

# 由中文雙音詞談中文電腦

● 魏凌雲

## 作者簡介

魏凌雲教授為交大（重慶九龍坡）電信研究所第一屆碩士（1946），現任母校計算機工程研究所客座教授。

現代一般人，提到中文，多半搖頭三嘆：難，難，難，（難唸，難寫，難記），下列事實足為佐證：（一）大學生中文程度，普遍低落。（二）中文電腦輸入困難，達官貴人，匹夫匹婦都不會用。（三）中國文化在海外日趨消沉。

難道中文及中國文化真的是：「夕陽無限好，只是近黃昏」嗎？我的看法則不然。我堅信一個定律，（無以名之，姑稱為魏氏定律）：「天下任何人，任何事，任何學問，不會盡善盡美，有缺點，亦必有優點，反之亦然」，根據這個定律，中文縱然有不少缺點，但必定有優點，我們中國人的責任，第一是要找中文的優點在那裏，第二是怎樣發揮這些優點，使得到最高的效益，嘉惠人羣。

依我多年來的研究，中文結構上最大

的特色，是雙音詞的普遍性及卓越性，所謂雙音詞，是兩個字結合成為一個慣用詞，如：天下，國家，社會，漂亮，能幹，雖然，所以，台北等，在中國文化中，雙音詞趨於普遍化，實在是由於社會結構變化所促成。上古時代，以農業為主。農人除了種田，吃飯以外，所遇到的事，不外乎生老病死，年節，豐收，災亂等，因此他們的交談，用簡單的詞句，便可以完全表達他們的思想。但當社會進化到有商業行為，政治軍事，學術研究的時候，中文的單字，由於同音字太多，對於複雜的事體，便難得明白地表達，於是雙音詞的滋生，一方面減少同音的模稜性，另一方面增加詞彙的廣度與深度。因此中國文字的表現，便豐潤華美，多采多姿。隨社會的進步，雙音詞在中文所佔的比例愈來愈

大，為了證明這個假設，可依時代的順序，引幾段文字來看看。

## （一）孔子論語

子曰：學而時習之，不亦悅乎；有朋自遠方來，不亦樂乎；人不知而不愠，不亦君子乎？

在上段卅二個字中，僅「遠方」及「君子」是雙音詞，所佔比例為  $4/32=12.5\%$ 。

## （二）崔之顥（唐朝）黃鶴樓

昔人已乘黃鶴去，此地空餘黃鶴樓，黃鶴一去不復返，白雲千載空悠悠。晴川歷歷漢陽樹，芳草萋萋鸚鵡洲，日暮鄉關何處是，煙波江上使人愁。

詩中雙音詞所佔比例為  $36/56=64\%$ 。

## （三）李後主（南唐）虞美人

春花秋月何時了，往事知多少？小樓昨夜又東風，故國不堪回首月明中。雕闌玉砌應猶在，祇是朱顏改。問君能有幾多愁，恰似一江春水向東流。

雙音詞字數所佔比例 =  $42/56=75\%$ 。

## （四）現代講詞

今天是新年元旦，兄弟謹向各位拜年。恭喜大家發財，萬事如意，新年快樂。今天兄弟聊備水酒，敬請各位務必光臨，共渡佳節，並祝國運昌隆。

上段中雙音詞字數所佔比例 =  $46/55=84\%$ 。

由上看來，從孔子到現在兩千多年，中國文字的表達，由古樸、典雅到口語化，雙音詞比例愈來愈大，竟高達 80% 以上。

我們再從國語日報辭典，更得到一個具體的證明，根據該辭典，我們發現：

詞類	數目	比例
單音(字)	9098	22 %
雙音	25552	62 %
三音	3270	8 %
四音	3094	7.5 %
其他	207	0.5 %

總結來說，雙音詞是中文結構的骨幹，它有音有意，而且意義明確肯定，表達豐富深入，活潑經濟，凡是學中文的，用中文的人，對於這一點，必須要有深刻的認識。

從發音學來看，中文字的單音（包括四聲）共有 1281 個，如果不分四聲的話

，則有語意的音節僅有 417 個，因此中文的同音字非常多，這是中文語音上先天的缺憾，在後天方面，補救的方法，是採用四聲，以增廣音域，但四聲非常接近，學習與辨別，都有困難，再者中國由於幅員遼闊，往昔交通不便，致造成千百的方言，俗稱「南腔北調」，使中文發音，更是光怪陸離，洋洋大觀。

上述的原因，使中文實在「難唸」。於是有些人咒罵方块文，而對蠶形文字，歌頌不已，難道我們就如此認輸？幸好，我們有雙音詞，正好解救上述三大困難：(1)同音字(2)四聲(3)南腔北調。舉例來說，我唸一個「中」字，你不知道究竟是下列字中的那一個：鐘，終，盅，衷，中，忠，松，充，沖，沖，衝，宗，樞，棕，踪，縱，愜，粽，綜……，但是如果我唸「中國」，你一聽就懂，決不會想到任何其他同音詞，縱然有幾個方言不同的人用國語說「中國」，音調有差別，你也會聽出是「中國」，而不會誤想成「棕果」或「終過」。由此我們可以看出雙音詞的好處：第一，它的組合音（雙音），難有相同的。第二，它的意義，明白肯定，毫不含糊，所以一個雙音詞，事實上是一組簡明而確切的「音意」結構，聽起來，即令音不準，它的涵義可以輔助，這正是補救「單音」的缺點。雙音詞在中文，特別

是在交談中佔的比例很大，多半歸功於它的優越「音意」結構。

用洋文來翻譯，中文的雙音詞，相當於洋文的單字，例如：中國：China，文化：culture，美麗：beautiful，一個：a。但是從發音來說，雙音節僅有兩個音節，而相當的洋文，則可以是一個音節或多音節，因此如作語音辨認，中文雙音詞比洋文單字容易得多。

為要證明中文雙音詞在「音意」結構上的優越性，我們曾用 IBM PC/AT 電腦作中文注音符號，單字及雙音詞語音的辨認，所得的結果如下：

- (1) 59 個注音符號：辨認率(R.R.)= 76.3%
- (2) 兩組注音符號：子音組 R.R. = 78.6%  
鼻音組 R.R. = 83.3%
- (3) 40 個單字 R.R. = 95 %
- (4) 100 組雙音詞 R.R. = 99 %

以上的結果顯示，機器辨認雙音詞，幾乎絲毫不爽，這有雙重意義：

(一)它肯定中文雙音詞「音意」結構的優越性。

(二)機器辨認中文的可行性，遠比一般人預期為高。

第二項包括幾種涵意，第一，中文語音輸入電腦的實現，為期當不在遠，屆時中文鍵盤輸入法，在專業用途方面，必然淘汰。第二，語音辨認的科技與應用，中

文可能會優於洋文。第三，中國文字的優點，如能發掘，並且充份發揮，可以助長中國科學的進步。

對於以上的涵意，需要作進一步的解釋，有人擔心，中文雙音詞數量太大，如要滿足大眾會話、閱讀及寫作各種用途，當在十萬以上；若要存入電腦，記憶容量及查尋速度，都成問題。關於這個問題，有兩種解決方法，第一，專業用途方面，雙音詞需求量，不但不大，而且相當有限。茲以電腦用作火車訂票來說明，在這方面，用得最多的詞是地名，都是雙音詞（如台北，高雄），其次是固定用語，如：時刻，票價，車種（莒光，自強，復興等），起站，終站，動身，到達，換車，上午，下午，夜間，準時，誤點，停留，不停等，不僅都是雙音詞，而且數目有限，再者，訂票所用的「句型」，都可預先規定。因此一個人可在家中打電話到火車站訂票，他祇須說「訂票」，則電話直接接通電腦，在他講完訂票用語以後，電腦立即回答：(1)車次時刻及座位號碼(2)票價(3)訂購號碼。上述回答可用聲音或直接顯於詢問者個人電腦螢幕上。訂購者需在一小時內去火車站（或指定代售處），報訂購號碼，付款取票，如果一小時內不購票，則電腦自動銷號，即是訂票作廢。上述火

車訂票電腦，所需字彙，一千足夠，祇要雙音詞辨認很準的話，則「語法分析」和「語意分析」都可大為簡化，因此這種電腦資料庫及程式設計，諒無困難，如能將「台語雙音詞」及「國語雙音詞」兩套辨認率做到 97% 以上，則火車訂票服務，便可由電話完成。

至於一般性的用途，我們可採中央電腦站，它的資料庫，自可藏十萬以上的雙音詞。光碟的問世，將可解決儲存的問題，任何人以個人電腦接通中央電腦站，便可予取予求（得付費），達到他的欲望。

其次，我要談到語音辨認的科技與應用，中文可能會優於洋文的推斷。中國文字有兩大先天優點：第一，每字是單音節。第二，雙音詞質優量大，這兩大優點，是世界上最任何文字無法媲美的，由於單音節及雙音節，機器分割中文連續語音，將遠比洋文容易，例如英文說 Don't you，聽起來就像一個字，分不出字與字間的界限，再如說 Where were you while we were away, 或 She eats fish sandwich，英文稍差的人，聽來簡直莫明其妙，機器辨認困難，自不待言，這種困難，主要在洋文的(1)連音（Coarticulation）(2)音節長短不一，在中文方面，連音的情形（例如：將“司安”唸成“三”）不多，音意肯定的雙音詞佔大多數（例如：今天我

去台北參加朋友結婚典禮)。而雙音詞可使同音字，四聲，南腔北調引起的困難，大為減輕。根據以上的理由，我敢大胆地說，語音辨認的科技與應用，中文可能會勝過洋文，這不是我們的本領比洋人強，而是我們祖宗的遺產，中文的單音節與雙音詞，是西洋文化永遠無法媲美的。

一般人認為中文太難，是因為他們祇看到中文（音形義）的缺點，學理工的人，不願碰或不會用中文電腦，所在皆是，問題在很多人（愛好文學者除外）不相

## 會議動態：

財團法人國立交通大學同學會教學基金會第六屆第二次董事會於七十五年十月二日假台北市大陸工程公司會議室舉行，會中報告摘要於下：

(一)本會比照學術基金會之組織辦理，在董事會之下設執行秘書一人，另分設文書組、會計組、及審核組三組，每組設主任幹事一人，若事忙得另聘幹事若干人助

信中文有優點，自不願花精力去探尋中文有無優點，如果說中文有優點，而且可能助長中國科技的發展，有人將認為是荒誕不經，癡人說夢。我這篇蕪文，在蘇散這類「貶己」的思想，我要鄭重說一句，單音節和雙音詞，是中國文字的雙寶，充分發揮它們的潛力，將使中文電腦，更上一層樓，傲視全球，如果不信，請看來日電腦域中，究竟是誰家天下！

寫於新竹一九八七年六月

理之，如此，以便利將來兩個基金會之合併。

(二)本會目前工作同仁由學術基金會相關人員兼辦。即執行秘書兼文書組主任幹事由汪學長寶書兼任，會計組主任幹事由何學長開藩兼任，審核組主任幹事由盧學長善棟兼任，另聘請賴密芬及何慶珍兩位小姐分別擔任文書及會計兩組幹事。

## 讀陳珮玲教授的「我評『耀真盃』」文後

●老學生

收到「交大友聲」第320期，讀至陳教授的評亞大第四屆英語演講比賽，得知本屆比賽的裁判，連他共有五位，都是鼎鼎大名的名教授，唯有他執筆為文，才讓友聲的讀者們知道「參加比賽的學生成績那麼好」——陳教授的文中評語。不過，在該文第三段中有言：「比賽第二部份即席演講，這方面學生的表現好像比較弱，詞話比較不流利，內容也缺乏組織。……語法的錯誤真是太多，特別是預備的演講，不應該有那麼多錯誤。」當我讀完這些評語，不禁感覺陳教授的文章，有如「春秋史筆」，褒貶均具。同時，我也想到陳教授目前是交大指導英語研習社的教授，他擔任裁判後，必定面授參與比賽的同學和研習班的學友，要如何改良語法上錯誤的缺點。

現在我以交大「老學生」——照張岳軍先生的說法，我今年還只有五歲，該稱「學生」，而在抗戰期間，我已在上海交大就讀，畢業後，還是到處都學，故名「老學生」——的身份，並以不揣冒昧的退休後來美又學了十年英文的經驗，轉告在校的校友，並請教於各位老師尊前。

首先我不談英語，僅談「語言」，我們都聽到或看到，社會中有多數人不會「寫」，却有很好的口才，能說會「吹」，其地位決不低於只會寫的人。且據我親歷的環境，是在民國五十年，我奉命去中東，歷時近半年，途經東南亞，因我只能說鄉音的國語，遇了許多講廣東話和閩南語的同胞，真像「啞巴望著狼」，有口不能開，可見「語言」是多麼重要！因此，我勸各位校友，趁年輕，多學幾種「語言」，以免像我一樣吃苦頭。

次談「講英語」，我很笨，耳又聾，在美住過東、南、西、北、中各大州，除了住醫院養病外，有機會總在學英語，却還是個「半吊子」。猛記在波士頓的MIT圖書館時，遇見一位中國留學生，整天在圖書館看書，首先我以為他找參考書，後來才看到他除看圖書館的參考書外，多半在勤讀教授所講的書籍，原因是他初到美國聽不慣英語，不如自己用心看書更透切，只是自己太苦了點，另有一位太太是台籍，只唸過小學三年級，就嫁與我們一位學長，廿多年前他帶太太來到美國，先