

從語域及借詞觀點探討台語文寫作風格

楊允言

花蓮 大漢技術學院資訊工程系

ungian@ms01.dahan.edu.tw

http://203.64.42.21

摘要

本文利用台語語詞中移借語層詞彙的使用情形來探討台語文的寫作風格，將移借語層的詞彙視為日語借詞、華語借詞和教會用語三個主要不同的語域(register)。把不同年代的文本拿來比較，可以看出其差異性。

本文的實驗，以李勤岸的研究為基礎，其語料為 1920 年代及 1990 年代的台語文小說作品，他將語料中出現的詞彙分出所屬的各語域。本文則利用各語域在台語詞彙，探討台語文的寫作風格。除上述語料，並加上卓緞女士白話字歌詩，以計算詞彙所屬語域的方式，來量化寫作風格，並說明台語文的特殊性，這特殊性基於政治環境及宗教因素等。此外，並針對實驗結果，台語文的語域、文類與寫作風格的關係，以及書面語和漢字制約問題對台語文的影響等，做進一步討論。

如果利用台語詞彙，從語域及借詞觀點來探討台語文寫作風格的方法可行，那麼本實驗應可做為進一步研究的基礎。

關鍵詞 台語文(Written Taiwanese)、語域(Register)、借詞(Loanword)、寫作風格(Writing Style)、文類(Genre)、台語羅馬字(Taiwanese Romanization)

1. 台語文的特色與特殊性

在台灣，台語是許多人的母語，然而大部分人使用台語僅止於「聽」與「說」，並不包括「讀」和「寫」，究其原因，與台語所身處的歷史背景或政治現實有很大關係。

若以漢字角度來談，台語一字多音現象比華語更加明顯。以台語線上字典的資料為例，此線上字典共有 22,080 個項目（entry，每個項目包括一個漢字及其讀音），11,635 個不同的漢字，平均每個漢字有將近兩個讀音（最高紀錄是 1 個漢字 12 個不同讀音）；如果只考慮 BIG-5 的常用字集，則比例更高（13,176 個項目，5,337 個不同漢字，平均每個漢字將近 2.5 個讀音）。¹其主要原因為，台語的位階較低，主體性較差，導致在不同時期需吸收不同的漢字讀音。

上述問題，若以詞的角度來看，兩音節或以上的詞彙，部分的讀音模糊性問題得以解決（但不是全部）。而台語的詞彙又是另外一個錯綜複雜的問題。在我們日常生活中所使用的台語，夾雜了不少華語、日語，²還包括原住民語，此外，更底層的百越語、中國歷代的官話系統也在其中。1935 年，郭一舟發表〈福佬話〉一文，形容台語的詞彙現象是「九重糶」，不管是從歷時或共時的角度看皆如此。

另外一個問題是詞彙變遷，因為政治環境變動大，以及台語位階低，其詞彙變遷速度遠比華語快。變遷包括同一個詞，現在和從前的用法不同，以及從前使用的詞彙現在被另一個新詞彙取代。

台語文自然語言處理，其中一個重要的基礎就是台語詞彙的整理，然而，要把「九重糶」中的每一層次逐層抽離出來，實非易事。

¹ 台語線上字典的資料以 1913 年出版，甘為霖的《廈門音新字典》為主要來源，這本字典應該是台灣最通行的台語字典，有其代表性。台語線上字典於 2003 年 1 月上線，可以用漢字查發音（羅馬拼音），也可以反向查詢，網址為 <http://203.64.42.21/TG/jitian>。

² 在台灣，語碼混合(code-mixing)的情形普遍，不過，所夾雜的日語或華語能不能算是台語，有一個簡單的檢驗規則：對於不諳日語或華語的台語使用者(native speaker)是否使用這個華語或日語詞，如果是，則

2. 台語詞彙變遷的研究與寫作風格

李勤岸(2000)曾對台語的詞彙變遷做較深入的研究。他認為某一語言和其它語言的接觸，一開始是語碼轉換(code-switching)或是語碼混合(code-mixing)的方式出現，之後成為借詞，再漸漸深入本土語層；以目前而言，借詞的主要來源為廈門話、日語和華語。他將台語的詞彙分為兩大語層：本土語層(local layer)及非本土語層(non-local layer)（又稱為移借語層，loanword layer），本土語層包括澳亞語層、南島語層、古漢語以及中古漢語，移借語層則主要包括教會用語(church register)、日語借詞和華語借詞。日語借詞和華語借詞很明顯是因為台灣的歷史、政治因素，而教會用語則較為特殊，屬於宗教因素，而教會用語又可分為廈門話和禮拜儀式用語(liturgical register)。

教會用語的特殊性在於，日語借詞和華語借詞是兩個語言接觸時，強勢語言（日語、華語）影響弱勢語言（台語）的口語及書面語，教會用語卻是同一類語言（如果把廈門話和台語視為方言的關係）以書面語來影響口語，而如果以台語和華語相比較，書面語影響口語的現象，在台語非常明顯。

他所用的語料包括 1916 年及 1935 年出版的台語聖經（用來分析教會用語中的廈門話），³1920 年代鄭溪泮及賴仁聲用台語羅馬字寫的台語小說（用來分析日語借詞），以及 1990 年代陳雷及陳明仁用漢羅合用方式⁴所寫的台語小說（用來分析華語借詞），將這些語料逐一打字建檔校對；利用斷詞系統將語料斷詞，⁵並經過人工校對。接著將其中非本土語層的詞彙挑出來，日語借詞和華語借詞較易分辨，而教會用語中的廈門話則

此詞彙可視為台語，如日語的“khi-mo”、“tha-ma-to”，或華語的「大哥大」、「小姐[che]」等。

³ 這是巴克禮(Thomas Barclay)翻譯的版本，1916 年出版新約，1935 年出版舊約，雖然是以廈門話寫成，部分詞彙和語法和台語不同，不過主要閱讀人口都是台灣人。真正「台語」版本的聖經，在 1972 年才由高積煥和陳邦鎮以漳州腔編寫出版，稱為紅皮聖經，卻被當時政府以影響國語推行的理由而沒收。

⁴ 漢羅合用，指的是台語文的一種書寫方式，目前可算是台語文書寫的主流，基本觀念是，若這個漢字一般人很容易知道台語要怎麼讀，就用漢字，否則就用台語羅馬字。

⁵ 如果是台語羅馬字的語料，詞彙之間以空白分隔，沒有斷詞的問題。

需細加斟酌，除了本身為台語使用者的語言學者的語感和直覺，還需藉助相關資料。⁶

李勤岸的語料量是：1920 年代有 112,964 個詞次 (word token)、12,941 個詞型 (word type)，1990 年代有 92,539 個詞次、12,969 個詞型。這些書面語中，移借語層詞彙的數量及變化請參看表一：

表一 兩個時代移借語層詞彙數量的變化

	詞型數	教會用語	日語借詞	華語借詞	總計
1920 年代文本	12,941	563	178	27	768
1990 年代文本	12,969	97	239	202	538

資料來源：李勤岸(2000) Table 4.4、Table 5.3、Table 5.16

說明：以詞型來計算，非詞次。

台語詞彙在這段時間的變化，教會用語大量減少，日語借詞增加小部分，華語借詞則大幅度增加，而整體來說，使用移借語層詞彙的數量減少。⁷

另外一個度量是詞彙豐富度 (lexical richness)，計算方式為：

$$\text{詞彙豐富度} = \frac{\text{詞型(word type)}}{\text{詞次(word token)}}$$

其計算出的結果請參看表二：

表二 兩個時代詞彙豐富度的比較

	詞型	詞次	詞彙豐富度
1920 年代	12,941	112,964	11.46%
1990 年代	12,969	92,539	14.01%
變化情形			+2.55%

資料來源：李勤岸(2000) Table 5.31

我們可以說，這些是以小說為語料所算出來的結果。

李勤岸的研究，主要是從移借語層詞彙的使用情形來探討台語語詞變遷。而本文則

⁶ 根據王毯仁的研究，閩南語 2,000 個常用詞中，有 200 多個台語不使用；而閩南語一般以廈門話為代表。1916 年及 1935 年所出版的台語聖經，其實是以廈門話為本，當中所用的詞彙，有些並未通行於台灣。不過，是因為此詞彙從來沒有在台灣被使用，還是以前曾經有人使用而目前已被別的詞彙取代，這又是另一值得深入探討的題目。王毯仁的研究引用自李勤岸(2000)。

⁷ 這裡要說明的是，兩個不同年代有時空背景的差異。1920 年代，台語羅馬字的閱讀人口主要為天主教及基督教徒，而 1990 年代台語羅馬字已經世俗化，非教徒的閱讀人口應該遠超過教徒，不過如前所述，這時候的文字形式為漢羅合用。

是利用移借語層詞彙的使用，來探討寫作風格。由於作者基本背景的差異，這個差異包括世代、宗教信仰等，導致其作品偏好使用的移借語層詞彙有所不同，從而產生不同的寫作風格。

3. 特殊的教會用語寫作風格

這是頗特殊的例子。2001年2月，報端刊載一則百歲人瑞卓緞女士過生日的消息。對地方而言，人瑞可傳為美談，不過，報導中還提到，卓緞女士沒有受過體制內教育，全然不識漢字，⁸卻在51歲無師自通地使用在教會所習得的台語羅馬字（又稱為教會羅馬字、白話字）創作白話字歌詩，並持續到她九十多歲時。其寫作時間為1950年代至1990年代。

因為地利之便，我與卓緞女士家屬取得聯繫，並著手進行白話字歌詩的整理。經過半年多，總共整理25首白話字歌詩及兩篇自傳性的文章。⁹

卓緞女士的白話字歌詩與一般台灣民間的七字仔類似，七字一句，有押韻，教會稱為七言文，早期的《台灣教會公報》裡有不少這種文類的文章，¹⁰我認為這是當時基督教為融入台灣社會所發展出的一種文學創作方式。以下舉一小段白話字歌詩為例，其中左邊是原文，右邊是以漢羅合用方式整理後的文稿。

Gō-ko sìn-gióng chin kian-tēng	五哥信仰真堅定
Kóng-ōe sì-chià ⁿ chòe kng-teng	講話四正做光燈
lā ũ un-jiū ê sèng-chêng	lā 有溫柔 ê 性情
Hòk-sāi Siōng-tè iā khiân-sêng	服侍上帝 iā 虔誠

⁸ 相關報導刊登在東部發行的更生日報，以及自由時報、中國時報、聯合報的花蓮地方版。其中，中國時報的標題，竟然出現「目不識丁」。台語羅馬字也是文字，不識漢字並不表示目不識丁。這點令我感受到漢字教育所造成的觀念偏差。

⁹ 因為卓緞女士不識漢字，也聽不懂華語及日語，家中只有幾本泛黃的台語聖經、聖詩，這其間，我常常走訪，並將手邊的台語羅馬字資料放大影印供她閱讀，而她非常高興。卓緞女士的作品，目前已完成台語羅馬字原文和台語漢羅合用的兩個版本，家屬希望還有華文、日文及英文三個版本的整理再行出版，不過我傾向加註華文註解，若翻譯成別種語文，可能風味盡失。目前正尋找出版社出版。

¹⁰ 《台灣教會公報》於1885年開始發行，原名為《台灣府城教會報》，是台灣最早發行的報紙。歷經幾次改名和整併，目前名稱為《台灣教會公報》。使用台語羅馬字書寫，在日本時代曾因太平洋戰爭而被迫停刊；戰後復刊，但後來政府推行國語政策，於1970年被迫全面改成華文。以上資料引用自張裕宏(2001)。

如果讀者並非教徒，會發現其中有些詞彙較少出現在其日常用語中，例如「信仰」、「光燈」、「服侍」、「上帝」、「虔誠」等。而短短 4 句、28 個音節，總共 17 個詞彙中，就出現了 5 個這樣的詞彙，比例之高，很容易令人察覺到其特殊的寫作風格，即大量運用教會用語的寫作風格。¹¹

本文利用李勤岸整理的移借語層詞彙，從這些詞彙使用所透露的訊息，來探討卓緞女士作品的寫作風格，並與李勤岸的語料所得結果做比較，探討相關問題。

4. 實驗步驟及結果

實驗的步驟如下：

- (1) 語料為我已經整理好的卓緞女士白話字歌詩 25 首，因為是台語羅馬字寫作，不用經過斷詞的處理；
- (2) 計算的結果，總共有 4,594 個音節，¹²2,878 個詞次，829 個詞型；
- (3) 將這 829 個詞，與李勤岸整理出來的移借語層的所有詞彙逐一比對，其中發現，除了大量的教會用語外，只有極少數的日語再借詞（如「環境」、「博士」等），¹³和華語借詞；
- (4) 比對結果，將這 829 個詞型標記類別，區分為教會用語（標記為「C」）和本土語層詞彙（標記做「L」）。

表三列出詞頻較高的前 30 個詞型（詞頻 \geq 14），表四則列出教會用語中詞頻較高的

¹¹另外，若以創作題材而言，除了和一般七字仔相同，都有勸世的題材外，卓緞女士白話字歌詩還有許多慶祝節慶（聖誕、生日、新年）的，七字仔則幾乎沒有這類題材。這應該是宗教氣氛環境的影響。

¹² 如果討論華文，一字就是一音節，不過台語文若用羅馬字書寫，一音節包括好幾個字元，用「字」來等同音節並不妥當。

¹³ 日語再借詞（deep-relexified Japanese loanwords）指的是由古漢文創造出來的日語新詞，華語和台語再從日語借用過來，也算是台語和華語的共通詞（但是台語不是從華語借用過來），屬於移借語層中的日語借詞。而卓緞女士的作品中，只有出現三個日語再借詞，沒有一般的日語借詞。而詞頻出現最高的「博士」（5 次），經由台語文語詞索引程式(<http://203.64.42.21/TG/concordance>)的查詢結果，「博士」一詞在台語聖經中出現 23 次。這說明，雖然「博士」一詞是日語再借詞，不過卓緞女士應該就是由教會的台語羅馬字資料學到這個詞彙。

前 28 個詞型（詞頻 ≥ 4 ）。

表三 詞頻較高的30個詞型

詞頻	詞型	漢羅	類別
107	Chú	主	C
107	Siōng-tè	上帝	C
77	ê	ê	L
68	chin	真	L
66	lâi	來	L
60	lán	咱	L
54	ū	有	L
47	hō	hō	L
44	góa	我	L
37	hoa ⁿ -hí	歡喜	L
37	i	伊	L
31	chin-chià ⁿ	真正	L
30	chiok-hok	祝福	L
26	sī	是	L
25	kám-siā	感謝	L
23	kiù-chú	救主	C
22	beh	beh	L
22	hok-khì	福氣	L
21	chòe	做	L
21	tāi-ke	大家	L
21	tiòh	著	L
20	chhōa	chhōa	L
20	Sè-kan	世間	L
19	lâng	人	L
18	bô	無	L
18	o-ló	呵佬	L
17	kiá ⁿ	行	L
16	hó	好	L
16	tī	tī	L
14	chiū	就	(C)

表四 教會用語詞頻較高的28個詞型

詞頻	詞型	漢羅	類別
107	Chú	主	C
107	Siōng-tè	上帝	C
23	kiù-chú	救主	C
14	chiū	就	(C)
11	Bók-su	牧師	C
11	ōe	會	(C)
10	êng-kng	榮光	C
8	lâ-sō	耶穌	C
8	sūn-hók	順服	C
7	hok-im	福音	C
7	sèng-kiá ⁿ	聖罔	C
7	thi ⁿ -pē	天父	C
6	Ha-le-lu-iah	(哈里路亞)	C
6	ín-chhōa	弓 chhōa	C
5	chóng-hōe	總會	(C)
5	siók-hōe	贖回	C
5	Thian-kok	天國	C
4	bók-chiá	牧者	C
4	gîm-si	吟詩	C
4	kî-tó	祈禱	C
4	kiù-un	救恩	(C)
4	kng-teng	光燈	C
4	lâi-kiám	利劍	C
4	mô-kúí	魔鬼	(C)
4	sip-jī-kè	十字架	C
4	siú ⁿ -sù	賞賜	(C)
4	tiú ⁿ -ló	長老	C
4	tōa-tek-sèng	大得勝	C

其中，類別標記為「(C)」的詞型，沒有出現在李勤岸整理的移借語層詞彙中，不過我認為是教會用語。¹⁴會有這樣的分歧，主要原因包括：(i) 在他整理的語料中沒有出現此詞彙，(ii) 「會」、「就」兩詞，如果用漢字書寫，我會將之視為本土語層詞彙，然而卓

¹⁴ 標記為「(C)」的詞彙，除了表四中所列出的，還有詞頻為 3 的「該哉[kai-chài]」、「交託[kau-thok]」、「禮物[lé-mih]」等 3 個，詞頻為 2 的「滅無[biat-bô]」、「至微細[chì-bî-sòe]」、「尊崇[chun-chóng]」、「近倚[kūn-óa]」、「跟 tè[kun-tè]」...等 12 個，詞頻為 1 的共有 29 個。

緞女士的作品以台語羅馬字書寫，一般台語使用者，這兩個詞分別唸成”ē”、”tòh”，而非”ōe”、¹⁵”chiū”，(iii) 分類很難避免模糊性的問題。

表五列出卓緞女士作品中各類詞彙的數量及所佔比例。

表五 卓緞女士作品中各類詞彙的數量及比例

	本土語層		移借語層			總計
			教會用語	日語借詞	華語借詞	
詞型	677	81.66%	144	3	5	829
詞次	2,291	79.60%	569	8	10	2,878

5. 相關問題探討

本文中採用的卓緞女士作品的文本，或許較特殊，與相同文類的其它作品相比是個極端（使用過多的教會用語）。不過，這樣的文本卻也提供了一個視野，讓我們感受到，在幾乎不受日語及華語的影響下，台語文可以表現出什麼樣的風格。以下打算針對前述的實驗結果，做相關問題的探討，包括實驗結果的進一步分析討論，語域、文類與寫作風格的關係，以及台語文受書面語及漢字制約的問題。

5.1 不同文本移借語層詞彙使用比較

表六列出三個不同文本移借語層詞彙的數量及比例。

表六 三個文本移借語層數量及比例

	詞型數	教會用語		日語借詞		華語借詞		總計	
1920 年代文本	12,941	563	4.35%	178	1.37%	27	0.21%	758	5.93%
卓緞作品	829	144	17.37%	3	0.36%	5	0.60%	152	18.35%
1990年代文本	12,969	97	0.75%	239	1.85%	202	1.56%	538	4.15%

部分資料來源：李勤岸(2000) Table 4.4、Table 5.3、Table 5.16

說明：以詞型來計算。

若以 1920 年代文本及 1990 年代文本做比較，我們發現：(i)華語借詞大量增加，這是政治因素；(ii)日語借詞不減反增，但若詳加檢視，兩個時代所用的日語借詞差異頗大，

¹⁵ 「會」讀做”ē”或”ōe”是方言差的問題，事實上就我所知，目前仍有少部分泉州腔讀做”ōe”，但是我很確定，許多教會人士是因為聖經等台語書面語的關係而讀成”ōe”，而他們的腔調是”ē”。

1920年代的日語借詞應以政治因素居多，而這些詞彙戰後可能被台語對等詞彙或華語借詞取代（例如「辯護士」被「律師」取代、「探偵」被「偵探」取代、「無料」被「免費」取代、「案內」被「招待」、「引 chhōa」取代...等），1990年代的日語借詞，應該是整個社會、經濟因素所造成；(iii)教會用語大量減少，這與台語羅馬字的社會脈動有關，1920年代文本的讀者主要為教徒，1990年代文本的讀者為一般社會大眾，顯示台語羅馬字在整個台灣的台語文運動中，有世俗化的趨勢，不再僅限於教會人士。

若考慮卓緞女士的作品，除了大量使用教會用語，令人感受到很不一樣的寫作風格外，極少量的日語、華語借詞，也多少反映出的她的生活背景。若對照她的自傳性文章，確實發現她的生命歷程很艱辛，憑藉著宗教信仰而得到心靈的寄託。¹⁶

一般人眼中，卓緞女士生活在社會階層最底層，他們沒有機會受教育，導致不識漢字，失去了書寫媒介，更不易被社會大眾接受或關注。卓緞女士卻因為有機會接觸基督教，得以學習台語羅馬字，並在30年後以此為媒介，將自己的際遇、心情呈現出來。她是一個特例，讓我們得以一探他們所處的生活環境，而且是由自己書寫而非透過他人轉述。她所提供的語料之所以特別，在於我們所建構出的書面語語料極少這樣的材料，但是處於這樣生活環境的人卻不算少數。

5.2 詞彙豐富度的探討

詞彙豐富度的比較請參考表七：

¹⁶在她出生後三天母親難產死亡，父親無力撫養，於是轉給貧苦人家做童養媳，二十歲左右在花蓮做牧師的五哥想起曾有這一位妹妹，找到她後，將撫養她的家庭一起從桃園帶來花蓮鳳林，五哥引導她接觸基督教，卻被其養父母禁止；後來生了一場大病，養父母只讓她接受「王祿仔仙」的治療、喝「符仔水」，拖了一年多未見起色，曾痛苦得想要自殺，後來幾位哥哥共同出錢籌措醫藥費，帶她去花蓮醫院醫治才得以痊癒，而後養父母才不再禁止她接觸基督教。婚姻不美滿，歷經許多波折，40多歲時帶著兩名年幼的女兒來花蓮，為了養家活口必須身兼數個工作，直到女兒長大有工作後才得以稍稍喘息。而面臨生活上的困頓，宗教信仰提供她心靈的唯一寄託。

表七 詞彙豐富度的比較

	詞型	詞次	詞彙豐富度
1920年代文本	12,941	112,764	11.48%
卓緞作品	829	2,878	28.80%
1990年代文本	12,969	92,539	14.01%

部分資料來源：李勤岸(2000) Table 5.31

卓緞女士沒有受過體制內教育，作品中極少日語和華語借詞，但是詞彙豐富度仍有1990年代文本的兩倍以上。這個結果與我們預期的不同，沒有受過體制內教育，應會導致所使用詞彙受限而降低詞彙豐富度，但是大量的教會用語卻彌補了這項缺憾，反而大大提高詞彙豐富度。從這裡可以看出台語書面語的強大影響力。

當然，這牽涉到不同文類(genre)的問題，卓緞女士的作品為歌詩，作品中所使用的虛詞(function word)應該比小說少得多，而詞頻較高的詞彙以虛詞較多，減少虛詞使用自然容易使詞彙豐富度增加。

關於這部分，應該要以相同文類不同作者的文本做進一步計算，才能釐清這個問題。

5.3 語域、文類與寫作風格的計算

語域、文類等名詞的定義，在社會語言學界並不一致。本文中，語域指的是同一語言在不同世代、不同領域、不同行業中的語言變體(language variation)，並把範圍縮小到書面語；而文類指的是書面語的各種體裁類型，例如小說、詩歌、學術論文等。

不同的語域或文類，可以產生不同的寫作風格(writing style)。風格是抽象的概念，我們將這些文本中的詞彙抽出，利用計量的方式，量化成文本的風格。而計量的方式，就是先將詞彙做分類，尤其著重移借語層的分類。

台語曾經歷兩次有系統、全面性的壓制，一次是日語，一次是華語，這是政治力量促成的變化，比自然演變的速度快很多，影響的層面除了詞彙，還包括句法，不過詞彙的影響最為顯著。

以各種不同類型的借詞來區分不同世代的語域，對台語是有用的。舉例來說，曾受日本統治的老一輩台灣人，其日常用語中所使用的日語詞彙會比較多，年輕一輩則較能運用華語詞彙。這在口語應可充分表現出來，而在書面語呈現時可能會被淡化，¹⁷不過仍可由上述的分析中看出端倪。從華語借詞的數量可以清楚看出 1920 年代和 1990 年代文本的差異。而卓緞女士作品所呈現出來的詞彙運用，顯然是沒有受到體制內教育影響的結果，對照不同世代的教育普及率，這是較早世代的寫作風格。

5.4 書面語及漢字問題

從卓緞女士的作品，我們清楚看到台語羅馬字書面語所產生的影響。這樣的例子雖然不多，但是並非唯一，只是沒有透過漢字或是華文的媒介，這些人、這些文本不易引人側目。

較特殊的是，台灣雖然已經遠離日本統治，但是 90 年代文本的日語借詞不減反增，我們假設台灣與日本的交流頻繁導致日語持續走入台語的日常用語中，可是對照同樣在台灣發展的華語，這個情形卻不明顯。這個現象應該是華語受漢字制約的原因，漢字的特性導致其不易借用外來語。台語文以漢羅合用方式，羅馬字的部分很適合用來處理借詞；此外，台語相較於華語，有較多音節，¹⁸這是有利於吸收外語借詞的因素。事實上，某些行業如建築業、製造業等日語的借詞特別多。若有詳盡的台語外來語辭典可供參考，這部分可做進一步的探討。

漢字制約另一個情形是，目前台語文作家，幾乎都受過完整且長時間的華文教育，華文的運用毫無問題，並在日常生活中大量接受華文訊息：報紙、書籍或是電視字幕等。慣用華語思考，極可能不自覺將許多華文詞彙直接帶入其台語文作品中，除非以純粹的

¹⁷ 因為書面語在書寫的過程中，作者可能會儘量過濾掉這些借詞。

¹⁸ 台語的音節數，我們以台語線上字典為例，共有 2,728 個音節；華語的部分，以詞庫小組技術報告 98-01 詞頻詞典的索引來計算，共有 1,081 個音節；台語約是華語的 2.5 倍。音節的組成單位是聲母、韻母及聲

台語羅馬字書寫。如果把台語分為口語和書面語兩部分，日語借詞的數量，應該是口語多於書面語，而華語借詞的數量則相反，應該是書面語多於口語。這部分若要進一步探討，需藉助台語口語語料庫的建立。

6. 我們還需要什麼？

本文只做了一個初步的實驗，如果將李勤岸整理的1920年代、1990年代台語文文本視為那個時代台語書面語一般的寫作風格，則其它文本，若有其特殊性，可以透過比較的方式，說明其特殊點。如果從語域、借詞來看台語文寫作風格這個觀點可行，我們需要更完整的資料來完成這件事。這些資料包括：

- ⇒ 完整的詞彙資料：除了基本的詞性等訊息之外，還要有語源的資料，最起碼能像李勤岸將本土語層和移借語層區分出來，並將移借語層再做區分，除了上述的日語借詞、華語借詞和教會用語外，還會有英語借詞、南島語借詞、客語借詞...等。¹⁹最理想的情形是把「九重糕」的每一層都剝離出來。這是大工程，期待關心台語的語言學者有此雄心壯志。若從另一個角度考慮，如果某些詞彙實在很難斬釘截鐵地確認其為某一語域專用的詞彙，或許可以用統計語料的方式，來計算此詞彙在每一語域出現的比例，這是計算語言學者可以發揮的地方。
- ⇒ 完整的台語書面語語料庫：目前正在建立中，不過離「完整」的目標還有一段遙遠的距離。²⁰目前可找到的最早的台語文資料是1885年所書寫的，²¹如果資

調。有較多的音節，意謂著對於借詞，更容易找到較相近的音節來處理。

¹⁹ 英語借詞不多說；南島語借詞，例如我在花蓮，麵包果我們稱為 *pa-chi-lu*，這是阿美語；客語借詞，例如 *se-moe*（小姐）、*an-chu-se*（非常謝謝）等。

²⁰ 台語文語料庫，由台灣羅馬字協會發起，從2003年1月開始著手建立台語文語料庫，分台語羅馬字文本及漢羅合用文本兩大部分，截至2003年8月底止，分別蒐集到了92萬音節及164萬音節的語料。目前提供一個簡單的語詞檢索程式(Concordancer)供使用者查詢，有8千6百多次的查詢人次記錄。期待未來以申請計畫的方式，加速此台語文語料庫的建置速度。請參考 <http://203.64.42.21/TG/guliaukhou>

²¹ 中研院的語言典藏計畫中，台語文獻追溯到16世紀的歌仔冊。歌仔冊對於台語的保存，確實有其不可磨滅的貢獻，不過那個時候的書面語和口語有相當大的差距。這裡所說的1885年，是指用台灣的台語（而非廈門話等地的閩南語）所出版的《台灣府城教會報》，這份書面語書寫台語的口語。

料完全建檔，我們可以更詳細地討論同文類歷時(diachronic)的差異，或是不同文類共時(synchronic)的差異，或以性別、時間等因素做為語域的區分等。目前世界上較有名的語料庫，其文類的分類標準並不一致，我們還可以計算同一文類中文本的差異和不同文類文本的差異，來檢視這個分類是否恰當。

- ⇒ 相關的工具：例如斷詞工具、詞性標示工具、語詞索引工具...等。
- ⇒ 台語口語語料庫：因為至今台語文還沒有一套官方承認的正式書寫系統，這導致書面語的寫法不一致，也使口語少了文字的制約。所以，相較於華語，台語的書面語和口語之間的差異可能要大些。如果同時有書面語語料庫及口語語料庫可供比對，也許我們能更清楚台語的語言現象。

誌謝

本文寫作期間，感謝哈佛東亞語言文明系李勤岸助理教授、台東師院語教系張學謙助理教授、台語文作家吳景裕（陳雷）醫師對本文部分論點提出討論及建議。

參考資料

- Douglas Biber, 1995, 《Dimensions of register variation》, Cambridge Univ. Press.
- 楊允言, 2001, 〈白話字傳奇〉, 《台文通訊》86 期頁 1-3, 2001/4。(<http://203.64.42.21/iug/Ungian/Chokphin/sanbun/POJthoanki/poj-ama.htm>)
- 楊允言, 2002, 〈七字仔 tú 著上帝—論卓緞 ê 白話字歌詩〉, 《第四屆台灣語言及其教學國際學術研討會》頁 489-497, 高雄：中山大學中文系, 2002/4/27-28。(<http://203.64.42.21/iug/Ungian/Chokphin/Lunbun/7jia/7ji-a-Tionsan.htm>)
- 楊允言, 2002, 〈論卓緞白話字歌詩 ê 寫作風格〉, 《第八屆北美洲台灣研究學會年會》, 美國：芝加哥大學, 2002/6/27-30。(<http://203.64.42.21/iug/Ungian/Chokphin/Lunbun/7jia/7ji-a-NATSC.htm>)
- 李勤岸, 2000, 《Lexical Change and Variation in Taiwanese Literary Texts, 1916--1998 —A Computer-Assisted Corpus Analysis》, 美國夏威夷大學語言學研究所博士論文。

(2003 年 9 月由金安出版社出版)

郭一舟, 1935, 〈福佬話〉, 《台灣文藝》二卷 6 號、10 號, 1935/6 & 10.

黃昌寧、李涓子, 2002, 《語料庫語言學》, 北京：商務印書館。

黃佳惠, 2000, 《白話字資料中的台語文學研究》, 台南師範學院鄉土文化研究所碩士論文。(論文全文可至全國博碩士論文網下載 <http://datas.ncl.edu.tw/cgi-bin/theabs/flywebi.cgi?p=3535&i=5229789&t=603&o=v5>)

張學謙, 1998, 〈Ho-lo 台語的語層及語用〉, 《第二屆台灣語言暨語言學國際研討會論文選集》頁 451-463, 台北：文鶴。

張裕宏, 2001, 《白話字基本論：台語文對應&相關的議題淺說》, 台北：文鶴。(第一章導論有上網：<http://203.64.42.21/iug/Ungian/patlang/POJkpl/POJkpl01.htm>)

張德明, 1995, 《語言風格學》, 高雄：麗文文化。

卓, 1972, 《卓公宗慶子孫家譜—北部設教百週年暨卓家信主八十週年紀念》。

卓緞手稿，未出版。