

台文華文線上辭典建置技術及使用情形探討

楊允言

大漢技術學院資訊工程系

ungian@ms01.dahan.edu.tw

摘要

一個免費使用的台文華文線上辭典，至今已經上線超過兩年，預估有十萬人次以上的查詢。使用者可以用華文或台文來查詢；用台文查詢，除了用漢字，也可使用台語羅馬字（白話字）。本文中，將介紹台文華文線上辭典的發展過程及功能，並討論技術層面相關問題，包括開發此線上辭典所使用的平台、前端介面及後端的程式和資料庫，以及華文查台文、台文查華文的方式，在資料庫中如何做比對；並提出善用使用者資源的觀念，系統紀錄下相關的查詢資訊，如使用者查詢方式、查詢人次、查詢點等的統計，並將對查詢失敗的原因做探討；從這裡試圖來了解使用者的需求，做為系統改進的依據。最後，提出未來改進的方向，包括使用者修改、增加詞條，加入語音合成介面，白話字的模糊查詢，及更多連結等，希望此系統功能更強、更好用，讓更多的使用者受惠。

關鍵詞 線上辭典、檢索、台語文、白話字、台語羅馬字、華文

1 台文華文線上辭典緣起

從前學習英文，必備的是英漢辭典，後來電子辭典盛行，在許多場合取代了攜帶較不便、查閱較費時的紙本辭典。網路盛行後，線上提供英漢翻譯的服務，讓使用者更加方便。

在台灣，台語日漸受到社會的重視，雖然現有許多紙本台語辭典，但若能在網路上提供免費、查詢快速的線上辭典，對於台語的發展，絕對有正面的意義。

基於這樣的想法，2000年7月開始，由夏威夷大學東亞語文所鄭良偉教授提供資料，開始進行台文華文線上辭典的開發工作。

原先的電子檔，包括台文羅馬拼音（又稱台語羅馬字或白話字，以下稱白話字）、¹台文漢羅合用、²華文對譯、詞性、台文詞頻、華文詞頻、...等欄位，筆者

¹ 「白話字」這個詞，是爲了要區別三種當時存在於台灣的不同書寫法，用來書寫文言的漢詩、漢文的，叫做「孔子字」，用來書寫當時清國白話文的，叫做「唐人字」，而用來書寫台灣人日常

取前三項來使用。經過一段時間的資料整理、系統分析及程式撰寫、測試後，台文華文線上辭典於 2000 年 11 月上線 (<http://203.64.42.21/TG/sutian>)，之後並逐步加以修改、加強功能。

2 台華線上辭典功能說明

以下針對台文華文線上辭典的功能做一說明。

2.1 查詢方式

每一詞條有台文的白話字、漢羅合用，以及華文三個欄位，三欄位都可提供查詢。使用者可用華文的「蜥蜴」查出台文的「四腳蛇(sì-kha-chôa)」、「四腳杜定(sì-kha-tō-tēng)」、...等；也可由台文的「稀微」查出華文的「淒涼」、「落魄」；如果熟悉白話字，對於不確定漢字如何寫的台語詞，例如輸入「leng2-ki-ki」可查到「冷-ki-ki / 冷支支(冷颼颼/涼颼颼)」。

台語的拼音系統有很多套，³使用白話字，是因為白話字有一百多年的歷史，相關的資源豐富。⁴不過為方便使用者輸入，採用數字聲調、「ⁿ」用「N」、「^o」用「ou」的簡便方式，這樣的方式，事實上在台語文書寫網路社群中廣泛被使用，⁵而顯示時仍然以正式的白話字呈現。⁶此外，TLPA與白話字差別不大，除了聲母的「c/ch」與白話字的「ch/chh」相衝突無法處理，以及需要加上連字符(“-“)外，其它部分如「nn(N)」、「ua/ue(oa/oe)」、「ing(eng)」，本系統也可接受，兼顧TLPA的使用者。⁷以上例而言，輸入「ling2-kil-kil」和「leng2-ki-ki」會得到相同查詢結果。

2.2 比對方式

用語，既不是高深的漢文，也不是清國的官話，所以叫做「白話字」。資料引用自 1885 年的《台灣府城教會報》，另可參考 <http://203.64.42.21/iug/Ungian/POJ/SimmihPOJ/SimmihPOJ.htm>

²漢羅合用，指的是以漢字為主，遇到有爭議的漢字用字時，改用白話字。

³ 本文不擬對各套系統多做說明，此問題可參考楊允言、張學謙〈台灣福佬話非漢字拼音符號的回顧與分析〉，發表在《第一屆台灣母語文化重生與再建學術研討會》，1999.6.19，<http://203.64.42.21/iug/Ungian/Chokphin/Lunbun/Huho/huho-0.htm>

⁴ 白話字的問題，請參考張裕宏《白話字基本論：台語文對應&相關的議題淺說》，台北：文鶴，2001 年 3 月。

⁵ 台語文網路書寫社群於 1991 年成立，基本上是透過架設mailserver來建立此網路社群，一開始，伺服器設在交大電子物理系，後來在交大資訊工程系，之後移到美國(formosa.org)，目前設在Yahoo，網址是<http://tw.groups.yahoo.com/group/taigubang/>，管理者為陳鄭弘堯和賴柏年。

⁶ 這樣的好處是資料交換的方便，此外，詞條排序問題也可一併解決。

⁷ 根據筆者於 2002 年擔任國民中小學鄉土語言（閩南語）教學支援人員檢核筆試閱卷委員的經驗，有不少應考者並不清楚白話字和TLPA的差別，例如他聲明使用TLPA，但是卻寫出白話字。白話字與TLPA的問題不在本文的討論範圍，可參考楊允言〈台語符號競爭——以TLPA kah白話字做例〉，《2002 台灣羅馬字教學kap研究國際學術研討會》，2002.7.14，<http://203.64.42.21/iug/Ungian/Chokphin/Lunbun/huhokengcheng/huhokengcheng.htm>。

以紙本的英漢辭典為例，查詢是單向而非雙向（英文查華文，無法華文查英文），且多半以詞首排序，極少部分辭典以詞尾排序，兩種排列方式只能擇其一。

利用電腦來處理這些詞條，呈現的方式就顯得多樣，我們可以雙向查詢（台文華文互查），也可以從詞頭找、詞尾找，甚至從詞中間找。

本系統提供完全比對、部分比對、詞頭、詞尾四種方式來搜尋詞條。如果使用者很清楚所要查詢的詞彙，可以用完全比對；如果所找的詞彙，其中某些字不確定如何寫，或不知道完整的詞彙，或希望順便找出相關詞彙，可考慮用部分比對；此外，我們可以用詞頭或詞尾來找尋疊詞或是詞綴，例如我們可以用台文的「烏」、詞頭比對，找到台文的「烏暗暗」、「烏墨墨(bák)」、「烏臭臭」、「烏汁汁(chiap)」、「烏-chut-chut」、「烏蔭蔭(im)」、「烏-khâm-khâm」、「烏金金」、「烏鬼鬼」、「烏茹茹(lu)」、「烏-mà-mà」、...等，也可以用台文拼音的「hau2」、詞尾比對，找到台文的「哀哀吼」、「嚼(chiàuh)嚼吼」、「嘩(hōa)嘩吼」、「𦵏(hôa)𦵏吼」、「轟轟吼」、「好(hó")好吼」、「hu-hu-吼」、...等。雖然我們無法避免會找出一些不相干的詞彙，詞頭、詞尾的搜尋方式已經幫我們做了一些過濾的工作。

2.3 連結到台文/華文線上 Concordance 程式

因為本系統的詞條沒有例句，於是想到可以利用線上 Concordance 程式來彌補這一缺憾。

華文Concordance程式由中研院資訊所詞庫小組開發，1996 年提供Web介面供一般使用者查詢；⁸台語文語料庫於 2003 年由台灣羅馬字協會負責建置，分漢羅合用文本及白話字文本兩部分，白話字文本的語料量尚不足，截至 2003 年 3 月底，漢羅合用文本約有 1,465,500 音節，白話字文本約有 329,600 音節。⁹

是否連結到線上 Concordance 程式可由使用者選擇，若選擇要，則在查詢結果的詞彙上點選，就會根據使用者所點選的欄位，分別連結到台文的白話字、漢羅合用及華文的線上 Concordance 程式，查看語料庫中此詞彙的前後文。

2.4 其它功能

這套線上辭典還有促成台語文標準化的功能。詞條等於提出台文用字建議。

3 台華線上辭典建置技術

⁸ 請參考中央研究院平衡語料庫的說明，<http://www.sinica.edu.tw/ftms-bin/kiwi.sh?>。

⁹ 以音節為單位而不以字做單位，是因為雖然漢字一字元是一音節，但白話字一音節有數個字元。另外，音節數是估計數，請看台語文Concordance程式的說明，<http://203.64.42.21/TG/concordance/soatbeng.asp>。

以下針對線上辭典的建置及程式技術方面做說明。

3.1 平台、前端及後端

本系統為 Web 介面的資料庫查詢系統，在 Windows 2000 Server 作業系統加上 ASP 程式，後端使用 Access 資料庫。雖然辭典資料上網，但利用伺服器端的安全設定，使用者無法直接下載，確保資料安全性。

3.2 比對查詢方式

利用 SQL 語法來擷取使用者要查的詞條。有三個欄位，四種比對方式，應該要下十二個查詢，不過，三個欄位可用變數來處理，這樣只需下四個 SQL 查詢。表 1 列出這四個查詢的 SQL 語法：

表 1 比對的 SQL 語法

完全比對	SELECT * FROM <TABLE> WHERE <FIELD>=<INPUT>
部分比對	SELECT * FROM <TABLE> WHERE <FIELD> LIKE '%<INPUT>%'
詞頭	SELECT * FROM <TABLE> WHERE <FIELD> LIKE '<INPUT>%'
詞尾	SELECT * FROM <TABLE> WHERE <FIELD> LIKE '%<INPUT>'

其中，<TABLE>資料表名稱是固定的，<FIELD>欄位名稱用變數處理，依使用者選擇，可為白話字、漢羅合用或華文，<INPUT>則為使用者輸入的詞彙。

將 TLPA 拼音方式轉成白話字，用字串取代函數即可處理。

如果使用者輸入漢字，問題較小，¹⁰白話字部分則需進一步處理。SQL查詢時英文字母是不分大小寫，所以查「in」時，「iⁿ (iN)」也會被查出，另外查「ia」時，帶聲調的（如「iá (ia2)、ià (ia3)、...」等）、韻母含「ia」的（如「ian、iang、iaⁿ (iaN)、...」等）和帶聲母的（如「pia、hia、chia、...」等）也會被找出，而這些都是多餘的訊息。

解決方法是利用字串處理的技巧將 SQL 查詢結果再做一次過濾，首先找出輸入字串中的音節界線（連字符號“-”及字串的頭尾），然後以音節為單位，將 SQL 查詢結果的資料再做一次比對。

3.3 顯示方式

顯示包括白話字及漢字，一般電腦漢字顯示皆無問題；白話字部分，有少數符號目前尚未被Unicode納入，¹¹且Windows作業系統並未完全支援Unicode所有

¹⁰ 漢字的情形，若考慮異體字（「台/臺」、「體/体」，若考慮繁簡體字更複雜）或台文漢字用字未標準化的因素，其實也有問題，不過以實際狀況來看，這問題並未浮現。

¹¹ 相關問題，請參考 <http://203.64.42.21/Unicode>。

功能，所以暫以安裝字型來解決顯示問題。

2001 年劉杰岳先生開發Taiwanese Package（簡稱TP），可在瀏覽器中正常顯示白話字；2002 年，陳鄭弘堯先生用ASP寫一支Asc2TP數字轉調符的公用程式，上述軟體皆可免費取得。¹²本系統的資料，白話字部分以數字標調方式存放，顯示時，先執行Asc2TP做轉碼，再以TP的字型顯示，使用者只需安裝TP，即可順利瀏覽白話字。

3.4 其它

此系統並將使用者的查詢情形記錄下來，包括使用者用哪個欄位查詢、查詢的詞彙、使用者端的 IP 位址、...等，這些資料，多少可了解使用者的使用情形，並可進一步做為系統改進的重要參考。

4 使用情形及錯誤分析

統計功能在 2002 年 12 月加上來，主要考量是，我們應善用使用者提供的訊息，來加強系統功能。使用者的意見反應可以透過他們直接與系統管理者溝通，但是這需要使用者主動提供；另一可行的方式是，記錄使用者查詢過的訊息，從中發現使用者的需求。

以下把使用者查詢的相關訊息做初步分析，以便能改進系統。這些統計的資料，都計算到 2003 年 3 月底。¹³

4.1 查詢方式

由表 2 可看出，使用華文查詢（55.24%）多於使用台文查詢（44.76%），台文的部分，漢羅合用方式查詢（77.73%）則遠多於白話字查詢（22.27%）。

表 2 查詢方式統計

查詢方式	次數	百分比(%)	台文查詢方式	次數	百分比(%)
華文	15,711	55.24	漢羅合用	9,895	77.73
台文	12,730	44.76	白話字	2,835	22.27

4.2 查詢人次、查詢點統計

從 2002 年 12 月到 2003 年 3 月底，共有兩萬八千多人次的查詢紀錄，來自一千五百多個不同的 IP 位址。單日最高紀錄為 1,446 人次（2003 年 1 月 1 日）。

¹² TP請至 <http://www.phahng.idv.tw> 下載，Asc2TP請至 <http://lomaji.com/INFO/BOOKS/NNG-THE/asc2tp/>。

¹³ 上網觀看可得到截至目前為止最新的統計結果。
<http://203.64.42.21/iug/Ungian/Soannteng/chil/thongke.asp>。

統計結果超過估計。¹⁴由此推估，此系統上線至今，至少有 10 萬人次以上的查詢。這個情形說明社會上確實有此需求，且這個需求可能因為九年一貫鄉土語言教育的實施，以及各大專校院陸續成立台語文相關系所而更加迫切。

4.3 詞條狀況及被檢索情形

目前本系統有 53,071 個詞條，過去四個月的查詢，只有 1,909 個詞條從未被使用者查閱過。計畫未來要透過使用者的協助，修正詞條錯誤並增加詞條數。

4.4 查詢失敗原因分析

系統將使用者的查詢記錄下來，包括查詢失敗的訊息，以便找出原因，並做改進。查詢失敗的比例，請參考表 3。

表 3 查詢失敗統計

查詢方式	查詢次數	查詢失敗次數	查詢失敗比例
台文漢羅	9,895	2,246	22.70%
台文白話字	2,385	802	33.63%
華文	15,711	2,364	15.05%

查詢失敗的原因，主要包括下列幾種情形：

- (1) 輸入的不是詞：有時使用者輸入整句話或一句話的片段，或輸入的字串不是一個詞，此時查詢失敗的機率很高。慣用漢字書寫者，對於詞的概念也許較為淡薄，這必須透過教育來改變。
- (2) 詞條數不足：若辭典中沒有某個詞，則可能導致查詢失敗。例如辭典中，華文欄位沒有「小孩」，但是有「小孩子」這個詞，若用完全比對的方式查「小孩」會失敗，但如果用部分比對則找得到。
此問題的解決方法，就是增加詞條。
- (3) 查詢方式錯誤：尤其以台文白話字的部分佔多數（超過 80%），使用者選擇用白話字查，但卻輸入漢字；¹⁵也有用台文漢羅查詢卻輸入華文，或用華文查卻輸入台文漢羅。對於選白話字卻輸入漢字，可考慮利用程式來檢查，以減低查詢失敗的次數；或者當查詢失敗時，就用其它兩個欄位查查看。
- (4) 拼寫錯誤：大部分發生在白話字查詢，主要原因是省略了連字符號(“-”)。

¹⁴因為這期間，經過農曆新年（1 月 25 日到 2 月 9 日這 16 天共只有 1,166 人次查詢），以及電腦病毒肆虐（2 月及 3 月約有 20 天系統處於斷線或無法正常運作的狀態）。

¹⁵為了減少此錯誤，我們在查詢畫面中，白話字後註明「台語羅馬字」，不過此情形依然存在。

或聲調（如輸入”gin-a”或”gina”，正確查詢方式應為”gin2-a2”[囡仔/孩子]）；少部分發生在使用漢字的查詢，如數字的「一」打成注音符號的「一」等。白話字拼寫錯誤源於使用者習慣，下文將做較詳細討論。

若要減少查詢失敗，上述的(2)、(3)、(4)點都提示了系統改進的方向。

4.5 從使用者查詢情形看需求

由查詢人次來看，查詢台文詞彙這個需求是存在的，相關的政府單位或民間機構其實有必要投入這類系統的開發。行政院客家委員會委託民間廠商開發了一套客語線上辭典，¹⁶相對於人口數佔優勢的福佬人的需求，沒有理由不受到相關單位的重視。

另外，從查詢方式來看，以華文查詢的次數較多，可是坊間紙本的台語字辭典，絕大多數是台文詞條、華文解釋，這與使用者的需求或許沒有完全相符。即使對一個熟練台語的人而言，因為整個大環境的關係，他不見得對台文用字掌握得很好。在這種情形下，出版社若要出版紙本的台語辭典，或可考慮以華台辭典為優先考量。

5 未來展望

我們希望系統的功能更好，更能滿足使用者的需求，除了從使用者查詢中得到啟發之外，也希望透過其它的技術或觀念來改進。以下提出一些想法，這些也是本系統未來的改進方向。

5.1 使用者參與修改詞條及加詞

線上辭典除了技術，資料的豐富及正確更重要，透過使用者參與，可以改進辭典內容。

本系統打算提供一項功能，當使用者查詢後，發現某個詞條的資料有誤，或是使用者發現線上辭典未收錄的詞條，可以線上填寫表單，將此訊息放入資料庫，並同時以 Email 通知管理者，經過確認後，將詞條修改或新加詞條。

5.2 加入語音合成介面

主要針對台語部分，台文的寫法尚未標準化，有些使用者對白話字或台語漢字不熟悉，看到詞條未必讀得出來。如果此系統能夠結合語音合成研究的成果，利用詞條中的白話字，輸出聲音檔，對使用者幫助很大。另外的方法是錄音，聲

¹⁶行政院客家委員會客語語音辭典，http://dictionary.ihakka.net/kaga_flash/，聲音部分是用錄製的，不過是由華文查客語發音。

音比較好聽，不過後來新增的詞條需另外錄製，比較麻煩。

5.3 白話字的模糊查詢

在 1990 年代初期，漢字在網路上還不方便使用時，就已經有使用白話字做為溝通管道的網路社群，其所使用的白話字，姑且稱為簡便版的白話字，省略了聲調，有時甚至省略連字符(“-“)”，如「Tâi-ôan」(台灣)寫成「Taioan」；此外，有人雖然會拼寫白話字，可是對聲調掌握得不是很好。針對這些使用者，系統希望能提供用「Taioan」查到「Tâi-ôan」這個詞，這就是本文所指稱的模糊查詢。

技術上，有兩種方法可做到模糊查詢，一是將使用者輸入的字串拆成一個一個字母來查詢，以「Taioan」為例，查詢的 SQL 指令片段是「... WHERE <FIELD> LIKE 'T%a%i%o%a%n%'」；我們也可加入規則來加快查詢，例如聲母不能單獨成立一個音節（如上例的「T」和「n」），所以查詢可改成「... WHERE <FIELD> LIKE 'Ta%i%o%an%'」。另外一種方法是在資料表中增加「不帶聲調沒有連字符的白話字」這個欄位，直接比對速度較快，資料檔卻變大。若以時間為考量，第二個方法應該較為可行。

5.4 更多的連結

利用超連結的特性，台華辭典可做更多的延伸。目前，陳鄭弘堯正在開發「暗光鳥 ê 日台辭典」，如果做好連接的介面，使用者可以查到相對應的日文詞彙。另外，華文的資源遠比台文豐富，我們一樣可以由華文的欄位做橋樑，連結到線上的華文/英文辭典。

這樣的連結並非沒有任何代價，更多的連結意味著更長的等待時間；此外，多一個中介的語文，就增加了找出不必要訊息的機會。不過在資源有限的情形下，不失為一個簡便的方式。

另外，未來若有某些辭典的例句上線，我們也可以來做連結。

參考資料

- [Lâu 2002] 劉杰岳：〈Sú-iōng Unicode lâi chhú-lí POJ〉，《2002 台灣羅馬字教學 kap 研究國際學術研討會論文集》，台灣羅馬字協會，2002 年 7 月。
<http://203.64.42.21/iug/ungian/POJ/siausit/2002/2002POJGTH/lunbun/E4-Kiatgak.htm>
- [Lâu et 2002] 劉杰岳、楊允言：〈白話字電腦文書處理 ê 研究〉，《第四屆台灣語言及其教學國際學術研討會論文集》頁 341-349，高雄：中山大學中文系，2002 年 4 月。
<http://203.64.42.21/iug/Ungian/Chokphin/Lunbun/POJtiannau/POJtiannau-TIongsan.htm>

[Tân 2002] 陳鄭弘堯：〈計算台語功能詞 ê 頻率—「台語網」白話字文獻語料庫 ê 應用之一〉，《2002 台灣羅馬字教學 kap 研究國際學術研討會論文集》，台灣羅馬字協會，2002 年 7 月。<http://203.64.42.21/iug/ungian/POJ/siausit/2002/2002POJGTH/lunbun/B1-Henry.htm>

[Têng et 2003] 鄧文淵、陳惠貞、陳俊榮：《ASP 與網頁資料庫設計》，台北：松崗，2003 年 1 月。

[Tiuⁿ 2001] 張裕宏：《白話字基本論：台語文對應&相關的議題淺說》，台北：文鶴，2001 年 3 月。

相關網站

[AKC] 暗光鳥 ê 台日辭典，<http://lomaji.com/Su-khou/index.html>

[AT] Asc2TP 公用程式，<http://lomaji.com/INFO/BOOKS/NNG-THE/asc2tp/>

[HC] 華語文 Concordance 程式，<http://www.sinica.edu.tw/ftms-bin/kiwi.sh?qtype=1>

[HK] 行政院客家委員會客語語音辭典，http://dictionary.ihakka.net/kaga_flash/

[TC] 台語文 Concordance 程式，<http://203.64.42.21/TG/Concordance/>

[TGB] 台語網，<http://tw.groups.yahoo.com/group/taigubang/>

[TH] 台文華文線上辭典，<http://203.64.42.21/TG/sutian>

[TP] Taiwanese Package 文書處理軟體，<http://www.phahng.idv.tw/>

圖 1 台文華文線上辭典的首頁畫面

用白話字查 [thiong] e 結果

Chhih 白話字iah 是台語漢羅會tàng連去台語文Concordance程式，去看台語文語料庫內底這個詞前後文；華語詞連去中研院華語平衡語料庫。

	白話字	台語漢羅	華語		白話字	台語漢羅	華語
1	âm-thiong	暗暢	竊喜	2	lòk-thiong	樂暢	樂觀
3	lòk-thiong	樂暢	樂觀	4	lòk-thiong	樂暢	樂陶陶
5	lòk-thiong	樂暢	歡暢	6	su-thiong	舒暢	舒暢
7	thiong	暢	暢	8	thiong	暢	痛快
9	thiong	暢	快活	10	thiong	暢	酣暢
11	thiong-but-but	暢勿勿	樂滋滋	12	thiong-pán	暢板	愉快
13	thiong-pán	暢板	快意	14	thiong-phut-phut	暢-phut-phut	喜滋滋
15	thiong-sià-kiá ⁿ	暢舍国	花花公子	16	thiong-sià-kiá ⁿ	暢舍国	實梁子弟




圖 2 台文華文線上辭典查詢結果範例

台華線頂辭典使用統計

[\[轉去台華線頂辭典\]](#) [\[Ún-giân e 首頁\]](#) [\[台語線頂字典/統計資訊\]](#) [\[台語文語料庫/台語文Concordance/統計資訊\]](#)

一、使用者

(1) 查詢方式

- 用台語漢字 查lóng總有 11344 pái 34.97% 
 - 用白話字 查lóng總有 3394 pái 10.46% 
 - 用華語 查lóng總有 17699 pái 54.56% 
- lóng總 32437 pái

(2) 查詢e詞頻(Word Frequency)

- 使用白話字查詢 khah chōe pái e 詞：

詞	幾pái	詞	幾pái	詞	幾pái	詞	幾pái	詞	幾pái	詞	幾pái
á	18	hng	16	e	14	吃飯	13	ā ⁿ	10	ê	10
Tui	9	gō-ê	8	什麼	8	認真	8	chhia	7	gōa	7
ho	7	ta ⁿ	7	我	7	海爾默	7	a ⁿ	6	beh	6
HAM	6	he	6	m	6	ma	6	sng	6	芋謏	6
										ài	5

(3) 查詢日期統計

今 á 日	2003/4/21	154 pái	四工前	2003/4/17	328 pái	這個月	2003/4	4562 pái
Cháng	2003/4/20	355 pái	五工前	2003/4/16	284 pái	頂個月	2003/3	3310 pái
Chō--jít	2003/4/19	537 pái	六工前	2003/4/15	173 pái	頂頂個月	2003/2	3073 pái
Lòh-chō--jít	2003/4/18	373 pái	七工前	2003/4/14	199 pái	頂頂頂個月	2003/1	8182 pái

圖 3 台文華文線上辭典統計資訊 (部分)

查詢[欣羨] ê 結果—台語文Concordance 程式			
Beh chhē ē 詞：欣羨			
1	beh代理維他露的經銷權a？」阿梅真	欣羨	：「客戶拍拍=咧，放互lin阿添伊
2	阿瑛kham-khiat的身世。細漢	欣羨	別人食飽穿好，beh買啥，錢討=咧就
3	崧聯考考tiâu，分發去台北讀，真	欣羨	及替伊歡喜。當然，上歡喜猶是阿瑛，這
4	chhōe 物食，全款彼呢仔相好，實在予人	欣羨	，嘛誠予人o-ló！好好仔kā
5	；chit 對戀人結合了後，真是予人	欣羨	，ti 我 心內嘛一直kā
6	講著有夠力，是hō 人	欣羨	ē 才情，「某某人有夠力，拜託伊就
7	chit-ē 五、六十歲 序大，感覺	欣羨	起畏寒，一世人ē 消費，tō m̄
8	bē 肥」，m̄ 是好話，mài	欣羨	。
9	，bē giàn 〈癮〉bē	欣羨	，假那一個佛，你講叨chit 位卡高
10	chhit-thō 物，只有惹起困仔ē	欣羨	：
11	á,án-ni kā in a公	欣羨	講：
12	仙。」這個gin-á已經leh	欣羨	，是án-chóa ^a ？因為i 生

圖 4 台文線上 Concordance 程式查詢範例