工程技術師範學院學報》，第25卷第1期，44-46。[Xu, Hai-mei. 2009. Practical Analysis of Applying Lexical Chunk to Second Language Acquisition. *Journal of Jilin Teachers Institute of Engineering and Technology* 25.1:44-46.]
- 張金蘭。2012。〈對外華語文教學中的書面語教學——以臺灣華語文教材為主所做的探討〉。政治大學華語文教學中心「錦華工作坊」。[Zhang, Jin-lan. 2012. Teaching in Written Chinese as a Foreign Language-Base on Teaching Materials of Chinese as Second Language in Taiwan. *Chinese Language Center JinHua Workshop of National Chengchi University*.]
- 張莉萍、陳鳳儀。2005。〈華語詞彙分級初探〉，《第六屆漢語詞彙語義學研討會論文集》，250-260。[Chang, 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。[Chang, Li-ping. 2007. The development and status of Test

- of Proficiency-Huayu. *Proceedings of the International Conference of Foreign Language Proficiency Tests Trends*, 185-195.]
- 張文賢、路雲、李曉琪。2012。〈基於口語語料庫的漢語口語自動化考試詞表的研製〉，*《中文教學現代化學報》*，第 1 卷第 2 期，37-44。[Zhang, Wen-xian, Yun Lu, and Xiao-qi Li. 2012. A Corpus-Based Vocabulary Lists Study For Automated Test of Spoken Chinese. *Journal of Modernization of Chinese Language Education* 1.2:37-44]
- 常敬宇。1986。〈漢語口語詞彙的特點〉，《邏輯與語言學習》，第 4 期，36-37。[Chang, Jing-yu. 1986. Features of Chinese Spoken vocabulary. *Logic and Language Learning* 4:36-37.]
- 陳露、韋漢。2005。〈英語口語語料庫在英語口語教學中的作用〉，《外語電化教學》，第 103 期，23-26。[Chen, Lu, and Wei Han. 2005. The Role of Spoken English Corpora in Spoken English Teaching. *Journal of Computer-Assisted Foreign Language Education* 103:23-26.]
- 曹煒。2003。〈現代漢語口語詞和書面語詞的差異初探〉，《語言教學與研究》，第 6 期，39-44。[Cao, Wei. 2003. A Preliminary Study on the Differences between the Colloquial and the Literary in Modern Chinese. *Language Teaching and Linguistic Studies* 6:39-44.]
- 曹合建（編）。2008。《基於語料庫的商務英語研究》。北京：對外經濟貿易大學出版社。[Cao, He-jian (ed). 2008. *A Corpus-Driven Study of Business English*. Beijing: University of International Business and Economics Press]
- 葉蜚聲、徐通鏘。1993。《語言學綱要》。臺北市：書林。[Ye, Fei-sheng, and Tong-qiang Xu. 1993. *An Outline of Linguistics*. Taipei: Bookman Bookstore]
- 楊俊萱。1984。〈口語和書面語〉，《語言教學與研究》，第 1 期，137-146。[Yang, Jun-xuan. 1984. Spoken and Written Language. *Language Teaching and Linguistic Studies* 1:137-146.]
- 楊惠媚、陳浩然、潘依婷。2014。〈兩岸華語詞表之比較及選詞建議〉，《華語文教學研究》，第 11 卷第 1 期，67-98。[Yang, Hui-mei, Hao-jan H. Chen, and I-ting Pan. 2014. Comparison of TOCFL and HSK Wordlists and Suggestions on Word-selection. *World Chinese Language Association* 11.1:67-98.]

- 翟瑩。2012。〈淺談幼師英語詞塊教學〉，《成功（教育）》，第5期，104-104。  
[Di-Ying, 2012, Discuss in the Teaching of English Chunk, *Success: Education* 5:104-104.]
- 劉玫芳。2013。〈字詞搭配教學對臺灣國中學生字詞搭配能力發展之效益研究〉。  
國立高雄師範大學碩士論文。[Liu, Mei-fang. 2013. Effects of Collocation Instruction on Junior High School Students' Collocation Competence in Taiwan. National Kaohsiung Normal University MA Thesis]
- 錢旭菁。2008。〈漢語詞塊研究初探〉，《北京大學學報（哲學社會科學版）》，第45卷第5期，139-146。[Qian, Xu-jing. 2009. A Preliminary Study on Chinese Chunk, *Journal of Peking University: Philosophy & Social Sciences* 45.5:139-146.]
- 謝智香。2011。〈論現代漢語口語詞的特點〉，《西南石油大學學報（社會科學版）》，第3期，103-106。[Xie, Zhi-xiang. Characteristics of Colloquial Words of Modern Chinese. *Journal of Southwest Petroleum university: Social Sciences Edition* 3:103-106.]
- 顧安達、朱志平。2005。〈口語和書面語教學目標的衝突與漢語教學的課程改革〉，《海外華文教育》，第2005卷第2期，42-47。[Gu, An-da, and Zhi-ping Zhu. 2005. Instructional Objective Discrepancy between Spoken and Written Language VS Curriculum Reform of TCFL. *Overseas Chinese Education* 2005.2:42-47.]

[審查：2014.7.11 修改：2014.8.19 接受：2015.1.15]

楊惠媚

Hui-Mei YANG

106 臺北市和平東路一段 162 號

Department of English, National Taiwan Normal University

No. 162, Heping East Road, Section 1, Taipei, Taiwan 106

[huimei.yang2009@gmail.com](mailto:huimei.yang2009@gmail.com)

陳浩然

Howard Hao-Jan CHEN

106 臺北市和平東路一段 162 號

Department of English, National Taiwan Normal University

分析華語口語語料庫高頻詞之特點並對 TOCFL 詞表提出建議

No. 162, Heping East Road, Section 1, Taipei, Taiwan 106

[hjchen@ntnu.edu.tw](mailto:hjchen@ntnu.edu.tw)

潘依婷

I-ting PAN

106 臺北市和平東路一段 162 號

Department of English, National Taiwan Normal University

No. 162, Heping East Road, Section 1, Taipei, Taiwan 106

[leopardpan@gmail.com](mailto:leopardpan@gmail.com)

附錄一 口語語料收集來源

| 編號  | 電影/電視劇名稱    | 字數    | 編號  | 電影/電視劇名稱 | 字數    |
|-----|-------------|-------|-----|----------|-------|
| 001 | 雞排英雄        | 11715 | 064 | 幸福的麵包    | 5176  |
| 002 | 非誠勿擾 2      | 10814 | 065 | 桃姊       | 11617 |
| 003 | 蔡李佛         | 13577 | 066 | 陣頭       | 13423 |
| 004 | 帶一片風景走      | 6137  | 067 | 犀利人妻最終回  | 8606  |
| 005 | 命運化妝師       | 7687  | 068 | 愛的麵包魂    | 11856 |
| 006 | 寶島曼波        | 8902  | 069 | 新天生一對    | 8452  |
| 007 | 殺手歐陽盆栽      | 11096 | 070 | 聽風者      | 8349  |
| 008 | 翻滾吧阿信       | 7495  | 071 | 逆光飛翔     | 7923  |
| 009 | 那些年我們一起追的女孩 | 10544 | 072 | 一頁臺北     | 7259  |
| 010 | 燃燒吧歐吉桑      | 10950 | 073 | 人在囧途     | 10420 |
| 011 | 星空          | 4857  | 074 | 全城戒備     | 8735  |
| 012 | 戀愛恐慌症       | 12109 | 075 | 刀見笑      | 5904  |
| 013 | 對不起我愛你      | 6712  | 076 | 全城熱戀     | 8895  |
| 014 | 愛到底         | 8562  | 077 | 初戀紅豆冰    | 7081  |
| 015 | 練戀舞         | 6575  | 078 | 劍雨       | 7449  |
| 016 | 新魯冰花        | 8684  | 079 | 唐山大地震    | 9764  |
| 017 | 亂青春         | 3231  | 080 | 大兵小將     | 5676  |
| 018 | 不能沒有妳       | 7154  | 081 | 孔子       | 10651 |
| 019 | 聽說          | 7006  | 082 | 志明與春嬌    | 13040 |
| 020 | 夏天協奏曲       | 9227  | 083 | 戀愛通告     | 9469  |
| 021 | 帶我去遠方       | 6238  | 084 | 海洋天堂     | 5834  |
| 022 | 變身          | 10964 | 085 | 東風破      | 8848  |
| 023 | 風聲          | 7380  | 086 | 歲月神偷     | 12170 |
| 024 | 刺陵          | 7238  | 087 | 父後七日     | 7963  |
| 025 | 色戒          | 13116 | 088 | 獵艷       | 7875  |
| 026 | 不能說的秘密      | 7240  | 089 | 當愛來的時候   | 9527  |
| 027 | 練習曲         | 8507  | 090 | 第四張畫     | 9479  |
| 028 | 刺青          | 4987  | 091 | 第 36 個故事 | 10625 |
| 029 | 六號出口        | 8555  | 092 | 精武風雲陳真   | 9029  |
| 030 | 沉睡的青春       | 4469  | 093 | 維多利亞一號   | 7225  |
| 031 | 夏天的尾巴       | 5176  | 094 | 美麗密令     | 12814 |
| 032 | 基因決定我愛你     | 8473  | 095 | 鎗王之王     | 12112 |
| 033 | 大尾鱸鰻        | 13294 | 096 | 魚狗       | 6313  |
| 034 | 志氣          | 12567 | 097 | 九降風      | 8840  |

分析華語口語語料庫高頻詞之特點並對 TOCFL 詞表提出建議

| 編號  | 電影/電視劇名稱 | 字數    | 編號  | 電影/電視劇名稱         | 字數     |
|-----|----------|-------|-----|------------------|--------|
| 035 | 富春山居圖    | 6700  | 098 | 功夫灌籃             | 8206   |
| 036 | 聖誕玫瑰     | 12331 | 099 | 冏男孩              | 9779   |
| 037 | 不二神探     | 9226  | 100 | 長江七號             | 5792   |
| 038 | 盲探       | 12697 | 101 | 非誠勿擾             | 14295  |
| 039 | 長恨歌      | 9609  | 102 | 保持通話             | 10098  |
| 040 | 生日快樂     | 9400  | 103 | 海角七號             | 11899  |
| 041 | 門徒       | 9614  | 104 | 絕代雙嬌             | 21088  |
| 042 | 早熟       | 15170 | 105 | 幫幫我愛神            | 4281   |
| 043 | 頭文字 D    | 8751  | 106 | 鬪茶               | 7600   |
| 044 | 七劍       | 9902  | 107 | 臺北異想             | 4919   |
| 045 | 黑社會      | 8335  | 108 | 國士無雙             | 12337  |
| 046 | 再說一次我愛你  | 7288  | 109 | 一年之初             | 6711   |
| 047 | 殺破狼      | 4056  | 110 | 盛夏光年             | 6411   |
| 048 | 如果愛      | 7501  | 111 | 詭絲               | 6574   |
| 049 | 千機變      | 6393  | 112 | 龍虎門              | 5653   |
| 050 | 戀上你的床    | 10860 | 113 | 天生一對             | 13036  |
| 051 | 向左走向右走   | 7548  | 114 | 世界第一麥方           | 9350   |
| 052 | 大隻佬      | 5766  | 115 | 聖誕玫瑰             | 12335  |
| 053 | 無間道 2    | 8394  | 116 | 花樣年華             | 7867   |
| 054 | 尋找周杰倫    | 5895  | 117 | 傷城               | 7546   |
| 055 | 無間道 3    | 9478  | 118 | 功夫               | 6344   |
| 056 | 單身男女     | 7992  | 119 | 2046             | 11405  |
| 057 | 大笑江湖     | 8543  | 120 | 痞子英雄電影版          | 14905  |
| 058 | 一路向西     | 16935 | 121 | 光陰的故事(全套 107 集)  | 614945 |
| 059 | 十二生肖     | 15115 | 122 | 我在墾丁天氣晴(全套 20 集) | 98709  |
| 060 | 女朋友男朋友   | 8245  | 123 | 熱血青春(全套 20 集)    | 111852 |
| 061 | 太極 1     | 6065  | 124 | 赴宴(全套 22 集)      | 86687  |
| 062 | 太極 2     | 6874  | 125 | 痞子英雄(全套 24 集)    | 87522  |
| 063 | 臺北飄雪     | 8051  |     |                  |        |

附錄二 713 筆不在 TOCFL 詞表的口語高頻詞 (詞彙下方數字為詞頻)

|            |            |            |            |            |            |             |            |             |             |            |
|------------|------------|------------|------------|------------|------------|-------------|------------|-------------|-------------|------------|
| 不要<br>3772 | 不是<br>2322 | 耶<br>2124  | 不會<br>2040 | 一下<br>1957 | 回來<br>1312 | 好不好<br>1176 | 不能<br>1108 | 沒事<br>939   | 看到<br>932   | 有沒有<br>871 |
| 等一下<br>703 | 怎樣<br>673  | 想要<br>608  | 阿公<br>592  | 不行<br>538  | 阿爸<br>529  | 快點<br>527   | 要不要<br>524 | 囉<br>510    | 很多<br>500   | 回家<br>488  |
| 過來<br>459  | 走了<br>418  | 小小<br>415  | 吃飯<br>413  | 在一起<br>386 | 找到<br>345  | 聽到<br>342   | 都是<br>335  | 來了<br>327   | 會不會<br>319  | 跆拳道<br>297 |
| 不好<br>291  | 那麼多<br>290 | 談戀愛<br>290 | 跆拳道<br>277 | 女朋友<br>272 | 跆拳道<br>265 | 這麼多<br>259  | 不對<br>256  | 可不可以<br>249 | 知不知道<br>243 | 社長<br>241  |
| 很好<br>237  | 臺北<br>235  | 不可以<br>233 | 好的<br>229  | 笨蛋<br>226  | 不可能<br>225 | 黑客<br>216   | 咧<br>204   | 有事<br>204   | 一輩子<br>202  | 阿嬤<br>202  |
| 沒錯<br>200  | 去了<br>200  | 老爸<br>199  | 村子<br>195  | 唸書<br>194  | 死了<br>193  | 金牌<br>193   | 忘了<br>189  | 大姊<br>187   | 上面<br>184   | 墾丁<br>178  |
| 爸媽<br>175  | 好多<br>172  | 太多<br>170  | 到時候<br>170 | 後面<br>164  | 你好<br>164  | 前面<br>159   | 小孩子<br>159 | 男朋友<br>158  | 拿去<br>156   | 有人<br>156  |
| 你看<br>155  | 沒辦法<br>152 | 沒問題<br>152 | 還不是<br>149 | 跑去<br>148  | 想想<br>147  | 家裡<br>146   | 麵攤<br>144  | 交給<br>142   | 總動員<br>141  | 不准<br>138  |
| 學長<br>136  | 說了<br>133  | 女孩<br>132  | 叫做<br>131  | 局長<br>131  | 小時候<br>130 | 分局<br>129   | 心裡<br>125  | 通通<br>124   | 女孩子<br>121  | 還給<br>119  |
| 年輕人<br>118 | 怎麼說<br>118 | 會長<br>118  | 想說<br>117  | 不見了<br>117 | 過了<br>116  | 考上<br>115   | 們<br>112   | 哭了<br>110   | 上來<br>109   | 唸<br>109   |
| 越來越<br>107 | 打了<br>106  | 做了<br>106  | 組長<br>105  | 收到<br>104  | 超<br>104   | 變得<br>101   | 牛肉麵<br>101 | 走開<br>100   | 說清楚<br>100  | 過分<br>99   |
| 上去<br>97   | 賺錢<br>97   | 多久<br>97   | 試試看<br>96  | 好人<br>95   | 能不能<br>94  | 還可以<br>94   | 打給<br>93   | 上車<br>93    | 放在<br>93    | 嫁給<br>92   |
| 參賽<br>91   | 帶走<br>88   | 自個兒<br>87  | 不在<br>87   | 在家<br>87   | 好久<br>87   | 好不容易<br>87  | 回到<br>86   | 喝酒<br>86    | 聯絡<br>86    | 好朋友<br>85  |
| 便當<br>85   | 韓國<br>85   | 準備好<br>84  | 拿到<br>84   | 搞不好<br>84  | 雞婆<br>84   | 閉嘴<br>83    | 不只<br>83   | 開門<br>82    | 沒用<br>82    | 跑到<br>82   |
| 拿來<br>82   | 都在<br>82   | 打贏<br>81   | 不應該<br>81  | 聽話<br>81   | 嚇死<br>80   | 南區<br>79    | 到了<br>79   | 寫信<br>79    | 好玩<br>79    | 跑來<br>79   |
| 姊夫<br>79   | 多了<br>78   | 送給<br>78   | 白癡<br>78   | 馬子<br>78   | 乖乖<br>77   | 說過<br>77    | 壞人<br>77   | 留下來<br>77   | 不理<br>77    | 衝浪<br>76   |
| 都有<br>75   | 高雄<br>75   | 胡說<br>75   | 想過<br>74   | 站在<br>74   | 當成<br>74   | 小聲<br>74    | 不停<br>74   | 吃了<br>73    | 打死<br>73    | 放開<br>72   |
| 長官<br>72   | 一邊<br>71   | 氣死<br>71   | 偏偏<br>70   | 留下<br>70   | 做生意<br>70  | 待在<br>70    | 真相<br>70   | 對手<br>70    | 有夠<br>70    | 很久<br>69   |
| 吃完<br>68   | 不可<br>68   | 說真的<br>68  | 留在<br>67   | 路上<br>67   | 混蛋<br>67   | 見到<br>66    | 做錯<br>66   | 神經病<br>66   | 坐下<br>65    | 還在<br>65   |
| 來到<br>65   | 放下<br>65   | 打敗<br>65   | 出賽<br>65   | 拿給<br>65   | 長得<br>65   | 好聽<br>65    | 好久不見<br>65 | 夥伴<br>65    | 騙人<br>65    | 董事長<br>65  |



分析華語口語語料庫高頻詞之特點並對 TOCFL 詞表提出建議

|      |     |     |     |     |     |      |      |     |      |     |
|------|-----|-----|-----|-----|-----|------|------|-----|------|-----|
| 64   | 64  | 64  | 64  | 64  | 63  | 63   | 63   | 63  | 62   | 62  |
| 總務長  | 老頭  | 做好  | 走到  | 退學  | 噁心  | 帶著   | 生日快樂 | 等等  | 一塊兒  | 沒有用 |
| 62   | 61  | 61  | 61  | 61  | 61  | 61   | 61   | 60  | 60   | 60  |
| 說出來  | 早餐  | 善良  | 廁所  | 才是  | 十幾  | 一家人  | 出國   | 拿出來 | 哈囉   | 做到  |
| 60   | 60  | 60  | 60  | 60  | 60  | 59   | 59   | 59  | 59   | 58  |
| 莫名其妙 | 好棒  | 坐下來 | 拿著  | 龜   | 行不行 | 夠了   | 少爺   | 團長  | 一點點  | 讓開  |
| 58   | 57  | 57  | 57  | 57  | 56  | 56   | 56   | 56  | 55   | 55  |
| 晚安   | 蠻   | 一路  | 完蛋  | 愛上  | 留給  | 行了   | 有一點  | 也有  | 香港   | 來過  |
| 55   | 55  | 55  | 55  | 54  | 54  | 54   | 54   | 53  | 53   | 52  |
| 您好   | 殺人  | 打來  | 說到  | 無敵  | 上海  | 隊長   | 臺大   | 住在  | 騎車   | 對不對 |
| 52   | 52  | 52  | 52  | 52  | 52  | 52   | 52   | 51  | 51   | 51  |
| 抓到   | 一陣子 | 老了  | 之外  | 妳好  | 出了  | 完了   | 你們倆  | 警官  | 聽聽   | 不夠  |
| 51   | 51  | 50  | 50  | 50  | 50  | 50   | 50   | 50  | 49   | 49  |
| 站起來  | 外國人 | 村口  | 囉唆  | 好事  | 弄得  | 鑑識   | 屁    | 瞧瞧  | 打到   | 大了  |
| 49   | 49  | 49  | 48  | 48  | 48  | 48   | 48   | 47  | 47   | 47  |
| 買的   | 錯了  | 更何況 | 亂講  | 不一定 | 聽過  | 解散   | 想起   | 吃醋  | 兇手   | 社員  |
| 47   | 47  | 47  | 46  | 46  | 46  | 46   | 46   | 46  | 46   | 46  |
| 招    | 睡了  | 看清楚 | 講到  | 拚   | 警察局 | 問問   | 說說   | 教我  | 看不出來 | 沒人  |
| 46   | 45  | 45  | 44  | 44  | 44  | 43   | 43   | 43  | 43   | 43  |
| 計劃   | 啥   | 鬧了  | 不敢  | 喝醉  | 借過  | 買了   | 死人   | 才有  | 屁事   | 今兒  |
| 43   | 42  | 42  | 42  | 42  | 42  | 42   | 42   | 42  | 42   | 41  |
| 事兒   | 哈   | 坐好  | 輸了  | 美女  | 想吃  | 亂七八糟 | 線索   | 借我  | 記不記得 | 阿伯  |
| 41   | 41  | 41  | 41  | 41  | 41  | 41   | 41   | 40  | 40   | 40  |
| 開店   | 上了  | 拚命  | 亂說  | 帥哥  | 聽清楚 | 外人   | 不懂   | 出走  | 撞到   | 年前  |
| 40   | 39  | 39  | 39  | 39  | 39  | 39   | 39   | 39  | 39   | 39  |
| 不對勁  | 頂嘴  | 拿出  | 死掉  | 偷看  | 沒了  | 下車   | 有了   | 地上  | 牛肉湯  | 王子  |
| 39   | 38  | 38  | 38  | 38  | 38  | 38   | 38   | 38  | 38   | 38  |
| 檢察官  | 臭小子 | 莫   | 晚了  | 開槍  | 喝茶  | 花錢   | 很忙   | 見過  | 最愛   | 很像  |
| 38   | 38  | 38  | 37  | 37  | 37  | 37   | 37   | 37  | 37   | 37  |
| 餵    | 警方  | 牛肉  | 黑道  | 條子  | 哈哈  | 手上   | 帶回來  | 喝水  | 更多   | 看醫生 |
| 37   | 37  | 37  | 37  | 36  | 36  | 36   | 36   | 36  | 36   | 36  |
| 好心   | 看完  | 喝了  | 生了  | 很難  | 找了  | 有出息  | 法國   | 線報  | 拿了   | 很棒  |
| 36   | 36  | 36  | 36  | 36  | 36  | 36   | 36   | 36  | 35   | 35  |
| 沒空   | 聽得懂 | 贏了  | 很大  | 不怕  | 親愛的 | 負責任  | 三更半夜 | 雞蛋  | 爽    | 就有  |
| 35   | 35  | 35  | 35  | 35  | 35  | 35   | 35   | 35  | 35   | 35  |
| 就在   | 輕輕  | 拿走  | 眼前  | 好歹  | 找人  | 日本人  | 剛剛好  | 大夥兒 | 懂了   | 忙著  |
| 35   | 34  | 34  | 34  | 34  | 34  | 34   | 33   | 33  | 33   | 33  |
| 好笑   | 打完  | 離家  | 所有人 | 清白  | 自卑  | 男子漢  | 聯考   | 唉唷  | 很愛   | 好想  |
| 33   | 33  | 33  | 33  | 33  | 33  | 33   | 33   | 32  | 32   | 32  |
| 丟掉   | 停下來 | 只能  | 毀了  | 無關  | 詛咒  | 不要臉  | 喜歡上  | 男孩  | 媽咪   | 公主  |
| 32   | 32  | 32  | 32  | 32  | 32  | 32   | 32   | 32  | 32   | 32  |

|            |           |           |           |               |               |           |           |            |            |           |
|------------|-----------|-----------|-----------|---------------|---------------|-----------|-----------|------------|------------|-----------|
| 逞強<br>32   | 白痴<br>32  | 呐<br>31   | 躲在<br>31  | 今晚<br>31      | 跑掉<br>31      | 去哪<br>31  | 拿過來<br>31 | 有的時候<br>31 | 胡思亂想<br>31 | 眼神<br>31  |
| 臺中<br>31   | 絕招<br>31  | 聊聊<br>30  | 唄<br>30   | 掰<br>30       | 減肥<br>30      | 不要緊<br>30 | 打人<br>30  | 搞錯<br>30   | 害死<br>30   | 解開<br>30  |
| 說完<br>30   | 看好<br>30  | 對得起<br>30 | 甘願<br>30  | 自首<br>30      | 良心<br>30      | 團隊<br>30  | 尿尿<br>29  | 老外<br>29   | 在忙<br>29   | 殺了<br>29  |
| 想清楚<br>29  | 壞掉<br>29  | 瘋了<br>29  | 下面<br>29  | 搞清楚<br>29     | 壞事<br>29      | 難聽<br>29  | 嚇到<br>29  | 亂來<br>29   | 送去<br>29   | 這陣子<br>29 |
| 從小到大<br>29 | 哥兒們<br>29 | 制服<br>29  | 菩薩<br>29  | 初戀<br>29      | 掩護<br>29      | 神明<br>29  | 哪有<br>29  | 痞子<br>29   | 偶爾<br>29   | 作夢<br>28  |
| 很少<br>28   | 帶去<br>28  | 掉了<br>28  | 男女<br>28  | 急事<br>28      | 不配<br>28      | 帶到<br>28  | 不算<br>28  | 收好<br>28   | 重案組<br>28  | 介意<br>28  |
| 回報<br>28   | 晚飯<br>28  | 跟蹤<br>28  | 阿弟<br>28  | 上場<br>28      | 北區<br>28      | 才剛<br>28  | 讚<br>28   | 噓<br>27    | 那有<br>27   | 又是<br>27  |
| 忙了<br>27   | 查到<br>27  | 看錯<br>27  | 壞了<br>27  | 講了<br>27      | 站好<br>27      | 活著<br>27  | 好喝<br>27  | 一早<br>27   | 找出<br>27   | 坐著<br>27  |
| 生下來<br>27  | 想通<br>27  | 重來<br>27  | 自己人<br>27 | 幾十<br>27      | 中華<br>27      | 聊聊天<br>26 | 說出<br>26  | 累了<br>26   | 躲起來<br>26  | 搞定<br>26  |
| 帶回去<br>26  | 闖禍<br>26  | 吵醒<br>26  | 好痛<br>26  | 一 模 一 樣<br>26 | 比如說<br>26     | 一家子<br>26 | 大事<br>26  | 義氣<br>26   | 變態<br>26   | 蠢<br>26   |
| 豬肉<br>26   | 水溝<br>26  | 懲罰<br>26  | 才對<br>26  | 大小姐<br>26     | 大一<br>26      | 簡訊<br>26  | 傻傻<br>25  | 男孩子<br>25  | 明兒<br>25   | 吼<br>25   |
| 想好<br>25   | 被人<br>25  | 讓人<br>25  | 笑死<br>25  | 讓給<br>25      | 吃點<br>25      | 煮飯<br>25  | 通電話<br>25 | 停車<br>25   | 急著<br>25   | 打球<br>25  |
| 抱著<br>25   | 丟下<br>25  | 太大<br>25  | 幹活<br>25  | 上廁所<br>25     | 甘心<br>25      | 發呆<br>25  | 囂張<br>25  | 上山<br>25   | 自責<br>25   | 才會<br>25  |
| 攤<br>25    | 老媽<br>24  | 老太婆<br>24 | 求你<br>24  | 收下<br>24      | 累死<br>24      | 帶了<br>24  | 好意<br>24  | 約好<br>24   | 去過<br>24   | 找不著<br>24 |
| 感人<br>24   | 心願<br>24  | 監視器<br>24 | 啞巴<br>24  | 表哥<br>24      | 店裡<br>24      | 包包<br>24  | 呼叫<br>24  | 叔<br>24    | 開了<br>23   | 吃到<br>23  |
| 想了<br>23   | 下山<br>23  | 沒錢<br>23  | 少了<br>23  | 往後<br>23      | 抽煙<br>23      | 留著<br>23  | 亂動<br>23  | 借錢<br>23   | 做完<br>23   | 真愛<br>23  |
| 搶劫<br>23   | 想去<br>23  | 不想<br>23  | 聽好<br>23  | 就是了<br>23     | 在 一 塊 兒<br>23 | 最大的<br>23 | 今兒個<br>23 | 牌<br>23    | 一般人<br>23  | 高麗菜<br>23 |
| 出氣<br>23   | 立正<br>23  | 嫌棄<br>23  | 請坐<br>23  | 玩玩<br>22      | 男子<br>22      | 想聽<br>22  | 新來<br>22  | 不說<br>22   | 昨晚<br>22   | 做不到<br>22 |
| 搬走<br>22   | 坐車<br>22  | 抓住<br>22  | 送來<br>22  | 並肩<br>22      | 拿回去<br>22     | 多好<br>22  | 慢走<br>22  | 沒出息<br>22  | 再一次<br>22  | 話說<br>22  |
| 照樣<br>22   | 巷子<br>22  | 里長<br>22  | 山上<br>22  | 早安<br>22      | 豬頭<br>22      | 找找<br>21  | 試試<br>21  | 辛辛苦苦<br>21 | 膽子<br>21   | 煩死<br>21  |

分析華語口語語料庫高頻詞之特點並對 TOCFL 詞表提出建議

|          |          |           |           |          |           |           |            |           |            |          |
|----------|----------|-----------|-----------|----------|-----------|-----------|------------|-----------|------------|----------|
| 講好<br>21 | 拍到<br>21 | 罵人<br>21  | 懂不懂<br>21 | 不像<br>21 | 死心<br>21  | 寫了<br>21  | 趁熱<br>21   | 偷吃<br>21  | 看得出來<br>21 | 稀罕<br>21 |
| 不多<br>21 | 存錢<br>21 | 我幫你<br>21 | 怎麼著<br>21 | 阿母<br>21 | 背叛<br>21  | 父母親<br>21 | 大海<br>21   | 鏡頭<br>21  | 勸勸<br>20   | 遠遠<br>20 |
| 唉呦<br>20 | 賣了<br>20 | 跟上<br>20  | 拿錢<br>20  | 更好<br>20 | 寄給<br>20  | 掉下來<br>20 | 喜不喜歡<br>20 | 很痛<br>20  | 賣給<br>20   | 亂丟<br>20 |
| 很快<br>20 | 輪到<br>20 | 關掉<br>20  | 講出來<br>20 | 抽菸<br>20 | 找出來<br>20 | 找麻煩<br>20 | 一通<br>20   | 內心<br>20  | 一人<br>20   | 甜蜜<br>20 |
| 粉絲<br>20 | 嗓門<br>20 | 奧運<br>20  | 晚餐<br>20  | 心痛<br>20 | 窗戶<br>20  | 傻了<br>19  | 真好<br>19   | 感受到<br>19 |            |          |

## **Analyzing the Features of the High-Frequency Words on Chinese Spoken Corpus and Offering the Word-recruiting Suggestion to TOCFL Wordlist**

**Hui-Mei YANG   Howard Hao-Jan CHEN   I-ting PAN**  
**Mandarin Training Center & Department of English,**  
**National Taiwan Normal University**

### **Abstract**

Wordlist serves as a reference for second language teaching, and also guides second language learners to evaluate what words they need to acquire. The TOCFL wordlist is one of the common learning materials for learners preparing for Chinese proficiency test. However, words included in the TOCFL wordlist were largely selected from a written corpus, whereas words extracted from a spoken corpus were limited. Because written and spoken corpora are presumably different, it is necessary to include words in both registers and to emphasize the differences while teaching. To balance the proportions of written and spoken words in the TOCFL wordlists, this study first established a native spoken corpus by extracting subtitles from Mandarin movies and TV series, and then compiled a list of high-frequency spoken words as an amendment to the TOCFL wordlist. Comparison between this spoken wordlist with the TOCFL wordlist showed that the most frequently used 713 words in the corpus were not covered in the TOCFL wordlist. We then suggested a list of the top 238 high-frequency words to be included in the TOCFL wordlist. The 713 high-frequency spoken words were further classified into six groups based on their features, and some key findings were summarized as follows: (1) the majority of the items are word chunks, (2) the spoken words are characterized as multi-syllable words, and (3) there are large numbers of word combinations of *bu* and *mei* in the list. We hope that the provision of this commonly used spoken wordlist can increase the proportion of spoken words in the TOCFL wordlist, which can offer learners more authentic materials to meet their oral communication needs.

**Keywords:** Chinese learning, Chinese wordlist, Spoken corpus, high-frequency words