

# 對外漢語字集

張莉萍

國立台灣師範大學

## 摘要

本論文檢視三個以漢語為外語學習的字集，來探討華語字集的問題。這三個字集，一是漢語水平考試（HSK）大綱的甲級字表 800 個，二是台灣華語文能力測驗(TOP-Huayu)的初級字表（從初等詞彙 1500 個中抽取出來）863 個字，三是法國中文字表 804 個字。前兩個字表都是以詞彙為前提，從常用詞彙分析出來的漢字；法國字表則是字本位。

結果顯示，同時出現在三個字表中的字共 593 個，共現率佔每個字表的百分之七十左右。這 593 個字是不是可以做為優先學習的字集，需要進一步探討。在初步觀察同時，也檢視這些字表中是否包括黃沛榮(2001)提出的 72 個具有優先學習價值的整字部件/部首，大部分出現在共同字中，TOP 字表中涵蓋量較其他兩個字表多，但以下 18 個具有優先學習價值的整字部首/部件並沒有出現在三個字表中的任何一個：「巾、田、竹、寸、弓、戈、欠、止、犬、皿、禾、穴、羽、舟、貝、邑、阜、虫」。

本文亦探討對外漢語字集的功能與內涵，提出對應於 CEFR 制訂字彙量的建議，A2-500 字；B1-900 字；B2-1800 字；C1-3000 字。

關鍵字：對外漢語、漢字、字集、優先學習價值

## 一、研究動機

今年（民國 97 年）國家華語測驗推動工作委員會（華測會）受教育部委託執行華語文能力指標計畫，這個計畫的緣起是去年教育部要華測會評估「歐洲語言學習、教學、評量共同參考架構(CEFR)」與華語文能力測驗(TOP)接軌問題。近年來，國際上重要的能力考試皆致力於對應於這個六級能力架構(A1/ A2/ B1/ B2/ C1/ C2)，除了劍橋大學研發的各種英檢考試、日本語能力考試、漢語水平考試外，歐洲其他語言，例如，法語檢定考試(DELF)成立於 1985 年，自 2005 年 9 月 1 日起，該考試配合 CEFR，將考試級數重新規劃。之後，各種法語檢定考試皆對應於 CEFR 所制訂的六級語言程度，每一級數包含聽、說、讀、寫四種能力

測驗。在這個接軌評估中，我們思考一些問題，包括現今華語文教學的現狀。目前華語文教學在台灣並沒有統一教學綱領，應用 CEFR 來建構台灣華語文教學架構是一個值得嘗試的路，待語言能力架構、能做描述的系統(can-do system)得到教學者、學習者的認同，並落實在台灣的華語教學中，教學、學習、評量與國際接軌，例如，在台灣學習一年華語的學生可以達到 CEFR 的 B1-B2 程度，之後再來談 CEFR 與華語測驗的連結還不嫌晚。否則現在貿然進行 CEFR 與測驗的對應，反而形成教學、學習與測驗評量之間的鴻溝。

在思考的過程中，CEFR 能不能做為華語文教學的綱領，這個問題得到大部分專家學者的肯定。原因很簡單，CEFR 以廣泛的方式描述溝通所需要的能力、相關的知識和技能、溝通的情況與領域，描述語言學習者必須學習做什麼才能使用語言進行溝通，學習者必須發展什麼樣的知識與技能才能有效地行動，整個發展是從七 0 年代的英語能力門檻(English Threshold Level)開始，走到目前的完整架構，長達二十多年，中間的工作包括了語言學習理論上概念的整合以及實務可行性的修正，這個架構提供語言教學者、學習者與評量者三方面清晰完整的指標性能力描述。學者們深信一個製作嚴謹的外語能力架構具備共通性，加上建立語言能力指標是一項相當耗費人力與時間的工作，與其重新建立華語文能力指標，不如套用此架構，只要針對漢語語言的特徵做一微調或細部描述即可。

一旦華語文能力指標的架構確立，華語教學者最關心的問題不外是在這些抽象的能力說明之下，如果要達到各等級所描述的能力，學生應該具備多少詞彙、甚至哪些詞彙、哪些語法等實際問題，也就是必須給課程設計與教材編寫者一個指導性的原則，類似台灣九年一貫國語文分段能力指標中，指出 1-3 年級要認識 1000-1200 個常用中國字；4-6 年級要認識 2200-2700 個常用中國字；7-9 年級要認識 3500-4500 個中國字。

由於這個架構的敘述，強調的是學習者能做到什麼，能完成什麼樣的任務。要把行動導向的語言能力架構和學習者內在語言能力（例如，語音、語彙、語法等等知識）連結起來，需要許多專家共同參與，不是短期內可以達到的目標。雖然如此，華測會無法逃避這個問題，這也是為什麼，在面對大眾殷切的詢問與期盼下，華測會不得不在網站上，公布 TOP 所參考的八千詞彙的內容。站在教學者、學習者或考生的立場而言，他們需要的只是一個可以操作性的原則或是一個規範，甚至是準備考試的一個方向。對於專家、學者而言，詞彙量及詞彙內容應該是多少或是什麼，每人心中有自己的想法，見仁見智。

既然制訂華語文能力指標，就是為了讓華語教學有一指導性的原則，那麼每一個等級需要教授的內涵或質量，是應該制定出標準，讓教學機構、教材編輯者有所依循。以漢語的特性而言，詞彙量和字彙量是一般對規範的基本要求。不過，對外華語教學幾乎都是以詞彙為基本單位<sup>1</sup>，所以前人做過教材中常用詞彙的統

<sup>1</sup> 除了為華裔子弟（只會聽、說，不識字）特別設計的課程較不一樣，一開始就加強寫字、認字。

計研究(葉德明 1995)，但還未見針對台灣華語教學的字彙量或字彙分級研究文獻。大陸漢語水平考試則是於 1992 年規範了四級漢字量與內容，甲級 800 字，乙級 804 字，丙級 601 字，丁級 700 字，總共 2905 個字。本文初步擬擷取八千詞彙中的初等詞彙 1500 詞，拆解成漢字，與其他兩個對外漢語的字表對照，藉由觀察對照來討論對外漢語教學字彙相關問題。

## 二、對外漢語字集的內涵與功能

所謂字集是指爲了某一使用目的而選定的一群字(黃沛榮 2001:37)，本文所討論的對外漢語字集，是爲學習現代漢語的外籍人士選定一群字，也就是對他們而言，有學習價值的字。關於漢字的學習價值，由於華語教學的特性，一般人較容易忽視漢字的學習價值。黃沛榮(2001)一書中特別強調漢字教學的目標，主張從識字、寫字、用字這三方面來分析字的優先學習價值。對外漢字教學究竟是詞本位？還是字本位？我個人認爲不是一切爲二的單純，而是必須從課程設計、教學對象、教學目標等方面來考量。如果像早期歐洲漢語教學就是爲了讓學生能讀報、看懂文言文，著重翻譯或閱讀的能力，那麼以字本位出發教學是可行的方向。也就是黃沛榮(2001)一書中提出漢字學習的三個目標：

1. 掌握收字多、字義明顯的部首
2. 掌握組字力強的部件
3. 掌握構詞率高的單字

如果學生學習的是現代漢語，是以人際間的互動、溝通爲語言學習的目標，那麼哪些字應該優先學習，或許還是應該以常用詞彙爲出發點，幫助學生可以在短時間內看懂或寫出便條、留言等等溝通時使用的書面文字。所以在上述的三個學習目標上，我想還可以加上一條「掌握常用的單字詞」，例如「叫」「是」「的」「我」等等，是常用詞也是一個「單字」，教學時可以從這些詞的讀寫入手，再從分析部件或部首，讓學生進一步理解漢字結構，表形、表音、表義等特徵，以培養日後對漢字讀和寫的能力。

在 CEFR 的能力架構之下，學習就是爲了和人群溝通，因此本文從溝通功能來討論對外漢語字集，希望這個字集未來可以做爲教材編輯者或評量者的參考依據。但限於篇幅與時間的關係，本文沒辦法探討字彙分級或各級字彙量的問題，只能檢視對外漢語字表，嘗試探討基礎/初級的字彙內容。在檢視之前，我們先看看 CEFR 對於基礎使用者在書寫能力方面的描述。以下爲 CEFR 的 A2 能力說明，包含 A1 所應具備的能力，也就是基礎使用者所應該具備的書寫表達與互動的能力。從下面的敘述中，可見基礎使用者運用語言書寫的領域，多在個人範疇，而且是個人熟悉的主題。而這些情境也是初級華語教學的情境設計，只是在一般課堂評量時，教學者較少針對這些書寫功能設計課程，而是著重培養學生聽和說的能力。

## A2 書寫表達能力

- 能寫出一連串簡單片語及句子，使用簡單的連接詞，像「和」、「可是」、「因為」。
- 能針對事件、過去的活動以及個人經驗，寫出簡短、基本的描述。
- 能寫出簡短、簡單的、富想像力的自傳。
- 針對家庭、生活條件、教育背景、現在或是最近從事的工作，能寫出一連串簡單的片語及句子。
- 針對自己以及想像的人物、居住地點及從事的職業，能寫出簡單的短語及句子。(A1)
- 能寫出簡單、獨立的短語和句子。(A1)

## A2 寫作互動能力

- 能對於切身的需求，寫出簡短的、簡單的定式語言(formulaic)的便條。
- 能寫出非常簡單的個人信件，表達感謝與歉意。
- 在可以要求對方重複或以他種方式表達的情況下，能記錄簡短簡單的訊息。
- 能寫出一張簡短簡單的明信片。(A1)
- 能寫出數字與日期、自己的名字、國籍、地址、年齡、出生日期、或入境時間等等，例如旅館住宿登記表格。(A1)

所以在下一節中，我們採用初等詞彙表中的 1500 個詞來分析漢字，與其他兩個字表對照觀察，是否能找出多數共同字彙。不過，從這些能力指標也可以想像，學習者要完成這些任務所需的字彙不一定在字集規範中，例如，學生姓名，如果非常用字，就不可能在字集中，而是學習者自己的字彙能力。另一個人們漸漸關心的議題是，由於電腦文書處理的強大功能，不管是拼音文字的語言還是非拼音文字的語言都面臨大眾書寫能力的退化，新新人類拼錯字或寫錯字的現象日趨嚴重，所謂漢字能力可能因為使用不同媒介而產生不同結果，例如，如果學生認讀能力強，透過智慧型電腦輸入軟體寫作，一樣可以寫出選字正確的短文；如果採手寫形式，可能沒法順利完成。這些問題，還有繁簡字認讀問題，未來在編訂字集時都得釐清、處理。

### 三、對外漢語字集檢視

從上一節的討論，本文假設TOP初等詞彙的內容是符合初級學習者的情境，因此在制訂字集時，先從這個詞表出發，將詞彙拆解為「字」，做成爲字表，得出 863 個不同的漢字<sup>2</sup>。然後檢視另外兩個以華語爲外語學習的字集，對照異同。一是漢語水平考試（HSK）大綱的甲級字表（即初級）800 個字，二是法國中文字表 804 個字<sup>3</sup>。TOP字表與HSK字表都是以詞彙爲前提，從常用詞彙分析出來的漢字（劉英林等 1992），屬於初級程度的字彙；法國字表則是字本位<sup>4</sup>，選字的主要標準大致是以該字的構詞能力和使用頻率爲考量。這 804 個字是法國學習中文的高中生在畢業以前必須學習的字，其中 505 個是主動字，也就是要會寫，其它的要求會認即可。

三個字表統計分析出來的簡要結果如表一，總共有 1062 個不同字形的漢字，其中 593 個是三個字表共同出現的字（詳見附錄一），有 219 個字出現兩次。也有只出現在一個字表中的字，在 HSK 字表中有 50 個字；TOP 則是有 102 個字；法國字表有 98 個字。這 593 個共同字約佔個別字表的百分之七十左右，也就是說，雖然是三個不同地方所做出來的字集，在初級階段選用共同字的比例不低。

表一

	TOP	HSK	France
總字數	863	800	804
共同字%	68.7% (593/863)	74.1% (593/800)	73.8% (593/804)
只出現 1 次字	102	50	98

比較有趣的現象是，在法國字表中「王、李、林、華」等在台灣常見的姓氏或專有名稱，只出現在法國字表中，可見在編輯字集時，不能單以常用詞彙爲標準，否則這些常見的姓氏和地名漢字反而不在學生的學習教材之中，例如「台」「灣」這兩個字。當然，我們也不可能把專有名稱都列在字集中，如何取捨，還需要進一步討論。白樂桑教授去年來台灣的演講中即提到，在 CEFR 的 A1 能力敘述中涵蓋學習者寫簡單明信片的能力，他問如果這個學生在中國的爨縣旅遊，那麼以 A1 程度的學生有辦法寫出「爨」這麼複雜的字嗎？當然未來在規範這些標準時，都要考量這些問題，目前初步想法是，初級程度學習者(A1,A2)，應該具有摹寫的能力，即使只知道發音，也可以允許部分漢字使用拼音的方式寫出。

<sup>2</sup> 字表是爲此報告而抽取，所以沒有對外發表。至於華語八千詞的內容則已經放置在測驗網站上：[www.sc-top.org.tw](http://www.sc-top.org.tw)

<sup>3</sup> 雖然法國教育部文獻上列出 805 個字，但是經過本文比對，其中「著」出現兩次。實際只有 804 個。

<sup>4</sup> 根據法國教育部中文教學總督白樂桑教授 2007 年在台灣演講稿記錄，高中畢業的時候，學生規定必須要會 805 個字，...字表的選字標準除了該字的構詞能力和使用頻率之外，還考慮加入個別特殊字，並不一定是高頻字，但口語常用，例如“餓”。

在觀察這三個字表的異同時，本文也檢視這三個字表中是否包括黃沛榮 (2001) 提出的 72 個具有優先學習價值的整字部件/部首，所謂整字部件/部首是指這個字本身既可以單獨使用，也是一個重要部件/部首。結果顯示，這些具有優先學習價值的整字部件/部首超過半數出現在共同字中，共有 41 個：

「一、人、八、力、十、又、口、大、女、子、小、山、工、心、戶、手、支、斤、方、日、月、水、火、牛、白、目、米、羊、肉、行、衣、見、言、走、足、車、門、雨、非、馬、魚」。

有 7 個整字部件/部首則是出現在兩個字表中「食、金、角、示、瓜、立、刀」，而在這三個字表中，TOP 字表對整字部件/部首涵蓋量較其他兩個字表多了 6 個字「土、木、玉、石、耳、鳥」，但以下 18 個具有優先學習價值的整字部首/部件並沒有出現在三個字表中的任何一個：「巾、田、竹、寸、弓、戈、欠、止、犬、皿、禾、穴、羽、舟、貝、邑、阜、虫」。

觀察結果顯示，法國字表是否以字本位出發，似乎值得商榷。如果都是詞為出發點，大致可以知道為什麼這 18 個整字部件/部首都沒有出現在這些集合中，一是因為這些整字大部分難以單獨使用（也就是做為單字詞），一是初級語言溝通中罕用這些字。但我想並不表示這些部件/部首學生沒有概念，例如 TOP 初等詞表中有「張、和、歡、船、管、對、男、臭」等等，這些字詞包含了「弓、禾、欠、舟、竹、寸、田、犬」等整字概念。是不是一定要先學「犬」，才能學「臭」呢？初級學生在日常會話中要聽到「犬」的機會很少，在初級情境教學中也不太可能接觸到。但是學習「狗」「臭」等漢字時，可能教師已經把「犬」的概念帶入，學生因此已經建立了這個字的概念。所以這樣性質的字需不需要列入初級字集中呢？還有討論的空間。

#### 四、建議與未來方向

透過上述三個字集的觀察討論，希望對華語字集的制訂原則有一點點撥雲見日的效果。除了分析探討現有的對外漢語字集外，建議未來的研究者也可以參考教育部、中央研究院等機構所產出的常用字表或常用字頻率資訊，制訂進階程度字集的量及涵蓋內容，因為學習者一旦進入進階或高級課程，多方的閱讀仍是增進閱讀能力的關鍵。那麼，進階的字彙就不一定要以華語教材或溝通情境為本，高頻出現的字彙可能也是外籍學生日常所接觸的。關於字彙量的部分，如果以目前 TOP 考試對應 CEFR 等級的初步構想來規劃，基礎級相當於 A2，初等考試相當於 B1，中等考試相當於 B2，高等考試相當於 C1，對應於 TOP 詞彙量，可以朝此方向發展為 A2-500 字；B1-900 字；B2-1800 字；C1-3000 字。

參考書目：

法國國家教育部（2003）《高中課程參考文集—中文》。法國：國家教育資料中心。

黃沛榮（2001）《漢字教學的理論與實踐》。台北市：樂學書局。

國家對外漢語教學領導小組辦公室漢語水平考試部（1992）《漢語水平詞彙與漢字等級大綱》。北京語言學院出版社。

鄭昭明（1997）漢語水平考試的定位、編製及「字彙」與「詞彙」使用的問題，《華文世界》第 85 期，86 年 9 月，42-47 頁。

葉德明（1995）《華語文常用詞彙頻率等級統整研究（附錄一）》，行政院國家科學委員會專題研究計畫成果報告，民國 84 年。

劉英林、宋紹周（1992）論漢語教學字詞的統計與分級。收錄於《漢語水平詞彙與漢字等級大綱》，第 1-25 頁。

張莉萍、陳鳳儀（2006）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》（蘇新春、王惠主編），250-260 頁，新加坡：新加坡中文與東方語言信息處理學會出版。

ACTFL (1986) ACTFL Chinese Proficiency Guidelines.

The Common European Framework of Reference for Languages is published in English by Cambridge University Press - ISBN Hardback 0521803136 Paperback: 0521005310.

Quirk, R. (1981) International communication and the concept of nuclear English. In Smith, ed., 1981:151-165.

Smith, L. E. ed. (1981) English for Cross-Cultural Communication. London: Macmillan.

附錄一、優先學習字集（593 字）

一 元 冬 任 考 更 姓 近 洗 師 送 教 魚 訴 萬 算 誰 還  
七 內 出 先 肉 步 定 長 活 庭 酒 族 備 貴 節 精 論 雖  
九 六 加 全 自 每 店 門 爲 息 院 晚 最 買 經 綠 賣 點  
了 公 包 共 色 求 往 雨 界 拿 除 喝 跑 腳 實 禮  
二 分 北 再 行 決 念 青 相 拿 馬 條 望 緊 輾 簡  
人 切 半 同 衣 沒 怕 非 看 旁 高 涼 聞 鞋 舊  
八 化 去 各 西 汽 或 亮 科 旅 停 清 語 餓 醫  
力 午 可 向 位 房 信 秋 時 假 深 認 器 離  
十 友 右 名 住 所 便 穿 書 做 球 說 學 雙  
又 天 叫 合 何 拉 前 紅 健 理 現 輕 整 雞  
三 太 只 吃 但 抱 南 苦 動 產 遠 樹 顏  
下 少 史 因 作 你 走 足 提 幾 需 橋 題  
上 心 旬 回 你 足 換 幾 領 餃 機 騎  
久 戶 四 地 低 身 敢 提 話 領 麼 歷 壞  
也 手 外 在 冷 車 期 換 黃 餃 齊 燈 藥  
千 支 奶 多 別 那 湖 敢 黑 黃 齊 興 識  
口 文 它 好 利 事 些 增 寫 廣 親 辦 邊  
大 斤 左 她 助 些 使 增 寫 廣 親 辦 邊  
女 方 市 如 吧 告 呀 困 坐 完 希 床 弟 忘 快 我 把 找 改  
子 日 平 必 打 本 正 母 民 生 用 白 目 交 休 件  
小 月 必 打 本 正 母 民 生 用 白 目 交 休 件  
山 比 本 正 母 民 生 用 白 目 交 休 件  
工 毛 水 火 父 片 牛 世 主 以 他 代  
己 水 火 父 片 牛 世 主 以 他 代  
已 才 不 中 互 五 什 今 介

建議加入：王、李、林、華、台、灣、耳、石、玉、木、鳥、土