中華心理學刊 民101，54卷，2期，185-201
Chinese Journal of Psychology 2012, Vol.54, No.2, 185-201

中文版「語文探索與字詞計算」詞典之建立

黃金蘭[1] Cindy K. Chung[2] Natalie Hui[3] 林以正[4] 謝亦泰[2] Ben C. P. Lam[3] 程威銓[4]
Michael H. Bond[3] James W. Pennebaker[2]

[1]國立台灣科技大學人文社會學科

[2]Department of Psychology, University of Texas at Austin

[3]香港中文大學心理學系

[4]國立台灣大學心理學系

日常生活語言的使用，包括書寫與口語，反映了個人內在的心理狀態、思考型態甚至人格特質。它為心理學研究提供了一扇探索心靈的窗口。Pennebaker研究團隊（回顧見Pennebaker, 2011）採用字詞計算的方式進行語文特性的分析，並發展出電腦程式「語文探索與字詞計算」（Linguistic Inquiry and Word Count, LIWC）。LIWC的核心在於其詞典，目前LIWC2007已可分析80個字詞類別，具有相當好的信、效度。本研究的主要目的即在建立中文版LIWC詞典，並比較其與英文LIWC2007詞典的對等性及效度檢驗。研究一依據原英文版LIWC詞典建立之基本程序，並經各階段嚴謹的反覆檢驗，並依據中文特性加上一些中文特有詞類，最後編修成包含30個語文類別與42個心理類別的中文版LIWC詞典。研究二則蒐集了100篇中、英文對照文章，進行分析比對。結果發現中文LIWC與LIWC2007所偵測到詞類百分比，大部分有極高的相關，可見兩個版本具有相當的對等性。研究三分析憂鬱文本的特性，發現相對於對照組，憂鬱文本使用較多的第一人稱單數代名詞、較少的第一人稱複數代名詞與較多的負向情緒詞。此一結果與使用LIWC分析的英文相關文獻結果一致，顯示中文LIWC也具有一定的效度。整體而言，本論文透過三個研究建立了與LIWC2007具相當對等性且具有效度的中文LIWC詞典。相信中文LIWC日後可提供華語文使用的心理特性分析一項研究利器。

關鍵詞：文本分析、情緒書寫、語文探索與字詞計算

語言與思考的對等性，一直是哲學家與心理學家所關切的議題。姑且不論可否將之劃上等號，兩者的密切相關是不容質疑的。其實，日常生活語言的使用，包括書寫與口語，反映了個人的思考型態、內在的心理狀態甚至人格特質。語文使用的分析可以為心理學研究提供一個探索心靈的途徑。早在20世紀初期，Freud的口誤（slips of the tongue）分析與夢的解析就是利用口語分析探索個體內在需求或潛意識狀態。又如，臨床心理學常用的投射測驗，諸如羅夏克式墨漬測驗（Rorschach Inkblot Test），主題統覺測驗（Thematic Apperception Test）等，都是提供模糊圖形或模糊情境，請個案進行自由聯想或說故事，並就

初稿收件：2011/02/11；一修：2011/09/19；二修：2011/10/03；正式接受：2011/10/07
通訊作者：黃金蘭（chinlanhuang@gmail.com）10607台北市大安區基隆路四段43號 國立台灣科技大學人文社會學科
* Michael H. Bond現職Department of Applied Social Sciences, Hong Kong Polytechnic University; Ben C. P. Lam現職Department of Psychology, Iowa State University, Iowa。

致謝：作者感謝三位匿名審查人的細心評閱，提出各項批評與建議，使本論文在論述上有更加詳盡與精確的修正。本研究部分經費蒙行政院國家科學委員會核予第一作者（計畫編號NSC98-2410-H-001-003, NSC99-2918-I-011-004）之研究計畫，以及核予第五作者（計畫編號NSC99-2410-H-002 -082, NSC99-2918-I-002-024）之研究計畫支持，於德州大學短期進修期間完成本研究，特此致謝。感謝中央研究院中文詞知識庫小組授權使用《中央研究院漢語平衡語料資料庫詞集及詞頻統計》之資料。感謝本詞典修訂過程中參與討論之研究團隊成員：林瑋芳、呂婕、王宇廷、張菀軒、賴慧洵、鍾函君、柯昀青、唐思柔、翁若珊及劉貴全等人。也感謝鄭怡與黃郁喬協助蒐集網路部落格文章。

© 2012 Airiti Press Inc. & Taiwanese Psychology Association

06-黃金蘭.indd 185 2012/6/24 上午 10:58:20

其聯想與故事的語文特性等，進行分析診斷的工具。但這類的投射測驗最為人所詬病的缺點包括需受過專業訓練的評分員評分與缺乏客觀性等。其實專業評分員的工作主要是從個案的敘說文本中，根據某些特定的向度或關鍵詞進行計分。以現在發達的電腦計算環境，將這些工作由電腦程式所取代應可大大提升其客觀性與應用價值。本研究的主要目的即在於延伸發展一項適用於分析中文文本的工具，其應用範圍則可廣泛的用於各種書寫與口語的文本分析。

（一）情緒書寫

口語及書寫文字的使用除了反映個體內在心理狀態外，有時還具有療癒效果。1980年代開始，許多研究陸續發現情緒書寫對於個體的生理與心理健康有所助益。例如，研究發現個體連續四天對創傷經驗進行15分鐘的書寫，可減少其後續看病次數及阿斯匹靈的使用量（Pennebaker & Beall, 1986）；創傷經驗書寫也可以提升個體的免疫力（Pennebaker, Kiecolt-Glaser, & Glaser, 1988）。情緒書寫也有助於減低大學生的思鄉及焦慮情緒且可提升其學期成績（Pennebaker, Colder, & Sharp, 1990; 回顧整理見Pennebaker, 1997；Pennebaker & Chung, in press）。

在確認書寫的療癒效果之後，研究者更應關切的是究竟書寫的哪些特性可反映此種正面效果的機制？文字內容的文本分析提供了最直接的線索，也最能貼近書寫者當下的思緒，因此其內容分析應能提供珍貴的資訊（金樹人，2010）。當然，關鍵的因素還是在於如何由文本分析來探索內在心理機制。欲達此目的，建立有效的分析策略與工具顯然是關鍵的必要條件。文本屬於質性的資料，量化研究如何在不得不犧牲豐富的整體內容與特性評估並轉換成為具體的量化指標之後，仍能夠有效的標示心理特性，一直是一個艱難的挑戰。過去投射測驗的計分與書寫的文本分析就面臨如何有效地將質性資料轉換成量化指標的困境。一般的分析往往需要數名評分員逐篇閱讀文章並評分，不但耗費時間精力，且最常碰到的問題就是評分者信度不高（見Pennebaker & Francis, 1996）。這也正是書寫文本分析最常被質疑缺乏客觀性的主因。有些情緒書寫的研究甚至發現評分員的評分結果並不能有效預測健康（如Graybeal, Sexton, & Pennebaker, 2002），使得主觀評分的效度受到質疑。面對此一挑戰，Pennebaker及其同僚於1990年初開始著手發展電腦程式進行語文特性的分析，開發出「語文探索與字詞計算」（Linguistic Inquiry and Word Count，簡稱LIWC）軟體（回顧見Pennebaker, 2011）。現在該工具已展現出穩定的信、效度，可將質性文本資料有效轉換為量化資料以利統計分析，是研究方法上一個極為重要的突破。

（二）LIWC

Pennebaker的實驗室早期發展之語文字詞分析程式相當簡單，其目的就是在計算書寫文本的不同類別，諸如代名詞、情緒詞及認知詞，的使用百分比。該程式經過十餘年的逐步擴充後越形豐富與穩定，歷經LIWC（Francis & Pennebaker, 1993）、LIWC2001（Pennebaker, Francis, & Booth, 2001）、至目前的LIWC2007（Pennebaker, Booth, & Francis, 2007; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007）已經發展成包含4個一般描述性類別（總詞數、每句詞數等）、22個語文特性類別（如代名詞、冠詞等百分比）、23個心理特性類別（如情感詞、認知詞等）、7個個人化類別（如工作、休閒活動等）、3個副語言學（paralinguistic）類別（如應和詞、停頓詞等）以及12個標點符號類別（如句點、逗點等），總計擁有80個字詞類別的電腦程式。以下介紹均以LIWC2007為藍本加以說明。LIWC至今已具有相當不錯的信、效度，且已被廣泛使用在各類的書寫文本或口語語文特性的分析（文獻整理見Tausczik & Pennebaker, 2010）。

1. LIWC程式簡介

LIWC程式可以分成兩個主要的部分，一個是程式主體（使用手冊見Pennebaker, Booth, et al., 2007），另一則是詞典（Pennebaker, Chung, et al., 2007）。該程式可以一次處理多個文字檔案並計算其內含語文類別的百分比。它運作的方式是每次打開一個文字檔，從第一個詞開始逐詞與詞典進行比對。若該字詞與詞典內的某類別所定義的字詞相同，程式會增加一次該類別的計數。若同一字詞（如我）分屬數個不同類別（如第一人稱單數代名詞、人稱代名詞、代名詞、及功能詞），LIWC程式將會一一在每個類別均增加其計數。如此反覆將該文字檔案的全文或分段進行逐詞比對完後，LIWC程式就會計算每個類別內計數所得的詞數佔總詞數的百分比，並予輸出。輸出之檔案可以供Excel、SPSS或SAS等統計軟體做進一步分析。

LIWC核心部分在於其詞典。LIWC詞典定義了類別名稱以及隸屬該類別的字詞列表。LIWC2007詞

典已含有64個定義類別，在這些類別中有些是相對單純客觀的語文向度類別，如單數第一人稱代名詞、介係詞；有些則是相對主觀的類別，如情緒詞、頓悟詞。這些相對主觀的類別均經過嚴格的程序建立。LIWC2007內的每一個主觀類別的建立程序大約可分成四個階段進行。

第一個階段是候選詞的產生。LIWC的每一個主觀類別之建立均先個別產生一組候選詞。以情緒詞為例，係由研究者參考各類情緒測量問卷（如正向情感與負向情感量表；Watson, Clark, & Tellegen, 1988）、辭典及同義詞典等蒐集相關詞彙。最後由每次至少三至六名研究者針對所蒐集的字詞，進行腦力激盪產生其他相關詞，並將之彙整而得一組候選詞。

第二個階段是評分者評量階段。每個類別均有三個評分者，各自獨立評量每一個候選詞歸屬於該類別的適宜性及是否有其他字詞需加入該類別。最後決定字詞的增刪，其原則為某字詞若獲得至少兩位評分員一致的判斷，就依此判斷來決定該字詞是否保留、刪除或加入類別。接下來，對於那些具有階層特性的類別（如情緒詞類底下可分正向情緒詞與負向情緒詞，負向情緒詞底下又可分焦慮詞、生氣詞及悲傷詞），評分員還需針對上層類別（如情緒詞）的所有字詞逐一判斷該字詞是否應包含在下層類別中（如正向情緒或負向情緒），若還有更下一層的類別則也需判斷該字詞是否應包含在更下層類別（例如，若屬負向情緒，則需再判斷是否屬焦慮詞、生氣詞或悲傷詞）。最後的決定原則為若該字詞有至少兩位評分員有一致的判斷可分在某下層類別，才將該字詞加入所屬下層類別。

第三階段則為心理計量評估階段。從1991年開始建立最初級版的LIWC到1997年，約有總計近八百萬字詞曾被輸入處理。LIWC團隊依據其資料庫中使用率較低（低於 .03百分比）以及信、效度過低的類別進行檢討決定是否將之刪除，同時增加其他類別，如社會歷程類，相對性類別等。新增的類別字詞亦均需經過前兩個階段而產生。最後，並依照資料庫將詞頻太低的字詞予以刪除。

第四階段為更新與擴充階段。最新版LIWC2007詞典亦經過較大幅度的修改。透過包括書面及口語資料的數億字詞分析，再進一步刪除使用率低的字詞，及信、效度低的類別。其次，經由四名評分員檢查詞頻序在前2,000的詞之所屬類別，又據此增加了常用動詞與各類的功能詞類別以及標點符號等類別。最後形成了具有64個類別，總計約4,500個詞與詞幹（word-stem）的LIWC2007詞典。

值得注意的是，LIWC詞典建立的主要目的在於偵測語文使用者之心理特性，而非建立一個涵蓋所有字詞的詞典。如同前文所述，早期LIWC的建立主要以情緒詞與認知詞類為主，之後才逐漸因應不同的研究需求加入其他詞類。究竟哪些詞類會被放入LIWC內，也有其重要的考量。一般而言，語文依其文法特性可以大分為內容詞（或稱實詞）與功能詞（或稱虛詞）。過去的語文分析多著重在攜帶重要主題訊息的內容詞，包括名詞和動詞。由於隨著主題與情境的不同，內容詞的使用變異極大。除了常見的be動詞與助動詞外，很難找到某類名詞或動詞會在不同的主題間穩定且頻繁的出現，因此，目前的LIWC詞典並沒有針對名詞或動詞作比較詳細的類別處理。相對的，LIWC對於功能詞的處理則較為完備。功能詞本身並不攜帶特定的內容意涵，其作用主要在串連句中的內容詞以表達完整的句義。近年來，Pennebaker的研究團隊發現在語言溝通上的語言風格（language style），亦即如何將內容詞加以組織表達的型態，卻是決定於過去常被忽略的、甚至被稱為「垃圾詞」（junk word）的功能詞（Chung & Pennebaker, 2007; Ireland & Pennebaker, 2010; Tausczik & Pennebaker, 2010）。英文的功能詞本身數目雖少於400個詞，但是，卻涵蓋了一半以上的日常語言的使用。近年使用各類功能詞來標記心理特性的相關研究也不斷攀升。儘管目前的LIWC2007比較缺乏名詞與動詞的相關類別，但已經涵蓋了較為完整的各種功能詞類別，且該詞典所涵蓋總數約4,500個詞，就已經大約可以偵測到一般書寫文字的百分之八十五以上的字詞，亦可顯現功能詞在語文使用上扮演的重要角色。

2. LIWC信度檢驗

LIWC作為一個偵測語文心理特性的工具，必然需經信、效度的考驗。一般而言，自評式問卷在題目設計時，會就某個心理特性編制好幾個不同情境或不同描述方式的題目，並假設若這些題目均能夠反映同樣的心理特性，那麼題目間的評量一致性應該相當高，並據此計算內部一致性係數Cronbach's α。通常Cronbach's α大約要在 .7以上才能算是具有可接受的內部一致性。再測信度與折半信度基本上也是建立在相似的原理上。以此邏輯放在LIWC類別的信度檢驗上，固然可以就每一個類別（例如，第一人稱單數）內的

詞求其出現次數的內部一致性，或者是在兩次相同主題（如自我介紹）的書寫或口語文本中，求其相關。但需特別注意的是，語文的使用與問卷的自評方式並不相同。在單一次書寫中，我們不常見到類別內的詞交互使用的情況（例如，使用「我」頻率高，並不一定會也同時得到使用「本人」，「在下」，「自己」等的頻率也高）；在兩次同一主題的書寫或談話中，我們可能也會傾向讓著重點有所變化，而不太會使用相同的字詞內容（如兩次的自我介紹，如果內容都相同，可能反而會覺得太過制式化）。由上述說明，我們不難理解語文分析工具不同於問卷的測量，在內部一致性和信度的測量上可能也無法如一般問卷的標準來評斷。儘管如此，LIWC2007英文版的詞典也進行了內部一致性檢驗程序。所有類別的內部一致性大致落在 .02到 .75之間，若採用比較寬鬆的二分法（每個詞出現與否）的方式則大致落在 .14到 .97之間。以生氣詞這個類別為例，該類別共含有184個與生氣相關的字詞，兩種計算方式得到內部一致性分別為 .55與 .92。除了少數幾個類別偏低外，多數類別的內部一致性都落在可接受範圍（參見Pennebaker, Chung, et al., 2007, Table 1）。

3. LIWC效度檢驗

仿照量表建立的效度檢驗，LIWC的效度當然也可以就每個類別透過自評問卷、評分員評定結果以及LIWC變項進行相關分析檢驗之。LIWC最早的效度檢驗研究在比較不同書寫主題的內容特性及其效應（Pennebaker & Francis, 1996）。該研究將大一新生隨機分配在兩組實驗情境中，請他們就指定的主題進行連續三天的書寫。實驗組被要求書寫他們進大學後最深層的想法與感受；控制組則被要求以不帶情緒的方式書寫一個物件或事件。最後研究者請四位評分員對每一篇書寫針對情緒性、認知及內容等向度進行評量。結果發現評分員的評量與LIWC輸出之相對應的變項間大致上具有可接受程度的相關存在。例如，LIWC可以成功地偵測正、負向情緒與認知策略等（參見Pennebaker, Chung, et al., 2007, Table 1）。不過，同樣需要特別注意的是，LIWC的建立本來就是在解決自評及評分員評定缺乏客觀性與一致性的缺點，因此LIWC與書寫者自評及評分員評定結果的相關通常不會很高。

之後，LIWC效度檢驗多著重在其效標關連效度上。例如，憂鬱症患者使用較多的負向詞與較多的第一人稱單數代名詞（Rude, Gortner, & Pennebaker, 2004）。此外，也有研究發現使用越多正向情緒詞、中等程度的負向情緒詞、以及逐漸增多的因果詞與頓悟詞，書寫者的身體越健康（Pennebaker, Mayne, & Francis, 1997）。

4. LIWC的應用

從1991開始著手建立LIWC程式、2001商業版推出、到最近的2007版發行，LIWC已經被廣泛的應用在心理、商業乃至政治語言的分析。目前已累計有上百篇已發表的相關研究（整理見Tausczik & Pennebaker, 2010）。

偵測語言的使用特性，可以知道當事人的注意力所在。例如正向政治性的廣告使用較多的自我相關詞（我、我們），負向政治性廣告則使用較多的他人相關詞（他、他們）（Gunsch, Brownlow, Haynes, & Mabe, 2000）。也有研究發現說謊時會使用較多的負向情緒詞、動作詞、以及較少的第一人稱單數與排除詞（Newman, Pennebaker, Berry, & Richards, 2003）。研究論文的討論部分因需要整合與分析過去研究結果，在語言使用上要有更複雜的句型，因此得到綜合討論比緒論及摘要使用更多介係詞（Hartley, Pennebaker, & Fox, 2003）的現象。偵測語言使用特性也可以反映當事人心理狀態的變化。Cohn、Mehl及Pennebaker（2004）分析911事件前後網路書寫特性的變化。他們發現，儘管書寫內容可能與911事件無關，在恐怖攻擊後網友書寫較多負向情緒詞，較多認知詞與社會詞。研究也發現前紐約市長朱利安尼（Rudy Giuliani）在面臨其人生風暴時使用較多的認知詞（Pennebaker & Lay, 2002）。

使用LIWC也可以探討個別差異。例如，女性使用較多的社會詞與他人相關詞；男性的語言型態則傾向較為複雜（Newman, Groom, Handelman, & Pennebaker, 2008）。自殺詩人在其作品中使用較多的第一人稱單數代名詞與死亡詞（Stirman & Pennebaker, 2001）。高外向性者是用較多社會詞、正向情緒詞，較少的負向情緒詞與複雜詞（Mehl, Gosling, & Pennebaker, 2006; Pennebaker & King, 1999）。

近年來，Ireland與Pennebaker透過LIWC之功能詞類別的使用一致性，計算出兩篇文本（或對話）的綜合性語文風格匹配度（language style matching，簡稱LSM，Ireland & Pennebaker, 2010）指標。相關研究發現在小團體溝通時，成員間彼此的LSM可以預測團體的凝聚力及其後續合作表現（Gonzales, Hancock, & Pennebaker, 2010）。LSM也可以預測在快速約會

中，雙方的好感度；並且可以有效預測情侶三個月後的關係穩定性，而這些指標甚至比許多研究參與者自陳的感受更具有對實際行為的預測力（Ireland et al., 2011）。

（三）研究目的

LIWC程式最特別之處在於它具有相當大的彈性。研究者可以依其研究目的與需求，建立個人化的詞典，匯入其程式，即可進行個人化的類別與字詞百分比分析。利用這項特色，目前已經有多國語文的詞典紛紛建立，如法文、俄文、義大利文、阿拉伯文、西班牙文、德文、韓文等。本研究的主要目的即在建立中文版的LIWC詞典及其心理計量特性探討。

中、英文的語文使用，縱然有其各自的使用特色，也有許多共通的特性。因此，欲建立相對應的中文版LIWC詞典，可以使用兩種不同的策略。其一為從零開始，仿照英文LIWC的建立過程，或從特定類別的定義開始蒐集在中文中屬於該類別的特定語詞，或從最常見的前數千個詞開始進行類別的建立。另一個策略則是以LIWC2007為基礎，並針對中英文的語文使用特性予以增刪類別以及類別內所屬語詞。本文採取第二種策略，主要的考量有二。首先，LIWC2007已經花了十數年建立，且已經具備相當好的信、效度，應可將之作為語言共通性的基礎，再加入中文語言特性的考量。其次，以LIWC2007為本，建立相應的中文版，亦有利於在相近的基礎上進行跨文化比較。據此，本文透過三個研究以達上述研究目的。

研究一：中文版LIWC詞典的建立

研究一的主要目的為建立中文版的LIWC詞典。本研究以LIWC2007所使用的詞典為主要依據，先透過翻譯以及類別與其所含字詞適當性之確認。然後，更重要的是依照中文特性進行類別以及類別內語詞的增刪。例如，增加中文特有的第二人稱單／複數、量詞、語尾助詞等類別；雖然因中文沒有動詞時態的差異，刪去現在式、過去式與未來式等類別，但也另建了相對應概念的各類時態標記詞；由於華人特別重視家族關係，因而在家族詞（family）也加入了親屬關係的字詞。在增刪類別後，還需經過斷詞系統的處理，確認分詞的適當性。建立過程亦經過多次的小組討論，逐詞確認所使用之語詞與其所屬類別的適當性。整個中文版LIWC（簡稱CLIWC）的建立歷時逾半年才完成。茲將建立過程分為六個階段詳述如下：

第一階段——直接翻譯：如前文所述，LIWC2007詞典由64個類別，將近4500個不同的英文詞及詞幹所組成。這些詞依其語文詞性或意義隸屬於各個類別，且部分類別有上下層的結構關係存在。由此可知，在LIWC詞典中，單一的詞或詞幹可以隸屬於多個類別。考量中、英文詞義並非一對一的對應，我們決定對LIWC2007詞典中每一個詞或詞幹在其所屬各類別均分別進行翻譯。首先，我們整理並合併各類別所包含的詞與詞幹數，總計得到將近一萬零六百個英文詞次與詞幹次。由於在中文並沒有冠詞，也沒有動詞的過去式、現在式以及未來式的區別，我們先將分屬於這四個類別的詞先予以刪除，再針對剩餘詞進行翻譯。基於翻譯的周延性考量，我們將這些詞依照字母序排列大致平均分成五組，每組由兩位大學生個別進行翻譯。如此安排下，每一個詞都有兩位翻譯者翻譯。在正式翻譯前，由作者之一對翻譯者進行工作說明，讓翻譯者充分瞭解其任務是必須依照每個英文詞及其所屬類別之意涵進行翻譯，若同一英文詞可以翻譯成相同意義的不同中文詞也要就其所知一一加入。在翻譯過程中，翻譯者可以自行查詢各類英漢／漢英辭典。本階段經翻譯與整理共得一萬八千餘中文詞次。

第二階段——類別確認：第一階段所獲得一萬八千餘詞次係由英文直接翻譯而得。為了謹慎確認所翻譯得之中文詞在其所屬類別的適宜性，本階段分幾個步驟進行類別確認。首先，為使每個詞的類別適宜性都可以獲得三個判斷反應，本研究另外招募九位大學生進行字詞的類別判斷。每位大學生需針對六千餘個詞，一一就其歸屬於某特定類別的適宜性進行判斷，詞組與判斷員的安排使每個詞均可得到三個判斷反應。同樣的，在正式進行類別判斷前，由研究者之一對判斷員進行工作說明，讓其充分了解每個類別的意涵。本階段經整理，共有一萬三千多詞獲得一致的類別判斷。我們依據這些詞所獲之一致性判斷來決定其保留於詞典或予以刪除。若三位判斷員均認為該詞適宜分在該類別則予以保留，若三位判斷員均認為該詞不適合分在該類別則予以刪除。第二步驟則針對未得一致性判斷的六千多詞，再行招募六位大學生，每人判斷三千餘詞。將這六千多詞在前兩個步驟所得的判斷反應併入整理，每詞共得六個反應。其中有將近五千個詞可依照多數反應來決定是否保留於詞典中，但仍有約一千五百個詞在六個反應中不分軒輊。由於

仍有相當比例的詞未達區辨標準，為求嚴謹，本研究針對這些詞又進行進一步的類別確認程序。第三步驟則由本研究的其中三位作者與另兩位熟悉本研究之研究生，針對此一千五百餘個詞進行類別適當性判斷，最後並以多數反應來決定是否保留於詞典中。經上述三個步驟的類別確認，本階段共整理得各類別總數約一萬五千餘詞次，其中包含六千六百多個不同的詞。

第三階段——斷詞確認：在翻譯的過程中，研究者發現許多因為語言特性差異所可能造成的問題。首先，有些英文的單詞翻譯成中文時可能會需要兩個詞來表示其完整意義。例如，在人稱代名詞的類別中有my、mine、your、yours、his等，在中文直譯為「我的、你的、他的」，這些翻譯分別均由兩個詞（「我／你／他）」與「的」）組成。又如，on、in在中文直譯可為「在……上面、在……裡面」，也是分別由兩個詞「在」與「上面／裡面」所組成。此階段的主要任務即在確認前兩個階段所整理得的六千六百多個詞，是否為單一詞。本研究使用中央研究院中文詞知識庫小組（1998b）所發展之線上中文斷詞系統（http://ckipsvr.iis.sinica.edu.tw/）作為斷詞判斷之依據。該系統為目前在國內經常被使用的一個具備未知詞偵測以及語法詞類預測能力的中文斷詞系統（Ma & Chen, 2003a, 2003b）。上述六千六百多詞，經該中文斷詞系統分詞後，研究者再將被斷成兩個詞（或以上）的翻譯詞依據其所屬類別進行修改為適切的詞。例如前例中on在介係詞類別則修改為「在」，在空間類別則修改為「上面」。本階段依上述程序最後重新整理得六千五百多個詞。

第四階段——逐詞確認、加詞及詞幹建立：上述階段所得之六千五百多個詞均以英文詞典為本，經翻譯後，再透過上述方式整理而產生。為了進一步確認類別定義的精確性及其所屬詞的合宜性，且避免一些英文需用片語或數個詞表達但中文卻可以用單詞表達相同意義的常見詞（例如，according to、as long as、after all……等等；又如the day after tomorrow、last year等）遺漏於詞典中，本階段經過多次的團體討論與腦力激盪，逐類確認類別定義後，逐詞檢查其適宜性並加詞於該類別。參與討論的成員均為共同進行文本分析的研究者，他們對於整個研究架構都有相當的瞭解。每次的討論均至少有五位成員（其中包含三名本文作者）參與，每位成員將有疑義的詞或是想到可以加入的詞提出討論，最後不論是刪詞或加詞均需得到全體共識。

本階段還有一項重要的程序，即詞幹建立。LIWC程式可以利用詞幹加星號的方式來偵測同詞幹的詞群。例如，以infinit*即可偵測到所有以infinit為首的詞，如infinite、infinitely、infinitive、infinitude、infinity等詞。由於中文以詞，而非單字，為基本的意義單位，故相對於英文的以前幾個字母為詞幹，中文將以詞首字為詞幹（或稱詞素）來標示相對應的概念。中文單字本身具有特定意涵，更能以詞幹來偵測相關意涵的詞，例如，可用悲*來偵測悲傷、悲痛、悲哀、悲慟、悲苦等等負向情緒詞。本階段利用已經建立的詞典為依，確認建立詞幹的可能性。本階段討論的過程亦參考國語辭典電子版（國語推行委員會，2007）或斷詞系統以確認其適當性。判斷程序大致為小組成員經討論，並共同查詢辭典確認包含該詞首的詞均具該類別意涵才會將該詞幹列入CLIWC詞典中。期間或有一、兩個詞幹有少數例外者，也在考量它可以捕捉更高比例的相似詞，足以容忍少數誤差者，才會納入使用。

第五階段——增加類別：如前文所述，源於中文的獨特性，有些在中文常用但英文不存在的詞類也應加入CLIWC。例如，第二人稱複數代名詞（如你們）；中文有各種數量詞單位（如個、位、條、張、顆……等等）；又如，中文有各種句尾助詞（如啊、了、呀、喔、呢等等）。還有一些多功能詞，例如「的」，可以放在句尾，可以加在具體名詞或代名詞之後可表示所有格；也可以加在抽象名詞之後可表示形容詞；「地」加在抽象名詞之後可表示副詞…等等。相對於英文，這些都是中文獨有的特性，應被放入中文版的詞典中，提供相關研究者進行其使用特性的探討。此外，中文雖然沒有動詞時態之分，可是在文句中卻仍有一些標定時態的語詞（如，現在、過去、已經），也可分類放入詞典中。

研究者參考了一些語法資料及語詞資料庫，發現「中央研究院漢語平衡語料資料庫詞集及詞頻統計 」（Word List with Accumulated Word Frequency in Sinica Corpus 3.0；中央研究院中文詞知識庫小組，1998a）的資料庫提供了相當豐富的詞類資訊。該資料庫依據五百萬個詞次整理其出現頻率，得到近十五萬個帶有詞類、詞頻及累積頻率標記的詞。本研究在徵得該單位之授權後，參考該資料庫提供的詞頻類別，在本詞典中加入第二人稱複數代名詞（YouPl，如你們、您們等）、後置詞（PrepEnd，如之前、以上等）、特指定詞（SpecArt，如某、該等）、量詞（QuantUnit，

如箱、顆等）、語尾助詞／嘆詞（Interjunction，如喔、嗎等）、多功能詞（MultiFun，如的、地、是、有等）等。上述類別內的字詞均選自中研院詞頻資料庫中詞頻前5000之字詞。本研究亦增加時態標定詞（TenseM），其中又細分四個子類，過去式（PastM，如以前、昨天等）、現在式（PresentM，如今天，目前等）、未來式（FutureM，如今後，即將等）以及持續或完成式（ProgM，如了、已經等）。

至此，中文版LIWC詞典總計增加十一個類別。除此之外，我們將前四階段所產生的詞典與上述詞頻資料庫中使用頻率最高的前兩千詞進行比對，發現覆蓋率達78.4%，其中有106詞符合本詞典所定義的功能詞類別，但尚未包含在本詞典中，在本階段也一併將之加入詞典，以提升其文本的偵測率。

第六階段——詞類結構檢查及整體再確認。如前文所述，LIWC詞典中的部分類別間存在有階層關係，為了避免在前面幾個階段的翻譯、類別判斷及加詞等過程中，破壞了階層關係的完整性，本階段再度進行詞類結構的檢查，以確保其完整性。本階段將有隸屬階層關係的類別分別整理成不同檔案，並依照字詞排序後進行階層比對。比對的標準是屬於下層類別的字詞一定要包含在上層類別內，倘若有屬於上層類別但不屬與任何下層類別的字詞，則再進一步由研究團隊的3～5位成員判斷其是否需刪除、保留、或可以加入下層類別中。這個階段同時也可以對於整體詞典中各個詞所屬類別及詞幹建立適宜性進行再確認。

本研究經上述六階段程序，最後編制完成中文版的LIWC詞典，其中包含語文特性30類，心理特性42類，共有72個類別（如表1所示），總計6,862個詞，在各類別使用次數總計為15,175詞次。本研究完成之詞典對於中央研究院漢語平衡語料詞頻統計資料庫中最常使用的前一千詞之偵測率為83.5%，前二千詞之偵測率為76.2%，前五千詞之偵測率則為58.7%。需特別注意的是，在詞類資料庫中，隨選取詞頻的下降，名詞類的詞數會大幅增加，但LIWC詞典並未包含名詞，故其偵測率有明顯的下降。上述數據顯示，本詞典對於常用的語詞，具有相當不錯的偵測率。

研究二：對等性探討

為了探討CLIWC與LIWC2007的對等性（equivalence），我們自網路上蒐集了100篇中、英文對照文章，進行分析比對。這些文章的來源及型態涵蓋了雜誌、散文、新聞報導、演講稿、機構文宣品及電影對白腳本等等，其中50篇文章的原文為中文，另50篇文章的原文則為英文。

（一）結果

1. 資料前置處理

由於LIWC程式分析時係以空格作為字詞邊界之標定，然而中文書寫並沒有以空格作為字詞之邊界，故將文本匯入LIWC分析前需先進行斷詞。本研究先將所有中文的文本檔案一一放入中研院之網路斷詞系統進行斷詞，並將其輸出檔之字詞類別標記刪除，同時也將全形標點符號轉換為半形。經上述處理之檔案即可整批匯入LIWC程式，以CLIWC詞典進行分析。英文的文本檔案則以LIWC2007內建詞典進行分析。

2. 資料分析

本研究首先將中、英文的LIWC輸出的各個對應變項兩兩計算其簡單相關（如表1所示）。結果發現在77個比較變項中，僅有5個變項未達顯著相關（分別為阿拉伯數字、停頓贅詞、頓號、縮寫符號及其他標點符號）。依據Cohen（1992）的標準，在這些顯著相關中，有55個變項達高效果量，9個變項為中等效果量。本研究發現，以相同內容的文本分析得到的LIWC輸出變項，在中、英文有高比率高相關。此結果顯示CLIWC與LIWC2007具有相當高程度的對等性。

進一步以t檢定則發現，在總數77個差異檢定，中、英文對應變項間有顯著差異者達54個變項。面對此結果，我們需先注意的是，同一篇文章在翻譯過程中，由於語言特性的差異，必然不是逐字句的翻譯，而是在保留完整語意下，依其語言特性以適當的方式表達，因此一定不可能得到完全一對一的對等翻譯。也就是不可能得到變項的完全對等。此外，由於這些都是樣本數為100的配對*t*檢定，很小的差異（如在本研究中差異在 .06%）就有可能達顯著水準。若改以差異分數來看，比較大的差異來自於英文詞典可以偵測到文本的百分比顯著高於中文版詞典（80.3% vs. 70.7%）。此外，相較於中文版詞典，英文詞典可以偵測較高百分比的功能詞、代名詞、助動詞、介係詞與社會詞；其中前四者為語文文法向度，中文版詞典在編制時乃以英文詞典為主要藍本翻譯得，並非由中文文法為依歸，因此在文法向度上英文版涵蓋較高的偵測率是可以理解的，至於為何在心理向度上的社會

表1 中英文LIWC詞典對照及其對等性檢驗在各變項之平均值及其檢定表

變項名稱	簡寫	英文平均數（標準差）	中文平均數（標準差）	相關值	*t*檢定值	總詞數	範例
系統自行產生							
總詞數	WC	509.7(212.1)	494.6(197.9)	.90***	1.60		
每句詞數	WPS	17.3(7.8)	22.7(15.5)	.72***	-4.74***		
超過六字母字詞	Sixltr	20.9(7.3)	0.3(0.5)	.28**	28.80***		
抓取率	Dic	80.3(7.8)	70.7(9.2)	.85***	19.70***		
阿拉伯數字	Numerals	0.6(0.7)	0.0(0.0)	-	8.37***		
句號	Period	7.5(4.8)	6.4(4.5)	.95***	7.49***		
逗號	Comma	5.2(1.6)	8.8(3.1)	.50***	-13.45***		
冒號	Colon	0.1(0.2)	0.3(0.3)	.38***	-3.91***		
分號	SemiC	0.2(0.7)	0.3(0.8)	.96***	-2.85**		
問號	QMark	0.8(1.3)	0.7(1.1)	.97***	2.03*		
驚嘆號	Exclam	1.0(2.3)	0.6(1.8)	.94***	3.76***		
破折號	Dash	1.6(1.7)	0.1(0.3)	-.03	8.90***		
引號	Quote	0.8(1.1)	1.4(1.5)	.67***	-5.02***		
縮寫符號	Apostro	2.3(2.1)	0.0(0.0)	-.04	11.02***		
括弧	Parenth	0.1(0.2)	0.2(0.3)	.29**	-1.81		
其他標點符號	OtherP	0.1(0.3)	0.1(0.2)	.19	1.62		
所有標點符號	AllPct	19.2(8.0)	18.9(4.9)	.81***	0.58		
中英詞典共有類別							
功能詞	funct	50.9(7.4)	47.2(8.7)	.87***	8.56***	878	或許、許多、那些
代名詞	pronoun	11.3(7.1)	7.8(6.1)	.97***	17.28***	65	你、她們、在下
特定人稱代名詞	ppron	6.8(6.1)	5.2(5.0)	.95***	8.08***	35	他、大家、你們
第一人稱單數代名詞	i	2.3(3.1)	2.1(2.7)	.96***	1.87	9	本人、自己、我
第一人稱複數代名詞	we	0.8(1.3)	0.6(1.0)	.93***	4.34***	7	我們、我倆、咱們
第二人稱代名詞	you	1.8(2.6)	1.2(1.9)	.95***	6.45***	5	你、妳、您
第三人稱單數代名詞	shehe	1.2(1.6)	0.9(1.1)	.89***	4.56***	6	他、她、這位
第三人稱複數代名詞	they	0.7(0.7)	0.3(0.4)	.57***	8.70***	2	她們、他們
非特定人稱代名詞	ipron	4.5(1.9)	2.9(1.6)	.68***	11.34***	33	一切、這些、其他
動詞	verb	12.2(5.0)	11.2(3.3)	.73***	3.07**	296	分享、開車、聽
助動詞	auxverb	7.8(3.2)	2.3(1.2)	.49***	19.47***	32	不必、可能、應該
副詞	adverb	3.4(1.5)	6.7(2.5)	.71***	-17.99***	199	曾經、漸漸、那麼
介係詞	preps	13.6(2.4)	6.3(1.5)	.56***	36.27***	67	從、依照、把
連接詞	conj	5.5(1.7)	5.2(1.7)	.60***	1.93	94	和、一旦、不僅
否定詞	negate	1.3(1.1)	1.7(1.0)	.68***	-4.31***	37	不要、未必、沒有
概數詞	quant	2.6(1.0)	2.2(1.1)	.28**	3.60***	101	一些、所有、眾多
數字	number	1.0(0.7)	2.0(1.0)	.37***	-9.55***	107	一、八百、兩萬
髒話	swear	0.03(0.1)	0.04(0.1)	.45***	-1.26	55	去你的、智障、機車
社會歷程詞	social	8.9(5.2)	6.7(4.0)	.92***	10.52***	587	家人、接納、打招呼
家族詞	family	0.4(0.5)	0.4(0.6)	.80***	-1.06	136	公婆、兄弟、孫女
朋友詞	friend	0.2(0.3)	0.1(0.2)	.47***	2.93**	46	同伴、朋友、麻吉
人類詞	humans	0.9(0.9)	1.3(1.0)	.73***	-5.25***	54	人民、成員、群眾
情感歷程詞	affect	5.3(2.3)	4.6(2.3)	.74***	4.25***	1,304	氣憤、感恩、失望
正向情緒詞	posemo	3.5(1.8)	2.6(1.5)	.64***	6.21***	476	信心、滿足、祝福

表1　中英文LIWC詞典對照及其對等性檢驗在各變項之平均值及其檢定表（續）

變項名稱	簡寫	英文平均數（標準差）	中文平均數（標準差）	相關值	*t*檢定值	總詞數	範例
負向情緒詞	negemo	1.8(1.4)	1.3(1.2)	.83***	5.56***	693	擔憂、猜疑、報復
焦慮詞	anx	0.3(0.4)	0.3(0.4)	.80***	0.29	111	不安、掙扎、緊繃
生氣詞	anger	0.6(0.7)	0.3(0.4)	.60***	5.55***	249	可惡、抱怨、破壞
悲傷詞	sad	0.4(0.5)	0.2(0.3)	.75***	5.81***	128	心痛、沮喪、無力
認知歷程詞	cogmech	14.2(3.4)	17.5(3.6)	.74***	-12.84***	1,058	理解、選擇、質疑
洞察詞	insight	1.9(1.3)	3.1(1.5)	.55***	-8.94***	287	了解、恍然大悟、體會
因果詞	cause	1.6(0.9)	1.7(1.0)	.52***	-1.41	129	引起、使得、變成
差距詞	discrep	1.3(0.9)	2.3(1.2)	.58***	-9.80***	83	不足、納悶、期待
暫訂詞	tentat	1.9(1.1)	2.2(1.1)	.62***	-2.97*	164	大約、未定、差不多
確切詞	certain	1.3(1.0)	1.5(1.1)	.66***	-2.36**	119	不容置疑、必然、保證
限制詞	inhib	0.8(0.7)	0.9(0.7)	.42***	-1.43	177	廢止、不准、規則
包含詞	incl	4.7(1.7)	4.9(1.4)	.48***	-1.26	70	包括、附近、添加
排除詞	excl	1.7(1.1)	1.8(1.2)	.65***	-0.96	39	取消、但是、除外
感知歷程詞	percept	2.1(1.4)	1.9(1.2)	.57***	1.77	478	溫暖、經驗、注視
視覺詞	see	0.7(0.7)	0.8(0.6)	.40***	-0.50	134	外貌、閃亮、綠色
聽覺詞	hear	0.6(0.7)	0.5(0.6)	.61***	0.21	89	大叫、聽見、談話
感覺詞	feel	0.5(0.6)	0.3(0.4)	.55***	4.37***	123	平滑、彈性、觸摸
生理歷程詞	bio	1.9(1.8)	1.5(1.7)	.88***	4.70***	844	頭暈、流汗、擁抱
身體詞	body	0.5(0.7)	0.4(0.6)	.75***	1.64	257	脖子、皮膚、腸胃
健康詞	health	0.7(0.7)	0.4(0.6)	.75***	4.97***	377	失眠、醫生、運動
性詞	sexual	0.1(0.3)	0.1(0.3)	.71***	0.81	116	上床、性慾、裸體
攝食詞	ingest	0.5(1.2)	0.5(1.2)	.92***	1.38	171	消化、吃、煮菜
相對詞	relativ	14.0(3.2)	12.9(3.0)	.46***	5.21***	1,064	以前、相比、達到
移動詞	motion	2.0(1.2)	2.7(1.3)	.40***	-5.48***	260	通過、靠近、參加
空間詞	space	6.8(2.3)	5.6(2.0)	.53***	5.51***	393	裡面、街道、台上
時間詞	time	5.1(1.7)	4.1(1.8)	.62***	6.38***	398	期間、過去、秋天
工作詞	work	2.7(2.2)	2.9(2.7)	.84***	-1.17	577	工廠、面試、薪水
成就詞	achieve	2.5(1.3)	2.1(1.4)	.33**	2.66**	352	擅長、盡責、高手
休閒詞	leisure	1.3(1.6)	1.1(1.5)	.84***	2.18*	327	唱歌、輕鬆、假期
家庭詞	home	0.4(0.6)	0.4(0.5)	.63***	1.22	159	房子、家人、寵物
金錢詞	money	0.9(1.4)	0.7(1.5)	.92***	2.65**	257	富有、年薪、折扣
宗教詞	relig	0.5(1.3)	0.6(1.9)	.90***	-1.34	278	上帝、慈悲、信仰
死亡詞	death	0.2(0.4)	0.2(0.3)	.69***	1.66	130	亡故、自殺、遺囑
應和詞	assent	0.1(0.3)	3.8(1.9)	.38***	-20.22***	41	瞭解、真的、好
停頓贅詞	nonfl	0.1(0.2)	0.5(0.5)	-.05	-6.63***	7	呃、然後、那
填充贅詞	filler	0.2(0.2)	0.7(0.6)	.30**	-8.27***	8	就、像是、話說
中文詞典特有類別							
第二人稱複數代名詞	youpl					5	你們、妳們、汝等
後置詞	prepEnd					41	之中、以上、為止
特指定詞	specArt					17	本、該、每
量詞	quanUnit					121	條、頭、枝
語助詞	interjunction					34	呢、嗎、吧
多用途詞	multiFun					11	的、有、是

表1 中英文LIWC詞典對照及其對等性檢驗在各變項之平均值及其檢定表（續）

變項名稱	簡寫	英文平均數（標準差）	中文平均數（標準差）	相關值	*t*檢定值	總詞數	範例
時態標定詞	tenseM					84	已經、之前、日後
過去時態標定詞	pastM					27	去年、剛才、昔日
現在時態標定詞	presntM					25	目前、現今、平時
未來時態標定詞	futureM					22	之後、即將、將來
延續時態標定詞	progMp					10	了、至今、近期
英文詞典特有類別							
冠詞	article						
過去式	past						
現在式	present						
未來式	future						

*p < .05. **p < .01. ***p < .001.

詞的偵測率也較高，則可能是一個有趣的文化議題，有待進一步研究釐清。另外一方面，我們也發現在每句平均詞數、逗號、副詞、認知詞與應和詞，則是中文版詞典的偵測百分比高於英文版詞典。前兩者的差異可以反映語文使用習慣的差異，中文的句型通常比較複雜，或者甚至省略主詞將兩句併為一句表達等均頗為常見，因此在每句平均詞數與句內使用逗號均較高。在副詞與認知詞及應和詞的使用上，在詞典編譯之初，研究團隊在討論時就透過腦力激盪增加了許多屬於該類的詞彙，是否因此使得其偵測率增加，或者是它也顯現了文化差異，亦有待進一步釐清。除了上述的差異較為明顯外，其他的顯著差異值均小於百分之二。

眾所周知，中文和英文是兩種相當不同的語言系統，在使用習慣上也有相當大的差異。基於上述結果與討論，在語文對等性分析上，兩詞典的偵測率與相關程度，大致都達令人滿意的程度。一些使用百分比的差異則可能反映了兩種語言使用的特性差異，有待進一步探索。

研究三：效度檢驗

用於研究語文使用及心理特性的詞典之實用性判準中，高偵測率並非最關鍵的指標，能夠有效測量到使用者的心理特性才是最重要的判準。研究三嘗試以憂鬱者的書寫特性作為CLIWC的效度檢驗。Stirman與Pennebaker（2001）曾分析近300篇的詩作，他們發現相對於其他詩人，自殺詩人的作品使用較多的第一人稱單數代名詞（I、my、me）、較少具有社會連結意涵的第一人稱複數代名詞（we、ours、us），但是在正、負向情緒詞的使用則無差異。Rude等人（2004）依照BDI（Beck Depression Inventory; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961）及IDD-L（Inventory to Diagnose Depression-Lifetime; Zimmerman & Coryell, 1987）分數，將大學生分成憂鬱組與非憂鬱組，並將非憂鬱組進一步分成先前憂鬱組與從未憂鬱組。結果發現憂鬱的大學生比非憂鬱學生使用較多第一人稱單數代名詞，較多的負向情緒詞。整體而言，先前憂鬱組的第一人稱代名詞與未曾憂鬱組並無顯著差異，但若將書寫文本分段分析，則發現先前憂鬱組的文章隨著段落第一人稱單數代名詞使用逐漸增加，到文本末段則顯著高於未曾憂鬱組。上述結果均符合Beck（1967）的憂鬱認知模型以及Pyszczynski與Greenberg（1987）的自我焦點（self-focus）模型。

依據上述使用LIWC進行文本分析的研究結果，本研究預期，倘若CLIWC的詞典類別也具有偵測心理特性的效果，應可重複驗證過去的典型發現。也就是，在台灣蒐集得的憂鬱書寫文本也應該可以得到上述特性。亦即，相對於對照組，憂鬱組的書寫文本有較高的第一人稱單數代名詞，較少的第一人稱複數代名詞以及較多的負向情緒詞。在本研究中，對照組的選取則亦以強調情緒書寫，分享經驗與心情為原則，避免純粹學術討論，或單純資訊蒐集的討論版。各類網站貼文的類型繁多，經搜尋後決定採用打工版，該版多為打工經驗的分享，可能涵蓋各種經驗與心情，符合本研究的要求。基於打工版的經驗分享特性，本研究進一步預測，相對於憂鬱組，對照組有較多的工作詞、成就詞與金錢詞。

（一）研究方法

本研究從台灣相當流行的電子公佈欄（Bulletin Board System, BBS）中的批踢踢（PTT）站蒐集文本進行分析。PTT的使用族群年齡平均約21歲，多為高中職以上及大學生為主（引自張惠蓉、何玫樺與黃倩茹，2008）。本研究的主要探討族群為憂鬱患者，並以打工經驗分享的族群為對照組，故從PTT站的prozac版（憂鬱版）與打工版各選取文章30篇進行分析。文章選取的標準為文章長度至少一百個詞，且有經驗或情緒分享，而非純粹資訊獲取的文章。由於PTT並未要求貼文者在文章中註明性別與年齡，因此我們並不能區辨每篇文章作者的性別與年齡，故本研究並不對此二變項進行分析或控制。

（二）資料分析與結果

同樣地，本研究在文章蒐集後、匯入LIWC分析前，需要先將每篇文章的錯別字，標點符號等進行修正，並透過網路使用中研院斷詞系統進行斷詞處理，去除其詞類標記及標點符號全半形轉換等前置性處理。經前置處理後的文本再匯入以CLIWC為詞典的LIWC程式分析。

本研究首先將兩組文本的LIWC輸出變項中的第一人稱單數代名詞、第一人稱複數代名詞及負向情緒詞進行獨立t檢定（如表2所示）。雖然有些變項並不符合變異數相等的假設，但經調整自由度後，本研究結果仍如預測。在第一人稱單數代名詞，憂鬱組顯著高於對照組，$t(58) = -7.56$, $p < .001$；在第一人稱複數代名詞，則為憂鬱組顯著低於對照組，$t(32.02) = 3.38$, $p < .01$；在負向情緒詞，憂鬱組亦顯著高於對照組，$t(40.58) = -7.64$, $p < .001$。上述結果均與過去文獻對於憂鬱特性的發現一致。接下來，則針對工作詞、成就詞與金錢詞進行兩組獨立t檢定。結果發現三者均為對照組（即打工組）高於憂鬱組，檢定值分別為$t(58) = 4.38$, $p < .001$；$t(58) = 3.66$, $p < .001$；$t(35.83) = 5.39$, $p < .001$。此結果亦符合打工版經驗分享的特性。

表2　研究二之檢驗變項在各組之描述統計與檢定值

	打工組		憂鬱組		*t*檢定值
	平均數	標準差	平均數	標準差	
第一人稱單數	3.35	1.93	7.50	2.30	-7.56***
第一人稱複數	0.53	0.68	0.10	0.15	3.38**
負向情緒詞	1.38	0.97	4.65	2.13	-7.64***
工作詞	3.81	2.00	1.81	1.50	4.38***
成就詞	2.32	1.03	1.44	0.83	3.66***
金錢詞	1.29	1.05	0.20	0.36	5.39***

**$p < .01$. ***$p < .001$.

綜合討論

語文使用可被視為一種內隱式的行為指標，它可以避免許多使用自陳式問卷所面臨參與者的覺察猜測、反應偏向或社會讚許等介入污染的疑慮。語文的使用分析為心理特性的探討開了一扇窗，也為心理特性與機制研究提供了一個重要的行為指標。過去的語文分析多需要依賴複雜的人工編碼系統，耗時費力，且客觀性也頗受質疑。各式電腦化的語文分析工具遂因應而生，LIWC即是在國外心理學界已被廣為應用且仍深具潛力的一種字詞分析工具。最近的LIWC2007版已經是一個具有極佳的信效度之分析工具，也已有多國詞典紛紛建立，但是尚無透過前述嚴謹編製程序的相應中文詞典。本研究依據原英文版LIWC詞典建立之基本程序，並經各階段嚴謹的反覆檢驗，並加上一些中文特有詞類，編修建立中文版LIWC詞典。

從中、英文對照的翻譯文本分析發現CLIWC與LIWC2007所偵測到大部分詞類百分比間有極高的相關，可見兩個版本具有相當的對等性。以此為基礎，本研究進一步分析憂鬱文本的特性，發現相對於對照組，憂鬱文本使用較多的第一人稱單數代名詞（我、自己）、較少的第一人稱複數代名詞（我們、咱們）與較多的負向情緒詞（難過、討厭）。此一結果也與使用LIWC分析的其他相關文獻結果一致（Rude et al., 2004; Stirman & Pennebaker, 2001），顯示CLIWC也具有一定的效度。研究者及同僚曾經測試數百篇不同類型的故事或情緒書寫文本，CLIWC詞典的偵測率大約都可在百分之八十以上，顯示此工具已經具有相當高的文本分類分析的抓取能力。以下將針對「字詞計數」研究取向之優缺點以及未來CLIWC之詞類的擴充加以說明討論。

（一）「字詞計數」研究取向之優缺點

在諸多語文分析的取向中，LIWC採取的是量化的、字詞計數(word count)的取向（Pennebaker, Mehl, & Niederhoffer, 2003）。不同於質性分析取向，或是評分員加入脈絡考量判斷之方式，字詞計數取向既然取

其客觀量化之特點，就無可避免會遺漏上下文脈絡周延性的考量，造成誤差的介入。舉例而言，LIWC初期使用者，多少會經驗到少數錯誤歸類的情況。例如，由於英文詞具有一詞多義的特性，LIWC抓取特定類別之單詞有時會出現錯誤歸類的情況，在資料分析上就會增加其誤差，減低其效果之敏感度。不過，正如同其他的量化研究，LIWC既然是量化取向，又以類別為單位進行計數，必然需經誤差容忍度的考驗。倘若在存有誤差的情況下，仍具有其預測效果，此取向應仍有其存在之價值。

同樣的，在使用CLIWC除了會碰到上述相類似的問題之外，中文文本在匯入LIWC程式之前，還需要經過斷詞的處理。目前常用的斷詞系統雖無法保證斷詞的絕對正確性，但大抵仍按照國人常用的字詞使用方式進行，就像所有的分析工具一樣，誤差勢必存在，但仍有一定敏感度。不過，斷詞後若不經過人工再確認過程，難免會增加誤差量，而減低其效果檢驗之敏感度。如果研究者可以在斷詞後進行人工閱讀之再確認，校正明顯的斷詞錯誤，應可減低誤差，提高檢驗之敏感度。這樣的人工化處理當然耗時費力，因此研究者可依據樣本大小，決定使用何種策略。樣本大者，可以容忍斷詞錯誤的誤差；樣本小者，則可以承擔重新校正的人力消耗。

LIWC以詞為單位進行分析，因此在情緒詞的分類上，並無法區分負面情緒詞的否定應該是正面情緒，正面情緒詞的否定應該是負面情緒。LIWC會將之分成一個否定詞加上一個負（正）面情緒詞。這樣的分類誤差也是使用者常會萌生的疑慮。從某方面來說，這固然可以說是電腦化語詞計數取向的缺點，是LIWC需忍受的誤差。但是，從另一個角度看，也提供了語言使用心理特性的探討的一項途徑。以情緒表達為例，倘若某人對A有負面印象，當他使用語文表達時選擇使用「我不喜歡A」或「我討厭A」，其實也代表了心理特性的差異。也就是說，傾向使用否定詞加上正面情緒或否定詞加上負面情緒，而非採用肯定句直接表達的個體，也顯現了特定之心理特性。LIWC含有「否定詞」類別，亦可從此類別與情緒類別或其他變項的相關性，一窺此心理特性。

此外，需特別說明的是，語文的使用僅為一種標記（marker），反映書寫者的內在心理狀態與特性。因此，其研究邏輯是必須由上而下（top-down）的推衍，由理論出發建立心理歷程假設，再從假設推衍出特定的語文特性；並非使用由下而上（bottom-up）的歸納邏輯。倘若由理論出發推導出憂鬱症患者具有負向情緒的陷落，以自我為關切點，那麼在語文特性分析時就只需要針對正、負向情緒以及第一人稱單數代名詞進行檢定即可，不需對其他變項進行檢驗。例如，在研究三中，我們並不需要假設憂鬱症與打工經驗書寫在知覺詞或常用動詞上有特定的差異。因此，儘管LIWC提供了數十個類別的輸出變項，在資料分析層次並不需要就所有類別都一一進行檢驗。

（二）CLIWC之詞類擴充

如前文所述，在國外，針對LIWC分析的研究已經累積相當的成果，並且也已經建立不同語言系統下之修訂，其中包含了歐陸語系的西班牙語、丹麥語、法語、德語、義大利語、土耳其語與阿拉伯語，連亞洲地區的韓語也包含在內。這些研究對於部分類別（如各類代名詞，情緒詞等）的心理特性探討均顯示相當穩定的結果與預測力（如Lee, Park, & Seo, 2006)。更重要的是，世界十大常用語系裡，僅剩葡萄牙文尚未加入此一研究行列，日文版本正在發展中。做為中文版LIWC的發展，本論文著實負有銜接此一主流研究之使命，並為華人與西方字詞研究開啟一可能的對話橋梁。因此，儘管本版本的詞典涵蓋之類別或未臻完善，但作者認為本文所建立之詞典仍應及早公開（有興趣使用此詞典者，可自下列網頁獲取相關資訊：https://sites.google.com/site/chineseliwc/），提供有興趣的研究者作為其分析工具。此外，本文建立之詞典係以英文版LIWC2007為藍本建立，在進行跨文化的語文使用心理特性的研究上，也提供了一項可以相互比較的工具。

當然，目前中文版的中文詞典仍屬於初步完成的階段，基於此套分析工具增添詞典類別的彈性相當大，未來還可以再逐步增加或進行細部的修正，並陸續公布於CLIWC網頁上。未來相關研究者若就其研究需求建立特定之語詞類別，CLIWC編制小組亦會在徵得其同意下，增加額外的詞典或加入於CLIWC內，並公布於網頁，在不違反智慧財產權為前提下，供相關研究者使用。未來，除了需要有更多類型的文本與不同領域的研究分析，以更加確定此工具在中文語文環境有良好的適用性之外，研究者認為增加一些更能反映華人心理特性的語詞類別，也是未來使此詞典能夠更切合本土文化的努力方向。

舉例而言，中文的語文使用的獨特性，包括為數眾多的量詞（如件、顆、頭、條、個、本、次、回等等）、語尾助詞（如啊、嗎、喔、呢、咧、呀、等等）以及疊詞（如亮晶晶、水汪汪、雪花片片、濕漉漉）等，都是英文所沒有的特性。其使用之心理特性探討應該也是相當有趣的議題。前兩類詞業已納入本版本之CLIWC詞典，已經可以供有興趣的研究者進行探討；至於後者，研究者正準備著手進行相關語詞的收集，待建立完成將陸續在網頁公布相關資訊。

又如，針對本土文化的情緒表現特色來增修類別應更能夠反映華人的情緒特性。在情緒書寫研究中，情緒詞的使用分析尤顯其重要性。因此，情緒詞類別也是LIWC最早建立的類別。在個別差異的研究也常發現在情緒詞使用上的差異。例如Pennebaker與King（1999）發現，外向性高的人會傾向使用較多的正向情緒詞，較少的負向情緒詞。憂鬱症患者則使用較多的負向情緒詞（Rude et al., 2004）。目前LIWC2007的詞典僅將情緒詞大分為兩類，正向情緒與負向情緒；負向情緒再細分為焦慮，生氣及悲傷等三類。近期跨文化比較的研究指出，相對於西方文化追求高激發正向情緒，東亞文化的理想情緒則是低激發正向情緒（Tsai, 2007；Tsai, Knutson, & Fung, 2006）。在一項比較基督教與佛教文本的研究中，也發現相對於基督教文本有較多的高激發正向詞，佛教文本則有較多的低激發正向詞（Tsai, Miao, & Seppala, 2007）。該研究雖也使用LIWC程式進行文本分析，但由於LIWC詞典僅有正向情緒詞類別，他們只能自行建立低激發正向情緒與高激發正向情緒類別。很可惜的，他們並沒有嚴格的遵循LIWC詞典類別建立的客觀程序來建立類別詞典。LIWC情緒類別的再細分之重要性可見一斑。

研究者認為在CLIWC詞典可建立正式的情緒次類別，將情緒類別從激發狀態與情緒價兩個向度區分，如此便可得四種情緒類別，高激發正向，高激發負向，低激發正向與低激發負向。除了上述兩向度之區分外，鄭昭明、陳英孜、卓淑玲、陳學志及梁庚辰（2011）提出華人情緒類別的加性樹狀結構的情緒分類，也是一個重要的分類方式。該研究從傳統華人的七情說出發，將華人日常生活所使用的305個情緒詞分成30類，並利用「加性樹狀結構」進行最近鄰分析，得到屬於華人特有的情緒結構。這個情緒的加性樹狀結構具有層級關係，本身就可以作為一個適用於華人的完整情緒結構。研究者目前也正在著手進行此情緒結構的分類特性檢驗。未來若可提供更能夠反映華人情緒特性的類別之後，研究者相信，在心理健康以及諮詢治療上的運用上，LIWC應更能夠成為有力的分析工具。

又例如，自從Peng與Nisbett（1999）提出華人辯證觀的概念之後，許多相關研究因運而生。不論是延續使用辯證思維（dialectical thinking）一詞的相關研究（如Spencer-Rodgers, Peng, & Wang, 2009；Spencer-Rodgers, Williams, & Peng, 2010）或者是全局思維（holistic thinking）的相關研究（Choi, Koo, & Choi, 2007），甚至是中庸（Zong-Yong）的相關研究（楊中芳，2010），三者雖有概念重點上的差異，但都涵蓋了華人對於變動為常態及對立共存的概念。這方面的本土研究已經逐漸的指出可能的研究方向，這類的思維方式是否也會反映在書寫的文本上？例如，居安思危，否極泰來，悲喜交加，愛恨交織等字詞的使用。目前LIWC詞典並沒有可以反映這類思維的類別，未來詞典的修訂可以納入這類具有華人思維特性的認知詞類，這將對於LIWC在華人社會的應用提供更大的文化敏感度。

華人本土心理學在過去數十年間建立了不少具有本土文化特色的理論，其中質化研究提供了非常重要的概念基礎，過程中也有非常多的研究者累積了極為大量的文本資料。此外，在諮商、臨床及工商領域等更是累積了豐富的質性訪談或歷程記錄。中文版LIWC詞典的建立將可提供這些研究者更進一步的將其質化資料轉化成為可供統計分析的量化資料。這類研究的推展可對於質化研究與量化研究之間互相對話提供積極的貢獻。而由於LIWC具有非常大的詞典編修彈性，正可以適切的針對理論來增加類別以因應本土研究的需求，以佐證各種具有文化特色的理論。本研究目前已經建立了基本的詞典架構，未來應可有更大的發揮空間。

在全球化的浪潮中，中文使用者的語文特性分析已經是一個全球不可忽視的研究議題。本研究已初步確立了CLIWC的心理計量特性及其適用性，應能為中文語文研究以及心理特性探討提供一項重要的研究利器。當然，這並不表示CLIWC已具完備性。LIWC2007經過近十年的不斷修正改版，使用四千五百多詞在本論文研究一的文本偵測率就高達百分之八十，可說是一個相當完備的詞典。相對而言，目前CLIWC雖也有近百分之七十一的偵測率，但也有相當的提升空間，有待日後進一步修改成為更符合偵測華語文使用者特性的LIWC詞典。

參考文獻

中央研究院中文詞知識庫小組（1998a）：《線上中文斷詞系統》。2010年8月26日，摘自中文詞知識庫小組網站，http://ckipsvr.iis.sinica.edu.tw/ [Chinese Knowledge Information Processing Group. (1998a). *Chinese word segmentation online system*. Retrieved August 26, 2010, from Chinese Knowledge Information Processing Group Web site: http://ckipsvr.iis.sinica.edu.tw/]

中央研究院中文詞知識庫小組（1998b）：《中央研究院漢語平衡語料資料庫詞集及詞頻統計》。台北：中華民國計算語言學學會。[Chinese Knowledge Information Processing Group. (1998b). *Word list with accumulated word frequency in sinica corpus 3.0*. Taipei: The Association of Computational Linguistics and Chinese Language Processing.]

金樹人（2010）：〈心理位移之結構特性及其辯證現象之分析：自我多重面向的敘寫與敘說〉。《中華輔導與諮商學報》，*28*，187-228。[Jin, S. R. (2010). Structure characteristics of psychological displacement and its dialectical phenomenon: Narratives of the multidimensional self. *Chinese Journal of Guidance and Counseling*, *28*, 187-228.]

國語推行委員會（2007）：《重編國語辭典（修訂本）中華民國九十六年十二月臺灣學術網路四版ver.2》。2010年8月26日，摘自教育部網站，http://dict.revised.moe.edu.tw/ [National Language Committee. (2007). *Taiwan Ministry of Education online dictionary version 2*, Retrieved August 26, from Ministry of Education Web site: http://dict.revised.moe.edu.tw/]

張惠蓉、何玫樺、黃倩茹（2008）：〈線上社會支持類型探討：以PTT精神疾病版及整型美容版為例〉。《新聞學研究》，*94*，61-105。[Chang, H. J., Ho, M. H., & Huang, C. J. (2008). Typology of online social support: A study of bulletin board system in Taiwan. *Mass Communication Research*, *94*, 61-105.]

楊中芳（2010）：〈中庸實踐思維體系探研的初步進展〉。《本土心理學研究》，*34*，3-96。[Yang, Z. F. (2010). Multiplicity of Zhong Yong studies. *Indigenous Psychological Research in Chinese Societies*, *34*, 3-96.]

鄭昭明、陳英孜、卓淑玲、陳學志、梁庚辰（2011，9月）：〈華人情緒類別的加性樹狀結構〉。第七屆華人心理學家學術研討會發表，中央研究院，台北市。[Cheng, C. M., Cheng, J., Cho, S.-L., Chern, H.-C., & Liang, K. C. (2011, September). *The additive-tree structure of Chinese emotions*. Paper presented at the Seventh Chinese Psychologist Conference, Academia Sinica, Taipei.]

Beck, A. T. (1967). *Depression: Clinical, experimental and theoretical aspects*. New York: Hoeber Medical Division, Harper & Row.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561-571.

Choi, I., Koo, M., & Choi, J. A. (2007). Individual differences in analytic versus holistic thinking. *Personality and Social Psychological Bulletin*, *33*, 691-705.

Chung, C., & Pennebaker, J. (2007). The psychological functions of function words. In K. Fiedler (Ed.), *Social communication: Frontiers of social psychology* (pp. 343-359). New York: Psychology Press.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*, 687-693.

Francis, M. E., & Pennebaker, J. W. (1993). *LIWC: Linguistic inquiry and word count*. (Tech. Rep.). Dallas, TX: Southern Methodist University.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*, 3-19.

Graybeal, A., Sexton, J. D., & Pennebaker, J. W. (2002). The role of story-making in disclosure writing: The psychometrics of narrative. *Psychology & Health*, *17*, 571-581.

Gunsch, M. A., Brownlow, S., Haynes, S. E., & Mabe, Z. (2000). Differential linguistic content of various forms

of political advertising. *Journal of Broadcasting & Electronic Media*, *44*, 27-42.

Hartley, J., Pennebaker, J. W., & Fox, C. (2003). Abstracts, introductions and discussions: How far do they differ in style? *Scientometrics*, *57*, 389-398.

Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and Poetry. *Journal of Personality and Social Psychology*, *99*, 549-571.

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*, 39-44.

Lee, C. H., Park, J., & Seo, Y. S. (2006). An analysis of linguistic styles by inferred age in TV dramas. *Psychological Reports*, *99*, 351-356.

Ma, W. Y., & Chen, K. J. (2003a). Introduction to CKIP Chinese word segmentation system for the first international Chinese Word Segmentation Bakeoff. *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*, *17*, 168-171.

Ma, W. Y., & Chen, K. J. (2003b). A bottom-up merging algorithm for Chinese unknown word extraction. *Proceedings of the Second SIGHAN Workshop on Chinese Language Processing*, *17*, 31-38.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*, 862-877.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*, 211-236.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*, 665-675.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, *54*, 741-754.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological Science*, *8*, 162-166.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. New York: Bloomsbury Press.

Pennebaker, J. W., & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, *95*, 274-281.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Operator's manual -- Linguistic inquiry and word count: LIWC2007*. Austin, TX: LIWC.net.

Pennebaker, J. W., & Chung, C. K. (in press). Expressive writing and its links to mental and physical health. In H. S. Friedman (Ed.), *Oxford handbook of health psychology*. New York: Oxford University Press.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.

Pennebaker, J. W., Colder, M., & Sharp, L. K. (1990). Accelerating the coping process. *Journal of Personality and Social Psychology*, *58*, 528-537.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition and Emotion*, *10*, 601-626.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC2001*. Mahwah, NJ: Erlbaum.

Pennebaker, J. W., Kiecolt-Glaser, J., & Glaser, R. (1988). Disclosure of traumas and immune function: Health implications for psychotherapy. *Journal of Consulting and Clinical Psychology*, *56*, 239-245.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, *77*, 1296-1312.

Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, *36*, 271-282.

Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology*, *72*, 863-871.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547-577.

Pyszczynski, T., & Greenberg, J. (1987). Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of reactive depression. *Psychological Bulletin*, *102*, 122-138.

Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*, 1121-1133.

Spencer-Rodgers, J., Peng, K., & Wang, L. (2009). Dialecticism and the co-occurrence of positive and negative emotions across cultures. *Journal of Cross-Cultural Psychology*, *41*, 109-115.

Spencer-Rodgers, J., Williams, M. J., & Peng, K. (2010). Cultural differences in expectations of change and tolerance for contradiction: A decade of empirical research. *Personality and Social Psychology Review*, *14*, 296-312.

Stirman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, *63*, 517-522.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24-54.

Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, *2*, 242-259.

Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, *90*, 288-307.

Tsai, J. L., Miao, F. F., & Seppala, E. (2007). Good feelings in Christianity and Buddhism: Religious differences in ideal affect. *Personality and Social Psychology Bulletin*, *33*, 409-421.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063-1070.

Zimmerman, M., & Coryell, W. (1987). The inventory to diagnose depression, lifetime version. *Acta Psychiatrica Scandinavica*, *75*, 495-499.

06-黃金蘭.indd 200 2012/6/24 上午 10:58:23

The Development of the Chinese Linguistic Inquiry and Word Count Dictionary

Chin-Lan Huang,[1] Cindy K. Chung,[2] Natalie Hui,[3] Yi-Cheng Lin,[4] Yi-Tai Seih,[2] Ben C. P. Lam,[3] Wei-Chuan Chen,[4] Michael H. Bond,[3] and James W. Pennebaker[2]

[1]Department of Humanities and Social Sciences, National Taiwan University of Science and Technology

[2]Department of Psychology, University of Texas at Austin

[3]Department of Psychology, The Chinese University of Hong Kong

[4]Department of Psychology, National Taiwan University

The analysis of natural language opens a window to the exploration of thoughts, feelings, and personalities. To analyze word use, Pennebaker and his colleagues (Pennebaker, 2011) developed a computer software, Linguistic Inquiry and Word Count (LIWC). LIWC reports on the percentage of words in a text file that are in the grammatical, psychological, and content categories of its dictionary, and was created by judges who rated whether each entry belonged in a category. This paper describes the development of the Chinese LIWC dictionary and establishes its reliability and validity. Study 1 involved the translation of terms from the English LIWC dictionary to Chinese, and judges' ratings of whether each entry belonged in a category. Furthermore, some categories unique to the Chinese language were added to the Chinese LIWC dictionary. These resulted in a total of approximately 6,800 words across 30 linguistic categories and 42 psychological categories. In Study 2, we analyzed one hundred texts and their translations using both the English LIWC dictionary and Chinese LIWC dictionary. Fifty texts of varied genres and authors were written in English and translated to Chinese; others were written in Chinese and translated to English. Reliable correlations were found between the English LIWC and Chinese LIWC categories, indicating acceptable equivalence between the two dictionaries. In Study 3, we analyzed 30 bulletin board messages from a site for people suffering from depression, and 30 messages from a site on a control topic (personal experiences with part-time jobs). Similar to what past researchers have found in English, depressed people used more first person singular pronouns, fewer first person plural pronouns, and more negative emotion words than non-depressed people, confirming the validity of the Chinese LIWC. Just as the English LIWC has led to discoveries in the social sciences, the Chinese LIWC now opens new windows to the psychology of Chinese speakers and authors by providing an efficient means by which to analyze open-ended texts. Future research, applications, and revisions of the Chinese LIWC dictionary are discussed.

Keywords: *expressive writing, LIWC (Linguistic Inquiry and Word Count), text analysis*

06-黃金蘭.indd 201 2012/6/24 上午 10:58:23

airiti press®