

## 附錄2 用詞分析

### 壹、用詞統計方法

詞是最小有意義且可以自由使用的語言單位，詞彙是語言組成的要素之一，使用的詞彙越多，則表現越豐富。然詞彙的分析相當複雜，如何分析詞語，成為歷來研究中文語言文法的重要議題。

一般多利用詞典收錄的詞和文本比對進行分析，然在語言不斷遞嬗的情形下，無法切分新詞成為研究分析的問題之一。據此，中央研究院開發以：「自動抽取新詞建立領域用詞或線上即時分詞功能。為一具有新詞辨識能力並附加詞類標記的選擇性功能」<sup>1</sup>之中文斷詞系統，提供外界使用。

本研究應用中央研究院於網路提供之中文斷詞系統<sup>2</sup>以及Microsoft Excel，協助分析高中及高職二年級的學生在書寫本次非選擇題：「評論短文」及「創作短文」的用詞情形，其正確率約達96%<sup>3</sup>。由於中文的詞彙組成複雜繁多，故本研究主要以高分群（評論短文3級分、創作短文5-6級分）的學生為主進行詞彙的分析統計，分析篇章數如下表。

|      | 高中評論<br>(3級分) | 高職評論<br>(3級分) | 高中創作<br>(5-6級分) | 高職創作<br>(5-6級分) |
|------|---------------|---------------|-----------------|-----------------|
| 篇數   | 149           | 84            | 145             | 74              |
| 用字總數 | 26176         | 14916         | 60986           | 32188           |

### 貳、高中高職用詞統計結果

本次施測以斷詞系統分析後，統計高中、高職評論短文及創作短文用詞數量，詳如【表 42】。依本次施測統計之用詞頻率<sup>4</sup>詳如【表 43】，統計可見：無論評論或創作短文，用詞頻率皆集中在1-100頻次內，顯示無論高中、職，他們用以表達相同題目的文章裡，大多使用不同的用詞。

<sup>1</sup> 創中央研究院斷詞系統簡介。民國102年4月15日，取自：<http://ckipsvr.iis.sinica.edu.tw/>

<sup>2</sup> 中央研究院用詞統計網站出處。民國102年4月15日，取自：<http://ckipsvr.iis.sinica.edu.tw/>

<sup>3</sup> 據斷詞系統網站提供資料說明：斷詞正確率約為95%-96%，因此斷詞結果中仍會有部分錯誤。

<sup>4</sup> 分別依本次施測評論短文3級分、創作短文5-6級分之高中、高職分別統計的結果，頻次越高代表使用的次數越多。

【表 42】詞頻分析篇數及詞目統計

|      | 高中評論<br>(3級分) | 高職評論<br>(3級分) | 高中創作<br>(5-6級分) | 高職創作<br>(5-6級分) |
|------|---------------|---------------|-----------------|-----------------|
| 篇數   | 149           | 84            | 145             | 74              |
| 統計詞目 | 2310          | 1523          | 5161            | 2894            |

【表 43】詞頻分析頻次統計

| 頻次         | 1-100  | 101-200 | 201-300 | 301-500 | 501以上 |
|------------|--------|---------|---------|---------|-------|
| 高中評論用詞     | 2289   | 17      | 2       | 1       | 1     |
| 高中評論用詞頻次比例 | 99.09% | 0.74%   | 0.09%   | 0.04%   | 0.04% |
| 高職評論用詞     | 1514   | 6       | 2       | 0       | 1     |
| 高職評論用詞頻次比例 | 99.41% | 0.39%   | 0.13%   | 0.00%   | 0.07% |
| 高中創作用詞     | 5116   | 23      | 9       | 6       | 7     |
| 高中創作用詞頻次比例 | 99.13% | 0.45%   | 0.17%   | 0.12%   | 0.14% |
| 高職創作用詞     | 2868   | 14      | 6       | 1       | 4     |
| 高職創作用詞頻次比例 | 99.14% | 0.48%   | 0.21%   | 0.03%   | 0.14% |

### 參、與教育部〈詞頻總表〉比對

教育部《國語辭典簡編本編輯資料字詞頻統計報告》中有〈詞頻總表〉<sup>5</sup>，民國86年時配合簡編編輯目標，希望能符合國民中小學生及一般社會人士的語文使用實況。其中所收錄的「詞」較為廣義，是指多音節的詞及短語、諺語等。依「政經」、「科學」、「醫學」、「文學」、「兒童文學」、「綜合」、「教科書」及「其他」等八大類，隨機抽樣選取八十一年度及八十二年度發行的報刊雜誌、八十二年度暢銷排行榜書籍部分內容及國小國語課本、國中國文課本全文，作為主要統計資料。該表經過分詞後，統計得 64326 詞。

本研究與〈詞頻總表〉比對結果如【表 44】。另外，依用詞頻次分別統計後可以發現在本次施測中，無論高中或高職高分群學生在「評論短文」及「創作短文」對照〈詞頻總表〉詞頻序號<sup>6</sup>，超過半數落在1-10000序號中。其中高中創作短文的部分，約有三成五在10001詞頻總表序號以上，相對於高職創作用詞來看，可見高中生的用詞量較高職生高，詳如【表 45】。

<sup>5</sup> 《國語文教育叢書21·國語辭典簡編本編輯資料字詞頻統計調查報告》(民國86年3月臺灣學術網路三版)，《國語辭典簡編本編輯資料字詞頻統計報告·統計表·詞頻總表》。民國101年9月25日，取自：[http://www.edu.tw/files/site\\_content/M0001/pin/biau1.htm?open](http://www.edu.tw/files/site_content/M0001/pin/biau1.htm?open)

<sup>6</sup> 詞頻序號為〈詞頻總表〉依統計結果，由使用率大至小排序的編號，編號越大代表使用的次數越低。

【表 44】與〈詞頻總表〉比對結果

|           | 高中評論<br>(3級分) | 高職評論<br>(3級分) | 高中創作<br>(5-6級分) | 高職創作<br>(5-6級分) |
|-----------|---------------|---------------|-----------------|-----------------|
| 斷詞系統統計詞目  | 2310          | 1523          | 5161            | 2893            |
| 與〈詞頻總表〉比對 | 1642          | 1057          | 3477            | 1963            |

【表 45】與〈詞頻總表〉比對統計

| 〈詞頻總表〉詞頻序號 | 1-10000 | 10001-<br>20000 | 20001-<br>30000 | 30001-<br>40000 | 40001-<br>50000 | 50001以上 |
|------------|---------|-----------------|-----------------|-----------------|-----------------|---------|
| 高中評論用詞     | 1235    | 212             | 83              | 48              | 31              | 33      |
| 高中評論用詞頻次比例 | 75.21%  | 12.91%          | 5.05%           | 2.92%           | 1.89%           | 2.01%   |
| 高職評論用詞     | 849     | 98              | 53              | 25              | 16              | 16      |
| 高職評論用詞頻次比例 | 80.32%  | 9.27%           | 5.01%           | 2.37%           | 1.51%           | 1.51%   |
| 高中創作用詞     | 2290    | 586             | 273             | 134             | 73              | 121     |
| 高中創作用詞頻次比例 | 65.86%  | 16.85%          | 7.85%           | 3.85%           | 2.10%           | 3.48%   |
| 高職創作用詞     | 1433    | 280             | 99              | 58              | 35              | 58      |
| 高職創作用詞頻次比例 | 73.00%  | 14.26%          | 5.04%           | 2.95%           | 1.78%           | 2.95%   |

#### 肆、評論短文用詞統計

分別統計高中、高職3級分評論短文用詞後，再與〈詞頻總表〉進行比對。並以〈詞頻總表〉為基準對照比較高中及高職學生的用詞結果，其中符號【#N/A】表示比對後未出現符合的詞。以下列舉【表 46】本次施測高中評論短文前20個常用詞為例，可以發現在高中評論序號1-20的詞目中，高職排序亦略相符合。其中「男人」、「女人」、「袋鼠」這些詞目，因受本次施測內容影響，故在高中及高職評論短文中使用的頻率較高。

【表 47】以〈詞頻總表〉為基準，比較評論短文高中（高職）使用，但是高職（高中）未使用的語詞，並依〈詞頻總表〉的序號統計。結果顯示這部分的語詞多集中在1-10000序號內，高於10001序號外高中或高職的比例相差不多。

【表 46】高中評論詞頻序號1-20

| 高中詞頻序號 | 詞目 | 高中評論頻次 | 高職詞頻序號 | 高職詞頻 | 比對<br>〈詞頻總表〉 |
|--------|----|--------|--------|------|--------------|
| 1      | 男人 | 415    | 1      | 214  | 412          |
| 2      | 女人 | 172    | 2      | 102  | 286          |
| 3      | 袋鼠 | 139    | 5      | 64   | 9636         |
| 4      | 世界 | 99     | 8      | 51   | 23           |
| 5      | 社會 | 99     | 4      | 76   | 19           |
| 6      | 可以 | 84     | 6      | 64   | 3            |
| 7      | 懷孕 | 83     | 12     | 43   | 4345         |
| 8      | 他們 | 81     | 25     | 27   | 5            |
| 9      | 小孩 | 79     | 3      | 86   | 528          |
| 10     | 女性 | 77     | 13     | 43   | 244          |
| 11     | 男性 | 75     | 16     | 32   | 862          |
| 12     | 同意 | 75     | 15     | 34   | 788          |
| 13     | 如果 | 71     | 14     | 42   | 9            |
| 14     | 因為 | 65     | 22     | 29   | 7            |
| 15     | 辛苦 | 60     | 19     | 29   | 1578         |
| 16     | 平等 | 58     | 10     | 46   | 2207         |
| 17     | 男女 | 57     | 9      | 49   | 1082         |
| 18     | 男生 | 54     | 7      | 59   | 2403         |
| 19     | 所以 | 53     | 21     | 29   | 37           |
| 20     | 戰爭 | 52     | 39     | 18   | 1030         |

【表 47】評論短文用詞高中、高職比較

| 與詞頻表比對        | 1-10000 | 10001-20000 | 20001-30000 | 30001-40000 | 40001-50000 | 50001以上 |
|---------------|---------|-------------|-------------|-------------|-------------|---------|
| 高中評論-高職未使用的詞目 | 660     | 169         | 69          | 33          | 27          | 32      |
| 評論-高職未使用的詞目比例 | 66.67%  | 17.07%      | 6.97%       | 3.33%       | 2.73%       | 3.23%   |
| 高職評論-高中未使用的詞目 | 274     | 55          | 39          | 10          | 12          | 15      |
| 評論-高中未使用的詞目比例 | 67.65%  | 13.58%      | 9.63%       | 2.47%       | 2.96%       | 3.70%   |

### 伍、創作短文用詞統計

分別統計高中、高職5-6級分評論短文用詞後，再與〈詞頻總表〉進行比對。並以〈詞頻總表〉為基準對照比較高中及高職學生的用詞結果，其中符號【#N/A】表示比對後未出現符合的詞。以下列舉【表48】本次施測高中評論短文前20個常用詞為例，可以發現在高中創作序號1-10的詞目中，高職排序亦略相符合。其中「等待」、「心情」兩個詞目，因本次施測試題為「等待的心情」，故在高中及高職評論短文中使用的頻率較高。

【表 49】以〈詞頻總表〉為基準，比較創作短文高中（高職）使用，但是高職（高中）未使用的語詞，並依〈詞頻總表〉的序號統計。結果顯示這部分的語詞多集中在1-20000序號內，高於20001序號外高中或高職的比例相差不多。

【表 48】高中創作詞頻序號1-20

| 高中詞頻序號 | 詞目   | 高中創作頻次 | 高職詞頻序號 | 高職詞頻 | 比對<br>〈詞頻總表〉 |
|--------|------|--------|--------|------|--------------|
| 1      | 1199 | 等待     | 1      | 743  | 1056         |
| 2      | 437  | 心情     | 2      | 286  | 585          |
| 3      | 203  | 時間     | 3      | 129  | 28           |
| 4      | 175  | 自己     | 4      | 116  | 4            |
| 5      | 152  | 我們     | 8      | 60   | 1            |
| 6      | 149  | 期待     | 5      | 91   | 1077         |
| 7      | 138  | 因為     | 7      | 64   | 7            |
| 8      | 101  | 個人     | 6      | 75   | 181          |
| 9      | 94   | 不同     | 10     | 53   | 18552        |
| 10     | 83   | 結果     | 13     | 45   | 112          |
| 11     | 80   | 可以     | 9      | 57   | 3            |
| 12     | 77   | 什麼     | 11     | 50   | 13           |
| 13     | 61   | 總是     | 15     | 43   | 667          |
| 14     | 61   | 知道     | 12     | 47   | 38           |
| 15     | 61   | 沒有     | 17     | 40   | 6            |
| 16     | 56   | 緊張     | 14     | 43   | 699          |
| 17     | 54   | 成績     | 51     | 20   | 416          |
| 18     | 53   | 還是     | 29     | 27   | 322          |
| 19     | 52   | 一樣     | 19     | 36   | 88           |
| 20     | 51   | 感覺     | 23     | 32   | 174          |

【表 49】創作短文用詞高中、高職比較

| 與詞頻表比對        | 1-10000 | 10001-20000 | 20001-30000 | 30001-40000 | 40001-50000 | 50001以上 |
|---------------|---------|-------------|-------------|-------------|-------------|---------|
| 高中創作-高職未使用的詞目 | 1189    | 454         | 237         | 114         | 65          | 106     |
| 創作-高職未使用的詞目比例 | 54.92%  | 20.97%      | 10.95%      | 5.27%       | 3.00%       | 4.90%   |
| 高職創作-高中未使用的詞目 | 330     | 148         | 63          | 38          | 27          | 43      |
| 創作-高中未使用的詞目比例 | 50.85%  | 22.80%      | 9.71%       | 5.86%       | 4.16%       | 6.63%   |

## 陸、〈詞頻總表〉未收錄詞

與〈詞頻總表〉比對，並初步篩檢出多音節的詞及短語後，以高中創作舉例如表【表 50】。可以發現無論在評論或創作，均有未列於表上的詞。其一顯示新的語詞不斷出現：亦即部分語詞中出現了與高中高職學生的生活息息相關的語詞，如「基測」、「學測」、「手機」等。

【表 50】〈詞頻總表〉未收錄之詞目

| 高中創作詞目 | 高中創作使用頻次 |
|--------|----------|
| 很多     | 35       |
| 不會     | 27       |
| 為了     | 25       |
| 隨著     | 22       |
| 不停     | 19       |
| 有的     | 17       |
| 不能     | 15       |
| 基測     | 15       |
| 考完     | 14       |
| 手機     | 13       |
| 不要     | 12       |
| 一下     | 11       |
| 何時     | 11       |
| 越來越    | 11       |
| 不可能    | 9        |
| 不已     | 8        |
| 太多     | 8        |
| 更加     | 8        |

另外，有部分語詞在日常中可見，如：「不會」、「不能」等。而〈詞頻總表〉為《國語辭典簡編本》之依據，值得關注是這些常見詞在《國語辭典簡編本》中部分是可查詢得到其解釋意涵的，但並未列於〈詞頻總表〉中，顯示詞頻的調查並未隨著辭典的編修而更新。

由於〈詞頻總表〉調查目的為符合國民中小學生及一般社會人士的語文使用實況，然自民國86年迄今並未能逐年修訂，可能會影響字辭典的完整及共時性，間接影響課程內容對於語文教材詞彙的擴充。建議教育部應延請專家學者，針對〈詞頻總表〉再行修訂調整，以提供教學者和學習者更完善的學習平臺。