

四川省立教育科學館

專題研究報告

常用字選

第一冊
第一冊
第一冊
第一冊
第一冊

立人

陳行可題

目 次

序

常用字研究

常用字選

贈送

序

現代的美國，是一個民主政治，自由社會的國家。美國建國不過百餘年，所以能達到這個階段，我們從美國歷史，知道美國有下面幾件事的努力：

(一) 一七八〇年以後，各州努力強化教育基礎。

(二) 一七八二年紐約的克林頓州長致力普遍教育運動。他說：「一個自由國家的高級職位，是公開給每一階段的公民的，故其政府有設立學校，傳播公眾信任所必須的教育的特殊責任。」意即民主的自治政府，需求受過教育的選民。

(三) 同時傑斐遜總統，對於政治的改進，亦稱：「最要緊的，我希望平民教育受到注意，保存適度自由的必要，受到認識。」

(四) 一七八五年土地法的規定，以數百萬畝的公地，分給公立學校。

美國對於廣泛的平民教育，有了以上的措施，於是美國一般的教育，蒙到普遍的發展，民智增高，政治增進，所以到了現在完成了世界上，惟一的民主政治，自由社會的領袖國家。

我們要學習美國，完成中國的民主自由。我們應該看看我們的環境，和明白我們應負的責任，與努力的方向。

我們的國家現在情形是這樣：

(一) 不識字的文盲占全國百分之八十以上，假若要實行民主政治，實到不可能之事。

(二) 目前社會人士高懸民主口號，但對於掃除文盲的工作，毫不在意，捨本逐末，莫此為甚，身知一羣民衆，既不識字，又無知識，民主政治自然難實現。

(三) 我國抗戰幸得最後勝利，目前建設，千絲萬縷，惟掃除文盲實為一切建設之基本工作。

(四) 就四川省說為抗戰根據地，要於抗戰貢獻極大，有賴於教育以發使命至重難鉅。中央將四川定為建設實驗區，一切建設自應迎頭趕上，掃除文盲的工作，當然不容置人後。故四川省今後施政方針早經省府屢次宣佈以掃除文盲為第一要項。

(五) 掃除文盲為一艱巨偉大之工作，其實施方案，必須由專家擬訂於前，其教材教法，又必須經專家研究於后，每一步驟，每一問題必須用教育科學的方法解決，然後才能收事半功倍之效。

因此行可於去歲承省府聘請主持川省教育科學館之後，深感掃除文盲問題之重大，特請教育心理學專家蔡樂生博士商訂掃除文盲計劃（見四川教育通訊第三期）蔡氏為國際知名之學者，曾任美國大學教授，蜚聲海外，其漢字心理之研究已有二十年之歷史，故請其專題研

究選擇常用字彙作為掃除文盲之工具，此種艱巨之偉大工作，過去有人費十餘年之工夫而無成就者，蔡氏本其二十年來之心得，用謹嚴之科學方法，朝夕研究，并經本館劉鶴笙唐宗郭陳儒蘭等十餘人協助統計，耗時一載即告成此項工作。其報告內容，字字有所根據，讀者自能見到，無庸贅述。竊以掃除文盲之偉大工作，目前已成為國際問題，南美印度及世界其他各國亦正在開展此項工作，蓋全世界四分之三之人類尚不識字，欲求世界永久和平，殊非易事。蔡氏此種研究，其內容須屬國字，其方法實可應用於其他各國，對於國家建設與世界文化，貢獻誠匪淺鮮，故於其報告付印之日，聊綴數語，是為序。

陳行可 二十五年八月一日

常用字研究

我國的教育很不普及，文盲的百分數非常之高，一般民衆的知識水準極低，他們多數既然不能讀書閱報，而且對於國家的事情，社會的動態，和世界的走勢總是从透澈了解，當然對於高深的學術和科學的常識亦都無法研究和吸收。因此我們要實行民主政治和建設一個現代化的國家，實在是一件不可思議的事。所以建國第一步最重要的工作就是在普及教育掃除文盲。本來教育是提高學術水準和普及基本常識兩方面，在民主國家裏，後者似乎比較前者更為急迫和重要；何況登高自卑，高深的學術研究是要從基本的知識入手的。教育越普及，從事高深研究的人亦越多。教育完全普及之後，整個民族的精神力量才能全部發揮起來。普及教育的方法固然包括電影播音展覽表演等，但文字書為求獲知識最基本的工具和最重要的媒介。

有人認為我國教育不普及文盲難於掃除的主要原因是中國文字太過艱難。西洋人學習漢文時常感覺到『漢字難學，難認，難記，難讀，難懂，難譯，難拼，難查，難寫，難打』。拿破崙的字真真一難字，而中國的字真真似乎每個都是難字！俗稱我國的文字為單音制的視覺文字，而西洋的文字為拼音制的聽覺文字。單音制的視覺文字，音無專形，一字一形，千萬字千萬形，每個字必須單獨學，單獨認，單獨記，不像西洋拼音制的聽覺文字，只須學會廿六個字母和拼法，便會看個任何字都能讀，想寫任何字都能寫。記得有一個故事說中國有一位牧師懂得一點英文，廿六個字母固能背誦，但因程度太差，普通英語，只能聽而不能說，教會為造就人材起見，派他到美國研究神學。每逢禮拜天總是到美國禮拜堂做禮拜。有一次美國牧師講道完畢，為表示尊敬中國牧師起見，請他登臺講道，他想自己的英文程度太低，辭不達意，但自己既係牧師，新講事不能推却，所以不得已勉強登臺，閉眼俯首，開口背誦英文ABCDEFGHIJKLMN OPQRSTU VWXYZ 廿六個字母，末後一句用中國話說：『上帝啊，請你自己去拼罷！』因為一切英文字句，無非都由廿六個字母拼綴而成的！因此有人主張完全廢除漢字，採用拼音制內注音符號，國語羅馬字，或拉丁化新文字。但是漢字有悠久的歷史，而且通行全國。上下數千年，縱橫千萬里，漢字無疑地是統一我們中華民族的一個最大因素，因為各地的方言不同，拼出來的音彼此不能會意，寫出字來，大家才能彼此了解，可見語言統一非拼音制文字的先決條件。我們應該加速發展交通，運用播音，電影，語劇等促進語言統一，然後才談得到廢除漢字，全盤改用拼音制的文字。改後如何繼承歷史的文化，發掘古書的寶藏，和對於寫讀科學的專著是否適用，都是值得研究的問題。所以我們覺得漢字一時尚不致於廢除，而注音符號和國語羅馬字的功用，目前恐只在準確地注音漢字而已。其實漢字不一定難學難認難記難讀，況且視覺文字有它的優點。根本文字就是感覺的，語言才是聽覺的，現代心理學證明默讀比朗誦的效率寫得多。所以我們目前的問題並不是漢字的存廢問題，而是漢字的教學問題。教學如果得法，事半功倍，文盲掃除，教育普及，然後民主政治才能實現，現代化的國家才能建設出來！

有人以爲中國字的數量太多，學不勝學，記不勝記，難怪我國文盲的百分數如此之高。但我國字數究竟有多少？我們似應實地調查，切實解答。兩千年來漢字數量遞增的情形，我們可從歷代的字彙看出大概。

表一 歷代字彙遞增長

書 名	編 著	時 代	字 數
倉頡篇	李 斯 等	秦	三，三〇〇
說文解字	楊 雄	漢	五，三四〇
廣雅	許 慎	漢	九，三五三
類編	李 登	魏	一一，五〇〇
廣雅	張 揖	魏	一八，一五〇
玉篇	顧 野 王	梁	二二，七二六
廣韻	孫 愐	唐	二六，一九四
韻海鏡原	顏 真 卿	唐	二六，九一一
類篇	司 馬 光	宋	三一，三一九
字彙	梅 膺 祚	明	三三，一七九
正字通	張 自 烈	明	三三，四四〇
康熙字典	張 玉 書 等	清	四二，一七四
中華大字典	歐陽詢春等	民國	四四，九〇八

我們如果把上列各字彙編著的時代化爲公年，然後畫成一條曲線，便可立即看出兩千年來我國字彙遞增之速率是極有規則的。大概字數的增加和歷史的歲月成爲一個正比例。這樣一條曲線代表一個偉大民族在兩千年中文字數的發展情形，在世界上恐怕找不出第二條曲線像它那樣年代悠久和有那麼多的準確材料作爲根據。

我們既發現這種正比例的關係，便要進一步，研究我中華民族兩千年來字彙遞增的過程是否和兒童學習說話字彙增加的過程相彷彿。關於我國兒童學習說話字彙的發展，國人還沒有系統的研究，我們現在只得根據美國心理學家 M. B. Smith 測量美國兒童所得的結果畫成曲線，以便比較。

表二 美國兒童學習說話字彙遞增表

兒童年齡 (以月計)	學話字彙 (平均數)
18	22
24	272
30	446
36	896
42	1,222

48	1,540
54	1,870
60	2,072
66	2,289
72	2,562

上面種種曲線的形狀十分相似，一個民族和一個兒童獲得字彙的過程竟不謀而合，這點發現，從科學的眼光看來，很有興趣。

綜計我國目前漢字的數量，應有盡有，不外乎四萬五千個字。這個數目還比不上英文字彙的一半，可見我們的字數並不多。況且我國文字經過數千年的新陳代謝，其中一部分已隨時代變遷，在社會上應用日罕，此外還有許多同音，同義而異形的字體，古體，正體，俗體等重複字，所以目前通用各種字典所收的字數平均不過九千零七十二個字，只占總數五分之一而已。

現代通用各種字典字數比較表

(徐則敏)

字典	編者	書局	字數
1 平民字典	馮 瀛	商務	4,431
2 小學國語字典	馮 俊 如	中華	5,832
3 注音國語學生字典		三民	6,985
4 中華新式字典		中華	6,987
5 中華注音國語字典	孫 繼 懋	中華	7,183
6 學工字典	陸 爾 奎	商務	7,247
7 新體國語大字典		世界	7,825
8 新體學生國語字典		世界	7,970
9 注音符號學生國語字典	馮 國 音	世界	7,970
10 國語學生字典	陸 衣 言	中華	8,290
11 新國語學生字典	陸 衣 言	中華	8,435
12 四角號碼學生字典	王 雲	商務	8,643
13 校改國語學生字典	方 毅	商務	9,035
14 縮小新字典	陸 爾 奎	商務	9,586
15 詞性分類新式中國字典	黃 鏡 瀛	商務	10,021
16 中國萬字字典	沈	中華	10,482
17 校改國音字典		商務	11,902
18 實用學生字典	陸 爾 奎	商務	13,469
19 新體學生大字典	秦 同 培	世界	14,175
20 五身五下字典	王 雲	商務	15,111

平均字數 9,072

但是字典所以廣從生字之用，所文字數決非一般民衆所應認識。我國常用的字究竟數量多少？我們又得查問調查，非下一番苦工實地考察研究不可。在五十年前的我國有兩位著名學者羅振玉和錢坫曾與周壽 Villiers de Villehervé 的專心研究，每日苦勞，統計我國當時常用之漢字總數，並將此字常用之普及，請其並表印刷局所承印的刊物一百一十一萬六千四百七十三個字中尋出五十一萬三千個不同的漢字，按此字常用之次第列成等次。這項大規模調查統計台工，實爲全球是所罕見『常用字表』的嚮矢。因此我國常用字表的編定應正仿法編定。我國之書，可惜這種有價值的工本，結果只供印刷局安排字位之用，不能引起社會一般人士的注意，以及普及民間。廿年前陳望學撰『白話書經，紅樓夢，白話等韻等』，編定語體文應用字表，從五十二萬四千四百七十八個字中找出四千二百六十一個生字。此外模範漢字的研究員等，現在爲着更於比較起見，把中外人士所編定的常用字彙列成兩表如下：

綜觀上表所列中外各家編訂的漢文常用字彙，最少者到一千餘字，最多者爲五千餘字。又據十三家集字作者所得結果，四書裏面共二千三百二十八個生字；五經裏面另外還有二千四百二十六個生字；十三經全部生字共六千五百四十四個，除去四書裏面常用字或數字九百二十八個之外共有五千六百一十六個不同的字。又據表若會印刷局所得結果新的語體字數共一十七萬三千一百六十四個，裏面

字彙編訂之常用字彙

編訂者	研 究 對 象	生 字 數
洪 深	基本標準	1,100
平 放	平民千字表	1,300
彭 仁 山	民十卷	2,134
陳 人 哲	民衆文件	2,301
周 祖 訓	初小國語教本	2,340
杜佐周，蔣成望	商人文件與讀物	2,358
王 文 新	初小教科書	2,516
教 育 部	小學初級字彙	2,711
杜佐周，蔣成望	民衆文件與讀物	2,774
杜佐周，蔣成望	兒童讀物與作品	3,654
王 文 新	初小高小教科書	3,799
杜佐周，蔣成望	民衆商人兒童文件及讀物	4,117
陳 鶴 琴	白話書報	4,261
放 弘 德	普通讀物	4,389
莊 澤 宣	綜合數字結果	5,262
商務印書館	華文打字機	5,372

外僑編定之漢字常用字彙

編訂者	生字數
W. A. P. Martin	2,000
Walter Hillier	2,000
J. E. Moncrieff	2,000
Adam Grainger	3,756
P. Kranz	4,000
C. H. Fenn	5,000
William Gamble	5,150

李榕綜合前人研究

最常用字彙	占普通語彙百分數
1,000	89
1,500	94
2,000	97
2,500	98
3,000	99
4,000	99.8

不同的字只有二千七百一十三個，舊約總詞守數為五十五萬三千六百六十三個。裏面生字不過三千九百四十六個；新舊約全部聖書共有二十七萬六千八百二十七個字，裏面只有四千一百八十二個不同的字而已。到底社會上常用字一共有多少？洪深的基本漢字乃模仿西洋基本英語辦法，起初擬選千字，後來因為字數太少不敷應用，遂增添百字，湊成一千一百個字。但仍有些常用之字不在此基本字彙之內。洪深的補救辦法是用基本字代替非基本字。例如以「女弟」二字代替「妹」字，以「兒」的走聲「兒」字代替「媳」字，以「海當中的一座小山」八個字代替「島」字，以「那從地裏冒出來的水」九字代替「泉」字。這種辦法即不致引起笑話，亦嫌冗長。最大的毛病是在破壞語言文字的一致性，迂迴而不自然。總之，無論如何，一千一百個字在日常生活上總覺不敷應用。不敷用的一千三百一十九個字亦嫌太少。那麼一般民衆應當學習五千字麼？查華文打字機雖有五千餘字，但常用字彙上不過二千四百零一個字而已。莊澤宣綜合數家字彙雖共有五千餘字，然常用字不過二千八百二十七個，其餘一千二百四十一個為備用字，一千一百九十四個為罕用字。William Gamble九十年前的文言字彙雖有五千多字，然 W. A. P. Martin 於一八六二年（距今八十三年前）早具卓識而編印他的「常字雙千」了。Paulken 的字彙雖有四千字，然計分六級，而以首兩千字為最重要最常用。李榕綜合前人研究結果，發現最常用的一千字只占普通語彙百分之八十九，而最常用的兩千字則占普通語彙百分之九十九。再學多兩千個字（由兩千到四千）不過增加百分之二點八而已。國父孫中山先生對時代的著作「三民主義」，包括民黨民權民生各

方面，範圍既極廣博，內容又極豐富，然據彭仁山氏之統計，其中只有二千一百三十四個生字。可見兩千個字如果選擇得好，在社會上已經大致够用，一般民衆如能學會這兩千個字，那麼文盲掃除，教育普及了。

關於字彙的選擇，有人用客觀的方法，有人用主觀的方法。有的方法是科學的，有的方法是不科學的，根據太平廣記所說，中國著名的千字文裏面的字是這樣子隨意選擇編排的：「梁武帝於大玉書中撮一千字不重者，每字片紙，雜碎無序，召周興嗣，謂之曰，「卿有才思，爲我韻之。」興嗣一夕編綴進上，鬢髮皆白。」像這樣選擇字彙和編排次第的方法是客觀的和不科學的。科學的方法是要客觀地從社會上流行的書報和日常生活上應用的文件，查那些字是常用的，並且按照各字常用的次數排列等第。現在我們把各家選擇字彙的經過，評如下：

一、William Gamble 的教會文言字彙。遠在九十年前，比英美德法和西班牙各國著名的大規模字彙統計還早些，上海長善會印刷局的執事 William Gamble 先生請兩位無名英雄的中國學者實地調查統計當時文言的常用字彙，從四十一百六十六頁的文言刊物（包括全部聖經和該局所印刷的其他二十七種書牘）共一百一十六萬六千四百七十三個字中找出五千一百五十個不同的漢字。Gamble 的目的是在根據各字常用次數的比例去鑄鉛字粒和把最常用的字粒擺近排字工人的手邊以節省該印刷局的人力物力。因為他曉得在西洋的印刷局中常用字母的鉛粒總是擺近排字工人的手邊而且每套活字鉛粒中各字母的數量大概按照下面常用次數的比例：

e	12,000
t	9,000
a	8,500
i	8,000
n	8,000
o	8,000
s	8,200
r	6,200
k	800
q	500
j	400
x	400
z	200

Gamble 把他的字彙按照各字常用的次數列成十五等級，最後一級包括從別處找來比較罕用（見二十五次以下）的八百五十字，湊成整個六千字的字彙。

等 級	不同字數	最低到最高常用次數	用次總數
1	18	10,000—40,000 以上	214,156

2	224	1 000—10,000	568,126
3	16	900—1,000	15,150
4	19	800—900	16,844
5	44	700—800	82,757
6	47	600—700	81,167
7	59	500—600	82,116
8	99	400—500	44,258
9	112	300—400	88,024
10	197	200—300	47,982
11	400	100—200	55,627
12	207	75—100	17,687
13	301	50—75	18,508
14	549	25—50	19,497
15	8,715	1—25	19,036
	6,000		1,156,386

從上表我們可以看出最常用的十三個字占全部用次總數六分之一。第一和第二等級只有二百三十七個字共見七十七萬七千二百八十二次占全部半數以上。頭八個等級共五百二十一字，見九十四萬九千零七十四次占全部十一分之九。第十五等級（除去不該列入的八百五十字）共二千八百六十五個字見一萬九千零三十六次只占全部六十分之一。第十四十五兩個等級共三千四百一十二個字見三萬八千五百三十一次占全部三十分之一。第一等級中「之」字一字見四萬二千零六十八次，比第十四十五兩等級中三千四百一十二個字的用次還多。

Gamble 的字彙雖共有五千個字，但是如果只包括頭十四等級，那麼只有二千二百八十五個字。如果把見五十次以上的字才算常用字，那麼頭十三個等級只共有一千七百三十八個字，所以 W. A. P. Martin 折衷選擇前面所採用的兩千個字編了一本『常用字雙千』於一八六三年發表。我們對於他們兩家共同的批評是（一）基督教的刊物不能代表社會上一般的書報文件，（二）九十年前的文字字彙和目前夾有語體文的通俗字彙不同。我們一看 Gamble 字彙中最常用的十三個字（不之人以或非於曰爾爾者而耶）便知道他的字彙現在不大適用了。

二、陳鶴琴的語體文應用字彙。自從民國九年開始，陳氏和他的助理九人先後費了二三年的工夫，檢查了九十萬二千六百五十八個字。這個研究分兩次作成功的。初次用六種材料計有（一）兒童用書（二）新聞報（三）雜誌（四）小學生課外作品（五）古今小說（六）雜類等，共五十五萬四千四百七十八個字。其中有四千二百六十一個異樣的單字。那時誠恐搜羅的材料不廣，所以繼續增加三十四萬八千八百八十個字的材料，求出與第一次不同的單字四百五十八個。惟第二次的材料於民國十二年冬東南大學口字房遭火時燒毀，故陳氏於民國十七年六月初版的『語體文應用字彙』只是第一次的研究報告。

據陳氏自己的批評，這個研究還有兩個缺點。第一，所搜集的材料不廣，第二，所彙集的字數不足。材料不廣，所得的字有所偏，字數不足，所得之字又不能稱為完備。要補救這兩個缺點，非再廣收材料繼續研究不可，邱椿和杜佐周等批評陳氏「於採字時曾採用了一些描寫貴族生活（如紅樓夢）和西洋古代生活（如聖經）的作品，對於現在中國的平民生活既不相干，同時對於一般兒童的需要也覺不能適應。」但我們站在第三者客觀的立場來檢討，陳氏搜羅紅樓夢的材料只有四回，連擇選幾段新舊兩約、聖經一共只占全部字數百分之六點二二而已。裏面最大部分字還是社會上一般的字，也許可能有些專門描寫貴族生活（如紅樓夢）和西洋古代生活（如聖經）的特別字樣，但陳氏卷首的凡例第三條說得很清楚「姓名，地方，以及一切其他專門名詞，概不收入在內。」後來於弘德繼續陳氏的工作，從四萬六千八百四十七字材料中求出單字二千三百三十九個，較陳氏還多七十八個單字。這也許是因為他對於姓名地名等不加放棄的緣故，但杜佐周等認為「雖然他所根據的材料的内容，沒有陳先生那種毛病，但是他僅從五種材料，四萬六千餘字中，求得四千三百餘字，這在材料根據的分量上，覺得似乎有點不大充分。」

三、杜佐周蔣成堃的『兒童與成人常用字彙之調查及比較』，這個報告是廈門大學教育學院研究叢刊的一種，民國二十二年六月出版，他們的目標「但求用最普通最常用的字彙而且在比較兒童和成人的常用字彙是否大體相同，他們利用「既選」課程編製」班上的學生二十左右人，每人分擔一部分徵集材料和分類統計的工作，先後歷時約兩載。從兒童和成人的各種普通讀物和文字作品共二十二萬一千三百五十四個字中求出四千一百一十七個不同的單字。分析起來，兒童方面被調查的字數為十一萬八千八百四十四個字其中三千六百五十四個不同的單字。成人方面分為民衆和商人兩種。民衆方面被調查的字共七萬零九百八十六個單字得二千七百七十四個不同的單字。商人方面被調查的字有三萬一千二百二十八個，其中不同的單字為二千三百五十八個。每個字按照在三方面的文件中被採用的次數多少列成等第，然後又把在三方面發現的次數總和起來計算普通常用字的次數和等第。根據比較的結果（一）兒童最常用的百字與普通最常用的百字，其間相同的數目為八二，不同的數目為一八；（二）民衆最常用的百字與普通最常用的百字，其間相同的數目為六〇，不同的數目為四〇；（三）商人最常用的百字與普通最常用的百字，其間相同的數目為二二，而不同的數目為七八。差異的原因「雖或不免因為成人方面（尤其是商人）所根據統計的材料不如兒童方面之充分，但事實上一般成人的日常活動，普遍總比兒童來得廣，因之，其用字的變化，也比兒童來得大。至於商人，更因其業務上的關係，其日常活動，比之一般民衆既然有所不同，因之，其用字的變化，亦自不免有顯著的差異了。」我們對於杜氏等關於民衆和商人的字彙認為有地方性，例如「廈」字在商人的字彙中列為第十五最常用的字，在民衆的字彙中列為第二十八最常用的字。如果把這些地方性太重的字除去，他們的字彙便更加圓滿了。

四、王文新的小學分級字彙研究。民書局出版。原書因抗戰時期校方無法找到，從間接方面得悉「他輯好幾種小學國語教科書及各省小學兒童的作業卷子，加以分析和統計，其結果：關於作文方面從二十萬七千二百四十六字的材料中，共得單字二千九百五十四個」關

於教科書方面，從三十萬三千九百四十一字的材料中，共得單字四千二百七十九個。」他按小學分級把字彙分爲六個等級。現在把他統計所得小學各級學生作文用字數量列表如下：

年 級	生字總數
2	1,208
3	1,848
4	2,387
5	2,708
6	3,008

因爲原書不在身邊，不好隨便批評。杜佐周等認爲「雖是一種精心研究的結果；但他的對象只限於小學各年級的兒童用字。對於一般成人或平民生活方面，毫無涉及」。

五、Paul Kranz 的常用字錄。上海長老會印刷局一九〇三年初版，一九一〇年再版。編者 Kranz 氏爲德國的聖羅基教徒，自一八九二年來華任傳教士，其四千字錄在初版時原分四級，每級一千字，再版時又將頭兩千字分爲四級，每級五百字。連第三千字及第四千字共六級。四千個字的來源如下：（一）第一千字包括 Dr. D. Z. Sheffield 所設計之中國打字機字盤由實地調查所得最常用之七百二十六字加上次常用之一三八六字中的頭二百七十四個，（二）第二千字從 Sheffield 的次常用字中取來減去一二個不太常用的字。（三）第三千字從 X. W. Baller 的新約字典取七百字再加上從 W. U. Soothill 的袖珍字典第二部選來的三百字。（四）第四千字從 Soothill 的字典和 K. Hemeling 的『南京官話』中得來。於此可見 Kranz 的四千字錄來源複雜，有一點東牽西牽而成的。

六、莊澤宣的基本字彙。中華書局出版，工作於民國十九年在廣州中山大學教育學研究所完成。因得彭仁山協助研究，費時間二人合計約在六百小時以上。他們查到六個字彙：（一）陳敦字彙計四千三百三十九字（陳字原得四二六一字），（二）華文打字機所用字計二千三百九十九字，（三）商務出版的平民字典所採字計四千五百餘字（四）西人學華文德人 P. Kranz 所編的常用四千字錄，（五）平民袖珍字典稿本所採的計三千八百八十二字，（六）王文新編定的小學各級字彙共得五二六二個不同的單字，分爲常用字二八二七個，備用字一四一四一個，罕用字一一九四個。至於那些字是屬於甲類（即常用）必須識而能用，何者屬於乙類（即備用）在某種情形之下可以遇見面熟，自是難定。「但我們姑且用主觀的方法假定，其中主觀的程度亦各不同，計：多數計彙中有的，有變成通用之可能的，完全由主觀而定的」。「至六種中有一兩種有的字我們都放在丙類（即罕用）裏面。所以丙類的罕用字，祇是曾在五種字彙中見過一種或兩種而就我們主觀看來也不關重要除了讀古書或某種專著寫古典文字外很少用到的，列入本書供研究文字教育的人的參考罷了，並不包含一切罕用的字」。杜佐周等認爲莊氏的基本字彙，「雖然採攝各種字彙而成，雖然在內容上有常用，備用，及罕用之分；但他的方法，却是姑且用主觀的方法假定」的「雖然他說「其中主觀的程度亦各不同」，但我們知道「嚴格的科學的研究，是不能「姑且用主觀的方法」的。」

七、教育部的小學初級分級暫用字彙。民國廿四年發布，共二千七百一十一個生字。

爲四年級：第一年級五六六字，第二年級六四四字，第三年級七三七字，第四年級七六兩字。選擇的標準：根據初級小學國語教科書十部，算術教科書五部，常識教科書五部分統計其所用生字，取其總次數較多的字；更和王文新所編的小學分級字彙，陳鶴琴所編的語體文應用字彙，平民教育促進會所編的基本通用字彙，莊澤宣所編的基本字彙，杜佐周所編的兒童與成人常用字彙參照斟酌選定。分級的標準：依據廿部教科書各年級所用生字統計表選取被用次數較多的字，并參照王文新分級字彙各字應列年級，及陳鶴琴語體文應用字彙各字次第，分別排定適用的年級。在教育部身面從事字彙工作的魏冰心先生最近發表他對於這個字彙的批評有二。（一）許多很常用的字不見在這字彙之內，而裏面倒有好些不常用的字應該刪去，（二）各字分級未盡妥善。

八、C. H. Fenn 的五千字的補修字典。一九二六年六月出版。本書可以說是北平華北協合語言學校所用單字卡片的索引。編者在華駐三十餘年，根據平日經驗認爲此五千字足夠外國教士之用，因爲這個字彙包括下列幾本書所有的字：（一）華言學校課本，（二）Baller 的國語初步，（三）Matcoer 的國語讀本，（四）官話指南，（五）新舊約聖經，和（六）Bludget-Goodrich 的聖詩。按照常用的程度，Fenn 把他的五千字分爲 ABCDEI FGHij 十個等級，每級五百字。W. B. Pettis 在一九三二年該書第三版的序文中說道五千元字是用科學方法選擇得來的。西洋人在華面幾年所學和所用的文字材料曾經實地調查過，但調查的方法和各字常用的次數 Fenn 本人并無涉及，總之，這個字彙是專門爲西洋傳教士用的，對於普通一般民衆是否適用，恐怕是一個問題。

九、平教會的平民千字課。平教會的基本字彙共有一千三百一十九個字，據說是實地統計得來的。但是陶知行在陳鶴琴的語體文應用字彙序文中說陳氏這冊報告未付印前已經做了平民千字課用字的根據。其實平教會所辦的平民千字課和市民千字課每種用字都是超過這個基本數字。例如農民千字課初級四本已有生字一〇八〇個，再加上高級兩本六七七個生字，合計有一千七百五十七個不同的生字，超過原來基本數字四百三十九個，超過名稱數字（千字課）百分之七十五以上。至於市民千字課初級四冊中即有一三二八個生字，亦已超過原定基本數目。因此他們又補充三百九十個字，和原來的一千三百一十九個字合計得一千七百零九個字。後來他們認爲基本字彙和補充字彙這個辦法不好，所以索性把原來的基本字彙修訂，加上一百九十個字，擴充成爲一千五百零九個字的平民基本字彙。在這個修改的字彙裏面，他們注重農民日常的需要而不只靠每字常用的次數，因此原來客觀調查得來的字彙現在加上主觀的成分了。

十、J. E. Moncrief 的國語兩千字彙編。一九三八年十月發表。編者積十二年的工夫分析二十二種的材料，其中十一種爲西人學習中國語入門，一種爲教育部編的短期小學課本，十種爲與教會有關各家所編的千字課或字彙。從二萬零三百四十九個字中求得三千二百五十三個不同的字。他根據每字在各家字彙中發現次數的多少和在各種課本中發現次序的前後分別斟酌給予積點。然後把積點比較高的兩千字編爲常用字彙，其次的二百字編爲補充字彙。每個字都有他的積點和次序。我們認爲 Moncrief 採取綜合的方式定出每字的次序這種辦法

在現在是很好的，不過他計算積點的方法(例如根據每字在各課本中發現次序的先後作為常用性和重要性的標準)似乎有一點問題。而且，我們看他從各種課本和字表取材的字數有十二

	材 料 來 源 課 本	積 點 估 計	取 材 字 數	被 採 用 字 數		被 撰 棄 字 數
				常用表	補充表	
1	Hillier 漢文學習法	16-2	1615	1392	49	174
2	教育部編短期小學課本	10-2	1506	1369	22	115
3	AlDrich 實用漢文	10-2	1230	1137	10	88
4	Warren 漢文會話課本	13-2	1000	935	16	49
5	Baller 國語入門	13-2	850	838	6	6
6	金陵華言學校課本	11-2	810	795	6	9
7	McHugh 初步國語課本	8-2	783	765	5	13
8	Amunbson 華西國語捷徑	6-2	680	661	13	6
9	董景安 六百字編	4-1	600	592	0	8
10	華北語言學校課本	7-1	352	352	0	0
11	Mateer 初級國語課本簡編	6-1	300	300	0	0
12	Baller 成語課本字表	5-1	250	250	0	0
13	Fenn 五千字字典	15-3	2500	1942	131	427
14	Kranz 常用四千字錄	8-2	2000	1657	80	253
15	青年協會千字課	5	1000	999	0	1
16	商務印書館千字課	5	1000	979	0	21
17	畢來思基督教千字文	5	1000	1000	0	0
18	芳泰瑞福音習用千字表	5	1000	1000	0	0
19	Grainger 華西國語	3-1	877	825	34	18
20	Lyon 國語最常用字	10-2	500	500	0	0
21	成都教士訓練學校字表	6-1	300	300	0	0
22	Giles, 百字表(第一第二)	4-1	196	196	0	0
		175-1	20,349	18,794	372	1133

家不到一千字(從一九六字到八七七字平均五五八字左右)，五家剛滿一千字，只有五家超過一千字的。所以他綜合的結果，我們可以斷定頭五百字是沒有問題的，第二的五百字也還不錯第三的五百字比較勉強，再往下去，因為原料越來越少，恐怕所選的字就不大準確了。

綜觀上列十種字彙，有的是客觀的，有的是主觀的；有的是自己調查的，有的是綜合多家的；有的是排列次第的，有的是分為等級的；有的是文言的，有的是白話的；有的是成人的，有的是兒童的；有的是農民的，有的是商人的；有的是是一二千字的，有的是四五千字的；有的是為教會外僑的，有的是為普通民衆的。我們相信客觀的方法比主觀的方法來得科學

·綜合多家的結果比一家調查的結果來得準確，排列次第的辦法比分為等級的辦法來得精細，語文兼收的字彙比文言或語體的字彙來得通用，兩千個字的數目比四五千字的數目來得基本，為一般民衆用的比為特殊階級用的來得普通。所以本研究的目的是在採用客觀的方法，綜合各家統計的結果，根據每字平均常用的次數或次第，編排成爲一種民衆通用的兩千字基本常用字彙。此外我們想編幾種特殊的補充字彙，如姓名地名字彙，小學兒童字彙，各界職業字彙，書寫簡體字彙等。我們亦想把我們用客觀方法選來的兩千個字參照各家辦法，分爲幾個等級。同時我們已經着手編纂一本常用詞典，包括現代流行的詞彙和術語。這些都要等到工作完畢後才陸續報告。本篇不過是這一類研究的第一種報告而已。

在上列十種字彙中，按照每字常用的次數或讀點的多寡排列次第的，計有陳鶴琴的語體文應用字彙，杜佐周蔣戈堃的兒童與成人常用字彙之調查及比較，和 Moncrief 的國語兩千常字編。因此我們頭一步的工作就是把三家共有的字選擇出來，然後將每字在各家字彙中所占次數的數目加起來，再根據總和次第的高低排列成爲一個綜合的兩千常用字彙。這個字彙編成之後，我們發現好些常用字，因爲只是兩家共有的，所以不能入選；反過來，有些在陳杜兩家的字彙中所占的次第很低（三千以後）只要亦在 Moncrief 的字彙之內（包括補充字彙共二千二百字）便居然列入這個綜合的常用字彙。原因是 Moncrief 的字彙太短，後面那些字又因取材不廣不頂準確（見前文），所以有這種限制。後來我們索性直接採取兩家常用次數的總和作爲標準，依次排列，如果總數相等，則根據數家平均次第定其先後。這樣做法，結果準確得多了。因爲兩家合起來共有七十七萬五千八百三十二個字的材料，從這裏面按照客觀發現次數的多少，選出兩千個最常用的字，材料更加豐富方法更加科學，結果自然更加準確了。杜氏等的材料除了幾個有地方性的字（如廈字）外，其他都是兒童，成人，民衆和商人的普通用字。但是我們援引陳氏字彙的凡例，這些字有地方性的概不列入。至於陳氏的材料，有人批評他曾採用了一些描寫貴族生活（如紅樓夢）和西洋古代生活（如聖經）的作品，現在我們把他全部的材料仔細分析一下。

材料種類	字 數	百 分 數
1. 兒童用書類	127,298	22.98
2. 報刊類	153,344	27.65
3. 雜誌類	90,142	16.25
4. 小學生課外著作類	51,807	9.34
5. 古今小說類	71,267	12.85
6. 雜類	60,625	10.93
	554,478	100.00
第五類中紅樓夢四回	24,086	4.34
第六類中聖經一部分	10,453	1.88
	34,519	6.22
陳杜材料共 775,882	34,519	4.45

從上表我們看出陳氏從《樓夢》中取材四回共二四〇六六字，占百分之四點三四，從聖經取材一部分共一〇四五三字，占百分之一點八八。兩種材料共計三四五一九字，在陳氏兩家綜合材料總數七七五八三二個字中，只占百分之四點四五而已。在這個百分數很低的材料中，縱然有些描寫貴族生活和西洋古代生活的字眼，亦因陳氏不收專門名詞的凡例而被棄，何況我們只從陳氏字彙中四千二百一十六個字選取最常用的一千九百餘字（本字彙有些字爲陳氏字彙所無）其餘一大半的字假設都是描寫貴族生活和西洋古代生活的比較不常用的字，亦和我們的字彙無關。當然，最理想的辦法是自己重新調查統計數百萬字的現代材料，但是這恐怕是時間、精神和經濟所不容易許可。經過十年的工夫調查統計出來的結果，到那個時候工業發達，社會情形改變，這樣辛辛苦苦得來的字彙，是否還是全部照樣常用。我們急於掃除文盲，決不能多等待十年。我們相信我們這個綜合的基本字彙可以適應目前迫切的需要，而且百分的九十以上是不致於受改變的，現在把本字彙二千個字依照次第或部首筆劃分別列爲兩表附錄於后。

常用字的數目，選擇，和排列既經確定之後，我們應當更進一步從事常用字的分析和教學法的研究工作。我們首先要分析這兩千個常用字的字形，用偏旁、部首等做單位，歸納得多少『字根』，先學會這些『字根』，那麼兩千個字，便類旁通，迎刃而解。其次，我們要分析這兩千個字的聲音。同音的字，彼此相註，同時學習，或依照母音子音，加以分類，編成押韻的詩歌詞句，既便學習，又易記憶。還有，根據國語統一籌備委員會所得結果，標準國語，只共有四百一十一音。現在我們如果選取四百一十一個不同音的最容易學，容易認，容易記，容易寫的常用字，每字代表他所有的一切同音異形的字（例如才字代表材財豺柴裁差猜彩採菜蔡等字，四聲用符號標明），那麼一個人學會了這四百一十一個標準別字或白字，一切的話都能寫成字了。但這些別字要標準化，就是「別字連篇」的詞亦都要標準化，經過教育部立案，承認和頒佈。並且應該編有別字詞典，以備查用。每一個詞連寫起來成爲一個單位，比較容易辨認，因爲我國雖有許多同音異義的字，但是很少有同音異義的詞。吳稚暉先生寫的別字篇似乎有點提倡的意思，可惜到如今還沒有人真真下一番工夫，把別字和別詞標準化，徹底研究，看看這個辦法到底是不是辦得通。當然，在方言不同的地方便要發生困難，但註音符號，國語羅馬字和拉丁化新文字都有同一樣的困難。據說中國農民和一般民衆都有愛好學習中國字的心理，他們認爲注音符號，國語羅馬字和所謂拉丁化新文字都不是中國字，所以不願意熱心去學習，這也許是這些『新文字』失敗原因的心理基礎。用注音符號或羅馬字來註音漢字，幫助讀音準確和促進國語統一，我們是很贊成的；但是拿來代替漢字，目前尚非其時，恐怕辦不通。至於註音的方式我們認爲注音銅模的辦法，在讀物漢字的右邊註上一行擾亂視線和鼓勵依賴性的注音符號，是不合學習心理的原則的，其實只要在每個生字介紹的時候，於生字欄注音便得了。

吳稚暉先生寫的別字篇（詞兒連寫）

「衆話命鍋 低伊柯 糞通 交 巽問； 低雷柯 交 冤師款； 低散柯 交 里怨
關； 低思柯 交 奉果丈」。