

## 第四章 結果分析

本研究採用語料庫語言學的研究方法，探討台語詞彙在學術和非學術書面語的使用差異。建立學術類語料 101,349 音節，非學術類語料 102,335 音節，合計 203,684 音節，經過電腦斷詞、人工校正、詞頻統計後，得到學術詞頻統計表與非學術詞頻統計表各一份，做為分析台華共通詞、台語詞彙豐富度、台語羅馬字詞彙、台語平均詞長等四項台語詞彙特性的使用差異。分析的結果分成四節做說明：第一節台華共通詞的使用差異分析；第二節台語詞彙豐富度分析；第三節台語羅馬字詞彙分析；第四節台語平均詞長分析。

### 第一節 台華共通詞的使用差異分析

本節分成詞型、詞次、詞型與詞次比較分析、小結四個部分討論台華共通詞在學術與非學術書面語的使用差異，在詞型和詞次的部分再細分成覆蓋率 100%、覆蓋率 80%、不同覆蓋率的比較以及小結四個項目討論。

#### 一、詞型分析

分成覆蓋率 100%、覆蓋率 80%、覆蓋率比較分析、小結四個部分討論。

(一) 覆蓋率 100%

表 20 台華共通詞與台語特別詞使用差異統計表 (詞型, 覆蓋率 100%)

項目	學術類		非學術類	
	詞型	比例	詞型	比例
台華共通詞	4,588	66.91%	5,003.5	55.09%
台語特別詞	2,269	33.09%	4,079.5	44.91%
合計	6,857	100.00%	9,083.0	100.00%

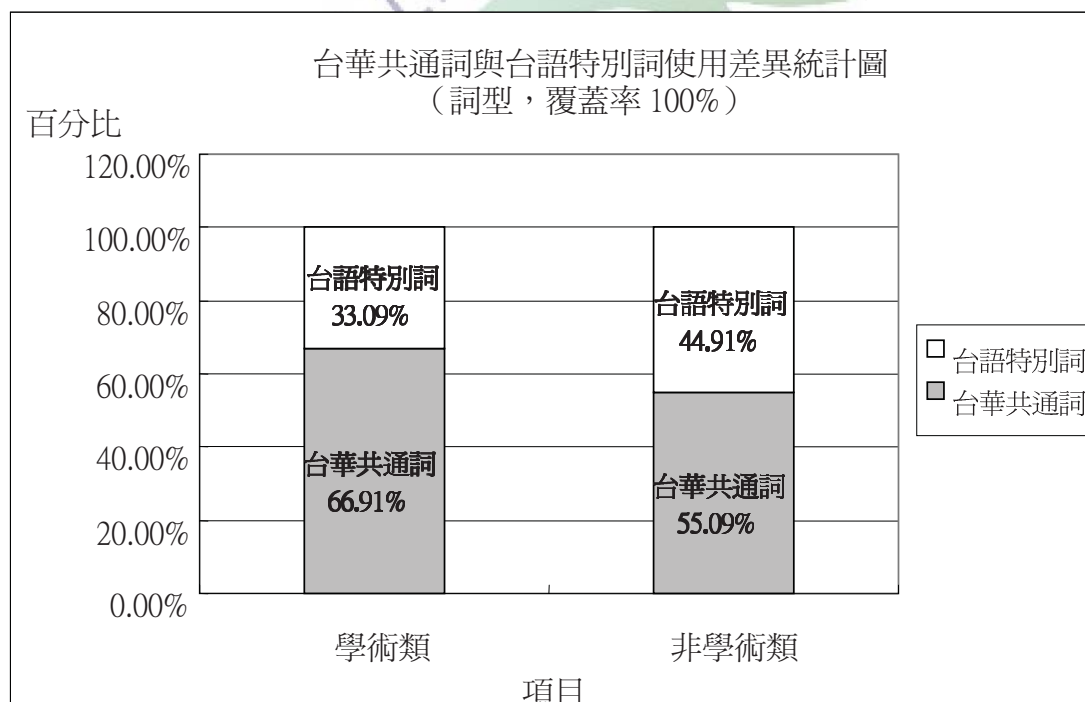


圖 3 台華共通詞與台語特別詞使用差異統計圖 (詞型, 覆蓋率 100%)

從上表和上圖可知：台華共通詞在學術類佔 66.91%，非學術佔 55.09%，

學術類比非學術類多出 11.82%；兩種文類所佔的比例都超過 50%，學術類達到近 70%的水準。統計的結果支持本研究第一項假設：台華共通詞的使用比例，學術類多於非學術類。

(二) 覆蓋率 80%

表 21 台華共通詞與台語特別詞使用差異統計表（詞型，覆蓋率 80%）

項目	學術類		非學術類	
	詞型	比例	詞型	比例
台華共通詞	1,030	82.40%	1,088.5	65.69%
台語特別詞	220	17.60%	568.5	34.31%
合計	1,250	100.00%	1,657.0	100.00%

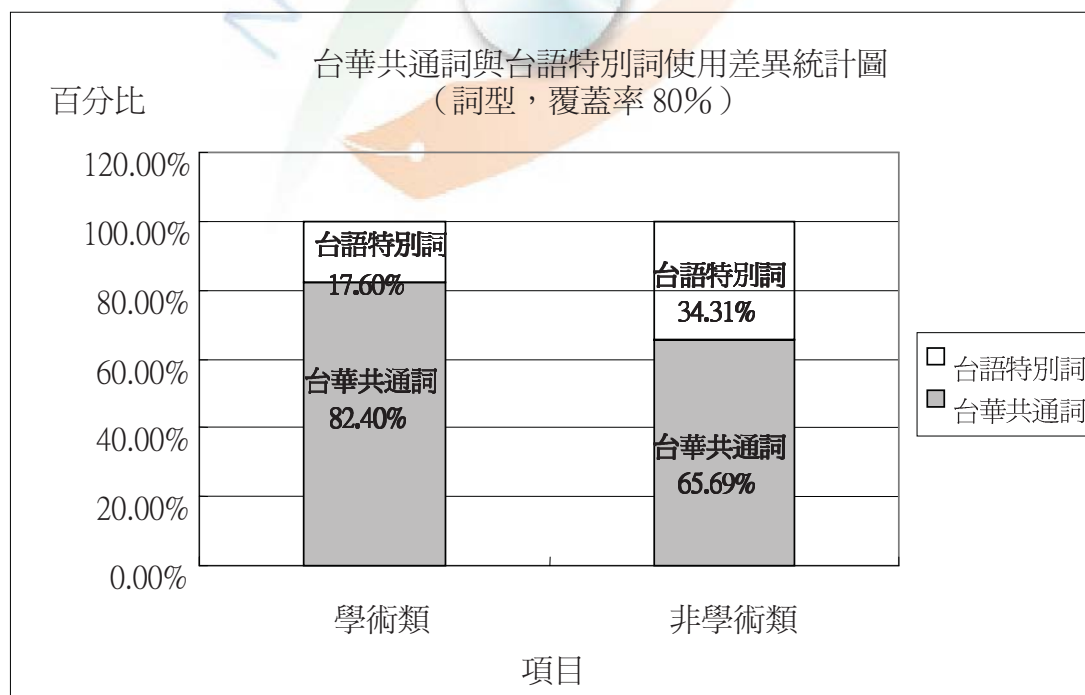


圖 4 台華共通詞與台語特別詞使用差異統計圖（詞型，覆蓋率 80%）

從上表和上圖可知：台華共通詞在學術類佔 82.40%，非學術佔 65.69%，學術類比非學術類多出 16.71%。兩種文類所佔的比例都超過 65%，學術類達到 80% 以上的水準。統計的結果支持本研究第一項假設：台華共通詞的使用比例，學術類多於非學術類。

### (三) 覆蓋率比較分析

本小節透過比較台華共通詞不同覆蓋率的比例，分析低頻詞對統計結果的影響；並與鄭良偉（1987）、黃宣範（1998）的研究比較，探討台華共通詞實際的使用情形。

表 22 台華共通詞覆蓋率 100%VS 覆蓋率 80%比較表（詞型）

項目	覆蓋率 100%	覆蓋率 80%
學術類	66.91%	82.40%
非學術類	55.09%	65.69%

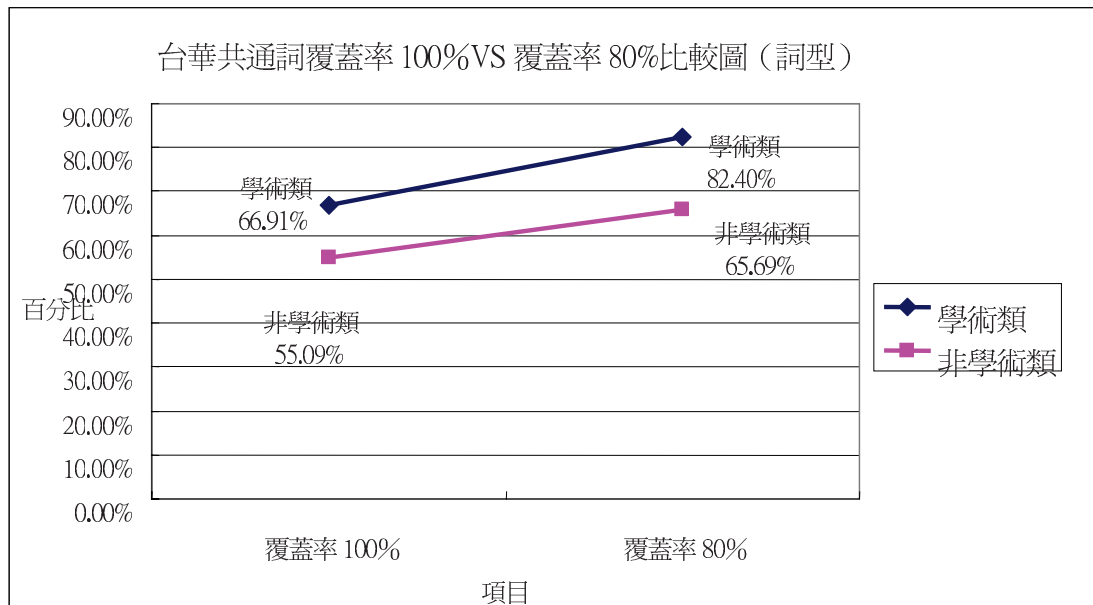


圖 5 台華共通詞覆蓋率 100%VS 覆蓋率 80%比較圖 (詞型)

從上表及上圖可知：排除低頻詞後（即覆蓋率 100%調整至覆蓋率 80%），台華共通詞在學術類的比例由（覆蓋率 100%）66.91%提升至（覆蓋率 80%）82.40%，增加 15.49%；在非學術類的比例由（覆蓋率 100%）55.09%提升至（覆蓋率 80%）65.69%，增加 10.60%。

由上述可知，計算台華共通詞在學術類與非學術類的比例時，低頻詞對於統計結果會產生影響，影響的程度至少超過 10%。可能的原因是低頻詞中的電腦雜訊是以台語特別詞計算，而且以詞型計算，低頻詞和高頻詞的比重是一樣的，因此增加了低頻詞對統計結果的影響程度。

將兩個不同覆蓋率的台華共通詞使用比例與鄭良偉（1987）、黃宣範（1998）的統計做比較，探討台華共通詞的使用情形。比較結果如下表：

表 23 台華共通詞使用比較表（詞型）

項目	鄭良偉	黃宣範	覆蓋率 100%		覆蓋率 80%	
			學術	非學術	學術	非學術
台華共通詞	65.6%	73%	66.91%	55.09%	82.40%	65.69%
台語特別詞	34.4%	27%	33.09%	44.91%	17.60%	34.31%

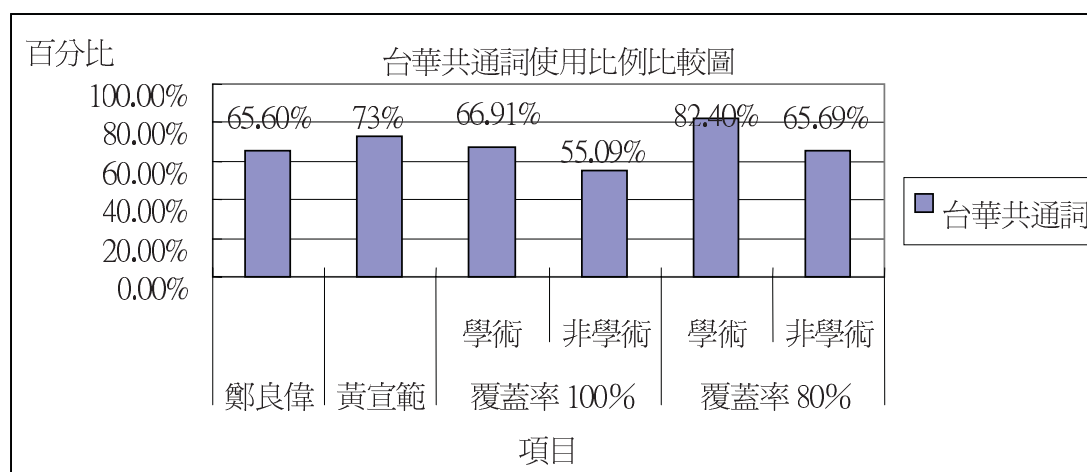


圖 6 台華共通詞使用比例比較圖

從上表及上圖可知：鄭、黃兩位學者台華共通詞的比例約介於 66%~73%；本研究台華共通詞覆蓋率 100%的比例介於 55.09%（非學術）~66.91%（學術），台華共通詞覆蓋率 80%的比例介於 65.69%（非學術）~82.40%（學術）。

鄭、黃兩位學者統計「現代閩南語辭典」台華共通詞是沒有分文類，普遍性的調查。如果以鄭、黃兩位學者的統計為比較基準，學術類的文本對書面語標準化的程度要求比較高，使用台華共通詞的比例應該會高於鄭、黃兩位學者的調查結果；非學術性文本含蓋的文體種類較廣，使用台華共通詞的比例應該和鄭、黃

兩位學者的調查差不多。因此，我們可以歸納出一個簡單的比較方式，了解那一個台華共通詞的覆蓋率比較接近實際的使用情形，為比較方便，以數學符號「>」表示「大於、多於」，以「=」表示「等於」。敘述如下：

由上述的推論可知，學術類文本 > 鄭、黃  $\geq$  非學術類文本是比較接近台華共通詞實際的使用情形，因此以台華共通詞覆蓋率 100% 與覆蓋率 80% 與之比較就可以很容易看出那一個覆蓋率是比較符合實際的使用情形。

根據上述方式比較台華共通詞覆蓋率 100%、覆蓋率 80%，結果如下：

覆蓋率 100%：鄭、黃（66%~73%） $\geq$  學術類文本（66.91%）> 非學術類文本（55.09%）。

覆蓋率 80%：學術類文本（82.40%）> 鄭、黃（66%~73%） $\geq$  非學術類文本（65.69%）。

由上述的比較可知，台華共通詞覆蓋率 80% 的統計結果（65.69%~82.40%）是比較符合實際的使用情形。

#### （四）小結

覆蓋率 100% 和覆蓋率 80% 的統計結果皆支持本研究的假設：台華共通詞的使用比例，學術類多於非學術類。

低頻詞對統計結果支持本研究假設方面沒有造成影響；但是在計算台華共通詞的比例時，對於學術類或非學術類大約有 10% 以上的影響，排除低頻詞影響後（覆蓋率 80%）所得到的結果是比較接近實際的使用情形。

## 二、詞次分析

分成覆蓋率 100%、覆蓋率 80%、覆蓋率比較分析以及小結四個部分：

(一) 覆蓋率 100%

表 24 台華共通詞與台語特別詞使用差異統計表（詞次，覆蓋率 100%）

項目	學術類		非學術類	
	詞次	比例	詞次	比例
台華共通詞	43,301	74.58%	44,289	56.81%
台語特別詞	14,756	25.42%	33,666	43.19%
合計	58,057	100.00%	77,955	100.00%

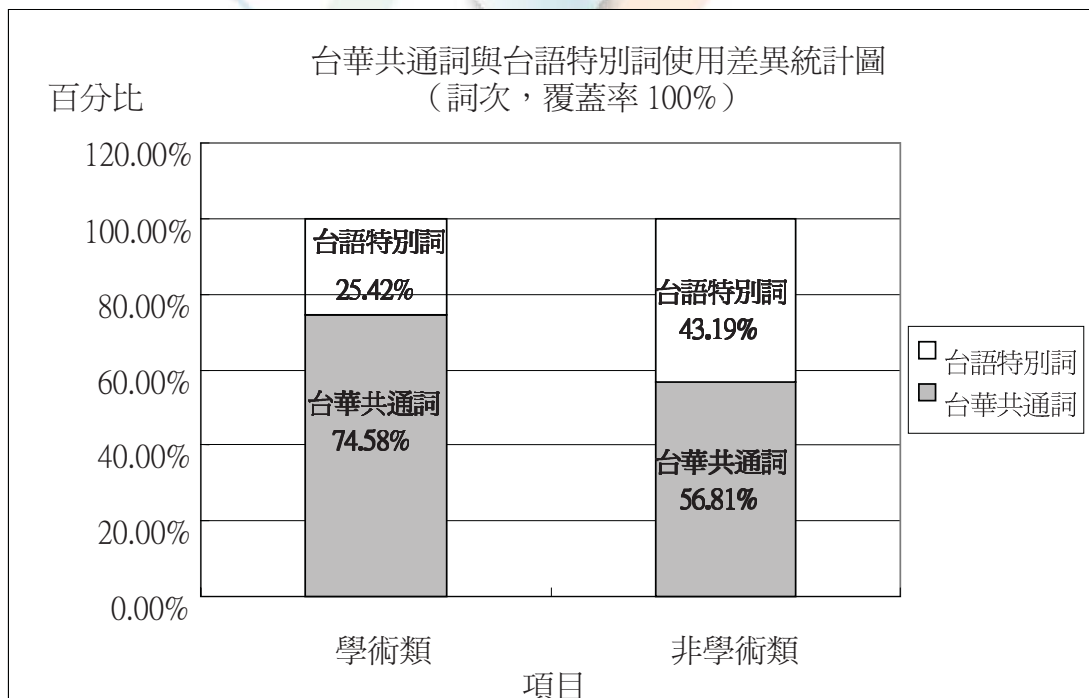


圖 7 台華共通詞與台語特別詞使用差異統計圖（詞次，覆蓋率 100%）



從上表和上圖可知：台華共通詞在學術類佔 74.58%，非學術佔 56.81%，學術類比非學術類多出 17.77%；兩種文類所佔的比例都超過 55%，學術類達到 70% 以上的水準。統計的結果支持本研究第一項假設：台華共通詞的使用比例，學術類多於非學術類。

(二) 覆蓋率 80%

表 25 台華共通詞與台語特別詞使用差異統計表（詞次，覆蓋率 80%）

項目	學術類		非學術類	
	詞次	比例	詞次	比例
台華共通詞	35,469	75.98%	36,197	57.01%
台語特別詞	11,215	24.02%	27,298	42.99%
合計	46,684	100.00%	63,495	100.00%

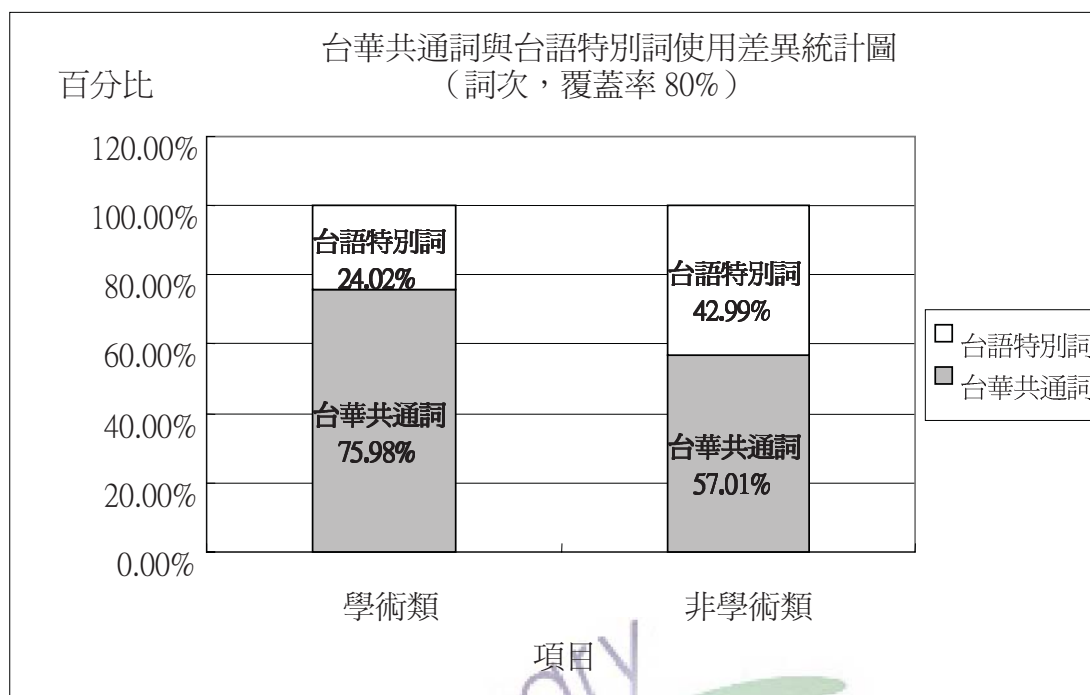


圖 8 台華共通詞與台語特別詞使用差異統計圖 (詞次, 覆蓋率 80%)

從上表和上圖可知：台華共通詞在學術類佔 75.98%，非學術佔 56.68%，學術類比非學術類多出 19.30%；兩種文類所佔的比例都超過 55%，學術類達到 75% 以上的水準。統計的結果支持本研究第一項假設：台華共通詞的使用比例，學術類多於非學術類。

### (三) 覆蓋率比較分析

表 26 台華共通詞覆蓋率 100%VS 覆蓋率 80%比較表 (詞次)

項目	覆蓋率 100%	覆蓋率 80%
學術類	74.58%	75.98%
非學術類	56.81%	57.01%

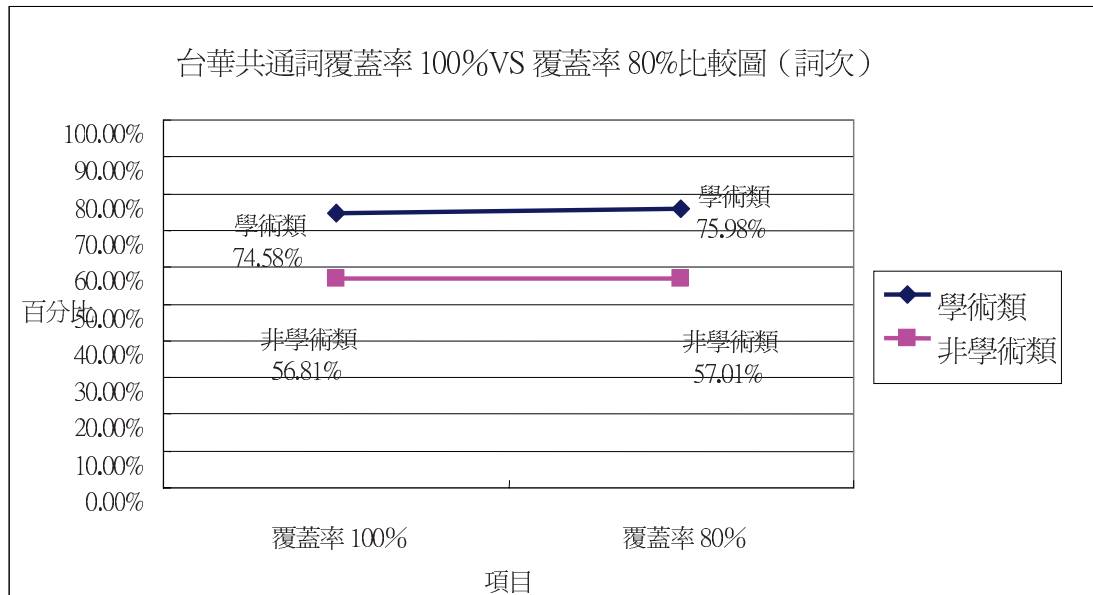


圖 9 台華共通詞覆蓋率 100%VS 覆蓋率 80%比較圖（詞次）

從上表及上圖可知：排除低頻詞後（即覆蓋率 100%調整至覆蓋率 80%），台華共通詞在學術類的比例由（覆蓋率 100%）74.58%提升至（覆蓋率 80%）75.98%，增加 1.4%；在非學術類的比例由（覆蓋率 100%）56.81%提升至（覆蓋率 80%）57.01%，增加 0.2%。

由上述可知，低頻詞在計算台華共通詞的比例時產生的影響很小。可能的原因是低頻詞中電腦雜訊雖然以台語特別詞計算，在總詞次中所佔比例不高，因此產生的影響有限。

將統計結果與 Sander 及 Hsieh（1981）的研究做比較，探討台華共通詞的使用情形，比較結果如下表：

表 27 台華共通詞使用比較表（詞次）

項目	新聞報告 (散文)	可愛的仇人 (小說)	情歌 (詩)	諺語及 俗語	學術 覆蓋率 80 %	非學術 覆蓋率 80 %
共通語層	77.0%	67.5%	67.4%	67.5%	75.98%	57.01%
本土語層	23.0%	32.5%	32.6%	32.5%	24.02%	42.99%
合計	100.0%	100.0%	100.0%	100.0%	100.00%	100.00%

資料來源：Sander 及 Hsieh (1981)；轉引自張學謙 (1998)。

說明：1.本研究之台華共通詞和 Sander 及 Hsieh 研究之共通語層是指同一概念的詞彙；台語特別詞和本土語層亦同。

2.覆蓋率 80%是比較接近實際的使用情形，因此以之做為比較。

從上表可知：Sander 及 Hsieh 新聞報告共通語層佔 77%、可愛的仇人佔（小說）67.5%、歌詩佔 67.4%、諺語及俗語佔 67.5%。新聞報告是屬於比較正式的文體，其他三類屬於比較非正式的文體；共通語層的使用比例，新聞報告（正式）多於小說、歌詩、諺語及俗語（非正式）。因此 Sander 及 Hsieh 的調查支持本研究的第一項假設：台華共通詞的使用比例，學術類多於非學術類。

再來比較各類文體台華共通詞的比例，新聞報告和學術類同屬於比較正式的文體，因此兩類文體台華共通詞所佔的比例差不多（77%和 75.98%）。非學術類和小說、歌詩、諺語及俗語同屬於非正式的文體，理論上台語特別詞的使用比例應該也是差不多，但是 Sander 及 Hsieh 調查的三類文體皆比本研究少了約 10%。可能是 Sander 及 Hsieh 的調查樣本數較少，單一作者個人的用詞習慣所造成的差異。

本研究非學術台語特別使用比例的統計結果（42.99%），和鄭良偉預估的使用比例差不多：日常會話、詩歌、俗語有時佔文章的一半左右，散文大約佔百分之二十到四十之間（轉引自張學謙 1998）。

#### （四）小結

覆蓋率 100%和覆蓋率 80%的統計結果皆支持本研究的假設：台華共通詞的使用比例，學術類高於非學術類。低頻詞沒有影響統計結果支持本研究假設，在計算台華共通詞的比例時也沒有太大的影響。

### 三、詞型 VS 詞次比較分析

詞型是文本中詞彙的類型，詞次是文本中詞型的實際使用次數或頻率；本研究透過比較兩者的差異探討台華共通詞詞型與詞次實際使用的情形。

以下分成覆蓋率 100%、覆蓋率 80%兩部分討論。

#### （一）覆蓋率 100%

表 28 台華共通詞詞型 VS 詞次比較表（覆蓋率 100%）

項目	詞型	詞次
學術類	66.91%	74.58%
非學術類	55.09%	56.81%

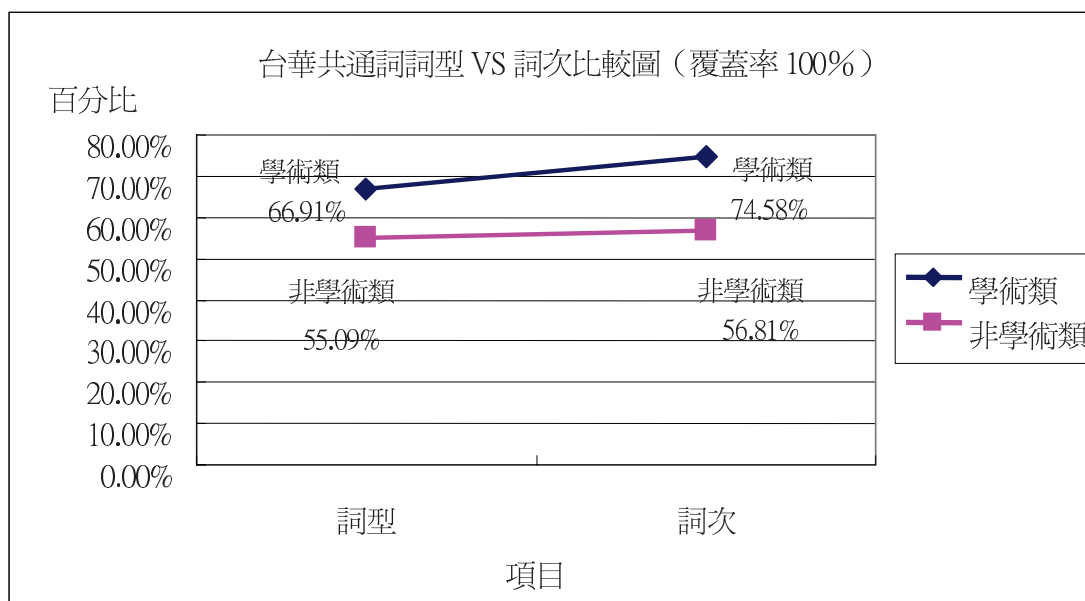


圖 10 台華共通詞詞型 VS 詞次比較圖 (覆蓋率 100%)

從上表及上圖可知：學術類詞次比詞型多 7.67%；非學術類詞次比詞型多 1.72%。

學術類與非學術類詞次多於詞型的原因可能是：

1. 低頻詞的影響：低頻詞包含許多電腦雜訊，統計時是以台語特別詞計算，因此會低估台華共通詞詞型與詞次的比例，但是對詞型的影響遠大於對詞次。
2. 台華共通詞在文本中實際的使用頻率比較高。

因為電腦雜訊是以台語特別詞計算，若以覆蓋率 80% (排除電腦雜訊後) 來看，台華共通詞的詞型應該會大幅提升，詞次提升的幅度應該會低於詞型。

## (二) 覆蓋率 80%

表 29 台華共通詞詞型 VS 詞次比較表 (覆蓋率 100%)

項目	詞型	詞次
學術類	82.40%	75.98%
非學術類	65.69%	57.01%

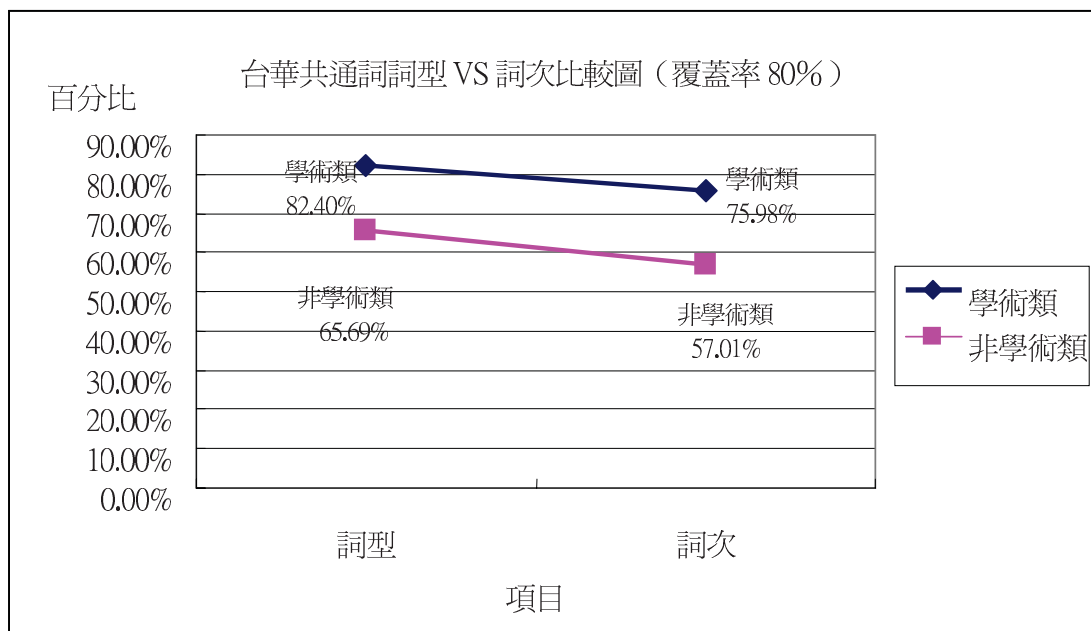


圖 11 台華共通詞詞型 VS 詞次比較圖 (覆蓋率 80%)

從上表及上圖可知：學術類詞型比詞次多 6.42%；非學術類詞型比詞次多 8.68%；亦即台華共通詞在實際文本中的使用比例低於詞型。覆蓋率 80% 已是排除電腦雜訊影響後的統計結果，因此，原因可能是一部分的台華共通詞是屬於使用頻率低的詞彙，非學術類又比學術類有更多的低頻詞。

#### 四、小結

本研究分別以詞型和詞次以及不同覆蓋率探討台華共通詞的使用比例，並與

鄭良偉(1987)、黃宣範(1998)以及 Sander 及 Hsieh (1981)的調查做比較，結果皆支持本研究的第一項假設：台華共通詞的使用比例，學術類多於非學術類。低頻詞並不會影響統計結果支持本研究假設，但是在計算台華共通詞詞型的比例時會有影響的程度比較大，排除低頻詞後，以覆蓋率 80%的計算結果比較符合實際的使用情形。台華共通詞詞型的比例高於詞次，也就是說台華共通詞有一部分可能是屬於低頻詞。

## 第二節 台語詞彙豐富度分析

本研究詞彙豐富度的計算方式為：詞型 ÷ 詞次，分別計算學術與非學術類的詞彙豐富度後比較其中的差異。低頻詞中包含有電腦雜訊，因此本研究計算五個不同覆蓋率的詞彙豐富度，降低電腦雜訊對研究結果的影響。

研究的結果五個覆蓋率的詞彙豐富度都是學術類多於非學術類，沒有支持本研究假設；因為差距不大，比例都低於 0.25%，因此研究者另外計算此差距在非學術類佔有多少比例（即差異比），以探討假設沒有得到支持的原因，計算方式如下：

$$\text{差異比} = \frac{\text{學術類 WR} - \text{非學術類 WR}}{\text{非學術類 WR}}$$

說明：WR 為 word richness 的縮寫，即詞彙豐富度。

詞彙豐富度與差異比的統計結果如下表與下圖：



表 30 詞彙豐富度比較表

項目	覆蓋率 100%	覆蓋率 95%	覆蓋率 90%	覆蓋率 85%	覆蓋率 80%
學術類	11.81%	7.17%	4.77%	3.44%	2.63%
非學術類	11.65%	7.00%	4.62%	3.26%	2.40%
差異	0.16%	0.17%	0.15%	0.18%	0.23%
差異比	1.37%	2.40%	3.24%	5.56%	9.58%

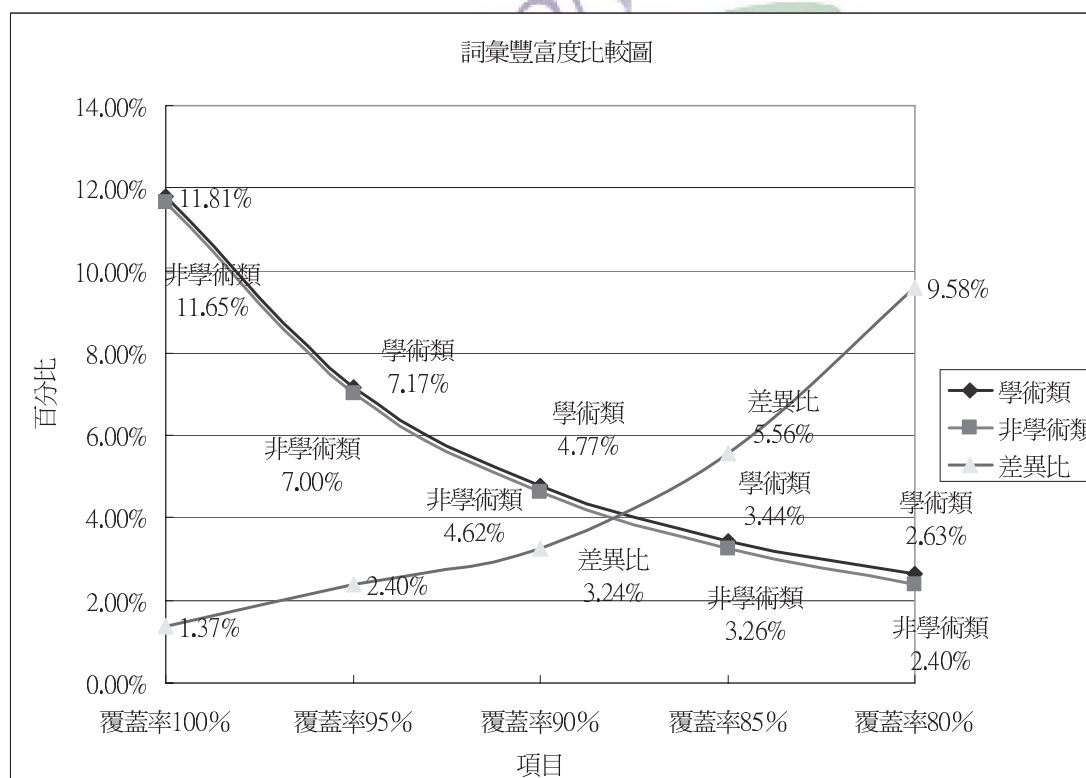


圖 12 詞彙豐富度比較圖

從上表及上圖可知：五個不同覆蓋率的詞彙豐富度都是學術類多於非學術類，相差的比例都低於 0.25%。統計結果並沒有支持本研究的第二項假設：台

語詞彙豐富度非學術類多於學術類。

本研究假設雖然並沒有得到支持，但是五個覆蓋率的差異都不大（差距都少於 0.25%，差異比介於 1.37%~9.58%），本研究試圖分析假設沒有得到支持的原因：

一、本研究已用不同覆蓋率計算的方式將電腦雜訊可能造成的影響降低，因此統計結果未能支持研究假設，初步可以排除電腦雜訊的因素。

二、樣本文體種類：本研究非學術類文體僅有小說、散文、劇本三類，未能包含詩、傳記、演講稿等其他文體，可能是研究假設沒有得到支持的因素之一。

三、樣本內部差異：《風俗通》(7.89%) 詞彙豐富度多於《孟子》(5.40%)<sup>36</sup>，文本雖然分屬不同時代、且《孟子》出於多人手，但是差異比達到 46%，初步可知小說和口語兩種文體存在一定程度的差異。劇本在內容上，口語佔有相當的比例；因此，小說、散文和劇本的內部差異，可能是研究結果未能支持假設的因素之一。

四、樣本語料量：本研究的語料量約 20 萬音節（學術、非學術約各 10 萬），可能是少數特殊的文本影響研究結果，以致研究假設未能得到支持。

以上是研究未能得到支持的可能因素，如能排除文本內部差異，廣泛蒐集各種文體，並且有更多的語料量再行研究，對於學術類與非學術類的詞彙豐富會有更全面性的了解。

### 第三節 台語羅馬字詞彙分析

---

<sup>36</sup> 見本研究第二章第三節台語詞彙豐富度部分。

本節分成詞型與詞次兩個部分探討台語羅馬字詞彙在學術類與非學術類的  
使用差異。

### 一、詞型

表 31 台語羅馬字詞彙使用差異統計表（詞型）

項目	學術	非學術
詞型個數	861	1,882
詞型總數	6,857	9,083
百分比	12.56%	20.72%

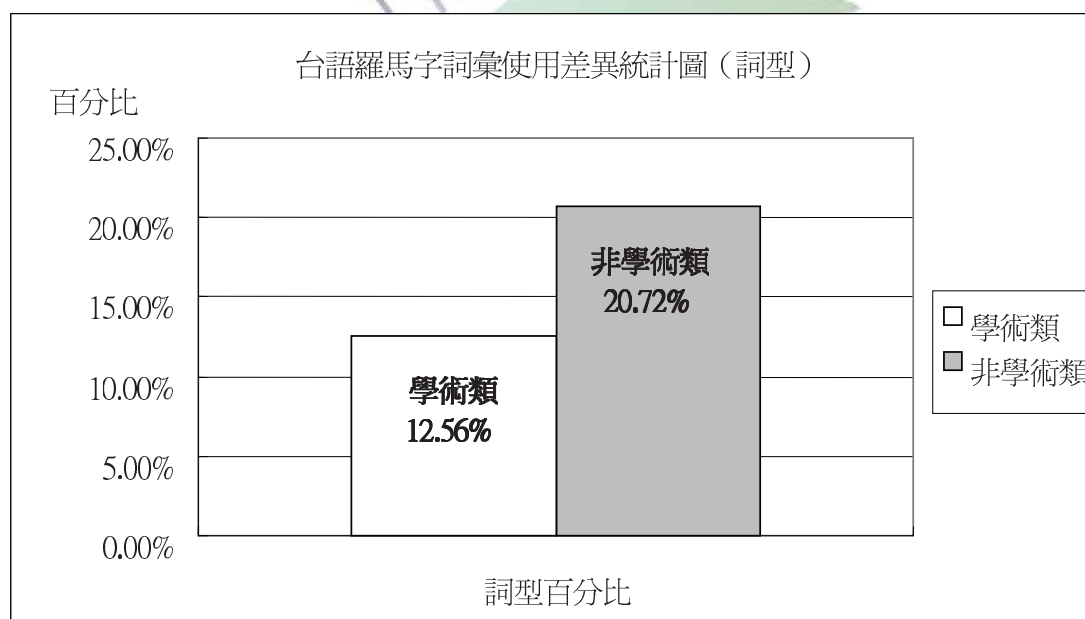


圖 13 台語羅馬字詞彙使用差異統計圖（詞型）

從上表及上圖中可知：學術類台語羅馬字詞彙的使用比例佔 12.56%，非學術類台語羅馬字詞彙的使用比例佔 20.72%，非學術類比學術類在台語羅馬字詞

彙的使用上多 8.16%。兩種文類使用台語羅馬字詞彙的比例皆超過 10%以上，非學術類達到 20%以上的水準。

上述的統計結果支持本研究的第四項假設：假設台語羅馬字詞彙使用比例非學術類高於學術類。

## 二、詞次

表 32 台語羅馬字詞彙使用差異統計表（詞次）

項目	學術	非學術
詞次個數	9,073	23,725
詞次總數	58,057	77,955
百分比	15.63%	30.43%

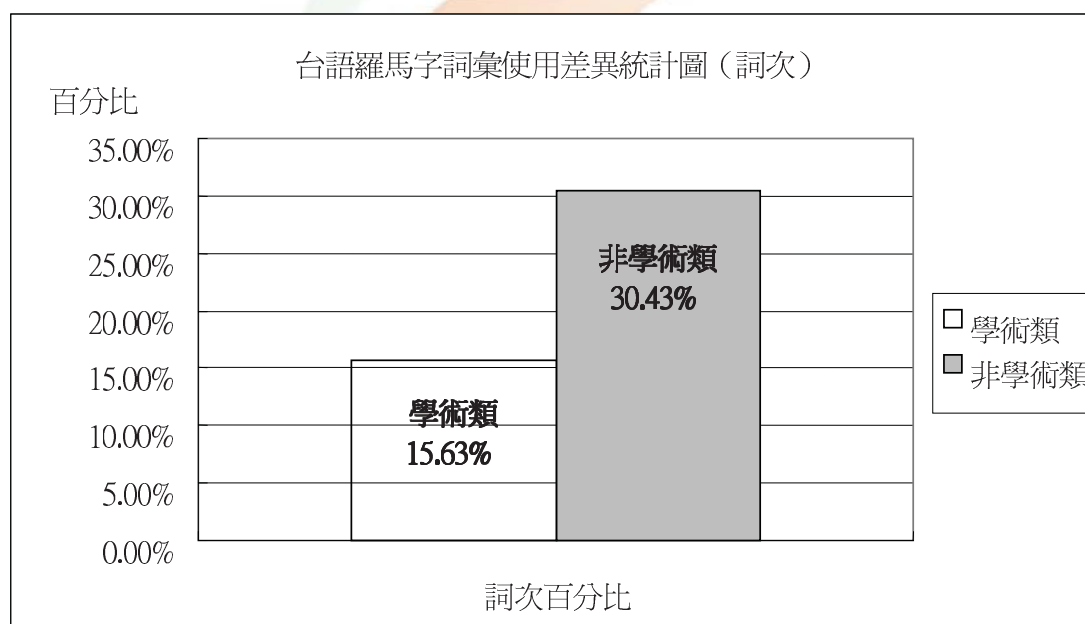


圖 14 台語羅馬字詞彙使用差異統計圖（詞次）

從上表及上圖中可知：學術類台語羅馬字詞彙的使用比例佔 15.63%，非學術類台語羅馬字詞彙的使用比例佔 30.43%，非學術類比學術類在台語羅馬字詞彙的使用上多出 14.80%。兩種文類使用台語羅馬字詞彙皆超過 15%以上，非學術類達到 30%以上的水準。

上述的統計結果支持本研究的第四項假設：台語羅馬字詞彙使用比例非學術類高於學術類。

### 三、詞型 VS 詞次比較分析

本研究透過比較詞型與詞次的差異探討台語羅馬字詞彙實際使用的情形。

表 33 台語羅馬字詞彙詞型 VS 詞次比較表

項目	詞型	詞次
學術類	12.56%	15.63%
非學術類	20.72%	30.43%

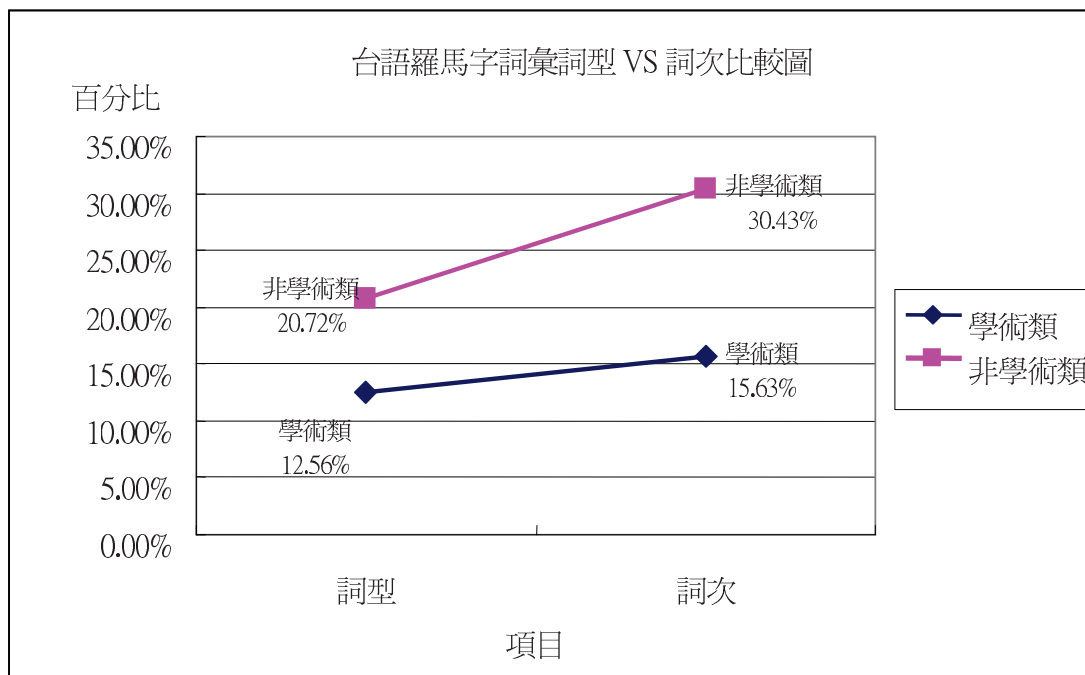


圖 15 台語羅馬字詞彙詞型 VS 詞次比較圖

從上表及上圖中可知：學術類詞次比詞型多 3.07%；非學術類詞次比詞型多 9.71%。可能的原因是：有一部分的台語羅馬字詞彙在文本中實際的使用頻率比較高；非學術類又比學術類有更多的高頻詞。

#### 第四節 台語平均詞長分析

本研究計算平均詞長的方式為：音節數 ÷ 詞次，分成四個部分比較學術類與非學術類的平均詞長：覆蓋率 100%、覆蓋率 80%、覆蓋率 100%VS 覆蓋率 80%、台語 VS 華語平均詞長比較分析。

## 一、覆蓋率 100%

表 34 台語平均詞長統計表（覆蓋率 100%）

項目	學術類	非學術類	台語詞彙
音節數	101,349	102,335	203,684
詞次	58,057	77,955	136,012
平均詞長	1.75	1.31	1.50

從上表可知：學術類平均詞長為 1.75，非學術類平均詞長為 1.31，學術類比非學術類多 0.44。統計結果支持本研究的第四項假設：台語平均詞長，學術類多於非學術類。

## 二、覆蓋率 80%

表 35 台語平均詞長統計表（覆蓋率 80%）

項目	學術類	非學術類	台語詞彙
音節數	66,617	77,863	144,480
詞次	46,452	62,368	108,820
平均詞長	1.43	1.25	1.33

從上表可知：學術類平均詞長為 1.43，非學術類平均詞長為 1.25，學術類比非學術類多 0.18。統計結果支持本研究的第四項假設：台語平均詞長，學術類多於非學術類。

排除低頻詞後的平均詞長比較符合實際情況。

### 三、覆蓋率 100%VS 覆蓋率 80%

表 36 台語平均詞長覆蓋率 100%VS 覆蓋率 80%比較表

項目	覆蓋率 100%	覆蓋率 80%
學術類	1.75	1.43
非學術類	1.31	1.25
台語詞彙	1.5	1.33



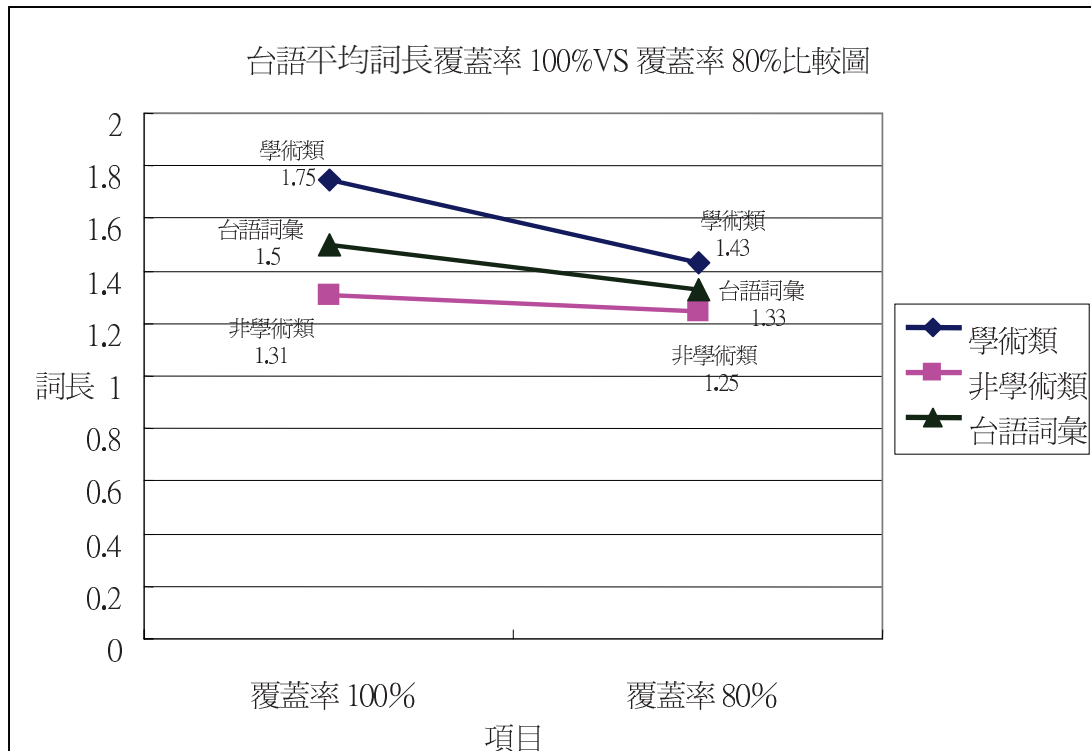


圖 16 台語平均詞長覆蓋率 100%VS 覆蓋率 80%比較圖

從上表及上圖可知：

- (一) 覆蓋率 100%與覆蓋率 80%的平均詞長都向下調整。
- (二) 平均詞長學術類多於台語詞彙多於非學術類。

#### 四、台語 VS 華語平均詞長比較分析

上表台語平均詞長為 1.50，此數據是以本研究的語料音節總數 ÷ 詞次總數計算得到的，雖然學術類與非學術類各文體並非以 1：1 的比例抽樣，亦未能涵蓋其他文體，但仍可做為初步了解台語平均詞長的參考。茲將本研究數據與漢語做比較，以了解其間的差異，結果如下表：

表 37 台語 VS 華語平均詞長比較表

項目	台語詞彙 覆蓋率 80%	現代漢語詞語 頻率辭典	兩岸三地漢語語料庫 (台灣、中國、香港)
平均詞長	1.33	2.0928	2.2706

從上表可知：台語平均詞長：1.33；《現代漢語詞語頻率辭典》：2.0928；  
《兩岸三地漢語語料庫》：2.2706。平均詞長《現代漢語詞語頻率辭典》多於《兩  
岸三地漢語語料庫》多於台語詞彙。以上結果支持本研究以音節數多寡推論比較  
平均詞長的論點。

