



異讀率和字集數負相關現象所反映的《廣韻》異讀成因

2018/12/15 顧國林

【提要】 針對《廣韻》異讀（多音字）的成因，提出一條統計性的規律：異讀率和字集數呈顯著負相關，即在字越多的語音單元裏，異讀發生率越低，根據其顯著程度，可證實“文字的近音借用”是《廣韻》異讀主要來源。

一、方法原理

《廣韻》異讀的成因，學者們各有看法，有方言層次說、詞義派生說、文白讀說、假借說、形態說等。“異讀率和字集數呈顯著負相關”現象顯示，《廣韻》異讀的主體來源是“文字的近音借用”，屬於先秦的用字現象。

文字借用，也稱假借，即用 A 字的音寫 B 詞（如豆本是器皿，後來用豆的音記植物果實）。A 和 B 如果不完全同音，是近音假借。近音假借發生後，A 字就有兩個音，成為異讀。從大眾心理來說，假借應同音字優先，近音假借會給閱讀帶來誤會。

（1）何為“異讀率和字集數呈負相關”？

舉個直白的例子：中古 A 韻轄 100 字，B 韻轄 200 字，如果前者有非 A 韻的異讀 10 次，那麼後者有非 B 韻的異讀多少次呢？根據統計，其平均數量不是平行的 20 次，而是在 15 次左右，有點違背直覺。根據下文的統計，韻的字數越多，該韻的異讀發生率越低（指趨勢）。

這是“近音假借”留下的痕跡。

古人為寫字、記詞尋找記音字時，在字數越少的語音單元裏，越難找到合適的同音字，這時，書寫者姑且使用近音字的發生率越高，對後人來說，異讀率越高。反之，在字數越多的語音單元裏，越容易找到合適的同音字，使用近音字的發生率越低，對後人來講，異讀率越低。

舉個聲調的例子，古代平聲字多，上、去兩聲的字少，古人寫文章用記音字，平聲的詞更有優勢，因為同音字更多。對它們的異讀率統計是：平聲 23.8%、上聲 36.8%、去聲 41.7%，平聲字的異讀最低。

（2）異讀率和字集數的約定

異讀率：異讀次數占總字數的比例。比如中古幫母有 770 字（異讀算一字，下同），存在 206 次非幫母的異讀，那麼幫母異讀率為 $206/770=26.8\%$ 。由於異讀記的是次數，一字可以有 multiple 異讀，所以理論上可以超過 100%。

字集數：指異讀的統計範圍。比如，在“幫母異讀率”中字集數指幫母總字數，在“歌韻異讀率”中字集數是歌韻總字數，以此類推。

（3）相關性計算方法

兩組數據，是否存在相關性（第一組增大時，對應第二組是否也相應增大或相應減少），一般採用皮爾森相關性方法。文中給出所有數據的散點圖和計算結果。相關性不等同因果性，對因果的定性，應當排查所有可能的原因，見第四節。



二、統計分組

(1) 異讀概貌

廣韻有多少異讀字？統計結果是：

廣韻去重總字數為 19499 個（異讀只記一次），其中 14778 字只有一個音，4721 字有兩個或兩個以上音，異讀占總字數 24.2%，即每 4 個字裏就有 1 個是異讀，這是針對字頭的統計。

表 1：《廣韻》字頭異讀數量和比例

廣韻	去重字數	單音字	雙音字	三音字	四音字	五音字	六音字	七音字
字數	19499	14778	3787	784	125	22	2	1
占比	100%	75.8%	19.4%	4.0%	0.6%	0.1%	0.0%	0.0%

(2) 統計分組

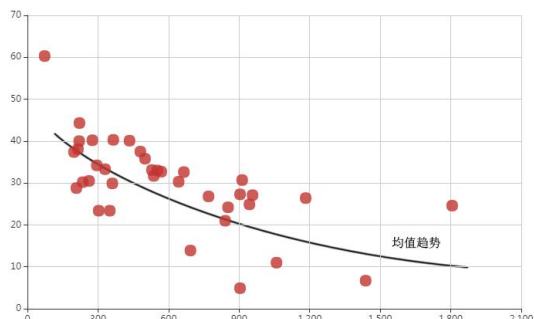
下文統計共分四組，分別為：

- ①《廣韻》37 個基本聲類的異讀率對字數的分佈。
目的：檢測古人記音字的語音準確度對 37 個基本聲類的字數是否敏感。
 - ②《廣韻》206 個基本韻類的異讀率對字數的分佈。
目的：檢測古人記音字的語音準確度對 206 個基本韻類的字數是否敏感。
 - ③《廣韻》240 個語言單元的異讀率對字數的分佈（37 聲類以等、開合、重紐¹展開為 240 組）。
目的：檢測古人記音字的語音準確度對音節前半段是否敏感（模糊對應 skwrans 的 skwr-部分）。
 - ④《廣韻》405 個語言單元的異讀率對字數的分佈（206 韻類以等、開合、重紐展開為 405 組）。
目的：檢測古人記音字的語音準確度對音節後半段是否敏感（模糊對應 skwrans 的-wrans 部分）。
- 《廣韻》雖然是中古音，但語音單元是從上古繼承而來的，音位上仍能反映上古音。

¹ 重紐範圍指傳統的支、脂、祭、真、仙、宵、侵、鹽八類，幫組、見系存在 A 類和 B 類對立，其他無對立的聲類部分，不歸入 A 類或 B 類，而是另成一類，所以本文的重紐韻分為“三種顆粒”。

三、統計結果

圖 1: 37 聲類異讀率 (縱軸) 對字數 (橫軸) 的分佈

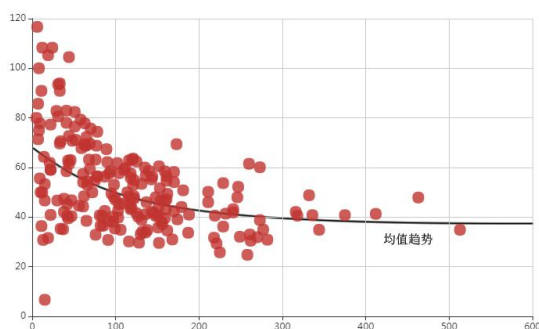


①37 個基本聲類對異讀率對字數的分佈

皮爾森相關性數值: -0.644 (強負相關)

在《廣韻》37 個基本聲類中，隨著每個聲類字數的增加，異讀率呈下降趨勢（負值），絕對值 0.644 顯示“強相關”，表明這種趨勢是非常顯著的。異讀率均值從約 35% 下降至約 10%。

圖 2: 206 韻類異讀率 (縱軸) 對字數 (橫軸) 的下降分佈

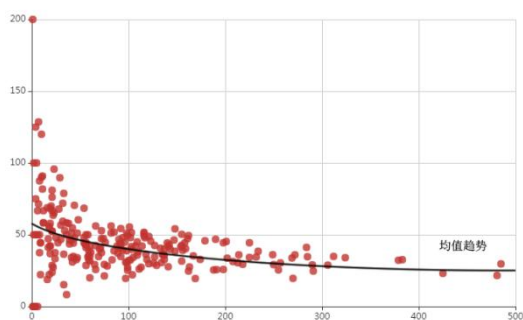


②206 個基本韻類的異讀率對字數的分佈

皮爾森相關性數值: -0.460 (中等負相關)

在《廣韻》206 基本韻類中，隨著每個韻類字數的增加，異讀率呈下降趨勢（負值），絕對值 0.460 顯示為“中等相關”，表明這種趨勢是明顯的，異讀率均值從約 65% 下降至約 40%。

圖 3: 240 語言單元異讀率 (縱軸) 對字數 (橫軸) 的下降分佈

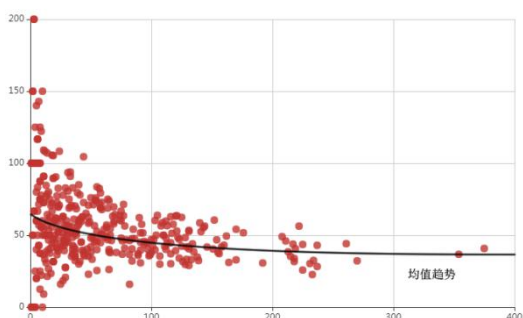


③240 個語言單元的異讀率對字數的分佈

皮爾森相關性數值: -0.313 (負相關)

在《廣韻》240 個語言單元中，隨著每個單元字數的增加，異讀率呈下降趨勢（負值），絕對值 0.313 顯示“相關”，由於 240 個樣本數量可觀，相關性具有很高的可信度。異讀率均值從約 55% 下降至約 25%。

圖 4: 405 語言單元異讀率 (縱軸) 對字數 (橫軸) 的下降分佈



④405 個語言單元的異讀率對字數的分佈

皮爾森相關性數值: -0.299 (負相關)

在《廣韻》405 個語言單元中，隨著每個單元字數的增加，異讀率呈下降趨勢（負值），絕對值 0.299 顯示“相關”，由於 240 個樣本數量可觀，相關性具有很高的可信度。異讀率均值從約 65% 下降至約 40%。



以上四項統計，不管對中古音節進行何種方式的切分，總是存在異讀率和字集數呈負相關現象，可以證實兩點：

- (1) 假借行為會優先找同音字，在沒有合適同音字的情況，會放棄音節中某一個部位的音準。
- (2) 這個被放棄音準的部位是多樣的，可以是聲母，也可以是韻母、聲調（從中古音視角）。

四、異讀率的兩個層次

異讀率的高低，主要由歷史語音關係決定，這是第一層，近音假借是第二層，覆蓋其上，這篇文章探索的是第二層。

異讀率的第一層：以聲類舉例，比如明母獨立性很強，除曉母外很少對外接觸，異地率很低；定母不僅與同類端母、透母頻繁接觸，還和以母頻繁接觸，異讀率就很高。語音關係是否接近，對異讀率的影響是很大，但它對字數不敏感，對字數來說是無規律的。

異讀率的第二層：近音假借是異讀率的第二層，疊加在第一層之上，這個層次對字數敏感，會讓異讀率分佈的均值呈現下降趨勢，在雜亂中形成規律。

五、成因篩查

相關性不全是因果關係，有必要對各種原因進行檢查，看看哪個對解釋“異讀率和字集數負相關”更直接、有效、可信，下面把產生異讀的各種原因羅列如下：

- (1) 權威語影響和文白讀。比如說話用本地音，讀書用通語音，造成一字兩音，如普通話剝、血等。
- (2) 方言接觸。比如兩個語言相鄰，居民頻繁來往，語言混雜，某些字詞會混雜兩種口音。
- (3) 同源詞、詞義派生。比如“種”在種地、種子裏是兩個音，屬同源詞。
- (4) 形態。比如動詞的使動、動用語音不同，但字是同一個。
- (5) 同形字。兩字碰巧一樣，造成一字兩音，如吳語的片（指一片）和古文字的片（床的初文）。
- (6) 訓讀，讀為別的同義詞並成為社會習慣。
- (7) 錯讀，即聲符誤導，如普通話的溪、墟、譜等。
- (8) 其他數量較少的音變，如避諱、詞間感染音變、高頻詞弱化、小稱兒化等。
- (9) 近音假借。

這些異讀的產生原因中，(1) - (8) 都不能解釋“異讀率和字集數負相關”，它們的異讀產生率和字集數沒有明顯的相關性，只有(9)近音假借能解釋這個現象。

六、總結

從負相關現象，可看到《廣韻》擁有一個近音假借的層次，從顯著性推測，近音假借層的字數是非常龐大的，龐大到足以影響到圖 1-圖 4 的均值趨勢線的明顯下降，這是判斷《廣韻》異讀主體來自近音假借的主要依據。

從文字發展史來看，近音假借主要發生在先秦。歷史進入春秋戰國時期，書寫開始普及，文字開始記錄風土人情、市井聲響，新詞大量入文，從出土簡帛看，當時的書寫文化是很務實的，文字記音功能強大，假借頻繁，異文突出。假借是對已有漢字的選擇，能否找到合適的同音字，對該語音單元內的字



數是非常敏感的，會形成“異讀率和字集數呈負相關”的統計規律。

《廣韻》異讀主體來自近音假借，這個事實對上古音會產生什麼啟發呢？下面簡要談談。

(1) 《廣韻》的眾多異讀記音是可靠的、可信的，它們大多數是上古遺留，不能推給方言，其主體是先秦音，通過歷代讀書人繼承（雖然他們並不一定知道為什麼要這樣讀）。

(2) 《廣韻》對異讀字的釋義大部分無別，這是不可靠的。在中古時代，它們的假借面貌已被遺忘，異讀對應的準確語義失收。

(3) 假借型異讀的上古擬音應遵循“聽感上比較相似”的原則，還原記音字的樸素追求。

(4) 文字假借的一個副作用是：很多字發生了借而不還的現象，本音本義失傳，特別是筆劃簡單的常用字、部件字，不能輕信《廣韻》和《說文》。

(5) 假借也會造成“語音分化”的假像，一個很單純的韻部，經過較多近音假借之後，會造成“分化”的假像，讓人產生“方言層次”的錯覺。

(6) 上古異讀更發達，《廣韻》失收字音是情理中事，甚至失收本音也很正常，在擬測古音時，如果遇到細節上的參差和矛盾，應當把失收也考慮進去。

參考資料：

- 《宋本廣韻》 江蘇教育出版社 2008 年
李方桂 1980 年 《上古音研究》 商務印書館
鄭張尚芳 2008 年 《上古音系》 上海教育出版社
潘悟雲 2000 《漢語歷史音韻學》 上海教育出版社
周法高等 1975 年 《金文詁林》 香港中文大學
張世超等 1996 年 《金文形義通解》 中文出版社
白於藍 2017 年 《簡帛古書通假字大系》 福建人民出版社